Aan Mad, Max en Lars
het licht in mijn leven

Ter nagedachtenis aan mijn ouders
die me altijd aangemoedigd hebben
CONTENTS

List of illustrations xi
List of tables xiii
Preface xv
Course outline xxi

Part I  Preliminaries 1

1  Introduction and preliminaries 3
   1.1 Motivation: Real-world networks 4
   1.2 Random graphs and real-world networks 9
   1.3 Random graph models 11
      1.3.1 Erdős-Rényi random graph 11
      1.3.2 Inhomogeneous random graphs 13
      1.3.3 Configuration model 18
      1.3.4 Switching algorithms for uniform random graphs 23
      1.3.5 Preferential attachment models 27
      1.3.6 A Bernoulli preferential attachment model 30
      1.3.7 Universality of random graphs 31
   1.4 Power-laws and their properties 32
   1.5 Notation 33
   1.6 Notes and discussion 35
   1.7 Exercises for Chapter 1 35

2  Local convergence of random graphs 37
   2.1 The metric space of rooted graphs 38
   2.2 Local weak convergence of graphs 39
      2.2.1 Definition of local weak convergence 39
      2.2.2 Criterion for local weak convergence and tightness 41
      2.2.3 Local weak convergence and completeness of the limit 43
      2.2.4 Examples of local weak convergence 44
   2.3 Local convergence of random graphs 45
      2.3.1 Definition of local convergence of random graphs 46
      2.3.2 Criterion for local convergence of random graphs and tightness 48
      2.3.3 Local convergence and completeness of the limit 49
      2.3.4 Local convergence of random regular graphs 50
      2.3.5 Local convergence of Erdős-Rényi random graphs 51
   2.4 Consequences of local weak convergence: local functionals 55
      2.4.1 Local weak convergence and convergence of neighborhoods 55
      2.4.2 Local weak convergence and clustering coefficients 56
      2.4.3 Neighborhoods of edges and degree-degree dependencies 59
      2.4.4 PageRank and local weak convergence 63
   2.5 The giant component is almost local 64
      2.5.1 Asymptotics of the giant 65
      2.5.2 Properties of the giant 68
## Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.5.3</td>
<td>The condition (2.5.7) revisited</td>
<td>70</td>
</tr>
<tr>
<td>2.5.4</td>
<td>The giant in Erdős-Rényi random graphs</td>
<td>72</td>
</tr>
<tr>
<td>2.5.5</td>
<td>Outline of the remainder of this book</td>
<td>77</td>
</tr>
<tr>
<td>2.6</td>
<td>Notes and discussion</td>
<td>78</td>
</tr>
<tr>
<td>2.7</td>
<td>Exercises for Chapter 2</td>
<td>80</td>
</tr>
</tbody>
</table>

### Part II  Connected components in random graphs  

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>Phase transition in general inhomogeneous random graphs</td>
<td>89</td>
</tr>
<tr>
<td>3.1</td>
<td>Motivation</td>
<td>89</td>
</tr>
<tr>
<td>3.2</td>
<td>Definition of the model</td>
<td>91</td>
</tr>
<tr>
<td>3.2.1</td>
<td>Examples of inhomogeneous random graphs</td>
<td>94</td>
</tr>
<tr>
<td>3.3</td>
<td>Degree sequence of inhomogeneous random graphs</td>
<td>96</td>
</tr>
<tr>
<td>3.3.1</td>
<td>Degree sequence of finitely-types case</td>
<td>97</td>
</tr>
<tr>
<td>3.3.2</td>
<td>Finite-type approximations of bounded kernels</td>
<td>97</td>
</tr>
<tr>
<td>3.3.3</td>
<td>Degree sequences of general inhomogeneous random graphs</td>
<td>99</td>
</tr>
<tr>
<td>3.4</td>
<td>Multi-type branching processes</td>
<td>101</td>
</tr>
<tr>
<td>3.4.1</td>
<td>Multi-type branching processes with finitely many types</td>
<td>101</td>
</tr>
<tr>
<td>3.4.2</td>
<td>Survival vs. extinction of multi-type branching processes</td>
<td>102</td>
</tr>
<tr>
<td>3.4.3</td>
<td>Poisson multi-type branching processes</td>
<td>103</td>
</tr>
<tr>
<td>3.5</td>
<td>Local weak convergence for inhomogeneous random graphs</td>
<td>108</td>
</tr>
<tr>
<td>3.5.1</td>
<td>Local weak convergence: finitely many types</td>
<td>108</td>
</tr>
<tr>
<td>3.5.2</td>
<td>Local weak convergence: infinitely many types</td>
<td>112</td>
</tr>
<tr>
<td>3.5.3</td>
<td>Comparison to branching processes</td>
<td>113</td>
</tr>
<tr>
<td>3.5.4</td>
<td>Local weak convergence to unimodular trees for GRG_n(w)</td>
<td>116</td>
</tr>
<tr>
<td>3.6</td>
<td>The phase transition for inhomogeneous random graphs</td>
<td>117</td>
</tr>
<tr>
<td>3.7</td>
<td>Related results for inhomogeneous random graphs</td>
<td>121</td>
</tr>
<tr>
<td>3.8</td>
<td>Notes and discussion</td>
<td>124</td>
</tr>
<tr>
<td>3.9</td>
<td>Exercises for Chapter 3</td>
<td>125</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>The phase transition in the configuration model</td>
<td>129</td>
</tr>
<tr>
<td>4.1</td>
<td>Local weak convergence to unimodular trees for CM_n(d)</td>
<td>130</td>
</tr>
<tr>
<td>4.2</td>
<td>Phase transition in the configuration model</td>
<td>135</td>
</tr>
<tr>
<td>4.2.1</td>
<td>The giant component is almost local</td>
<td>139</td>
</tr>
<tr>
<td>4.2.2</td>
<td>Finding the largest component</td>
<td>140</td>
</tr>
<tr>
<td>4.2.3</td>
<td>Analysis of the algorithm for CM_n(d)</td>
<td>141</td>
</tr>
<tr>
<td>4.2.4</td>
<td>Proof of Theorem 4.4</td>
<td>146</td>
</tr>
<tr>
<td>4.2.5</td>
<td>The giant component of related random graphs</td>
<td>150</td>
</tr>
<tr>
<td>4.3</td>
<td>Connectivity of CM_n(d)</td>
<td>152</td>
</tr>
<tr>
<td>4.4</td>
<td>Related results for the configuration model</td>
<td>158</td>
</tr>
<tr>
<td>4.5</td>
<td>Notes and discussion</td>
<td>161</td>
</tr>
<tr>
<td>4.6</td>
<td>Exercises for Chapter 4</td>
<td>162</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>Connected components in preferential attachment models</td>
<td>167</td>
</tr>
<tr>
<td>5.1</td>
<td>Exchangeable random variables and Pólya urn schemes</td>
<td>167</td>
</tr>
<tr>
<td>5.2</td>
<td>Local weak convergence of preferential attachment models</td>
<td>177</td>
</tr>
<tr>
<td>5.2.1</td>
<td>Local weak convergence of PAMs with fixed number of edges</td>
<td>177</td>
</tr>
<tr>
<td>5.3</td>
<td>Proof local convergence for preferential attachment models</td>
<td>189</td>
</tr>
</tbody>
</table>
5.3.1 Local weak convergence of related models 205
5.3.2 Local weak convergence Bernoulli preferential attachment 205
5.4 Connectivity of preferential attachment models 206
5.4.1 Giant component for Bernoulli preferential attachment 207
5.5 Further results for preferential attachment models 209
5.6 Notes and discussion 209
5.7 Exercises for Chapter 5 210

Part III Small-world properties of random graphs 215

6 Small-world phenomena in inhomogeneous random graphs 219
6.1 Small-world effect in inhomogeneous random graphs 219
6.2 Lower bounds on typical distances in IRGs 223
6.2.1 Logarithmic lower bound distances for finite-variance degrees 223
6.2.2 log log lower bound on distances for infinite-variance degrees 228
6.3 The log log upper bound 233
6.4 Path counting and log upper bound for finite-variance weights 238
6.4.1 Path-counting techniques 239
6.4.2 Logarithmic distance bounds for finite-variance weights 245
6.4.3 Distances for IRGn(κ): Proofs of Theorems 3.16 and 6.1(ii) 251
6.5 Related results on distances for inhomogeneous random graphs 254
6.5.1 The diameter in inhomogeneous random graphs 254
6.5.2 Distance fluctuations for GRGn(w) with finite-variance degrees 256
6.5.3 Distances GRGn(w) for τ = 3 256
6.6 Notes and discussion 256
6.7 Exercises for Chapter 6 258

7 Small-world phenomena in the configuration model 263
7.1 The small-world phenomenon in CMn(d) 264
7.2 Proofs of small-world results CMn(d) 265
7.2.1 Branching process approximation 265
7.2.2 Path-counting techniques 266
7.2.3 A log log upper bound on the diameter core τ ∈ (2, 3) 271
7.3 Branching processes with infinite mean 275
7.4 The diameter of the configuration model 281
7.4.1 The diameter of the configuration model: logarithmic case 281
7.4.2 Diameter of the configuration model for τ ∈ (2, 3): log log case 283
7.5 Related results for the configuration model 288
7.5.1 Distances for infinite-mean degrees 288
7.5.2 Fluctuation of distances for finite-variance degrees 290
7.5.3 Fluctuation of distances for infinite-variance degrees 291
7.6 Notes and discussion 293
7.7 Exercises for Chapter 7 295

8 Small-world phenomena in preferential attachment models 297
8.1 Logarithmic distances in preferential attachment trees 298
8.2 Small-world phenomena in preferential attachment models 307
8.3 Path counting in preferential attachment models 309
8.4 Small-world effect in PAMs: logarithmic lower bounds for $\delta \geq 0$
  8.4.1 Logarithmic lower bounds on distances for $\delta > 0$
  8.4.2 Extension to lower bounds on distances for $\delta = 0$ and $m \geq 2$
8.5 Typical distance in PA-models: log log-lower bound for $\delta < 0$
8.6 Small-world PAMs: logarithmic upper bounds for $\delta \geq 0$
  8.6.1 Logarithmic upper bounds for $\delta > 0$
8.7 Small-world effect in PAMs: loglog upper bounds for $\delta < 0$
  8.7.1 The diameter of the core for $\delta < 0$
  8.7.2 Connecting the periphery to the core for $\delta < 0$
8.8 Diameters in preferential attachment models
8.9 Further results on distances in preferential attachment models
  8.9.1 Distance evolution for $\delta < 0$
  8.9.2 Distances in the Bernoulli PAM
8.10 Notes and discussion
8.11 Exercises for Chapter 8

Part IV Related models and problems

9 Related Models
9.1 Two real-world network examples
  9.1.1 Citation networks
  9.1.2 Terrorist networks
  9.1.3 General real-world network models
9.2 Directed random graphs
  9.2.1 Directed inhomogeneous random graphs
  9.2.2 The directed configuration model
  9.2.3 Directed preferential attachment models
9.3 Random graphs with community structure: global communities
  9.3.1 Stochastic blockmodel
  9.3.2 Degree-corrected stochastic blockmodel
  9.3.3 Configuration models with global communities
  9.3.4 Preferential attachment models with global communities
9.4 Random graphs with community structure: local communities
  9.4.1 Inhomogeneous random graphs with communities
  9.4.2 Configuration models with community structure
  9.4.3 Random intersection graphs
  9.4.4 Exponential random graphs and maximal entropy
9.5 Spatial random graphs
  9.5.1 Small-world model
  9.5.2 Hyperbolic random graphs
  9.5.3 Geometric inhomogeneous random graphs
  9.5.4 Spatial preferential attachment models
  9.5.5 Complex network models on the hypercubic lattice
9.6 Notes and discussion
9.7 Exercises for Chapter 9

Appendix Some facts about the metric space structure of rooted graphs
## Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>References</td>
<td>445</td>
</tr>
<tr>
<td>Index</td>
<td>457</td>
</tr>
<tr>
<td>Todo list</td>
<td>461</td>
</tr>
</tbody>
</table>
## Illustrations

<table>
<thead>
<tr>
<th>Illustration</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1 Degree sequence of the Autonomous Systems graph in 2014</td>
<td>6</td>
</tr>
<tr>
<td>1.2 log-log plot of the degree sequence in the Internet Movie Data base</td>
<td>7</td>
</tr>
<tr>
<td>1.3 The probability mass functions of the in- and out-degree sequence in the WWW from various data sets</td>
<td>7</td>
</tr>
<tr>
<td>1.4 AS and Internet hopcount</td>
<td>8</td>
</tr>
<tr>
<td>1.5 Typical distances in the Internet Movie Data base</td>
<td>9</td>
</tr>
<tr>
<td>1.6 Forward and backward switchings</td>
<td>25</td>
</tr>
<tr>
<td>6.1 A 12-step self-avoiding path connecting vertices $i$ and $j$.</td>
<td>224</td>
</tr>
<tr>
<td>6.2 A 10-step good path connecting $i$ and $j$ and the upper bounds on the weight of its vertices. The height of a vertex is high for vertices with large weights.</td>
<td>229</td>
</tr>
<tr>
<td>6.3 An example of a pair of paths $(\pi, \rho)$ and its corresponding shape.</td>
<td>242</td>
</tr>
<tr>
<td>7.1 Number of AS traversed in hopcount data (blue) compared to the model (purple) with $\tau = 2.25, n = 10,940$.</td>
<td>263</td>
</tr>
<tr>
<td>7.2 Empirical probability mass function of the $\text{dist}_{CM,\alpha}(o_1, o_2)$ for $\tau = 1.8$ and $n = 10^3, 10^4, 10^5$.</td>
<td>289</td>
</tr>
<tr>
<td>9.1 (a) Number of publication per year (logarithmic Y axis). (b)Loglog plot for the in-degree distribution tail in citation networks</td>
<td>359</td>
</tr>
<tr>
<td>9.2 Degree distribution for papers from 1984 over time.</td>
<td>359</td>
</tr>
<tr>
<td>9.3 Time evolution for the number of citations of samples of 20 randomly chosen papers from 1980 for PS and EE, and from 1982 for BT.</td>
<td>360</td>
</tr>
<tr>
<td>9.4 Average degree increment over a 20-years time window for papers published in different years. PS presents an aging effect different from EE and BT, showing that papers in PS receive citations longer than papers in EE and BT.</td>
<td>361</td>
</tr>
<tr>
<td>9.5 Distribution of the age of cited papers for different citing years.</td>
<td>361</td>
</tr>
<tr>
<td>9.6 The WWW according to Broder et al.</td>
<td>364</td>
</tr>
<tr>
<td>9.7 Tail probabilities of the degrees, community sizes $s$ and the inter-community degrees $k$ in real-world networks. a) AMAZON co-purchasing network, b) GOWALLA social network, c) English word relations, d) GOOGLE web graph. Pictures taken from Stegehuis et al. (2016b).</td>
<td>387</td>
</tr>
<tr>
<td>9.8 The relation between the denseness of a community $2c_m/(s^2 - s)$ and the community size $s$ can be approximated by a power law. a) AMAZON co-purchasing network, b) GOWALLA social network, c) English word relations, d) GOOGLE web graph. Pictures taken from Stegehuis et al. (2016b).</td>
<td>388</td>
</tr>
<tr>
<td>9.9 The degree distribution of a household model follows a power law with a smaller exponent than the community size distribution and outside degree distribution</td>
<td>395</td>
</tr>
<tr>
<td>9.10 A hyperbolic embedding of the Internet at Autonomous Systems level by Boguná et al. (2010) (see (Boguná et al., 2010, Figure 3))</td>
<td>408</td>
</tr>
</tbody>
</table>
3.1 The five rows correspond to the following real-life networks: (1) Model of the protein-protein connections in blood of people with Alzheimer-disease. (2) Model of the protein-protein connections in blood of people with Multiple Sclerosis. (3) Network of actors in 2011 in the Internet Movie Data base (IMDb), where actors are linked if they appeared in the same movie. (4) Collaboration network based on DBLP. Two scientist are connected if they ever collaborated on a paper. (5) Network where nodes represent zebras. There is a connection between two zebras if they ever interacted with each other during the observation phase.

4.1 The rows in the above table represent the following real-world networks: The California road network consists of nodes representing intersections or endpoints of roads. In the Facebook network nodes represent the users and the edges Facebook friendships. Hyves was a Dutch social media platform. Nodes represent the users, and an edge between nodes represents a friendship. The arXiv astro-physics network represents authors of papers within the astro-physics section of arXiv. There is an edge between authors if they have ever been co-authors. Models the high voltage power network in western USA. The nodes denote transformers substations, and generators and the edges represent transmission cables. The data set consists of jazz-musicians and there exists a connection if they ever collaborated.
In this book, we study connected components and small-world properties of random graph models for complex networks. This is Volume 2 of a sequel of two books. Volume 1 describes the preliminary topics of random graphs as models for real-world networks. Since 1999, many real-world networks have been investigated. These networks turned out to have rather different properties than classical random graph models, for example in the number of connections that the elements in the network make. As a result, a wealth of new models was invented to capture these properties. Volume 1 studies the models as well as their degree structure. This book summarizes the insights developed in this exciting period that focus on the connected components and small-world properties of the proposed random graph models.

While Volume 1 is intended to be used for a master level course where students have a limited prior knowledge of special topics in probability, Volume 2 describes more involved notions that have been the focus of attention in the past two decades. Volume 2 is intended to be used for a PhD level course, a reading seminar, or for researchers wishing to obtain a consistent and extended overview of the results and methodology that has been developed in this area. Volume 1 includes many of the preliminaries, such as convergence of random variables, probabilistic bounds, coupling, martingales and branching processes, and we frequently refer to these results. The series of Volumes 1 and 2 aims to be self-contained. In Volume 2 we do briefly repeat some of the preliminaries on random graphs, including the introduction of some of the models and the key results on their degree distributions as discussed in great length in Volume 1. In Volume 2, we give the results concerning connected components, connectivity transitions, as well as the small-world nature of the random graph models introduced in Volume 1. We aim to give detailed and complete proofs. When we do not give proofs of our results, we provide extensive pointers to the literature. We further discuss several more recent random graph models that aim to provide for more realistic models for real-world networks, as they incorporate their directed nature, their community structure or their spatial embedding.

The field of random graphs was pioneered in 1959-1960 by Erdős and Rényi (1959; 1960; 1961a; 1961b), in the context of the Probabilistic Method. The initial work by Erdős and Rényi on random graphs has incited a great amount of follow up in the field, initially mainly in the combinatorics community. See the standard references on the subject by Bollobás (2001) and Janson, Luczak and Ruciński (2000) for the state of the art. Erdős and Rényi (1960) give a rather complete picture of the various phase transitions that occur in the Erdős-Rényi random graph. This initial work did not aim to realistically model real-world networks. In the period after 1999, due to the fact that data sets of real-world networks became abundantly available, their structure has attracted enormous attention in mathematics as well as various applied domains. This is for example reflected in the fact that one of the first articles in the field by Albert and Barabási
(2002) has attracted over 33000 citations. One of the main conclusions from this overwhelming body of work is that many real-world networks share two fundamental properties. The first is that they are highly inhomogeneous, in the sense that vertices play rather different roles in the networks. This property is exemplified by the degree structure of the real-world networks obeying power laws: these networks are scale-free. The scale-free nature of real-world networks prompted the community to come up with many novel random graph models that, unlike the Erdős-Rényi random graph, do have power-law degree sequences. This was the key focus in Volume 1, where three models were presented that can have power-law degree sequences.

In this book, we pick up on the trail left in Volume 1, and we now focus on the connectivity structure between vertices. Connectivity can be summarized in two key aspects of real-world networks: the fact that they are highly connected, as exemplified by the fact that they tend to have one giant component containing a large proportion of the vertices (if not all of them), and their small-world nature, quantifying the fact that most pairs of vertices are separated by short paths. We discuss the available methods for these proofs, including path counting techniques, branching process approximations, exchangeable random variables and De Finetti’s theorems. We pay particular attention to a recent technique, called local weak convergence, that makes the statement that random graphs ‘locally look like trees’ precise. This technique is extremely powerful, and we believe that its full potential has not yet been reached.

This book consists of four parts. In Part I, consisting of Chapters 1-2, we repeat some definitions from Volume 1, including the random graph models studied in this book, which are inhomogeneous random graphs, the configuration model and preferential attachment models. We also discuss general topics that are important in random graph theory, such as power-law distributions and their properties. In Chapter 2, we discuss local-weak convergence, a method that plays a central role in the theory of random graphs and in this book. In Part II, consisting of Chapters 3–5, we discuss large connected components in random graph models. In Chapter 3, we further extend the definition of the generalized random graph to general inhomogeneous random graphs. In Chapter 4 we discuss the large connected components in the configuration model, and in Chapter 5, we discuss the connected components in preferential attachment models. In Part III, consisting of Chapters 3–5, we study the small-world nature in random graphs, starting with inhomogeneous random graphs, continuing with the configuration model and ending with the preferential attachment model. In Part IV, consisting of Chapter 9, we study related random graph models and their structure. Along the way, we give many exercises that help the reader to obtain a deeper understanding of the material by working on their solutions. These exercises appear in the last section of each of the chapters, and when applicable, we refer to them at the appropriate place in the text.

I have tried to give as many references to the literature as possible. How-
ever, the number of papers on random graphs is currently exploding. In MathSciNet (see http://www.ams.org/mathscinet), there were, on December 21, 2006, a total of 1,428 papers that contain the phrase ‘random graphs’ in the review text, on September 29, 2008, this number increased to 1614, to 2346 on April 9, 2013, and to 2986 on April 21, 2016, and to 12038 on October 5, 2020. These are merely the papers on the topic in the math community. What is special about random graph theory is that it is extremely multidisciplinary, and many papers using random graphs are currently written in economics, biology, theoretical physics and computer science. For example, in Scopus (see http://www.scopus.com/scopus/home.uri), again on December 21, 2006, there were 5,403 papers that contain the phrase ‘random graph’ in the title, abstract or keywords, on September 29, 2008, this increased to 7,928, to 13,987 on April 9, 2013, to 19,841 on April 21, 2016 and to 30,251 on October 5, 2020. It can be expected that these numbers will continue to increase, rendering it impossible to review all the literature.

In June 2014, I decided to split the preliminary version of this book up into two books. This has several reasons and advantages, particularly since Volume 2 is more tuned towards a research audience, while the first part is more tuned towards an audience of master students with varying backgrounds. The pdf-versions of both Volumes 1 and 2 can be obtained from http://www.win.tue.nl/~rhofstad/NotesRGCN.html

For further results on random graphs, or for solutions to some of the exercises in this book, readers are encouraged to look there. Also, for a more playful approach to networks for a broad audience, including articles, videos, and demos of many of the models treated in this book, we refer all readers to the NetworkPages at http://www.networkpages.nl. The NetworkPages are an interactive website developed by and for all those that are interested in networks. One can find demos for some of the models discussed here, as well as of network algorithms and processes on networks.

This book, as well as Volume 1 of it, would not have been possible without the help and encouragement of many people. I thank Gerard Hooghiemstra for the encouragement to write it, and for using it at Delft University of Technology almost simultaneously while I used it at Eindhoven University of Technology in the Spring of 2006 and again in the Fall of 2008. I particularly thank Gerard for many useful comments, solutions to exercises and suggestions for improvements of the presentation throughout the book. Together with Piet Van Mieghem, we entered the world of random graphs in 2001, and I have tremendously enjoyed exploring this field together with you, as well as with Henri van den Esker, Dmitri Znamenski, Mia Deijfen and Shankar Bhamidi, Johan van Leeuwaarden, Júlia Komjáthy, Nelly Litvak and many others.

I thank Christian Borgs, Jennifer Chayes, Gordon Slade and Joel Spencer for joint work on random graphs that are alike the Erdős-Rényi random graph, but
do have geometry. Special thanks go to Gordon Slade, who has introduced me to
the world of percolation, which is closely linked to the world of random graphs
(see also the classic on percolation by Grimmett (1999)). It is peculiar to see that
two communities work on two so closely related topics with different methods and
even different terminology, and that it has taken such a long time to build bridges
between the subjects. I am very happy that these bridges are now rapidly appear-
ing, and the level of communication between different communities has increased
significantly. I hope that this book helps to further enhance this communication.
Frank den Hollander deserves a special mention. Frank, you have been important
as a driving force throughout my career, and I am very happy now to be working
with you on fascinating random graph problems!

Further I thank Marie Albenque, Gianmarco Bet, Shankar Bhamidi, Finbar
Bogerd, Marko Boon, Francesco Caravenna, Rui Castro, Kota Chisaki, Mia Dei-
jén, Michel Dekking, Henri van den Esker, Lorenzo Federico, Lucas Gerin, Jesse
Goodman, Rajat Hazra, Markus Heydenreich, Frank den Hollander, Yusuke Ide,
Svante Janson, Lancelot James, Martin van Jole, Willeminen Kets, Bas Kleijn,
Júlia Komjáthy, John Lapeyre, Lasse Leskelä, Nelly Litvak, Norio Konno, Ab-
bas Mehrabian, Mislav Mišković, Mirko Moscatelli, Jan Nagel, Sidi harshan Nair,
Alex Olssen, Mariana Olvera-Cravioto, Helena Peña, Rounak Ray, Nathan Ross,
Christoph Schumacher, Karoly Simon, Clara Stegehuis, Dominik Tomecki, Nicola
Turchi, Viktória Vadon, Thomas Vallier, Xiaotin Yu for remarks and ideas that
have improved the content and presentation of these notes substantially. Wouter
Kager has entirely read the February 2007 version of this book, giving many ideas
for improvements of the arguments and of the methodology. Artém Sapozhnikov,
Maren Eckhoff and Gerard Hooghiemstra read and commented the October 2011
version.

I especially thank Dennis Timmers, Eefje van den Dungen, Joop van de Pol and
Rowel Gündlach, who, as, my student assistants, have been a great help in the
development of this book, in making figures, providing solutions to the exercises,
checking the proofs and keeping the references up to date. Maren Eckhoff and
Gerard Hooghiemstra also provided many solutions to the exercises, for which I
am grateful! Sándor Kolumbán and Robert Fitzner helped me to turn all pictures
of real-world networks as well as simulations of network models into a unified
style, a feat that is beyond my \TeX skills. A big thanks for that! Also my thanks
for suggestions and help with figures to Marko Boon, Alessandro Garavaglia,
Dimitri Krioukov, Vincent Kusters, Lourens Touwen, Clara Stegehuis, Piet Van
Mieghem and Yana Volkovich.

This work would not have been possible without the generous support of the
Netherlands Organisation for Scientific Research (NWO) through VIDI grant
639.032.304, VICI grant 639.033.806 and the Gravitation NETWORKS grant
024.002.003.
Preface

Remco van der Hofstad  
Eindhoven
The relation between the chapters in Volumes 1 and 2 of this book is as follows:
Here is some more explanation as well as a possible itinerary of a master or PhD course on random graphs, including both Volumes 1 and 2, in a course outline:

- Start with the introduction to real-world networks in [Volume 1, Chapter 1], which forms the inspiration for what follows. Continue with [Volume 1, Chapter 2], which gives the necessary probabilistic tools used in all later chapters, and pick those topics that your students are not familiar with and that are used in the later chapters that you wish to treat. [Volume 1, Chapter 3] introduces branching processes, and is used in [Volume 1, Chapters 4, 5], as well as in most of Volume 2.

- After these preliminaries, you can start with the classical Erdős-Rényi random graph as covered in [Volume 1, Chapters 4–5]. Here you can choose the level of detail, and decide whether you wish to do the entire phase transition or would rather move on to the random graphs models for complex networks. It is possible to omit [Volume 1, Chapter 5] before moving on.

- Having discussed the Erdős-Rényi random graph, you can make your own choice of topics from the world of random graphs. There are three classes of models for complex networks that are treated in quite some detail in this book. You can choose how much to treat in each of these models. You can either treat few models and discuss many aspects, or instead discuss many models at a less deep level. The introductory chapters about the three models, [Volume 1, Chapter 6] for inhomogeneous random graphs, [Volume 1, Chapter 7] for the configuration model, and [Volume 1, Chapter 8] for preferential attachment models, provide a basic introduction to them, focussing on their degree structure. These introductory chapters need to be read in order to understand the later chapters about these models (particularly the ones in Volume 2). The parts on the different models can be read independently.

- For readers that are interested in more advanced topics, one can either take one of the models and discuss the different chapters in Volume 2 focussing on them. [Volume 2, Chapters 3 and 6] discuss inhomogeneous random graphs, [Volume 2, Chapters 4 and 7] discuss the configuration model, while [Volume 2, Chapters 5 and 8] focus on preferential attachment models. The alternative is to take one of the topics, and work through them in detail. [Volume 2, Part II] discusses the largest connected components or phase transition in our random graph models, while [Volume 2, Part III] treats their small-world nature.

When you have further questions and/or suggestions about course outlines, then feel free to contact me.
Part I

Preliminaries
Chapter 1
INTRODUCTION AND PRELIMINARIES

Abstract
In this chapter, we draw motivation from real-world networks, and formulate random graph models for them. We focus on some of the models that have received the most attention in the literature, namely, the Erdős-Rényi random graph, Inhomogeneous random graphs, the configuration model and preferential attachment models. We also discuss some of their extensions that have the potential to yield more realistic models for real-world networks. We follow van der Hofstad (2017), which we refer to as [Volume 1], both for the motivation as well as for the introduction of the random graph models involved.

Looking back, and ahead
In Volume 1 of this pair of books, we have discussed various models having flexible degree sequences. The generalized random graph and the configuration model give us static flexible models for random graphs with various degree sequences. Preferential attachment models give us a convincing explanation of the abundance of power-law degree sequences in various applications. In [Volume 1, Chapters 6–8], we have focussed on the properties of the degrees of such graphs. However, we have noted in [Volume 1, Chapter 1] that many real-world networks not only have degree sequences that are rather different from the ones of the Erdős-Rényi random graph, also many examples are small worlds and have a giant connected component.

In Chapters 3–8, we shall return to the models discussed in [Volume 1, Chapters 6–8], and focus on their connected components as well as on the distances in these random graph models. Interestingly, a large chunk of the non-rigorous physics literature suggests that the behavior in various different random graph models can be described by only a few essential parameters. The key parameter of each of these models in the power-law degree exponent, and the physics literature predicts that the behavior in random graph models with similar degree sequences is similar. This is an example of the notion of universality, a notion which is central in statistical physics. Despite its importance, there are only few example of universality that can be rigorously proved. In Chapters 3–8, we investigate the level of universality present in random graph models.

We will often refer to Volume 1. When we do, we write [Volume 1, Theorem 2.17] to mean that we refer to Theorem 2.17 in van der Hofstad (2017).

Organisation of this chapter
This chapter is organised as follows. In Section 9.1.3, we discuss real-world networks the inspiration that they provide. In Section 1.2, we then discuss how graph sequences, where the size of the involved graphs tends to infinity, aim at describing large complex networks. In Section 1.3, we recall the definition of several random
Introduction and preliminaries

In Section 1.5, we recall some of the standard notion used in this pair of books. We close this chapter with notes and discussion in Section 1.6, and with exercises in Section 1.7.

1.1 Motivation: Real-world networks

In the past two decades, an enormous research effort has been performed on modeling of various real-world phenomena using networks.

Networks arise in various applications, from the connections between friends in friendship networks, the connectivity of neurons in the brain, to the relations between companies and countries in economics and the hyperlinks between web-pages in the World-Wide web. The advent of the computer era has made many network data sets available, and around 1999-2000, various groups started to investigate network data from an empirical perspective. See Barabási (2002) and Watts (2003) for expository accounts of the discovery of network properties by Barabási, Watts and co-authors. Newman et al. (2006) bundle some of the original papers detailing the empirical findings of real-world networks and the network models invented for them. The introductory book by Newman (2010) lists many of the empirical properties of, and scientific methods for, networks. See also [Volume 1, Chapter 1] for many examples of real-world networks and the empirical findings for them. Here we just give some basics.

Graphs

A graph $G = (V, E)$ consists of a collection $V$ of vertices, also called vertex set, and a collection of edges $E$, often called edge set. The vertices correspond to the objects that we model, the edges indicate some relation between pairs of these objects. In our settings, graphs are usually undirected. Thus, an edge is an unordered pair $\{u, v\} \in E$ indicating that $u$ and $v$ are directly connected. When $G$ is undirected, if $u$ is directly connected to $v$, then also $v$ is directly connected to $u$. Thus, an edge can be seen as a pair of vertices. When dealing with social networks, the vertices represent the individuals in the population, while the edges represent the friendships among them. We mainly deal with finite graphs, and then, for simplicity, we take $V = [n] := \{1, \ldots, n\}$. The degree $d_u$ of a vertex $u$ is equal to the number of edges containing $u$, i.e.,

$$d_u = \#\{v \in V : \{u, v\} \in E\}. \quad (1.1.1)$$

Often, we deal with the degree of a random vertex in $G$. Let $o \in [n]$ be a vertex chosen uniformly at random in $[n]$, then the typical degree is the random variable $D_n$ given by

$$D_n = d_o. \quad (1.1.2)$$

It is not hard to see that the probability mass function of $D_n$ is given by

$$\mathbb{P}(D_n = k) = \frac{1}{n} \sum_{i \in [n]} 1_{\{d_i = k\}}. \quad (1.1.3)$$
Exercise 1.1 asks you to prove (1.1.3).

We next discuss some of the common features that many real-world networks turn out to have, starting with the high variability of the degree distribution:

**Scale-free phenomenon**

The first, maybe quite surprising, fundamental property of many real-world networks is that the number of vertices with degree at least \( k \) decays slowly for large \( k \). This implies that degrees are highly variable, and that, even though the average degree is not so large, there exist vertices with extremely high degree. Often, the tail of the empirical degree distribution seems to fall off as an inverse power of \( k \). This is called a ‘power-law degree sequence’, and resulting graphs often go under the name ‘scale-free graphs’. It is visualized for the AS graph in Figure 1.1, where the degree distribution of the AS graph is plotted on a log-log scale. Thus, we see a plot of \( \log k \mapsto \log n_k \), where \( n_k \) is the number of vertices with degree \( k \). When \( N_k \) is proportional to an inverse power of \( k \), i.e., when, for some normalizing constant \( c_n \) and some exponent \( \tau \),

\[
 n_k \approx c_n k^{-\tau}, \tag{1.1.4}
\]

then

\[
 \log n_k \approx \log c_n - \tau \log k, \tag{1.1.5}
\]

so that the plot of \( \log k \mapsto \log n_k \) is close to a straight line. This is the reason why degree sequences in networks are often depicted in a log-log fashion, rather than in the more customary form of \( k \mapsto n_k \). Here, and in the remainder of this section, we write \( \approx \) to denote an uncontrolled approximation. The power-law exponent \( \tau \) can be estimated by the slope of the line in the log-log plot. Naturally, we must have that

\[
 \sum_k n_k = n < \infty, \tag{1.1.6}
\]

so that it is reasonable to assume that \( \tau > 1 \).

Let us define the degree distribution by \( p_k^{(n)} = n_k/n \), so that \( p_k^{(n)} \) equals the probability that a uniformly chosen vertex has degree \( k \). Then, (1.1.4) can be rephrased as

\[
 p_k^{(n)} \approx c k^{-\tau}, \tag{1.1.7}
\]

where again \( \approx \) denotes an uncontrolled approximation.

Vertices with extremely high degrees go under various names, indicating their importance in the field. They are often called hubs, as the hubs in airport networks. Another name for them is super-spreader, indicating the importance of the high-degree vertices in spreading information, or diseases. The hubs quantify the amount of inhomogeneity in the real-world networks, and a large part
of these notes is centered around rigorously establishing the effect that the high-degree vertices have on various properties of the graphs involved, as well as on the behavior of stochastic processes on them.

For Internet, log-log plots of degree sequences first appeared in a paper by the Faloutsos brothers (1999) (see Figure 1.1 for the degree sequence in the Autonomous Systems graph). Here the power-law exponent is estimated as $\tau \approx 2.15 - 2.20$. Figure 1.2 displays the degree distribution in the Internet Movie Data base (IMDb), in which the vertices are actors and two actors are connected when they have played together in a movie. Figure 1.3 displays the degree-sequence for both the in- as well as the out-degrees in various World-Wide Web data bases.

TO DO 1.1: Refer to recent discussion on power-laws in real-world networks.

![Figure 1.1](image)

Figure 1.1 (a) Log-log plot of the probability mass function of the degree sequence of Autonomous Systems (AS) on April 2014 on a log-log scale from Krioukov et al. (2012) (data courtesy of Dmitri Krioukov).

After the discussion of degrees in graphs, we continue with graph distances.

**Small-world phenomenon**

The first fundamental network property observed in many real-world networks is the fact that typical distances between vertices are small. This is called the ‘small-world’ phenomenon (see e.g. the book by Watts (1999)). In particular, such networks are highly connected: their largest connected component contains a significant proportion of the vertices. Many networks, such as the Internet, even consist of one connected component, since otherwise e-mail messages could not be delivered. For example, in the Internet, IP-packets cannot use more than a threshold of physical links, and if distances in the Internet would be larger than this threshold, then e-mail service would simply break down. Thus, the graph of the Internet has evolved in such a way that typical distances are relatively small,
1.1 Motivation: Real-world networks

![Log-log plot of the degree sequence in the Internet Movie Data base in 2007.](image1)

**Figure 1.2** Log-log plot of the degree sequence in the Internet Movie Data base in 2007.

![The probability mass function of the in- and out-degree sequences in the Berkeley-Stanford and Google competition graph data sets of the WWW in Leskovec et al. (2009).](image2)

**Figure 1.3** The probability mass function of the in- and out-degree sequences in the Berkeley-Stanford and Google competition graph data sets of the WWW in Leskovec et al. (2009). (a) in-degree; (b) out-degree.

even though the Internet itself is rather large. For example, as seen in Figure 1.4(a), the number of Autonomous Systems (AS) traversed by an e-mail data set, sometimes referred to as the AS-count, is typically at most 7. In Figure 1.4(b),
the proportion of routers traversed by an e-mail message between two uniformly chosen routers, referred to as the hopcount, is shown. It shows that the number of routers traversed is at most 27, while the distribution resembles a Poisson probability mass function. Figure 1.5 shows typical distances in the IMDB, where distances are quite small despite the fact that the network contains more than one million vertices.

We can imagine that the small-world nature of real-world networks is significant. Indeed, in small-worlds, news can spread quickly as relatively few people are needed to spread it between two typical individuals. This is quite helpful in Internet, where e-mail messages hop along the edges of the network. At the other side of the spectrum, it also implies that infectious diseases can spread quite fast, as few infections carry it to large parts of a population. This implies that diseases have a larger potential of becoming pandemic, and the fact that human society becomes a ‘smaller world’ due to the more extensive traveling of virtually everyone is a continuous threat to health care workers throughout the population.

Let us continue by introducing graph distances, as displayed in Figures 1.2–1.5, formally. For $u, v \in [n]$ and a graph $G = ([n], E)$, we let the graph distance $\text{dist}_G(u, v)$ between $u$ and $v$ be equal to the minimal number of edges in a path linking $u$ and $v$. When $u$ and $v$ are not in the same connected component, we set $\text{dist}_G(u, v) = \infty$. We are interested in settings where $G$ has a high amount of connectivity, so that many pairs of vertices are connected to one another by short paths. In order to describe how large distances between vertices typically are, we draw $o_1$ and $o_2$ uniformly at random from $[n]$, and we investigate the random variable

$$\text{dist}_G(o_1, o_2).$$ (1.1.8)

The quantity in (1.1.8) is a random variable even for deterministic graphs due to the occurrence of the two, uniformly at randomly, vertices $o_1, o_2 \in [n]$. Fig-
1.2 Random graphs and real-world networks

Figures 1.2–1.5 display the probability mass function of this random variable for some real-world networks. Often, we will consider $\text{dist}_G(o_1, o_2)$ conditionally on $\text{dist}_G(o_1, o_2) < \infty$. This means that we consider the typical number of edges between a uniformly chosen pair of connected vertices. As a result, $\text{dist}_G(o_1, o_2)$ is sometimes referred to as the typical distance.

The nice property of $\text{dist}_G(o_1, o_2)$ is that its distribution tells us something about all possible distances in the graph. An alternative and frequently used measure of distances in a graph is the diameter $\text{diam}(G)$, defined as

$$\text{diam}(G) = \max_{u,v \in [n]} \text{dist}_G(u, v).$$

However, the diameter has several disadvantages. First, in many instances, the diameter is algorithmically more difficult to compute than the typical distances (since one has to measure the distances between all pairs of vertices and maximize over them). Second, it is a number instead of the distribution of a random variable, and therefore contains far less information that the distribution of $\text{dist}_G(o_1, o_2)$. Finally, the diameter is highly sensitive to small changes of the graph. For example, adding a string of connected vertices to a graph may change the diameter dramatically, while it hardly influences the typical distances.

1.2 Random graphs and real-world networks

In this section, we discuss how random graph sequences can be used to model real-world networks. We start by discussing graph sequences:

Graph sequences

Motivated by the previous section, in which empirical evidence was discussed that many real-world networks are scale free and small worlds, we set about the question of how to model them. Since many networks are quite large, mathemat-
Introduction and preliminaries

ically, we model real-world networks by graph sequences \((G_n)_{n \geq 1}\), where \(G_n\) has size \(n\) and we take the limit \(n \to \infty\). Since most real-world networks are such that the average degree remains bounded, we will focus on the sparse regime. In the sparse regime, it is assumed that

\[
\limsup_{n \to \infty} \mathbb{E}[D_n] = \limsup_{n \to \infty} \frac{1}{n} \sum_{i \in [n]} d_i < \infty. \tag{1.2.1}
\]

Furthermore, we aim to study graphs that are asymptotically well behaved. For example, we will often either assume, or prove, that the typical degree distribution converges, i.e., there exists a limiting degree random variable \(D\) such that

\[
D_n \overset{d}{\to} D, \tag{1.2.2}
\]

where \(\overset{d}{\to}\) denotes weak convergence of random variables. Also, we will assume that our graphs are small worlds, which is often translated in the asymptotic sense that there exists a constant \(K\) such that

\[
\lim_{n \to \infty} \mathbb{P}(\text{dist}_G(o_1, o_2) \leq K \log n) = 1. \tag{1.2.3}
\]

In what follows, we will discuss random graph models that share these two features.

There are many more features that one could take into account when modeling real-world networks. See e.g., [Volume 1, Section 1.4] for a slightly expanded discussion of such features. Other features that many networks share, or rather form a way to distinguish them, are the following:

(a) their degree correlations, measuring the extent to which high-degree vertices tend to be connected to high-degree vertices, or rather to low-degree vertices (and vice versa);

(b) their clustering, measuring the extent to which pairs of neighbors of vertices are neighbors themselves as well;

(c) their community structure, measuring the extent to which the networks have more dense connected subparts;

See e.g., the book by Newman (2010) for an extensive discussion of such features, as well as the algorithmic problems that arise from them.

Random graphs as models for real-world networks
Real-world networks tend to be quite complex and unpredictable. This is quite understandable, since connections often arise rather irregularly. We model such irregular behavior by letting connections arise through a random process, thus leading us to study random graphs. By the previous discussion, our graphs will be large and their size will tend to infinity. In such a setting, we can either model the graphs by fixing their size to be large, or rather by letting the graphs grow to infinite size in a consistent manner. We refer to these two settings as static and
1.3 Random graph models

Both are useful viewpoints. Indeed, a static graph is a model for a snapshot of a network at a fixed time, where we do not know how the connections arose in time. Many network data sets are of this form. A dynamic setting, however, may be more appropriate when we know how the network came to be as it is. In the static setting, we can make model assumptions on the degrees so that they are scale free. In the dynamic setting, we can let the evolution of the graph be such that they give rise to power-law degree sequences, so that these settings may provide explanations for the frequent occurrence of power-laws in real-world networks.

Most of the random graph models that have been investigated in the (extensive) literature are caricatures of reality, in the sense that one cannot with dry eyes argue that they describe any real-world network quantitatively correctly. However, these random graph models do provide insight into how any of the above features can influence the global behavior of networks, and thus provide for possible explanations of the empirical properties of real-world networks that are observed. Also, random graph models can be used as null models, where certain aspects of real-world networks are taken into account, while others are not. This gives a qualitative way of investigating the importance of such empirical features in the real world. Often, real-world networks are compared to uniform random graphs with certain specified properties, such as their number of edges or even their degree sequence. We will come back to how to generate random graphs uniformly at random from the collection of all graphs with these properties below.

In the next section, we describe four models of random graphs, three of which are static and one of which is dynamic.

1.3 Random graph models

We start with the most basic and simple random graph model, which has proved to be a source of tremendous inspiration, both for its mathematical beauty, as well as providing a starting point for the analysis of random graphs.

1.3.1 Erdős-Rényi random graph

The Erdős-Rényi random graph is the simplest possible random graph. In it, we make every possible edge between a collection of \( n \) vertices open or closed with equal probability. Thus, Erdős-Rényi random graph has vertex set \( [n] = \{1, \ldots, n\} \), and, denoting the edge between vertices \( s, t \in [n] \) by \( st \), \( st \) is occupied or present with probability \( p \), and vacant or absent otherwise, independently of all the other edges. The parameter \( p \) is called the edge probability. The above random graph is denoted by \( ER_n(p) \).

The model is named after Erdős and Rényi, since they have made profound contributions in the study of this model. See, in particular, Erdős and Rényi (1959, 1960, 1961a,b), where Erdős and Rényi investigate a related model in which a collection of \( m \) edges is chosen uniformly at random from the collection of
possible edges. The model just defined was first introduced though by Gilbert (1959), and was already investigated heuristically by Solomonoff and Rapoport (1951). Informally, when \( m = p \binom{n}{2} \), the two models behave very similarly. We remark in more detail on the relation between these two models at the end of this section. Exercise 1.2 investigates the uniform nature of \( \text{ER}_n(p) \) with \( p = 1/2 \). Alternatively speaking, the null model where we take no properties of the network into account is the \( \text{ER}_n(p) \) with \( p = 1/2 \). This model has expected degree \( (n-1)/2 \), which is quite large. As a result, this model is not sparse at all. Thus, we next make this model sparser by making \( p \) smaller.

Since each edge is occupied with probability \( p \), we obtain that

\[
\mathbb{P}(D_n = k) = \binom{n-1}{k} p^k (1-p)^{n-k} = \mathbb{P}(\text{Bin}(n-1, p) = k), \tag{1.3.1}
\]

where \( \text{Bin}(m, p) \) is a binomial random variable with \( m \) trials and success probability \( p \). Since

\[
\mathbb{E}[D_n] = (n-1)p, \tag{1.3.2}
\]

for this model to be sparse, we need that \( p \) becomes small with \( n \). Thus, we take

\[
p = \frac{\lambda}{n}, \tag{1.3.3}
\]

and study the graph as \( \lambda \) is fixed while \( n \to \infty \). In this regime, we know that

\[
D_n \xrightarrow{d} D, \tag{1.3.4}
\]

where \( D \sim \text{Poi}(\lambda) \). It turns out that this result can be strengthened to the statement that the proportion of vertices with degree \( k \) also converges to the probability mass function of a Poisson random variable (see [Volume 1, Section 5.4]), i.e., for every \( k \geq 0 \),

\[
P(n) \xrightarrow{p} p_k \equiv \frac{e^{-\lambda} \lambda^k}{k!}. \tag{1.3.5}
\]

It is well known that the Poisson distribution has very thin tails, even thinner than any exponential, as you are requested to prove in Exercise 1.3. We conclude that Erdős-Rényi random graph is not a good model for real-world networks with their highly-variable degree distributions. In the next section, we discuss inhomogeneous extensions of Erdős-Rényi random graphs which can have highly-variable degrees.

Before doing so, let us make some useful final remarks about the Erdős-Rényi random graph. Firstly, we can also view it as percolation on the complete graph. Percolation is a paradigmatic model in statistical physics describing random failures in networks (see Grimmett (1999) for an extensive overview of percolation theory focussing on \( \mathbb{Z}^d \)). Secondly, the model described here as the Erdős-Rényi random graph was actually not invented by Erdős and Rényi, but
1.3 Random graph models

rather by Gilbert (1959). Erdős and Rényi (1959, 1960, 1961b), instead, considered the closely related combinatorial setting where a uniform sample of $m$ edges is added to the empty graph. In the latter case, the proportion of edges is $2m/n(n-1) \approx 2m/n^2$, so we should think of $m \approx 2\lambda n$ for a fair comparison. Note that when we condition the total number of edges to be equal to $m$, the law of the Erdős-Rényi random graph is equal to the model where a collection of $m$ uniformly chosen edges is added, explaining the close relation between the two models. Due to the concentration of the total number of edges, we can indeed roughly exchange the binomial model with $p = \lambda/m$ with the combinatorial model with $m = 2\lambda n$. The combinatorial model has the nice feature that it produces a uniform graph from the collection of all graphs with $m$ edges, and thus could serve as a null model for a real-world network in which only the number of edges is fixed.

1.3.2 Inhomogeneous random graphs

In inhomogeneous random graphs, we keep the independence of the edges, but make the edge probabilities different for different edges. A general format for such models is in the seminal work of Bollobás et al. (2007). We will discuss such general inhomogeneous random graphs in Chapter 3 below. We start with one key example, that has attracted the most attention in the literature so far, and is also discussed in great detail in [Volume 1, Chapter 6].

Rank-1 inhomogeneous random graphs

The simplest inhomogeneous random graph models are sometimes referred to as rank-1 models, since the edge probabilities are (close to) products of vertex weights. This means that the expected number of edges between vertices, when viewed as a matrix, is (close to) a rank-1 matrix. We start by discussing one of such models, which is the so-called generalized random graph and was first introduced by Britton et al. (2006).

In the generalized random graph model, the edge probability of the edge between vertices $i$ and $j$, for $i \neq j$, is equal to

$$p_{ij} = P_{ij}^{\text{GRG}} = \frac{w_i w_j}{\ell_n + w_i w_j}, \quad (1.3.6)$$

where $w = (w_i)_{i \in [n]}$ are the vertex weights, and $\ell_n$ is the total weight of all vertices given by

$$\ell_n = \sum_{i \in [n]} w_i. \quad (1.3.7)$$

We denote the resulting graph by $\text{GRG}_n(w)$. In many cases, the vertex weights actually depend on $n$, and it would be more appropriate, but also more cumbersome, to write the weights as $w^{(n)} = (w_i^{(n)})_{i \in [n]}$. To keep notation simple, we refrain from making the dependence on $n$ explicit. A special case of the generalized random graph is when we take $w_i \equiv \frac{n\lambda}{n-\lambda}$, in which case $p_{ij} = \lambda/n$ for all $i, j \in [n]$, so that we retrieve the Erdős-Rényi random graph $\text{ER}_n(\lambda/n)$. 

The generalized random graph GRG\(_n(w)\) is close to many other inhomogeneous random graph models, such as the random graph with given prescribed degrees or Chung-Lu model, where instead
\[
p_{ij} = p_{ij}^{(CL)} = \min(w_iw_j/\ell_n, 1),
\]
and which has been studied intensively by Chung and Lu (2002a,b, 2003, 2006a,b). A further adaptation is the so-called Poissonian random graph or Norros-Reittu model (introduced by Norros and Reittu (2006)), for which
\[
p_{ij} = p_{ij}^{(NR)} = 1 - \exp\left(-w_iw_j/\ell_n\right).
\]
See Janson (2010a) or [Volume 1, Sections 6.7 and 6.8] for conditions under which these random graphs are asymptotically equivalent, meaning that all events have equal asymptotic probabilities.

Naturally, the topology of the generalized random graph depends sensitively upon the choice of the vertex weights \(w = (w_i)_{i \in [n]}\). These vertex weights can be rather general, and we both investigate settings where the weights are deterministic, as well as where they are random. In order to describe the empirical proportions of the weights, we define their empirical distribution function to be
\[
F_n(x) = \frac{1}{n} \sum_{i \in [n]} 1\{w_i \leq x\}, \quad x \geq 0.
\]
We can interpret \(F_n\) as the distribution of the weight of a uniformly chosen vertex in \([n]\) (see Exercise 1.4). We denote the weight of a uniformly chosen vertex \(o\) in \([n]\) by \(W_n = w_o\), so that, by Exercise 1.4, \(W_n\) has distribution function \(F_n\).

The degree distribution can only converge when the vertex weights are sufficiently regular. We often assume that the vertex weights satisfy the following regularity conditions, which turn out to imply convergence of the degree distribution in the generalized random graph:

**Condition 1.1** (Regularity conditions for vertex weights) There exists a distribution function \(F\) such that, as \(n \to \infty\) the following conditions hold:

**(a) Weak convergence of vertex weight.** As \(n \to \infty\),
\[
W_n \overset{d}{\to} W,
\]
where \(W_n\) and \(W\) have distribution functions \(F_n\) and \(F\), respectively. Equivalently, for any \(x\) for which \(x \mapsto F(x)\) is continuous,
\[
\lim_{n \to \infty} F_n(x) = F(x).
\]

**(b) Convergence of average vertex weight.** As \(n \to \infty\),
\[
\mathbb{E}[W_n] \to \mathbb{E}[W],
\]
where \(W_n\) and \(W\) have distribution functions \(F_n\) and \(F\), respectively. Further, we
1.3 Random graph models

assume that $\mathbb{E}[W] > 0$.

(c) Convergence of second moment vertex weight. As $n \to \infty$,

$$\mathbb{E}[W_n^2] \to \mathbb{E}[W^2].$$

(1.3.14)

Condition 1.1(a) guarantees that the weight of a ‘typical’ vertex is close to a random variable $W$ that is independent of $n$. Condition 1.1(b) implies that the average weight of the vertices in $\text{GRG}_n(w)$ converges to the expectation of the limiting weight variable. In turn, this implies that the average degree in $\text{GRG}_n(w)$ converges to the expectation of the limit random variable of the vertex weights. Condition 1.1(c) ensures the convergence of the second moment of the weights to the second moment of the limiting weight variable.

Remark 1.2 (Regularity for random weights) Sometimes we will be interested in cases where the weights of the vertices are random themselves. For example, this arises when the weights $w = (w_i)_{i \in [n]}$ are realizations of i.i.d. random variables. When the weights are random variables themselves, also the function $F_n$ is a random distribution function. Indeed, in this case $F_n$ is the empirical distribution function of the random weights $(w_i)_{i \in [n]}$. We stress that $\mathbb{E}[W_n]$ is then to be interpreted as $\frac{1}{n} \sum_{i \in [n]} w_i$, which is itself random. Therefore, in Condition 1.1, we require random variables to converge, and there are several notions of convergence that may be used. As it turns out, the most convenient notion of convergence is convergence in probability.

Let us now discuss some canonical examples of weight distributions that satisfy the Regularity Condition 1.1.

Weights moderated by a distribution function

Let $F$ be a distribution function for which $F(0) = 0$ and fix

$$w_i = [1 - F]^{-1}(i/n),$$

(1.3.15)

where $[1 - F]^{-1}$ is the generalized inverse function of $1 - F$ defined, for $u \in (0, 1)$, by (recall [Volume 1, (6.2.14)–(6.2.15)])

$$[1 - F]^{-1}(u) = \inf\{x : [1 - F](x) \leq u\}. \quad (1.3.16)$$

For the choice in (1.3.15), we can explicitly compute $F_n$ as (see [Volume 1, (6.2.17)])

$$F_n(x) = \frac{1}{n} \left( \lfloor nF(x) \rfloor + 1 \right) \wedge 1. \quad (1.3.17)$$

It is not hard to see that Condition 1.1(a) holds for $(w_i)_{i \in [n]}$ as in (1.3.15), while Condition 1.1(b) holds when $\mathbb{E}[W] < \infty$ and Condition 1.1(c) when $\mathbb{E}[W^2] < \infty$, as can be concluded from Exercise 1.6.
Independent and identically distributed weights

The generalized random graph can be studied both with deterministic weights as well as with independent and identically distributed (i.i.d.) weights. Since we often deal with ratios of the form $\frac{w_i w_j}{\left(\sum_{k \in [n]} w_k\right)}$, we assume that $\mathbb{P}(w = 0) = 0$ to avoid situations where all weights are zero.

Both models, i.e., with weights $(w_i)_{i \in [n]}$ as in (1.3.15), and with i.i.d. weights $(w_i)_{i \in [n]}$, have their own merits. The great advantage of i.i.d. weights is that the vertices in the resulting graph are, in distribution, the same. More precisely, the vertices are completely exchangeable, like in the Erdős-Rényi random graph $\text{ER}_n(p)$. Unfortunately, when we take the weights to be i.i.d., then in the resulting graph the edges are no longer independent (despite the fact that they are conditionally independent given the weights). In the sequel, we focus on the setting where the weights are prescribed. When the weights are deterministic, this changes nothing, when the weights are i.i.d., this means that we work conditionally on the weights.

Degrees in generalized random graphs

We write $d_i$ for the degree of vertex $i$ in $\text{GRG}_n(w)$. Thus, $d_i$ is given by

$$d_i = \sum_{j \in [n]} X_{ij}, \quad (1.3.18)$$

where $X_{ij}$ is the indicator that the edge $ij$ is occupied. By convention, we set $X_{ij} = X_{ji}$. The random variables $(X_{ij})_{1 \leq i < j \leq n}$ are independent Bernoulli variables with $\mathbb{P}(X_{ij} = 1) = p_{ij}$ as defined in (1.3.6).

For $k \geq 0$, we let

$$P_k^{(n)} = \frac{1}{n} \sum_{i \in [n]} \mathbb{1}_{\{D_i = k\}} \quad (1.3.19)$$

denote the degree sequence of $\text{GRG}_n(w)$. We denote the probability mass function of a mixed Poisson distribution by $p_k$, i.e., for $k \geq 0$,

$$p_k = \mathbb{E}\left[e^{-W} W^k/k!\right], \quad (1.3.20)$$

where $W$ is a random variable having distribution function $F$ from Condition 1.1. The main result concerning the vertex degrees, which is [Volume 1, Theorem 6.10] is as follows:

**Theorem 1.3** (Degree sequence of $\text{GRG}_n(w)$) Assume that Conditions 1.1(a)-(b) hold. Then, for every $\varepsilon > 0$,

$$\mathbb{P}\left(\sum_{k=0}^{\infty} |P_k^{(n)} - p_k| \geq \varepsilon\right) \to 0, \quad (1.3.21)$$

where $(p_k)_{k \geq 0}$ is given by (1.3.20).
1.3 Random graph models

Consequently, with $D_n = d_o$ denoting the degree of a random vertex, we obtain

$$D_n \xrightarrow{d} D,$$

where $\mathbb{P}(D = k) = p_k = \mathbb{E}[e^{-W} W^k]$, as shown in Exercise 1.7.

Recall from Section 9.1.3 that we are interested in scale-free random graphs, i.e., random graphs for which the degree distribution obeys a power law. We see from Theorem 1.3 that this is true precisely when $D$ obeys a power law. This, in turn, occurs precisely when $W$ obeys a power law, i.e., when, for $w$ large,

$$\mathbb{P}(W > w) = c w^{\tau - 1}(1 + o(1)),$$

and then also, for $w$ large,

$$\mathbb{P}(D > w) = c w^{\tau - 1}(1 + o(1)).$$

Generalized random graph conditioned on its degrees

The generalized random graph with its edge probabilities as in (1.3.6) is rather special. Indeed, when we condition on its degree sequence, then the graph has a uniform distribution over the set of all graphs with the same degree sequence. For this, note that $\text{GRG}_n(w)$ can be equivalently encoded by $(X_{ij})_{1 \leq i < j \leq n}$, where $X_{ij}$ is the indicator that the edge $ij$ is occupied. Then, $(X_{ij})_{1 \leq i < j \leq n}$ are independent Bernoulli random variables with edge probabilities as in (1.3.6). By convention, let $X_{ii} = 0$ for every $i \in [n]$, and $X_{ji} = X_{ij}$ for $1 \leq i < j \leq n$ In terms of the variables $X = (X_{ij})_{1 \leq i < j \leq n}$, let $d_i(X) = \sum_{j \in [n]} X_{ij}$ be the degree of vertex $i$. Then, the uniformity is equivalent to the statement that, for each $x = (x_{ij})_{1 \leq i < j \leq n}$ such that $d_i(x) = d_i$ for every $i \in [n]$,

$$\mathbb{P}(X = x \mid d_i(X) = d_i \forall i \in [n]) = \frac{1}{\#\{y : d_i(y) = d_i \forall i \in [n]\}},$$

that is, the distribution is uniform over all graphs with the prescribed degree sequence. This will turn out to be rather convenient, and thus we state it formally here:

**Theorem 1.4** (GRG conditioned on degrees has uniform law) The GRG with edge probabilities $(p_{ij})_{1 \leq i < j \leq n}$ given by

$$p_{ij} = \frac{w_i w_j}{\ell_n + w_i w_j},$$

conditioned on $\{d_i(X) = d_i \forall i \in [n]\}$, is uniform over all graphs with degree sequence $(d_i)_{i \in [n]}$.

**Proof** This is [Volume 1, Theorem 6.15].

In Chapter 3 below, we discuss a far more general setting of inhomogeneous random graphs. The analysis of general inhomogeneous random graphs is substantially more challenging than the rank-1 case. As explained in more detail in the next chapter, this is due to the fact that they are no longer locally described by single-type branching processes, but rather by multitype branching processes.
1.3.3 Configuration model

The configuration model is a model in which the degrees of vertices are fixed beforehand. Such a model is more flexible than the generalized random graph. For example, the generalized random graph always has a positive proportion of vertices of degree 0, 1, 2, etc, as easily follows from Theorem 1.3.

Fix an integer \( n \) that will denote the number of vertices in the random graph. Consider a sequence of degrees \( d = (d_i)_{i \in [n]} \). The aim is to construct an undirected (multi)graph with \( n \) vertices, where vertex \( j \) has degree \( d_j \). Without loss of generality, we assume throughout this chapter that \( d_j \geq 1 \) for all \( j \in [n] \), since when \( d_j = 0 \), vertex \( j \) is isolated and can be removed from the graph.

One possible random graph model is then to take the uniform measure over such undirected and simple graphs. Here, we call a multigraph \textit{simple} when it has no self-loops, and no multiple edges between any pair of vertices. However, the set of undirected simple graphs with \( n \) vertices where vertex \( j \) has degree \( d_j \) may be empty. For example, in order for such a graph to exist, we must assume that the total degree

\[
\ell_n = \sum_{j \in [n]} d_j
\]  

(1.3.27)

is even. We wish to construct a simple graph such that \( d = (d_i)_{i \in [n]} \) are the degrees of the \( n \) vertices. However, even when \( \ell_n = \sum_{j \in [n]} d_j \) is even, this is not always possible.

Since it is not always possible to construct a simple graph with a given degree sequence, instead, we construct a multigraph, that is, a graph possibly having self-loops and multiple edges between pairs of vertices. One way of obtaining such a multigraph with the given degree sequence is to pair the half-edges attached to the different vertices in a uniform way. Two half-edges together form an edge, thus creating the edges in the graph. Let us explain this in more detail.

To construct the multigraph where vertex \( j \) has degree \( d_j \) for all \( j \in [n] \), we have \( n \) separate vertices and incident to vertex \( j \), we have \( d_j \) half-edges. Every half-edge needs to be connected to another half-edge to form an edge, and by forming all edges we build the graph. For this, the half-edges are numbered in an arbitrary order from 1 to \( \ell_n \). We start by randomly connecting the first half-edge with one of the \( \ell_n - 1 \) remaining half-edges. Once paired, two half-edges form a single edge of the multigraph, and the half-edges are removed from the list of half-edges that need to be paired. Hence, a half-edge can be seen as the left or the right half of an edge. We continue the procedure of randomly choosing and pairing the half-edges until all half-edges are connected, and call the resulting graph the \textit{configuration model with degree sequence} \( d \), abbreviated as \( \text{CM}_n(d) \).

A careful reader may worry about the order in which the half-edges are being paired. In fact, this ordering turns out to be completely irrelevant since the random pairing of half-edges is completely \textit{exchangeable}. It can even be done in a \textit{random} fashion, which will be useful when investigating neighborhoods in the
configuration model. See e.g., [Volume 1, Definition 7.5 and Lemma 7.6] for more details on this exchangeability.

Interestingly, one can compute rather explicitly what the distribution of $\text{CM}_n(d)$ is. To do so, note that $\text{CM}_n(d)$ is characterized by the random vector $(X_{ij})_{1 \leq i \leq j \leq n}$, where $X_{ij}$ is the number of edges between vertex $i$ and $j$. Here $X_{ii}$ is the number of self-loops incident to vertex $i$, and $d_i = X_{ii} + \sum_{j \in [n]} X_{ij}$ (1.3.28)

In terms of this notation, and writing $G = (x_{ij})_{i,j \in [n]}$ to denote a multigraph on the vertices $[n]$, $P(\text{CM}_n(d) = G) = \frac{1}{(\ell_n - 1)!} \prod_{i \in [n]} d_i! \prod_{1 \leq i \leq j \leq n} x_{ij}!$. (1.3.29)

See e.g., [Volume 1, Proposition 7.7] for this result. In particular, $P(\text{CM}_n(d) = G)$ is the same for each simple $G$, where $G$ is simple when $x_{ii} = 0$ for every $i \in [n]$ and $x_{ij} \in \{0, 1\}$ for every $1 \leq i < j \leq n$. Thus, the configuration model conditioned on simplicity is a uniform random graph with the prescribed degree distribution. This is quite relevant, as it gives a convenient way to obtain such a uniform graph, which is a highly non-trivial fact.

Interestingly, the configuration model was invented by Bollobás (1980) to study uniform random regular graphs (see also (Bollobás, 2001, Section 2.4)). The introduction was inspired by, and generalized the results in, the work of Bender and Canfield (1978). The original work allowed for a careful computation of the number of regular graphs, using a probabilistic argument. This is the probabilistic method at its best, and also explains the emphasis on the study of the probability for the graph to be simple as we will see below. The configuration model, as well as uniform random graphs with a prescribed degree sequence, were further studied in greater generality by Molloy and Reed (1995, 1998). This extension is quite relevant to us, as the scale-free nature of many real-world applications encourages us to investigate configuration models with power-law degree sequences.

The uniform nature of the configuration model partly explains its popularity, and it has become one of the most highly studied random graph models. It also implies that, conditioned on simplicity, the configuration model is the null model for a real-world network where all the degrees are fixed. It thus allows one to distinguish the relevance of the degree inhomogeneity and other features of the network, such as its community structure, clustering, etc.

As for the $\text{GRG}_n(w)$, we again impose regularity conditions on the degree sequence $d$. In order to state these assumptions, we introduce some notation. We denote the degree of a uniformly chosen vertex $\omega$ in $[n]$ by $D_n = d_\omega$. The random variable $D_n$ has distribution function $F_n$ given by

$$F_n(x) = \frac{1}{n} \sum_{j \in [n]} 1(d_j \leq x),$$ (1.3.30)
which is the empirical distribution of the degrees. We assume that the vertex degrees satisfy the following regularity conditions:

**Condition 1.5** (Regularity conditions for vertex degrees)

(a) **Weak convergence of vertex weight.** There exists a distribution function $F$ such that, as $n \to \infty$,

$$D_n \xrightarrow{d} D,$$

where $D_n$ and $D$ have distribution functions $F_n$ and $F$, respectively.

Equivalently, for any $x$,

$$\lim_{n \to \infty} F_n(x) = F(x).$$

Further, we assume that $F(0) = 0$, i.e., $P(D \geq 1) = 1$.

(b) **Convergence of average vertex degrees.** As $n \to \infty$,

$$\mathbb{E}[D_n] \to \mathbb{E}[D],$$

where $D_n$ and $D$ have distribution functions $F_n$ and $F$ from part (a), respectively.

(c) **Convergence of second moment vertex degrees.** As $n \to \infty$,

$$\mathbb{E}[D_n^2] \to \mathbb{E}[D^2],$$

where again $D_n$ and $D$ have distribution functions $F_n$ and $F$ from part (a), respectively.

The possibility to obtain a non-simple graph is a major disadvantage of the configuration model. There are two ways of dealing with this complication:

(a) **Erased configuration model**

The first way of dealing with multiple edges is to erase the problems. This means that we replace $\text{CM}_n(d) = (X_{ij})_{1 \leq i \leq j \leq n}$ by its erased version $\text{ECM}_n(d) = (X_{ij}^{(er)})_{1 \leq i < j \leq n}$, where $X_{ii}^{(er)} = 0$, while $X_{ij}^{(er)} = 1$ precisely when $X_{ij} \geq 1$. In words, we remove the self-loops and merge all multiple edges to a single edge. Of course, this changes the precise degree distribution. However, [Volume 1, Theorem 7.10] shows that only a small proportion of the edges is erased, so that the erasing does not change the degree distribution. See [Volume 1, Section 7.3] for more details.

Of course, the downside of this approach is that the degrees are changed by the procedure, while we would like to keep the degrees precisely as specified.

Let us describe the degree distribution in the erased configuration model in more detail, to study the effect of the erasure of self-loops and multiple edges. We denote the degrees in the erased configuration model by $D^{(er)} = (D^{(er)}_i)_{i \in [n]}$, so that

$$D^{(er)}_i = d_i - 2s_i - m_i,$$

(1.3.35)
1.3 Random graph models

where \((d_i)_{i \in [n]}\) are the degrees in \(CM_n(d)\), \(s_i = x_{ii}\) is the number of self-loops of vertex \(i\) in \(CM_n(d)\), and

\[
m_i = \sum_{j \neq i} (x_{ij} - 1) \mathbb{1}_{\{x_{ij} \geq 2\}}
\]

is the number of multiple edges removed from \(i\).

Denote the empirical degree sequence \((p_k^{(n)})_{k \geq 1}\) in \(CM_n(d)\) by

\[
p_k^{(n)} = \mathbb{P}(D_n = k) = \frac{1}{n} \sum_{i \in [n]} \mathbb{1}_{\{d_i = k\}}
\]

and denote the related degree sequence in the erased configuration model \((P_k^{(er)})_{k \geq 1}\) by

\[
P_k^{(er)} = \frac{1}{n} \sum_{i \in [n]} \mathbb{1}_{\{D_i^{(er)} = k\}}
\]

From the notation it is clear that \((p_k^{(n)})_{k \geq 1}\) is a deterministic sequence when \(d = (d_i)_{i \in [n]}\) is deterministic, while \((P_k^{(er)})_{k \geq 1}\) is a random sequence, since the erased degrees \((D_i^{(er)})_{i \in [n]}\) form a random vector even when \(d = (d_i)_{i \in [n]}\) is deterministic.

Now we are ready to state the main result concerning the degree sequence of the erased configuration model:

**Theorem 1.6** (Degree sequence of erased configuration model with fixed degrees)

For fixed degrees \(d\) satisfying Conditions 1.5(a)-(b), the degree sequence of the erased configuration model \((P_k^{(er)})_{k \geq 1}\) converges in probability to \((p_k)_{k \geq 1}\). More precisely, for every \(\varepsilon > 0\),

\[
P\left(\sum_{k=1}^{\infty} |P_k^{(er)} - p_k| \geq \varepsilon\right) \to 0,
\]

where \(p_k = \mathbb{P}(D = k)\) as in Condition 1.5(a).

Theorem 1.6 indeed shows that most of the edges are kept in the erasure procedure, see Exercise 1.9.

**(b) Configuration model conditioned on simplicity**

The second solution to the multigraph problem of the configuration model is to throw away the result when it is not simple, and to try again. Therefore, this construction is sometimes called the repeated configuration model (see Britton et al. (2006)). It turns out that, when Conditions 1.5(a)-(c) hold, then (see [Volume 1, Theorem 7.12])

\[
\lim_{n \to \infty} \mathbb{P}(CM_n(d) \text{ is a simple graph}) = e^{-\nu/2 - \nu^2/4},
\]

where \(\nu = \frac{1}{2} \sum_{i \in [n]} (d_i^2 - d_i)\).
where

\[
\nu = \frac{\mathbb{E}[D(D - 1)]}{\mathbb{E}[D]} \tag{1.3.41}
\]

is the expected forward degree. Thus, this is a realistic option when \(\mathbb{E}[D^2] < \infty\). Unfortunately, this is not an option when the degrees obey an asymptotic power law with \(\tau \in (2, 3)\), since then \(\mathbb{E}[D^2] = \infty\). Note that, by (1.3.29), \(\text{CM}_n(d)\) conditioned on simplicity is a uniform random graph with the prescribed degree sequence. We denote this random graph by \(\text{UG}_n(d)\). We will return to the difficulty of generating simple graphs with infinite-variance degrees later in this chapter.

Relation GRG and CM

Since \(\text{CM}_n(d)\) conditioned on simplicity yields a uniform (simple) random graph with these degrees, and by (1.3.25), also \(\text{GRG}_n(w)\) conditioned on its degrees is a uniform (simple) random graph with the given degree distribution, the laws of these random graph models are the same. As a result, one can prove results for \(\text{GRG}_n(w)\) by proving them for \(\text{CM}_n(d)\) under the appropriate degree conditions, and then proving that \(\text{GRG}_n(w)\) satisfies these conditions in probability. See [Volume 1, Section 7.5], where this is worked out in great detail. We summarize the results in Theorem 1.7 below, as it will be frequently convenient to derive results for \(\text{GRG}_n(w)\) through those for appropriate \(\text{CM}_n(d)\)’s.

A further useful result in this direction is that the weight regularity conditions in Conditions 1.1(a)-(c) imply the degree regularity conditions in Conditions 1.5(a)-(c):

**Theorem 1.7** (Regularity conditions weights and degrees) Let \(d_i\) be the degree of vertex \(i\) in \(\text{GRG}_n(w)\), and let \(d = (d_i)_{i \in [n]}\). Then, \(d\) satisfies Conditions 1.5(a)-(b) in probability when \(w\) satisfies Conditions 1.1(a)-(b), where

\[
P(D = k) = \mathbb{E}\left[\frac{W^k}{k!}e^{-W}\right] \tag{1.3.42}
\]

denotes the mixed-Poisson distribution with mixing distribution \(W\) having distribution function \(F\) in Condition 1.1(a). Further, \(d\) satisfies Conditions 1.5(a)-(c) in probability when \(w\) satisfies Conditions 1.1(a)-(c).

**Proof** This is [Volume 1, Theorem 7.19]. The weak convergence in Condition 1.5(a) is Theorem 1.3. \(\square\)

Theorem 1.7 allows us to prove many results for the generalized random graph by first proving them for the configuration model, and then extending them to the generalized random graph. See [Volume 1, Sections 6.6 and 7.5] for more details. This will prove to be a convenient proof strategy to deduce results for the generalized random graph from those for the configuration model that will also be frequently used in this book.
1.3 Random graph models

1.3.4 Switching algorithms for uniform random graphs

So far, we have focused on obtaining a uniform random graph with a prescribed degree sequence by conditioning the configuration model on being simple. As explained above, this does not work so well when the degrees have infinite variance. Another setting where this method fails to deliver is when the average degree is large rather than bounded, so that the graph is no longer sparse in the strict sense. An alternative method to produce a sample from the uniform distribution is by using a switching algorithm. A switching algorithm is a Markov chain on the space of simple graphs, where, in each step, some edges in the graph are rewired. The uniform distribution is the stationary distribution of this Markov chain, so letting the switching algorithm run infinitely long, we obtain a perfect sample from the uniform distribution. Let us now describe in some more detail how this algorithm works. Switching algorithms can also be used rather effectively to compute probabilities of certain events for uniform random graphs with specified degrees, as we explain afterwards. As such, switching methods form an indispensable tool in studying uniform random graphs with prescribed degrees. We start by explaining the basic switching algorithms and its relation to uniform sampling.

The switch Markov chain

The switch Markov chain is a Markov chain on the space of simple graphs with prescribed degrees given by \( d \). Fix a simple graph \( G = ([n], E(G)) \) for which the degree of vertex \( i \) equals \( d_i \) for all \( i \in [n] \). We assume that such a simple exists. In order to describe the dynamics of the switch chain, choose two edges \( \{u, v\} \) and \( \{x, y\} \) uniformly at random from \( E(G) \). The possible switches of these two edges are (1) \( \{u, x\} \) and \( \{v, y\} \); (2) \( \{v, x\} \) and \( \{u, y\} \); and (3) \( \{u, v\} \) and \( \{x, y\} \) (so that no change is made). Choose each of these three options with probability equal to \( \frac{1}{3} \), and write the chosen edges as \( e_1, e_2 \). Accept the switch when the resulting graph with edges \( \{e_1, e_2\} \cup (E(G) \setminus \{\{u, v\}, \{x, y\}\}) \) is simple, and reject the switch otherwise (so that the graph remains unchanged under the dynamics).

It is not very hard to see that the resulting Markov chain isaperiodic and irreducible. Further, the switch chain is doubly stochastic, since it is reversible. As a result, its stationary distribution is the uniform random graph with prescribed degree sequence \( d \) which we denoted by \( UG_n(d) \).

The above method works rather generally, and will, in the limit of infinitely many switches, produce a sample from \( UG_n(d) \). As a result, this chain is the method of choice to produce a sample of \( UG_n(d) \) when the probability of simplicity of the configuration model vanishes. However, it is unclear how often one needs to switch in order for the Markov chain to be sufficiently close to the uniform (and thus stationary) distribution. See Section 1.6 for a discussion of the history of the problem, as well as the available results about its convergence to the stationary distribution.
Switching methods for random graphs with prescribed degrees

Switching algorithms can also be used to prove properties about uniform random graphs with prescribed degrees. Here, we explain how switching can be used to estimate the connection probability between vertices of specific degrees in a uniform random graph. Recall that \( \ell_n = \sum_{i \in [n]} d_i \). We write \( \{u \sim v\} \) for the event that vertex \( u \) is connected to \( v \). In \( UG_n(d) \), there can be at most one edge between two vertices. Then, the edge probabilities for \( UG_n(d) \) are given in the following theorem:

**Theorem 1.8** (Edge probabilities for uniform random graphs with prescribed degrees) Assume that the empirical distribution \( F_n \) of \( d_i \) satisfies

\[
1 - F_n(x) \leq C_F x^{-(\tau - 1)}, \tag{1.3.43}
\]

for some \( C_F > 0 \) and \( \tau \in (2, 3) \). Assume further that \( \ell_n/n \not\to 0 \). Let \( U \) denote a set of unordered pairs of vertices and let \( E_U \) denote the event that \( \{x, y\} \) is an edge for every \( \{x, y\} \in U \). Then, assuming that \( |U| = O(1) \), for every \( \{u, v\} \notin U \),

\[
P(u \sim v \mid E_U) = (1 + o(1)) \frac{(d_u - |U_u|)(d_v - |U_v|)}{\ell_n + (d_u - |U_u|)(d_v - |U_v|)}, \tag{1.3.44}
\]

where \( U_x \) denote the set of pairs in \( U \) that contain \( x \).

**Remark 1.9** (Relation to ECM\(_n(d)\) and GRG\(_n(w)\)) Theorem 1.8 shows that when \( d_u d_v \gg \ell_n \), then

\[
1 - P(u \sim v) = (1 + o(1)) \frac{\ell_n}{d_u d_v}. \tag{1.3.45}
\]

In the erased configuration model, on the other hand,

\[
1 - P(u \sim v) \leq e^{-d_u d_v / \ell_n}. \tag{1.3.46}
\]

Thus, the probability that two high-degree vertices are not connected is much smaller for ECM\(_n(d)\) than for UG\(_n(d)\). Instead, \( P(u \sim v) \approx \frac{d_u d_v}{\ell_n + d_u d_v} \), as it would be for GRG\(_n(w)\) when \( w = d \), which indicated once more that GRG\(_n(w)\) and UG\(_n(d)\) are closely related.

We now proceed to prove Theorem 1.8. We first prove a useful lemma about the number of two-paths starting from a specified vertex.

**Lemma 1.10** (The number of two-paths) Assume that \( d \) satisfies (1.3.43) for some \( C > 0 \) and \( \tau \in (2, 3) \). For any graph \( G \) whose degree sequence is \( d \), the number of two-paths starting from any specified vertex is \( o(n) \).

**Proof** Without loss of generality we may assume that \( d_1 \geq d_2 \geq \cdots \geq d_n \). For every \( 1 \leq i \leq n \), the number of vertices with degree at least \( d_i \) is at least \( i \). By (1.3.43), we then have

\[
C_F n d_i^{1-\tau} \geq n[1 - F_n](d_i) \geq i \tag{1.3.47}
\]
for every \(i \in [n]\). Thus, \(d_i \leq (C_F n/i)^{1/(\tau-1)}\). Then the number of two-paths from any vertex is bounded by 
\[
\sum_{i=1}^{d_1} d_i \leq (C_F n)^{1/(\tau-1)} \sum_{i=1}^{d_1} i^{-1/(\tau-1)} = O\left(n^{1/(\tau-1)}\right) \left(\frac{d_1}{\tau^2}\right) O\left(n^{(2\tau-3)/(\tau-1)^2}\right),
\]
\[= O\left(n^{1/(\tau-1)}\right) \left(\frac{d_1}{\tau^2}\right),
\]
since \(d_1 \leq (C_F n)^{1/(\tau-1)}\). Since \(\tau \in (2, 3)\), the above is \(o(n)\).

**Proof of Theorem 1.8.** To estimate \(Pr(u \sim v \mid E_U)\), we will switch between two classes of graphs \(S\) and \(\bar{S}\). \(S\) consists of graphs where all edges in \(\{u, v\} \cup U\) are present, whereas \(\bar{S}\) consists of all graphs where every \(\{x, y\} \in U\) is present, but \(\{u, v\}\) is not. Note that 
\[
Pr(u \sim v \mid E_U) = \frac{|S|}{|S| + |\bar{S}|} = \frac{1}{1 + |S|/|\bar{S}|},
\]
In order to estimate the ratio \(|\bar{S}|/|S|\), we will define an operation called a *forward switching* which converts a graph in \(G \in S\) to a graph \(G' \in \bar{S}\). The reverse operation converting \(G'\) to \(G\) is called a *backward switching*. Then we estimate \(|\bar{S}|/|S|\) by counting the number of forward switchings that can be applied to a graph \(G \in S\), and the number of backward switchings that can be applied to a graph \(G' \in \bar{S}\). Since we wish to have control on whether \(\{u, v\}\) is present or not, we change the switching considerably compared to the switching discussed above.

The forward switching is defined by choosing two edges and specifying their ends as \(\{x, a\}\) and \(\{y, b\}\). The choice must satisfy the following constraints:

1. None of \(\{u, x\}, \{v, y\}\), or \(\{a, b\}\) is an edge;
2. \(\{x, a\}, \{y, b\} \notin U\);
3. All of \(u, v, x, y, a,\) and \(b\) must be distinct except that \(x = y\) is permitted.

Given a valid choice, the forward switching replaces the three edges \(\{u, v\}, \{x, a\}\), and \(\{y, b\}\) by \(\{u, x\}, \{v, y\}\), and \(\{a, b\}\), while ensuring that the graph after switching is simple. Note that the forward switching preserves the degree sequence, and converts a graph in \(S\) to a graph in \(\bar{S}\). The inverse operation of a forward switching is called a backward switching. See Figure 1.6 for an illustration.
Next, we estimate the number of ways to perform a forward switching to a graph $G$ in $S$, denoted by $f(G)$, and the number of ways to perform a backward switching to a graph $G'$ in $S$, denoted by $b(G)$. The number of total switchings between $S$ and $\bar{S}$ is equal to $|S|E[f(G)] = |\bar{S}|E[b(G')]$, where the expectation is over a uniformly random $G \in S$ and $G' \in \bar{S}$ respectively. Consequently,

$$\frac{|\bar{S}|}{|S|} = \frac{E[f(G)]}{E[b(G')]}.$$  \hspace{1cm} (1.3.50)

Given an arbitrary graph $G \in S$, the number of ways of carrying out a forward switching is at most $\ell_n^2$, since there are at most $\ell_n$ ways to choose $\{x, a\}$, and at most $\ell_n$ ways to choose $\{y, b\}$. Note that choosing $\{x, a\}$ for the first edge and $\{y, b\}$ for the second edge results in a different switching than vice versa. To find a lower bound on the number of ways of performing a forward switching, we subtract from $L_n^2$ an upper bound on the number of invalid choices for $\{x, a\}$ and $\{y, b\}$. These can be summarized as follows:

(a) At least one of $\{u, x\}, \{a, b\}, \{v, y\}$ is an edge,
(b) At least one of $\{x, a\}$ or $\{y, b\}$ is in $U$,
(c) Any vertex overlap other than $x = y$ (i.e. if one of $a$ or $b$ is equal to one of $x$ or $y$, or if $a = b$, or if one of $u$ or $v$ are one of $\{a, b, x, y\}$).

To find an upper bound for (a), note that any choice in case (a) must involve a single edge, and a two-path starting from a specified vertex. By Lemma 1.10, the number of choices for (a) then is upper bounded by $3 \cdot o(\ell_n) \cdot \ell_n = o(\ell_n^2)$. The number of choices for case (b) is $O(\ell_n)$ as $|U| = O(1)$, and there are at most $L_n$ ways to choose the other edge which is not restricted to be in $U$. To bound the number of choices for (c), we investigate each case:

(C1) $a$ or $b$ is equal to $x$ or $y$; or $a = b$. In this case, $x, y, a, b$ forms a two-path. Thus, there are at most $5 \cdot n \cdot o(\ell_n) = o(\ell_n^2)$ choices (noting that $n = O(\ell_n)$), where $n$ is the number of ways to choose a vertex, and $o(\ell_n)$ bounds the number of two-paths starting from this specified vertex;

(C2) one of $u$ and $v$ is one of $\{a, b, x, y\}$. In this case, there is one two-path starting from $u$ or $v$, and a single edge. Thus, there are at most $8 \cdot \ell_n d_{\text{max}} = o(\ell_n^2)$ choices, where $d_{\text{max}}$ bounds the number of ways to choose a vertex adjacent to $u$ or $v$ and $\ell_n$ bounds the number of ways to choose a single edge.

Thus, the number of invalid choices for $\{x, a\}$ and $\{y, b\}$ is $o(\ell_n^2)$, so that the number of forward switchings which can be applied to any $G \in S$ is $(1 + o(1))\ell_n^2$. Thus,

$$E[f(G)] = \ell_n^2(1 + o(1)).$$  \hspace{1cm} (1.3.51)

Given a graph $G' \in \bar{S}$, consider the backward switchings that can be applied to $G'$. There are at most $\ell_n(d_u - |U_u|)(d_v - |U_v|)$ ways to do the backward switching, since we are choosing an edge which is adjacent to $u$ but not in $U$, an edge which is adjacent to $v$ but not in $U$, and another “oriented” edge $\{a, b\}$ (oriented in the
sense that each edge has two ways to specify its end vertices as $a$ and $b$). For a lower bound, we consider the following forbidden choices:

(a') at least one of $\{x, a\}$ or $\{y, b\}$ is an edge,
(b') $\{a, b\} \in U$,
(c') any vertices overlap other than $x = y$ (i.e., when $\{a, b\} \cap \{u, v, x, y\} \neq \emptyset$).

For (a'), suppose that $\{x, a\}$ is present, giving the two-path $\{x, a\}, \{a, b\}$ in $G'$. There are at most $(d_u - |U_u|)(d_v - |U_v|)$ ways to choose $x$ and $y$. Given any choice for $x$ and $y$, there are at most $o(\ell_n)$ ways to choose a two-path starting from $x$, and hence $o(\ell_n)$ ways to choose $a, b$. Thus, the total number of choices is at most $o((d_u - |U_u|)(d_v - |U_v|)) n^{-1}$. The case that $\{y, b\}$ is an edge is symmetric.

For (b'), there are $O(1)$ choices for choosing $\{a, b\}$ since $|U| = O(1)$, and at most $(d_u - |U_u|)(d_v - |U_v|)$ choices $x$ and $y$. Thus, the number of choices for case (b') is $O((d_u - |U_u|)(d_v - |U_v|)) = o((d_u - |U_u|)(d_v - |U_v|)) n^{-1}$. For (c'), the case that $a$ or $b$ is equal to $x$ or $y$ corresponds to a two-path starting from $u$ or $v$ together with a single edge from $u$ or $v$. Since $o(\ell_n)$ bounds the number of two-paths starting from $u$ or $v$ and $d_u - |U_u| + d_v - |U_v|$ bounds the number of ways to choose the single edge, there are $o(\ell_n(d_v - |U_v|)) + o(L_n(d_u - |U_u|))$ total choices. If $a$ or $b$ is equal to $u$ or $v$, there are $(d_u - |U_u|)(d_v - |U_v|)$ ways to choose $x$ and $y$, and at most $d_u + d_v$ ways to choose the last vertex as a neighbor of $u$ or $v$. Thus, there are $O((d_u - |U_u|)(d_v - |U_v|)) d_{\text{max}} = o((d_u - |U_u|)(d_v - |U_v|)) \ell_n$ total choices, since $d_{\text{max}} = O(n^{1/(\gamma - 1)}) = o(n) = o(L_n)$. This concludes that the number of backward switchings that can be applied to any graph $G' \in S'$ is $(d_u - |U_u|)(d_v - |U_v|) \ell_n(1 + o(1))$, so that also

$$E[b(G')] = (d_u - |U_u|)(d_v - |U_v|) \ell_n(1 + o(1)).$$ (1.3.52)

Combining (1.3.50), (1.3.51) and (1.3.52) results in

$$|\bar{S}|/|S| = (1 + o(1)) \frac{\ell_n^2}{(d_u - |U_u|)(d_v - |U_v|) \ell_n},$$ (1.3.53)

and thus (1.3.49) yields

$$P(u \sim v \mid E_* \cup U) = \frac{1}{1 + |\bar{S}|/|S|} = (1 + o(1)) \frac{(d_u - |U_u|)(d_v - |U_v|)}{\ell_n + (d_u - |U_u|)(d_v - |U_v|)}. \quad (1.3.54)$$

1.3.5 Preferential attachment models

Most networks grow in time. Preferential attachment models describe growing networks, where the numbers of edges and vertices grow linearly with time. Preferential attachment models were first introduced by Barabási and Albert (1999), whose model we will generalize. Bollobás, Riordan, Spencer and Tusnády (2001) studied the model by Barabási and Albert (1999), and later many other papers followed on this, and related, models. See [Volume 1, Chapter 8] for details. Here we give a brief introduction.
The model that we investigate produces a graph sequence that we denote by \((PA_t^{(m,\delta)})_{t \geq 1}\) and which, for every time \(t\), yields a graph of \(t\) vertices and \(mt\) edges for some \(m = 1, 2, \ldots\) We start by defining the model for \(m = 1\) when the graph consists of a collection of trees. In this case, \(PA_1^{(1,\delta)}\) consists of a single vertex with a single self-loop. We denote the vertices of \(PA_t^{(1,\delta)}\) by \(v_1^{(1)}, \ldots, v_t^{(1)}\). We denote the degree of vertex \(v_i^{(1)}\) in \(PA_t^{(1,\delta)}\) by \(D_i(t)\), where, by convention, a self-loop increases the degree by 2.

We next describe the evolution of the graph. Conditionally on \(PA_t^{(1,\delta)}\), the growth rule to obtain \(PA_{t+1}^{(1,\delta)}\) is as follows. We add a single vertex \(v_{t+1}^{(1)}\) having a single edge. This edge is connected to a second end point, which is equal to \(v_i^{(1)}\) with probability \((1 + \delta)/(t(2 + \delta) + (1 + \delta))\), and to vertex \(v_j^{(1)}\) in \(PA_t^{(1,\delta)}\) with probability \((D_i(t) + \delta)/(t(2 + \delta) + (1 + \delta))\) for each \(i \in [t]\), where \(\delta \geq -1\) is a parameter of the model. Thus,

\[
\mathbb{P}(v_{t+1}^{(1)} \rightarrow v_i^{(1)} \mid PA_t^{(1,\delta)}) = \begin{cases} 
\frac{1 + \delta}{t(2 + \delta) + (1 + \delta)} & \text{for } i = t + 1, \\
\frac{D_i(t) + \delta}{t(2 + \delta) + (1 + \delta)} & \text{for } i \in [t].
\end{cases} \tag{1.3.55}
\]

The above preferential attachment mechanism is called affine, since the attachment probabilities in (1.3.55) depend in an affine way on the degrees of the random graph \(PA_t^{(1,\delta)}\).

The model with \(m > 1\) is defined in terms of the model for \(m = 1\) as follows. Fix \(\delta \geq -m\). We start with \(PA_{mt}^{(1,\delta/m)}\), and denote the vertices in \(PA_{mt}^{(1,\delta/m)}\) by \(v_1^{(1)}, \ldots, v_{mt}^{(1)}\). Then we identify or collapse the \(m\) vertices \(v_1^{(1)}, \ldots, v_{mt}^{(1)}\) in \(PA_{mt}^{(1,\delta/m)}\) to become vertex \(v_1^{(m)}\) in \(PA_t^{(m,\delta)}\). In doing so, we let all the edges that are incident to any of the vertices in \(v_1^{(1)}, \ldots, v_{mt}^{(1)}\) be incident to the new vertex \(v_1^{(m)}\) in \(PA_t^{(m,\delta)}\). Then, we collapse the \(m\) vertices \(v_{m+1}^{(1)}, \ldots, v_{2m}^{(1)}\) in \(PA_{mt}^{(1,\delta/m)}\) to become vertex \(v_2^{(m)}\) in \(PA_t^{(m,\delta)}\), etc. More generally, we collapse the \(m\) vertices \(v_{(j-1)m+1}^{(1)}, \ldots, v_{jm}^{(1)}\) in \(PA_{mt}^{(1,\delta/m)}\) to become vertex \(v_j^{(m)}\) in \(PA_t^{(m,\delta)}\). This defines the model for general \(m \geq 1\). The resulting graph \(PA_t^{(m,\delta)}\) is a multigraph with precisely \(t\) vertices and \(mt\) edges, so that the total degree is equal to \(2mt\). The original model by Barabási and Albert (1999) focused on the case \(\delta = 0\) only, which is sometimes called the proportional model. The inclusion of the extra parameter \(\delta > -1\) is relevant though, as we will see later.

The preferential attachment model \((PA_t^{(m,\delta)})_{t \geq 1}\) is increasing in time, in the sense that vertices and edges, once they have appeared, remain there forever. Thus, the degrees are monotonically increasing in time. Moreover, vertices with a high degree have a higher chance of attracting further edges of later vertices. Therefore, the model is sometimes called the rich-get-richer model. It is not hard to see that \(D_i(t) \xrightarrow{a.s.} \infty\) (see Exercise 1.10). As a result, one could also call the preferential attachment model the old-get-richer model, which may be more appropriate.

Let us continue to discuss the degree structure in \((PA_t^{(m,\delta)})_{t \geq 1}\).
Degrees of fixed vertices

We start by investigating the degrees of fixed vertices as \( t \to \infty \), i.e., we study \( D_i(t) \) for fixed \( i \) as \( t \to \infty \). To formulate our results, we define the Gamma-function \( t \mapsto \Gamma(t) \) for \( t > 0 \) by

\[
\Gamma(t) = \int_0^\infty x^{t-1}e^{-x}dx. \tag{1.3.56}
\]

The following theorem describes the evolution of the degree of fixed vertices (see [Volume 1, Theorem 8.2 and (8.3.11)]):

**Theorem 1.11** (Degrees of fixed vertices) \( \text{Fix } m \geq 1 \) and \( \delta > -m \). Then, \( D_i(t)/t^{1/(2+\delta/m)} \) converges almost surely to a random variable \( \xi_i \) as \( t \to \infty \).

It turns out that also \( t^{-1/(2+\delta/m)} \max_{i \in [t]} D_i(t) \overset{a.s.}{\to} M \) for some limiting positive and finite random variable \( M \) (see [Volume 1, Section 8.7]). In analogy to i.i.d. random variables, the fact that \( t^{-1/(2+\delta/m)} \max_{i \in [t]} D_i(t) \overset{a.s.}{\to} M \) suggests that the degree of a random vertex satisfies a power law with power-law exponent \( \tau = 3 + \tau/m \), and that is our next item on the agenda.

The degree sequence of the preferential attachment model

The main result in this section establishes the scale-free nature of preferential attachment graphs. In order to state it, we need some notation. We write

\[
P_k(t) = \frac{1}{t} \sum_{i=1}^{t} \mathbb{1}_{\{D_i(t) = k\}} \tag{1.3.57}
\]

for the (random) proportion of vertices with degree \( k \) at time \( t \). For \( m \geq 1 \) and \( \delta > -m \), we define \( (p_k)_{k \geq 0} \) by \( p_k = 0 \) for \( k = 0, \ldots, m-1 \) and, for \( k \geq m \),

\[
p_k = \frac{(2 + \delta/m) \Gamma(k+\delta)\Gamma(m+2+\delta+\delta/m)}{\Gamma(m+\delta)\Gamma(k+3+\delta+\delta/m)} \tag{1.3.58}
\]

It turns out that \( (p_k)_{k \geq 0} \) is a probability mass function (see [Volume 1, Section 8.3]). The probability mass function \( (p_k)_{k \geq 0} \) arises as the limiting degree distribution for \( \text{PA}^{(m,\delta)}_t \), as shown in the following theorem:

**Theorem 1.12** (Degree sequence in preferential attachment model) \( \text{Fix } m \geq 1 \) and \( \delta > -m \). There exists a constant \( C = C(m,\delta) > 0 \) such that, as \( t \to \infty \),

\[
P \left( \max_k |P_k(t) - p_k| \geq C \sqrt{\frac{\log t}{t}} \right) = o(1). \tag{1.3.59}
\]

We next investigate the scale-free properties of \( (p_k)_{k \geq 0} \) by investigating the asymptotics of \( p_k \) for \( k \) large. By (1.3.58) and Stirling’s formula, as \( k \to \infty \),

\[
p_k = c_{m,\delta}k^{-\tau}(1 + O(1/k)), \tag{1.3.60}
\]
where
\[
\tau = 3 + \frac{\delta}{m} > 2,
\quad \text{and} \quad c_{m,\delta} = \frac{(2 + \frac{\delta}{m}) \Gamma(m + 2 + \delta + \frac{\delta}{m})}{\Gamma(m + \delta)}.
\]
(1.3.61)

Therefore, by Theorem 1.12 and (1.3.60), the asymptotic degree sequence of \(PA_t^{(m,\delta)}\) is close to a power law with exponent \(\tau = 3 + \frac{\delta}{m}\). We note that any exponent \(\tau > 2\) is possible by choosing \(\delta > -m\) and \(m \geq 1\) appropriately.

Extensions to the preferential attachment rule
In this book, we also sometimes investigate the related \(PA_t^{(m,\delta)}(b)\) model, in which the self-loops for \(m = 1\) in (1.3.55) are not allowed, so that
\[
P(v^{(1)}_{t+1} \rightarrow v^{(1)}_i \mid PA_t^{(m,\delta)}(b)) = \frac{D_i(t) + \delta}{t(2 + \delta)}
\quad \text{for} \quad i \in [t].
\]
(1.3.62)
The model for \(m \geq 2\) is again defined in terms of the model \((PA_t^{(1,\delta)}(b))_{t \geq 1}\) for \(m = 1\) by collapsing blocks of \(m\) vertices. The advantage of \((PA_t^{(m,\delta)}(b))_{t \geq 1}\) compared to \((PA_t^{(m,\delta)})_{t \geq 1}\) is that \((PA_t^{(m,\delta)}(b))_{t \geq 1}\) is naturally connected, while \((PA_t^{(m,\delta)})_{t \geq 1}\) may not be.

Another adaptation of the preferential attachment rule is when no self-loops are ever allowed while the degrees are updated when the \(m\) edges incident to the new vertex is being attached. We denote this model by \((PA_t^{(m,\delta)}(c))_{t \geq 1}\). In this case, the model for \(m = 1\) is the same as \((PA_t^{(1,\delta)}(b))_{t \geq 1}\), while for \(m \geq 2\) and \(j \in \{0, \ldots, m - 1\}\), we attach the \((j + 1)\text{st}\) edge of vertex \(v^{(m)}_{t+1}\) to vertex \(v^{(m)}_i\) for \(i \in [t]\) with probability
\[
P(v^{(m)}_{t+1,j+1} \rightarrow v^{(m)}_i \mid PA_t^{(m,\delta)}(c)) = \frac{D_i(t, j) + \delta}{t(2m + \delta)}
\quad \text{for} \quad i \in [t].
\]
(1.3.63)
Here, \(D_i(t, j)\) is the degree of vertex \(v^{(m)}_i\) after the connection of the edges incident to the first \(t\) vertices, as well as the first \(j\) edges incident to vertex \(v^{(m)}_{t+1}\). Many other adaptations are possible, and have been investigated in the literature, such as settings where the \(m\) edges incident to \(v^{(m)}_{t+1}\) are independently connected as in (1.3.63) when \(j = 0\), but we refrain from discussing these. It is not hard to verify that Theorem 1.12 remains to hold for all these adaptations, which explains why authors have often opted for the version of the model that is most convenient for them. On the other hand, in Theorem 1.11 there will be minor adaptations, particularly since the limiting random variables \((\xi_i)_{i \geq 1}\) do depend on the precise model.

1.3.6 A Bernoulli preferential attachment model
In this section, we discuss a model that is quite a bit different from the other preferential attachment models discussed above. The main difference is that in this model, the number of edges is not fixed, but instead there is much more independence in the edge attachments. A preferential attachment models with
conditionally independent edges is investigated by Dereich and Mörters (2009, 2011, 2013). We call this model the Bernoulli preferential attachment model, as the attachment indicators are all conditionally independent Bernoulli variables. Let us now give the details.

Fix a preferential attachment function $f : \mathbb{N}_0 \mapsto (0, \infty)$. Then, the graph evolves as follows. Start with $	ext{BPA}_1^{(f)}$ being a graph containing one vertex $v_1$ and no edges. At each time $t \geq 2$, we add a vertex $v_t$. Conditionally on $G(t-1)$, and independently for every $i \in [t-1]$, we connect this vertex to $i$ by a directed edge with probability

$$\frac{f(D_i(t-1))}{t-1},$$

(1.3.64)

where $D_i(t-1)$ is the in-degree of vertex $i$ at time $t-1$. This creates the random graph $\text{BPA}_t^{(f)}$. Note that the number of edges in the random graph process $(\text{BPA}_t^{(f)})_{t \geq 1}$ is not fixed, and equal a random variable. In particular, it makes a difference whether we use the in-degree in (1.3.64).

We consider functions $f : \mathbb{N} \mapsto (0, \infty)$ that satisfy that $f(k+1) - f(k) < 1$ for every $k \geq 0$. Under this assumption and when $f(0) \leq 1$, Mörters and Dereich show that the empirical degree sequence converges as $t \to \infty$, i.e.,

$$P_k(t) \equiv \frac{1}{t} \sum_{i \in [t]} 1_{\{D_i(t) = k\}} \xrightarrow{p} p_k,$$

where

$$p_k = \frac{1}{1+f(k)} \prod_{l=0}^{k-1} \frac{f(l)}{1+f(l)}.$$

(1.3.65)

In particular, $\log(1/p_k)/\log(k) \to 1+1/\gamma$ when $f(k)/k \to \gamma \in (0, 1)$ (see Exercise 1.12). Remarkably, when $f(k) = \gamma k + \beta$, the power-law exponent of the degree distribution does not depend on $\beta$. The restriction that $f(k+1) - f(k) < 1$ is needed to prevent the degrees from exploding. Further, $\log(1/p_k) \sim k^{1-\alpha}/(\gamma(1-\alpha))$ when $f(k) \sim \gamma k^\alpha$ for some $\alpha \in (0, 1)$ (see Exercise 1.13). Interestingly, Mörters and Dereich also show that when $\sum_{k \geq 1} 1/f(k)^2 < \infty$, then there exists a persistent hub, i.e., a vertex that has maximal degree for all but finitely many times. When $\sum_{k \geq 1} 1/f(k)^2 = \infty$, this does not happen.

1.3.7 Universality of random graphs

There are tons of other graph topologies where one can expect similar results as in the random graphs discussed above. We will discuss a few related models in Chapter 9 below, where we include several aspects that are relevant in practice, such as directed graphs, adding community structure to random graphs, and geometry. The random graph models that we investigate are inhomogeneous, and one can expect that the results depend sensitively on the amount of inhomogeneity present. This is reflected in the results that we will prove, where the precise asymptotics is different when the vertices have heavy-tailed degrees rather than light-tailed degrees. However, interestingly, what is ‘heavy tailed’ and what is ‘light tailed’ depends on the precise model at hand. Often, as we will see, the distinction depends on how many moments the degree distribution has.
We have proposed many random graph models for real-world networks. Since these models are aiming at describing similar real-world networks, one would hope that they also give similar answers. Indeed, for a real-world network with power-law degree sequences, we could model its static structure by the configuration model with the same degree sequence, and its dynamical properties by the preferential attachment model with similar scale-free degrees. How to interpret the modeling when these attempts give completely different predictions?

*Universality* is the phrase physicists use when different models display similar behavior. Models that show similar behavior are then in the same *universality class*. Enormous effort is going into decided whether various random graph models are in the same universality class, or rather in different ones, and why. We will see that for some features the degree distribution decides the universality class for a wide range of models, as one might possibly hope. This also explains why the degree distribution plays such a dominant role in the investigation of random graphs. Let us discuss some related graphs that have attracted substantial attention in the literature as models for networks. In Chapter 9, we discuss several related random graph models that have attracted the attention in the literature.

### 1.4 Power-laws and their properties

In this text, we frequently deal with random variables having an (asymptotic) power-law distribution. For such random variables, we often need to investigate truncated moments. We study two of such truncated moment bounds here. We start with the tail of the mean:

**Lemma 1.13 (Truncated moments)** Let $X$ be a non-negative random variable whose distribution function satisfies that for every $x \geq 1$,

\[
1 - F_X(x) \leq C_x x^{-(\tau-1)}. \tag{1.4.1}
\]

Then, there exists a constant $C = C_x(a)$ such that, for $a < \tau - 1$ and all $\ell \geq 1$,

\[
\mathbb{E}[X^a \mathbb{1}_{\{X > \ell\}}] \leq C \ell^{a-(\tau-1)}, \tag{1.4.2}
\]

while, for $a > \tau - 1$ and all $\ell \geq 1$,

\[
\mathbb{E}[X^a \mathbb{1}_{\{X \leq \ell\}}] \leq C \ell^{a-(\tau-1)}. \tag{1.4.3}
\]

**Proof** We note that for any cumulative distribution function $x \mapsto F(x)$ on the
non-negative reals, we have the partial integration identity
\[
\int_{u}^{\infty} f(x)F(dx) = f(u)[1 - F(u)] + \int_{u}^{\infty} [f(x) - f(u)]F(dx) \tag{1.4.4}
\]
\[
= f(u)[1 - F(u)] + \int_{u}^{\infty} \int_{0}^{x} f'(y)dyF(dx)
\]
\[
= f(u)[1 - F(u)] + \int_{u}^{\infty} f'(y)\int_{y}^{\infty} F(dx)dy
\]
\[
= f(u)[1 - F(u)] + \int_{u}^{\infty} f'(y)[1 - F(y)]dy.
\]
provided that either \( y \mapsto f'(y)[1 - F(y)] \) is absolutely integrable, or \( x \mapsto f(x) \) is either non-decreasing or non-increasing. Here, the interchange of the summation order is allowed by Fubini’s Theorem for non-negative functions (see (Halmos, 1950, Section 3.6, Theorem B)) when \( x \mapsto f(x) \) is non-decreasing, and by Fubini’s Theorem (Halmos, 1950, Section 3.6, Theorem C) when \( y \mapsto f'(y)[1 - F(y)] \) is absolutely integrable.

When \( D \geq 0 \), using (1.4.1) and (1.4.4), for \( a < \tau - 1 \) and \( \ell > 0 \),
\[
E\left[X^a\mathbb{1}_{\{X > \ell\}}\right] = x^aP(X > \ell) + \int_{\ell}^{\infty} ax^{a-1}P(X > x)dx \tag{1.4.5}
\]
\[
\leq C_\ell x^{a-(\tau-1)} + aC_\ell \int_{\ell}^{\infty} x^{a-1}x^{-(\tau-1)} \leq C_{a,\ell}x^{a-(\tau-1)}.
\]
as required.

An important notion in many graphs is the size-biased version \( X^* \) of a non-negative random variable \( X \) that is given by
\[
P(X^* \leq x) = \frac{E[X\mathbb{1}_{\{X \leq x\}}]}{E[X]} \tag{1.4.6}
\]
Let \( F^*_x \) denote the distribution function of \( X^* \). The following lemma gives bounds on the tail of the distribution function \( F^*_x \):

**Lemma 1.14** (Size-biased tail distribution) Let \( X \) be a non-negative random variable whose distribution function satisfies that for every \( x \geq 1 \),
\[
1 - F_x(x) \leq C_x x^{-(\tau-1)} \tag{1.4.7}
\]
Assume that \( \tau > 2 \), so that \( E[X] < \infty \). Further, assume that \( E[X] > 0 \). Then, there exists a constant \( C \) such that
\[
1 - F^*_x(x) \leq C_x x^{-(\tau-2)} \tag{1.4.8}
\]

**Proof** This follows immediately from (1.4.6), by using (1.4.2) with \( a = 1 \).

1.5 Notation

Let us introduce some standard notation used throughout this book.
Random variables

We use special notation for certain random variables. We write \( X \sim \text{Ber}(p) \) when \( X \) has a Bernoulli distribution with success probability \( p \), i.e., \( \Pr(X = 1) = 1 - \Pr(X = 0) = p \). We write \( X \sim \text{Bin}(n, p) \) when the random variable \( X \) has a binomial distribution with parameters \( n \) and \( p \), and we write \( X \sim \text{Poi}(\lambda) \) when \( X \) has a Poisson distribution with parameter \( \lambda \). We write \( X \sim \text{Exp}(\lambda) \) when \( X \) has an exponential distribution with mean \( 1/\lambda \). We write \( X \sim \text{Gam}(\lambda, r) \) when \( X \) has a Gamma distribution with density \( f_X(x) = \lambda^r x^{r-1} e^{-\lambda x}/\Gamma(r) \), where \( r, \lambda > 0 \). The random variable \( \text{Gam}(\lambda, r) \) has mean \( r/\lambda \) and variance \( r/\lambda^2 \). Finally, we write \( X \sim \text{Beta}(a, b) \) when \( X \) has a Beta distribution with parameters \( a, b > 0 \), so that \( X \) has density \( f_X(x) = x^{a-1}(1-x)^{b-1} \Gamma(a+b)/\Gamma(a)\Gamma(b) \).

We sometimes abuse notation, and write e.g., \( \Pr(\text{Bin}(n, p) = k) \) to denote \( \Pr(X = k) \) when \( X \sim \text{Bin}(n, p) \). We call a sequence of random variables \( (X_i)_{i \geq 1} \) independent and identically distributed (i.i.d.) when they are independent, and \( X_i \) has the same distribution as \( X_1 \) for every \( i \geq 1 \).

Convergence of random variables

We say that a sequence of events \( (E_n)_{n \geq 1} \) occurs with high probability (whp) when \( \lim_{n \to \infty} \Pr(E_n) = 1 \). We further write \( f(n) = O(g(n)) \) if \( |f(n)|/|g(n)| \) is uniformly bounded from above by a positive constant as \( n \to \infty \), \( f(n) = \Theta(g(n)) \) if \( f(n) = O(g(n)) \) and \( g(n) = O(f(n)) \), \( f(n) = \Omega(g(n)) \) if \( 1/f(n) = O(1/g(n)) \) and \( f(n) = o(g(n)) \) if \( f(n)/g(n) \) tends to 0 as \( n \to \infty \). We say that \( f(n) \gg g(n) \) when \( g(n) = o(f(n)) \).

For sequences of random variables \( (X_n)_{n \geq 1} \), we let \( X_n = \xrightarrow{d} X \) denote that \( X_n \) converges in distribution to \( X \), while \( X_n = \xrightarrow{a.s.} X \) denotes that \( X_n \) converges in probability to \( X \) and \( X_n = \xrightarrow{a.s.} X \) denotes that \( X_n \) converges almost surely to \( X \). We write that \( X_n = O_Y(Y_n) \) when \( |X_n|/Y_n \) is a tight sequence of random variables and \( X_n = O_Y(Y_n) \) when \( |X_n|/Y_n \xrightarrow{p} 0 \). Finally, we write that \( X_n = o_Y(Y_n) \) when \( X_n/Y_n \xrightarrow{p} 0 \).

Trees

In this book, we will often deal with trees, and then it is important to be clear about what we mean exactly with a tree. Trees are rooted and ordered. It will be convenient to think of a tree \( t \) with root \( \emptyset \) as being labelled in the Ulam-Harris way, so that a vertex \( v \) in generation \( k \) has a label \( \emptyset v_1 \cdots v_k \), where \( a_i \in \mathbb{N} \). Naturally, there are some restrictions in that if \( \emptyset v_1 \cdots v_k \in V(t) \), then also \( \emptyset v_1 \cdots v_{k-1} \in V(t) \) and \( \emptyset v_1 \cdots (v_k - 1) \in t \) when \( v_k \geq 2 \). We refer to [Volume 1, Chapter 3] for details.

It will sometimes also be useful to explore trees in a breadth-first exploration. This corresponds to the lexicographical ordering in the Ulam-Harris encoding of the tree. Ulam-Harris trees are also sometimes called plane trees (see e.g., (Drmota, 2009, Chapter 1)).

Let us now make the breadth-first ordering of the tree precise. For \( v \in V(t) \), let
1.6 Notes and discussion

|v| be its height. Thus |v| = k when \( v = \varnothing v_1 \cdots v_k \) and \( |\varnothing| = 0 \). Let \( u, v \in V(t) \). Then \( u < v \) when either \( |u| < |v| \), or when \( |u| = |v| \) and \( u = \varnothing u_1 \cdots u_k \) and \( v = \varnothing v_1 \cdots v_k \) are such that \( (u_1, \ldots, u_k) < (v_1, \ldots, v_k) \) in the lexicographic sense.

We next explain the breadth-first exploration of \( t \). For a tree \( t \) of size \(|V(t)| = t\), we let \((a_k)_{k=0}^t\) be the elements of \( V(t) \) ordered according to the breadth-first ordering of \( t \). For \( i \geq 1 \), let \( x_i \) denote the number of children of vertex \( a_i \). Thus, with \( d_v \) denoting the degree of \( v \in V(t) \) in the tree \( t \), we have \( x_1 = d_{a_0} = d_{\varnothing} \) and \( x_i = d_{a_i} - 1 \) for \( i \geq 2 \). The recursion

\[
 s_i = s_{i-1} + x_i - 1 \quad \text{for } i \geq 1, \quad \text{with} \quad s_0 = 1, \tag{1.5.1}
\]

describes the evolution of the number of unexplored vertices in the breadth-first exploration. For a finite tree \( t \) of size \(|V(t)| = t\), we thus have that \( s_i > 0 \) for \( i \in \{0, \ldots, t-1\} \), and \( s_t = 0 \). The sequence \((x_i)_{i=1}^t\) gives an alternative encoding of the tree \( t \) that will often be convenient.

1.6 Notes and discussion

Notes on Sections 9.1.3–1.3

These sections are summaries of chapters in Volume 1, to which we refer for notes and discussion. The exception is Section 1.3.4 on switching algorithms for uniform random graphs with prescribed degrees. Switching algorithms have a long history, dating back at least to McKay (1981), see also Gao and Wormald (2016); McKay and Wormald (1990), as well as McKay (2011) and the references therein for an overview. The material in Section 1.3.4 is an adaptation of some of the material in Gao et al. (2018), where Theorem 1.8 was used to compute the number of triangles in uniform random graphs with power-law degree distributions of infinite variance.

Notes on Section 1.4

1.7 Exercises for Chapter 1

Exercise 1.1 (Probability mass function typical degree) \hspace{1cm} Prove (1.1.3).

Exercise 1.2 (Uniform random graph) \hspace{1cm} Consider \( ER_n(p) \) with \( p = 1/2 \). Show that the result is a uniform graph, i.e., it has the same distribution as a uniform choice from all the graphs on \( n \) vertices.

Exercise 1.3 (Thin tails Poisson) \hspace{1cm} Show that, for every \( \alpha > 0 \), and with \( p_k = e^{-\lambda \frac{k^\alpha}{k!}} \) the Poisson probability mass function,

\[
 \lim_{k \to \infty} e^{\alpha k} p_k = 0. \tag{1.7.1}
\]

Exercise 1.4 (Weight of uniformly chosen vertex) \hspace{1cm} Let \( o \) be a vertex chosen uniformly at random from \([n]\). Show that the weight \( w_o \) of \( o \) has distribution function \( F_n \).
Exercise 1.5 (Maximal weight bound) Assume that Conditions 1.1(a)-(b) hold. Show that \( \max_{i \in [n]} w_i = o(n) \). Further, show that \( \max_{i \in [n]} w_i = o(\sqrt{n}) \) when Conditions 1.1(a)-(c) hold.

Exercise 1.6 (Domination weights) Let \( W_n \) have distribution function \( F_n \) from (1.3.17). Show that \( W_n \) is stochastically dominated by the random variable \( W \) having distribution function \( F \).

Exercise 1.7 (Degree of uniformly chosen vertex in \( GRG_n(w) \)) Prove that, under the conditions of Theorem 1.3, (1.3.22) holds.

Exercise 1.8 (Power-law degrees in generalized random graphs) Prove that, under the conditions of Theorem 1.3, (1.3.24) follows from (1.3.23). Does the converse also hold?

Exercise 1.9 (Number of erased edges) Assume that Conditions 1.5(a)-(b) hold. Show that Theorem 1.6 implies that the number of erased edges is \( o_P(n) \).

Exercise 1.10 (Degrees grow to infinity a.s.) Fix \( m = 1 \) and \( i \geq 1 \). Prove that \( D_i(t) \xrightarrow{a.s.} \infty \), by using that \( \sum_{s=1}^{t-1} I_s \leq D_i(t) \), where \( (I_t)_{t \geq 1} \) is a sequence of independent Bernoulli random variables with \( \mathbb{P}(I_t = 1) = (1+\delta)/(t(2+\delta)+1+\delta) \). What does this imply for \( m > 1 \)?

Exercise 1.11 (Degrees of fixed vertices) Prove Theorem 1.11 for \( m = 1 \) and \( \delta > -1 \) using the martingale convergence theorem and the fact that

\[
M_i(t) = \frac{D_i(t) + \delta}{1 + \delta} \prod_{s=i-1}^{t-1} \frac{(2 + \delta)s + 1 + \delta}{(2 + \delta)(s + 1)}
\] (1.7.2)

is a martingale.

Exercise 1.12 (Degrees distribution of affine Bernoulli PAM) Show that \( p_k \sim c_{\gamma, \beta} k^{-(1+1/\gamma)} \) when \( f(k) = \gamma k + \beta \). What is \( c_{\gamma, \beta} \)?

Exercise 1.13 (Degrees distribution of sublinear Bernoulli PAM) Show that \( \log(1/p_k) \sim k^{1-\alpha}/(\gamma(1-\alpha)) \) when \( f(k) \sim k^\alpha \) for some \( \alpha \in (0,1) \).

Exercise 1.14 (Power-law degree sequence) Prove (1.3.61) by using Stirling’s formula.
Chapter 2
LOCAL CONVERGENCE
OF RANDOM GRAPHS

Abstract
In this chapter, we discuss local weak convergence, as introduced by Benjamini and Schramm (2001) and independently by Aldous and Steele (2004). Local weak convergence described the intuitive notion that a finite graph, seen from the perspective of a typical vertex, looks like a certain limiting graph. This is already useful to make the notion that a finite cube in $\mathbb{Z}^d$ with large side length is locally much alike $\mathbb{Z}^d$ itself precise. However, it plays an even more profound role in random graph theory. For example, local convergence to some limiting tree, which often occurs in random graphs as we will see throughout this book, is referred to as locally tree-like behavior. In this chapter, we discuss local weak convergence in general.

We give general definitions of local convergence in several probabilistic senses, and show that local convergence is equivalent to the convergence of subgraph counts. Then we discuss several implications of local convergence, concerning local neighborhoods, clustering, assortativity and PageRank. We further investigate the relation between local convergence and the size of the giant, making the statement that the giant is almost local precise.

Organisation of this chapter
This chapter is organised as follows. In Section 2.1, we start by discussing the metric space of rooted graphs, that will play a crucial role in local weak convergence. In Section 2.2, we give the formal definition of local weak convergence. There, we think of our graphs as being deterministic and we discuss their convergence to some (surprisingly possibly random) limit. Here, the randomness originates from the fact that we consider the graph rooted at a random vertex, and this randomness may persist in the limit. In Section 2.3, we then extend the notion of local convergence to random graphs, for which there are several notions of convergence alike for convergence of real-valued random variables, such as local weak convergence and local convergence in probability. In Section 2.4, we discuss consequences of local convergence to local functionals, such as clustering and assortativity. For the latter, we extend the convergence of neighborhoods of a uniformly chosen vertex to those of a uniformly chosen edge. In Section 2.5, we discuss the consequences of local convergence on the giant component. While the proportion of vertices in the giant is not a continuous functional, one could argue that it is ‘almost local’, in a way that can be made precise. We close this chapter with notes and discussion in Section 2.6, and with exercises in Section 2.7.
2.1 The metric space of rooted graphs

Local weak convergence is a notion of weak convergence for finite graphs. In general, weak convergence is equivalent to convergence of expectations of continuous functions. For continuity, one needs a topology. Therefore, we start by discussing the topology of rooted graphs that is at the center of local weak convergence. We start with some definitions:

**Definition 2.1** (Locally finite and rooted graphs) A rooted graph is a pair \((G, o)\), where \(G = (V(G), E(G))\) is a graph with vertex set \(V(G)\) and edge set \(E(G)\), and \(o \in V(G)\) is a vertex. Further, a rooted or non-rooted graph is called locally finite when each of its vertices has finite degree (though not necessarily uniformly bounded).

In Definition 2.1, graphs can have finitely or infinitely many vertices, but we will always have graphs that are locally finite in mind. Also, in the definitions below, the graphs are deterministic and we will clearly indicate when we move to random graphs instead.

**Definition 2.2** (Neighborhoods as rooted graphs) For a rooted graph \((G, o)\), we let \(B_r(o) = B_r^{(o)}(o)\) denote the (rooted) subgraph of \((G, o)\) of all vertices at graph distance at most \(r\) away from \(o\). Formally, this means that \(B_r(o) = ((V(B_r^{(o)}(o)), E(B_r^{(o)}(o)), o), where\)

\[
V(B_r^{(o)}(o)) = \{u : d_o(o, u) \leq r\},
\]

\[
E(B_r^{(o)}(o)) = \{\{u, v\} \in E(G) : d_o(o, u), d_o(o, v) \leq r\}.
\]

We continue by introducing the notion of isomorphism between graphs, which basically describes that graphs ‘look the same’. Here is the formal definition:

**Definition 2.3** (Graph isomorphism) Two (finite or infinite) graphs \(G_1 = (V(G_1), E(G_1))\) and \(G_2 = (V(G_2), E(G_2))\) are called isomorphic, which we write as \(G_1 \simeq G_2\), when there exists a bijection \(\phi : V(G_1) \mapsto V(G_2)\) such that \(\{u, v\} \in E(G_1)\) precisely when \(\{\phi(u), \phi(v)\} \in E(G_2)\).

Similarly, two rooted (finite or infinite) graphs \((G_1, o_1)\) and \((G_2, o_2)\) where \(G_i = (V(G_i), E(G_i))\) for \(i \in \{1, 2\}\) are called isomorphic, which we write as \((G_1, o_1) \simeq (G_2, o_2)\), when there exists a bijection \(\phi : V(G_1) \mapsto V(G_2)\) such that \(\phi(o_1) = o_2\) and \(\{u, v\} \in E(G_1)\) precisely when \(\{\phi(u), \phi(v)\} \in E(G_2)\).

Exercises 2.1 and 2.2 below investigate the notion of graph isomorphism. We let \(\mathcal{G}^*\) denote the space of rooted graphs modulo isomorphisms, as introduced by Aldous and Steele (2004). This means that \(\mathcal{G}^*\) contains the set of equivalence classes of rooted connected graphs. We often omit the equivalence classes, and write \((G, o) \in \mathcal{G}^*\), bearing in mind that all \((G', o')\) such that \((G', o') \simeq (G, o)\) are considered to be the same. Thus, formally we deal with equivalence classes of rooted graphs. In the literature, the equivalence class containing \((G, o)\) is sometimes denoted as \([G, o]\).
2.2 Local weak convergence of graphs

These notions allow us to turn the space of connected rooted graphs into a metric space:

**Definition 2.4 (Metric on rooted graphs)** Let \((G_1, o_1)\) and \((G_2, o_2)\) be two rooted connected graphs, and write \(B^{G_1}_r(o_1)\) for the neighborhood of vertex \(o_1\) in \(G_1\). Let

\[
R^* = \sup \{ r : B^{G_1}_r(o_1) \cong B^{G_2}_r(o_2) \}. \tag{2.1.2}
\]

Define

\[
d_{\mathcal{G}}((G_1, o_1), (G_2, o_2)) = \frac{1}{R^* + 1}. \tag{2.1.3}
\]

The space \(\mathcal{G}\) of rooted graphs is a nice metric space under the metric \(d_{\mathcal{G}}\) in (2.1.3), in that it is separable and thus Polish. Here we recall that a metric space is called separable when there exists a countable dense subset of elements. We will see later on that such a countable dense set can be created by looking at finite rooted graphs. Since graphs that agree up to distance \(r\) are at distance at most \(1/(r + 1)\) from each other (see Exercise 2.3), this is indeed a dense countable subset. We discuss the metric structure of the space of rooted graphs in more detail in Appendix A.2. See Exercises 2.4 and 2.5 below for two exercises that study such aspects.

The value \(R^*\) is the largest value of \(r\) for which \(B^{G_1}_r(o_1)\) is isomorphic to \(B^{G_2}_r(o_2)\). When \(R^* = \infty\), then \(B^{G_1}_r(o_1)\) is isomorphic to \(B^{G_2}_r(o_2)\) for every \(r \geq 1\), and then the rooted graphs \(G_1\) and \(G_2\) are the same apart from an isomorphism, see Lemma A.9 in the appendix where this is worked out in detail.

2.2 Local weak convergence of graphs

In this section, we discuss local weak convergence of deterministic graphs. This section is organised as follows. In Section 2.2.1, we give the definitions of local weak convergence of (possibly disconnected) finite graphs. In Section 2.2.2, we provide a convenient criterion to prove local weak convergence and discuss tightness. In Section 2.2.3, we show that when the limit has full support on some subset of rooted graphs, that then convergence can be restricted to that set. Finally, in Section 2.2.4, we discuss two examples of graphs that converge locally weakly.

2.2.1 Definition of local weak convergence

Above, we have worked with connected graphs. When dealing with local weak convergence, we often wish to apply this to disconnected graphs. For such examples, we think of the corresponding rooted connected graph \((G_n, o_n)\) as corresponding to the connected component \(C(o_n)\) of \(o_n\) in \(G_n\). Here, we define, similarly
to (2.1.1), the rooted graph \( \mathcal{C}(o_n) = ((V(\mathcal{C}(o_n)), E(\mathcal{C}(o_n)), o_n) \), where

\[
\begin{align*}
V(\mathcal{C}(o_n)) &= \{ u : d_{oc}(o, u) < \infty \}, \\
E(\mathcal{C}(o_n)) &= \{ \{ u, v \} \in E(G) : d_{oc}(o, u), d_{oc}(o, v) < \infty \}.
\end{align*}
\]

Then, for \( h : \mathcal{G} \to \mathbb{R} \), by convention, we extend the definition to all (not-necessarily connected) graphs by letting

\[
h(G_n, o_n) \equiv h(\mathcal{C}(o_n)).
\]

Using this, we next define local weak convergence of finite graphs:

**Definition 2.5** (Local weak convergence) Let \( G_n = (V(G_n), E(G_n)) \) denote a finite (possibly disconnected) graph. Let \( (G_n, o_n) \) be the rooted graph obtained by letting \( o_n \in V(G_n) \) be chosen uniformly at random and restricting \( G_n \) to the connected component of \( o_n \) in \( G_n \). We say that \( (G_n, o_n) \) converges in the local weak sense to the connected rooted graph \( (G, o) \), which is a (possibly random) element of \( \mathcal{G} \) having law \( \mu \), when, for every bounded and continuous function \( h : \mathcal{G} \to \mathbb{R} \),

\[
E[h(G_n, o_n)] \to E_\mu[h(G, o)],
\]

where the expectation on the right-hand side of (2.2.3) is w.r.t. \((G, o)\) having law \( \mu \), while the expectation \( E_n \) is w.r.t. the random vertex \( o_n \).¹ We denote the above convergence by \( (G_n, o_n) \xrightarrow{d} (G, o) \).

Of course, the values \( h(G_n, o_n) \) only give you information about \( \mathcal{C}(o_n) \), which may only be a small portion of the graph with \( G_n \) is disconnected. However, since we are sampling \( o_n \in V(G_n) \) uniformly at random, we actually may 'see' every connected component, so in distribution we do observe the graph as a whole.

Since we will later apply local weak convergence ideas to random graphs, we strive to be absolutely clear about what we take the expectation with. Indeed, the expectation in (2.2.3) is equal to

\[
E[h(G_n, o_n)] = \frac{1}{|V(G_n)|} \sum_{u \in V(G_n)} h(G_n, u).
\]

The notion of local weak convergence is hard to grasp, and it also may appear to be rather weak. In the sequel, we discuss examples of graphs that converge locally weakly. Further, in Section 2.4 we discuss examples of how local weak convergence may be used to obtain interesting consequences for graphs, such as its clustering and its degree-degree dependences, measured through the assortativity coefficient. The notion of local weak convergence will also play a central role in this book. We continue by discussing a convenient criterion for proving local weak convergence.

¹ In the next section, our graphs may be random, and then we need to be very clear as to what we take the expectation.
2.2 Local weak convergence of graphs

2.2.2 Criterion for local weak convergence and tightness

We next provide a convenient criterion for local weak convergence:

**Theorem 2.6** (Criterion for local weak convergence) The sequence of finite rooted graphs $((G_n, o_n))_{n \geq 1}$ converges in the local weak sense to $(G, o)$ having law $\mu$ precisely when the sequence $((G_n, o_n))_{n \geq 1}$ is tight, and, for every rooted graph $H_\ast \in \mathcal{H}$ and all integers $r \geq 0$,

$$p^{(G_n)}(H_\ast) = \frac{1}{|V(G_n)|} \sum_{u \in V(G_n)} 1_{\{B^{(G_n)}_r(u) \supseteq H_\ast\}} \to \mu(B^{(G)}_r(o) \simeq H_\ast), \quad (2.2.5)$$

where $B^{(G_n)}_r(u)$ is the rooted $r$-neighborhood of $u$ in $G_n$, and $B^{(G)}_r(o)$ is the rooted $r$-neighborhood of $o$ in the limiting graph $(G, o)$.

Note that local weak convergence implies that (2.2.5) holds, since we can take $h(G, o) = 1_{\{B^{(G)}_r(o) \simeq H_\ast\}}$ and $h: \mathcal{H} \to \{0, 1\}$ is continuous (see Exercise 2.6).

**Proof** This is a standard weak convergence argument. By tightness, every subsequence of $((G_n, o_n))_{n \geq 1}$ has a further subsequence that converges in distribution. We will work along that subsequence, and note that the limiting law is that of $(G, o)$, since the laws of $B^{(G)}_r(o)$ uniquely identify the law of $(G, o)$ (see Proposition A.13 in Appendix A.2.5). Since this is true for every subsequence, the local weak limit is $(G, o)$.

Theorem 2.6 shows that the proportion of vertices in $G_n$ whose neighborhood looks like $H_\ast$ converges to a (possibly random) limit. In the literature, it is claimed that the tightness condition on $(G_n)_{n \geq 1}$ is not necessary, this would follow from the convergence in (2.2.5). We hope to return to this in the near future.

See Exercise 2.8, where you are asked to construct an example where the local weak limit of a sequence of deterministic graphs actually is random. You are asked to prove local weak convergence for some examples in Exercises 2.9 and 2.10.

Tightness is discussed in more detail in Appendix A.2.6. We next derive a convenient tightness criterion for local weak convergence. We recall that a sequence $(X_n)_{n \geq 1}$ of random variables is called uniformly integrable when

$$\lim_{K \to \infty} \limsup_{n \to \infty} \mathbb{E}[|X_n| 1_{\{|X_n| > K\}}] = 0. \quad (2.2.6)$$

We continue by giving a tightness criterion for local weak convergence:

**Theorem 2.7** (Tightness criterion for local weak convergence) Let $(G_n)_{n \geq 1}$ be a sequence of graphs with $V(G_n)$. Let $d^{(o_n)}_n$ denote the degree of $o_n$ in $G_n$, where $o_n$ is chosen uniformly at random from the vertex set $V(G_n)$ of $G_n$. Then $((G_n, o_n))_{n \geq 1}$ is tight when $(d^{(o_n)}_n)_{n \geq 1}$ forms a uniformly integrable sequence of random variables.

**Proof** Let $A$ be a family of finite graphs. For a graph $G$, let $U_G$ denote a random vertex drawn uniformly at random from $V(G)$, and let $U(G) = (G, U_G)$ be the rooted graph obtained by rooting $G$ at $U_G$. We need to show that if
\{\text{deg}_G(U_G) : G \in \mathcal{A}\} is a uniformly integrable sequence of random variables, then the family \(\mathcal{A}\) is tight. Let

\[ f(d) = \sup_{G \in \mathcal{A}} E[\text{deg}_G(U_G) \mathbb{1}_{(\text{deg}_G(U_G) > d)}]. \tag{2.2.7} \]

By assumption, \(\lim_{d \to \infty} f(d) = 0\). Write \(m(G) = E[\text{deg}_G(U_G)]\). Thus, \(1 \leq m(G) \leq f(0) < \infty\). Write \(D(G)\) for the degree-biased probability measure on \(\{(G, v) : v \in V(G)\}\), that is,

\[ D(G)[(G, v)] = \frac{\text{deg}_G(v)}{m(G)} \cdot U(G)[(G, v)], \tag{2.2.8} \]

and \(D_G\) for the corresponding root. Since \(U(G) \leq m(G)D(G) \leq f(0)D(G)\), it suffices to show that \(\{D(G) : G \in \mathcal{A}\}\) is tight. Note that \(\{\text{deg}_G(D_G) : G \in \mathcal{A}\}\) is tight.

For \(r \in \mathbb{N}\), let \(F_r^M(v)\) be the event such that there is some vertex at distance at most \(r\) from \(v\) whose degree is larger than \(M\). Let \(X\) be a uniform random neighbor of \(D_G\). Because \(D(G)\) is a stationary measure for a simple random walk, \(F_r^M(D_G)\) and \(F_r^M(X)\) have the same probability. Also,

\[ \mathbb{P}\left(F_r^M(D_G) \mid \text{deg}_G(D_G)\right) \leq \text{deg}_G(D_G) \mathbb{P}\left(F_r^M(X) \mid \text{deg}_G(D_G)\right). \tag{2.2.9} \]

We claim that, for all \(r \in \mathbb{N}\) and \(\varepsilon > 0\), there exists \(M < \infty\) such that

\[ \mathbb{P}\left(F_r^M(X)\right) \leq \varepsilon \tag{2.2.10} \]

for all \(G \in \mathcal{A}\). This clearly implies that \(\{D(G) : G \in \mathcal{A}\}\) is tight. We prove the claim by induction on \(r\).

The statement for \(r = 0\) is trivial. Given that the property holds for \(r\), let us now show it for \(r + 1\). Given \(\varepsilon > 0\), choose \(d\) so large that \(\mathbb{P}(\text{deg}_G(D_G) > d) \leq \varepsilon/2\) for all \(G \in \mathcal{A}\). Also, choose \(M\) so large that \(\mathbb{P}(F_r^M(D_G)) \leq \varepsilon/(2d)\) for all \(G \in \mathcal{A}\). Write \(F\) for the event such that \(\text{deg}_G(D_G) > d\). Then, by conditioning on \(\text{deg}_G(D_G)\), we see that

\[ \mathbb{P}\left(F_{r+1}^M(D_G)\right) \leq \mathbb{P}(F) + E[\mathbb{1}_{F} \mathbb{P}\left(F_{r+1}^M(D_G) \mid \text{deg}_G(D_G)\right)] \tag{2.2.11} \]

\[ \leq \varepsilon/2 + E[\mathbb{1}_{F} \text{deg}_G(D_G) \mathbb{P}\left(F_{r+1}^M(D_G) \mid \text{deg}_G(D_G)\right)] \]

\[ \leq \varepsilon/2 + E[\mathbb{1}_{F} \text{deg}_G(D_G) \mathbb{P}\left(F_{r+1}^M(D_G) \mid \text{deg}_G(D_G)\right)] \]

\[ \leq \varepsilon/2 + d \mathbb{P}\left(F_{r+1}^M(D_G)\right) \]

\[ \leq \varepsilon/2 + d \varepsilon/(2d) = \varepsilon, \]

for all \(G \in \mathcal{A}\), which proves the claim. \(\square\)

The needed uniform integrability in Theorem 2.7 is quite suggestive. Indeed, in many random graph models, such as for example the configuration model, the degree of a random neighbor of a vertex has the size-biased degree distribution. Previously, we have often written \(D_n = d_n^{(n)}\) for the degree of a uniform vertex
in our graph. When \((d_{o_n}^{(e)})_{n \geq 1}\) forms a uniformly integrable sequence of random variables, there exists a subsequence along which \(D_n^{(e)}\), the size-biased version of \(D_n = d_{o_n}^{(e)}\), converges in distribution (see Exercise 2.11 below).

For the configuration model, Conditions 1.5(a)-(b) imply that \((d_{o_n}^{(e)})_{n \geq 1}\) is a tight sequence of random variables (see Exercise 2.12). Further, [Volume 1, Theorem 7.25] discussed how Conditions 1.5(a)-(b) imply the convergence of the degrees of \(o_n\) to \(\bar{\mu}\). Let us explain this in more detail.

Therefore, \(\bar{\mu} \neq \mu\). Fix this set. Then we bound
\[
\limsup_{n \to \infty} P(B_r^{(e)}(o_n) \notin \mathcal{H}_s(r)) \leq 1 - \liminf_{n \to \infty} P(B_r^{(e)}(o_n) \in \mathcal{H}_s(r,m)) \leq 1 - \bar{\mu}(\mathcal{H}_s(r,m)) = 1 - (1 - \varepsilon) = \varepsilon.
\]

2.2.3 Local weak convergence and completeness of the limit

For many settings, the local weak limit is almost surely contained in a smaller set of rooted graphs. In this case, it turns out to be enough to prove the convergence on that subset. Then it turns out that it suffices to study convergence on that subset.

**Theorem 2.8** (Local weak convergence and subsets) Let \((G_n)_{n \geq 1}\) be a sequence of rooted graphs. Let \((G, o)\) be a random variable on \(\mathcal{G}\) having law \(\mu\). Let \(\mathcal{H} \subset \mathcal{G}\) be a subset of the space of rooted graphs. Assume that \(\mu(G \in \mathcal{H}) = 1\). Then, \((G_n, o_n) \overset{d}{\to} (G, o)\) when (2.3.9) holds for all \(H \in \mathcal{H}_s(r)\) and all \(r \geq 1\).

We will apply Theorem 2.8 in particular when the limit is a.s. a tree. Then, Theorem 2.8 implies that we only have to investigate \(H_s\) that are trees themselves.
Local convergence of random graphs

Since $\varepsilon > 0$ is arbitrary, we conclude that $\mathbb{P}(B_r^{(G_n)}(o_n) \notin \mathcal{T}_*(r)) \to 0$. In particular, this means that, for any $H_\ast \notin \mathcal{T}_*(r)$,
\[
\mathbb{P}(B_r^{(G_n)}(o_n) \simeq H_\ast) \to 0 = \mu(B_r^{(G)}(o) \simeq H_\ast).
\] (2.2.14)

Thus, when the required convergence holds for every $H_\ast \in \mathcal{T}_*(r)$, it follows for every $H_\ast \notin \mathcal{T}_*(r)$ with limit zero when $\mu((G, o) \in \mathcal{T}_*) = 1$.

2.2.4 Examples of local weak convergence

We close this section by discussing two relevant examples of local weak convergence. We start with uniform points in large boxes in $\mathbb{Z}^d$, and then discuss the local weak limit of finite trees.

**Local weak convergence of boxes in $\mathbb{Z}^d$**

Consider the nearest-neighbor box $[n]^d$, where $x = (x_1, \ldots, x_d) \in \mathbb{Z}^d$ is a neighbor of $y \in \mathbb{Z}^d$ precisely when there is a unique $i \in [d]$ such that $|x_i - y_i| = 1$. Take a root uniformly at random, and denote the resulting graph by $(G_n, o_n)$. We claim that $(G_n, o_n) \xrightarrow{d} (\mathbb{Z}^d, o)$, which we now prove. We rely on Theorem 2.6, which shows that we need to prove tightness and convergence of subgraph proportions. We use Theorem 2.7 to see that $(G_n, o_n)$ is tight, since $(G_n, o_n)$ has uniformly bounded degree.

Let $\mu$ be the point measure on $(\mathbb{Z}^d, o)$, so that $\mu(B_{r}^{(G)}(o) \simeq B_{r}^{(z_d)}(o)) = 1$. Thus, by Theorem 2.8, it remains to show that $p^{(G_n)}(B_{r}^{(z_d)}(o)) \to 1$ (recall (2.2.5)).

For this, we note that $B_{r}^{(G_n)}(o_n) \simeq B_{r}^{(z_d)}(o)$ unless $o_n$ happens to lie within distance strictly smaller than $r$ from one of the boundaries of $[n]^d$. This means that either one of the coordinates of $o_n$ is in $[r-1]$, or in $[n] \setminus [n-r+1]$. Since the latter occur with vanishing probability, the claim follows.

In the above case, we see that the local weak limit is deterministic, as one would have expected. One can generalize the above to convergence of tori as well.

**Local weak convergence of truncated trees**

Recall the notion of a tree in Section 1.5. Fix a degree $d$, and a height $n$ that we will take to infinity. We now define the regular tree $\mathbb{T}_{d,n}$ truncated at height $n$. The graph $\mathbb{T}_{d,n}$ has vertex set
\[
V(\mathbb{T}_{d,n}) = \{\emptyset\} \cup \bigcup_{k=1}^{n} \{\emptyset\} \times [d] \times [d-1]^{k-1},
\] (2.2.15)
and edge set as follows. Let $v = \emptyset v_1 \cdots v_k$ and $u = \emptyset u_1 \cdots u_{\ell}$ be two vertices. We say that $u$ is the parent of $v$ when $\ell = k - 1$ and $u_i = v_i$ for all $i \in [k-1]$. Then we say that two vertices $u$ and $v$ are neighbors when $u$ is the parent of $v$ or vice versa. This is a graph with
\[
|V(\mathbb{T}_{d,n})| = 1 + d + \cdots + d(d-1)^{n-1}
\] (2.2.16)
2.3 Local convergence of random graphs

Let \( o_n \) denote a vertex chosen uniformly at random from \( V(T_{d,n}) \). We consider \( (T_{d,n}, o_n) \) and its local weak limit, that we now describe. We first consider the so-called canopy tree. For this, we take the graph \( T_{d,n} \), root it at a leaf and take the limit of \( n \to \infty \). Remove the label of the root leaf, and denote this graph by \( T_d \), which we consider to be an unrooted graph. This graph has a unique infinite path from the root-leaf. Let \( o_{\ell} \) be the \( \ell \)th vertex on this infinite path (the root leaf having label 0), and consider \( (T_d, o_{\ell}) \). Define the limiting measure \( \mu \) by

\[
\mu((T_d, o_{\ell})) \equiv \mu_{\ell} = (d - 2)(d - 1)^{-\ell + 1}, \quad \ell \geq 0. \tag{2.2.17}
\]

We claim that \( (G_n, o_n) \equiv (T_{d,n}, o_n) \overset{d}{\to} (G, o) \) with law \( \mu \). We again rely on Theorem 2.6, which shows that we need to prove tightness and convergence of subgraph proportions. We use Theorem 2.7 to see that \( (T_{d,n}, o_n) \) is tight, since \( (T_{d,n}, o_n) \) has uniformly bounded degree.

Further, by Theorem 2.8, it remains to show that \( p^{(G_n)}(B^{(T_{d,n})}_{o_{\ell}}(o_n)) \to \mu_{\ell} \) (recall (2.2.5)). When \( n \) is larger than \( r \) (which we now assume), we have that \( B^{(G_n)}_{o_{\ell}}(o_n) \simeq B^{(T_{d,n})}_{o_{\ell}}(o_n) \) precisely when \( o_n \) has distance \( \ell \) from the closest leaf. There are

\[
d(d - 1)^{n-1-k}
\]

vertices at distance \( k \) from the root, out of a total of \( |V(T_{d,n})| = d(d - 1)^n/(d - 2)(1 + o(1)) \). Thus,

\[
p^{(G_n)}(B^{(T_{d,n})}_{o_{\ell}}(o_n)) = \frac{d(d - 1)^{n-1-k}}{|V(T_{d,n})|} \to (d - 2)(d - 1)^{-\ell + 1} = \mu_{\ell}, \tag{2.2.19}
\]

as required.

We see that, for the truncated tree, the local weak limit is random, where the randomness clearly originates from the choice of the random root of the graph. In this case, this is more particularly due to the choice how far away the chosen root is from the leaves of the finite tree.

Having discussed the notion of local weak convergence for deterministic graphs, we now move on to random graphs. Here the situation becomes even more delicate, as now we have double randomness, both in the random root as well as the random graph. This gives rise to surprisingly subtle matters.

2.3 Local convergence of random graphs

We next discuss settings of random graphs. This section is organised as follows. In Section 2.3.1, we define what it means for a sequence of random graphs to converge in the local weak sense, as well as which different versions exist thereof. In Section 2.3.2, we then give a useful criterion to verify local weak convergence of random graphs that shows that the proportion of vertices whose neighborhood
counts are isomorphic to a certain rooted graph converges for all rooted graphs. In Section 2.2.3 we prove the completeness of the limit by showing that when the limit is supported on a subset of rooted graphs, then one only needs to verify the convergence for that subset. In many examples that we will encounter in this book, this subset is the collection of trees. We close in Section 2.3.5 by presenting the example of local weak convergence in probability of the Erdős-Rényi random graph $\text{ER}_n(\lambda/n)$.

2.3.1 Definition of local convergence of random graphs

Even for random variables, there are different notions of convergence that are relevant, such as convergence in distribution and in probability. Also for local weak convergence, there are several related notions of convergence that we may consider:

**Definition 2.9** (Local weak convergence of random graphs) Let $G_n = (V(G_n), E(G_n))$ denote a finite (possibly disconnected) random graph. Then,

(a) we say that $G_n$ converges locally weakly to $(\bar{G}, \bar{o})$ having law $\bar{\mu}$ when

$$\mathbb{E}[h(G_n, o_n)] \to \mathbb{E}_{\bar{\mu}}[h(\bar{G}, \bar{o})],$$

for every bounded and continuous function $h: \mathcal{G} \to \mathbb{R}$, while the expectation $\mathbb{E}$ is w.r.t. the random vertex $o_n$ and the random graph $G_n$. This is equivalent to $(G_n, o_n) \xrightarrow{d} (G, o)$;

(b) we say that $G_n$ converges locally in probability to $(G, o)$ having (possibly random) law $\mu$ when

$$\mathbb{E}[h(G_n, o_n) \mid G_n] \xrightarrow{p} \mathbb{E}_{\mu}[h(G, o)].$$

for every bounded and continuous function $h: \mathcal{G} \to \mathbb{R}$. We write this as $G_n \xrightarrow{p} (G, o)$;

(c) we say that $G_n$ converges locally almost surely to $(G, o)$ having law $\mu$ when

$$\mathbb{E}[h(G_n, o_n) \mid G_n] \xrightarrow{a.s.} \mathbb{E}_{\mu}[h(G, o)],$$

for every bounded and continuous function $h: \mathcal{G} \to \mathbb{R}$. We write this as $G_n \xrightarrow{a.s.} (G, o)$.

As usual in convergence of random variables, the difference between these closely related definitions lies in what can be concluded from it, and how they can be proved. We discuss the differences in the limits denoted by $\bar{\mu}$ for convergence in distribution and $\mu$ for convergence in probability and almost surely in Corollary 2.11 below.

When we have local convergence in probability, then $\mathbb{E}[h(G_n, o_n) \mid G_n]$, which is a random variable due to the dependence on the random graph $G_n$, converges in probability to $\mathbb{E}_{\mu}[h(G, o)]$, which possibly is also a random variable in that $\mu$ might be a random probability distribution on $\mathcal{G}$. When we have local weak
2.3 Local convergence of random graphs

convergence, instead, then only expectations w.r.t. the random graph of the form $E[h(G_n, o_n)]$ converge.

Remark 2.10 (Local convergence in probability and rooted vs. unrooted graphs)
Usually, when we have a sequence of objects $x_n$ living in some space $X$, and $x_n$ converges to $x$, then $x$ also lives in the same space. In the above local convergence in probability and almost surely, however, we take a graph sequence $(G_n)_{n \geq 1}$ that converges locally in probability to a rooted graph $(G, o) \sim \mu$. One might have guessed that this is related to $(G_n, o_n) \xrightarrow{P} (G, o)$, but it is quite different. Indeed, $(G_n, o_n) \xrightarrow{P} (G, o)$ is a very strong and not so useful statement, since sampling $o_n$ gives rise to variability in $(G_n, o_n)$ that is hard to capture by the limit $(G, o)$.

The following observations turn the local convergence in probability into convergence of objects living on the same space, namely, the space of measures on rooted graphs. Denote the empirical neighborhood measure $\mu_n$ on $G_n$ by

$$\mu_n(\mathcal{H}_*) = \frac{1}{|V(G_n)|} \sum_{v \in V(G_n)} 1\{ (G_n, v) \in \mathcal{H}_* \}.$$  

(2.3.4)

for any measurable subset $\mathcal{H}_*$ of $\mathcal{G}_*$. Then, $(G_n)_{n \geq 1}$ that converges locally in probability to the random rooted graph $(G, o) \sim \mu$ when

$$\mu_n(\mathcal{H}_*) \xrightarrow{P} \mu(\mathcal{H}_*),$$  

(2.3.5)

for every measurable subset $\mathcal{H}_*$ of $\mathcal{G}_*$. Thus, it is equivalent to the convergence in probability of the empirical neighborhood measure. This explains why we view local convergence in probability as a property of the graph $G_n$, rather than the rooted graph $(G_n, o_n)$. A similar comment applies to local convergence almost surely.

The limits $\bar{\mu}$ for convergence in distribution and $\mu$ for convergence in probability and almost surely are closely related to each other:

Corollary 2.11 (Relation between local weak limits) Suppose that $(G_n, o_n) \xrightarrow{d} (G, \bar{o})$ having law $\bar{\mu}$, and in $(G_n)_{n \geq 1}$, converges locally probability to $(G, o)$ having law $\mu$. Then $\bar{\mu}(\cdot) = E[\mu(\cdot)]$. In particular,

$$E_{\bar{\mu}}[h(G, \bar{o})] = E,E_{\mu}[h(G, o)].$$  

(2.3.6)

Proof By construction,

$$E[h(G_n, o_n) | G_n] \rightarrow E_{\mu}[h(G, o)].$$  

(2.3.7)

Further, note that $E_{\mu}[h(G, o)]$ is a bounded random variable, and so is $E_n[h(G_n, o_n)]$. Therefore, by Dominated Convergence [Volume 1, Theorem A.1], also the expectations converge. Therefore,

$$E[h(G_n, o_n)] = E[E_n[h(G_n, o_n)]] \rightarrow E[E_{\mu}[h(G, o)]],$$  

(2.3.8)
which, together with \((2.3.7)\), completes the proof. \(\square\)

In most of our examples, the law \(\mu\) of the local weak limit in probability is actually deterministic, in which case \(\bar{\mu} = \mu\). However, there are some cases where this is not true. A simple example arises as follows. For \(\text{ER}_n(\lambda/n)\), the local weak limit in probability will turn out to be a \(\text{Poi}(\lambda)\) branching process. Therefore, when considering \(\text{ER}_n(X/n)\), where \(X\) is uniform on \([0, 2]\), the local weak limit in probability will be a \(\text{Poi}(X)\) branching process. Here, the mean of all the offsprings are random and related, as they are all equal to \(X\). This is not the same as a mixed-Poisson branching process with offspring distribution \(\text{Poi}(X)\), since for the local limit in probability of \(\text{ER}_n(X/n)\), we draw \(X\) only once. See Section 2.3.5 for more details on local convergence for \(\text{ER}_n(\lambda/n)\).

We have added the notion of local convergence in the almost sure sense, even though for random graphs this notion often is not highly useful. Indeed, almost sure convergence for random graphs is often already tricky, as for static models such as the Erdős-Rényi random graph and the configuration model, there is no obvious relation between the graphs of size \(n\) and those of size \(n + 1\). This of course is different for the preferential attachment model, which is a (consistent) graph process. However, even for preferential attachment models, local convergence is all about the neighborhood of a uniform vertex, and it is not obvious to relate the uniform choices for graphs of different sizes.

### 2.3.2 Criterion for Local Convergence of Random Graphs and Tightness

We next discuss a convenient criterion for local convergence, inspired by Theorem 2.6:

**Theorem 2.12** (Criterion for local convergence of random graphs) Let \((G_n)_{n \geq 1}\) be a sequence of rooted graphs. Then,

(a) \((G_n, o_n) \xrightarrow{d} (\bar{G}, \bar{o})\) having law \(\bar{\mu}\) when \(((G_n, o_n))_{n \geq 1}\) is tight and when, for every rooted graph \(H_* \in \mathcal{G}_*\) and all integers \(r \geq 0,\)

\[
\mathbb{E}[p^{(G_n)}(H_*)] = \frac{1}{|V(G_n)|} \sum_{v \in V(G_n)} \mathbb{P}(B_r^{(G_n)}(v) \simeq H_*) \rightarrow \bar{\mu}(B_r^{(G)}(\bar{o}) \simeq H_*). \quad (2.3.9)
\]

(b) \(G_n\) converges locally in probability to \((G, o)\) having law \(\mu\) when \(((G_n, o_n))_{n \geq 1}\) is tight and when, for every rooted graph \(H_* \in \mathcal{G}_*\) and all integers \(r \geq 0,\)

\[
p^{(G_n)}(H_*) = \frac{1}{|V(G_n)|} \sum_{v \in V(G_n)} \mathbb{1}_{(B_r^{(G_n)}(v) \simeq H_*)} \xrightarrow{p} \mu(B_r^{(G)}(o) \simeq H_*). \quad (2.3.10)
\]

(c) \(G_n\) converges almost surely in the local weak sense to \((G, o)\) having law \(\mu\) when \(((G_n, o_n))_{n \geq 1}\) is almost surely tight and when, for every rooted graph \(H_* \in \mathcal{G}_*\)
2.3 Local convergence of random graphs

and all integers \( r \geq 0 \),

\[
p^{(G_n)}(H_*) = \frac{1}{|V(G_n)|} \sum_{v \in V(G_n)} 1_{\{B_{(G_n)}^r(v) \geq H_*\}} \xrightarrow{a.s.} \mu(B_*(o) \approx H_*).
\] (2.3.11)

Proof. This follows directly from Theorem 2.6.

We next investigate tightness:

**Theorem 2.13 (Tightness criterion for local convergence of random graphs)**

Let \((G_n)_{n \geq 1}\) be a sequence of graphs with \(V(G_n)\). Let \(d_{o_n}^n\) denote the degree of \(o_n\) in \(G_n\), where \(o_n\) is chosen uniformly at random from the vertex set \(V(G_n)\) of \(G_n\). Then

(a) \(((G_n,o_n))_{n \geq 1}\) is tight when \((d_{o_n}^n)_{n \geq 1}\) forms a uniformly integrable sequence of random variables;

(b) \(((G_n,o_n))_{n \geq 1}\) is tight in probability when \((d_{o_n}^n)_{n \geq 1}\) forms a uniformly integrable sequence of random variables;

(c) \(((G_n,o_n))_{n \geq 1}\) is tight almost surely when \((d_{o_n}^n)_{n \geq 1}\) almost surely forms a uniformly integrable sequence of random variables.

Proof. Part (a) is simply Theorem 2.7. For part (b), we note that Theorem 2.7 implies that tightness in probability follows when

\[
\lim_{K \to \infty} \limsup_{n \to \infty} P\left(\mathbb{E}[d_{o_n}^n 1_{\{d_{o_n}^n \geq K\}} \mid G_n] > \varepsilon\right) = 0,
\] (2.3.12)

which, after applying the Markov inequality, follows from the uniform integrability of \((d_{o_n}^n)_{n \geq 1}\). For part (c), we again use Theorem 2.7, but now we need to show that the convergence occurs almost surely. That means that, for every \(\varepsilon > 0\), there exists a \(K = K(\varepsilon) > 0\) such that the event \(\{\mathbb{E}[d_{o_n}^n 1_{\{d_{o_n}^n \geq K\}} \mid G_n] > \varepsilon\}\) occurs only finitely often.

In what follows, we will be mainly interested in convergence in probability in the local weak sense, since this is the notion that is the most powerful in the setting of random graphs.

2.3.3 Local convergence and completeness of the limit

For many random graph models, the local limit is almost surely contained in a smaller set of rooted graphs. The most common example is when the random graph converges locally to a tree (recall that trees are rooted, see Section 1.5), but it can apply more generally. In this case, and similarly to Theorem 2.8, it turns out to be enough to prove the convergence in Definition 2.9 only for elements that the limit \((G,o)\) takes values in. Let us explain this in more detail.

Let \(\mathcal{T} \subset \mathcal{G}_*\) be a subset of the space of rooted graphs. Recall that \(\mathcal{T}^r \subset \mathcal{G}_*\) is the subset of \(\mathcal{T}_*\) of graphs for which the distance between any vertex and the root is at most \(r\). Then, we have the following result:
**Theorem 2.14** (Local weak convergence and subsets) Let \((G_n)_{n \geq 1}\) be a sequence of rooted graphs. Let \((\bar{G}, \bar{o})\) be a random variable on \(\mathcal{G}\) having law \(\bar{\mu}\). Let \(\mathcal{S} \subset \mathcal{G}\) be a subset of the space of rooted graphs. Assume that \(\bar{\mu}( (\bar{G}, \bar{o}) \in \mathcal{S}) = 1\). Then, \((G_n, o_n) \xrightarrow{d} (\bar{G}, \bar{o})\) when (2.3.9) holds for all \(H \in \mathcal{S}(r)\) and all \(r \geq 1\).

Similar extensions hold for local convergence in probability in (2.3.10) and almost surely in (2.3.11), with \(\bar{\mu}\) replaced by \(\mu\) and \((\bar{G}, \bar{o})\) by \((G, o)\).

**Proof** The proof for local weak convergence is identical to that of Theorem 2.8. The extensions to convergence in probability and almost surely follow similarly. \(\square\)

### 2.3.4 Local convergence of random regular graphs

In this section, we give the very first example of a random graph that converges locally in probability. This is the random regular graph, which is obtained by taking the configuration model \(CM_n(d)\), and letting \(d_i = d\) for all \(i \in [n]\). Here, we assume that \(nd\) is even. The main result is the following:

**Theorem 2.15** (Local weak convergence of random regular graphs) Fix \(d \geq 1\) and assume that \(nd\) is even. The random regular graph of degree \(d\) and size \(n\) converges locally in probability to rooted \(d\)-regular tree.

Looking back at the local weak limit of truncated trees discussed in Section 2.2.4, we see that a random regular graph is a much better approximation to a \(d\)-regular tree than a truncation of it. One could see this as another example of the probabilistic method, where a certain amount of randomization is useful to construct deterministic objects.

**Proof** Let \(G_n\) be the random regular graph, and let \((T_d, o)\) be the rooted \(d\)-regular tree. Since the limit is constant, convergence in probability follows from convergence in distribution, which we now do. We use the convenient criterion in Theorem 2.12. Tightness follows by Theorem 2.13, since the degrees in \(G_n\) are constant, so in particular bounded. We see that the only further requirement is to show that \(\mathbb{E}[p^{G_n}(B_r^{T_d}(o))] \to 1\).

Write

\[
1 - \mathbb{E}[p^{G_n}(B_r^{T_d}(o))] = \mathbb{P}(B_r^{G_n}(o_n) \text{ is not a tree}).
\] (2.3.13)

For \(B_r^{G_n}(o_n)\) not to be a tree, a cycle needs to occur within distance \(r\). We grow the neighborhood \(B_r^{G_n}(o_n)\) by pairing the half-edges incident to discovered vertices one by one. Since there is just a bounded number of unpaired half-edges incident to the vertices found at any moment in the exploration, and since we need to pair at most

\[
d + d^2 + \cdots + d^r
\] (2.3.14)

half-edges, the probability that any one of them creates a cycle vanishes. We conclude that \(1 - \mathbb{E}[p^{G_n}(B_r^{T_d}(o))] \to 0\), which completes the proof. \(\square\)
2.3 Local convergence of random graphs

2.3.5 Local convergence of Erdős-Rényi random graphs

In this section, we work out one example and show that the Erdős-Rényi random graph $ER_n(\lambda/n)$ converges locally weakly in probability to a Poisson branching process:

Theorem 2.16 (Local convergence of Erdős-Rényi random graph) Fix $\lambda > 0$. $ER_n(\lambda/n)$ converges locally in probability to a Poisson branching process with mean offspring $\lambda$.

Proof We start by reducing the proof to a convergence in probability statement.

Setting the stage of the proof

We start by using the convenient criterion in Theorem 2.12. Tightness follows by Theorem 2.13, since the degrees in $G_n = ER_n(\lambda/n)$ are uniformly integrable, by (1.3.5) (see also Theorem 1.3), and the fact that

$$\mathbb{E}[d'_{o_n}] = \frac{2}{n} \mathbb{E}[|E(G_n)|] = \frac{2}{n} \binom{n}{2} \frac{\lambda}{n} \to \lambda = \sum_{k \geq 0} k p_k. \quad (2.3.15)$$

Thus, we are left to prove (2.3.10) in Theorem 2.12.

We then rely on Theorem 2.14, and prove convergence of subgraph proportions as in (2.3.11) only for trees. Recall the discussion of trees in Section 1.5. There, we considered trees to be ordered, in the sense that every vertex has an ordered set of forward neighbors. Fix a tree $t$. Then, in order to prove that $ER_n(\lambda/n)$ converges locally weakly in probability to a Poisson branching process, we need to show that, for every tree $t$ and all integers $r \geq 0$,

$$p(G_n)(t) = \frac{1}{n} \sum_{u \in [n]} \mathbb{1}_{\{B^{(G_n)}(u) \simeq t\}} \overset{p}{\to} \mu(B^{(G)}(o) \simeq t), \quad (2.3.16)$$

where $G_n$ equals $ER_n(\lambda/n)$, and the law $\mu$ of $(G, o)$ is that of a Poisson branching process. We see that in this case, $\mu$ is deterministic, as it will be in most examples. In (2.3.20), we may without loss of generality assume that $t$ is a finite tree of depth at most $r$, since otherwise both sides are zero.

Ordering trees and subgraphs

Of course, for the event $\{B^{(G)}(o) \simeq t\}$ to occur, the order of the tree $t$ is irrelevant. Recall the breadth-first exploration of the tree $t$, which is described in terms of $(x_i)_{i=0}^t$ as in (1.5.1) and the corresponding vertices $(a_i)_{i=0}^t$. Further, note that $(G, o)$ is, by construction, an ordered tree, and therefore so is $B^{(G)}(o)$. Therefore, we can write $B^{(G)}(o) = t$ to indicate that the two ordered trees $B^{(G)}(o)$ and $t$ agree. In terms of this notation, one can compute

$$\mu(B^{(G)}(o) = t) = \prod_{i \in [t]: \text{dist}(o, a_i) < r} e^{-\lambda} \frac{\lambda^{x_i}}{x_i!}, \quad (2.3.17)$$
where dist(∅, v) is the tree distance between v ∈ V(t) and the root ∅ ∈ V(t). We note that B_r(o) ∼ t says nothing about the degrees of the vertices that are at distance exactly r away from the root ∅, which is why we restrict to vertices v with dist(∅, a_i) < r in (2.3.18). Further, for each ordered tree t’ that is isomorphic to the tree t, we have that μ(B_r(o) = t’) = μ(B_r(o) = t), since the root degrees and degree sequences of the non-root vertices are the same for all trees that are isomorphic to t, and the right-hand side of (2.3.17) only depends on the degree of the root, and the degrees of all other non-root vertices. Therefore,

$$\mu(B_r^{G_n}(o) \simeq t) = \#(t) \prod_{i \in |t| : \text{dist}(\emptyset, a_i) < r} e^{-\lambda x_i / x_i^r}$$  \hspace{1cm} (2.3.18)

where #(t) is the number of ordered trees that are isomorphic to t. This identifies the right hand side of (2.3.20). We note further that by permuting the labels of all the children of any vertex in t, we obtain a rooted tree that is isomorphic to t, and there are \( \prod_{i \in |t|} x_i! \) such permutations. However, not all of them may lead to distinct ordered trees. In our analysis, the precise value of #(t) is not relevant.

It will be convenient to also order the vertices in \( B_r^{G_n}(o_n) \), where \( G_n = \text{ER}_n(\lambda/n) \). This will be achieved by ordering the forward children of a vertex in \( B_r^{G_n}(o_n) \) according to their vertex labels. Then, we can again write \( B_r^{G_n}(o) = t \) to indicate that the two ordered graphs \( B_r^{G_n}(o) \) and \( t \) agree. This implies that \( B_r^{G_n}(o) \) is a tree (so there are no cycles within depth r), and that its ordered version is equal to the ordered tree t. Then, as in (2.3.18),

$$p^{G_n}(t) = \frac{1}{n} \sum_{u \in [n]} \mathbb{1}_{\{B_r^{G_n}(u) = t\}} = \#(t) \frac{1}{n} \sum_{u \in [n]} \mathbb{1}_{\{B_r^{G_n}(u) = t\}}.$$  \hspace{1cm} (2.3.19)

We will prove that

$$\frac{1}{n} \sum_{u \in [n]} \mathbb{1}_{\{B_r^{G_n}(u) = t\}} \xrightarrow{p} \mu(B_r^{G_n}(o) = t).$$  \hspace{1cm} (2.3.20)

**Second moment method: first moment**

To prove (2.3.20), we use a second moment method. Denote

$$N_n(t) = \sum_{u \in [n]} \mathbb{1}_{\{B_r^{G_n}(u) = t\}}.$$  \hspace{1cm} (2.3.21)

We will then apply a second moment method to show that \( N_n(t) \) is highly concentrated around \( n \mu(B_r(o) = t) \).

We start by investigating the first moment of \( N_n(t) \), which equals

$$\mathbb{E}[N_n(t)] = \sum_{u \in [n]} \mathbb{P}(B_r^{G_n}(u) = t) = n \mathbb{P}(B_r^{G_n}(1) = t),$$  \hspace{1cm} (2.3.22)

where the latter step uses the fact that the distribution of neighborhoods of all vertices is the same. We recall the breadth-first description of an ordered tree in Section 1.5. Let \( v_i \in [n] \) denote the vertex label of the i-th vertex that is explored
We conclude that \( v \) of yet explored, vertices. Then, \( B_r^{(G_n)}(1) = t \) occurs precisely when \((X_i, Y_i) = (x_i, 0)\) for all \(i \in [t]\). Therefore, 
\[
P(B_r^{(G_n)}(1) = t) = P((X_i, Y_i) = (x_i, 0) \forall i \in [t]) = P((X_{[t]}, Y_{[t]}) = (x_{[t]}, 0_{[t]}) \forall i \in [t]),
\]
where we abbreviate \(x_{[t]} = (x_1, \ldots, x_t)\). We condition to write this as 
\[
P(B_r^{(G_n)}(1) = t) = \prod_{i=1}^{t} P((X_i, Y_i) = (x_i, 0) \mid (X_{[i-1]}, Y_{[i-1]}) = (x_{[i-1]}, 0_{[i-1]})).
\]
(2.3.24)

Conditionally on \((X_{[i-1]}, Y_{[i-1]}) = (x_{[i-1]}, 0_{[i-1]})\), for all \(i\) for which \(v_i\) is at distance at most \(r - 1\) from vertex 1, 
\[
X_i \sim \text{Bin}(n_i, \lambda/n),
\]
(2.3.25)
where \(n_i = n - s_{i-1} - i + 1\), and \(X_i = 0\) otherwise. Here, we recall from (1.5.1) that \((s_i)_{i\geq 0}\) satisfies \(s_0 = 1\) and \(s_i = s_{i-1} + x_i - 1\) for \(i \geq 1\). Further, for all \(i \in [t]\), 
\[
Y_i \sim \text{Bin}(s_{i-1} - 1, \lambda/n),
\]
(2.3.26)
since there are \(s_{i-1}\) active vertices, and \(Y_i\) counts the number of edges between \(v_i\) and any other vertex. Finally, \(X_i\) and \(Y_i\) are conditionally independent given \((X_{[i-1]}, Y_{[i-1]}) = (x_{[i-1]}, 0_{[i-1]})\) due to the independence of edges in \(\text{ER}_n(\lambda/n)\). Note that the distance of \(v_i\) from vertex 1 is exactly equal to the distance of the corresponding vertex \(a_i \in V(t)\) to the root \(\emptyset \in V(t)\). Therefore, 
\[
P(B_r^{(G_n)}(1) = t) = \prod_{i \in [t]: \text{dist}(\emptyset, a_i) < r} P(\text{Bin}(n_i, \lambda/n) = x_i) \times \prod_{i \in [t]} (1 - \frac{\lambda}{n})^{s_{i-1}}
\rightarrow \prod_{i \in [t]: \text{dist}(\emptyset, a_i) < r} e^{-\lambda} \frac{\lambda^{x_i}}{x_i!} = \mu(B_r(o) = t).
\]
(2.3.27)

We conclude that 
\[
P(B_r^{(G_n)}(1) = t) \to \mu(B_r(o) = t),
\]
(2.3.28)
so that 
\[
\frac{1}{n} \mathbb{E}[N_n(t)] \to \mu(B_r(o) = t).
\]
(2.3.29)

**Second moment method: second moment**

For the second moment of \(N_n(t)\), we compute 
\[
\mathbb{E}[N_n(t)^2] = \sum_{u_1, u_2 \in [n]} P(B_r^{(G_n)}(u_1) = t, B_r^{(G_n)}(u_2) = t) = n P(B_r^{(G_n)}(1) = t) + n(n - 1) P(B_r^{(G_n)}(1) = t, B_r^{(G_n)}(2) = t).
\]
(2.3.30)
We have already computed the asymptotics of the first term, so we are left with the second term. We claim that
\[ P(B_r^{(G_n)}(1) = t, B_r^{(G_n)}(2) = t) \rightarrow \mu(B_r(o) = t)^2. \tag{2.3.31} \]

We are left to show (2.3.31). Let \( \text{dist}_{G_n}(1, 2) \) denote the graph distance between vertices 1 and 2 in \( G_n = \text{ER}_n(\lambda/n) \). We split
\[ P(B_r^{(G_n)}(1) = t, B_r^{(G_n)}(2) = t) \tag{2.3.32} \]
\[ = P(B_r^{(G_n)}(1) = t, B_r^{(G_n)}(2) = t, \text{dist}_{G_n}(1, 2) > 2r) \]
\[ + P(B_r^{(G_n)}(1) = t, B_r^{(G_n)}(2) = t, \text{dist}_{G_n}(1, 2) \leq 2r). \]

We bound these terms one by one. Using the fact that all vertices are exchangeable, so that vertex 2 has the same distribution as a uniform vertex unequal to vertex 1, the second term in (2.3.32) can be bounded by
\[ P(B_r^{(G_n)}(1) = t, B_r^{(G_n)}(2) = t, \text{dist}_{G_n}(1, 2) \leq 2r) \tag{2.3.33} \]
\[ = E\left[ 1_{\{B_r^{(G_n)}(1) = t\}} \frac{|B_r^{(G_n)}(1)| - 1}{n - 1} \right] \]
\[ \leq \frac{1}{n} E[|B_r^{(G_n)}(1)|]. \]

Write
\[ E[|B_r^{(G_n)}(1)|] = \sum_{k=0}^{2r} E[|\partial B_k^{(G_n)}(1)|]. \tag{2.3.34} \]

It is not hard to show that (see Exercise 2.18)
\[ E[|\partial B_k^{(G_n)}(1)|] \leq \lambda^k. \tag{2.3.35} \]

This implies that
\[ P(B_r^{(G_n)}(1) = t, B_r^{(G_n)}(2) = t, \text{dist}_{G_n}(1, 2) \leq 2r) = o(1). \tag{2.3.36} \]

We rewrite the first term in (2.3.32) as
\[ P(B_r^{(G_n)}(1) = t, \text{dist}_{G_n}(1, 2) > 2r) P(B_r^{(G_n)}(2) = t \mid B_r^{(G_n)}(1) = t, \text{dist}_{G_n}(1, 2) > 2r). \tag{2.3.37} \]

By (2.3.28) and (2.3.36),
\[ P(B_r^{(G_n)}(1) = t, \text{dist}_{G_n}(1, 2) > 2r) \rightarrow \mu(B_r(o) = t). \tag{2.3.38} \]

Further, recall that \( t = |V(t)| \) denotes the number of vertices in \( t \). The event that \( \text{dist}_{G_n}(1, 2) > 2r \) is the same as the event that \( B_r^{(G_n)}(2) \) is disjoint from \( B_r^{(G_n)}(1) \). We note that
\[ P(B_r^{(G_n)}(2) = t \mid B_r^{(G_n)}(1) = t, \text{dist}_{G_n}(1, 2) > 2r) \tag{2.3.39} \]
\[ = P(B_r^{(G_n)}(2) = t \mid \text{in } \text{ER}_{n-t}(\lambda/n)). \]

Since \( t \) is bounded, and as in (2.3.28),
\[ P(B_r^{(G_n)}(2) = t \mid B_r^{(G_n)}(1) = t, \text{dist}_{G_n}(1, 2) > 2r) \rightarrow \mu(B_r(o) = t), \tag{2.3.40} \]
which completes the proof of (2.3.31).
2.4 Consequences of local weak convergence: local functionals

Completion of the proof
The convergence in (2.3.31) implies that $\frac{N_n(t)}{\mathbb{E}[N_n(t)]} \xrightarrow{p} 1,$

$$\frac{N_n(t)}{\mathbb{E}[N_n(t)]} \xrightarrow{p} 1,$$  \hspace{1cm} (2.3.41)

In turn, by (2.3.29), this implies that

$$\frac{1}{n} N_n(t) \xrightarrow{p} \mu(B_r(o) = t),$$  \hspace{1cm} (2.3.42)

which implies that $p^{(\omega)}(t) = \frac{#(t)N_n(t)}{n} \xrightarrow{p} \frac{#(t)\mu(B_r(o) = t)}{\mu(B_r(o) = t)}$ as required.

In Chapters 3, 4 and 5 below, we will extend the above analysis to inhomogeneous random graphs, the configuration model and preferential attachment models. In many cases, the steps taken are a lot like the above. We always combine a second moment method for $N_n(t)$ with explicit computations that allow us to show the appropriate adaptation of (2.3.42).

2.4 Consequences of local weak convergence: local functionals

In this section, we discuss some consequences of local weak convergence that will prove to be useful in the sequel, or describe how network statistics are determined by the local weak limit.

2.4.1 Local weak convergence and convergence of neighborhoods

We start by describing that the number of vertices at distance up to $m$ from a uniform vertex weakly converges to the neighborhood sizes of the limiting rooted graph:

**Corollary 2.17** (Weak convergence of neighborhood sizes) Let $(G_n)_{n \geq 1}$ be a sequence of graphs whose sizes $|V(G_n)|$ tend to infinity.

(a) Assume that $G_n$ converges locally weakly to $(\bar{G}, \bar{o}) \sim \bar{\mu}$ on $\mathcal{G}$. Then, for every $m \geq 1$,

$$\left( |\partial B_r^{(G_n)}(o_n)| \right)_{r=0}^{m} \xrightarrow{d} \left( |\partial B_r^{(\bar{G})}(\bar{o})| \right)_{r=1}^{m},$$  \hspace{1cm} (2.4.1)

(b) Assume that $G_n$ converges locally in probability to $(G, o) \sim \mu$ on $\mathcal{G}$. Then, for every $m \geq 1$, with $o_n^{(1)}, o_n^{(2)}$ two independent uniformly chosen vertices in $V(G_n)$,

$$\left( |\partial B_r^{(G_n)}(o_n^{(1)})|, |\partial B_r^{(G_n)}(o_n^{(2)})| \right)_{r=1}^{m} \xrightarrow{d} \left( |\partial B_r^{(G)}(o^{(1)})|, |\partial B_r^{(G)}(o^{(2)})| \right)_{r=1}^{m},$$  \hspace{1cm} (2.4.2)

where the two limiting neighborhood sizes are independent given $\mu$.

**Proof** Part (a) follows immediately, since the function

$$h(G, o) = 1_{\{ |\partial B_r(o)| = \ell, \forall r \in [m] \}}$$  \hspace{1cm} (2.4.3)
is a bounded continuous function for every $m$ and $\ell_1, \ldots, \ell_m$ (see Exercise 2.21). The proof of part (b) in (2.4.2) follows by noting that
\[
P\left(B_m^{(\infty)}(o_{n}^{(1)}) \simeq t_1, B_m^{(\infty)}(o_{n}^{(2)}) \simeq t_2 \mid G_n\right) = p^{(\infty)}(t_1)p^{(\infty)}(t_2),
\]
by the independence of $o_{n}^{(1)}, o_{n}^{(2)}$. Therefore,
\[
P\left(B_m^{(\infty)}(o_{n}^{(1)}) \simeq t_1, B_m^{(\infty)}(o_{n}^{(2)}) \simeq t_2 \mid G_n\right) \xrightarrow{n \to \infty} \mu(B_m^{(\infty)}(o_{1}^{(1)}) \simeq t_1)\mu(B_m^{(\infty)}(o_{2}^{(2)}) \simeq t_2).
\]
Taking the expectation proves the claim (where you are asked to provide the fine details of this argument in Exercise 2.22).

In the above argument, it is crucial to note that when $\mu$ is a random probability measure on $\mathcal{G}$, that the limits in (2.4.2) are corresponding to two independent $(G,o)$ having law $\mu$, but with the same $\mu$. It is here that the randomness of $\mu$ manifests itself. Recall also the example below Corollary 2.11.

We continue by showing that local weak convergence implies that the graph distance between two uniform vertices tends to infinity:

**Corollary 2.18** (Large distances) Let $(G_n)_{n \geq 1}$ be a sequence of graphs whose sizes $|V(G_n)|$ tend to infinity. Let $o_n^{(1)}, o_n^{(2)}$ be two vertices chosen independently and uniformly at random from $[n]$. Assume that $G_n$ converges in distribution in the local weak sense to $(\bar{G}, \bar{o}) \sim \bar{\mu}$. Then,
\[
dist_{G_n}(o_{n}^{(1)}, o_{n}^{(2)}) \xrightarrow{n \to \infty} \infty.
\]

**Proof** It suffices to prove that, for every $r \geq 1$,
\[
P(\dist_{G_n}(o_{n}^{(1)}, o_{n}^{(2)}) \leq r) = o(1).
\]

For this, we use that $o_{n}^{(2)}$ is chosen uniformly at random from $V(G_n)$ independently of $o_{n}^{(1)}$, so that
\[
P(\dist_{G_n}(o_{n}^{(1)}, o_{n}^{(2)}) \leq r) = \mathbb{E}[|B_r^{(\infty)}(o_{n}^{(1)})|/n] = \mathbb{E}[|B_r^{(\infty)}(o_n)|/n].
\]
By Corollary 2.17, $|B_r^{(\infty)}(o_n)|$ is a tight random variable, so that $|B_r^{(\infty)}(o_n)|/n \xrightarrow{n \to \infty} 0$. Further, $|B_r^{(\infty)}(o_n)|/n \lesssim 1$ a.s. Thus, by Dominated Convergence ([Volume 1, Theorem A.1]), $\mathbb{E}[|B_r^{(\infty)}(o_n)|/n] = o(1)$ for every $r \geq 1$, so that the claim follows.

### 2.4.2 Local weak convergence and clustering coefficients

In this section, we discuss the convergence of various local and global clustering coefficients when a random graph converges locally weakly.

We start by recalling what the global clustering coefficient is, following [Volume 1, Section 1.5]. For a graph $G_n = (V(G_n), E(G_n))$, we let
\[
W_{G_n} = \sum_{i,j,k \in V(G_n)} 1_{\{i,j,k \in E(G_n)\}} = \sum_{v \in V(G_n)} D_v(D_v - 1)
\]
denote two times the number of wedges in the graph $G_n$. The factor of two comes
from the fact that the wedge $ij, jk$ is the same as the wedge $kj, ji$, but it is counted twice in (2.4.9). We further let
\[ \Delta_{G_n} = \sum_{i,j,k \in V(G_n)} \mathbb{1}_{\{i,j,k \in E(G_n)\}} \]  
(2.4.10)
denote three times the number of triangles in $G_n$. The (global) clustering coefficient $CC_{G_n}$ in $G_n$ is defined as
\[ CC_{G_n} = \frac{\Delta_{G_n}}{W_{G_n}}. \]  
(2.4.11)
Informally, the global clustering coefficient measures the proportion of wedges for which the closing edge is also present. As such, it can be thought of as the probability that from a randomly drawn individual and two of its friends, the two friends are friends themselves. The following theorem describes when the clustering coefficient converges:

**Theorem 2.19 (Convergence of global clustering coefficient)** Let $(G_n)_{n \geq 1}$ be a sequence of graphs whose sizes $|V(G_n)|$ tend to infinity. Assume that $G_n$ converges locally in probability to $(G, o) \sim \mu$. Further, assume that $D_n = d_{G_n}^o$ is such that $D_n^2$ is uniformly integrable, and that $\mu(d_n > 1) > 0$. Then
\[ CC_{G_n} \xrightarrow{p} \frac{\mathbb{E}_\mu[\Delta_G(o)]}{\mathbb{E}_\mu[d(o)(d_n - 1)]}, \]  
(2.4.12)
where $\Delta_G(o) = \sum_{u,v \in \partial B_1(o)} \mathbb{1}_{\{(u,v) \in E(G)\}}$ denotes twice the number of triangles in $G$ that contain $o$ as a vertex.

**Proof** We write
\[ CC_{G_n} = \frac{\mathbb{E}[\Delta_{G_n}(o_n) \mid G_n]}{\mathbb{E}[d_{G_n}^o(d_{G_n}^o - 1) \mid G_n]}, \]  
(2.4.13)
where the expectation is with respect to the uniform choice of $o_n \in [n]$, and $d_{G_n}^o$ denotes the degree of $o_n$ in $G_n$, while $\Delta_{G_n}(o_n) = \sum_{u,v \in \partial B_1(o_n)} \mathbb{1}_{\{(u,v) \in E(G_n)\}}$ denotes twice the number of triangles that $o_n$ is part of.

By local weak convergence, $\Delta_{G_n}(o_n) \xrightarrow{d} \Delta_G(o)$ and $d_{G_n}^o(d_{G_n}^o - 1) \xrightarrow{d} d_n(d_n - 1)$. However, both are not bounded functionals, so that the convergence of their expectations over $o_n$ does not follow immediately. It is here that we need to make use of the uniform integrability of $D_n^2$, where $D_n = d_{G_n}^o$. We split
\[ \mathbb{E}[d_{G_n}^o(d_{G_n}^o - 1) \mid G_n] \]  
(2.4.14)
\[ = \mathbb{E}[d_{G_n}^o(d_{G_n}^o - 1)1_{\{d_{G_n}^o \leq K\}} \mid G_n] + \mathbb{E}[d_{G_n}^o(d_{G_n}^o - 1)1_{\{d_{G_n}^o > K\}} \mid G_n]. \]

By local convergence in probability (recall Corollary 2.17),
\[ \mathbb{E}[d_{G_n}^o(d_{G_n}^o - 1)1_{\{d_{G_n}^o \leq K\}} \mid G_n] \xrightarrow{p} \mathbb{E}_\mu[d(o)(d_n - 1)]1_{\{d \leq K\}}, \]  
(2.4.15)
since $h(G, o) = d(o)(d_n - 1)1_{\{d \leq K\}}$ is a bounded continuous function. Further, by uniform integrability of $(d_{G_n}^o)^2$ and with $\mathbb{E}$ denoting the expectation with respect
Then we observe that since $\Delta$ which can be treated as above. This completes the proof.

Then, by the Markov inequality,

$$
\mathbb{P}\left( \mathbb{E}[d_{G_n}^2] \geq \varepsilon \right) \leq \frac{1}{\varepsilon} \mathbb{E}[d_{G_n}^2] \leq \varepsilon.
$$

It follows that $\mathbb{E}[d_{G_n}^2] \to \mathbb{E}_\mu[d_o(d_o - 1)]$, as required. Since $\mu(d_o > 1)$, also $\mathbb{E}_\mu[d_o(d_o - 1)] > 0$, so that also $1/\mathbb{E}[d_{G_n}^2] \to 1/\mathbb{E}_\mu[d_o(d_o - 1)]$.

The proof that $\mathbb{E}_n[\Delta_{G_n}(o_n)] \to \mathbb{E}_\mu[\Delta_G(o)]$ is similar, where now we split

$$
\mathbb{E}[\Delta_{G_n}(o_n) | G_n] = \mathbb{E}[\Delta_{G_n}(o_n)1_{\{d_{G_n}^2 \leq K\}} | G_n] + \mathbb{E}[\Delta_{G_n}(o_n)1_{\{d_{G_n}^2 > K\}} | G_n].
$$

Then, we observe that since $\Delta_{G_n}(o_n) \leq d_{G_n}^2(d_{G_n}^2 - 1)$, again the first term is the expectation of a bounded and continuous functional, and therefore converges in probability. The second term, on the other hand, is bounded by $\mathbb{E}[(d_{G_n}^2)^2]1_{\{d_{G_n}^2 > K\}}$, which can be treated as above. This completes the proof.

We see that in order to obtain convergence of the clustering coefficient, we need an additional uniform integrability condition on the degree distribution. Indeed, more precisely, we need that $D_n^2$ is uniformly integrable, with $D_n = d_{G_n}^2$ the degree of a uniform vertex. This will also be a recurring theme below. We next discuss a related clustering coefficient, where such additional assumptions are not needed. For this, we define the local clustering coefficient for vertex $v \in [n]$ to be

$$
\text{CC}_{G_n}(v) = \frac{\Delta_{G_n}(v)}{d_o(d_o - 1)},
$$

where $\Delta_{G_n}(v) = \sum_{s,t \in \partial B_G(v)} \mathbb{1}_{\{s,t \in E(G_n)\}}$ is again twice the number of triangles that $v$ is part of. Then, we let

$$
\text{CC}_{G_n} = \frac{1}{n} \sum_{v \in [n]} \text{CC}_{G_n}(v)
$$

be the local clustering coefficient. Here, we can think of $\frac{\Delta_{G_n}(v)}{d_o(d_o - 1)}$ as the proportion of edges present between neighbors of $v$, and then (2.4.20) takes the average of this. The following theorem implies its convergence without any further uniform integrability conditions:

**Theorem 2.20 (Convergence of local clustering coefficient)** Let $(G_n)_{n \geq 1}$ be a sequence of graphs whose sizes $|V(G_n)|$ tend to infinity. Assume that $G_n$ converges locally in probability to $(G, o) \sim \mu$. Then

$$
\text{CC}_{G_n} \to \mathbb{E}_\mu\left[\frac{\Delta_G(o)}{d_o(d_o - 1)}\right].
$$
Proof. We now write
\[
\mathbb{C}C_{G_n} = \mathbb{E}\left[\frac{\Delta_{G_n}(o_n)}{d_{o_n}(d_{o_n} - 1)} \mid G_n\right],
\]  
\[\tag{2.4.22}\]
and note that \(h(G,o) = \frac{\Delta_G(o)}{d_o(d_o - 1)}\) is a bounded continuous functional. Therefore,
\[
\mathbb{E}\left[\frac{\Delta_{G_n}(o_n)}{d_{o_n}(d_{o_n} - 1)} \mid G_n\right] \xrightarrow{p} \mathbb{E}\mu\left[\frac{\Delta_G(o)}{d_o(d_o - 1)}\right],
\]  
\[\tag{2.4.23}\]
as required. \[\square\]

There are more versions of the clustering coefficient. A discussion of the convergence of the so-called clustering spectrum can be found in Section 2.6.

2.4.3 Neighborhoods of edges and degree-degree dependencies

In this section, we discuss the convergence of degree-degree dependencies when a random graph converges locally weakly. Often, this is described in terms of the so-called assortativity coefficient. For this, it is crucial to also discuss the convergence of the local neighborhood of a uniformly chosen edge. We will again see that an extra uniform integrability condition is needed for the assortativity coefficient to converge.

We start by defining the neighborhood structure of edges. For this, it will be convenient to consider directed edges. We let \(e = (u,v)\) be an edge directed from \(u\) to \(v\), and let \(\mathcal{E}(G_n) = \{(u,v) : (u,v) \in E(G_n)\}\) denote the collection of directed edges. This will be convenient, as it assigns a (root-)vertex to an edge.

For \(H \in \mathcal{G}_*\), let
\[
p_e^{(G_n)}(H) = \frac{1}{2|\mathcal{E}(G_n)|} \sum_{(u,v) \in \mathcal{E}(G_n)} 1_{\{B_{\mathcal{E}(G_n)}(u) \simeq H\}}.
\]  
\[\tag{2.4.24}\]
Note that
\[
p_e^{(G_n)}(H) = \mathbb{P}(B_{\mathcal{E}(G_n)}(e) \simeq H \mid G_n),
\]  
\[\tag{2.4.25}\]
where \(e = (\xi, \xi)\) is a uniformly chosen directed edge from \(\mathcal{E}(G_n)\). Thus, \(p_e^{(G_n)}(H)\) is the edge-equivalent of \(p^{(G_n)}(H)\) in (2.2.5). We next study its asymptotics:

**Theorem 2.21** (Convergence neighborhoods of edges) Let \((G_n)_{n \geq 1}\) be a sequence of graphs whose sizes \(|V(G_n)|\) tend to infinity. Assume that \(G_n\) converges locally in probability to \((G,o) \sim \mu\). Assume that \((d_{o_n}^{(G_n)})_{n \geq 1}\) is a uniformly integrable sequence of random variables, and that \(\mu(d_o \geq 1) > 0\). Then, for every \(H \in \mathcal{G}_*\),
\[
p_e^{(G_n)}(H) \xrightarrow{p} \frac{\mathbb{E}_\mu[d_o 1_{\{B_{\mathcal{E}(G)}(o) \simeq H\}}]}{\mathbb{E}_\mu[d_o]}.
\]  
\[\tag{2.4.26}\]
Therefore, since \((d_{\mu_n}^{G_n})_{n \geq 1}\) is uniformly integrable, local convergence in probability implies that

\[
\frac{2}{|V(G_n)|} |E(G_n)| \xrightarrow{p} \mathbb{E}_\mu[d_o].
\]  

(2.4.28)

Since \(\mu(d_o \geq 1) > 0\), it follows that \(\mathbb{E}_\mu[d_o] > 0\).

Further, we rewrite

\[
\frac{1}{|V(G_n)|} \sum_{(u,v) \in E(G_n)} \mathbb{1}_{\{B_n^{(G_n)}(u) \geq H_s\}} = \frac{1}{|V(G_n)|} \sum_{u \in V(G_n)} d_{\mu_n}^{G_n} \mathbb{1}_{\{B_n^{(G_n)}(u) \geq H_s\}} \quad (2.4.29)
\]

\[
= \mathbb{E} \left[ d_{\mu_n}^{G_n} \mathbb{1}_{\{B_n^{(G_n)}(o) \geq H_s\}} | G_n \right].
\]

where \(o_n\) is uniformly chosen in \(V(G_n)\). Again, since \((d_{\mu_n}^{G_n})_{n \geq 1}\) is uniformly integrable and by local convergence in probability,

\[
\mathbb{E} \left[ d_{\mu_n}^{G_n} \mathbb{1}_{\{B_n^{(G_n)}(o) \geq H_s\}} | G_n \right] \xrightarrow{p} \mathbb{E}_\mu[d_o \mathbb{1}_{\{B(G) \geq H_s\}}].
\]  

(2.4.30)

Therefore, by (2.4.31), taking the ratio of the terms in (2.4.28)–(2.4.30) proves the claim.

We continue by considering the degree-degree distribution, again following [Volume 1, Section 1.5]. For this, it is convenient to consider directed edges \(e = (u, v)\), and write \(\overrightarrow{e} = u\) and \(\overleftarrow{e} = v\) for the starting and endpoint of \(e\). The degree-degree distribution is given by

\[
p_{k,l}^{(G_n)} = \frac{1}{2|E(G_n)|} \sum_{e} \mathbb{1}_{\{d_{\mu_n}^{G_n} = k, d_{\mu_n}^{G_n} = l\}}.
\]  

(2.4.31)

Thus, \(p_{k,l}^{(G_n)}\) is the probability that a random directed edge connects a vertex of degree \(k\) with one of degree \(l\). By convention, we define \(p_{k,0}^{(G_n)} = 0\) when \(k = 0\). The following theorem proves that the degree-degree distribution converges when the graph locally converges in probability:

**Theorem 2.22 (Degree-degree convergence)** Let \((G_n)_{n \geq 1}\) be a sequence of graphs whose sizes \(|V(G_n)|\) tend to infinity. Let \(G_n\) converge locally in probability to \((G, o) \sim \mu\). Assume that \((d_{\mu_n}^{G_n})_{n \geq 1}\) is a uniformly integrable sequence of random variables, and that \(\mu(d_o \geq 1) > 0\). Then, for every \(k, l\) with \(k \geq 1\),

\[
p_{k,l}^{(n)} \xrightarrow{p} k\mu(d_o = k, d_V = l),
\]  

(2.4.32)

where \(V\) is a neighbor of \(o\) chosen uniformly at random.
2.4 Consequences of local weak convergence: local functionals

Proof Recall (2.4.28). We rewrite
\[
\frac{1}{|V(G_n)|} \sum_e \mathbbm{1}\{d^{(G_n)}_e=k,d^{(G_n)}_v=l\} = k\frac{1}{n} \sum_{u \in [n]} \mathbbm{1}\{d_u=k\} \left( \frac{1}{k} \sum_{v : v \sim u} \mathbbm{1}\{d^{(G_n)}_v=l\} \right) 
\]
where \( V \) is a uniformly chosen neighbor of \( o \), which itself is uniformly chosen in \([n]\). Again, by local weak convergence in probability and the fact that the distribution \( V \) is a neighbor of \( o \). Therefore, by (2.4.31), taking the ratio of the terms in (2.4.28)–(2.4.34) proves the claim.

We next discuss the consequences for the assortativity coefficient (recall [Volume 1, Section 1.5]). We now write the degrees in \( G_n \) as \((d_o,d_v)\) conditionally on \((G,o)\) is a deterministic function of \( B^{(2)}_2(n)\),
\[
\frac{1}{|V(G_n)|} \sum_e \mathbbm{1}\{d^{(G_n)}_e=k,d^{(G_n)}_v=l\} \xrightarrow{p} k\mu(d_o=k,d_v=l).
\]

Therefore, by (2.4.31), taking the ratio of the terms in (2.4.28)–(2.4.34) proves the claim.

Theorem 2.23 (Assortativity convergence) Let \((G_n)_{n \geq 1}\) be a sequence of graphs whose sizes \(|V(G_n)|\) tend to infinity. Assume that \( G_n \) converges locally in probability to \((G,o) \sim \mu\). Assume that \( D_n = d^{(G_n)}_o \) is such that \( D_n \) is uniformly integrable, and that \( \mu(d_o \geq 1) > 0 \). Then,
\[
\rho_{G_n} \xrightarrow{p} \frac{\mathbb{E}_\mu[d^2_o d_V] - \mathbb{E}_\mu[d^2_o] \mathbb{E}_\mu[d_v]}{\mathbb{E}_\mu[d^2_o] - \mathbb{E}_\mu[d^2_v] \mathbb{E}_\mu[d_o]},
\]
where \( V \) is a neighbor of \( o \) chosen uniformly at random.

Proof We start with (2.4.36), and consider the various terms. We divide all sums
by \( n \). Then, by local weak convergence in probability and uniform integrability of \((d_{o_n}^{(G_n)})^3\), which implies that also \(d_{o_n}^{(G_n)}\) is uniformly integrable,

\[
\frac{1}{n} \mathbb{E}(G_n) = \mathbb{E}[d_{o_n}^{(G_n)} | G_n] \longrightarrow \mathbb{E}_\mu[d_o]. \tag{2.4.38}
\]

Again by local weak convergence and uniform integrability of \((d_{o_n}^{(G_n)})^3\), which implies that also \((d_{o_n}^{(G_n)})^2\) is uniformly integrable,

\[
\frac{1}{|V(G_n)|} \sum_{i \in V(G_n)} d_i^2 = \mathbb{E}[(d_{o_n}^{(G_n)})^2 | G_n] \longrightarrow \mathbb{E}_\mu[d_o^2]. \tag{2.4.39}
\]

Further, again by local weak convergence in probability and uniform integrability of \((d_{o_n}^{(G_n)})^3\),

\[
\frac{1}{|V(G_n)|} \sum_{i \in V(G_n)} d_i^3 = \mathbb{E}[(d_{o_n}^{(G_n)})^3 | G_n] \longrightarrow \mathbb{E}_\mu[d_o^3]. \tag{2.4.40}
\]

This identifies the limits of all but one of the sums appearing in (2.4.36). Details are left to the reader in Exercise 2.19.

We finally consider the last term involving the cross terms of the degrees across edges, i.e.,

\[
\frac{1}{|V(G_n)|} \sum_{ij \in E(G_n)} d_i d_j = \frac{1}{|V(G_n)|} \sum_{u \in V(G_n)} d_u^2 \left( \frac{1}{d_{u \sim v}} \sum_{v \sim u} d_v \right) \tag{2.4.41}
\]

\[= \mathbb{E}_{nu}[d_{o_n}^2 d_{V}], \]

where \( V \) is a random neighbor of \( o_n \). When the degrees are uniformly bounded, the functional \( h(G,o) = d_o^2 \mathbb{E}[d_V | G] \) is bounded and continuous, so that it would converge. However, the degrees are not necessarily bounded, so a truncation argument is needed.

We split

\[
\frac{1}{n} \sum_{ij \in E(G_n)} d_i d_j = \frac{1}{n} \sum_{ij \in E(G_n)} d_i d_j 1_{\{d_i \leq K, d_j \leq K\}} + \frac{1}{n} \sum_{ij \in E(G_n)} d_i d_j (1 - 1_{\{d_i \leq K, d_j \leq K\}}). \tag{2.4.42}
\]

We now rewrite, similarly as above,

\[
\frac{1}{n} \sum_{ij \in E(G_n)} d_i d_j 1_{\{d_i \leq K, d_j \leq K\}} = \mathbb{E}[d_{o_n}^2 d_V 1_{\{d_{o_n} \leq K, d_V \leq K\}} | G_n]. \tag{2.4.43}
\]

By local convergence in probability (or by Theorem 2.22), since the functional is now bounded and continuous,

\[
\frac{1}{|V(G_n)|} \sum_{ij \in E(G_n)} d_i d_j 1_{\{d_i \leq K, d_j \leq K\}} \longrightarrow \mathbb{E}_\mu[d_{o_n}^2 d_V 1_{\{d_{o_n} \leq K, d_V \leq K\}}]. \tag{2.4.44}
\]
We are left to show that the second contribution in (2.4.42) is small. We bound this contribution as
\[
\frac{1}{|V(G_n)|} \sum_{ij \in E(G_n)} d_i d_j (\mathbb{1}_{\{d_i > K\}} + \mathbb{1}_{\{d_j > K\}}) = \frac{2}{|V(G_n)|} \sum_{ij \in E(G_n)} d_i d_j \mathbb{1}_{\{d_i > K\}}.
\]
(2.4.45)

We now use Cauchy-Schwarz to bound this as
\[
\frac{1}{|V(G_n)|} \sum_{ij \in E(G_n)} d_i d_j (\mathbb{1}_{\{d_i > K\}} + \mathbb{1}_{\{d_j > K\}}) \leq \frac{2}{|V(G_n)|} \sqrt{\sum_{ij \in E(G_n)} d_i^2 \mathbb{1}_{\{d_i > K\}}} \sqrt{\sum_{ij \in E(G_n)} d_j^2}
\]
\[
= 2E[(d_{G_n}^{(G_n)})^3 | 3 \mathbb{1}_{\{d_{G_n}^{(G_n)}>K\}} | G_n]^{1/2}E[(d_{G_n}^{(G_n)})^3 | G_n]^{1/2}.
\]

By uniform integrability of $(d_{G_n}^{(G_n)})^3$, there exists $K = K(\varepsilon)$ and $N = N(\varepsilon)$ such that, for all $n \geq N$,
\[
E[(d_{G_n}^{(G_n)})^3 | \mathbb{1}_{\{d_{G_n}^{(G_n)}>K\}}] \leq \varepsilon^4/4.
\]
(2.4.47)

In turn, by the Markov inequality, this implies that
\[
P\left(E[(d_{G_n}^{(G_n)})^3 | \mathbb{1}_{\{d_{G_n}^{(G_n)}>K\}} > \varepsilon^4/4 | G_n] \right) \leq \frac{4}{\varepsilon^3} E[(d_{G_n}^{(G_n)})^3 | \mathbb{1}_{\{d_{G_n}^{(G_n)}>K\}}] \leq \varepsilon.
\]
(2.4.48)

As a result, with probability at least $1 - \varepsilon$ and for $\varepsilon > 0$ sufficiently small to accommodate the factor $E_n[(d_{G_n}^{(G_n)})^3]^{1/2}$ (which is uniformly bounded by the uniform integrability of $(d_{G_n}^{(G_n)})^3$),
\[
\frac{1}{|V(G_n)|} \sum_{ij \in E(G_n)} d_i d_j (\mathbb{1}_{\{d_i > K\}} + \mathbb{1}_{\{d_j > K\}}) \leq \varepsilon^3/2 E[(d_{G_n}^{(G_n)})^3 | G_n]^{1/2} \leq \varepsilon.
\]
(2.4.49)

This completes the proof.

\[\square\]

2.4.4 PageRank and local weak convergence

Recall the definition of PageRank from [Volume 1, Section 1.5]. Here we restrict our discussion to undirected graphs, see Section 9.2 for a discussion of the more interesting setting of directed graphs. Let $G_n = (V(G_n), E(G_n))$ denote an undirected graph where every vertex has degree at least one. Then, we let the vector of PageRanks $(R_i)_{i \in [n]}$ be the unique solution to the equation
\[
R_i = \alpha \sum_{\{i,j\} \in E(G_n)} \frac{R_j}{d_j} + 1 - \alpha,
\]
(2.4.50)
satisfying the normalization
\[
\sum_{i \in V(G_n)} R_i = |V(G_n)|.
\]
(2.4.51)
The parameter $\alpha \in (0, 1)$ is called the damping factor, and guarantees that (2.4.50) has a unique solution. This solution can be understood in terms of the stationary distribution of a bored surfer. Indeed, denote $\pi_i = R_i/n$, so that $(\pi_i)_{i \in [n]}$ is a probability distribution that satisfies a similar relation as $(R_i)_{i \in [n]}$ in (2.4.50), namely

$$\pi_i = \alpha \sum_{\{i,j\} \in E(G)} \frac{\pi_j}{d_j} + \frac{1 - \alpha}{|V(G_n)|},$$

Therefore, $(\pi_i)_{i \in V(G_n)}$ is the stationary distribution of a random walker that, with probability $\alpha$ jumps according to a simple random walk, i.e., it chooses any of the edges with equal probability, while with probability $1 - \alpha$, the walker jumps to a uniform location.

The damping factor $\alpha$ is quite crucial. When $\alpha = 0$, the stationary distribution is just $\pi_i = 1/n$ for every $i$, so that all pages have PageRank 1. This is not very informative. On the other hand, PageRank concentrates on dangling ends when $\alpha$ is close to one, and this is also not what we want. Experimentally, $\alpha = 0.85$ seems to work well and strikes a nice balance between these two extremes.

We next investigate the convergence of the PageRank distribution on an undirected graph sequence $(G_n)_{n \geq 1}$ that converges locally weakly:

**Theorem 2.24** (Existence of asymptotic PageRank distribution) Consider a sequence of undirected random graphs $(G_n)_{n \in \mathbb{N}}$ whose size tends to infinity.

(i) If $G_n$ converges locally weakly, then there exists a limiting distribution $R_\emptyset$, with $E_\mu[R_\emptyset] \leq 1$, such that

$$R_{(G_n)} \xrightarrow{d} R_\emptyset.$$  

(ii) If $G_n$ converges locally in probability, then there exists a limiting distribution $R_\emptyset$, with $E_\mu[R_\emptyset] \leq 1$, such that, for every $r > 0$,

$$\frac{1}{|V(G_n)|} \sum_{v \in V(G_n)} 1_{\{R_{(G_n)}(v) > r\}} \xrightarrow{P} \mu(R_\emptyset > r).$$

See Section 9.2 for a discussion of the more interesting setting of directed graphs. There, we also extend the discussion of local weak convergence to directed graphs, which is not as trivial as one might expect.

### 2.5 The giant component is almost local

We continue by investigating the size of the giant component when the graph converges in the local weak sense. Here, we will simplify the notation by assuming that $G_n = (V(G_n), E(G_n))$ is such that $|V(G_n)| = n$. 
2.5 The giant component is almost local

2.5.1 Asymptotics of the giant

Clearly, the proportion of vertices in the largest connected component $|C_{\text{max}}|/n$ is not continuous in the local convergence topology (see Exercise 2.23), as it is a global object. In fact, also $|C(o_n)|/n$ does not converge in distribution when $(G_n, o_n) \xrightarrow{d} (G, o)$. However, local convergence still tells us a useful story about the existence of a giant, as well as its size:

**Corollary 2.25** (Upper bound on the giant) Let $(G_n)_{n \geq 1}$ be a sequence of graphs whose sizes $|V(G_n)| = n$ tend to infinity. Assume that $G_n$ converges locally in probability to $(G, o) \sim \mu$. Write $\zeta = \mu(\{|C(o)| = \infty\})$ for the survival probability of the limiting graph $(G, o)$. Then, for every $\varepsilon > 0$ fixed,

$$\mathbb{P}( |C_{\text{max}}| \leq n(\zeta + \varepsilon)) \to 1. \quad (2.5.1)$$

In particular, Corollary 2.25 implies that $|C_{\text{max}}|/n \xrightarrow{p} 0$ when $\zeta = 0$ (see Exercise 2.24).

**Proof** Define

$$Z_{\geq k} = \sum_{v \in V(G_n)} 1_{\{|C(v)| \geq k\}}. \quad (2.5.2)$$

Assume that $G_n$ converges locally in probability to $(G, o)$. Then, we conclude that with $\zeta_{\geq k} = \mu(|C(o)| \geq k)$ (see Exercise 2.25),

$$\frac{Z_{\geq k}}{n} \xrightarrow{p} \zeta_{\geq k}. \quad (2.5.3)$$

For every $k \geq 1$,

$\{|C_{\text{max}}| \geq k\} = \{Z_{\geq k} \geq k\}, \quad (2.5.4)$

and, on the event that $Z_{\geq k} \geq 1$, also $|C_{\text{max}}| \leq Z_{\geq k}$. Thus, for every $k \geq 1$ and every $\varepsilon > 0$, and all $n$ large enough,

$$\mathbb{P}( |C_{\text{max}}| \geq n(\zeta_{\geq k} + \varepsilon/2)) = o(1). \quad (2.5.5)$$

Therefore, with $\zeta = \lim_{k \to \infty} \zeta_{\geq k} = \mu(|C(o)| = \infty)$, and for every $\varepsilon > 0$,

$$\mathbb{P}( |C_{\text{max}}| \geq n(\zeta + \varepsilon)) = o(1). \quad (2.5.6)$$

We conclude that while local convergence cannot determine the size of the largest connected component, it can prove an upper bound on $|C_{\text{max}}|$. In this book, we will often extend this to $|C_{\text{max}}|/n \xrightarrow{p} \zeta$, but this is no longer a consequence of local convergence alone. In Exercise 2.26, you are asked to give an example where $|C_{\text{max}}|/n \xrightarrow{p} a < \zeta$. Therefore, in general, more involved arguments must be used. We next prove that one, relatively simple, condition suffices:
Theorem 2.26 (The giant is almost local) Let $G_n = ([n], E(G_n))$ denote a finite (possibly disconnected) random graph. Assume that $G_n$ converges locally in probability to $(G, o) \sim \mu$. Assume that

$$\lim_{k \to \infty} \limsup_{n \to \infty} \frac{1}{n} \mathbb{E} \left[ \# \{(x, y) \in V(G_n): |\mathcal{C}(x)|, |\mathcal{C}(y)| \geq k, x \leftrightarrow y \} \right] = 0. \quad (2.5.7)$$

Then, with $\mathcal{C}_{\max}$ and $\mathcal{C}_{(2)}$ denoting the largest and second largest connected components (with ties broken arbitrarily),

$$\frac{|\mathcal{C}_{\max}|}{n} \xrightarrow{p} \zeta = \mu(|\mathcal{C}(o)| = \infty), \quad \frac{|\mathcal{C}_{(2)}|}{n} \xrightarrow{p} 0. \quad (2.5.8)$$

Theorem 2.26 shows that a relatively mild condition as in (2.5.7) suffices for the giant to have the expected limit. In fact, it is necessary and sufficient, as you can see in Exercise 2.28. Theorem 2.26 will be most useful when we can easily show that vertices with large clusters are likely to be connected.

We now start with the proof of Theorem 2.26. We first note that, by Corollary 2.25, it suffices to prove Theorem 2.26 for $\zeta > 0$, which we assume from now on.

We recall that the vector $(|\mathcal{C}(i)|)_{i \geq 1}$ denotes the cluster sizes ordered in size, from large to small with ties broken arbitrarily, so that $\mathcal{C}_{(1)} = \mathcal{C}_{\max}$. The following lemma gives a useful estimate on the sum of squares of these ordered cluster sizes. In its statement, we write $X_{n,k} = o_{k,p}(1)$ when

$$\lim_{k \to \infty} \limsup_{n \to \infty} \mathbb{P}(|X_{n,k}| > \varepsilon) = 0. \quad (2.5.9)$$

Lemma 2.27 (Convergence of sum of squares of cluster sizes) Under the conditions of Theorem 2.26,

$$\frac{1}{n^2} \sum_{i \geq 1} |\mathcal{C}(i)|^2 \xrightarrow{p} \zeta^2, \quad (2.5.10)$$

and

$$\frac{1}{n^2} \sum_{i \geq 1} |\mathcal{C}(i)|^2 \mathbb{1}_{(|\mathcal{C}(i)| \geq k)} = \zeta^2 + o_{k,p}(1). \quad (2.5.11)$$

Proof We use that, by local convergence in probability and for any $k \geq 1$ fixed (recall (2.5.3))

$$\frac{1}{n} \sum_{i \geq 1} |\mathcal{C}(i)| \mathbb{1}_{(|\mathcal{C}(i)| \geq k)} = \frac{1}{n} \sum_{v \in [n]} \mathbb{1}_{(|\mathcal{C}(v)| \geq k)} = \frac{1}{n} \mathcal{Z}_{\geq k} \xrightarrow{p} \zeta_{\geq k}, \quad (2.5.12)$$

where we recall that $\zeta_{\geq k} = \mathbb{P}(|\mathcal{C}(o)| \geq k)$. Then we obtain that

$$\frac{1}{n} \sum_{i \geq 1} |\mathcal{C}(i)| \mathbb{1}_{(|\mathcal{C}(i)| \geq k)} = \zeta + o_{k,p}(1). \quad (2.5.13)$$
Further,
\[
\frac{1}{n^2} \sum_{i,j \geq 1 \atop i \neq j} |\mathcal{E}_{(i)}||\mathcal{E}_{(j)}|1_{\{|\mathcal{E}_{(i)}|,|\mathcal{E}_{(j)}| \geq k\}}.
\]

By the Markov inequality,
\[
\lim_n \limsup_{k \to \infty} \mathbb{P} \left( \frac{1}{n^2} \sum_{i,j \geq 1 \atop i \neq j} |\mathcal{E}_{(i)}||\mathcal{E}_{(j)}|1_{\{|\mathcal{E}_{(i)}|,|\mathcal{E}_{(j)}| \geq k\}} \right) \geq \varepsilon \] (2.5.15)

by our main assumption in (2.5.7). As a result,
\[
\frac{1}{n^2} \sum_{i,j \geq 1 \atop i \neq j} |\mathcal{E}_{(i)}|^2 1_{\{|\mathcal{E}_{(i)}| \geq k\}} = \frac{1}{n^2} \sum_{i \geq 1} |\mathcal{E}_{(i)}|^2 1_{\{|\mathcal{E}_{(i)}| \geq k\}} + o_{k,p}(1) (2.5.16)
\]

We conclude that
\[
\frac{1}{n^2} \sum_{i \geq 1} |\mathcal{E}_{(i)}|^2 1_{\{|\mathcal{E}_{(i)}| \geq k\}} = \left( \frac{1}{n} \sum_{i \geq 1} |\mathcal{E}_{(i)}|1_{\{|\mathcal{E}_{(i)}| \geq k\}} \right)^2 + o_{k,p}(1) (2.5.17)
\]

by (2.5.12). This proves (2.5.11). Finally,
\[
\frac{1}{n^2} \sum_{i \geq 1} |\mathcal{E}_{(i)}|^2 1_{\{|\mathcal{E}_{(i)}| < k\}} \leq \frac{k}{n} \frac{1}{n^2} \sum_{i \geq 1} |\mathcal{E}_{(i)}|^2 1_{\{|\mathcal{E}_{(i)}| \geq k\}} \leq \frac{k}{n}, \quad (2.5.18)
\]

so that
\[
\frac{1}{n^2} \sum_{i \geq 1} |\mathcal{E}_{(i)}|^2 = \frac{1}{n^2} \sum_{i \geq 1} |\mathcal{E}_{(i)}|^2 1_{\{|\mathcal{E}_{(i)}| \geq k\}} + O \left( \frac{k}{n} \right), \quad (2.5.19)
\]

which, together with (2.5.11), completes the proof of (2.5.10). □

We are now ready to complete the proof of Theorem 2.26, and we start by explaining the ideas behind this. Define the (possibly random) probability measure \((q_{i,n})_{i \geq 1}\) by
\[
q_{i,n} = \frac{|\mathcal{E}_{(i)}|1_{\{|\mathcal{E}_{(i)}| \geq k\}}}{\sum_{j \geq 1} |\mathcal{E}_{(j)}|1_{\{|\mathcal{E}_{(j)}| \geq k\}}} \quad (2.5.20)
\]

By (2.5.12), \(\sum_{j \geq 1} |\mathcal{E}_{(j)}|1_{\{|\mathcal{E}_{(j)}| \geq k\}}/n = \zeta + o_{k,p}(1)\), and thus, by (2.5.11), \(q_{i,n}^2 = 1 + o_{k,p}(1)\). We conclude that we have a probability mass function such that its sum of squares is close to 1. This is only possible when \(\max_{i \geq 1} q_{i,n} = 1 + o_{k,p}(1)\). Since \(i \mapsto q_{i,n}\) is non-increasing, this means that
\[
q_{1,n} = \frac{|\mathcal{E}_{\text{max}}|1_{\{|\mathcal{E}_{\text{max}}| \geq k\}}}{\sum_{j \geq 1} |\mathcal{E}_{(j)}|1_{\{|\mathcal{E}_{(j)}| \geq k\}}} = 1 + o_{k,p}(1),
\]

which, together with \(\sum_{j \geq 1} |\mathcal{E}_{(j)}|1_{\{|\mathcal{E}_{(j)}| \geq k\}} = n\zeta(1+o_{k,p}(1))\) implies that \(|\mathcal{E}_{\text{max}}|/n \to \)
ζ. Further, \( q_2,n = o_{k,p}(1) \), which implies that \( |\mathcal{G}(o)|/n \xrightarrow{p} 0 \). This is what we aim to prove. We now fill in the details:

**Proof of Theorem 2.26.** By Lemma 2.27,

\[
\frac{1}{n^2} \sum_{i \geq 1} |\mathcal{G}(i)|^2 \mathbb{1}_{\{|\mathcal{G}(i)| \geq k\}} = \zeta^2 + o_{k,p}(1).
\]  

(2.5.21)

Denote \( x_{i,n} = |\mathcal{G}(i)| \mathbb{1}_{\{|\mathcal{G}(i)| \geq k\}} / n \). We wish to show that \( x_{1,n} = \zeta + o_{k,p}(1) \). Clearly \( x_{1,n} \leq \zeta + o_{k,p}(1) \). Since \( x_{1,n} \) is bounded, it has a subsequence that converges in probability. Now suppose by contradiction that, along such a subsequence, that \( x_{1,n} = |\mathcal{G}(1)| \mathbb{1}_{\{|\mathcal{G}(1)| \geq k\}} / n \xrightarrow{p} a_k \), where \( a_k \to a = \zeta - \delta \) for some \( \delta > 0 \), so that \( x_{1,n} = \zeta - \delta + o_{k,p}(1) \). Let us work along this subsequence. Then, also using (2.5.21),

\[
\sum_{i \geq 2} (x_{i,n})^2 = \sum_{i \geq 1} x_{i,n}^2 - x_{1,n}^2 = \zeta^2 - (\zeta - \delta)^2 + o_{k,p}(1) = \delta(2\zeta - \delta) + o_{k,p}(1). \tag{2.5.22}
\]

Also, \( \sum_{i \geq 2} x_{i,n} = \delta + o_{k,p}(1) \). As a result,

\[
\delta^2 + o_{k,p}(1) = \left( \sum_{i \geq 2} x_{i,n} \right)^2 = \sum_{i \geq 2} x_{i,n}^2 + \sum_{i,j \geq 2, i \neq j} x_{i,n}x_{j,n}. \tag{2.5.23}
\]

Clearly, by our main assumption (2.5.7),

\[
\sum_{i,j \geq 2, i \neq j} x_{i,n}x_{j,n} \leq \left( \sum_{i \geq 2} x_{i,n} \right)^2 - \sum_{i \geq 1} x_{i,n}^2 = o_{k,p}(1), \tag{2.5.24}
\]

so that combining (2.5.22)–(2.5.24) leads to \( \delta^2 = \delta(2\zeta - \delta) \) which gives \( \delta = \zeta \), which implies \( x_{1,n} = o_{k,p}(1) \). By (2.5.11) and \( x_{i,n} \leq x_{1,n} \), we get \( \zeta = 0 \), contradicting the assumption \( \zeta > 0 \) in the statement of the theorem. □

### 2.5.2 Properties of the giant

We next extend Theorem 2.26 somewhat, and investigate the structure of the giant. For this, we first let \( v_k(\mathcal{G}_{\text{max}}) \) denote the number of vertices with degree \( k \) in the giant component, and we recall that \( E(\mathcal{G}_{\text{max}}) \) denotes the number of edges in the giant component:

**Theorem 2.28** (Properties of the giant) Under the assumptions of Theorem 2.26, when \( \zeta = \mu(|\mathcal{G}(o)| = \infty) > 0 \),

\[
\frac{v_k(\mathcal{G}_{\text{max}})}{n} \xrightarrow{p} \mu(|\mathcal{G}(o)| = \infty, d_o = \ell). \tag{2.5.25}
\]

Further, assume that \( D_n = d^{(n)}_o \) is uniformly integrable. Then,

\[
\frac{|E(\mathcal{G}_{\text{max}})|}{n} \xrightarrow{p} \frac{1}{2} \mathbb{E}_n \left[ d_o \mathbb{1}_{|\mathcal{G}(o)| = \infty} \right]. \tag{2.5.26}
\]
Assume that \( G_n \) converges in probability in the local weak sense to \((G,o)\). Then, we conclude that with \( \zeta_{A,>k} = \mu(|\mathcal{E}(o)| \geq k, d_o \in A) \),

\[
\frac{Z_{A,>k}}{n} \xrightarrow{\mathbb{P}} \mu(|\mathcal{E}(o)| \geq k, d_o \in A).
\]  

(2.5.28)

Since \( |\mathcal{E}_{\max}| \geq k \) whp, we thus obtain, for every \( A \subseteq \mathbb{N} \),

\[
\frac{1}{n} \sum_{a \in A} v_a(\mathcal{E}_{\max}) \leq \frac{Z_{A,>k}}{n} \xrightarrow{\mathbb{P}} \mu(|\mathcal{E}(o)| \geq k, d_o \in A).
\]  

(2.5.29)

Applying this to \( A = \{\ell\}^c \), we obtain that, for all \( \varepsilon > 0 \),

\[
\lim_{n \to \infty} \mathbb{P}\left( \frac{1}{n} |\mathcal{E}_{\max} - v_\ell(\mathcal{E}_{\max})| \leq \mu(|\mathcal{E}(o)| \geq k, d_o \neq \ell) + \varepsilon/2 \right) = 1. \tag{2.5.30}
\]

We argue by contradiction. Suppose that, for some \( \ell \),

\[
\liminf_{n \to \infty} \mathbb{P}\left( \frac{v_\ell(\mathcal{E}_{\max})}{n} \leq \mu(|\mathcal{E}(o)| = \infty, d_o = \ell) - \varepsilon \right) = \kappa > 0.
\]  

(2.5.31)

Then, along the subsequence \((n_l)_{l \geq 1}\) that attains the liminf in (2.5.31), with strictly positive probability, and using (2.5.30),

\[
\frac{|\mathcal{E}_{\max}|}{n} = \frac{1}{n} |\mathcal{E}_{\max} - v_\ell(\mathcal{E}_{\max})| + \frac{v_\ell(\mathcal{E}_{\max})}{n} \leq \mu(|\mathcal{E}(o)| \geq k) - \varepsilon/2,
\]  

(2.5.32)

which contradicts Theorem 2.26. We conclude that (2.5.31) cannot hold, so that (2.5.25) follows.

For (2.5.26), we note that

\[
|E(\mathcal{E}_{\max})| = \frac{1}{2} \sum_{\ell \geq 1} v_\ell(\mathcal{E}_{\max}).
\]  

(2.5.33)

We divide by \( n \) and split the sum over \( \ell \) into small and large \( \ell \) as

\[
\frac{|E(\mathcal{E}_{\max})|}{n} = \frac{1}{2n} \sum_{\ell \in [K]} v_\ell(\mathcal{E}_{\max}) + \frac{1}{2n} \sum_{\ell > K} v_\ell(\mathcal{E}_{\max}).
\]  

(2.5.34)

For the first term in (2.5.34), by (2.5.25),

\[
\frac{1}{2n} \sum_{\ell \in [K]} v_\ell(\mathcal{E}_{\max}) \xrightarrow{\mathbb{P}} \frac{1}{2} \sum_{\ell \in [K]} \mu(|\mathcal{E}(o)| = \infty, d_o = \ell)
\]  

\[
= \frac{1}{2} \mathbb{E}_\mu \left[ d_o \mathbf{1}_{\{|\mathcal{E}(o)| = \infty, d_o \in [K]\}} \right].
\]  

(2.5.35)

For the second term in (2.5.34), we bound, with \( n_\ell \) the number of vertices in \( G_n \) of degree \( \ell \),

\[
\frac{1}{2n} \sum_{\ell > K} v_\ell(\mathcal{E}_{\max}) \leq \frac{1}{2} \sum_{\ell > K} \frac{n_\ell}{n} = \frac{1}{2} \mathbb{E}\left[ [d^\infty_\ell] \mathbf{1}_{\{d^\infty_\ell > K\}} \mid G_n \right].
\]  

(2.5.36)
By uniform integrability,

$$\lim_{K \to \infty} \limsup_{n \to \infty} \mathbb{E}[d_{on}^n \mathbf{1}_{\{d_{on}^n > K\}}] = 0.$$  \hspace{1cm} (2.5.37)

As a result, for every $\varepsilon > 0$, there exists a $K = K(\varepsilon) < \infty$ such that

$$
\mathbb{P}\left( \mathbb{E}_n [d_{on}^n \mathbf{1}_{\{d_{on}^n > K\}}] > \varepsilon \right) \to 0.
$$  \hspace{1cm} (2.5.38)

This completes the proof of (2.5.26).

It is not hard to extend the above analysis to the local convergence in probability of the giant, as well as its outside, as formulated in the following theorem:

**Theorem 2.29** (Local limit of the giant)  
Under the assumptions of Theorem 2.26, when $\zeta = \mu(|\mathcal{C}(o)| = \infty) > 0$,

$$
\frac{1}{n} \sum_{v \in \mathcal{C}_{\text{max}}} 1_{\{B(G_n)(v) = H_+\}} \xrightarrow{p} \mu(|\mathcal{C}(o)| = \infty, B_r^{(o)}(o) \simeq H_+),
$$  \hspace{1cm} (2.5.39)

and

$$
\frac{1}{n} \sum_{v \notin \mathcal{C}_{\text{max}}} 1_{\{B(G_n)(v) = H_+\}} \xrightarrow{p} \mu(|\mathcal{C}(o)| < \infty, B_r^{(o)}(o) \simeq H_+).
$$  \hspace{1cm} (2.5.40)

**Proof**  
The convergence in (2.5.40) follows from that in (2.5.39) combined with the fact that, by assumption,

$$
\frac{1}{n} \sum_{v \in V(G_n)} \mathbf{1}_{\{B_r^{(G_n)}(v) \in H_+\}} \xrightarrow{p} \mu(B_r^{(o)}(o) \simeq H_+).
$$  \hspace{1cm} (2.5.41)

The convergence in (2.5.40) can be proved as for Theorem 2.28, now using that

$$
\frac{1}{n} Z_{\mathcal{H}_k} = \frac{1}{n} \sum_{v \in [n]} \mathbf{1}_{(|\mathcal{C}(v)| \geq k, B_r^{(G_n)}(v) \in \mathcal{H}_k)} \xrightarrow{p} \mu(|\mathcal{C}(o)| \geq k, B_r^{(o)}(o) \in \mathcal{H}_k),
$$  \hspace{1cm} (2.5.42)

and, since $|\mathcal{C}_{\text{max}}|/n \xrightarrow{p} \zeta > 0$,

$$
\frac{1}{n} \sum_{v \in V(G_n)} \mathbf{1}_{B_r^{(G_n)}(v) \in \mathcal{H}_k} \leq Z_{\mathcal{H}_k}.
$$  \hspace{1cm} (2.5.43)

We leave the details to the reader.  \hspace{1cm} \square

**2.5.3 The condition (2.5.7) revisited**

The condition (2.5.7) is sometimes not so easy to verify, and we now give an alternative form that is often easier to check. That is the content of the following lemma:
2.5 The giant component is almost local

Lemma 2.30 (Condition (2.5.7) revisited) Under the assumptions of Theorem 2.26, the condition in (2.5.7) holds when

\[
\lim_{r \to \infty} \limsup_{n \to \infty} \frac{1}{n^2} \mathbb{E} \left[ \# \left\{ (x, y) \in V(G_n) : |\partial B_r(x)|, |\partial B_r(x)| \geq r, x \leftrightarrow y \right\} \right] = 0,
\]

(2.5.44)

when there exists \( r = r_k \to \infty \) such that

\[
\mu(|\mathcal{C}(o)| \geq k, |\partial B_r(o)| < r) \to 0, \quad \mu(|\mathcal{C}(o)| < k, |\partial B_r(o)| \geq r) \to 0.
\]

(2.5.45)

Proof Denote

\[
P_k = \# \left\{ (x, y) \in V(G_n) : |\mathcal{C}(x)|, |\mathcal{C}(y)| \geq k, x \leftrightarrow y \right\},
\]

(2.5.46)

\[
P^{(2)}_k = \# \left\{ (x, y) \in V(G_n) : |\partial B_r(x)|, |\partial B_r(x)| \geq r, x \leftrightarrow y \right\}.
\]

(2.5.47)

Then,

\[
|P_r - P^{(2)}_k| \leq 2n[Z_{r, \geq k} + Z_{r, < k}],
\]

(2.5.48)

where

\[
Z_{r, \geq k} = \sum_{v \in V(G_n)} 1_{\{|\partial B_r(v)| < r, |\mathcal{C}(v)| \geq k\}}, \quad Z_{r, < k} = \sum_{v \in V(G_n)} 1_{\{|\partial B_r(v)| \geq r, |\mathcal{C}(v)| < k\}}.
\]

(2.5.49)

Therefore, by local convergence in probability,

\[
\frac{1}{n^2} |P_k - P^{(2)}_k| \leq \frac{2}{n} \left[ Z_{r, \geq k} + Z_{r, < k} \right]
\]

(2.5.50)

\[
\overset{\mathcal{L}}{\to} 2\mu(|\mathcal{C}(o)| \geq k, |\partial B_r(o)| < r) + 2\mu(|\mathcal{C}(o)| < k, |\partial B_r(o)| \geq r).
\]

Take \( r = r_k \) as in (2.5.45). Then, the right hand side vanishes, so that, by Dominated Convergence [Volume 1, Theorem A.1] also

\[
\lim_{k \to \infty} \limsup_{n \to \infty} \frac{1}{n^2} \mathbb{E}[|P_k - P^{(2)}_k|] = 0.
\]

(2.5.51)

We arrive at

\[
\lim_{k \to \infty} \limsup_{n \to \infty} \frac{1}{n^2} \mathbb{E}[P_k] \leq \lim_{k \to \infty} \limsup_{n \to \infty} \frac{1}{n^2} \mathbb{E}[P^{(2)}_k] = 0,
\]

(2.5.52)

by (2.5.44) and since \( r_k \to \infty \) when \( k \to \infty \).

The assumption in (2.5.45) is often easily verified. For example, for the Erdős-Rényi random graph ER_n(\( \lambda/n \)) with \( \lambda > 1 \), to which we will apply it below, we can take \( r = k \) and use that on the event of survival (recall [Volume 1, Theorem 3.9]),

\[
\lambda^{-r} |\partial B^{(c)}_r(o)| \overset{a.s.}{\to} M.
\]

(2.5.53)

Here \((G, o)\) denotes a Poisson branching process with mean \( \lambda \) offspring, and where \( M > 0 \) on the event of survival by [Volume 1, Theorem 3.10]. Therefore, \( \mu(|\mathcal{C}(o)| \geq k, |\partial B^{(c)}_r(o)| < k) \to 0 \) as \( k \to \infty \). Further, \( \mu(|\mathcal{C}(o)| < k, |\partial B^{(c)}_r(o)| \geq k) = 0 \) trivially. However, there are examples where (2.5.45) fails, and then also the equivalence of (2.5.7) and (2.5.44) may be false.
2.5.4 The giant in Erdős-Rényi random graphs

In this section, we use the local convergence in probability of $\text{ER}_n(\lambda/n)$ in Theorem 2.16, combined with the fact that the ‘giant is almost local’ in Theorem 2.26, to identify the phase transition and size of the giant in the Erdős-Rényi random graph $\text{ER}_n(\lambda/n)$:

**Theorem 2.31** (Phase transition Erdős-Rényi random graph) Fix $\lambda > 0$, and let $C_{\text{max}}$ be the largest connected component of the Erdős-Rényi random graph $\text{ER}_n(\lambda/n)$, and $C_{(2)}$ the second largest connected component (breaking ties arbitrarily). Then,

\[
\frac{|C_{\text{max}}|}{n} \xrightarrow{p} \zeta_\lambda, \quad \frac{|C_{(2)}|}{n} \xrightarrow{p} 0,
\]

where $\zeta_\lambda$ is the survival probability of a Poisson branching process with mean offspring $\lambda$. In particular, $\zeta > 0$ precisely when $\lambda > 1$.

Further, for $\lambda > 0$, with $\eta_\lambda = 1 - \zeta_\lambda$ and for all $\ell \geq 0$,

\[
\frac{\nu_\ell(C_{\text{max}})}{n} \xrightarrow{p} e^{-\lambda} \frac{\lambda^\ell}{\ell!} [1 - \eta_\lambda^\ell],
\]

and

\[
\frac{E(C_{\text{max}})}{n} \xrightarrow{p} \frac{1}{2} \lambda [1 - \eta_\lambda^2].
\]

The law of large numbers for the giant in Theorem 2.31 for $\lambda > 1$ has also been proved in [Volume 1, Theorem 4.8], where a more precise bound was given on the convergence rate. There, the proof was given using explicit computations, here we show that it follows rather directly from local weak convergence considerations. We refer to [Volume 1, Section 4.6] for a discussion of the history of the phase transition for $\text{ER}_n(\lambda/n)$.

**Proof** The main work will reside in showing that the condition (2.5.7) in Theorem 2.26 holds. In turn, (2.5.7) in Theorem 2.26 can be replaced by (2.5.44) in Lemma 2.5.44, which turns out to be more convenient as stated there.

Indeed, local convergence in probability follows from Theorem 2.16. Then, the claim in (2.5.54) follows directly from Theorem 2.26, while (2.5.55)–(2.5.56) follow from Theorem 2.28, and the observations that, for a Poisson branching process with mean offspring $\lambda > 1$,

\[
\mu(|C(o)| = \infty, d_o = \ell) = e^{-\lambda} \frac{\lambda^\ell}{\ell!} [1 - \eta_\lambda^\ell],
\]

and thus

\[
E_{\mu} \left[ d_o \mathbb{1}_{|C(o)| = \infty} \right] = \sum_\ell \ell \mu(|C(o)| = \infty, d_o = \ell) = \sum_\ell \ell e^{-\lambda} \frac{\lambda^\ell}{\ell!} [1 - \eta_\lambda^\ell] = \lambda [1 - \eta_\lambda e^{-\lambda(1 - \eta_\lambda)}] = \lambda [1 - \eta_\lambda^2],
\]
2.5 The giant component is almost local

since \( \eta_\lambda \) satisfies \( \eta_\lambda = e^{-\lambda(1-\eta_\lambda)} \). Therefore, we are left to show that \( (2.5.44) \) holds.

The proof proceeds in several steps.

**Step 1: Concentration of binomials**

Note that, with \( o_1, o_2 \in [n] \) chosen independently and uniformly at random,

\[
\frac{1}{n^2} \mathbb{E} \left[ \# \{ (x, y) \in V(G_n) : |\partial B_r(x)|, |\partial B_r(x)| \geq r, x \leftrightarrow y \} \right] = \mathbb{P}(|\partial B_r(o_1)|, |\partial B_r(o_2)| \geq r, o_1 \leftrightarrow o_2).
\]

We let \( \mathbb{P}_r \) denote the conditional distribution of \( \text{ER}_n(\lambda/n) \) given \( |\partial B_r(o_1)| = b^{(i)}_0 \) with \( b^{(i)}_0 \geq r \), and \( |B_{r-1}(o_1)| = s^{(i)}_0 \), so that

\[
\mathbb{P}(|\partial B_r(o_1)|, |\partial B_r(o_2)| \geq r, o_1 \leftrightarrow o_2) = \sum_{b^{(i)}_0, b^{(2)}_0, s^{(i)}_0} \mathbb{P}_r(o_1 \leftrightarrow o_2) \mathbb{P}(|\partial B_r(o_1)| = b^{(i)}_0, |B_{r-1}(o_1)| = s^{(i)}_0, i \in \{1, 2\}).
\]

Our aim is to show that, for every \( \varepsilon > 0 \), we can find \( r = r_\varepsilon \) such that, for every \( b^{(i)}_0, b^{(2)}_0 \geq r \) and \( s^{(i)}_0, s^{(2)}_0 \) fixed,

\[
\limsup_{n \to \infty} \mathbb{P}_r(o_1 \leftrightarrow o_2) \leq \varepsilon.
\]

Under \( \mathbb{P}_r \),

\[
|\partial B_{r+1}(o_1)| \sim \text{Bin}(n^{(i)}_1, p^{(i)}_1),
\]

where

\[
n^{(i)}_1 = n - b^{(i)}_0 - s^{(i)}_0 - s^{(2)}_0, \quad p^{(i)}_1 = 1 - \left(1 - \frac{\lambda}{n}\right)^{b^{(i)}_0}.
\]

Here, we note that the vertices in \( \partial B_r(o_2) \) play a different role than those in \( \partial B_r(o_1) \), as they can be in \( \partial B_{r+1}(o_1) \), but those in \( \partial B_r(o_1) \) cannot. This explains the slightly asymmetric form with respect to vertices 1 and 2 in \( (2.5.63) \). We are lead to studying concentration properties of binomial random variables. For this, we will rely on the following lemma, which follows from [Volume 1, Theorem 2.21]:

**Lemma 2.32** (Concentration binomials)  Let \( X \sim \text{Bin}(m, p) \). Then, for every \( \delta > 0 \),

\[
\mathbb{P}(|X - \mathbb{E}[X]| \geq \delta \mathbb{E}[X]) \leq 2 \exp \left( -\frac{\delta^2 \mathbb{E}[X]}{2(1 + \delta/3)} \right).
\]

**Proof** This is a direct consequence of [Volume 1, Theorem 2.21].

Lemma 2.32 ensures that whp the boundary of \( |\partial B_{r+1}(o_1)| \) is close to \( \lambda|\partial B_r(o_1)| \), so that the boundary grows by a factor \( \lambda > 1 \). Further applications lead to the statement that \( |\partial B_{r+k}(o_1)| \approx \lambda^k |\partial B_r(o_1)| \). Thus, in roughly \( a \log_a n \) steps, the boundary will have expanded to \( n^a \) vertices. However, in order to make this precise, we need that (1) the sum of complementary probabilities in Lemma 2.32 is
still quite small; and (2) we have good control over the number of vertices in the boundaries, not just in terms of lower bounds, but also in terms of upper bounds, as that will give control over the number of vertices that have not yet been used. For the latter, we will also need to deal with the $\varepsilon$-dependence in (2.5.64).

We will prove (2.5.61) by first growing $|\partial B_{r+k}(o_1)|$ for $k \geq 1$ until $|\partial B_{r+k}(o_1)|$ is very large (much larger than $\sqrt{n}$ will do), and then, outside of $B_{r+k}(o_1)$ for the appropriate $k$, growing $|\partial B_{r+k}(o_2)|$ for $k \geq r$ until also $|\partial B_{r+k}(o_2)|$ is very large (now $\sqrt{n}$ will do). Then, it is very likely that there is a direct edge between the resulting boundaries. We next provide the details.

**Step 2: Erdős-Rényi neighborhood growth of first vertex**

To make the above analysis precise, we start by introducing some notation. We let $b_k^{(1)} = b_0^{(1)}[\lambda(1 + \varepsilon)]^k$ and $b_k^{(1)} = b_0^{(1)}[\lambda(1 - \varepsilon)]^k$ denote the upper and lower bounds on $|\partial B_{r+k}(o_1)|$, and let

$$s_k^{(i)} = s_0^{(i)} + \sum_{l=0}^{k-1} b_l^{(i)}$$

(2.5.65)

denote the resulting upper bound on $|B_{r+k-1}(o_1)|$. We let $k \leq k_\varepsilon = k_\varepsilon(n) = a \log_{\lambda(1-\varepsilon)} n$, and note that

$$s_k^{(i)} \leq \frac{b_0^{(i)}}{1 - \lambda(1 + \varepsilon)} n^{\lambda(1+\varepsilon)/(1-\varepsilon)}$$

(2.5.66)

uniformly in $k \leq k_\varepsilon$. We will choose $a \in (\frac{1}{2}, 1)$ so that $a(1+\varepsilon)/(1-\varepsilon) \in (\frac{1}{2}, 1)$.

Define the good event by

$$E_{r,[k]}^{(i)} = \bigcap_{l \in [k]} E_{r,l}^{(i)}, \quad \text{where} \quad E_{r,k}^{(i)} = \{b_k^{(i)} \leq |\partial B_{r+k}(o_1)| \leq b_k^{(i)} \}.$$ (2.5.67)

We then bound, for $k = k_\varepsilon$,

$$\mathbb{P}_r(E_{r,[k]}^{(i)}) = \prod_{l \in [k]} \mathbb{P}_r\left( E_{r,l}^{(i)} \mid E_{r,[l]}^{(i)} \right),$$

(2.5.68)

so that

$$\mathbb{P}_r(E_{r,[k]}^{(i)}) \geq 1 - \sum_{l \in [k]} \mathbb{P}_r\left( (E_{r,l}^{(i)})^c \mid E_{r,[l]}^{(i)} \right).$$

(2.5.69)

With the above choices, we note that, conditionally on $|\partial B_{r+l-1}(o_1)| = b_{l-1}^{(i)} \in [b_{l-1}^{(i)}, \overline{b}_{l-1}^{(i)}]$ and $|\partial B_{r+l-1}(o_1)| = s_{l-1}^{(i)} \leq s_{l-1}^{(i)}$,

$$|\partial B_{r+l}(o_1)| \sim \text{Bin}(n_l^{(i)}, p_l^{(i)}),$$

(2.5.70)

where

$$n_l^{(i)} = n - s_{l-1}^{(i)} - s_0^{(i)}, \quad p_l^{(i)} = 1 - \left(1 - \frac{\lambda}{n}\right)^{b_{l-1}^{(i)}}.$$ (2.5.71)
Further, \( n_1 \) will denote the corresponding neighborhoods by \( (b_{l-1}^{(1)} \lambda / n)^2 \leq \lambda b_{l-1}^{(1)} \). Therefore,
\[ n_1^{(1)} p_1^{(1)} \leq n b_{l-1}^{(1)} \lambda / n = \lambda b_{l-1}^{(1)}, \]
which provides the upper bound on \( n_1^{(1)} p_1^{(1)} \). For the lower bound, we use the lower bound in \((2.5.72)\) to note that \( p_1^{(1)} \geq (1 - \varepsilon/4) \lambda b_{l-1}^{(1)} / n \), since we are on \( E_{r,l}^{(1)} \).
Further, \( n_1^{(1)} \geq (1 - \varepsilon/4)n \) on \( E_{r,l}^{(1)} \), uniformly in \( l \leq k \). We conclude that
\[ n_1^{(1)} p_1^{(1)} \geq (1 - \varepsilon/4)^2 n b_{l-1}^{(1)} \lambda / n \geq (1 - \varepsilon/2) \lambda b_{l-1}^{(1)}. \]
By Lemma 2.32 with \( \delta = \varepsilon/2 \), therefore,
\[ \mathbb{P}_r \left( (E_{r,l}^{(1)})^c \mid E_{r,l}^{(1)} \right) \]
\[ \leq \mathbb{P}_r \left( \left| \partial B_{r+l}(o_1) \right| - \mathbb{E}[\left| \partial B_{r+l}(o_1) \right| \mid E_{r,l}^{(1)}] \right) \geq (\varepsilon/2) \mathbb{E}[\left| \partial B_{r+l}(o_1) \right| \mid E_{r,l}^{(1)}] \]
\[ \leq 2 \exp \left( -\varepsilon^2 (1 - \varepsilon/2) \lambda b_{l-1}^{(1)} / 8(1 + \varepsilon/6) \right) = 2 \exp \left( -q \lambda b_{l-1}^{(1)} \right), \]
where \( q = \varepsilon^2 (1 - \varepsilon/2)/[8(1 + \varepsilon/6)] > 0. \)
We conclude that
\[ \mathbb{P}_r \left( (E_{r,l}^{(1)})^c \right) \leq 1 - 2 \sum_{i=1}^{k-1} e^{-q \lambda b_{l-1}^{(1)}}. \]

**Step 3: Erdős-Rényi neighborhood growth of second vertex**

We next grow the neighborhoods from vertex 2 in a similar way, and we focus on the differences only. In the whole argument below, we condition on \( \left| \partial B_{r+l}(o_1) \right| = b_{l}^{(1)} \in [\lambda b_{l-1}^{(1)}, \overline{\lambda b_{l-1}^{(1)}}] \) and \( \left| B_{r+l}(o_1) \right| = s_{l}^{(1)} \leq \overline{s}_{l}^{(1)} \) for all \( l \leq k \). Further, rather than exploring \( (\partial B_{r+k}(o_2))_{k \geq 0} \), we will explore these neighborhoods outside of \( B_{r+k}(o_1) \), and we will denote the corresponding neighborhoods by \( (\partial B'_{r+k}(o_2))_{k \geq 0} \).
We define
\[ E_{r,k}^{(2)} = \bigcap_{l \leq k} E_{r,l}^{(2)}, \text{ where } E_{r,k}^{(2)} = \{b_k^{(2)} \leq \left| \partial B'_{r+k}(o_2) \right| \leq \overline{b}_k^{(2)} \}. \]
We then note that, conditionally on the above, as well as on \( \left| \partial B'_{r+k-1}(o_2) \right| = b_{k-1}^{(2)} \in [\overline{b}_{k-1}^{(2)}, \overline{b}_{k-1}^{(2)}] \) and \( \left| B'_{r+k-1}(o_2) \right| = s_{k-1}^{(2)} \leq \overline{s}_{k-1}^{(2)} \),
\[ \left| \partial B'_{r+k}(o_2) \right| \sim \text{Bin}(n_k^{(2)}, p_k^{(2)}), \]
where now
\[ n_k^{(2)} = n - s_{k-1}^{(2)} - s_{k-1}^{(2)}, \quad p_k^{(2)} = 1 - \left( 1 - \frac{\lambda}{n} \right) b_{k-1}^{(2)}. \]
Denote the \( q = \varepsilon^2 (1 - \varepsilon/2)/[8(1 + \varepsilon/6)] > 0 \).

**Completion of the proof of Theorem 2.31**

We use Lemma 2.5.44, and conclude that we need to show (2.5.61). Recall (2.5.80), and that \( b_0^{(i)} = b_0^{(i)} [\lambda(1 - \varepsilon)]^k \), where \( b_0^{(i)} \geq r = r_x \). Fix \( k = k_z = \log_{\lambda(1-\varepsilon)} n \) as above. For this, \( k \), take \( r = r_x \) so large that

\[
\sum_{l=1}^{k-1} \left[ e^{-q\lambda^{(i)}(1)} + e^{-q\lambda^{(i)}(2)} \right] \leq \varepsilon/2.
\]

Denote the **good event** by

\[
\mathcal{E}_{r_x, [k_z]} = \mathcal{E}_{r_x, [k_z]}^{(1)} \cap \mathcal{E}_{r_x, [k_z]}^{(2)},
\]

where we recall (2.5.67) and (2.5.77). Then,

\[
\lim_{n \to \infty} \mathbb{P}_r \left( \mathcal{E}_{r_x, [k_z]} \right) \geq 1 - \varepsilon.
\]

On \( \mathcal{E}_{r_x, [k_z]} \), we have that \( |\partial B_{r_x+k_z}(a_1)| \geq b_0^{(i)} \geq r_x n^a \), where we choose \( a \in (\frac{1}{2}, 1) \) so that \( a \in (\frac{1}{2}, 1) \). An identical bound holds for \( |\partial B_{r_x+k_z}(a_2)| \).

We use that, on \( \mathcal{E}_{r_x, [k_z]} \),

\[
|\partial B_{r_x+k_z}(a_1)| \geq b_0^{(i)} \geq r_x \left[ \lambda(1 - \varepsilon) \right]^{k_z} = r_x n^a,
\]

and the same bound holds for \( |\partial B_{r_x+k_z}(a_2)| \). Therefore, the number of direct edges between \( \partial B_{r_x+k_z}(a_1) \) and \( \partial B_{r_x+k_z}(a_2) \) is at least \( (r_x n^a)^2 \gg n \) when \( a > \frac{1}{2} \). Therefore,

\[
\mathbb{P}_r \left( \text{dist}_{\text{E}r, n} (a_1, a_2) > 2(k_z + r_x) + 1 \mid \mathcal{E}_{r_x, [k_z]} \right) \leq \left( 1 - \frac{\lambda}{n} \right)^{(r_x n^a)^2} = o(1).
\]

We conclude that

\[
\mathbb{P}_r \left( \text{dist}_{\text{E}r, n} (a_1, a_2) \leq 2(k_z + r_x) + 1 \mid \mathcal{E}_{r_x, [k_z]} \right) = 1 - o(1).
\]

Thus, by (2.5.83),

\[
\mathbb{P}_r (a_1 \leftrightarrow a_2) \leq \mathbb{P}_r (\mathcal{E}_{r_x, [k_z]}^c) + \mathbb{P}_r (a_1 \leftrightarrow a_2 \mid \mathcal{E}_{r_x, [k_z]}) \leq \varepsilon/2 + \varepsilon/2 \leq \varepsilon.
\]

Since \( \varepsilon > 0 \) is arbitrary, the claim in (2.5.61) follows.\[\square\]
2.5 The giant component is almost local

The small-world nature of $\text{ER}_n(\lambda/n)$

In the above proof, we have also identified an upper bound on the typical distance in $\text{ER}_n(\lambda/n)$, that is, the graph distance in $\text{ER}_n(\lambda/n)$ between $o_1$ and $o_2$, as formulated in the following theorem:

**Theorem 2.33** (Small-world nature Erdős-Rényi random graph) Consider $\text{ER}_n(\lambda/n)$ with $\lambda > 1$. Then, conditionally on $o_1 \leftrightarrow o_2$,

$$\lim_{n \to \infty} \frac{\text{dist}_{\text{ER}_n(\lambda/n)}(o_1, o_2)}{\log n} = \frac{1}{\log \lambda}. \quad (2.5.88)$$

**Proof** The lower bound follows directly from (2.3.35), which implies that

$$P(\text{dist}_{\text{ER}_n(\lambda/n)}(o_1, o_2) \leq k) = \mathbb{E}\left[\left|B_k(o_1)\right|/n\right] \leq \sum_{l=0}^{k} \frac{\lambda^l}{n} = \frac{\lambda^{k+1} - 1}{n(\lambda - 1)}. \quad (2.5.89)$$

Applying this to $k = (1 - \varepsilon) \log \lambda n$ shows that

$$P(\text{dist}_{\text{ER}_n(\lambda/n)}(o_1, o_2) \leq (1 - \varepsilon) \log \lambda n) \to 0 \quad (2.5.90)$$

For the upper bound, we start by noting that

$$P(o_1 \leftrightarrow o_2 \mid \text{ER}_n(\lambda/n)) = \frac{1}{n^2} \sum_{i \geq 1} |\mathcal{E}_{i,k}|^2 \xrightarrow{p} \zeta^2, \quad (2.5.91)$$

by (2.5.10) in Lemma 2.27. Thus, conditioning on $o_1 \leftrightarrow o_2$ is asymptotically the same as $o_1, o_2 \in \mathcal{E}_{\max}$. The upper bound then follows from the fact that the event $\mathcal{E}_{\varepsilon, [k]}$ in (2.5.82) holds with probability at least $1 - \varepsilon$, and, on the event $\mathcal{E}_{\varepsilon, [k]}$,

$$\text{dist}_{\text{ER}_n(\lambda/n)}(o_1, o_2) \leq 2(k\varepsilon + r\varepsilon) + 1, \quad (2.5.92)$$

whp by (2.5.86).

2.5.5 Outline of the remainder of this book

The results proved for the Erdős-Rényi random graph complete the preliminaries for this book in Part I. They further allow us to provide a brief glimpse into the content of the remainder of this book. There, we will prove results about local convergence, the existence of the giant component, and the small-world nature of various random graph models. We focus on inhomogeneous random graphs such as the generalized random graph, on the configuration model and its related uniform random graph with prescribed degrees, and the preferential attachment model.

In Part II, consisting of Chapters 3, 4 and 5, we focus on local convergence as in Theorem 2.16, but then applied to these models. Further, we investigate the size of the giant component as in Theorems 2.26 and 2.31. The proofs are often more involved than the corresponding results for the Erdős-Rényi random graph, since these random graphs models are inhomogeneous, and often also lack
the independence of the edge statuses. Therefore, the proofs will require detailed knowledge of the models in question.

In Part III, consisting of Chapters 6, 7 and 8, we focus on the small-world nature of these random graph models, as in Theorem 2.33. It turns out that the exact scaling of the graph distances depends on level of inhomogeneity present in the random graph model. In particular, we will see that when the second moment of the degrees remains bounded, then graph distances grow logarithmically as in Theorem 2.33. If, on the other hand, the second moment blows up with the graph size, then distances are smaller. In particular, these typical distances are doubly logarithmic when the degrees obey a power-law with exponent $\tau$ that satisfies $\tau \in (2,3)$, so that even a moment of order $2 - \varepsilon$ is infinite for some $\varepsilon > 0$. Anyone who has done some numerical work will realize that there, in practice, is little difference between $\log \log n$ and a constant, even when $n$ is quite large.

One of the main conclusions of the local convergence results in Part II is that the most popular random graph models for inhomogeneous real-world networks are locally tree-like, in that the majority of neighborhoods of vertices have no cycles. In many settings this is not realistic. Certainly in social networks, many triangles and even cliques of larger sizes exist. Therefore, in Part IV, consisting of Chapter 9, we investigate some adaptations of the models discussed in Parts II and III. These models may incorporate clustering, community structure, may be directed or be living in a geometric space. All these aspects have received tremendous attention in the literature. Therefore, with Part IV in hand, the reader will be able to access the literature more easily.

2.6 Notes and discussion

Notes on Section 2.1

The discussion in this and the following section follow Aldous and Steele (2004) and Benjamini and Schramm (2001). We refer to Appendix A.2 for proofs of various properties of the metric $d_{\mathcal{G}}$ on rooted graphs, including the fact that it turns $\mathcal{G}$ into a Polish space.

Notes on Section 2.2

Various generalizations of local weak convergence are possible. For example, Aldous and Steele (2004) introduce the notion of geometric rooted graphs, which are rooted graphs where each edge $e$ receives a weight $\ell(e)$, turning the rooted graph into a metric space itself. Benjamini, Lyons and Schramm (2015) allow for more general marks. These marks are associated to the vertices as well as the edges, and can take values in a general complete separable metric space, which are called marked graphs. Such more general set ups are highly relevant in many applications, for example when dealing with general inhomogeneous graphs. Let us explain this in some more detail. A marked graph is a (multi-)graph $G = (V(G), E(G))$ together with a complete separable metric space $\Xi$,
called the mark space. Here Ξ maps from \( V(G) \) and \( E(G) \) to \( \Xi \). Images in \( \Xi \) are called marks. Each edge is given two marks, one associated to (‘at’) each of its endpoints. The only assumption on degrees is that they are locally finite. We omit the mark maps from our notation for networks. We next extend the metric to the above setting of marked graphs. Let the distance between \((G_1, o_1)\) and \((G_2, o_2)\) be \(1/(1 + \alpha)\), where \(\alpha\) is the supremum of those \(r > 0\) such that there is some rooted isomorphism of the balls of (graph-distance) radius \(\lfloor r \rfloor\) around the roots of \(G_i\), such that each pair of corresponding marks has distance less than \(1/r\). For probability measures \(\mu_n, \mu\) on \(\mathcal{G}\), we write \(\mu_n \xrightarrow{d} \mu\) as \(n \to \infty\) when \(\mu_n\) converges weakly to \(\mu\) with respect to this metric. See Exercise 2.17 for an application of marked graphs to directed graphs.

Aldous and Lyons (2007) also study the implications for stochastic processes, such as percolation and random walks, on unimodular graphs.

The tightness statement in Theorem 2.7 is (Benjamini et al., 2015, Theorem 3.1). Benjamini et al. (2015) use the term network instead of a marked graph. We avoid the term networks here, as it may cause confusion with the complex networks in the real world that form inspiration for this book. A related proof can be found in Angel and Schramm (2003).

Aldous (1991) investigates local weak convergence in the context of finite trees, and calls the trees rooted at a uniform vertex fringe trees. For fringe trees, the uniform integrability of the degree of a random vertex is equivalent to tightness of the resulting tree in the local weak sense (see (Aldous, 1991, Lemma 4(ii))).

Dembo and Montanari (2010b) define a version of local weak convergence in terms of convergence of subgraph counts (see (Dembo and Montanari, 2010b, Definition 2.1)), and also states that this holds for several models, including ER\(_n(\lambda/n)\) and CM\(_n(d)\) under appropriate conditions, while Dembo and Montanari (2010a) provides more details. See e.g., (Dembo and Montanari, 2010a, Lemma 2.4) for a proof for the configuration model, and (Dembo and Montanari, 2010a, Proposition 2.6) for a proof for the Erdős-Rényi random graph.

Notes on Section 2.3

We thank Shankar Bhamidi for useful discussions about possibly random limits in Definition 2.9. As far as we could tell, a precise definition of local convergence in its various senses has not appeared explicitly in the literature.

An intuitive analysis of \(\mathbb{P}(B_t^{G_n}(1) = t)\) in (2.3.28) has already appeared in [Volume 1, Section 4.1.2], see in particular [Volume 1, (4.1.12)].

Notes on Section 2.4

While many of the results in this section are ‘folklore’, finding appropriate references for the results stated is not obvious.

Theorems 2.19 and 2.20 discuss the convergence of two clustering coefficients. In the literature, also the clustering spectrum has attracted attention. For this, we recall that \(n_k\) denotes the number of vertices with degree \(k\) in \(G_n\), and define
Local convergence of random graphs

the clustering coefficient of vertices of degree \( k \) to be

\[
    c_{G_n}(k) = \frac{1}{n_k} \sum_{v \in [n]: d^{(n)}_v = k} \frac{\Delta_{G_n}(v)}{k(k-1)}.
\]  

(2.6.1)

It is not hard to adapt the proof of Theorem 2.20 to show that, under its assumptions, \( c_{G_n}(k) \xrightarrow{P} c_G(k) \), where

\[
    c_G(k) = E \mu \left[ \frac{\Delta_G(o)}{k(k-1)} \mid d_o = k \right].
\]  

(2.6.2)

See Exercise 2.32.

The convergence of the assortativity coefficient in Theorem 2.23 is restricted to degree distributions that have uniformly integrable third moments. In general, an empirical correlation coefficient needs a finite variance of the random variables to converge to the correlation coefficient. Litvak and the author (see van der Hofstad and Litvak (2014) and (2013)) proved that when the random variables do not have finite variance, such convergence (even for an i.i.d. sample) can be to a proper random variable, that has support containing a subinterval of \([-1, 0]\) and a subinterval of \([0, 1]\), giving problems in the interpretation.

For networks, \( \rho \) in (2.4.36) is always well defined, and gives a value in \([-1, 1]\). However, also for networks there is a problem with this definition. Indeed, van der Hofstad and Litvak (2014) and (2013) prove that if a limiting value of \( \rho_{G_n} \) exists for a sequence of networks and the third moment of the degree of a random vertex is not uniformly integrable, then \( \liminf_{n \to \infty} \rho_{G_n} \geq 0 \), so no asymptotically disassortative graph sequences exist. Naturally, other ways of classifying the degree-degree dependence can be proposed, such as the correlation of their ranks. Here, for a sequence of numbers \( x_1, \ldots, x_n \) with ranks \( r_1, \ldots, r_n \), \( x_i \) is the \( r_i \)th largest of \( x_1, \ldots, x_n \). Ties tend to be broken by giving random ranks for the equal values. For practical purposes, maybe a scatter plot of the values might be the most useful way to gain insight into degree-degree dependencies.

Notes on Section 2.5

The results in this section are new.

2.7 Exercises for Chapter 2

Exercise 2.1 (Graph isomorphisms fix vertex and edge numbers)  Assume that \( G_1 \simeq G_2 \). Show that \( G_1 \) and \( G_2 \) have the same number of vertices and edges.

Exercise 2.2 (Graph isomorphisms fix degree sequence)  Let \( G_1 \) and \( G_2 \) be two finite graphs. Assume that \( G_1 \simeq G_2 \). Show that \( G_1 \) and \( G_2 \) have the same degree sequences. Here, for a graph \( G \), we let the degree sequence be \( (p_k(G))_{k \geq 0} \), where

\[
    p_k(G) = \frac{1}{|V(G)|} \sum_{v \in V(G)} \mathbb{1}_{\{d_v(G) = k\}}, \tag{2.7.1}
\]
where $d_v(G)$ is the degree of $v$ in $G$.

**Exercise 2.3** (Distance to rooted graph ball) Let the ball $B_r(o)$ around $o$ in the graph $G$ be defined as in (2.1.1). Show that $d_{G_r}(B_r(o),(G,o)) \leq 1/(r+1)$. When does equality hold?

**Exercise 2.4** (Countable number of graphs with bounded radius) Fix $k$. Show that there is a countable number of isomorphism classes of rooted graphs $(G,o)$ with radius at most $r$. Here, we let the radius $\text{rad}(G,o)$ of a rooted graph $(G,o)$ be equal to $\text{rad}(G,o) = \max_{v \in V(G)} d_G(o,v)$.

**Exercise 2.5** $(G_\star$ is separable) Use Exercise 2.4 above to show that the set of rooted graphs $G_\star$ has a countable dense set, and is thus separable. [See also Proposition A.10 in Appendix A.2.2.]

**Exercise 2.6** (Continuity of local neighborhood functions) Fix $H_\star \in \mathcal{G}_\star$. Show that $h: G_\star \mapsto \{0,1\}$ given by $h(G,o) = \mathbb{1}_{(B_r(o) \cong H_\star)}$ is continuous.

**Exercise 2.7** (Bounded number of graphs with bounded radius and degrees) Show that there are only a bounded number of isomorphism classes of rooted graphs $(G,o)$ with radius $r$ for which the degree of every vertex is at most $k$.

**Exercise 2.8** (Random local weak limit) Construct the simplest (in your opinion) possible example where the local weak limit of a sequence of deterministic graphs is random.

**Exercise 2.9** (Local weak limit of line and cycle) Let $G_n$ be given by $V(G_n) = [n], E(G_n) = \{\{i,i+1\}: i \in [n-1]\}$ be the line. Show that $G_n$ converges to $\mathbb{Z}$. Show that the same is true for the cycle, for which $E(G_n) = \{\{i,i+1\}: i \in [n-1]\} \cup \{\{1,n\}\}$.

**Exercise 2.10** (Local weak limit of finite tree) Let $G_n$ be the tree of depth $k$, in which every vertex except the $3 \times 2^{k-1}$ leaves have degree 3. Here $n = 3(2^k - 1)$. What is the local weak limit of $G_n$? Show that it is random, despite the fact that the graphs $G_n$ are deterministic.

**Exercise 2.11** (Uniform integrability and convergence of size-biased degrees) Show that when $(d_{o_n}^{(n)})_{n \geq 1}$ forms a uniformly integrable sequence of random variables, there exists a subsequence along which $D_\star$, the size-biased version of $D_n = d_{o_n}^{(n)}$, converges in distribution.

**Exercise 2.12** (Uniform integrability and degree regularity condition) For $\text{CM}_n(d)$, show that Conditions 1.5(a)-(b) imply that $(d_{o_n})_{n \geq 1}$ is a uniformly integrable sequence of random variables.

**Exercise 2.13** (Adding a small disjoint graph does not change the local weak limit) Let $G_n$ be a graph that converges in the local weak sense. Let $a_n \in \mathbb{N}$ be such that $a_n = o(n)$, and add a disjoint copy of an arbitrary graph of size $a_n$ to $G_n$. Denote the resulting graph by $G'_n$. Show that $G'_n$ has the same local weak limit of $G_n$. 
Exercise 2.14 (Local weak convergence does not imply uniform integrability of the degree of a random vertex) In the setting of Exercise 2.13, add a complete graph of size $a_n$ to $G_n$. Let $a_n^2 \gg n$. Show that the degree of a vertex chosen uniformly at random in $G_n'$ is not uniformly integrable.

Exercise 2.15 (Local weak limit of random 2-regular graph) Show that the configuration model $CM_n(d)$ with $d_i = 2$ for all $i \in [n]$ converges in the local weak sense in probability to $Z$. Conclude that the same applies to the random 2-regular graph.

Exercise 2.16 (Independent neighborhoods of different vertices) Let $G_n$ converge in probability in the local weak sense to $(G,o)$. Let $(o^{(1)}_n, o^{(2)}_n)$ be two independent uniformly chosen vertices in $[n]$. Show that $(G_n, o^{(1)}_n)$ and $(G_n, o^{(2)}_n)$ jointly converge to two independent copies of $(G,o)$.

Exercise 2.17 (Directed graphs as marked graphs) There are several ways to describe directed graphs as marked graphs. Give one.

Exercise 2.18 (Expected boundary of balls in Erdős-Rényi random graphs) Prove that, for $ER_n(\lambda/n)$ and every $k \geq 0$

$$E\left[|\partial B_{k}^{(G_n)}(1)|\right] \leq \lambda^k. \quad (2.7.2)$$

This can be done, for example, by showing that, for every $k \geq 1$

$$E\left[|\partial B_{k}^{(G_n)}(1)| \mid B_{k-1}^{(G_n)}(1)\right] \leq \lambda E[|\partial B_{k-1}^{(G_n)}(1)|], \quad (2.7.3)$$

and using induction.

Exercise 2.19 (Uniform integrability and moment convergence) Assume that $D_n = d_{o_n}^{(G_n)}$ is such that $D_n^3$ is uniformly integrable. Prove that $E_n[(d_{o_n}^{(G_n)})^3] \xrightarrow{p} E_{\mu}[d_{o}^3]$. Conclude that (2.4.39) and (2.4.40) hold. [Hint: You need to be very careful, as $E_{\mu}[d_{o}^3]$ may be a random variable when $\mu$ is a random measure.]

Exercise 2.20 (Example of weak convergence where convergence in probability fails) Construct an example where $G_n$ converges in distribution in the local weak sense to $(G,o)$, but not in probability.

Exercise 2.21 (Continuity of neighborhood functions) Fix $m \geq 1$ and $\ell_1, \ldots, \ell_m$. Show that

$$h(G,o) = 1_{\{(|\partial B_{k}(o)| = \ell_k \forall k \leq m\}} \quad (2.7.4)$$

is a bounded continuous function.

Exercise 2.22 (Proof of (2.4.2)) Let $G_n$ converge in probability in the local weak sense to $(G,o)$. Prove (2.4.2) using Exercise 2.16.

Exercise 2.23 (|C_{max}| is not a continuous functional) Show that $|C_{max}|/n$ is not a continuous functional in $G_n$.

Exercise 2.24 (Upper bound on |C_{max}| using LWC) Let $G_n = ([n],E(G_n))$ denote a finite (possibly disconnected) random graph. Let $(G,o)$ be a random
2.7 Exercises for Chapter 2

Exercise 2.25 (Convergence of the proportion of vertices in clusters of size at least k) Let $G_n$ converge in probability in the local weak sense to $(G,o)$. Show that $Z_{\geq k}$ in (2.5.2) satisfies $Z_{\geq k}/n \xrightarrow{P} \zeta_{\geq k} = \mathbb{P}(|C(o)| \geq k)$ for every $k \geq 1$.

Exercise 2.26 (Example where proportion in giant is smaller than survival probability) Construct an example where $G_n$ converges in probability in the local weak sense to $(G,o)$, while $|C_{\text{max}}|/n \xrightarrow{P} \eta < \zeta = \mathbb{P}(|C(o)| = \infty)$.

Exercise 2.27 (Local weak convergence with a subcritical limit) Let $G_n$ converge in probability in the local weak sense to $(G,o)$. Assume that $\zeta = \mathbb{P}(|C(o)| = \infty) = 0$. Use (2.5.6) to prove that $|C_{\text{max}}|/n \xrightarrow{P} 0$.

Exercise 2.28 (Sufficiency of (2.5.7) for almost locality giant) Under the conditions in Exercise 2.28, show that $\limsup_{n \to \infty} \mathbb{P}(|C(o)| > \varepsilon n) > 0$. (2.7.7)

Exercise 2.30 (Lower bound on graph distances in Erdős-Rényi random graphs) Use Exercise 2.18 to show that, for every $\varepsilon > 0$,

$$\mathbb{P}\left( \frac{\text{dist}_{\text{ER}}(\lambda/n)(o_1,o_2)}{\log n} \leq \frac{1 - \varepsilon}{\log \lambda} \right) = 0.$$ (2.7.8)

Exercise 2.31 (Lower bound on graph distances in Erdős-Rényi random graphs) Use Exercise 2.18 to show that

$$\lim_{K \to \infty} \limsup_{n \to \infty} \mathbb{P}\left( \text{dist}_{\text{ER}}(\lambda/n)(o_1,o_2) \leq \frac{\log n}{\log \lambda} - K \right) = 0,$$ (2.7.9)

which is a significant extension of Exercise 2.30.

Exercise 2.32 (Convergence of the clustering spectrum) Prove that, under the conditions of Theorem 2.20, the convergence of the clustering spectrum in (2.6.2) holds.
Part II

Connected components
in random graphs
Overview of Part II.

In this part, we study connected components in random graphs. In more detail, we investigate the connected components of uniform vertices, thus describing the local weak limits of these random graphs. Further, we study the structure of the largest connected component, sometimes also called the giant component, which contains a positive proportion of the graph. In many random graphs, such a giant component exists when there are sufficiently many connections, while the largest connected component is much smaller than the number of vertices when there are few connections. Thus, these random graphs satisfy a phase transition. We identify the size of the giant component, as well as its structure in terms of the degrees of its vertices. We also investigate when the graph is fully connected, thus establishing the connectivity threshold for the random graphs in question.

In more detail, this part is organised as follows. We study general inhomogeneous random graphs in Chapter 3, and the configuration model, as well the closely related uniform random graph with prescribed degrees, in Chapter 4. In the last chapter of this part, Chapter 5, we study the connected components and local weak convergence of the preferential attachment model.
Chapter 3
Phase transition in general inhomogeneous random graphs

Abstract
In this chapter, we introduce the general setting of inhomogeneous random graphs. The inhomogeneous random graph is a generalization of the Erdős-Rényi random graph $\text{ER}_n(p)$ as well as the inhomogeneous random graphs such as $\text{GRG}_n(w)$ studied in [Volume 1, Chapter 6]. In inhomogeneous graphs, the status of edges is independent with unequal edge occupation probabilities. While these edge probabilities are moderated by vertex weights in $\text{GRG}_n(w)$, in the general setting they are described in terms of a kernel. We start by motivating its choice, which is inspired by [Volume 1, Example 6.1]. The main results in this section concern the degree structure of such inhomogeneous random graphs, multi-type branching process approximations to neighborhoods and the phase transition in these random graphs. We also discuss various examples of such inhomogeneous random graphs, and indicate that they can have rather different behavior.

3.1 Motivation
In this chapter, we discuss general inhomogeneous random graphs, where the edge statuses are independent. We investigate their connectivity structure, and particularly whether they have a giant component. This is inspired by the fact that many real-world networks are highly connected, in the sense that their largest connected component contains a large proportion of the total vertices of the graph. See Table 3.1 for many examples.

Table 3.1 raises the question how one can view these settings where giant components exist. We know that there is a phase transition in the size of the giant component.

<table>
<thead>
<tr>
<th>Subject</th>
<th>% in Giant</th>
<th>Total</th>
<th>Original Source</th>
<th>Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>AD-blood</td>
<td>0.8542</td>
<td>96</td>
<td>Goñi et al. (2008)</td>
<td>Goñi et al. (2008)</td>
</tr>
<tr>
<td>MS-blood</td>
<td>0.8780</td>
<td>205</td>
<td>Goñi et al. (2008)</td>
<td>Goñi et al. (2008)</td>
</tr>
<tr>
<td>actors</td>
<td>0.8701</td>
<td>2180759</td>
<td>Boldi and Vigna (2004)</td>
<td>Boldi et al. (2011)</td>
</tr>
<tr>
<td>DBLP</td>
<td>0.8162</td>
<td>986324</td>
<td>Boldi and Vigna (2004)</td>
<td>Boldi et al. (2011)</td>
</tr>
<tr>
<td>Zebras</td>
<td>0.8214</td>
<td>28</td>
<td>Sundaresan et al. (2007)</td>
<td>Sundaresan et al. (2007)</td>
</tr>
</tbody>
</table>

Table 3.1 The five rows correspond to the following real-life networks:
2. Model of the protein-protein connections in blood of people with Multiple Sclerosis.
3. Network of actors in 2011 in the Internet Movie Data base (IMDb), where actors are linked if they appeared in the same movie.
4. Collaboration network based on DBLP. Two scientist are connected if they ever collaborated on a paper.
5. Network where nodes represent zebras. There is a connection between two zebras if they ever interacted with each other during the observation phase.
Phase transition in general inhomogeneous random graphs

component in the ER$_n(\lambda/n)$, recall [Volume 1, Chapter 4]. A main ingredient will be to investigate when a giant component is present in general inhomogeneous random graphs. This will occur precisely when the local weak limit has a positive probability of surviving (recall Section 2.5). Therefore, we also investigate local weak convergence in this chapter.

We will study much more general models where edges are present independently than we have done so far (recall the generalized random graph in [Volume 1, Chapter 6], see also Section 1.3.2). In the generalized random graph, vertices have weights associated to them, and the edge occupation probabilities are proportional to the product of the weights of the vertices that the edge connects. This means that vertices with high weights have relatively large occupation probabilities to all other vertices, a property that may not be always be appropriate. Let us illustrate this by an example, which is a continuation of [Volume 1, Example 6.1]:

**Example 3.1** (Population of two types: general setting) Suppose that we have a complex network in which there are $n_1$ vertices of type 1 and $n_2$ of type 2. Type 1 individuals have on average $m_1$ neighbors, type 2 individuals $m_2$, where $m_1 \neq m_2$. Further, suppose that the probability that a type 1 individual is a friend of a type 2 individual is quite different from the probability that a type 1 individual is a friend of a type 1 individual.

In the model proposed in [Volume 1, Example 6.3], the probability that a type $i$ individual is a friend of a type $j$ individual (where $i, j, \in \{1, 2\}$) is equal to $m_i m_j / (\ell_n + m_i m_j)$, where $\ell_n = n_1 m_1 + n_2 m_2$. Approximating this probability by $m_i m_j / \ell_n$, we see that the probability that a type 1 individual is friend of a type 2 individual is highly related to the probability that a type 1 individual is friend of a type 1 individual. Indeed, take two type 1 and two type 2 individuals. Then, the probability that the type 1 individuals are friends and the type 2 individuals are friends is almost the same as the probability that first type 1 individual is friend with the first type 2 individual, and the second type 1 individual is friend of the second type 2 individual. Thus, there is some, possibly unwanted and artificial, symmetry in the model. How can one create instances where the edge probabilities between vertices of the same type are much larger, or alternatively much smaller, than they would be for the generalized random graph? In the two extremes, we either have a bipartite graph where vertices are only connected to vertices of the other type, or a disjoint union of two Erdős-Rényi random graphs, consisting of the vertices of the two types and no edges between them. We aim to be able to obtain anything in between. In particular, the problem with the generalized random graph originates in the product structure of the edge probabilities. In this chapter, we will deviate from such a product structure.

As explained above, we wish to be quite flexible in our choices of edge probabilities. However, we also wish for settings where the random graph is sufficiently ‘regular’, as for example exemplified by its degree sequences converging to some deterministic distribution. In particular, we aim for settings where the random.
3.2 Definition of the model

graphs are sparse. As a result, we need to build this regularity into the precise structure of the edge probabilities. This will be achieved by introducing an appropriate kernel that moderates the edge probabilities, and that is sufficiently regular. Let us introduce the model in detail in the next section.

Organization of this chapter

This chapter is organised as follows. In Section 3.2, we introduce general inhomogeneous random graphs. In Section 3.3, we study the degree distribution in general inhomogeneous random graphs. In Section 3.4, we treat multi-type branching processes, the natural generalization of branching processes for inhomogeneous random graphs. In Section 3.5, we use multi-type branching processes to identify the local weak limit of inhomogeneous random graphs. In Section 3.6, we study the phase transition of inhomogeneous random graphs. In Section 3.7, we state some recent related results. We close this chapter with notes and discussion in Section 3.8, and exercises in Section 3.9.

3.2 Definition of the model

We assume that our individuals have types which are in a certain type space $S$. When there are individuals of just 2 types, as in Example 3.1, then it suffices to take $S = \{1, 2\}$. However, the model allows for rather general sets of types of the individuals, both finite as well as (countably or uncountably) infinite. An example of an uncountably infinite type space could be types related to the ages of the individuals in the population. Also the setting of the generalized random graph with $\nu_i$ satisfying (1.3.15) corresponds to the uncountable type-space setting when the distribution function $F$ is that of a continuous random variable $W$. We therefore also need to know how many individuals there are of a given type. This is described in terms of a measure $\mu_n$, where, for $A \subseteq S$, $\mu_n(A)$ denotes the proportion of individuals having a type in $A$.

In our general model, instead of vertex weights, the edge probabilities are moderated by a kernel $\kappa: S^2 \to [0, \infty)$. The probability that two vertices of types $x_1$ and $x_2$ are connected is approximately $\kappa(x_1, x_2)/n$, and different edges are present independently. Since there are many choices for $\kappa$, we arrive at a rather flexible model, where vertices have types and connection probabilities are related to the types of the vertices involved.

We start by making the above definitions formal, by defining what our ground space is and what a kernel is:

**Definition 3.2** (Setting: ground space and kernel) (i) A ground space is a pair $(S, \mu)$, where $S$ is a separable metric space and $\mu$ is a Borel probability measure on $S$.

(ii) A vertex space $V$ is a triple $(S, \mu, (x_n)_{n \geq 1})$, where $(S, \mu)$ is a ground space and, for each $n \geq 1$, $x_n$ is a random sequence $(x_1, x_2, \ldots, x_n)$ of $n$ points of $S$, such
that
\[ \mu_n(A) = \# \{i : x_i \in A \} / n \rightarrow \mu(A) \]  
(3.2.1)
for every \( \mu \)-continuity set \( A \subseteq S \). The convergence in (3.2.1) is denoted by \( \mu_n \xrightarrow{p} \mu \).

(iii) A kernel \( \kappa \) is a symmetric non-negative (Borel) measurable function on \( S^2 \). By a kernel on a vertex space \((S, \mu, (x_n)_{n \geq 1})\) we mean a kernel on \((S, \mu)\).

Before defining the precise random graph model, we state the necessary conditions on our kernels. We write \( E(G) \) for the number of edges in a graph \( G \). Note that for an inhomogeneous random graph with edge probabilities \( p = (p_{ij})_{1 \leq i < j \leq n} \),
\[ E[E(IRG_n(p))] = \sum_{i<j} p_{ij}, \]  
(3.2.2)
so that our model has bounded degree in expectation precisely when \( \frac{1}{n} \sum_{i<j} p_{ij} \) remains bounded. In our applications, we wish that the average degree per vertex in fact converges, as this corresponds to the random graph model being sparse (recall [Volume 1, Definition 1.3]). Further, we do not wish the vertex set to be subdivided into two sets of vertices where the probability to have an edge between them is zero, as this implies that the graph can be divided into two entirely disjoint graphs. This explains the main conditions we pose on the kernel \( \kappa \) in the following definition:

**Definition 3.3** (Setting: graphical and irreducible kernels) (i) A kernel \( \kappa \) is graphical if the following conditions hold:
(a) \( \kappa \) is continuous a.e. on \( S^2 \);
(b) \[ \int \int_{S^2} \kappa(x,y)\mu(dx)\mu(dy) < \infty; \]  
(3.2.3)
(c) \[ \frac{1}{n} E[E(IRG_n(\kappa))] \rightarrow \frac{1}{2} \int \int_{S^2} \kappa(x,y)\mu(dx)\mu(dy). \]  
(3.2.4)
Similarly, a sequence \( (\kappa_n)_{n \geq 1} \) of kernels is called graphical with limit \( \kappa \) when
\[ y_n \rightarrow y \quad \text{and} \quad z_n \rightarrow z \quad \text{imply that} \quad \kappa_n(y_n, z_n) \rightarrow \kappa(y, z), \]  
(3.2.5)
where \( \kappa \) satisfies conditions (a) and (b) above, and
\[ \frac{1}{n} E[E(IRG_n(\kappa_n))] \rightarrow \frac{1}{2} \int \int_{S^2} \kappa(x,y)\mu(dx)\mu(dy). \]  
(3.2.6)
(ii) A kernel \( \kappa \) is called reducible if
\[ \exists A \subseteq S \quad \text{with} \quad 0 < \mu(A) < 1 \quad \text{such that} \quad \kappa = 0 \quad \text{a.e. on} \quad A \times (S \setminus A); \]
otherwise \( \kappa \) is irreducible.
3.2 Definition of the model

We now discuss the above definitions. The assumptions in (3.2.3), (3.2.4), (3.2.6) imply that the expected number of edges is proportional to \( n \), and that the proportionality constant is precisely equal to \( \int \int S^2 \kappa(x, y)\mu(dx)\mu(dy) \). Thus, in the terminology of [Volume 1, Chapter 1], \( \text{IRG}_n(\kappa) \) is sparse. This sparsity also allows us to approximate graphical kernels by bounded ones in such a way that the number of removed edges is \( o(n) \), a fact that will be crucially used in the sequel. Indeed, bounded graphical kernels can be well approximated by step functions in a similar way as continuous functions on \( \mathbb{R} \) can be well approximated by step functions. In turn, such step functions on \( S \times S \) correspond to random graphs with finitely many different types.

We extend the setting to \( n \)-dependent sequences \( (\kappa_n)_{n \geq 1} \) of kernels in (3.2.5), as in many natural cases the kernels do indeed depend on \( n \). In particular, it allows us to deal with several closely related and natural notions of the edge probabilities at the same time (see e.g., (3.2.7), (3.2.8) and (3.2.9) below), showing that identical results hold in each of these cases.

Roughly speaking, \( \kappa \) is reducible if the vertex set of \( \text{IRG}_n(\kappa) \) can be split in two parts so that the probability of an edge from one part to the other is zero, and irreducible otherwise. When \( \kappa \) is reducible, then the random graph splits into two independent random graphs on the two disjoint subsets \( A \) and \( S \setminus A \). Therefore, we could have equally well started with each of them separately, explaining why the notion of irreducibility is quite natural.

In many cases, we shall take \( S = [0,1] \), \( x_i = i/n \) and \( \mu \) the Lebesgue-measure on \([0,1]\). Then, clearly, (3.2.1) is satisfied. In fact, Janson (2009b) shows that we can always restrict to \( S = [0,1] \) by suitably adapting the other choices of our model. However, for notational purposes, it is more convenient to work with general \( S \). For example, where \( S = \{1\} \) is just a single type, the model reduces to the Erdős-Rényi random graph, and in the setting where \( S = [0,1] \), this is slightly more cumbersome, as worked out in detail in Exercise 3.1.

Now we come to the definition of our random graph. Given a kernel \( \kappa \), for \( n \in \mathbb{N} \), we let \( \text{IRG}_n(\kappa) \) be the random graph on \([n]\), each possible edge \( ij \), where \( i, j \in [n] \), is present with probability

\[
p_{ij}(\kappa) = p_{ij} = \frac{1}{n} [\kappa(x_i, x_j) \land n],
\]

and the events that different edges are present are independent. Similarly, \( \text{IRG}_n(\kappa_n) \) is defined with \( \kappa_n \) replacing \( \kappa \) in (3.2.7). Exercise 3.2 shows that the lower bound in (3.2.4) always holds for \( \text{IRG}_n(\kappa) \) when \( \kappa \) is continuous. Further, Exercise 3.3 shows that (3.2.4) holds for \( \text{IRG}_n(\kappa) \) when \( \kappa \) is bounded and continuous.

Bollobás et al. (2007) also allow for the choices

\[
p_{ij}^{(NR)}(\kappa_n) = 1 - e^{-\kappa_n(x_i, x_j)/n},
\]

and

\[
p_{ij}^{(GRG)}(\kappa_n) = p_{ij} = \frac{\kappa(x_i, x_j)}{n + \kappa(x_i, x_j)}.
\]
Phase transition in general inhomogeneous random graphs

All results in Bollobás et al. (2007) remain valid for the choices in (3.2.8) and (3.2.9). When
\[ \sum_{i,j \in [n]} \kappa_n(x_i, x_j)^3 = o(n^{3/2}), \]  
(3.2.10)
this follows immediately from [Volume 1, Theorem 6.18], see Exercise 3.6.

For CL \( n(w) \) with \( w = (w_i)_{i \in [n]} \) as in (1.3.15), we take \( S = [0, 1] \), \( x_i = i/n \) and, with \( \psi(x) = [1 - F]^{-1}(x) \),
\[ \kappa_n(x, y) = \psi(x) \psi(y) n/\ell_n. \]  
(3.2.11)

For CL \( n(w) \) with \( w = (w_i)_{i \in [n]} \) satisfying Condition 1.1, instead, we take \( S = [0, 1] \), \( x_i = i/n \) and
\[ \kappa_n(i/n, j/n) = w_i w_j / \mathbb{E}[W_n]. \]  
(3.2.12)
Exercises 3.4–3.5 study the Chung-Lu random graph in the present framework.

In the next section, we discuss some examples of inhomogeneous random graphs.

3.2.1 Examples of inhomogeneous random graphs

The Erdős-Rényi random graph
If \( S \) is general and \( \kappa(x, y) = \lambda \) for every \( x, y \in S \), then the edge probabilities \( p_{ij} \) given by (3.2.7) are all equal to \( \lambda/n \) (for \( n > \lambda \)). Then \( IRG_n(\kappa) = ER_n(\lambda/n) \).

The simplest choice here is to take \( S = \{1\} \).

The homogeneous bipartite random graph
Let \( n \) be even, ket \( S = \{0, 1\} \), let \( x_i = 0 \) for \( i \in [n/2] \) and \( x_i = 1 \) for \( i \in [n] \setminus [n/2] \). Further, let \( \kappa \) be defined by \( \kappa(x, y) = 0 \) when \( x \neq y \) and \( \kappa(x, y) = \lambda \) when \( x = y \). Then \( IRG_n(\kappa) \) is the random bipartite graph with \( n/2 \) vertices in each class, where each possible edge between classes is present with probability \( \lambda/n \), independently of the other edges. Exercise 3.7 investigates the validity of Definitions 3.2–3.3 for homogeneous bipartite graphs.

The stochastic blockmodel
The stochastic block model generalizes the above setting. Let \( n \) be even, ket \( S = \{0, 1\} \), let \( x_i = 0 \) for \( i \in [n/2] \) and \( x_i = 1 \) for \( i \in [n] \setminus [n/2] \). Further, let \( \kappa \) be defined by \( \kappa(x, y) = b \) when \( x \neq y \) and \( \kappa(x, y) = b \) when \( x = y \). This means that vertices of the same type are connected with probability \( a/n \), while vertices with different types are connected with probabilities \( a/n \). A major research effort has been devoted to studying when it can be statistically detected that \( a \neq b \). Below, we also investigate more general stochastic blockmodels.

Homogeneous random graphs
We call an inhomogeneous random graph homogeneous when \( \lambda(x) = \int_S \kappa(x, y) \mu(dy) \equiv \lambda \). Thus, despite the inhomogeneity that is present, every vertex in the graph has
3.2 Definition of the model

(asymptotically) the same number of expected offspring. Exercise 3.8 shows that the Erdős-Rényi random graph, the homogeneous bipartite random graph and the stochastic block model are all homogeneous random graphs. In such settings, however, the level of inhomogeneity is limited.

Inhomogeneous random graphs with finitely many types

Fix \( r \geq 2 \) and suppose we have a graph with \( r \) different types of vertices. Let \( S = \{1, \ldots, r\} \). Let \( n_i \) denote the number of vertices of type \( i \), and let \( \mu_n(i) = n_i/n \). Let \( IRG_n(\kappa) \) be the random graph where two vertices of types \( i \) and \( j \), respectively, joined by an edge with probability \( n^{-1} \kappa(i, j) \) (for \( n \geq \max \kappa \)). Then \( \kappa \) is equivalent to an \( r \times r \) matrix, and the random graph \( IRG_n(\kappa) \) has vertices of \( r \) different types (or colors). The finite-types case has been studied by Söderberg (2002, 2003a,b,c). We conclude that our general inhomogeneous random graph covers the cases of a finite (or even countably infinite) number of types. Exercises 3.9–3.11 study the setting of inhomogeneous random graphs with finitely many types.

It will turn out that this case is particularly important, as many of the other settings can be arbitrarily well approximated by inhomogeneous random graphs with finitely many types. As such, this model will be the building block upon which most of the results are built.

Uniformly grown random graph

The uniformly grown random graph was proposed by Callaway, Hopcroft, Kleinberg, Newman, and Strogatz (2001). In their model, the graph grows dynamically as follows. At each time step, a new vertex is added. Further, with probability \( \delta \), two vertices are chosen uniformly at random and joined by an undirected edge. This process is repeated for \( n \) time steps, where \( n \) describes the number of vertices in the graph. Callaway et al. predict, based in physical reasonings, that in the limit of large \( n \), the resulting graph has a giant component precisely when \( \delta > 1/8 \), and the proportion of vertices in the giant component is of the order \( e^{-\Theta(1/\sqrt{8\delta - 1})} \). Such behavior is sometimes called an infinite order phase transition. Durrett (2003) discusses this model. We will discuss a variant of this model proposed and analyzed by Bollobás, Janson and Riordan (2005). Their model is an example of the general inhomogeneous random graphs as discussed in the previous section. We take as a vertex space \([n]\), and the edge \( ij \) is present with probability

\[
p_{ij} = \frac{c}{\max\{i, j\}},
\]

all edge statuses being independent random variables. Equivalently, we can view this random graph as arising dynamically, where vertex \( t \) connects to a vertex \( s < t \) with probability \( 1/t \) independently for all \( s \in [t-1] \).

Sum kernels

We have already seen that product kernels are special, as they give rise to the Chung-Lu model or its close brothers, the generalized random graph or the
96 Phase transition in general inhomogeneous random graphs

Norros-Reittu model. For sum kernels, instead, we take \( \kappa(x, y) = \psi(x) + \psi(y) \), so that \( p_{ij} = \min\{((\psi(i/n) + \psi(j/n))/n, 1) \} \).

We see that there are tons of examples of random graphs with independent edges that fall into the general class of inhomogeneous random graphs. In the sequel, we will investigate them in general. We start by investigating their degree structure.

3.3 Degree sequence of inhomogeneous random graphs

We now start by investigating the degrees of the vertices of IRG\(_n(\kappa_n)\). As we shall see, the degree of a vertex of a given type \( x \) is asymptotically Poisson with a mean

\[ \lambda(x) = \int_S \kappa(x, y) \mu(dy) \]  

(3.3.1)

that depends on \( x \). This leads to a mixed Poisson distribution for the degree \( D \) of a (uniformly chosen) random vertex of IRG\(_n(\kappa)\). We recall that \( N_k(n) \) denotes the number of vertices of IRG\(_n(\kappa)\) with degree \( k \). Our main result is as follows:

**Theorem 3.4** (The degree sequence of IRG\(_n(\kappa)\)) Let \((\kappa_n)\) be a graphical sequence of kernels with limit \( \kappa \) as described in Definition 3.3(i). For any fixed \( k \geq 0 \),

\[ N_k(n)/n \xrightarrow{p} \int_S \frac{\lambda(x)^k}{k!} e^{-\lambda(x)} \mu(dx) \]  

(3.3.2)

where \( x \mapsto \lambda(x) \) is defined by

\[ \lambda(x) = \int_S \kappa(x, y) \mu(dy) \]  

(3.3.3)

Equivalently,

\[ N_k/n \xrightarrow{p} \mathbb{P}(D = k), \]  

(3.3.4)

where \( D \) has the mixed Poisson distribution with distribution \( W_\lambda \) given by

\[ \mathbb{P}(W_\lambda \leq x) = \int_0^x \lambda(y) \mu(dy). \]  

(3.3.5)

In the remainder of this section, we prove Theorem 3.4. This proof is also a good example of how such proofs will be carried out in the sequel. Indeed, we start by proving Theorem 3.4 for the finite-types case, which is substantially easier. After this, we give a proof in the general case, for which we shall need to prove results on approximations of sequences of graphical kernels. These approximations will apply for bounded kernels, and thus we will also need to show that unbounded kernels can be well-approximated by bounded kernels. It is here that the assumption (3.2.4) will be crucially used.
3.3 Degree sequence of inhomogeneous random graphs

3.3.1 Degree sequence of finitely-types case

We start by proving Theorem 3.4 in the finite-types case. Take a vertex $v$ of type $i$, let $D_v$ be its degree, and let $D_{v,j}$ be the number of edges from $v$ to vertices of type $j \in [r]$. Then, clearly, $D_v = \sum_j D_{v,j}$.

Recall that, in the finite-types case, the edge probability between vertices of types $i$ and $j$ is denoted by $(\kappa_n(i,j) \wedge n) / n$. Further, (3.2.5) implies that $\kappa_n(i,j) \to \kappa(i,j)$ for every $i,j \in [r]$, while (3.2.1) implies that the number $n_i$ of vertices of type $i$ satisfies $n_i / n \to \mu_i$ for some probability distribution $(\mu_i)_{i \in [r]}$.

Assume that $n \geq \max_\kappa$. The random variables $(D_{v,j})_{j \in [r]}$ are independent, and $D_{v,j} \sim \text{Bin}(n_j - \delta_{ij}, \kappa(i,j)/n) \xrightarrow{d} \text{Poi}(\mu_j \kappa(i,j))$, where $n_j$ is the number of vertices with type $j$ and $\mu_j = \lim_{n \to \infty} n_j / n$. Hence,

$$D_v \xrightarrow{d} \text{Poi}\left(\sum_j \mu_j \kappa(i,j)\right) = \text{Poi}(\lambda(i)),$$

where $\lambda(i) = \int \kappa(i,j)d\mu(j) = \sum_j \kappa(i,j) \mu_j$. Consequently,

$$\mathbb{P}(D_v = k) \to \mathbb{P}(\text{Poi}(\lambda(i)) = k) = \frac{\lambda(i)^k}{k!} e^{-\lambda(i)}. \tag{3.3.7}$$

Let $N_{k,i}$ be the number of vertices in $\text{IRG}_n(\kappa_n)$ of type $i$ with degree $k$. Then, for fixed $n_1, \ldots, n_r$,

$$\frac{1}{n} \mathbb{E}[N_{k,i}] = \frac{1}{n} n_i \mathbb{P}(D_v = k) \to \mu_i \mathbb{P}(\text{Poi}(\lambda(i)) = k). \tag{3.3.8}$$

It is easily checked that $\text{Var}(N_{k,i}) = O(n)$ (see Exercise 3.12). Hence,

$$\frac{1}{n} N_{k,i} \xrightarrow{p} \mathbb{P}(\text{Poi}(\lambda(i)) = k) \mu_i = \mathbb{P}(N = k), \tag{3.3.9}$$

and thus, summing over $i$,

$$\frac{1}{n} N_k = \sum_i \frac{1}{n} N_{k,i} \xrightarrow{p} \sum_i \mathbb{P}(\text{Poi}(\lambda(i)) = k) \mu_i = \mathbb{P}(D = k). \tag{3.3.10}$$

This proves Theorem 3.4 in the finite-type case.

In order to prove Theorem 3.4 in the general case, we shall be approximating a sequence of graphical kernels $(\kappa_n)$ by appropriate regular finite kernels.

3.3.2 Finite-type approximations of bounded kernels

Recall that $S$ is a separable metric space, and that $\mu$ is a Borel measure on $S$ with $\mu(S) = 1$. Here the metric and topological structure of $S$ will be important. We refer to Appendix A for more details on metric spaces that will be assumed throughout this section.

In this section, we assume that $(\kappa_n)$ is a graphical sequence of bounded kernels with limit $\kappa$ as described in Definition 3.3(i). Thus, we assume that

$$\sup_{n \geq 1} \sup_{x,y \in S} \kappa_n(x,y) < \infty. \tag{3.3.11}$$
Our aim is to find finite-type approximations of $\kappa_n$ that bound $\kappa_n$ from above and below. It is here that the metric structure of $S$, as well as the continuity of $(x, y) \mapsto \kappa_n(x, y)$, is crucially used. This is the content of the next lemma:

**Lemma 3.5** (Finite-type approximations of general kernels) If $(\kappa_n)_{n \geq 1}$ is a graphical sequence of kernels on a vertex space $V$ with limit $\kappa$, then there exists a sequence $(\kappa_m)_{m \geq 1}$ of finite-type kernels on $V$ with the following properties:

(i) If $\kappa$ is irreducible, then so is $\kappa_m$ for all large enough $m$;

(ii) $\kappa_m(x, y) \nearrow \kappa(x, y)$ for every $x, y \in S$.

Let us now give some details. We find these finite-type approximations by giving a partition $P_m$ of $S$ on which $\kappa_n(x, y)$ is almost constant when $x$ and $y$ are inside cells of the partition. Fix $m \geq 1$ which will indicate the number of cells in the partition of $S$. Given a sequence of finite partitions $P_m = \{A_{m1}, \ldots, A_{mM_m}\}$, $m \geq 1$, of $S$ and an $x \in S$, we define the function $x \mapsto i_m(x)$ by requiring that

$$x \in A_{m,i_m(x)}.$$  \hfill (3.3.12)

Thus, $i_m(x)$ indicates the cell in $P_m$ that $x$ is in. For $A \subset S$ we write $\text{diam}(A)$ for $\sup\{|x - y|: x, y \in A\}$. By taking $P_m$ as the dyadic partition into intervals of length $2^{-m}$ in $S$, we easily see the following:

**Lemma 3.6** (Approximating partition) Fix $m \geq 1$. There exists a sequence of finite partitions $P_m = \{A_{m1}, \ldots, A_{mM_m}\}$ of $S$ such that

(i) each $A_{mi}$ is measurable and $\mu(\partial A_{mi}) = 0$;

(ii) for each $m$, $P_{m+1}$ refines $P_m$, i.e., each $A_{mi}$ is a union $\bigcup_{j \in J_{mi}} A_{m+1,j}$ for some set $J_{mi}$;

(iii) for a.e. $x \in S$, $\text{diam}(A_{m,i_m(x)}) \to 0$ as $m \to \infty$, where $i_m(x)$ is defined by (3.3.12).

Recall that a kernel $\kappa$ is a symmetric measurable function on $S \times S$. Fixing a sequence of partitions with the properties described in Lemma 3.6, we can define sequences of lower and upper approximations to $\kappa$ by

$$\kappa_m^-(x, y) = \inf\{\kappa(x', y') : x' \in A_{m,i_m(x)}, y' \in A_{m,i_m(y)}\},$$  \hfill (3.3.13)

$$\kappa_m^+(x, y) = \sup\{\kappa(x', y') : x' \in A_{m,i_m(x)}, y' \in A_{m,i_m(y)}\}. $$  \hfill (3.3.14)

We thus replace $\kappa$ by its infimum or supremum on each $A_{mi} \times A_{mj}$. As $\kappa_m^+$ might be $+\infty$, we only use it for bounded $\kappa_n$ as in (3.3.11). Obviously, $\kappa_m^-$ and $\kappa_m^+$ are constant on $A_{mi} \times A_{mj}$ for every $i, j$, so that $\kappa_m^-$ and $\kappa_m^+$ correspond to finite-type kernels (see Exercise 3.14).

By Lemma 3.6(ii),

$$\kappa_m^- \leq \kappa_{m+1}^- \quad \text{and} \quad \kappa_m^+ \geq \kappa_{m+1}^+. $$  \hfill (3.3.15)

Furthermore, if $\kappa$ is continuous a.e. then, by Lemma 3.6(iii),

$$\kappa_m^-(x, y) \to \kappa(x, y) \quad \text{and} \quad \kappa_m^+(x, y) \to \kappa(x, y) \quad \text{for a.e.} \ (x, y) \in S^2.$$  \hfill (3.3.16)
If \((κ_n)\) is a graphical sequence of kernels with limit \(κ\), we define instead
\[
k_m^-(x, y) := \inf \{(κ \land κ_n)(x', y') : x' \in A_{m, i_m(x)}, y' \in A_{m, i_m(y)}, n \geq m\},
\]
\[
k_m^+(x, y) := \sup \{(κ \lor κ_n)(x', y') : x' \in A_{m, i_m(x)}, y' \in A_{m, i_m(y)}, n \geq m\}.
\]
Again, by Lemma 3.6, we have \(κ_m^- \leq κ_{m+1}^-\), and, by Lemma 3.6(iii) and 3.3(ii),
\[
k_m^-(x, y) \nearrow κ(x, y) \quad \text{as} \quad m \to \infty, \quad \text{for} \quad \text{a.e.} \quad (x, y) \in S^2.
\]
Since \(κ_m^- \leq κ\), we can obviously construct our random graph so that \(\text{IRG}_n(κ_m^-) \subseteq \text{IRG}_n(κ_n)\), and in the sequel we assume this. See also Exercise 3.15. Similarly, we shall assume that \(\text{IRG}_n(κ_m^+) \supseteq \text{IRG}_n(κ_n)\) when \(κ_n\) is bounded as in (3.3.11). Moreover, when \(n \geq m\),
\[
κ_n \geq κ_m^-,
\]
and we may assume that \(\text{IRG}_n(κ_m^-) \subseteq \text{IRG}_n(κ_n)\). By the convergence of the sequence of kernels \((κ_n)\), we further obtain that also the number of edges converges. Thus, in bounding \(κ_n\), we do not create or destroy too many edges. This provides the starting point of our analysis, which we provide in the following section.

### 3.3 Degree sequence of general inhomogeneous random graphs

Now we are ready to complete the proof of Theorem 3.4 for general sequences of graphical kernels \((κ_n)\). Define \(κ_m^-\) by (3.3.17). Since we only use the lower bounding kernel \(κ_m^-\) (which always exists), we need not assume that \(κ_n\) is bounded.

Let \(ε > 0\) be given. From (3.2.6) and monotone convergence, there is an \(m\) such that
\[
\iint_{S^2} κ_m^-(x, y)μ(dx)μ(dy) > \iint_{S^2} κ(x, y)μ(dx)μ(dy) - ε.
\]
For \(n \geq m\), we have \(κ_m^- \leq κ_n\) by (3.3.20), so we may assume that \(\text{IRG}_n(κ_m^-) \subseteq \text{IRG}_n(κ_n)\). Then, by (3.2.6) and (3.3.21),
\[
\frac{1}{n} E\left(\text{IRG}_n(κ_n) \setminus \text{IRG}_n(κ_m^-)\right)
= \frac{1}{n} E\left(G(n, κ_n)\right) - \frac{1}{n} E\left(\text{IRG}_n(κ_m^-)\right)
\rightarrow \frac{1}{2} \iint_{S^2} κ(x, y)μ(dx)μ(dy) - \frac{1}{2} \iint_{S^2} κ_m^-(x, y)μ(dx)μ(dy) < \frac{ε}{2},
\]
so that, whp \(E\left(\text{IRG}_n(κ_n) \setminus \text{IRG}_n(κ_m^-)\right) < εn\). Let us write \(N_k^{(m)}\) for the number of vertices of degree \(k\) in \(\text{IRG}_n(κ_m^-)\). It follows that, whp,
\[
|N_k^{(m)} - N_k| < 2εn.
\]
Writing \(Ξ^{(m)}\) for the equivalent of \(Ξ\) defined using \(κ_m^-\) in place of \(κ\), by the proof for the finite-type case, \(N_k^{(m)}/n \xrightarrow{p} P(D^{(m)} = k)\). Thus, whp,
\[
|N_k^{(m)}/n - P(D^{(m)} = k)| < ε.
\]
Finally, we have \( E[\Xi] = \int_S \lambda(x) \mu(dx) = \int \int_{S^2} \kappa(x, y) \mu(dx) \mu(dy) \). Since \( \lambda^{(m)}(x) \leq \lambda(x) \), we can couple the branching processes such that \( \Xi^{(m)} \leq \Xi \), and thus
\[
\mathbb{P}(D \neq D^{(m)}) = \mathbb{P}(\Xi - \Xi^{(m)} \geq 1) \leq \mathbb{E}[\Xi - \Xi^{(m)}] = \int \int_{S^2} \kappa(x, y) \mu(dx) \mu(dy) - \int \int_{S^2} \kappa^{(m)}(x, y) \mu(dx) \mu(dy) < \varepsilon.
\] (3.3.25)

Combining (3.3.23), (3.3.24) and (3.3.25), we see that \(|N_k/n - \mathbb{P}(\Xi = k)| < 4\varepsilon\) whp.

Now that we have identified the limit of the degree distribution, let us discuss its proof as well as some of the properties of the limiting degree distribution.

**Bounded kernels**

First of all, the proof above will be exemplary of several proofs that we will use in this chapter, as well as Chapter 6. The current proof is particularly simple, as it only makes use of the lower bounding finite-type inhomogeneous random graph, while in many settings we also need the upper bound. This upper bound can only apply to bounded kernels \( \kappa_n \) as in (3.3.11). As a result, we will need to study the effect of bounding \( \kappa_n \), for example by approximating it by \( \kappa_n(x, y) \wedge K \).

**Tail properties of the degree distribution**

Let \( W = W_\lambda \) be the random variable \( \lambda(U) \), where \( U \) is a random variable on \( S \) having distribution \( \mu \). Then we can also describe the mixed Poisson distribution of \( D \) as \( \text{Poi}(W) \). Under mild conditions, the tail probabilities \( \mathbb{P}(D > t) \) and \( \mathbb{P}(W > t) \) agree for large \( t \). We state this for the case of power-law tails; many of these result generalizes to regularly-varying tails. Let \( N_{\geq k} \) be the number of vertices with degree at least \( k \).

**Corollary 3.7** (Power-law tails for the degree sequence) Let \( (\kappa_n) \) be a graphical sequence of kernels with limit \( \kappa \). Suppose that \( \mathbb{P}(W > t) = \mu(x: \lambda(x) > t) = at^{-(\tau-1)}(1 + o(1)) \) as \( t \to \infty \), for some \( a > 0 \) and \( \tau > 2 \). Then
\[
N_{\geq k}/n \xrightarrow{p} \mathbb{P}(D \geq k) \sim ak^{-(\tau-1)},
\] (3.3.26)
where the first limit is for \( k \) fixed and \( n \to \infty \), and the second for \( k \to \infty \). In particular, \( \lim_{n \to \infty} \mathbb{P}(D \geq k) \sim ak^{-(\tau-1)} \) as \( k \to \infty \).

**Proof** It suffices to show that \( \mathbb{P}(D \geq k) \sim ak^{-(\tau-1)} \); the remaining conclusions then follow from Theorem 3.4. For any \( \varepsilon > 0 \), as \( t \to \infty \),
\[
\mathbb{P}(\text{Poi}(W) > t \mid \Lambda > (1 + \varepsilon)t) \to 1, \quad (3.3.27)
\]
and \( \mathbb{P}(\text{Poi}(W) > t \mid W < (1 - \varepsilon)t) = o(t^{-(\tau-1)}) \).

It follows that \( \mathbb{P}(D > t) = \mathbb{P}(\text{Poi}(W) > t) = at^{-(\tau-1)}(1 + o(1)) \) as \( t \to \infty \). Exercise 3.16 asks you to fill the details on this argument.

This result shows that the general inhomogeneous random graph does include natural cases with power-law degree distributions. Recall that we have already
3.4 Multi-type branching processes

In order to study further properties of IRG\(_n(\kappa_n)\), we need to understand the neighborhood structure of vertices. This will be crucially used in the next section, where we study the local weak convergence properties of IRG\(_n(\kappa_n)\). For simplicity, let us restrict ourselves to the finite-types case. As we have seen, nice kernels can be arbitrarily well approximated by finite-type kernels, so this should be a good start. Then, for a vertex of type \(i\), the number of neighbors of type \(j\) is close to Poisson distributed with approximate mean \(\kappa(i, j)\mu(j)\). Even when we assume independence of the neighborhood structures of different vertices, we still do not arrive at a classical branching process as discussed in [Volume 1, Chapter 3]. Instead, we can describe the neighborhood structure with a branching process in which we keep track of the type of each of the vertices. For general \(\kappa\) and \(\mu\), we can even have a continuum of types. Such branching processes are called **multi-type branching processes**. In this section, we discuss some of the basics and we shall quickly go to the special case of multi-type branching processes where every offspring has a Poisson distribution.

### 3.4.1 Multi-type branching processes with finitely many types

A multi-type branching process can be analyzed using linear algebra. In order to do so, we first introduce some notation. We first assume that we are in the finite types case, and denote the number of types by \(r\). We let \(j = (j_1, \ldots, j_r) \in \mathbb{N}_0^r\) be a vector of non-negative integers, and denote by \(p_j^{(i)}\) the probability that an individual of type \(i\) gives rise to an offspring \(j\), i.e., \(j_1\) children of type 1, \(j_2\) children of type 2, etc. The offsprings of the different individuals are all mutually independent. Denote by \(Z_{n,j}^{(i)}\) the number of individuals of type \(j\) in generation \(n\) when starting from a single particle of type \(i\) and \(Z_n = (Z_{n,1}^{(i)}, \ldots, Z_{n,r}^{(i)})\). We are interested in the survival or extinction of multi-type branching processes, and in the growth of the generation sizes. In the multi-type case, we are naturally lead to a matrix setup. We now discuss the survival versus extinction of multi-type branching processes. We denote the survival probability of the multi-type branching process when starting from a single individual of type \(i\) by

\[
\zeta^{(i)} = \mathbb{P}(Z_{n}^{(i)} \neq 0 \text{ for all } n),
\]

and we let \(\zeta = (\zeta^{(1)}, \ldots, \zeta^{(r)})\). Our first aim is to investigate when \(\zeta = 0\).

#### Multi-type branching processes and generating functions

We write \(p(j) = (p_{j}^{(1)}, \ldots, p_{j}^{(r)})\) and we let

\[
G^{(i)}(s) = \sum_j p_{j}^{(i)} \prod_{a=1}^{r} s_{j_a}^{a}
\]
be the joint moment generating function of the offspring of an individual of type $i$. We write $G(s) = (G^{(1)}(s), \ldots, G^{(r)}(s))$ for the vector of generating functions. We now generalize Theorem 3.1 to the multi-type case. Let $\zeta$ satisfy $\zeta = 1 - G(1 - \zeta)$. By convexity of $s \mapsto G(s)$, there is at most one non-zero solution to the equation $s = G(s)$ which is not equal to 0. Define

$$G_n^{(i)}(s) = \mathbb{E}\left[\prod_{a=1}^{r} s_a Z_n^{(i)}(a)\right],$$

and $G_n(s) = (G_n^{(1)}(s), \ldots, G_n^{(r)}(s))$. Then, we have that $G_{n+1}(s) = G_n(G(s)) = G(G_n(s))$ and $\zeta = 1 - \lim_{n \to \infty} G_n(0)$. Naturally, the extinction probability depends sensitively on the type of the ancestor of the branching process. On the other hand, under reasonable assumptions, the positivity of the survival probability is independent of the initial type. A necessary and sufficient condition for this property is that, with positive probability, an individual of type $i$ arises as a descendent of an individual of type $j$ for each type $i$ and $j$. See Exercise 3.18. Exercise 3.19 relates this to the $l$th power of the mean offspring matrix $T_\kappa = (\kappa_{ij} \mu(j))_{i,j=1}^r$.

We note that when $G(s) = M s$ for some matrix $M$, then each individual in the Markov chain has precisely one offspring, and we call this case singular (see Exercise 3.20). When each particle has precisely one offspring, the multi-type branching process is equivalent to a Markov chain, and the process a.s. survives. Thus, in this case, there is no survival vs. extinction phase transition. We shall assume throughout the remainder that the multi-type branching process is non-singular.

### 3.4.2 Survival vs. extinction of multi-type branching processes

We continue to describe the survival versus extinction of multi-type branching processes in terms of the mean offspring. Let $\kappa_{ij}$ denote the expected offspring of type $j$ of a single individual of type $i$, and let $T_\kappa = (\kappa_{ij} \mu(j))_{i,j=1}^r$ be the matrix of expected offsprings. We shall assume that there exists an $l$ such that the matrix $T_\kappa^l$ has only strictly positive entries. This is sometimes called irreducibility, as it implies that the Markov chain of the number of individuals of the various types is an irreducible Markov chain. By the Perron-Frobenius theorem, the matrix $T_\kappa$ has a unique largest eigenvalue $\|T_\kappa\|$ with non-negative left-eigenvector $x_\kappa$, and the eigenvalue $\|T_\kappa\|$ can be computed as

$$\|T_\kappa\| = \sup_{x: \|x\|_2 \leq 1} \|T_\kappa x\|_2, \quad \text{where} \quad \|x\|_2 = \sqrt{\sum_{i=1}^{r} x_i^2}.$$  

(3.4.4)

We note that

$$\mathbb{E}[Z_{n+1}^{(i)} | Z_n^{(i)} = z] = T_\kappa z,$$

so that

$$\mathbb{E}[Z_{n+1}^{(i)}] = T_n^{(i)} e^{(i)}.$$

(3.4.5)
where $T_n^\kappa$ denotes the $n$-fold application of the matrix $T_\kappa$, and $e^{(i)}$ is the vector which has on the $i$th position a 1, and further only zeroes. The identifications in (3.4.5) and (3.4.6) have several important consequences concerning the phase transition of multi-type branching processes, as we shall now discuss in more detail.

First, when $\|T_\kappa\| < 1$, 
\[
E[Z_{n+1}^{(i)}] \leq \|T_\kappa\|^n \|e^{(i)}\|_2, \tag{3.4.7}
\]
which converges to 0 exponentially fast. Therefore, by the Markov inequality ([Volume 1, Theorem 2.17]), the multi-type branching process dies out a.s. When $\|T_\kappa\| > 1$, on the other hand, the sequence
\[
M_n = x_\kappa Z_{n+1}^{(i)} \|T_\kappa\|^{-n} \tag{3.4.8}
\]
is a non-negative martingale, by (3.4.5) and the fact that $x_\kappa$ is a left-eigenvector with eigenvalue $\|T_\kappa\|$, since $x_\kappa T_\kappa = \|T_\kappa\| x_\kappa$. By the Martingale convergence theorem (Theorem 2.24), the martingale $M_n$ converges a.s. When we further assume some further restrictions on $M_n$, for example that $M_n$ has finite second moment, then we obtain that $M_n \overset{a.s.}{\to} M_\infty$ and $E[M_n] \to E[M_\infty]$. More precisely, there is a multi-type analog of the Kesten-Stigum Theorem ([Volume 1, Theorem 3.10]). Since $E[M_n] = E[M_0] = x_\kappa e^{(i)} > 0$, we thus have that $Z_{n+1}^{(i)}$ grows exponentially with a strictly positive probability, which implies that the survival probability is positive. [Volume 1, Theorem 3.1] can be adapted to show that $Z_{n+1}^{(i)} \overset{p}{\to} 0$ when $\|T_\kappa\| = 1$. See e.g. (Harris, 1963, Sections II.6-II.7). We conclude that, for non-singular and irreducible multi-type branching processes, we have that $\zeta > 0$ precisely when $\|T_\kappa\| > 1$. This is the content of the following theorem:

**Theorem 3.8** (Survival vs. extinction of finite-type branching processes) Let $(Z_n^{(i)})_{n \geq 0}$ be a multi-type branching process with offspring matrix $T_\kappa$ on the type space $[r]$. Assume that there exists an $i$ such that the matrix $T_\kappa^{(i)}$ has only strictly positive entries, and that $\sum_j \kappa_{ij} \mu_j \neq 1$ for every $j \in [r]$. Then the following holds:

(a) The survival probability $\zeta$ is the largest solution to $\zeta = 1 - G(1 - \zeta)$, and $\zeta > 0$ precisely when $\|T_\kappa\| > 1$.

(b) Assume that $\|T_\kappa\| > 1$. Let $x_\kappa$ be the unique positive left-eigenvector of $T_\kappa$. Then, the martingale $M_n = x_\kappa Z_{n+1}^{(i)} \|T_\kappa\|^{-n}$ converges a.s. to a non-negative limit on the event of survival precisely when $E[Z_1^{(i)} \log(Z_1^{(i)})] < \infty$ for all $i \in [r]$, where $Z_1^{(i)} = \|Z_1^{(i)}\|_1$ is the total offspring of a type $i$ individual.

### 3.4.3 Poisson multi-type branching processes

We now specialize to Poisson multi-type branching processes as these turn out to be the most relevant in the random graph setting. We call a multi-type branching processes Poisson when all the number of children of each type are independent Poisson random variables. Thus, $Z^{(i)} = (Z_1^{(i)}, \ldots, Z_{n}^{(i)})$ is a vector of independent Poisson random variables with means $(\kappa_{1,i}, \ldots, \kappa_{n,i})$. As we see later, Poisson multi-type branching processes arise naturally when exploring a component of IRG$_n(\kappa)$ starting at a vertex of type $x$. This is directly analogous
to the use of the single-type Poisson branching process in the analysis of the Erdős-Rényi graph $\text{ER}_n(\lambda/n)$ as discussed in detail in [Volume 1, Chapters 4 and 5].

**Poisson multi-type branching processes with finitely many types**

For Poisson multi-type branching processes with finitely many types, we obtain that

$$G^{(i)}(s) = \mathbb{E}\left[\prod_{a=1}^{r} s_{a}^{Z_{a}^{(i)}}\right] = e^{\sum_{a=1}^{r} \kappa_{a,i} \mu(i)(s_{a}-1)} = e^{(T_{a}(s-1))_{i}}. \tag{3.4.9}$$

Thus, the vector of survival probabilities $\zeta$ satisfies

$$\zeta = 1 - e^{-T_{a} \zeta}, \tag{3.4.10}$$

where, for a matrix $M$, we recall that $e^{M} = \sum_{n \geq 0} M^{n}/n!$ denotes the matrix exponential. This leads us to the investigation of eigenfunctions of non-linear operators of the form $f \mapsto 1 - e^{-T_{a} f}$.

There is a beautiful property of Poisson random variables that allows us to construct a Poisson multi-type branching process in a particularly convenient way. This property follows from the following Poisson thinning property:

**Lemma 3.9 (A Poisson number of multinomial trials)** Let $X$ have a Poisson distribution with parameter $\lambda$. Perform $X$ multinomial trials, where the $i$th outcome appears with probability $p_{i}$ for some probabilities $(p_{i})_{i=1}^{r}$. Let $(X_{i})_{i=1}^{k}$, where $X_{i}$ denotes the total number of outcomes $i$. Then $(X_{i})_{i=1}^{k}$ is a sequence of independent Poisson random variables with parameters $(\lambda p_{i})_{i=1}^{r}$.

**Proof** Let $(x_{i})_{i=1}^{k}$ denote a sequence of non-negative integers, denote $x = \sum_{i=1}^{k} x_{i}$ and compute

$$\mathbb{P}(X_{i})_{i=1}^{k} = (x_{i})_{i=1}^{k} = \mathbb{P}(X = x)\mathbb{P}((X_{i})_{i=1}^{k} = (x_{i})_{i=1}^{k} | X = x) \tag{3.4.11}$$

$$= e^{-\lambda} \frac{\lambda^{x}}{x!} \left( \prod_{i=1}^{k} \frac{x}{x_{i}} \right) p_{i}^{x_{i}} \cdots p_{k}^{x_{k}} = \prod_{i=1}^{k} e^{-\lambda x_{i}} \frac{\lambda^{x_{i}}}{(x_{i})!}. \square$$

By Lemma 3.9, we can alternatively construct a Poisson branching process as follows. For an individual of type $i$, let its total number of offspring $Z_{i}$ have a Poisson distribution with parameter $\sum_{j \in [r]} \kappa_{ij} \mu(j)$. Then give each of the children a type $j$ with probability $\kappa_{ij} \mu(j)/\lambda(i)$. Let $Z_{ij}$ denote the total number of individuals of type $j$ thus obtained. Then, the offspring distribution $Z^{(i)}$ has the same distribution as $(Z_{ij})_{j \in [r]}$.

We now extend the above setting of finite-type Poisson multi-type branching processes to the infinite-type case. Again, we prove results in the infinite-types case by reducing to the finite-type case.
3.4 Multi-type branching processes

Let $\kappa$ be a kernel. We define the Poisson multi-type branching processes with kernel $\kappa$ as follows. Each individual of type $x \in S$ is replaced in the next generation by a set of individuals distributed as a Poisson process on $S$ with intensity $\kappa(x, y)\mu(dy)$. Thus, the number of children with types in a subset $A \subseteq S$ has a Poisson distribution with mean $\int_A \kappa(x, y)\mu(dy)$, and these numbers are independent for disjoint sets $A$ and for different particles; see e.g., Kallenberg (2002).

Let $\zeta_\kappa(x)$ be the survival probability of the Poisson multi-type branching process with kernel $\kappa$, starting from an ancestor of type $x \in S$. Set

$$
\zeta_\kappa = \int_S \zeta_\kappa(x)\mu(dx).
$$

Again, it can be seen in a similar way as above that $\zeta_\kappa > 0$ if and only if $\|T_\kappa\| > 1$, where now the linear operator $T_\kappa$ is defined, for $f : S \to \mathbb{R}$,

$$
(T_\kappa f)(x) = \int_S \kappa(x, y)f(y)\mu(dy),
$$

for any (measurable) function $f$ such that this integral is defined (finite or $+\infty$) for a.e. $x \in S$.

As we shall see, the analysis of multi-type branching processes with a possibly uncountable number of types is a bit more functional analytic. Similarly to the finite-type case in (3.4.4), we define

$$
\|T_\kappa\| = \sup \left\{ \|T_\kappa f\|_2 : f \geq 0, \|f\|_2 \leq 1 \right\} \leq \infty.
$$

When finite, $\|T_\kappa\|$ is the norm of $T_\kappa$ as an operator on $L^2(S)$; it is infinite if $T_\kappa$ does not define a bounded operator on $L^2$. The norm $\|T_\kappa\|$ is at most the Hilbert-Schmidt norm of $T_\kappa$:

$$
\|T_\kappa\| \leq \|T_\kappa\|_{HS} = \|\kappa\|_{L^2(S \times S)} = \left( \iint_{S^2} \kappa(x, y)^2\mu(dx)\mu(dy) \right)^{1/2}.
$$

We also define the non-linear operator $\Phi_\kappa$ by

$$
(\Phi_\kappa f)(x) = 1 - e^{-\langle T_\kappa f(x) \rangle}, \quad x \in S,
$$

for $f \geq 0$. Note that for such $f$ we have $0 \leq T_\kappa f \leq \infty$, and thus $0 \leq \Phi_\kappa f \leq 1$. We shall characterize the survival probability $\zeta_\kappa(x)$, and thus $\zeta_\kappa$, in terms of the non-linear operator $\Phi_\kappa$, showing essentially that the function $x \mapsto \zeta_\kappa(x)$ is the maximal fixed point of the non-linear operator $\Phi_\kappa$ (recall (3.4.10)). Again, the survival probability satisfies that $\zeta_\kappa > 0$ precisely when $\|T_\kappa\| > 1$, recall the finite-types case discussed in detail above.

We call a multi-type branching process supercritical when $\|T_\kappa\| > 1$, critical
when $\|T_\kappa\| < 1$, and subcritical when $\|T_\kappa\| > 1$. Then, the above discussion can be summarized by saying that a multi-type branching process survives with positive probability precisely when it is supercritical.

**Poisson branching processes and product kernels: the rank-1 case**

We continue by studying the rank-1 case, where the kernel is of product structure.

Let $\kappa(x, y) = \psi(x)\psi(y)$, so that $T_\kappa(x, y) = \psi(x)\psi(y)\mu(y)$. In this case, we see that $\psi$ is an eigenvector with eigenvalue

$$\|T_\kappa\| = \int_S \psi(y)^2 \mu(dy) = \|\psi\|_{L^2(\mu)}^2.$$  \hspace{1cm} (3.4.17)

Further, the left eigenvector equals $\mu(x)\psi(x)$. Thus, the rank-1 multi-type branching process is supercritical when $\|\psi\|_{L_2(\mu)}^2 > 1$, critical when $\|\psi\|_{L_2(\mu)}^2 = 1$, and subcritical when $\|\psi\|_{L_2(\mu)}^2 < 1$.

The rank-1 case is rather special, and not only since we can explicitly compute the eigenvectors of the operator $T_\kappa$. It also turns out that the rank-1 multi-type case reduces to a single type branching process with mixed Poisson offspring distribution. For this, we recall the construction right below Lemma 3.9. We compute that

$$\lambda(x) = \int_S \psi(x)\psi(y)\mu(dy) = \psi(x)\int_S \psi(y)\mu(y),$$  \hspace{1cm} (3.4.18)

so that an offspring of an individual of type $x$ receives mark $y$ with probability

$$p(x, y) = \frac{T_\kappa(x, y)\mu(y)}{\lambda(x)} = \frac{\psi(x)\psi(y)\mu(y)}{\psi(x)\int_S \psi(z)\mu(dz)} = \frac{\psi(y)\mu(y)}{\int_S \psi(z)\mu(dz)}.$$  \hspace{1cm} (3.4.19)

We conclude that every individual chooses its type independently of the type of its parent. This means that this multi-type branching process reduces to a single type branching process with offspring distribution that is $\text{Poi}(W_\lambda)$, where

$$\mathbb{P}(W_\lambda \in A) = \frac{\int_A \psi(y)\mu(dy)}{\int_S \psi(z)\mu(dz)}.$$  \hspace{1cm} (3.4.20)

This makes the rank-1 setting particularly appealing.

**Poisson branching processes and sum kernels**

For the sum kernel, the analysis becomes slightly more involved, but can still be solved. Recall that $\kappa(x, y) = \alpha(x) + \alpha(y)$ for the sum kernel. Anticipating a nice shape of the eigenvalues and -vectors, we let $\phi(x) = a\alpha(x) + b$, and verify the eigenvalue relation. This leads to

$$(T_\kappa\phi)(x) = \int_S \kappa(x, y)\phi(y)\mu(dy) = \int_S [\alpha(x) + \alpha(y)](a\alpha(y) + b)\mu(dy) \hspace{1cm} (3.4.21)$$

$$= \alpha(x)(a\|a\|_{L^1(\mu)} + b) + (a\|a\|_{L^1(\mu)} + b\|a\|_{L^1(\mu)}) = \lambda(a\alpha(x) + b).$$
3.4 Multi-type branching processes

Solving for $a, b, \lambda$ leads to $a\|a\|_{L^1(\mu)} + b = a\lambda, a\|a\|_{L^2(\mu)}^2 + b\|a\|_{L^1(\mu)} = \lambda b$, so that the vector $(a, b)^T$ is the eigenvector with eigenvalue $\lambda$ of the matrix

\[
\begin{bmatrix}
\|a\|_{L^1(\mu)} & 1 \\
\|a\|_{L^2(\mu)} & \|a\|_{L^1(\mu)}^2
\end{bmatrix}
\]

(3.4.22)

Solving this equation leads to eigenvalues

\[
\lambda = \|a\|_{L^1(\mu)} \pm \|a\|_{L^2(\mu)},
\]

(3.4.23)

and the corresponding eigenvectors $(1, \pm\|a\|_{L^2(\mu)})^T$. Clearly, the maximal eigenvalue equals $\lambda = \|a\|_{L^2(\mu)} + \|a\|_{L^1(\mu)}$, with corresponding $L^2(\mu)$-normalized eigenvector

\[
\phi(x) = \frac{\alpha(x) + \|a\|_{L^2(\mu)}}{2(\|a\|_{L^2(\mu)}^2 + \|a\|_{L^1(\mu)}\|a\|_{L^2(\mu)})}.
\]

(3.4.24)

All other eigenvectors can be chosen to be orthogonal to $a$ and $1$, so that this corresponds to a rank-2 setting.

**Unimodular Poisson branching processes**

In our results, we will be interested in multi-type Poisson branching processes that start from the type distribution $\mu$. Thus, we fix the root $\emptyset$ in the branching process tree, and give it a random type $Q$ satisfying that $P(Q \in A) = \mu(A)$, for any measurable $A \subseteq X$. This corresponds to the *unimodular* setting that is important in random graph settings. The idea is that the total number of vertices with types in $A$ is close to $n\mu(A)$, so that if we pick a vertex uniformly at random, it will have a type in $A$ with asymptotic probability equal to $\mu(A)$.

**Branching process notation**

We now introduce some notation that will be helpful. We let $\text{BP}_{\leq k}$ denote the branching process up to and including generation $k$, where of each individual $v$ in the $k$th generation, we record its type as $Q(v)$. It will be convenient to think of $T$ as being labelled in the Ulam-Harris way (recall Section 1.5), so that a vertex $v$ in generation $k$ has a label $a_1 \cdots a_k$, where $a_i \in \mathbb{N}$. When applied to $\text{BP}$, we will denote this process by $(\text{BP}(t))_{t \geq 1}$, where $\text{BP}(t)$ consists of precisely $t$ vertices (with $\text{BP}(1)$ equal to the root $\emptyset$, as well as its type $Q(\emptyset)$).

**Monotone approximations of kernels**

In what follows, we will often approximate general kernels by kernels with finitely many types, as described in Lemma 3.5. For monotone sequences, we can prove the following convergence result:

**Theorem 3.10** (Monotone approximations of multi-type Poisson branching processes) Let $(\kappa_n)$ be a sequence of kernels such that $\kappa_n(x, y) \nearrow \kappa(x, y)$. Let $\text{BP}_{\leq k}^{(\kappa)}$ denote the first $k$ generations of the Poisson multi-type branching process with kernel $\kappa_n$ and $\text{BP}_{\leq k}^{(\kappa)}$ that of the Poisson multi-type branching process with kernel $\kappa$. Then $\text{BP}_{\leq k}^{(\kappa)} \overset{d}{\longrightarrow} \text{BP}_{\leq k}$. Further, let $\zeta_{\geq k}(x)$ denote the probability that an individual of type $x$ has at least $k$ descendants. Then, $\zeta_{\geq k}^{(\kappa)}(x) \nearrow \zeta_{\geq k}(x)$. 
Phase transition in general inhomogeneous random graphs

Proof. Since \( \kappa_n(x, y) \nrightarrow \kappa(x, y) \), we can write

\[
\kappa(x, y) = \sum_{n \geq 1} \Delta \kappa_n(x, y), \quad \text{where} \quad \Delta \kappa_n(x, y) = \kappa_n(x, y) - \kappa_{n-1}(x, y).
\]

(3.4.25)

We can represent this by a sum of independent Poisson multi-type processes with intensities \( \Delta \kappa_n(x, y) \) and associate a label \( n \) to each particle that arises from \( \Delta \kappa_n(x, y) \). Then, the branching process \( \mathbb{BP}^{(n)}_{\leq k} \) is obtained by keeping all vertices with labels at most \( n \), while \( \mathbb{BP}^{(n)}_{\leq k} \) is obtained by keeping all vertices. Consequently, \( \mathbb{BP}^{(n)}_{\leq k} \xrightarrow{d} \mathbb{BP}^{(n)}_{\leq k} \) follows. Further, \( 1 - \zeta^{(n)}_{\geq k}(x) = \zeta^{(n)}_{\leq k}(x) \) is a continuous function of \( \mathbb{BP}^{(n)}_{\leq k} \) and thus also converges.

3.5 Local weak convergence for inhomogeneous random graphs

In this section, we prove the local weak convergence of \( \text{IRG}_n(\kappa_n) \) in general. Our main result is as follows:

**Theorem 3.11 (Locally tree-like nature \( \text{IRG}_n(\kappa_n) \))**  Assume that \( \kappa_n \) is an irreducible graphical kernel converging to some limiting kernel \( \kappa \). Then \( \text{IRG}_n(\kappa_n) \) converges in probability in the local weak sense to the unimodular multi-type marked Galton-Watson tree, where

- the root has offspring distribution \( \text{Poi}(W) \) and type \( Q \) with distribution

  \[
  \mathbb{P}(Q \in A) = \mu(A); \quad (3.5.1)
  \]

- a vertex of type \( x \) has offspring distribution \( \text{Poi}(\lambda(x)) \), with

  \[
  \lambda(x) = \int_S \kappa(x, y) \mu(dy), \quad (3.5.2)
  \]

  and each of its offspring receives an independent type with distribution \( Q(x) \) given by

  \[
  \mathbb{P}(Q(x) \in A) = \int_A \kappa(x, y) \mu(dy), \quad (3.5.3)
  \]

The proof of Theorem 3.11 follows a usual pattern. We start by proving Theorem 3.11 for the finite-type case, and then use finite-type approximations to extend the proof to the infinite-type case.

3.5.1 Local weak convergence: finitely many types

In order to get started for the proof of (2.3.10) for Theorem 3.11 in the finite-type case, we introduce some notation. Fix a rooted tree \( (t, q) \) of \( k \) generations, and where the vertex \( v \in V(t) \) has type \( q(v) \) for all \( v \in V(t) \). It will again be convenient to think of \( t \) as being labelled in the Ulam-Harris way, so that a vertex \( v \) in generation \( k \) has a label \( \emptyset a_1 \cdots a_k \), where \( a_i \in \mathbb{N} \).
3.5 Local weak convergence for inhomogeneous random graphs

Let

\[ N_n(t, q) = \sum_{v \in [n]} 1_{B_k^{(n, q)}(v) \simeq (t, q)} \]  

(3.5.4)

denote the number of vertices whose local neighborhood up to generation \( t \) including their types equals \((t, q)\). Here, in \( B_k^{(n, q)}(v) \), we record of the types of the vertices in \( B_k^{(G, \omega)}(v) \). Theorem 2.13 implies that in order to prove Theorem 4.1, we need to show that

\[ \frac{N_n(t, q)}{n} \xrightarrow{p} P(BP \leq k \simeq (t, q)). \]  

(3.5.5)

Indeed, since there are only a finite number of types, (3.5.5) also implies that

\[ \frac{N_n(t)}{n} \xrightarrow{p} \rho(t), \]  

(3.5.6)

is the probability that the branching process produces a certain tree (and ignoring the types). To this, we can then apply Theorem 2.13. Alternatively, we can consider (3.5.5) as convergence of marked graphs as discussed in Section 1.6, which is stronger, but we do not pursue this direction here further.

To prove (3.5.5), we will use a second moment method. We first prove that the first moment

\[ \mathbb{E}[N_n(t, q)]/n \rightarrow P(BP \leq k \simeq (t, q)), \]  

(3.5.7)

where \( o \in [n] \) is a vertex chosen uniformly at random. Our aim is to prove that

\[ P(BP \leq k \simeq (t, q)) = \prod_{v \in T : |v| \leq k-1} e^{-\lambda(q(v))} \frac{\lambda(q(v))}{d_v!} \prod_{j=1}^{d_v} \frac{\kappa(q(v), q(vj))}{\lambda(q(v))} = e^{-\lambda(q(v))} \frac{1}{d_v!} \prod_{j=1}^{d_v} \kappa(q(v), q(vj)) \mu(vj), \]  

(3.5.8)

since we first draw a Poisson \( \lambda(q(v)) \) number of children, assign an order to them and then assign a type \( q \) to each of them with probability \( \kappa(q(v), q)/\lambda(q(v)) \). This is true independently for all \( v \in V(t) \) with \( |v| \leq k - 1 \), so that

\[ P(BP \leq k \simeq (t, q)) = \prod_{v \in T : |v| \leq k-1} e^{-\lambda(q(v))} \frac{1}{d_v!} \prod_{j=1}^{d_v} \kappa(q(v), q(vj)) \mu(vj). \]  

(3.5.9)

Local weak convergence: first moment

We start by noting that

\[ \frac{1}{n} \mathbb{E}[N_n(t)] = P(B_k^{(n, q)}(o) \simeq (t, q)), \]  

(3.5.7)

where \( o \in [n] \) is a vertex chosen uniformly at random. Our aim is to prove that

\[ P(B_k^{(n, q)}(o) \simeq (t, q)) \rightarrow P(BP \leq k \simeq (t, q)). \]
110 Phase transition in general inhomogeneous random graphs

For a comparison to the graph exploration, it will turn out to be convenient to rewrite this probability slightly. Let \( t_{\leq k-1} = \{ v : |v| \leq k - 1 \} \) denote the vertices in the first \( k - 1 \) generations of \( t \) and let \( |t_{\leq k-1}| \) denote its size. We can order the elements of \( t_{\leq k-1} \) in their lexicographic ordering as \((v_i)_{i=1}^{|t_{\leq k-1}|}\). The lexicographic ordering it the ordering in which the vertices are explored in the breadth-first exploration. Then we can rewrite

\[
P(BP_{\leq k} \simeq (t, q)) = \prod_{i=1}^{|t_{\leq k-1}|} e^{-\lambda(q(v_i))} \frac{1}{d_{v_i}} \prod_{j=1}^{d_{v_i}} \kappa(q(v_i), q(v_j)) \mu(v_i, j). \tag{3.5.10}
\]

Let us now turn to IRG\(_n\)(\(\kappa_n\)). Fix a vertex \( v \in [n] \) of type \( q(v) \). The probability of seeing a sequence of \( d_v \) neighbors of (ordered) types \((q(v1), \ldots, q(vd))\) equals

\[
\frac{1}{d_v} \prod_{q \in \mathcal{S}} \left( 1 - \kappa_n(q(v), q) \right)^{n_q-m_q} \prod_{j=1}^{d_v} \kappa_n(q(v), q(vj)) \frac{n_q - m_{q(v)}}{n} \left[ n_q - m_{q(vj)}(j - 1) \right], \tag{3.5.11}
\]

where \( m_q = \# \{ i : q(vi) = q \} \) is the number of type \( q \) vertices in \((q(v1), \ldots, q(vd))\) and \( m_{q(vj)}(j) = \# \{ i \leq j : q(vi) = q \} \) is the number of type \( q \) vertices in \((q(v1), \ldots, q(vj))\). Here, the first factor is since we assign an ordering uniformly at random, the second since all other edges (except for the specified ones) need to be absent, and the third specifies that the edges to vertices of the (ordered) sequence of types are present. When \( n \to \infty \), and since \( n_q/n \to \mu(q) \), \( \kappa_n(q(v), q) \to \kappa(q(v), q) \) for every \( q \in \mathcal{S} \),

\[
\frac{1}{d_v} \prod_{q \in \mathcal{S}} \left( 1 - \kappa_n(q(v), q) \right)^{n_q-m_q} \prod_{j=1}^{d_v} \kappa_n(q(v), q(vj)) \frac{n_q - m_{q(v)}}{n} \left[ n_q - m_{q(vj)}(j - 1) \right] \to e^{-\lambda(q(v))} \frac{1}{d_v} \prod_{i=1}^{d_v} \kappa(q(v), q(vj)) \mu(q(vj)), \tag{3.5.12}
\]

as required. The above computation, however, ignores the depletion-of-points effect that fewer vertices participate in the course of the exploration. To describe this, recall the lexicographic ordering of the elements in \( t_{\leq k-1} \) as \((v_i)_{i=1}^{|t_{\leq k-1}|}\), and, for a type \( q \), let \( m_q(i) = \# \{ j \in [i] : q(vi) = q \} \) denote the number of type \( q \) individuals in \((t, q)\) encountered up to and including the \( i \)th exploration. Then,

\[
P(B_{k-q}^\ast) \simeq (t, q) = \prod_{i=1}^{|t_{\leq k-1}|} \frac{1}{d_{v_i}} \prod_{q \in \mathcal{S}} \left( 1 - \kappa_n(q(v_i), q) \right)^{n_q-m_q(i-1)} \tag{3.5.13}
\]

\[
\times \prod_{j=1}^{d_{v_i}} \kappa_n(q(v_i), q(v_j)) \frac{n_q - m_{q(vj)}(i + j - 1)}{n} \left[ n_q - m_{q(vj)}(i + j - 1) \right].
\]

As \( n \to \infty \), this converges to the right hand side of (3.5.10), as required. This completes the proof of (3.5.7), and thus the convergence of the first moment. \[\square\]
3.5 Local weak convergence for inhomogeneous random graphs

Here, we study the second moment of $N_n(t, q)$, and show that it is close to the first moment squared:

**Lemma 3.12** (Concentration of the number of trees) As $n \to \infty$,

$$\frac{\text{Var}(N_n(t, q)^2)}{n^2} \to 0. \quad (3.5.14)$$

Consequently, $N_n(t, q)/n \xrightarrow{p} \mathbb{P}(BP \leq k \simeq (t, q))$.

**Proof** We start by computing

$$\mathbb{E}[N_n(t, q)^2] = \mathbb{P}(B_{o_1, o_2}^{(n, q)}(k) = (t, q)), \quad (3.5.15)$$

where $o_1, o_2 \in [n]$ are two vertices chosen uniformly at random from $[n]$, independently.

We first claim that with high probability and any $k$ fixed, $\text{dist}_{\text{IRG}_n(kn)}(o_1, o_2) > 2k$. For this, we take the expectation with respect to $o_2$ to obtain that

$$\mathbb{P}(\text{dist}_{\text{IRG}_n(kn)}(o_1, o_2) \leq 2k) = \frac{1}{n} \mathbb{E}[\|B_{2k}^{(n, q)}(o_1)\|]. \quad (3.5.16)$$

By (3.5.13),

$$\mathbb{P}(B_{2k}^{(n, q)}(o_1) \simeq (t, q)) \leq \prod_{i=1}^{[\ell \leq k]} \frac{1}{d_{v_i}} \prod_{j=1}^{d_{v_i}} \kappa_n(q(v_i), q(v_{i, j})) \frac{n_{q_j}}{n}. \quad (3.5.17)$$

Denote the mean-offspring matrix $M^{(n)}$ by $M^{(n)}_{ij} = \kappa_n(q(v_i), q(v_{i, j})) n_{q_j}/n$, and write $\mu_i^{(n)} = n_i/n$. Then,

$$\mathbb{P}(\text{dist}_{\text{IRG}_n(kn)}(o_1, o_2) \leq 2k) \leq \frac{1}{n} \sum_{\ell \leq 2k} (\mu^{(n)})^T (M^{(n)})^\ell 1. \quad (3.5.18)$$

Let $\lambda_n$ denote the largest eigenvalue of $M^{(n)}$, so that

$$\mathbb{P}(\text{dist}_{\text{IRG}_n(kn)}(o_1, o_2) \leq 2k) \leq \frac{1}{n} \sum_{\ell \leq 2k} \|\mu^{(n)}\|_2 \lambda_n^\ell \|1\|_2 \quad (3.5.19)$$

$$= \frac{\|\mu^{(n)}\|_2 \|1\|_2}{n} \sum_{\ell \leq 2k} \lambda_n^\ell$$

$$= \frac{\|\mu^{(n)}\|_2 \|1\|_2}{n} \frac{\lambda_n^{2k+1} - 1}{\lambda_n - 1}.$$

As $n \to \infty$, $M^{(n)}_{ij} \to (T_{n})_{ij}$ for every $i, j \in [m]$, so that also $\lambda_n \to \lambda = \|T_n\|$. We conclude that, for every $k \in \mathbb{N}$ fixed, $\mathbb{P}(\text{dist}_{\text{IRG}_n(kn)}(o_1, o_2) \leq 2k) = o(1)$. See Exercises 3.23–3.24 for extensions of this result for distances in inhomogeneous random graphs with finitely many types.
We conclude that

$$\frac{\mathbb{E}[N_n(t,q)^2]}{n^2} = \mathbb{P}(B_k^{(n,\kappa)}(o_1), B_k^{(n,\kappa)}(o_2) \simeq (t, q), o_2 \notin B_{2k}^{(n)}(o_1)) + o(1).$$

(3.5.20)

We now condition on $B_{2k}^{(n,\kappa)}(k) \simeq (t, q)$, and write

$$\mathbb{P}(B_k^{(n,\kappa)}(o_1), B_k^{(n,\kappa)}(o_2) \simeq (t, q), o_2 \notin B_{2k}^{(n)}(o_1))$$

$$= \mathbb{P}(B_k^{(n,\kappa)}(o_2) \simeq (t, q) | B_{2k}^{(n)}(k) \simeq (t, q), o_2 \notin B_{2k}^{(n)}(o_1))$$

$$\times \mathbb{P}(B_k^{(n,\kappa)}(o_1) \simeq (t, q), o_2 \notin B_{2k}^{(n)}(o_1)).$$

(3.5.21)

We already know that $\mathbb{P}(B_k^{(n,\kappa)}(o_1) \simeq (t, q)) \rightarrow \mathbb{P}(BP_{\leq k} \simeq (t, q))$, so that also

$$\mathbb{P}(B_{2k}^{(n,\kappa)}(k) \simeq (t, q), o_2 \notin B_{2k}^{(n)}(o_1)) \rightarrow \mathbb{P}(BP_{\leq k} \simeq (t, q)).$$

(3.5.22)

In Exercise 3.25, you prove that indeed (3.5.22) holds.

We next investigate the conditional probability, by noting that, conditionally on $B_k^{(n,\kappa)}(o_1) \simeq (t, q)$ and $o_2 \notin B_{2k}^{(n)}(o_1)$, the probability that $B_k^{(n,\kappa)}(o_2) = (t, q)$ is the same as the probability that $B_k^{(n,\kappa)}(o_2) \simeq (t, q)$ in $\text{IRG}_n(\kappa_n)$ which is obtained by removing the vertices in $B_k^{(n,\kappa)}(o_1)$ as well as the edges from them. We conclude that the resulting random graph has $n' = n - |t|$ vertices, and $n'_q = n_q - m_q$ vertices of type $q \in [m]$, where $m_q$ is the number of type $q$ vertices in $(t, q)$. Further, $\kappa_n(i, j) = \kappa_n(i, j)n'/n$. The whole point is that $\kappa(i, j) \rightarrow \kappa(i, j)$ and $n'_q/n \rightarrow \mu(q)$ still hold. Therefore, also

$$\mathbb{P}(B_k^{(n,\kappa)}(o_2) \simeq (t, q) | B_k^{(n,\kappa)}(o_1) = (t, q), o_2 \notin B_{2k}^{(n)}(o_1))$$

$$\rightarrow \mathbb{P}(BP_{\leq k} \simeq (t, q)).$$

(3.5.23)

and we have proved that $\mathbb{E}[N_n(t,q)^2]/n^2 \rightarrow \mathbb{P}(BP_{\leq k} \simeq (t, q))^2$. From this, (3.5.14) follows directly since $\mathbb{E}[N_n(t,q)]/n \rightarrow \mathbb{P}(BP_{\leq k} \simeq (t, q))$. As a result, $\mathbb{E}[N_n(t,q)]/n$ is concentrated and thus $N_n(t,q)/n \rightarrow \mathbb{P}(BP_{\leq k} \simeq (t, q))$, as required.

Lemma 3.12 completes the proof of Theorem 3.11 in the finite-types case. ☐

3.5.2 Local weak convergence: infinitely many types

We next extend the proof of Theorem 3.11 to the infinite-types case. We follow the strategy in Section 3.3.3.

Fix a general sequence of graphical kernels $(\kappa_n)$. Again define $\kappa^{-}_n$ by (3.3.17), so that $\kappa_n \geq \kappa^{-}_n$. Couple $\text{IRG}_n(\kappa^{-}_n)$ and $\text{IRG}_n(\kappa_n)$ such that $E(\text{IRG}_n(\kappa^{-}_n)) \subseteq E(\text{IRG}_n(\kappa_n))$. Let $\varepsilon' > 0$ be given. Recall (3.3.22), which shows that, whp, we can take $m$ so large that the bound

$$E(\text{IRG}_n(\kappa^{-}_n)) \subseteq E(\text{IRG}_n(\kappa_n)) = \sum_{u \in [n]} (D_u - D_u^{(m)}) \leq \varepsilon'n$$

holds whp. We let $K$ denote the maximal degree in $T$. Let $N_n^{(m)}(t,q)$ denote $N_n(t,q)$ for the kernel $\kappa^{-}_m$ (and keep $N_n(t,q)$ for (3.5.4) for the kernel $\kappa_n$). If a vertex $v$ is such that $B_k^{(n,\kappa)}(v) \simeq (t,q)$ in $\text{IRG}_n(\kappa^{-}_m)$, but not in $\text{IRG}_n(\kappa_n)$, or
vice versa, then one of the vertices in $B_{k-1}^{(n;q)}(v)$ needs to have a different degree in $\text{IRG}_n(\kappa_m^-)$ than in $\text{IRG}_n(\kappa_n)$. Thus, 

$$|N_n^{(m)}(t, q) - N_n(t, q)| \leq \sum_{u,v} \mathbb{1}_{\{u \in B_{k-1}^{(n;q)}(v), B_k^{(n;q)}(v) \simeq (t,q) \text{ in IRG}_n(\kappa_m^-)\}} \mathbb{1}_{\{D_u \neq D_u^{(m)}\}} + \sum_{u,v} \mathbb{1}_{\{u \in B_{k-1}^{(n;q)}(v), B_k^{(n;q)}(v) \simeq (t,q) \text{ in IRG}_n(\kappa_n)\}} \mathbb{1}_{\{D_u^{(m)} \neq D_u\}}. \quad (3.5.25)$$

Recall that the maximal degree of any vertex in $T$ is $K$. Further, if $B_{k-1}^{(n;q)}(v) \simeq (t,q)$ and $u \in B_k^{(n;q)}(v)$, then all the vertices on the path between $u$ and $v$ have degree at most $K$. Therefore,

$$\sum_v \mathbb{1}_{\{u \in B_{k-1}^{(n;q)}(v), B_k^{(n;q)}(v) \simeq (t,q) \text{ in IRG}_n(\kappa_m^-)\}} \leq \sum_{\ell \leq k-1} K^\ell \leq \frac{K^k - 1}{K - 1}, \quad (3.5.26)$$

and in the same way,

$$\sum_v \mathbb{1}_{\{u \in B_{k-1}^{(n;q)}(v), B_k^{(n;q)}(v) \simeq (t,q) \text{ in IRG}_n(\kappa_n)\}} \leq \frac{K^k - 1}{K - 1}. \quad (3.5.27)$$

We thus conclude that whp

$$|N_n^{(m)}(t, q) - N_n(t, q)| \leq 2\frac{K^k - 1}{K - 1} \sum_{u \in [n]} \mathbb{1}_{\{D_u^{(m)} \neq D_u\}} \quad (3.5.28)$$

and

$$\leq 2\frac{K^k - 1}{K - 1} \sum_{u \in [n]} (D_u^{(m)} - D_u) \leq 2\frac{K^k - 1}{K - 1} \varepsilon' n. \quad (3.5.29)$$

Taking $\varepsilon' = \varepsilon(K-1)/[2(K^k - 1)]$, we thus obtain that

$$|N_n^{(m)}(t, q) - N_n(t, q)| \leq \varepsilon n, \quad (3.5.29)$$

as required.

For $N_n^{(m)}(t, q)$, we can use Theorem 3.11 in the finite-types case to obtain that

$$\frac{1}{n} N_n^{(m)}(t, q) \xrightarrow{p} \mathbb{P}(\text{BP}_{\leq k}^{(m)} \simeq (t,q)). \quad (3.5.30)$$

The fact that $m$ can be taken so large that $|\mathbb{P}(\text{BP}_{\leq k}^{(m)} \simeq (t,q)) - \mathbb{P}(\text{BP}_{\leq k} \simeq (t,q))| \leq \varepsilon$ follows from Theorem 3.10.

\[ \square \]

3.5.3 Comparison to branching processes

In this section, we describe a beautiful comparison of the neighborhoods of a uniformly chosen vertex in rank-1 inhomogeneous random graphs, such as the generalized random graph, the Chung-Lu model and the Norros-Reittu model, and a marked branching process. This comparison is particularly pretty when considering the Norros-Reittu model, where there is an explicit stochastic domination result of these neighborhoods are bounded by a unimodular branching.
We start by describing the result for the rank-1 setting, after which we extend it to kernels with finitely many types.

Stochastic domination of clusters by a branching process

We shall dominate the cluster of a vertex in the Norros-Reittu model by the total progeny of a two-stage branching processes with mixed Poisson offspring. This domination is such that we also control the difference, and makes the heuristic argument below Theorem 3.17 precise.

We now describe the cluster exploration of a uniformly chosen vertex \(o \in [n]\).

Define the mark distribution to be the random variable \(M\) with distribution
\[
P(M = m) = \frac{w_m}{\ell_n}, \quad m \in [n].
\]
This is given in (3.5.31).

Let \((X_w, M_w)\) be a collection of independent random variables, where

(a) the number of children of the root \(X_\emptyset\) has a mixed Poisson distribution with random parameter \(w_{M_\emptyset}\), where \(M_\emptyset\) is uniformly chosen in \([n]\);

(b) \(X_w\) has a mixed Poisson distribution with random parameter \(w_{M_w}\), where \((M_w)_{w \neq \emptyset}\) are i.i.d. random marks with distribution (3.5.31) independently of \(M_\emptyset\).

We call \((X_w, M_w)\) a marked mixed-Poisson branching process (MMPBP).

Clearly, \(w_\emptyset = w_{M_\emptyset}\) has distribution \(W_n\) defined in (1.3.10), while the distribution of \(w_{M_w}\) for each \(w\) with \(|w| \geq 1\) is i.i.d. with distribution \(w_{M_\emptyset}\) given by
\[
P(w_M \leq x) = \sum_{m=1}^{n} 1\{w_m \leq x\} P(M = m) = \frac{1}{\ell_n} \sum_{m=1}^{n} w_m 1\{w_m \leq x\}
\]
where \(W_n^*\) is the size-biased distribution of \(W_n\) and \(F_n^*\) is given by
\[
F_n^*(x) = \frac{1}{\ell_n} \sum_{m=1}^{n} w_m 1\{w_m \leq x\}.
\]

When we are only interested in numbers of individuals, then we obtain a unimodular branching process since the random variables \((X_w)\) are independent, and the random variables \((X_w)_{w \neq \emptyset}\) are i.i.d. However, in the sequel, we make explicit use of the marks \((M_w)_{w \neq \emptyset}\), as the complete information \((X_w, M_w)\) gives us a way to retrieve the cluster of the vertex \(M_\emptyset\), something that would not be possible on the basis of \((X_w)\) only.

In order to define the cluster exploration in \(NR_n(w)\), we introduce a thinning that guarantees that we only inspect a vertex once. We think of \(M_w\) as being the vertex label in \(NR_n(w)\) of the tree vertex \(w\), and \(X_w = \text{Poi}(w_{M_w})\) as its potential number of children. These potential children effectively become children when

\footnote{In van den Esker et al. (2008), the unimodular branching process is called a delayed branching process, and also the term two-stage branchiag process has been used.}
their marks correspond to vertices in $\text{NR}_n(w)$ that have not yet appeared. The thinning ensures this. To describe the thinning, we set $\varnothing$ unthinned, and, for $w$ with $w \neq \varnothing$, we thin $w$ when either (i) one of the tree vertices on the (unique) path between the root $\varnothing$ and $w$ has been thinned, or (ii) when $M_w = M_w'$ for some unthinned vertex $w' < w$. We now make the connection between the thinned marked mixed Poisson branching process and the cluster exploration precise:

**Proposition 3.13** (Clusters as thinned marked branching processes) The cluster of a uniformly chosen vertex $C(o)$ is equal in distribution to $\{M_w : w \text{ unthinned}\}$, the marks of unthinned vertices encountered in the marked mixed Poisson branching process up to the end of the exploration. Similarly, the set of vertices at graph distance $k$ from $o$ has the same distribution as

$$\{M_w : \text{unthinned}, |w| = k\}_{k \geq 0}.$$  

**(3.5.34)**

**Proof** We prove the two statements simultaneously. By construction, the distribution of $o$ is the same as that of $M_\varnothing$, the mark of the root of the marked mixed Poisson branching process. We continue by proving that the direct neighbors of the root $\varnothing$ agree in both constructions. In $\text{NR}_n(w)$, the direct neighbors are equal to $\{j \in [n] \setminus \{l\} : I_{lj} = 1\}$, where $(I_{lj})_{j \in [n] \setminus \{l\}}$ are independent $\text{Be}(p_{lj})$ random variables with $p_{lj} = 1 - e^{-w_lw_j/\ell_n}$.

We now prove that the same is true for the marked mixed Poisson branching process. Conditionally on $M_\varnothing = l$, the root has a $\text{Poi}(w_l)$ number of children, where these $\text{Poi}(w_l)$ offspring receive i.i.d. marks. We make use of the fundamental ‘thinning’ property of the Poisson distribution in Lemma 3.9. By Lemma 3.9, the random vector $(X_{\varnothing,j})_{j \in [n]}$, where $X_{\varnothing,j}$ is the number of offspring of the root that receive mark $j$, is a vector of independent Poisson random variables with parameters $w_lw_j/\ell_n$. Due to the thinning, a mark occurs precisely when $X_{\varnothing,j} \geq 1$. Therefore, the mark $j$ occurs, independently for all $j \in [n]$, with probability $1 - e^{-w_lw_j/\ell_n} = p_{jk}^{(NR)}$. This proves that the set of marks of children of the root in the MMPBD has the same distribution as the set of neighbors of the chosen vertex in $\text{NR}_n(w)$.

Next, we look at the number of new elements of $C(o)$ neighboring the vertex which has received word $w$. First, condition on $M_w = l$, and assume that $w$ is not thinned. Conditionally on $M_w = l$, the number of children of $w$ in the MMPBP has distribution $\text{Poi}(w_l)$. Each of these $\text{Poi}(w_l)$ children receives an i.i.d. mark. Let $X_{w,j}$ denote the number of children of $w$ that receive mark $j$.

By Lemma 3.9, $(X_{w,j})_{j \in [n]}$ is again a vector of independent Poisson random variables with parameters $w_lw_j/\ell_n$. Due to the thinning, a mark appears within the offspring of individual $w$ precisely when $X_{w,j} \geq 1$, and these events are independent. In particular, for each $j$ that has not appeared as the mark of an unthinned vertex, the probability that it occurs equals $1 - e^{-w_lw_k/\ell_n} = p_{jk}^{(NR)}$, as required. 

\qed
Stochastic domination by branching processes: finite-type case

The rank-1 setting described above is special, since the marks of vertices in the tree are independent random variables in that they do not depend on the type of their parent. This is in general not true. We next describe how the result can be generalized. We restrict to the finite-type case.

Let us introduce some notation. Recall that $n_i$ denotes the number of vertices of type $i \in [m]$, and write $n_{\leq i} = \sum_{j \leq i} n_j$. Define the intervals $I_i = [n_{\leq i}] \setminus [n_{< i}]$ (where, by convention, $I_1 = [n_1]$). We note that all vertices in the intervals $I_i$ play the same role, and this will be used crucially in our coupling.

We now describe the cluster exploration of a uniformly chosen vertex $o \in [n]$ of type $i$. Define the mark distribution to be the random variable $M(j)$ with distribution

$$P(M(j) = \ell) = \frac{1}{n_i}, \quad \ell \in I_j. \quad (3.5.35)$$

Let $(X_w)_{w}$ be a collection of independent random variables, where

(a) the number of children of the root $X_{\emptyset}$ has a mixed Poisson distribution with random parameter $\lambda_n(i) = \sum_j \kappa_n(i,j)$, and each of the children of $\emptyset$ receives a type $T_w$, where $T_w = j$ with probability $\kappa_n(i,j)/\lambda_n(i)$;

(b) given that a vertex $w$ has type $j$, it receives a mark $M_w(j)$ with distribution in (3.5.35).

(b) $X_w$ has a mixed Poisson distribution with random parameter $w_{M_w}$, where $(M_w)_{w \neq \emptyset}$ are i.i.d. random marks with distribution (3.5.31) independently of $M_\emptyset$.

We call $(X_w, T_w, M_w)_{w}$ a marked multi-type Poisson branching process. Then, the following extension of Proposition 3.13 holds:

**Proposition 3.14** (Clusters as thinned marked multi-type branching processes) The cluster of a vertex $\mathcal{C}(v)$ of type $i$ is equal in distribution to $\{M_w: w$ unthinned$, \}|_{w \neq \emptyset}$, the marks of unthinned vertices encountered in the marked multi-type Poisson branching process up to the end of the exploration. Similarly, the set of vertices at graph distance $k$ from $o$ has the same distribution as

$$(\{M_w: w$ unthinned$, |w| = k\})_{k \geq 0}. \quad (3.5.36)$$

You are asked to prove Proposition 3.14 in Exercise 3.28.

### 3.5.4 Local weak convergence to unimodular trees for GRG\(_n\)(\(w\))

We close this section by investigating the locally tree-like nature of the generalized random graph. Our main result is as follows:

**Theorem 3.15** (Locally tree-like nature GRG\(_n\)(\(w\))) Assume that Condition 1.1(a)-(b) holds. Then GRG\(_n\)(\(w\)) converges locally-weakly in probability to the
unimodular Galton-Watson tree with offspring distribution \((p_k)_{k \geq 0}\) given by
\[
p_k = \mathbb{P}(D = k) = \mathbb{E}\left[e^{-w} \frac{W^k}{k!}\right]. \tag{3.5.37}
\]
This result also applies to \(NR_n(w)\) and \(CL_n(w)\) under the same conditions.

Theorem 3.15 follows directly from Theorem 3.11. However, we will also give
an alternative proof by following a different route, by relying on the local tree-like
nature of \(CM_n(d)\) proved in Theorem 4.1, and the relation between \(GRG_n(w)\)
and \(CM_n(d)\) discussed in Section 1.3 and using Theorem 1.7. This approach is
interesting in itself, since it allows for general proofs for \(GRG_n(w)\) by proving
the result first for \(CM_n(d)\), and then merely extending it to \(GRG_n(w)\). We will
frequently rely on such a proof.

3.6 The phase transition for inhomogeneous random graphs

In this section, we discuss the phase transition in \(IRG_n(\kappa)\). The main result
shows that there is a giant component when the associated multi-type branching
process is supercritical, while otherwise there is not:

**Theorem 3.16 (Giant component of IRG)** Let \((\kappa_n)\) be a sequence of irreducible
graphical kernels with limit \(\kappa\), and let \(C_{\text{max}}\) denote the largest connected compo-
nent of \(IRG_n(\kappa_n)\). Then,
\[
\frac{|C_{\text{max}}|}{n} \xrightarrow{\mathbb{P}} \zeta_\kappa. \tag{3.6.1}
\]
In all cases \(\zeta_\kappa < 1\), while \(\zeta_\kappa > 0\) precisely when \(\|T_\kappa\| > 1\).

Theorem 3.16 is a generalization of the law of large numbers for the largest
connected component in [Volume 1, Theorem 4.8] for \(ER_n(\lambda/n)\) (see Exercise
3.31).

We do not give a complete proof of Theorem 4.8 in this chapter. The upper
bound follows directly from Theorem 3.11, together with the realization in (2.5.6).
For the lower bound, it suffices to prove this for kernels with finitely-many types,
by Lemma 3.5. This will be done in Chapter 6. We close this section by discussing
a few examples of Theorem 3.16:

**The bipartite random graph**

We let \(n\) be even and take \(S = \{1, 2\}\) and
\[
\kappa_n(x, y) = \kappa(x, y) = \lambda \mathbb{1}_{\{x \neq y\}} / 2. \tag{3.6.2}
\]
Thus, for \(i < j\), the edge probabilities \(p_{ij}\) given by (3.2.7) are equal to \(\lambda / (2n)\)
(for \(2n > \lambda\) when \(i \in [n/2]\) and \(j \in [n] \setminus [n/2]\).

In this case, \(\|T_\kappa\| = \lambda\) with corresponding eigenfunction \(f(x) = 1\) for all \(x \in S\).
Thus, Theorem 3.16 proves that there is a phase transition at \(\lambda = 2\). Furthermore,
the function \(\zeta_\lambda(x)\) reduces to the single value \(\zeta_\lambda / 2\), which is the survival probability
of a Poisson branching process with mean offspring $\lambda/2$. This is not surprising, since the degree of each vertex is $\text{Bin}(n/2, \lambda/n)$, so the bipartite random graph of size $n$ is quite closely related the Erdős-Rényi random graph of size $n/2$.

The finite-type case

The bipartite random graph can also be viewed as a random graph with two types of vertices (i.e., the vertices $[n/2]$ and $[n] \setminus [n/2]$). We now generalize the results to the finite-type case, in which we have seen that $\kappa_n$ is equivalent to an $r \times r$-matrix $(\kappa_n(i, j))_{i,j \in [r]}$, where $r$ denotes the number of types. In this case, IRG$_n(\kappa)$ has vertices of $r$ different types (or colors), say $n_i$ vertices of type $i$, with two vertices of type $i$ and $j$ joined by an edge with probability $n^{-1} \kappa_n(i, j)$ (for $n \geq \max \kappa_n$).

This case has been studied by Söderberg (2002, 2003a, b), who noted Theorem 3.16 in this case. Exercises 3.29–3.30 investigate the phase transition in the finite-type case.

The random graph with prescribed expected degrees

We next consider the Chung-Lu model or expected degree random graph, where $\kappa_n$ is given by (3.2.12), i.e., $\kappa_n(i/n, j/n) = w_i w_j / E[W_n]$.

We first assume that Condition 1.1(a)-(c) holds, so that in particular $E[W^2] < \infty$, where $W$ has distribution function $F$. A particular instance of this case is the choice $w_i = [1 - F]^{-1}(i/n)$ in (1.3.15). In this case, the sequence $(\kappa_n)$ converges to $\kappa$, where the limit $\kappa$ is given by (recall (3.9.2))

$$\kappa(x, y) = \psi(x) \psi(y) / E[W],$$

where $\psi(x) = [1 - F]^{-1}(x)$ Then, we note that for each $f \geq 0$ with $\|f\|_2 = 1$,

$$\langle T_\kappa f \rangle(x) = \psi(x) \int_S \psi(x) f(x) \mu(dx) / \int_S \psi(x) \mu(dx),$$

so that $\|T_\kappa f\|_2 = \int_S \psi(x) f(x) \mu(dx) / \int_S \psi(x) \mu(dx)$, which is maximal when $f(x) = \psi(x)/\|\psi\|_2$. We conclude that $\|T_\kappa\| = \|\psi\|_2^2 / \int_S \psi(x) \mu(dx) = E[W^2]/E[W]$. Thus,

$$\|T_\kappa\| = E[W^2]/E[W],$$

and we recover the results in Chung and Lu (2004, 2006b) in the case where $E[W^2] < \infty$. In the case where $E[W^2] = \infty$, on the other hand, we see that $\|T_\kappa f\|_2 = \infty$ for every $f$ with $\|f\|_2 = 1$ such that $\int_S \psi(x) f(x) \mu(dx) \neq 0$, so that $\|T_\kappa\| = \infty$, so that CL$_n(u)$ is always supercritical in this regime.

While Theorem 3.16 identifies the phase transition in IRG$_n(\kappa_n)$, for the rank-1 setting, we will prove quite some stronger results as we discuss now. In its statement, we denote the complexity of a component $\mathcal{C}$ to be equal to $E(\mathcal{C}) - V(\mathcal{C}) + 1$, which equals the maximal number of edges that need to be removed to turn $\mathcal{C}$ into a tree. The main result is as follows:

**Theorem 3.17** (Phase transition in generalized random graphs) Suppose that Condition 1.1(a)-(b) hold and consider the random graphs GRG$_n(u)$, CL$_n(u)$ or NR$_n(u)$, letting $n \to \infty$. Denote $p_k = P(\text{Poi}(W) = k)$ as defined below (1.3.22).
Let $\mathcal{C}_{\max}$ and $\mathcal{C}_{(2)}$ be the largest and second largest components of GRG$_n$($\mathbf{w}$), CL$_n$($\mathbf{w}$) or NR$_n$($\mathbf{w}$).

(a) If $\nu = \mathbb{E}[W^2]/\mathbb{E}[W] > 1$, then there exist $\xi \in (0, 1), \zeta \in (0, 1)$ such that

\[
|\mathcal{C}_{\max}|/n \xrightarrow{\nu} \zeta, \\
v_k(\mathcal{C}_{\max})/n \xrightarrow{\nu} p_k(1 - \xi^k), \text{ for every } k \geq 0, \\
|E(\mathcal{C}_{\max})|/n \xrightarrow{\nu} \frac{1}{2}\mathbb{E}[W](1 - \zeta^2).
\]

while $|\mathcal{C}_{(2)}|/n \xrightarrow{\nu} 0$ and $|E(\mathcal{C}_{(2)})|/n \xrightarrow{\nu} 0$. Further, $\frac{1}{2}\mathbb{E}[W](1 - \zeta^2) < \zeta$, so that the complexity of the giant is linear.

(b) If $\nu = \mathbb{E}[W^2]/\mathbb{E}[W] \leq 1$, then $|\mathcal{C}_{\max}|/n \xrightarrow{\nu} 0$ and $|E(\mathcal{C}_{\max})|/n \xrightarrow{\nu} 0$.

The above results apply to GRG$_n$($\mathbf{w}$) and CL$_n$($\mathbf{w}$) under the same conditions.

The proof of Theorem 3.17 is deferred to Section 4.2.3 in Chapter 4, where a similar result is proved for the configuration model. By the strong relation between the configuration model and the generalized random graph, this result can be seen to imply Theorem 3.17.

Let us discuss some implications of Theorem 3.17, focussing on the supercritical case where $\nu = \mathbb{E}[W^2]/\mathbb{E}[W] > 1$. In this case, the parameter $\xi$ is the extinction probability of a branching process with offspring distribution $p_k = \mathbb{P}(\text{Poi}(W^*) = k)$, where $W^*$ is the size-biased version of $W$. Thus,

\[
\xi = G_{\text{Poi}(W^*)}(\xi) = \mathbb{E}[e^{W^*(\xi-1)}], \tag{3.6.6}
\]

where $G_{\text{Poi}(W^*)}(s) = \mathbb{E}[s^{\text{Poi}(W^*)}]$ is the probability generating function of a mixed Poisson random variable with mixing distribution $W^*$.

Further, since $v_k(\mathcal{C}_{\max})/n \xrightarrow{\nu} p_k(1 - \xi^k)$ and $|\mathcal{C}_{\max}|/n \xrightarrow{\nu} \zeta$, it must be that

\[
\zeta = \sum_{k \geq 0} p_k(1 - \xi^k) = 1 - G_{\nu}(\xi), \tag{3.6.7}
\]

where $G_{\nu}(s) = \mathbb{E}[s^\nu]$ is the probability generating function of $D = \text{Poi}(W)$. We also note that $|E(\mathcal{C}_{\max})|/n \xrightarrow{\nu} \eta$ with $\eta = \frac{1}{2}\mathbb{E}[W](1 - \zeta^2)$, so that

\[
\eta = \frac{1}{2} \sum_{k \geq 0} kp_k(1 - \xi^k) = \frac{1}{2}\mathbb{E}[W] \sum_{k \geq 0} \frac{kp_k}{\mathbb{E}[W]}(1 - \xi^k) \tag{3.6.8}
\]

\[
= \frac{1}{2}\mathbb{E}[W](1 - \xi G_{\text{Poi}(W^*)}(\xi)) = \frac{1}{2}\mathbb{E}[W](1 - \zeta^2),
\]

as required.

We now compare the limiting total number of edges to the limiting total size of $\mathcal{C}_{\max}$. We note that since $f(k) = k$ and $g(k) = 1 - \xi^k$ are both increasing,

\[
\sum_{k \geq 0} kp_k(1 - \xi^k) > \sum_{k \geq 0} kp_k \sum_{k \geq 0} (1 - \xi^k) p_k = \mathbb{E}[W]\zeta, \tag{3.6.9}
\]

As a result,

\[
\eta > \frac{1}{2}\mathbb{E}[W]\zeta. \tag{3.6.10}
\]
Thus, the average degree $\eta/\zeta$ in the giant component is strictly larger than the average degree in the entire graph $\mathbb{E}[W]/2$.

We finally show that $\eta > \zeta$, so that the giant has linear complexity. By convexity of $x \mapsto x^{k-1}$ and the fact that $\xi < 1$, for $k \geq 1$,
\begin{equation}
\sum_{i=0}^{k-1} \xi^i \leq k(1 + \xi^{k-1})/2 \tag{3.6.11}
\end{equation}
with strict inequality for $k \geq 3$. Multiply by $1 - \xi$ to obtain
\begin{equation}
1 - \xi^k \leq k(1 - \xi)(1 + \xi^{k-1})/2, \tag{3.6.12}
\end{equation}
again for every $k \geq 1$, again with strict inequality for $k \geq 3$. Now multiply by $p_k$ and sum to get
\begin{equation}
\sum_k p_k (1 - \xi^k) \leq (1 - \xi) \sum_k k p_k (1 + \xi^{k-1})/2. \tag{3.6.13}
\end{equation}
The lhs of (3.6.13) equals $\zeta$ by (3.6.7). We next investigate the rhs of (3.6.13). Recall that
\begin{equation}
\sum_k k p_k = \mathbb{E}[W], \tag{3.6.14}
\end{equation}
and
\begin{equation}
\sum_k k p_k \mathbb{E}[W] \xi^{k-1} = \xi. \tag{3.6.15}
\end{equation}
Hence, the rhs of (3.6.13) is
\begin{equation}
(1 - \xi) (\mathbb{E}[W] + \mathbb{E}[W] \xi)/2 = \mathbb{E}[W] (1 - \xi^2)/2, \tag{3.6.16}
\end{equation}
which is the limit in probability of $|E(C_{\max})|/n$. $\square$

We close this section by discussing the consequences of the phase transition for the attack vulnerability of $\text{CL}_n(w)$:

**Attack vulnerability of $\text{CL}_n(w)$**

Suppose an adversary attacks a network by removing some of its vertices. A *clever* adversary would remove the vertices in a clever way, this is often referred to as a *deliberate* attack. On the other hand, the vertices might also be exposed to random failures, which is often referred to as a *random* attack. The results as stated above do not specifically apply to these settings, but do have intuitive consequences. We model a deliberate attack as the removal of a proportion of the vertices with highest weights, whereas a random attack is modeled by random removal of the vertices with a given probability. One of the aims is to quantify the effect of such attacks, and in particular the difference in random and deliberate attacks. We denote the proportion of removed vertices by $p$. We shall always assume that $\nu > 1$, so that a giant component exists, and we investigate under what conditions on $p$ and the graph $\text{CL}_n(w)$, the giant component remains to exist.
We start by addressing the case of random attack for the CL\textsubscript{$n$}(\textit{w}) model under Condition 1.1(a)-(c), where $E[W^2] < \infty$. One of the difficulties of the above set-up is that we remove vertices rather than edges, so that the resulting graph is no longer an IRG. In percolation jargon, we deal with site percolation rather than with edge percolation. We start by relating the obtained graph to an inhomogeneous random graph.

Note that when we explore a cluster of a vertex after an attack, then the vertex may not have been affected by the attack, which has probability $p$. After this, in the exploration, we always inspect an edge between a vertex which is unaffected by the attack and a vertex of which we do not yet know whether it has been attacked or not. As a result, for random attacks, the probability that it is affected is precisely equal to $p$. Therefore, it is similar to the random graph where $p_{ij}$ is replaced with $(1-p) \cdot p_{ij}$. For a branching process, this identification is exact, and we have that $\zeta_{\kappa,p} = (1-p) \cdot \zeta_{(1-p)\kappa}$, where $\zeta_{\kappa,p}$ denotes the survival probability of the branching process where each individual is killed with probability $p$ independently of all other randomness. For CL\textsubscript{$n$}(\textit{w}), this equality is only asymptotic. In the case where $E[W^2] < \infty$, so that $\nu < \infty$, this means that there exists a critical value $p_c = 1 - 1/\nu$, such that if $p < p_c$, the CL\textsubscript{$n$}(\textit{w}) where vertices are removed with probability $p$, the giant component persists, while if $p > p_c$, then the giant component is destroyed. Thus, when $E[W^2] < \infty$, the CL\textsubscript{$n$}(\textit{w}) is sensitive to random attacks. When $E[W^2] = \infty$, on the other hand, $\nu = \infty$, so that the giant component persists for every $p \in [0,1)$, and the graph is called robust to random attacks. Here we must note that the size of the giant component does decrease, since $\zeta_{\kappa,p} < p \zeta_{\kappa}$.

For a deliberate attack, we remove the proportion $p$ of vertices with highest weight. This means that $\textit{w}$ is replaced with $\textit{w}(p)$, which is equal to $w_i(p) = w_i \mathbbm{1}_{\{i > np\}}$, and we denote the resulting edge probabilities by

$$p_{ij}(p) = \max\{1, w_i(p)w_j(p)/\ell_n\}. \quad (3.6.17)$$

In this case, the resulting graph on $[n] \setminus [np]$ is again a Chung-Lu model, for which $\nu$ is replaced with $\nu(p)$ given by

$$\nu(p) = \frac{E[\psi(U)^2 \mathbbm{1}_{\{U > p\}}]}{E[W]}, \quad (3.6.18)$$

where $U$ is uniform on $[0,1]$ and we recall that we denote $\psi(u) = [1 - F]^{-1}(u)$. Now, for any distribution function $F$, $E[(1 - F)^{-1}(U)^2 \mathbbm{1}_{\{U > p\}}] < \infty$, so that, for $p$ sufficiently close to 1, $\nu(p) < 1$ (see Exercise 3.38). Thus, the CL\textsubscript{$n$}(\textit{w}) model is always sensitive to deliberate attacks.

### 3.7 Related results for inhomogeneous random graphs

In this section, we discuss some related results for inhomogeneous random graphs. While we give intuition about their proofs, we do not include them in full detail.
The largest subcritical cluster

For the classical random graph ER\(_n(\lambda/n)\), it is well known that in the subcritical case for which \(\lambda < 1\), the stronger bound \(|C_{\text{max}}| = \Theta(\log n)\) holds (see [Volume 1, Theorems 4.4–4.5]), and that in the supercritical case for which \(\lambda > 1\), \(|C_{(2)}| = \Theta(\log n)\). These bounds do not always hold in the general framework we are considering here, but if we add some conditions, then we can improve the estimates in Theorem 3.16 for the subcritical case to \(O(\log n)\):

**Theorem 3.18** (Subcritical phase and duality principle of inhomogeneous random graphs) Consider the inhomogeneous random graph IRG\(_n(\kappa_n)\), where \((\kappa_n)\) is a graphical sequence of kernels with limit \(\kappa\). Then,

(i) if \(\kappa\) is subcritical and \(\sup_{x,y,n} \kappa_n(x,y) < \infty\), then \(|C_{\text{max}}| = O_p(\log n)\).

(ii) if \(\kappa\) is supercritical, \(\kappa\) is irreducible, and either \(\inf_{x,y,n} \kappa_n(x,y) > 0\) or \(\sup_{x,y,n} \kappa(x,y) < \infty\), then \(|C_{(2)}| = O_p(\log n)\).

When \(\lim_{n \to \infty} \sup_{x,y} \kappa_n(x,y) = \infty\), the largest subcritical clusters can have rather different behavior, as we now show for the rank-1 case. Note that, by Theorem 3.16 as well as the fact that \(\|T_\kappa\| = \nu = \mathbb{E}[W^2]/\mathbb{E}[W]\), a rank-1 model can only be subcritical when \(\mathbb{E}[W^2] < \infty\), i.e., in the case of finite variance degrees. However, when \(W\) has a power-law tail, i.e., when \(P(W \geq w) \sim w^{-(\tau-1)}\), then the highest weight can be much larger than \(\log n\). When this is the case, then also the largest subcritical cluster is much larger than \(\log n\), as proved in the following theorem:

**Theorem 3.19** (Subcritical phase for rank-1 inhomogeneous random graphs) Let \(w\) satisfy Condition 1.1(a)-(c) with \(\nu = \mathbb{E}[W^2]/\mathbb{E}[W] < 1\), and, further, that there exist \(\tau > 3\) and \(c_2 > 0\) such that

\[
1 - F_n(x) \leq c_2 x^{-(\tau-1)}. \tag{3.7.1}
\]

Then, for \(NR_n(w)\) with \(\Delta = \max_{j \in [n]} w_j\),

\[
|C_{\text{max}}| = \frac{\Delta}{1 - \nu} + o_p(n^{1/(\tau-1)}). \tag{3.7.2}
\]

Theorem 3.19 is most interesting in the case where the limiting distribution function \(F\) in Condition 1.1 has a power-law tail. For example, for \(w\) as in (1.3.15), let \(F\) satisfy

\[
1 - F(x) = cx^{-(\tau-1)}(1 + o(1)). \tag{3.7.3}
\]

Then, \(\Delta = w_1 = [1 - F]^{-1}(1/n) = (cn)^{1/(\tau-1)}(1 + o(1)).\) Therefore,

\[
|C_{\text{max}}| = (cn)^{1/(\tau-1)}/(1 - \nu) + o(n^{1/(\tau-1)}). \tag{3.7.4}
\]

Thus, the largest connected component is much larger than for ER\(_n(\lambda/n)\) with \(\lambda < 1\).

Theorem 3.19 can be intuitively understood as follows. The connected component of a typical vertex is close to a branching process, so that it is with high probability bounded since the expected value of its cluster will be close to
1/(1 − ν). Thus, the best way to obtain a large connected component is to start with a vertex with high weight \( w_i \), and let all of its roughly \( w_i \) children be independent branching processes. Therefore, in expectation, each of these children is connected to another 1/(1 − ν) different vertices, leading to a cluster size of roughly \( w_i/(1 − ν) \). This is clearly largest when \( w_i = \max_{j \in [n]} w_j = \Delta \), leading to an intuitive explanation of Theorem 3.19.

Theorems 3.18 and 3.19 raise the question what the precise conditions for \( |C_{\text{max}}| \) to be of order \( \log n \) are. Intuitively, when \( \Delta \gg \log n \), then \( |C_{\text{max}}| = \Delta/(1 − ν)(1 + o_1(1)) \), whereas if \( \Delta = \Theta(\log n) \), then \( |C_{\text{max}}| = \Theta(\log n) \) as well. In Turova (2011), it was proved that \( |C_{\text{max}}|/\log n \) converges in probability to a finite constant when \( ν < 1 \) and the weights are i.i.d. with distribution function \( F \) with \( \mathbb{E}[e^{αW}] < ∞ \) for some \( α > 0 \), i.e., exponential tails are sufficient.

The critical behavior of rank-1 random graphs

We next discuss the effect of inhomogeneity on the size of the largest connected components in the critical case. As it turns out, the behavior is rather different depending on whether \( \mathbb{E}[W^3] < ∞ \) or not.

**Theorem 3.20** (The critical behavior with finite third moments) Fix the Norros-Reittu random graph with weights \( w(t) = w(1 + tn^{(τ−3)(τ−1)}) \). Assume that \( ν = 1 \), that the weight sequence \( w(1 + tn^{(τ−3)(τ−1)}) \) satisfies Condition 1.1(a)-(c), and further assume that

\[
\mathbb{E}[W_n] = \mathbb{E}[W] + o(n^{-1/3}), \quad \mathbb{E}[W_n^2] = \mathbb{E}[W^2] + o(n^{-1/3}), \quad \mathbb{E}[W_n^3] = \mathbb{E}[W^3] + o(1)
\]

(3.7.5)

Let \( (|C_{\text{i}}(t)|)_{i\geq 1} \) denote the clusters of \( \text{NR}_n(w(t)) \) with \( w(t) = (1 + tn^{-1/3})w \), ordered in size. Then, as \( n \to ∞ \), for all \( t \in \mathbb{R} \),

\[
(n^{-2/3}|C_{\text{i}}(t)|)_{i\geq 1} \overset{d}{\to} (\gamma_i^*(t))_{i\geq 1},
\]

in the product topology, for some limiting random variables \( (\gamma_i^*(t))_{i\geq 1} \).

The limiting random variables \( (\gamma_i^*(t))_{i\geq 1} \) are, apart from a multiplication by a constant and a time-rescaling, equal to those for \( \text{ER}_n(λ/n) \) in the scaling window (see Theorem 5.7).

When \( \mathbb{E}[W^{3−ε}] = ∞ \) for some \( ε > 0 \), it turns out that the scaling of the largest critical cluster is rather different:

**Theorem 3.21** (Weak convergence of the ordered critical clusters for \( τ \in (3, 4) \)) Fix the Norros-Reittu random graph with weights \( w(t) = w(1 + tn^{(τ−3)(τ−1)}) \) defined in (1.3.15). Assume that \( ν = 1 \) and that there exists a \( τ \in (3, 4) \) and \( 0 < c_τ < ∞ \) such that

\[
\lim_{x \to ∞} x^{τ−1}[1 − F(x)] = c_τ.
\]

(3.7.7)

Let \( (|C_{\text{i}}(t)|)_{i\geq 1} \) denote the clusters of \( \text{NR}_n(w(t)) \), ordered in size. Then, as \( n \to ∞ \), for all \( t \in \mathbb{R} \),

\[
(n^{−(τ−2)/(τ−1)}|C_{\text{i}}(t)|)_{i\geq 1} \overset{d}{\to} (\gamma_i(t))_{i\geq 1},
\]

(3.7.8)
Phase transition in general inhomogeneous random graphs

in the product topology, for some non-degenerate limit \((\gamma_i(t))_{i \geq 1}\).

In this chapter, we have already seen that distances depend sensitively on the finiteness of \(E[W^2]\). Now we see that the critical behavior is rather different when \(E[W^3] < \infty\) or \(E[W^3] = \infty\). Interestingly, in the power-law case as described in (3.7.7), the size of the largest clusters grows like \(n^{(\tau-2)/(\tau-1)}\), which is much smaller than the \(n^{2/3}\) scaling when \(E[W^3] < \infty\). The proof of Theorems 3.20 and 3.21 also reveals that the structure of large critical clusters is quite different. When \(E[W^3] < \infty\), then the vertex with largest weight is in the largest connected component with vanishing probability. Therefore, the largest connected component arises by many attempts to create a large cluster, and each trial has roughly the same probability. This can be formulated as power to the masses. In the other hand, for weights \(w\) as in (1.3.15) for which (3.7.7) holds, the vertices with largest weight are with probability bounded away from 0 and 1 in the largest cluster, while a vertex with small weight is in the largest cluster with vanishing probability. Thus, to find the largest clusters, it suffices to explore the clusters of the high-weight vertices: power to the wealthy!

3.8 Notes and discussion

Notes on Section 3.1.

Notes on Section 3.2.

The seminal paper Bollobás et al. (2007) studies inhomogeneous random graph in an even more general setting, where the number of vertices in the graph need not be equal to \(n\). In this case, the vertex space is called a generalized vertex space. We simplify the discussion here by assuming that the number of vertices is always equal to \(n\). An example where the extension to a random number of vertices is crucially used is in Turova and Vallier (2010), which studies an interpolation between percolation and ER\(_n\)(\(p\)).

Notes on Section 3.3.

Theorem 3.4 is a special case of (Bollobás et al., 2007, Theorem 3.13).

Notes on Section 3.4.

See (Athreya and Ney, 1972, Chapter V) or (Harris, 1963, Chapter III) for more background on multi-type branching processes.

Notes on Section 3.5.

Theorem 3.11 is novel to the best of our knowledge, even though Bollobás et al. (2007) prove various relations between inhomogeneous random graphs and branching processes. Proposition 3.13 appears first as (Norros and Reittu, 2006, Proposition 3.1), where the connection between NR\(_n\)(\(w\)) and Poisson branching processes were first exploited to prove versions of Theorem 6.3.
3.9 Exercises for Chapter 3

Notes on Section 3.6.

Theorem 3.17 is taken from Janson and Luczak (2009), where the giant component is investigated for the configuration model. We explain its proof in detail in Section 4.2, where we also prove how the result for the configuration model in Theorem 4.4 can be used to prove Theorem 3.17. Theorem 3.16 is a special case of (Bollobás et al., 2007, Theorem 3.1). Earlier versions for random graphs with given expected degrees or Chung-Lu model appeared in Chung and Lu (2002b, 2006b) (see also the monograph Chung and Lu (2006a)). I have learned the proof of the linear complexity of the giant in Theorem 3.17 from Svante Janson.

Bollobás et al. (2007) prove various other results concerning the giant component of $\text{IRG}_n(\kappa)$. For example, (Bollobás et al., 2007, Theorem 3.9) proves that the giant component of $\text{IRG}_n(\kappa)$ is stable in the sense that its size does not change much if we add or delete a few edges. Note that the edges added or deleted do not have to be random or independent of the existing graph, rather, they can be chosen by an adversary after inspecting the whole of $\text{IRG}_n(\kappa)$. More precisely, (Bollobás et al., 2007, Theorem 3.9) shows that, for small enough $\delta > 0$, the giant component of $\text{IRG}_n(\kappa)$ in the supercritical regime does change by more than $\varepsilon n$ vertices if we remove any collection of $\delta n$ edges.

Notes on Section 3.7.

Theorem 3.19 is (Janson, 2008, Corollary 4.4). Theorem 3.20 is proved in Bhamidi et al. (2010b), a related version with a different proof can be found in Turova (2013). Theorem 3.21 is proved in Bhamidi et al. (2012).

3.9 Exercises for Chapter 3

Exercise 3.1 (Erdős-Rényi random graph) Show that when $S = [0, 1]$ and $\rho_{ij} = \kappa(i/n, j/n)/n$ with $\kappa: [0, 1]^2 \rightarrow [0, \infty)$ being continuous, then the model is the Erdős-Rényi random graph with edge probability $\lambda/n$ precisely when $\kappa(x, y) = \lambda$. Is this also true when $\kappa: [0, 1]^2 \rightarrow [0, \infty)$ is not continuous?

Exercise 3.2 (Lower bound on expected number of edges) Show that when $\kappa: S \times S$ is continuous, then

$$\liminf_{n \to \infty} \frac{1}{n} \mathbb{E}[\text{E}(\text{IRG}_n(\kappa))] \geq \frac{1}{2} \iint_{S^2} \kappa(x, y)\mu(dx)\mu(dy),$$

so that the lower bound in (3.2.4) generally holds.

Exercise 3.3 (Expected number of edges) Show that when $\kappa: S \times S$ is bounded and continuous, then (3.2.4) holds.

Exercise 3.4 (The Chung-Lu model) Prove that when $\kappa$ is given by

$$\kappa(x, y) = [1 - F]^{-1}(x)[1 - F]^{-1}(y)/\mathbb{E}[W],$$

then $\kappa$ is graphical precisely when $\mathbb{E}[W] < \infty$, where $W$ has distribution function $F$. Further, $\kappa$ is always irreducible.
Exercise 3.5 (The Chung-Lu model repeated) Let \( \tilde{w}_i = [1 - F]^{-1}(i/n) \sqrt{\frac{nE(W)}{\ell_n}} \) and \( w_i = [1 - F]^{-1}(i/n) \) as in [Volume 1, (6.2.14)]. Then \( CL_n(\tilde{w}) \) and \( CL_n(w) \) are asymptotically equivalent whenever \( (\frac{E(W)}{\ell_n} - 1)^2 = o(n) \).

Exercise 3.6 (Asymptotic equivalence for general IRGs) Prove that the random graphs \( IRG_n(p) \) with \( p_{ij} \) as in (3.2.7) is asymptotically equivalent to \( IRG_n(p) \) with \( p_{ij} = p_{ij}^{GRG}(\kappa_n) \) and to \( IRG_n(p) \) with \( p_{ij} = p_{ij}^{GRG}(\kappa_n) \) when (3.2.10) holds.

Exercise 3.7 (Definitions 3.2-3.3 for homogeneous bipartite graph) Prove that Definitions 3.2-3.3 hold for the homogeneous bipartite graph.

Exercise 3.8 (Examples of homogeneous random graphs) Show that the Erdős-Rényi random graph, the homogeneous bipartite random graph and the stochastic blockmodel are all homogeneous random graphs.

Exercise 3.9 (Homogeneous bipartite graph) Prove that the homogeneous bipartite random graph is a special case of the finite-types case.

Exercise 3.10 (Irreducibility for the finite-types case) Prove that, in the finite-type case, irreducibility follows when there exists an \( m \) such that the \( m \)th power of the matrix \( (\kappa(i,j)\mu(j))_{i,j \in [r]} \) contains no zeros.

Exercise 3.11 (Graphical limit in the finite-types case) Prove that, in the finite-type case, (3.2.1) holds precisely when
\[
\lim_{n \to \infty} \frac{n_i}{n} = \mu_i.
\]

Exercise 3.12 (Variance of number of vertices of type \( i \) and degree \( k \)) Let \( IRG_n(\kappa_n) \) be a finite-type inhomogeneous random graph with graphical sequence of kernels \( \kappa_n \). Let \( N_{i,k} \) be equal to the number of vertices of type \( i \) and degree \( k \). Show that \( \text{Var}(N_{i,k}) = O(n) \).

Exercise 3.13 (Proportion of isolated vertices in inhomogeneous random graphs) Let \( IRG_n(\kappa_n) \) be an inhomogeneous random graph with graphical sequence of kernels \( \kappa_n \) that converge to \( \kappa \). Show that the proportion of isolated vertices converges to
\[
\frac{1}{n} N_0(n) \xrightarrow{p} p_0 = \int e^{-\lambda(x)} \mu(dx).
\]
Conclude that \( p_0 > 0 \) when \( \int \lambda(x) \mu(dx) < \infty \).

Exercise 3.14 (Upper and lower bounding finite-type kernels) Prove that the kernels \( \kappa^-_n \) and \( \kappa^+_n \) in (3.3.13) and (3.3.14) are of finite type.

Exercise 3.15 (Inclusion of graphs for larger \( \kappa \)) Let \( \kappa' \leq \kappa \) hold a.e. Show that we can couple \( IRG_n(\kappa) \) and \( IRG_n(\kappa) \) such that \( IRG_n(\kappa) \subseteq IRG_n(\kappa) \).

Exercise 3.16 (Tails of Poisson variables) Use stochastic domination of Poisson random variables with different parameters, as well as concentration properties of Poisson variables, to complete the proof of (3.3.27).
3.9 Exercises for Chapter 3

Exercise 3.17 (Power-laws for sum kernels) Let $\kappa(x, y) = \alpha(x) + \alpha(y)$ for a continuous function $\alpha: [0, 1] \rightarrow [0, \infty)$. Use Corollary 3.7 to identify when the degree distribution satisfies a power law. How is the tail behavior of $D$ related to that of $\alpha$?

Exercise 3.18 (Irreducibility of multi-type branching process) Show that the positivity of the survival probability $\zeta^{(i)}$ of an individual of type $i$ is independent of the type $i$ when the probability that an individual of type $j$ to have a type $i$ descendent is strictly positive.

Exercise 3.19 (Irreducibility of multi-type branching process (Cont.)) Prove that the probability that an individual of type $j$ to have a type $i$ descendent is strictly positive precisely when there exists an $l$ such that $T^l_{\kappa}(i, j) > 0$, where $T_{\kappa}(i, j) = \kappa_{ij} \mu_j$ is the mean offspring matrix.

Exercise 3.20 (Singularity of multi-type branching process) Prove that $G(s) = M_\kappa s$ for some matrix $M$ precisely when each individual in the Markov chain has exactly one offspring.

Exercise 3.21 (Erdős-Rényi random graph) Prove that $\mathcal{N}_n(w) = \mathcal{E}_n(\lambda/n)$ when $w$ is constant with $w_i = -n \log (1 - \lambda/n)$.

Exercise 3.22 (Homogeneous Poisson multi-type branching processes) Consider a homogeneous Poisson multi-type branching process with parameter $\lambda$. Show that the function $\phi(x) = 1$ is an eigenvector of $T_{\kappa}$ with eigenvalue $\lambda$. Conclude that $(Z_j)_{j \geq 0}$ is a martingale, where $(Z_j)_{j \geq 0}$ denotes the number of individuals in the $j$th generation.

Exercise 3.23 (Bounds on distances in the case where $\|T_{\kappa}\| > 1$) Fix $\varepsilon > 0$. Use (3.5.19) to show that $\text{dist}_{\mathcal{I} \mathcal{R}_n(\kappa_n)}(o_1, o_2) \geq (1 - \varepsilon) \log \|T_{\kappa}\| n)$ whp when $\|T_{\kappa}\| > 1$.

Exercise 3.24 (No large component when $\|T_{\kappa}\| < 1$) Use (3.5.19) to show that $\text{dist}_{\mathcal{I} \mathcal{R}_n(\kappa_n)}(o_1, o_2) = \infty$ whp when $\|T_{\kappa}\| < 1$.

Exercise 3.25 (Proof of no-overlap property in (3.5.22)) Prove that $P(B_{\kappa}^{(\alpha)}(o_1) = (t, q), o_2 \in B_{\kappa}^{(\alpha)}(o_1)) \rightarrow 0$, and conclude that (3.5.22) holds.

Exercise 3.26 (Branching process domination of Erdős-Rényi random graph) Show that Exercise 3.21 together with Proposition 3.13 imply that $|\mathcal{E}(o)| \leq T^*$, where $T^*$ is the total progeny of a Poisson branching process with mean $-n \log (1 - \lambda/n)$ offspring.

Exercise 3.27 (Local weak convergence of $\mathcal{E}_n(\lambda/n)$) Use Theorem 3.15 to show that also $\mathcal{E}_n(\lambda/n)$ converges locally weakly in probability to the Galton-Watson tree with Poisson offspring distribution with parameter $\lambda$.

Exercise 3.28 (Coupling to a multi-type Poisson branching process) Prove Proposition 3.14 by adapting the proof of Proposition 3.13.
Exercise 3.29 (Phase transition for $r = 2$) Let $\zeta_\kappa$ denote the survival probability of a two-types multi-type branching process. Compute $\zeta_\kappa$ and give necessary and sufficient conditions for $\zeta_\kappa > 0$ to hold.

Exercise 3.30 (The size of small components in the finite-types case) Prove that, in the finite-types case, when $(\kappa_n)$ converges, then $\sup_{x,y,n} \kappa_n(x,y) < \infty$ holds, so that the results of Theorem 3.18 apply in the sub- and supercritical cases.

Exercise 3.31 (Law of large numbers for $C_{\max}$ for $\text{ER}_n(\lambda/n)$) Prove that, for the Erdős-Rényi random graph, Theorem 3.16 implies that $|C_{\max}|/n \xrightarrow{p} \zeta_\lambda$, where $\zeta_\lambda$ is the survival probability of a Poisson branching process with mean $\lambda$ offspring.

Exercise 3.32 (Connectivity of uniformly chosen vertices) Suppose we draw two vertices uniformly at random from $[n]$ in $\text{IRG}_n(\kappa_n)$. Prove that Theorem 3.17 implies that the probability that the vertices are connected converges to $\zeta^2$.

Exercise 3.33 (The size of small components for $\text{CL}_n(w)$) Use Theorem 3.18 to prove that, for $\text{CL}_n(w)$ with weights given by (1.3.15) and with $1 < \nu < \infty$, the second largest cluster has size $|C_{(2)}| = O(\log n)$ when $W$ has bounded support or is a.s. bounded below by $\varepsilon > 0$, while if $\nu < 1$, $|C_{\max}| = O(\log n)$ when $W$ has bounded support. Here $W$ is a random variable with distribution function $F$.

Exercise 3.34 (Average degree in two populations) Show that the average degree is close to $pm_1 + (1-p)m_2$ in the setting of Example 3.1.

Exercise 3.35 (The phase transition for two populations) Show that the $\zeta > 0$ precisely when $[pm_1^2 + (1-p)m_2^2]/[pm_1 + (1-p)m_2] > 1$ in the setting of Example 3.1. Find an example of $p, m_1, m_2$ where the average degree is less than one, yet there exists a giant component.

Exercise 3.36 (Degree sequence of giant component) Consider $\text{GRG}_n(w)$ as in Theorem 3.17. Show that the proportion of vertices of the giant component $C_{\max}$ having degree $k$ is close to $p_k(1-\xi^k)/\zeta$.

Exercise 3.37 (Degree sequence of complement of giant component) Consider $\text{GRG}_n(w)$ as in Theorem 3.17. Show that when $\xi < 1$, the proportion of vertices outside the giant component $C_{\max}$ having degree $k$ is close to $p_k\xi^k/(1-\zeta)$. Conclude that the degree sequence of the complement of the giant component never satisfies a power law. Can you give an intuitive explanation for this?

Exercise 3.38 (Finiteness of $\nu(p)$) Prove that $\nu(p)$ in (3.6.18) satisfies that $\nu(p) < \infty$ for every $p \in (0, 1]$ and any distribution function $F$. 
Chapter 4

The Phase Transition in the Configuration Model

Abstract

In this chapter, we investigate the local weak limit of the configuration model. After this, we give a detailed proof of the phase transition in the configuration model by investigating its largest connected component. We identify when there is a giant component and find its size and degree structure. Further results include the connectivity transition of the configuration model as well as the critical behavior in the configuration model.

In this chapter, we study the connectivity structure of the configuration model. We focus on the local connectivity by investigating the local weak limit of the configuration model, as well as the global connectivity by identifying its giant component as well as when it is completely connected. In inhomogeneous random graphs, there always is a positive proportion of vertices that are isolated (recall Exercise 3.13). In many real-world examples, we observe the presence of a giant component (recall Table 3.1). However, in many of these examples, the giant is almost the whole graph, and sometimes it, by definition is the whole graph. For example, Internet needs to be connected to allow e-mail messages to be sent between any pair of vertices. In many other real-world examples, though, it is not at all obvious why the network needs to be connected.

<table>
<thead>
<tr>
<th>Subject</th>
<th>% in Giant</th>
<th>Size</th>
<th>Source</th>
<th>Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>California roads</td>
<td>0.9958</td>
<td>1,965,206</td>
<td>Leskovec et al. (2009)</td>
<td>Leskovec and Krevl (2014)</td>
</tr>
<tr>
<td>Facebook</td>
<td>0.9991</td>
<td>721m</td>
<td>Ugander et al. (2011)</td>
<td>Ugander et al. (2011)</td>
</tr>
<tr>
<td>Hyves</td>
<td>0.996</td>
<td>8,047,530</td>
<td>Corten (2012)</td>
<td>Corten (2012)</td>
</tr>
</tbody>
</table>

Table 4.1 The rows in the above table represent the following real-world networks:

The California road network consists of nodes representing intersections or endpoints of roads. In the Facebook network nodes represent the users and the edges Facebook friendships. Hyves was a Dutch social media platform. Nodes represent the users, and an edge between nodes represents a friendship. The arXiv astro-physics network represents authors of papers within the astro-physics section of arXiv. There is an edge between authors if they have ever been co-authors. Models the high voltage power network in western USA. The nodes denote transformers substations, and generators and the edges represent transmission cables. The data set consists of jazz-musicians and there exists a connected if they ever collaborated.
Table 4.1 invites us to think about what makes networks close to fully connected. We investigate this question here in the context of the configuration model. The advantage of the configuration model is that it is highly flexible in its degree structure, making it possible for all degrees to be at least a certain value. We will see that this allows the random graph to be connected, while at the same time remaining sparse, as is the case in many real-world networks.

Organization of this chapter

This chapter is organized as follows. In Section 4.1, we study the local weak limit of the configuration model. In Section 4.2, we state and prove the law of large numbers for the giant component in the configuration model, thus establishing the phase transition. In Section 4.3, we study when the configuration model is connected. In Section 4.4, we state further results in the configuration model. We close this chapter in Section 4.5 with notes and discussion and with exercises in Section 4.6.

4.1 Local weak convergence to unimodular trees for $\text{CM}_n(d)$

We start by investigating the locally tree-like nature of the configuration model. Our main result is as follows:

**Theorem 4.1** (Locally tree-like nature configuration model) Assume that Conditions 1.5(a)-(b) hold. Then $\text{CM}_n(d)$ converges locally weakly in probability to the unimodular Galton-Watson tree with root offspring distribution $(p_k)_{k\geq 0}$ given by $p_k = \mathbb{P}(D = k)$.

Before starting with the proof of Theorem 4.1, let us explain the above connection between local neighborhoods and branching processes. We note that the asymptotic offspring distribution at the root is equal to $(p_k)_{k\geq 0}$, where $p_k = \mathbb{P}(D = k)$ is the asymptotic degree distribution, since the probability that a random vertex has degree $k$ is equal to $p_k^{(n)} = \mathbb{P}(D_n = k) = n_k/n$, where $n_k$ denotes the number of vertices with degree $k$, which, by Condition 1.5(a), converges to $p_k = \mathbb{P}(D = k)$, for every $k \geq 1$. This explains the offspring of the root of our branching process approximation.

The offspring distribution of the individuals in the first and later generations is given by

$$p_k^* = \frac{(k+1)p_{k+1}}{\mathbb{E}[D]}.$$ (4.1.1)

We now heuristically explain this relation to branching processes by intuitively describing the exploration of a vertex chosen uniformly from the vertex set $[n]$. To describe the offspring of the direct neighbors of the root, we examine the degree of the vertex to which the first half-edge incident to the root is paired. By the uniform matching of half-edges, the probability that a vertex of degree $k$ is chosen is proportional to $k$. Ignoring the fact that the root and one half-edge
4.1 Local weak convergence to unimodular trees for $\text{CM}_n(d)$

have already been chosen (which does have an effect on the number of available or free half-edges), the degree of the vertex incident to the chosen half-edge equals $k$ with probability equal to $kp_k(n)/E[D_n]$, where again $p_k(n) = n_k/n$ denotes the proportion of vertices with degree $k$, and

\[ E[D_n] = \frac{1}{n} \sum_{i \in [n]} d_i = \frac{1}{n} \sum_{i \in [n]} \sum_{k=0}^\infty k \mathbb{1}_{\{d_i = k\}} = \sum_{k=0}^\infty kp_k(n) \quad (4.1.2) \]

is the average degree in $\text{CM}_n(d)$. Thus, $(kp_k(n)/E[D_n])_{k \geq 0}$ is a probability mass function. However, one of the half-edges is used up to connect to the root, so that, for a vertex incident to the root to have $k$ offspring, it needs to connect its half-edge to a vertex having degree $k + 1$. Therefore, the probability that the offspring of any of the direct neighbors of the root is equal to $k$ equals

\[ p_k^{(n)} = \frac{(k + 1)p_{k+1}^{(n)}}{E[D_n]} \quad (4.1.3) \]

Thus, $(p_k^{(n)})_{k \geq 0}$ can be interpreted as the forward degree of vertices in the cluster exploration. When Condition 1.5(a)-(b) hold, also $p_k^{(n)} \to p_k^*$, where $(p_k^*)_{k \geq 0}$ is defined in (4.1.1). As a result, we often refer to $(p_k^*)_{k \geq 0}$ as the asymptotic forward degree distribution.

The above heuristically argues that the number of vertices unequal to the root connected to any direct neighbor of the root has asymptotic law $(p_k^*)_{k \geq 0}$. However, every time we pair two half-edges, the number of free or available half-edges decreases by 2. Similarly to the depletion-of-points effect in the exploration of clusters for the Erdős-Rényi random graph $\text{ER}_n(\lambda/n)$, the configuration model $\text{CM}_n(d)$ suffers from a depletion-of-points-and-half-edges effect. Thus, by iteratively connecting half-edges in a breadth-first way, the offspring distribution changes along the way, which gives potential trouble. Luckily, the number of available half-edges that we start with equals $\ell_n - 1$, which is very large when Condition 1.5(a)-(b) hold, since then $\ell_n/n = E[D_n]/n \to E[D] > 0$. Thus, we can pair many half-edges before we start noticing that their number decreases. As a result, the degrees of different vertices in the exploration process is close to being i.i.d., leading to a branching process approximation. In order to prove Theorem 4.1, we will only need to pair a bounded number of edges.

In order to get started for the proof of (2.3.10) for Theorem 4.1, we introduce some notation. Fix a rooted tree $t$ with $k$ generations, and let

\[ N_n(t) = \sum_{v \in [n]} \mathbb{1}_{\{B_k^{(n)}(v) = t\}} \quad (4.1.4) \]

denote the number of vertices whose local neighborhood up to generation $t$ equals $t$. By Theorem 2.13, in order to prove Theorem 4.1, we need to show that

\[ \frac{N_n(t)}{n} \overset{p}{\to} \mathbb{P}(\text{BP} \leq k \simeq t). \quad (4.1.5) \]

Here, we also rely on Theorem 2.8 to see that it suffices to prove (4.1.5) for
trees, since the unimodular Galton-Watson tree is a tree with probability 1. For this, we will use a second moment method. We first prove that the first moment $\mathbb{E}[N_n(t)]/n \to \mathbb{P}(t(p,k) \simeq t)$, after which we prove that $\text{Var}(N_n(t)) = o(n^2)$. Then, by the Chebychev inequality [Volume 1, Theorem 2.18], (4.1.5) follows.

Local weak convergence for configuration model: first moment

We next relate the neighborhood in a random graph to a branching process where the root has offspring distribution $D_n$, while all other individuals have offspring distribution $D_n^* - 1$, where

$$
P(D_n^* = k) = \frac{k}{E[D_n]} P(D_n = k), \quad k \in \mathbb{N},
$$

is the size-biased distribution of $D_n$. Denote this branching process by $(\text{BP}_n(t))_{t \in \mathbb{N}_0}$. Here, $\text{BP}_n(t)$ denotes the branching process when it contains precisely $t$ vertices, and we explore it in the breadth-first order. Clearly, by Conditions 7.8(a)-(b), $\text{BP}_n(t) \overset{d}{\rightarrow} \text{BP}(t)$ for every $t$ finite, where $\text{BP}(t)$ is the restriction of the unimodular branching process $\text{BP}$ with offspring distribution $(p_k)_{k \geq 1}$ for which $p_k = P(D = k)$ to its first $t$ individuals (see Exercise 4.1). For future reference, denote the offspring distribution of the above branching process by

$$\quad p^*_k = P(D^* - 1 = k)
$$

Note that, when $t$ is a fixed rooted tree of $k$ generations, then $\text{BP}_{\leq k} \simeq t$ precisely when $\text{BP}(t_k) \simeq t$, where $t_k$ denotes the number of vertices in $t$.

We let $(G_n(t))_{t \in \mathbb{N}_0}$ denote the graph exploration process from a uniformly chosen vertex $o \in [n]$. Here $G_n(t)$ is the exploration up to $t$ vertices, in the breadth-first manner. In particular, from $(G_n(t))_{t \in \mathbb{N}_0}$ we can retrieve $(B_n(t))_{t \in \mathbb{N}_0}$ for every $t \geq 0$. The following lemma proves that we can couple the graph exploration to the branching process in such a way that $(G_n(t))_{0 \leq t \leq m_n}$ is equal to $(\text{BP}_n(t))_{0 \leq t \leq m_n}$ whenever $m_n \to \infty$ arbitrarily slowly. In the statement, we write $(\hat{G}_n(t), \hat{\text{BP}}_n(t))_{t \in \mathbb{N}_0}$ for the coupling of $(G_n(t))_{0 \leq t \leq m_n}$ and $(\text{BP}_n(t))_{0 \leq t \leq m_n}$:

**Lemma 4.2** (Coupling graph exploration and branching process) Subject to Conditions 1.5(a)-(b), there exists a coupling $(\hat{G}_n(t), \hat{\text{BP}}_n(t))_{t \in \mathbb{N}_0}$ of $(G_n(t))_{0 \leq t \leq m_n}$ and $(\text{BP}_n(t))_{0 \leq t \leq m_n}$ such that

$$\mathbb{P}\left((\hat{G}_n(t))_{0 \leq t \leq m_n} \neq (\hat{\text{BP}}_n(t))_{0 \leq t \leq m_n}\right) = o(1),$$

whenever $m_n \to \infty$ arbitrarily slowly. Consequently, $\mathbb{E}[N_n(t)]/n \to \mathbb{P}(\text{BP} \leq k \simeq t)$.

In the proof, we will see that any $m_n = o(\sqrt{n/d_{\max}})$ is allowed. Here $d_{\max} = \max_{i \in [n]} d_i$ is the maximal vertex degree in the graph, which is $o(n)$ when Conditions 7.8(a)-(b) hold.
4.1 Local weak convergence to unimodular trees for $CM_n(d)$

**Proof**  We let the offspring of the root of the branching process $\hat{D}_n$ be equal to $d_o$, which is the number of neighbors of the vertex $o \in [n]$ that is chosen uniformly at random. By construction, $\hat{D}_n = d_o$, so that also $\hat{G}_n(1) = \mathbb{B}\hat{P}_n(1)$. We next explain how to *jointly* construct $(\hat{G}_n(t), \mathbb{B}\hat{P}_n(t))_{0 \leq t \leq m}$ given that we have already constructed $(\hat{G}_n(t), \mathbb{B}\hat{P}_n(t))_{0 \leq t \leq m-1}$.

To obtain $(\hat{G}_n(t))_{0 \leq t \leq m}$, we take the first unpaired half-edge $x_m$. This half-edge needs to be paired to a uniform half-edge that has not been paired so far. We draw a uniform half-edge $y_m$ from the collection of all half-edges, independently of the past, and we let the $(m-1)^{st}$ individual in $(\mathbb{B}\hat{P}_n(t))_{0 \leq t \leq m-1}$ have precisely $d_{U_m} - 1$ children. Note that $d_{U_m} - 1$ has the same distribution as $D_n^* - 1$ and, by construction, the collection $(d_{U_m} - 1)_{t \geq 0}$ is i.i.d. When $y_m$ is still free, i.e., has not yet been paired in $(\hat{G}_n(t))_{0 \leq t \leq m-1}$, then we also let $x_m$ be paired to $y_m$, and we have constructed $(\hat{G}_n(t))_{0 \leq t \leq m}$. However, a problem arises when $y_m$ has already been in $(\hat{G}_n(t))_{0 \leq t \leq m-1}$, in which case we draw a uniform *unpaired* half-edge $y_m'$ and pair $x_m$ to $y_m$ instead. Clearly, this might give rise to a difference between $(\hat{G}_n(t))_{t \leq m}$ and $(\mathbb{B}\hat{P}_n(t))_{0 \leq t \leq m}$. We now provide bounds on the probability that an error occurs before time $m_n$.

There are two sources of differences between $(\hat{G}_n(t))_{t \geq 0}$ and $(\mathbb{B}\hat{P}_n(t))_{t \geq 0}$:

**Half-edge re-use** In the above coupling $y_m$ had already been paired and is being re-used in the branching process, and we need to redraw $y_m'$;

**Vertex re-use** In the above coupling, this means that $y_m$ is a half-edge that has not yet been paired in $(\hat{G}_n(t))_{0 \leq t \leq m-1}$, but it is incident to a half-edge that has already been paired in $(\hat{G}_n(t))_{0 \leq t \leq m-1}$. In particular, the vertex to which it is incident has already appeared in $(\hat{G}_n(t))_{0 \leq t \leq m-1}$ and it is being re-used in the branching process. In this case, a *copy* of the vertex appears in $(\mathbb{B}\hat{P}_n(t))_{0 \leq t \leq m}$, while a *cycle* appears in $(\hat{G}_n(t))_{0 \leq t \leq m}$.

We now provide a bound on both contributions:

**Half-edge re-use**

Up to time $m-1$, at most $2m-1$ half-edges are forbidden to be used by $(\hat{G}_n(t))_{t \leq m})$. The probability that the half-edge $q_m$ equals one of these two half-edges is at most

$$\frac{2m - 1}{\ell_n}.$$  \hspace{1cm} (4.1.9)

Hence the probability that a half-edge is being re-used before time $m_n$ is at most

$$\sum_{m=1}^{m_n} \frac{2m - 1}{\ell_n} \leq \frac{m_n^2}{\ell_n} = o(1),$$  \hspace{1cm} (4.1.10)

when $m_n = o(\sqrt{n})$.  


Vertex re-use

The probability that vertex \( i \) is chosen in the \( m \)th draw is equal to \( d_i / \ell_n \). The probability that vertex \( i \) is drawn twice before time \( m_n \) is at most

\[
\frac{m_n(m_n-1)d_i^2}{2\ell_n^2}.
\]  
(4.1.11)

By the union bound, the probability that there exists a vertex that is chosen twice up to time \( m_n \) is at most

\[
\frac{m_n(m_n-1)}{2\ell_n} \sum_{i \in [n]} d_i^2 \leq m_n d_{\text{max}} / \ell_n = o(1),
\]  
(4.1.12)

by Condition 7.8 when \( m_n = o(\sqrt{n/d_{\text{max}}}) \).

Completion of the proof

In order to show that \( \mathbb{E}[N_n(t)/n] \rightarrow \mathbb{P}(\text{BP} \leq k \simeq t) \), we let \( t_k \) denote the number of individuals in the first \( k-1 \) generations in \( t \), and let \( (t(t))_{t \in [t_k]} \) be its breadth-first exploration. Then,

\[
\mathbb{E}[N_n(t)/n] = \mathbb{P}((G_n(t))_{t \in [t_k]} = (t(t))_{t \in [t_k]}),
\]  
(4.1.13)

so that

\[
\mathbb{P}((G_n(t))_{t \in [t_k]} = (t(t))_{t \in [t_k]}) = \mathbb{P}((\text{BP}_n(t))_{t \in [t_k]} = (t(t))_{t \in [t_k]}) + o(1)
\]  
(4.1.14)

\[
= \mathbb{P}((\text{BP}(t))_{t \in [t_k]} = (t(t))_{t \in [t_k]}) + o(1)
\]

\[
= \mathbb{P}(\text{BP} \leq k \simeq t) + o(1),
\]

where the first equality is (4.1.8), while the second is the statement that \( \text{BP}_n(t) \xrightarrow{d} \text{BP}(t) \) for every \( t \) finite from Exercise 4.1. This proves the claim.

Local weak convergence for the configuration model: second moment

Here, we study the second moment of \( N_n(t) \), and show that it is close to the first moment squared:

**Lemma 4.3** (Concentration of the number of trees) Assume that Conditions 1.5(a)-(b) hold. Then,

\[
\frac{\mathbb{E}[N_n(t)^2]}{n^2} \rightarrow \mathbb{P}(\text{BP} \leq k \simeq t)^2.
\]  
(4.1.15)

Consequently, \( N_n(t)/n \xrightarrow{p} \mathbb{P}(\text{BP} \leq k \simeq t) \).

**Proof** We start by computing

\[
\frac{\mathbb{E}[N_n(t)^2]}{n^2} = \mathbb{P}(B_k^{(\alpha)}(o_1), B_k^{(\alpha)}(o_2) \simeq t),
\]  
(4.1.16)
4.2 Phase transition in the configuration model

where \( o_1, o_2 \in [n] \) are two vertices chosen uniformly at random from \([n]\), independently. Since \(|B_k^{(n)}(o_1)| \xrightarrow{d} |BP_{\leq k}| \), which is a tight random variable, \( o_2 \notin B_{2k}^{(n)}(o_1) \) with high probability, so that also

\[
\frac{\mathbb{E}[N_n(t)^2]}{n^2} = \mathbb{P}(B_k^{(n)}(o_1), B_k^{(n)}(o_2) \simeq t, o_2 \notin B_{2k}^{(n)}(o_1)) + o(1). \tag{4.1.17}
\]

We now condition on \( B_k^{(n)}(o_1) = t \), and write

\[
\mathbb{P}(B_k^{(n)}(o_1), B_k^{(n)}(o_2) \simeq t, o_2 \notin B_{2k}^{(n)}(o_1)) = \mathbb{P}(B_k^{(n)}(o_2) \simeq t \mid B_k^{(n)}(o_1) \simeq t, o_2 \notin B_{2k}^{(n)}(o_1)) \times \mathbb{P}(B_k^{(n)}(o_1) \simeq t, o_2 \notin B_{2k}^{(n)}(o_1)). \tag{4.1.18}
\]

We already know that \( \mathbb{P}(B_k^{(n)}(o_1) \simeq t) \to \mathbb{P}(BP_{\leq k} \simeq t) \), so that also

\[
\mathbb{P}(B_k^{(n)}(o_1) \simeq t, o_2 \notin B_{2k}^{(n)}(o_1)) \to \mathbb{P}(BP_{\leq k} \simeq t). \tag{4.1.19}
\]

In Exercise 4.2, you prove that indeed (4.1.19) holds.

We note that, conditionally on \( B_k^{(n)}(o_1) \simeq t \), \( o_2 \notin B_k^{(n)}(o_1) \), the probability that \( B_k^{(n)}(o_2) \simeq t \) is the same as the probability that \( B_k^{(n)}(o_2) \simeq t \) in \( \text{CM}_n^{(d')} \) which is obtained by removing all vertices in \( B_k^{(n)}(o_1) \). Thus, since \( B_k^{(n)}(o_1) \simeq t \), we have that \( n' = n - |V(T)| \) and \( d' \) is the corresponding degree sequence. The whole point is that the degree distribution \( d' \) still satisfies Conditions 7.8(a)-(b). Therefore, also

\[
\mathbb{P}(B_k^{(n)}(o_2) \simeq t \mid B_k^{(n)}(o_1) \simeq t, o_2 \notin B_{2k}^{(n)}(o_1)) \to \mathbb{P}(BP_{\leq k} \simeq t), \tag{4.1.20}
\]

and we have proved (4.1.15). Since \( \mathbb{E}[N_n(t)]/n \to \mathbb{P}(BP_{\leq k} \simeq t) \) and \( \mathbb{E}[N_n(t)^2]/n^2 \to \mathbb{P}(BP_{\leq k} \simeq t)^2 \), it follows that \( \text{Var}(N_n(t))/n \to 0 \), so that \( N_n(t)/n \) is concentrated. Since also \( \mathbb{E}[N_n(t)]/n \to \mathbb{P}(BP_{\leq k} \simeq t) \), we obtain that \( N_n(t)/n \xrightarrow{p} \mathbb{P}(BP_{\leq k} \simeq t) \), as required.

Lemma 4.3 completes the proof of Theorem 4.1.

4.2 Phase transition in the configuration model

In this section, we investigate the connected components in the configuration model. Alike for the Erdős-Rényi random graph, we identify when the configuration model whp has a giant component. Again, this condition has the interpretation that an underlying branching process describing the exploration of a cluster has a strictly positive survival probability.

We start by recalling some notation from [Volume 1, Chapter 7]. We investigate the configuration model \( \text{CM}_n(d) \), where in most cases, the degrees \( d = (d_i)_{i \in [n]} \) are assumed to satisfy Condition 1.5(a)-(b), and sometimes also Condition 1.5(c). We recall that \( D_n \) is the degree of a uniformly chosen vertex in \([n]\), i.e., \( D_n = d_o \),
where \( o \) is uniformly chosen from \([n]\). Equivalently,

\[
P(D_n = k) = \frac{n_k}{n},
\]

(4.2.1)

where \( n_k \) denotes the number of vertices of degree \( k \). For a graph \( G \), we write \( v_k(G) \) for the number of vertices of degree \( k \) in \( G \), and \( |E(G)| \) for the number of edges. The main result concerning the size and structure of the largest connected components of \( \text{CM}_n(d) \) is the following:

**Theorem 4.4** (Phase transition in \( \text{CM}_n(d) \)) Suppose that Condition 1.5(a)-(b) hold and consider the random graph \( \text{CM}_n(d) \), letting \( n \to \infty \). Assume that \( p_2 = P(D = 2) < 1 \). Let \( C_{\text{max}} \) and \( C_{(2)} \) be the largest and second largest components of \( \text{CM}_n(d) \) (breaking ties arbitrarily).

(a) If \( \nu = \mathbb{E}[D(D - 1)]/\mathbb{E}[D] > 1 \), then there exist \( \xi \in [0, 1), \zeta \in (0, 1] \) such that

\[
\frac{|C_{\text{max}}|}{n} \xrightarrow{p} \zeta,
\]

\[
\frac{v_k(C_{\text{max}})}{n} \xrightarrow{p} p_k(1 - \xi^k) \quad \text{for every } k \geq 0,
\]

\[
\frac{|E(C_{\text{max}})|}{n} \xrightarrow{p} \frac{1}{2} \mathbb{E}[D](1 - \xi^2).
\]

while \( \frac{|C_{(2)}|}{n} \xrightarrow{p} 0 \) and \( \frac{|E(C_{(2)})|}{n} \xrightarrow{p} 0 \).

(b) If \( \nu = \mathbb{E}[D(D - 1)]/\mathbb{E}[D] \leq 1 \), then \( |C_{\text{max}}|/n \xrightarrow{p} 0 \) and \( |E(C_{\text{max}})|/n \xrightarrow{p} 0 \).

Consequently, the same result holds for the uniform random graph with degree sequence \( d \) satisfying Condition 1.5(a)-(b), under the extra assumption that \( \sum_{i \in [n]} d_i^2 = O(n) \).

**Reformulation in terms of branching processes**

We start by interpreting the results in Theorem 4.4 in terms of branching processes as also arising in Section 4.1. As it turns out, we can interpret \( \xi \) as the extinction probability of a branching process, and \( \zeta \) as the survival probability of the unimodular Galton-Watson tree that appears in Theorem 4.1. Thus, \( \zeta \) satisfies

\[
\zeta = \sum_{k \geq 1} p_k(1 - \xi^k),
\]

(4.2.2)

where \( \xi \) is the extinction probability of the branching process with offspring distribution \((p_k^*)_{k \geq 0}\), which satisfies

\[
\xi = \sum_{k \geq 0} p_k^* \xi^k.
\]

(4.2.3)

Clearly, \( \xi = 1 \) precisely when

\[
\nu = \sum_{k \geq 0} kp_k^* \leq 1.
\]

(4.2.4)
By (4.1.1), we can rewrite
\[ \nu = \frac{1}{\mathbb{E}[D]} \sum_{k \geq 0} k(k + 1)p_{k+1} = \mathbb{E}[D(D - 1)]/\mathbb{E}[D], \] (4.2.5)
which explains the condition on \( \nu \) in Theorem 4.4(a). Further, to understand the asymptotics of \( v_k(C_{\text{max}}) \), we note that there are \( n_k = np_k^{(n)} \approx np_k \) vertices with degree \( k \). Each of the \( k \) direct neighbors of a vertex of degree \( k \) survives with probability close to \( 1 - \xi \), so that the probability that at least one of them survives is close to \( 1 - \xi^k \). When one of the neighbors of the vertex of degree \( k \) survives, the vertex itself is part of the giant component, which explains why \( v_k(C_{\text{max}})/n \xrightarrow{p} p_k(1 - \xi^k) \). Finally, an edge consists of two half-edges, and an edge is part of the giant component precisely when one of the vertices incident to it is, which occurs with probability \( 1 - \xi^2 \). There are in total \( \ell_n/2 = n\mathbb{E}[D_n]/2 \approx n\mathbb{E}[D]/2 \) edges, which explains why \( |E(C_{\text{max}})|/n \xrightarrow{p} 1/2 \mathbb{E}[D](1 - \xi^2) \). Therefore, all results in Theorem 4.4 have a simple explanation in terms of the branching process approximation of the connected component for CM \( n(d) \) of a uniform vertex in \( [n] \).

Reformulation in terms of generating functions
We next reformulate the results in terms of generating functions, which play a crucial role throughout our proof. Let
\[ G_D(x) = \sum_{k=0}^{\infty} p_k x^k = \mathbb{E}[x^D] \] (4.2.6)
be the probability generating function of the probability distribution \((p_k)_{k \geq 1}\). Recall that, for a non-negative random variable \( D \), the random variable \( D^* \) denotes its size-biased distribution. Define further
\[ G_D^*(x) = \mathbb{E}[x^{D^*}] = \sum_{k=1}^{\infty} p_k^* x^k = G_D'(x)/G_D'(1), \] (4.2.7)
\[ H(x) = \mathbb{E}[D|x(x - G_D^*(x))]. \] (4.2.8)
Note that \( G_D^*(1) = 1 \), and thus \( H(0) = H(1) = 0 \). Note also that
\[ H'(1) = 2\mathbb{E}[D] - \sum_k k^2 p_k = \mathbb{E}[2D - D^2] = -\mathbb{E}[D(D - 2)] \] (4.2.9)
For further properties of \( x \mapsto H(x) \), see Lemma 4.9 below. We conclude that if \( \mathbb{E}[D(D - 2)] = \sum_k k(k - 2)p_k > 0 \) and if \( p_1^* > 0 \), then there is a unique \( \xi \in (0, 1) \) such that \( H(\xi) = 0 \), or equivalently \( G_D^* (\xi) = \xi \), so that indeed \( \xi \) is the extinction probability of the branching process with offspring distribution \((p_k^*)_{k \geq 0}\). When \( p_1^* = 0 \), instead, \( \xi = 0 \) is the unique solution in \([0, 1]\) of \( H(\xi) = 0 \). The functions \( x \mapsto H(x) \) and \( x \mapsto G_D^*(x) \) play a central role in our analysis of the problem.

We prove Theorem 4.4 in Section 4.2.3 below. We now remark upon the result and on the conditions arising in it.
The phase transition in the configuration model

The condition $\mathbb{P}(D = 2) = p_2 < 1$

Because isolated vertices do not matter, without loss of generality, we may assume that $p_0 = 0$. The case $p_2 = 1$, for which $\nu = 1$ is quite exceptional. In this case, $H(x) = 0$ for all $x$. We give three examples showing that then quite different behaviors are possible.

Our first example is when $d_i = 2$ for all $i \in [n]$, so we are studying a random 2-regular graph. In this case, the components are cycles and the distribution of cycle lengths in $\text{CM}_n(d)$ is given by the Ewen’s sampling formula ESF(1/2), see e.g. Arratia et al. (2003). This implies that $|C_{\text{max}}|/n$ converges in distribution to a non-degenerate distribution on $[0, 1]$ and not to any constant (Arratia et al., 2003, Lemma 5.7). Moreover, the same is true for $|C_{(2)}|/n$ (and for $|C_{(3)}|/n$, ...), so in this case there are several large components. To intuitively see this result, we note that in the exploration of a cluster we start with one vertex with two half-edges. When pairing a half-edge, it connects to a vertex that again has two half-edges. Therefore, the number of half-edges to be paired is always equal to 2, up to the moment when the cycle is closed, and the cluster is completed. When there are $m = \alpha n$ free half-edges left, the probability of closing up the cycle equals $1/m = 1/(\alpha n)$, and, thus, the time this takes is of order $n$. A slight extension of this reasoning shows that the time it takes to close a cycle is $nT_n$, where $T_n$ converges to a limiting non-degenerate random variable (see Exercise 4.3).

Our second example with $p_2 = 1$ is obtained by adding a small number of vertices of degree 1. More precisely, we let $n_1 \to \infty$ be such that $n_1/n \to 0$, and $n_2 = n - n_1$. In this case, components can either be cycles, or strings of vertices with degree 2 terminated with two vertices with degree 1. When $n_1 \to \infty$, it is more likely to terminate a long string of vertices of degree 2 by a vertex of degree 1 than by closing the cycle, as for the latter we need to pair to a unique half-edge, while for the former, we have $n_1$ choices. Therefore, it is easy to see that this implies that $|C_{\text{max}}| = o_P(n)$ (see Exercise 4.4 for details).

Our third example with $p_2 = 1$ is obtained by instead adding a small number of vertices of degree 4 (i.e., $n_4 \to \infty$ such that $n_4/n \to 0$, and $n_2 = n - n_4$.) We can regard each vertex of degree 4 as two vertices of degree 2 that have been identified. Therefore, to obtain $\text{CM}_n(d)$ with this degree distribution, we can start from a configuration model having $N = n + n_4$ vertices, and uniformly identifying $n_4$ pairs of vertices of degree 2. Since the configuration model with $N = n + n_4$ vertices of degree 2 has many components having size of order $n$, most of these will merge into one giant component. As a result, $|C_{\text{max}}| = n - o_P(n)$, so there is a giant component containing almost everything, as you will prove yourself in Exercise 4.5.

We conclude that the case where $p_2 = \mathbb{P}(D = 2) = 1$ is quite sensitive to the precise properties of the degree structure that are not captured by the limiting distribution $(p_k)_{k \geq 1}$ only. In the sequel, we ignore the case where $p_2 = 1$. 
4.2 Phase transition in the configuration model

Reduction to the case where $P(D = 1) = p_1 > 0$

In our proof, it is convenient to assume that $p_1 = P(D = 1) > 0$. The extinction probability $\xi = 0$ and the survival probability $\zeta = 1$ when $p_1 = 0$, which causes technical difficulties in the proof. We now explain how we can reduce the case where $p_1 = 0$ to the case where $p_1 > 0$.

Let $d_{\text{min}} = \min\{k : p_k > 0\}$ be the minimum of the support of the asymptotic degree distribution $D$. Fix $\varepsilon > 0$, and assume that $\varepsilon < p_{d_{\text{min}}}$. Consider the configuration model with $\tilde{n} = n + 2d_{\text{min}}\varepsilon n$, and degree sequence $\tilde{d} = (\tilde{d}_i)_{i \in [n]}$ with $\tilde{n}_k = n_k$ for all $k > d_{\text{min}}$, $\tilde{n}_{d_{\text{min}}} = n_{d_{\text{min}}} - \varepsilon n$, $\tilde{n}_1 = 2d_{\text{min}}\varepsilon n$. This configuration model can be obtained from $\text{CM}_n(d)$ by replacing $\varepsilon n$ vertices of degree $d_{\text{min}}$ by $d_{\text{min}}$ vertices having degree 1, as if we have ‘forgotten’ that these vertices are actually equal.

Clearly, $\text{CM}_n(d)$ can be retrieved by identifying $\varepsilon n$ collections of $d_{\text{min}}$ vertices of degree 1 to a single vertex of degree $d_{\text{min}}$. When $d$ satisfies Condition 1.5(a)-(b), then so does $\tilde{d}$ with limiting degree distribution $\tilde{p}_1 = 2d_{\text{min}}\varepsilon / (1 + 2d_{\text{min}}\varepsilon)$, $\tilde{p}_{d_{\text{min}}} = (p_{d_{\text{min}}} - \varepsilon) / (1 + 2d_{\text{min}}\varepsilon)$, $\tilde{p}_k = p_k / (1 + 2d_{\text{min}}\varepsilon)$ for all $k > d_{\text{min}}$. The above procedure clearly makes $|\mathcal{C}_{\text{max}}|$ smaller. Further, with $\zeta_\varepsilon$ denoting the limit of $|\mathcal{C}_{\text{max}}| / \tilde{n}$ for $\tilde{d}$, we have that $\zeta_\varepsilon \to 1$ as $\varepsilon \downarrow 0$. As a result, Theorem 4.4 for $\zeta = 1, \xi = 0$ follows from Theorem 4.4 with $p_1 > 0$, for which $\zeta < 1$ and $\xi > 0$. In the remainder of the proof, we therefore without loss of generality assume that $\xi > 0$ and $\zeta < 1$.

Organization of the proof of Theorem 4.4

Theorem 4.4 is proved using a clever randomization scheme to explore the connected components one by one. This construction is explained terms of a simple continuous-time algorithm in Section 4.2.2 below. The algorithm describes the number of vertices of given degrees that have been found, as well as the total number of unpaired half-edges, at time $t > 0$. It is proved that, when $n \to \infty$, these quantities all converge in probability to deterministic functions described in terms of the functions $x \mapsto H(x)$ and $x \mapsto G_x^{\ast}(x)$ above. In particular, the number of unpaired half-edges is given in terms of $x \mapsto H(x)$, so that the first zero of this function gives the size of the giant component. In Section 4.2.3, the algorithm is analyzed by showing that when $\zeta > 0$, after a short initial period of exploring small clusters, the giant component is found, and the exploration explores it completely, after which no large component is left. When $\zeta = 0$, instead, only small clusters are found. A crucial aspect in the proof resides in how to deal with the depletion-of-points-and-half-edges effect.

4.2.1 The giant component is almost local

A useful truncation argument

To be added!
4.2.2 Finding the largest component

The components of an arbitrary finite graph or multigraph can be found by the following standard procedure. Pick an arbitrary vertex \( v \) and determine the component of \( v \) as follows: include all the neighbors of \( v \) in an arbitrary order; then add in the neighbors of the neighbors, and so on, until no more vertices can be added. The vertices included until this moment form the component of \( v \). If there are still vertices left in the graph, then pick any such vertex \( w \), and repeat the above to determine the second component (the component of vertex \( w \)). Carry on in this manner until all the components have been found.

The same result can be obtained in the following way, which turns out to be more convenient for the exploration of the giant component in the configuration model. Regard each edge as consisting of two half-edges, each half-edge having one endpoint. We will label the vertices as sleeping or awake (= used) and the half-edges as sleeping, active or dead; the sleeping and active half-edges are also called living. We start with all vertices and half-edges sleeping. Pick a vertex and label its half-edges as active. Then take any active half-edge, say \( x \) and find its partner \( y \) in the graph; label these two half-edges as dead. Further, if the endpoint of \( y \) is sleeping, label it as awake and all other half-edges of the vertex incident to \( y \) as active. Repeat as long as there are active half-edges. When there is no active half-edge left, we have obtained the first component. Then start again with another vertex until all components are found.

We apply this algorithm to \( \text{CM}_n(d) \) with a given degree sequence, revealing its edges during the process. We thus initially only observe the vertex degrees and the half-edges, but not how they are joined to form edges. Hence, each time we need a partner of an half-edge, it is uniformly distributed over all other living half-edges, with the understanding that the dead half-edges are the ones that are already paired into edges. It is here that we are using the specific structure of the configuration model, which simplifies the analysis substantially.

We make the random choices of finding a partner to the edges by associating i.i.d. random maximal lifetimes \( \tau_x \) to the half-edge \( x \), where \( \tau_x \) has an \( \text{Exp}(1) \) distribution. We interpret these lifetimes as clocks, and changes in our exploration process only occur when a clock of a half-edge rings. In other words, each half-edge dies spontaneously at rate 1 (unless killed earlier). Each time we need to find the partner of a half-edge \( x \), we then wait until the next living half-edge \( \neq x \) dies and take that one. This process in continuous-time can be formulated as an algorithm, constructing \( \text{CM}_n(d) \) and exploring its components simultaneously, as follows. Recall that we start with all vertices and half-edges sleeping. The exploration is then formalized in the following three steps:

**Step 1** When there is no active half-edge (as in the beginning), select a sleeping vertex and declare it awake and all its half-edges active. For definiteness, we choose the vertex by choosing a half-edge uniformly at random among all sleeping half-edges. When there is no sleeping half-edge left, the process stops; the remaining sleeping vertices are all isolated and we have explored all other components.
4.2 Phase transition in the configuration model

Step 2 Pick an active half-edge (which one does not matter) and kill it, i.e., change its status to dead.

Step 3 Wait until the next half-edge dies (spontaneously, as a result of its clock ringing). This half-edge is joined to the one killed in the previous step Step 2 to form an edge of the graph. When the vertex incident to it is sleeping, we change this vertex to awake and all other half-edges incident to it to active. Repeat from Step 1.

The above randomized algorithm is such that components are created between the successive times Step 1 is performed, where we say that Step 1 is performed when there is no active half-edge and, as a result, a new vertex is chosen.

The vertices in the component created during one of these intervals between the successive times Step 1 is performed are the vertices that are awakened during the interval. Note also that a component is completed and Step 1 is performed exactly when the number of active half-edges is 0 and a half-edge dies at a vertex where all other half-edges (if any) are dead. In the next section, we investigate the behavior of the key characteristics of the algorithm, such as the number of sleeping half-edges and the number of sleeping vertices of a given degree.

4.2.3 Analysis of the algorithm for $\text{CM}_n(d)$

We start by introducing the key characteristics of the exploration algorithm. Let $S(t)$ and $A(t)$ be the numbers of sleeping and active half-edges, respectively, at time $t$, and let

$$L(t) = S(t) + A(t)$$

be the number of living half-edges. For definiteness, we define these random functions to be right-continuous.

Let us first look at $L(t)$. We start with $\ell_n = \sum_{i \in [n]} d_i$ half-edges, all sleeping and thus living, but we immediately perform Step 1 and Step 2 and kill one of them. Thus, $L(0) = \ell_n - 1$. In the sequel, as soon as a living half-edge dies, we perform Step 3 and then (instantly) either Step 2 or both Step 1 and Step 2. Since Step 1 does not change the number of living half-edges while Step 2 and Step 3 each decrease it by 1, the total result is that $L(t)$ is decreased by 2 each time one of the living half-edges dies, except when the last living one dies and the process terminates. Because of this simple dynamics of $t \mapsto L(t)$, we can give sharp asymptotics of $L(t)$ when $n \to \infty$:

**Proposition 4.5** (Number of living half-edges) As $n \to \infty$, for any $t_0 \geq 0$ fixed,

$$\sup_{0 \leq t \leq t_0} |n^{-1} L(t) - E[D_n] e^{-2t}| \to 0. \quad (4.2.11)$$

**Proof** The process $t \mapsto L(t)$ satisfies $L(0) = \ell_n - 1$, and it decreases by 2 at rate $L(t)$. As a result, it is closely related to a death process. We study such processes in the following lemma:

**Lemma 4.6** (Asymptotics of death processes) Let $d, \gamma > 0$ be given and let
The proof follows by distinguishing several cases. First assume that $d = 1$ and that $x$ is an integer. In this case, the process is a standard pure death process taking the values $x, x - 1, x - 2, \ldots, 0$, describing the number of particles alive when the particles die independently at rate $\gamma > 0$. As is well known, and easily seen by regarding $N^{(x)}(t)$ as the sum of $x$ independent copies of the process $N^{(1)}(t)$, the process $(e^{\gamma t} N^{(x)}(t))_{t \geq 0}$, is a martingale starting in $x$. Furthermore, for every $t \geq 0$, the random variable $N^{(x)}(t)$ has a $\text{Bin}(x, e^{-\gamma t})$ distribution, since each of the $x$ particles has a probability of dying before time $t$ of $e^{-\gamma t}$, and the different particles die independently.

We rely on Doob’s martingale inequality, which states that for a martingale $(M_n)_{n \geq 0}$,

$$
\mathbb{E} \left[ \sup_{m \leq n} |M_m - \mathbb{E}[M_m]|^2 \right] \leq \text{Var}(M_n). \tag{4.2.13}
$$

Application of Doob’s martingale inequality yields

$$
\mathbb{E} \left[ \sup_{t \leq t_0} |N^{(x)}(t) - e^{-\gamma t} x|^2 \right] \leq 4 \mathbb{E} \left[ \sup_{t \leq t_0} |e^{\gamma t} N^{(x)}(t) - x|^2 \right] \leq 4 \mathbb{E} \left[ (e^{\gamma t} N^{(x)}(t_0) - x)^2 \right]
$$

$$
= 4(e^{\gamma t_0} \text{Var}(N^{(x)}(t_0))) \leq 4(e^{\gamma t_0} - 1)x. \tag{4.2.14}
$$

This proves the claim for $x$ being integer.

Next, we still assume $d = 1$, but let $x > 0$ be arbitrary. We can couple the two processes $(N^{(x)}(t))_{t \geq 0}$ and $(N^{(x)}(t))_{t \geq 0}$ with different initial values such that whenever the smaller one jumps by 1, so does the other. This coupling keeps

$$
|N^{(x)}(t) - N^{(x)}(t)| < 1 \tag{4.2.15}
$$

for all $t \geq 0$, and thus,

$$
\sup_{t \leq t_0} |N^{(x)}(t) - e^{-\gamma t} x| \leq \sup_{t \leq t_0} |N^{(x)}(t) - e^{-\gamma t} x| + 2, \tag{4.2.16}
$$

so that by (4.2.14), in turn,

$$
\mathbb{E} \left[ \sup_{t \leq t_0} |N^{(x)}(t) - e^{-\gamma t} x|^2 \right] \leq 8(e^{\gamma t_0} - 1)x + 8. \tag{4.2.17}
$$

Finally, for a general $d > 0$, we observe that $N^{(x)}(t)/d$ is a process of the same type with the parameters $(\gamma, d, x)$ replaced by $(\gamma d, 1, x/d)$, and the general result follows from (4.2.17) and (4.2.14). \qed

The proof of Proposition 4.5 follows from Lemma 4.6 with $d = 2, x = (\ell_n - 1) = n \mathbb{E}[D_n] - 1$ and $\gamma = 1$. \qed
We continue by considering the sleeping half-edges \( S(t) \). Let \( V_k(t) \) be the number of sleeping vertices of degree \( k \) at time \( t \), so that

\[
S(t) = \sum_{k=1}^{\infty} kV_k(t). \tag{4.2.18}
\]

Note that Step 2 does not affect sleeping half-edges, and that Step 3 implies that each sleeping vertex of degree \( k \) is eliminated (i.e., awakened) with intensity \( k \), independently of all other vertices. There are also some sleeping vertices eliminated by Step 1, though, which complicates the dynamics of \( t \mapsto V_k(t) \). It is here that the depletion-of-points-and-half-edges effect enters the analysis of the component structure of \( \text{CM}_n(d) \). This effect is complicated, but we will see it is quite harmless. This can be understood by noting that we only apply Step 1 when we have completed exploring an entire component. Since we will be mainly interested in settings where the giant component is large, we will see that we will not be using Step 1 very often before having completely explored the giant component. After having completed the exploration of the giant component, we start using Step 1 again quite frequently, but it will turn out that then it is very unlikely to be exploring any particularly large connected components. Thus, we can have a setting in mind where the number of applications of Step 1 is quite small.

With this intuition in mind, we first ignore the effect of Step 1 by letting \( \tilde{V}_k(t) \) be the number of vertices of degree \( k \) such that all its half-edges have maximal lifetimes \( \tau_x > t \). Thus, none of its \( k \) half-edges would have died spontaneously up to time \( t \), assuming they all escaped Step 1. Let us now explain in some more detail why it is reasonable to ignore the effect of Step 1 in the leading order when there exists a giant component. Indeed, we perform Step 1 until we hit the giant component, and then it takes a long time to find the entire giant component. When \( \zeta > 0 \), the number of times we perform Step 1 until we find the giant component will be small (probably a tight random variable), as each time we have a strictly positive probability of choosing a vertex in the giant component. Thus, intuitively, we expect the difference between \( V_k(t) \) and \( \tilde{V}_k(t) \) to be insignificant. Thus, we start by focussing on the dynamics of \( \tilde{V}_k(t) \), and later adapt for this omission.

For a given half-edge, we call the half-edges incident to the same vertex its sibling half-edges. Let further

\[
\tilde{S}(t) = \sum_k k\tilde{V}_k(t) \tag{4.2.19}
\]

denote the number of half-edges whose sibling half-edges have escaped spontaneous death up to time \( t \). Comparing with (4.2.18), we see that the process \( S(t) \) ignores the effect of Step 1 in an identical way as \( \tilde{V}_k(t) \).

Recall the functions \( G_D \), \( G_D^\star \) from (4.2.6)–(4.2.7), and define

\[
h(x) = x\mathbb{E}[D]G_D^\star(x). \tag{4.2.20}
\]
Then, we can identify the asymptotics of $(\tilde{V}_k(t))_{t \geq 0}$ in a similar way as in Proposition 4.5:

**Lemma 4.7** (Number of living vertices of degree $k$) Assume that Conditions 1.5(a)-(b) hold. Then, as $n \to \infty$, for any $t_0 \geq 0$ fixed

$$\sup_{t \leq t_0} |n^{-1} \tilde{V}_k(t) - p_k e^{-kt}| \stackrel{p}{\to} 0$$

(4.2.21)

for every $k \geq 0$ and

$$\sup_{t \leq t_0} |n^{-1} \sum_{k=0}^{\infty} \tilde{V}_k(t) - G_D(e^{-t})| \stackrel{p}{\to} 0,$$  

(4.2.22)

$$\sup_{t \leq t_0} |n^{-1} \tilde{S}(t) - h(e^{-t})| \stackrel{p}{\to} 0.$$  

(4.2.23)

**Proof** The statement (4.2.21) again follows from Lemma 4.6, now with $\gamma = k$, $x = n_k$, and $d = 1$. The case $k = 0$ is trivial, with $\tilde{V}_0(t) = n_0$ for all $t$. We can replace $p'_k(n) = n_k/n$ by $p_k$ by Condition 1.5(a).

By Condition 1.5(b), the sequence of random variables $(D_n)_{n \geq 1}$ is uniformly integrable, which means that for every $\varepsilon > 0$ there exists $K < \infty$ such that for all $n$, $\sum_{k > K} k n_k/n = \mathbb{E}[D_n|D_n > k] < \varepsilon$. We may further assume (or deduce from Fatou’s inequality) that $\sum_{k > K} k p_k < \varepsilon$, and obtain by (4.2.21) that, whp,

$$\sup_{t \leq t_0} |n^{-1} \tilde{S}(t) - h(e^{-t})| = \sup_{t \leq t_0} \left| n^{-1} \sum_{k=1}^{\infty} k(n^{-1} \tilde{V}_k(t) - p_k e^{-kt}) \right|$$

$$\leq \sum_{k=1}^{K} k \sup_{t \leq t_0} |n^{-1} \tilde{V}_k(t) - p_k e^{-kt}| + \sum_{k > K} k \left( \frac{n_k}{n} + p_k \right)$$

$$\leq \varepsilon + \varepsilon + \varepsilon,$$

proving (4.2.23). An almost identical argument yields (4.2.22). \hfill \Box

Remarkably, the difference between $S(t)$ and $\tilde{S}(t)$ is easily estimated. The following result can be viewed as the key to why this approach works. Indeed, it gives a uniform upper bound on the difference due to the application of Step 1:

**Lemma 4.8** (Effect of Step 1) Let $d_{\max} := \max_{i \in [n]} d_i$ be the maximum degree of $\text{CM}_n(d)$. Then

$$0 \leq \tilde{S}(t) - S(t) < \sup_{0 \leq s \leq t} (\tilde{S}(s) - L(s)) + d_{\max}.$$  

(4.2.24)

The process $(\tilde{S}(t))_{t \geq 0}$ runs on scale $n$ (see e.g., the related statement for $(L(t))_{t \geq 0}$ in Proposition 4.5). Further, $d_{\max} = o(n)$ when Conditions 1.5(a)-(b) hold. Finally, one can expect that $\tilde{S}(s) \leq L(s)$ holds, since the difference is related to the number of active half-edges. Thus, intuitively, $\sup_{0 \leq s \leq t} (\tilde{S}(s) - L(s)) = \tilde{S}(0) - L(0) = 0$. We will make that argument precise after we have proved Lemma
4. Phase transition in the configuration model

4.8. We then conclude that \( \tilde{S}(t) - S(t) = o_\infty(n) \), and so they have the same limit after rescaling by \( n \). Let us now prove Lemma 4.8:

**Proof.** Clearly, \( V_n(t) \leq \tilde{V}_k(t) \), and thus \( S(t) \leq \tilde{S}(t) \). Furthermore, \( \tilde{S}(t) - S(t) \) increases only as a result of Step 1. Indeed, Step 1 acts to guarantee that \( A(t) = L(t) - S(t) \geq 0 \), and is only performed when \( A(t) = 0 \).

If Step 1 is performed at time \( t \) and a vertex of degree \( j > 0 \) is awakened, then Step 2 applies instantly and we have \( A(t) = j - 1 < d_{\max} \), and consequently
\[
\tilde{S}(t) - S(t) = \tilde{S}(t) - L(t) + A(t) < \tilde{S}(t) - L(t) + d_{\max}. \tag{4.2.25}
\]
Furthermore, \( \tilde{S}(t) - S(t) \) is never changed by Step 2 and either unchanged or decreased by Step 3. Hence, \( \tilde{S}(t) - S(t) \) does not increase until the next time Step 1 is performed. Consequently, for any time \( t \), if \( s \) was the last time before (or equal to) \( t \) that Step 1 was performed, then \( \tilde{S}(t) - S(t) \leq \tilde{S}(s) - S(s) \), and the result follows by (4.2.25). \( \square \)

Let us now set the stage for taking the limits of \( n \to \infty \). Recall that \( A(t) = L(t) - S(t) \) denotes the number of awakened vertices and let
\[
\tilde{A}(t) = L(t) - \tilde{S}(t) = A(t) - (\tilde{S}(t) - S(t)) \tag{4.2.26}
\]
denote the number of awakened vertices ignoring the effect of Step 1. Thus, \( \tilde{A}(t) \leq A(t) \) since \( S(t) \leq \tilde{S}(t) \). We will use \( \tilde{A}(t) \) as a proxy for \( A(t) \) in s similar way as \( \tilde{S}(t) \) is used a a proxy for \( S(t) \).

Recall the definition of \( H(x) \) in (4.2.8). By Lemmas 4.5 and 4.7 and the definition that \( \tilde{A}(t) = L(t) - \tilde{S}(t), \) for any \( t_0 \geq 0 \),
\[
\sup_{t \leq t_0} |n^{-1} \tilde{A}(t) - H(e^{-t})| \to 0. \tag{4.2.27}
\]
Lemma 4.8 can be rewritten as
\[
0 \leq \tilde{S}(t) - S(t) < - \inf_{s \leq t} \tilde{A}(s) + d_{\max}. \tag{4.2.28}
\]
By (4.2.26) and (4.2.28),
\[
\tilde{A}(t) \leq A(t) < \inf_{s \leq t} \tilde{A}(s) + d_{\max}, \tag{4.2.29}
\]
which, perhaps, illuminates the relation between \( A(t) \) and \( \tilde{A}(t) \). Recall that connected components are explored between subsequent zeros of the process \( t \to A(t) \). The function \( t \to H(e^{-t}) \), which acts as the limit of \( \tilde{A}(t) \) (and thus hopefully also of \( A(t) \)), is strictly positive in \((0, -\log \xi)\) and \( H(1) = H(\xi) = 0 \). Therefore, we expect \( \tilde{A}(t) \) to be positive for \( t \in (0, -\log \xi) \), and, if so, \( \inf_{s \leq t} \tilde{A}(s) = 0 \).

This would prove that indeed \( \tilde{A}(t) \) and \( A(t) \) are close on this entire interval, and the exploration on the interval \( t \in (0, -\log \xi) \) will turn out to correspond to the exploration of the giant component.

The idea is to continue our algorithm in Step 1-Step 3 until the giant component has been found, which implies that \( A(t) > 0 \) for the time of exploration of
the giant component, and $A(t) = 0$ for the first time when we have completed the exploration of the giant component, which is $t = -\log \xi$. Thus, the term $\inf_{s \leq t} \bar{A}(s)$ in (4.2.29) ought to be negligible. When Conditions 1.5(a)-(b) hold, we further have that $d_{\max} = o(n)$, so that one can expect $\bar{A}(t)$ to be a good approximation of $A(t)$. The remainder of the proof makes this intuition precise. We start by summarizing some useful analytical properties of $x \mapsto H(x)$ that we rely upon in the sequel:

**Lemma 4.9** (Properties of $x \mapsto H(x)$) Suppose that Conditions 1.5(a)-(b) hold and let $H(x)$ be given by (4.2.8). Suppose also that $p_2 < 1$.

(i) If $\nu = \mathbb{E}[D(D - 1)]/\mathbb{E}[D] > 1$ and $p_1 > 0$, then there is a unique $\xi \in (0, 1)$ such that $H(\xi) = 0$. Moreover, $H(x) < 0$ for all $x \in (0, \xi)$ and $H(x) > 0$ for all $x \in (\xi, 1)$.

(ii) If $\nu = \mathbb{E}[D(D - 1)]/\mathbb{E}[D] \leq 1$, then $H(x) < 0$ for all $x \in (0, 1)$.

**Proof** As remarked earlier, $H(0) = H(1) = 0$ and $H'(1) = -\mathbb{E}[D(D - 2)]$. Furthermore, if we define $\phi(x) := H(x)/x$, then $\phi(x) = \mathbb{E}[D](x - G_0^x(x))$ is a concave function on $(0, 1)$, and it is strictly concave unless $p_k = 0$ for all $k \geq 3$, in which case $H'(1) = -\mathbb{E}[D(D - 2)] = p_1 > 0$. Indeed, $p_1 + p_2 = 1$ when $p_k = 0$ for all $k \geq 3$. Since we assume that $p_2 < 1$, we thus obtain that $p_1 > 0$ in this case.

In case (ii), we thus have that $\phi$ is concave and $\phi'(1) = H'(1) - H(1) \geq 0$, with either the concavity or the inequality strict, and thus $\phi'(x) > 0$ for all $x \in (0, 1)$, whence $\phi(x) < \phi(1) = 0$ for $x \in (0, 1)$.

In case (i), $H'(1) < 0$, and thus $H(x) > 0$ for $x$ close to 1. Further, when $p_1 > 0$, $H'(0) = -h'(0) = -p_1 < 0$, and thus $H(x) \leq 0$ for $x$ close to 0. Hence, there is at least one $\xi \in (0, 1)$ with $H(\xi) = 0$, and since $H(x)/x$ is strictly concave and also $H(1) = 0$, there is at most one such $\xi$ and the result follows. $
$}

Now we are in the position to complete the proof of Theorem 4.4 in the following section.

### 4.2.4 Proof of Theorem 4.4

We start with the proof of Theorem 4.4(i). Let $\xi$ be the zero of $H$ given by Lemma 4.9(i) and let $\theta = -\log \xi$. Then, by Lemma 4.9, $H(e^{-t}) > 0$ for $0 < t < \theta$, and thus $\inf_{t \leq \theta} H(e^{-t}) = 0$. Consequently, (4.2.27) implies

$$n^{-1} \inf_{t \leq \theta} \bar{A}(t) = \inf_{t \leq \theta} n^{-1} \bar{A}(t) - \inf_{t \leq \theta} H(e^{-t}) \xrightarrow{p} 0. \tag{4.2.30}$$

Further, by Condition 1.5(b), $d_{\max} = o(n)$, and thus $n^{-1}d_{\max} \to 0$. Consequently, (4.2.28) and (4.2.30) yield

$$\sup_{t \leq \theta} n^{-1} |A(t) - \bar{A}(t)| = \sup_{t \leq \theta} n^{-1} |\bar{S}(t) - S(t)| \xrightarrow{p} 0. \tag{4.2.31}$$

Thus, by (4.2.27),

$$\sup_{t \leq \theta} |n^{-1} A(t) - H(e^{-t})| \xrightarrow{p} 0. \tag{4.2.32}$$
This will be the work horse of our argument. By Lemma 4.9, we know that \( t \mapsto H(e^{-t}) \) is positive on \((0, -\log \xi)\) when \( \nu > 1 \). Thus, the exploration in the interval \((0, -\log \xi)\) will find the giant component. We now make this intuition precise. In particular, we need to show that no large connected component is found before or after this interval (showing that the giant is unique), and we need to investigate the properties of the giant, in terms of number of edges, vertices of degree \( k \) etc. We now provide these details.

Let \( 0 < \varepsilon < \theta/2 \). Since \( H(e^{-t}) > 0 \) on the compact interval \([\varepsilon, \theta - \varepsilon]\), (4.2.32) implies that \( A(t) \) remains whp positive on \([\varepsilon, \theta - \varepsilon]\), and thus no new component is started during this interval.

On the other hand, again by Lemma 4.9(i), \( H(e^{-(\theta+\varepsilon)}) < 0 \) and (4.2.27) implies that \( n^{-1}A(\theta + \varepsilon) \to H(e^{-(\theta+\varepsilon)}) \), while \( A(\theta + \varepsilon) \geq 0 \). Thus, with \( \Delta = |H(e^{-\theta-\varepsilon})|/2 > 0 \), whp

\[
\hspace{1cm} S(\theta + \varepsilon) - S(\theta) = A(\theta + \varepsilon) - \tilde{A}(\theta + \varepsilon) \geq -\tilde{A}(\theta + \varepsilon) > n\Delta,
\]

while (4.2.31) yields that \( S(\theta) - S(\theta) < n\Delta \) whp. Consequently, whp \( S(\theta + \varepsilon) - S(\theta + \varepsilon) > S(\theta - S(\theta)) \), so whp Step 1 is performed between the times \( \theta \) and \( \theta + \varepsilon \).

Let \( T_1 \) be the last time Step 1 was performed before time \( \theta/2 \). Let \( T_2 \) be the next time Step 1 is performed (by convention, \( T_2 = \infty \) if such a time does not exist). We have shown that for every \( \varepsilon > 0 \), and whp \( 0 \leq T_1 \leq \varepsilon \) and \( \theta - \varepsilon \leq T_2 \leq \theta + \varepsilon \).

In other words, \( T_1 \to 0 \) and \( T_2 \to \theta \). We conclude that we have found one component that is explored between time \( T_1 \to 0 \) and time \( T_2 \to \theta \). This is our candidate for the giant component, and we continue to study its properties, i.e., its size, its number of edges and its number of vertices of degree \( k \). These properties are stated separately in the next proposition, so that we are able to reuse them later on:

**Proposition 4.10 (Connected component properties)** Let \( T_1^* \) and \( T_2^* \) be two random times when Step 1 is performed, with \( T_1^* \leq T_2^* \), and assume that \( T_1^* \to t_1 \) and \( T_2^* \to t_2 \) where \( 0 \leq t_1 \leq t_2 \leq \theta < \infty \). If \( C^* \) is the union of all components explored between \( T_1^* \) and \( T_2^* \), then

\[
\begin{align*}
v_k(C^*)/n & \to p_k(e^{-kt_1} - e^{-kt_2}), \quad k \geq 0, \\
|C^*|/n & \to G_D(e^{-t_1}) - G_D(e^{-t_2}), \\
|E(C^*)|/n & \to \frac{1}{2} h(e^{-t_1}) - \frac{1}{2} h(e^{-t_2}).
\end{align*}
\]

In particular, if \( t_1 = t_2 \), then \( |C^*|/n \to 0 \) and \( |E(C^*)| \to 0 \).

Below, we apply Proposition 4.10 to \( T_1 = o(1) \) and \( T_2 \), where \( T_2 = \theta + o(1) \). We can identify the values of the above constants for \( t_1 = 0 \) and \( t_2 = \theta \) as \( e^{-kt_1} = 1, e^{-kt_2} = \xi, G_D(e^{-t_1}) = 1, G_D(e^{-t_2}) = 1 - \zeta \), \( h(e^{-t_1}) = 2E[D], h(e^{-t_2}) = 2E[D] \xi^2 \) (see Exercise 4.6).
The phase transition in the configuration model

By Proposition 4.10 and Exercise 4.6, Theorem 4.4(i) follows when we prove that the connected component found between times $T_1$ and $T_2$ is indeed the giant component. This is proved after we complete the proof of Proposition 4.10:

**Proof.** The set of vertices $\mathcal{C}^*$ contains all vertices awakened in the interval $[T_1^*, T_2^*]$ and no others, and thus (writing $V_k(t-) = \lim_{s \searrow t} V_k(s)$)

$$v_k(\mathcal{C}^*) = V_k(T_1^*) - V_k(T_2^*), \quad k \geq 1. \quad (4.2.37)$$

Since $T_2^* \xrightarrow{\mathcal{P}} t_2 \leq \theta$ and $H$ is continuous, we obtain that $\inf_{t \leq t_2} H(e^{-t}) \xrightarrow{\mathcal{P}} \inf_{t \leq t_2} H(e^{-t}) = 0$, where the latter equality follows since $H(1) = 0$. Now, (4.2.27) and (4.2.28) imply, in analogy with (4.2.30) and (4.2.31), that $n^{-1} \inf_{t \leq t_2} A(t) \xrightarrow{\mathcal{P}} 0$ and thus also

$$\sup_{t \leq t_2} n^{-1} |\bar{S}(t) - S(t)| \xrightarrow{\mathcal{P}} 0. \quad (4.2.38)$$

Since $\bar{V}_j(t) \geq V_j(t)$ for every $j$ and $t \geq 0$,

$$\bar{V}_k(t) - V_k(t) \leq k^{-1} \sum_{j=1}^{\infty} j(\bar{V}_j(t) - V_j(t)) = k^{-1}(\bar{S}(t) - S(t)), \quad k \geq 1. \quad (4.2.39)$$

Hence (4.2.38) implies, for every $k \geq 1$, $\sup_{t \leq t_2} |\bar{V}_k(t) - V_k(t)| = o(n)$. Consequently, using Lemma 4.7, for $j = 1, 2$,

$$V_k(T_j^*) = \bar{V}_k(T_j^*) + o(n) = n p_1 e^{-kT_j^*} + o(n) = np_1 e^{-kt_1} + o(n), \quad (4.2.40)$$

and (4.2.34) follows by (4.2.37). Similarly, using $\sum_{k=0}^{\infty} (\bar{V}_k(t) - V_k(t)) \leq \bar{S}(t) - S(t)$,

$$|\mathcal{C}^*| = \sum_{k=1}^{\infty} (V_k(T_1^*) - V_k(T_2^*)) = \sum_{k=1}^{\infty} (\bar{V}_k(T_1^*) - \bar{V}_k(T_2^*)) + o(n) \quad (4.2.41)$$

$$= nG_{\alpha}(e^{-T_1^*}) - nG_{\alpha}(e^{-T_2^*}) + o(n),$$

and

$$2|E(\mathcal{C}^*)| = \sum_{k=1}^{\infty} k(V_k(T_1^*) - V_k(T_2^*)) = \sum_{k=1}^{\infty} k(\bar{V}_k(T_1^*) - \bar{V}_k(T_2^*)) + o(n) \quad (4.2.42)$$

$$= nh(e^{-T_1^*}) - nh(e^{-T_2^*}) + o(n),$$

and (4.2.35) and (4.2.36) follow from the convergence $T_i^* \xrightarrow{\mathcal{P}} t_i$ and the continuity of $t \mapsto G_{\alpha}(e^{-t})$ and $t \mapsto h(e^{-t})$. \hfill \Box

We are now ready to complete the proof of Theorem 4.4:

**Proof of Theorem 4.4.** Let $\mathcal{C}'_{\text{max}}$ be the component created at time $T_1$ and explored until time $T_2$, where we recall that $T_1$ is the last time Step 1 was performed before time $\theta/2$ and let $T_2$ be the next time it is performed if this occurs and $T_2 = \infty$ otherwise. Then, $T_1 \xrightarrow{\mathcal{P}} 0$ and $T_2 \xrightarrow{\mathcal{P}} \theta$. The cluster $\mathcal{C}'_{\text{max}}$ is our candidate for
the giant component \( \mathcal{C}_{\text{max}} \), and we next prove that indeed it is, whp, the largest connected component.

By Proposition 4.10, with \( t_1 = 0 \) and \( t_2 = \theta \),
\[
|v_k(\mathcal{C}'_{\text{max}})|/n \xrightarrow{\Pr} p_k(1 - e^{-\xi t}),
\]
\[
|\mathcal{C}'_{\text{max}}|/n \xrightarrow{\Pr} G_\alpha(1) - G_\alpha(e^{-\theta}) = 1 - G_\alpha(\xi),
\]
\[
|E(\mathcal{C}'_{\text{max}})|/n \xrightarrow{\Pr} \frac{1}{2}(h(1) - h(e^{-\theta})) = \frac{1}{2}(h(1) - h(\xi)) = \frac{E[D]}{2}(1 - \xi^2),
\]
using Exercise 4.6. We have found one large component \( \mathcal{C}'_{\text{max}} \) with the claimed numbers of vertices and edges. It remains to show that whp there is no other large component. The basic idea is that if there exists another component that has at least \( \eta \ell_n \), half-edges in it, then it should have a reasonable chance of actually being found quickly. Since we can show that the probability of finding a large component before \( T_1 \) or after \( T_2 \) is small, there just cannot be any other large connected component. Let us now make this intuition precise.

**No early large component**

Here we first show that it is unlikely that a large component different from \( \mathcal{C}'_{\text{max}} \) is found before time \( T_1 \). For this, let \( \eta > 0 \), and apply Proposition 4.10 to \( T_0 = 0 \) and \( T_1 \), where \( T_1 \) was defined to be the last time Step 1 was performed before time \( \theta/2 \). Then, since \( T_1 \xrightarrow{\Pr} 0 \), the total number of vertices and edges in all components found before \( \mathcal{C}'_{\text{max}} \), i.e., before time \( T_1 \), is \( o(n) \). Hence, recalling that \( \ell_n = \Theta(n) \) by Condition 1.5(b),
\[
\mathbb{P} \text{(a component } \mathcal{C} \text{ with } |E(\mathcal{C})| \geq \eta \ell_n \text{ is found before } \mathcal{C}'_{\text{max}} \text{)} \to 0. \tag{4.2.46}
\]
We conclude that whp no component containing at least \( \eta \ell_n \) half-edges is found before \( \mathcal{C}'_{\text{max}} \) is found.

**No late large component**

In order to study the probability of finding a component containing at least \( \eta \ell_n \) edges after \( \mathcal{C}'_{\text{max}} \) is found, we start by letting \( T_3 \) be the first time after time \( T_2 \) that Step 1 is performed. Since \( \tilde{S}(t) - S(t) \) increases by at most \( d_{\text{max}} = o(n) \) each time Step 1 is performed, we obtain from (4.2.38) that
\[
\sup_{t \leq T_3} (\tilde{S}(t) - S(t)) \leq \sup_{t \leq T_2} (\tilde{S}(t) - S(t)) + d_{\text{max}} = o(n).
\]
Comparing this to (4.2.33), for every \( \varepsilon > 0 \) and whp, we have that \( \theta + \varepsilon > T_3 \). Since also \( T_3 > T_2 \xrightarrow{\Pr} \theta \), it follows that \( T_3 \xrightarrow{\Pr} \theta \). If \( \mathcal{C}' \) is the component created between \( T_2 \) and \( T_3 \), then Proposition 4.10 applied to \( T_2 \) and \( T_3 \) yields \( |E(\mathcal{C}')|/n \xrightarrow{\Pr} 0 \) and \( |E(\mathcal{C}')| \xrightarrow{\Pr} 0 \).

On the other hand, if there would exist a component \( \mathcal{C} \neq \mathcal{C}'_{\text{max}} \) in \( \text{CM}_n(d) \) with at least \( \eta \ell_n \) edges that has not been found before \( \mathcal{C}'_{\text{max}} \), then with probability at least \( \eta \), the vertex chosen at random by Step 1 at time \( T_2 \) starting the component \( \mathcal{C}' \) would belong to \( \mathcal{C} \). When this occurs, we clearly have that \( \mathcal{C} = \mathcal{C}' \).
Consequently,
\[ \mathbb{P}(\text{a component } C \text{ with } |E(C)| \geq \eta \ell_n \text{ is found after } C''_{\max}) \leq \eta^{-1} \mathbb{P}(|E(C'')| \geq \eta \ell_n) \to 0, \]

since \( |E(C'')| \xrightarrow{p} 0 \).

Completion of the proof of Theorem 4.4(i)

Combining (4.2.46) and (4.2.48), we see that whp there is no component except \( C''_{\max} \) that has at least \( \eta \ell_n \) edges. As a result, we must have that \( C''_{\max} = C_{\max} \), where \( C_{\max} \) is the largest component. Further, again whp, \( |E(C''_{(2)})| < \eta \ell_n \). Consequently, the results for \( C_{\max} \) follow from (4.2.43)-(4.2.45). We have further shown that \( |E(C(2))|/\ell_n \xrightarrow{p} 0 \), which implies that \( |E(C''_{(2)})|/n \xrightarrow{p} 0 \) and \( |C_{(2)}|/n \xrightarrow{p} 0 \) because \( \ell_n = \Theta(n) \) and \( |C_{(2)}| \leq |E(C''_{(2)})| + 1 \). This completes the proof of Theorem 4.4(i).

Proof of Theorem 4.4(ii)

The proof of Theorem 4.4(ii) is similar to the last step in the proof for Theorem 4.4(i). Indeed, let \( T_1 = 0 \) and let \( T_2 \) be the next time Step 1 is performed, or \( T_2 = \infty \) when this does not occur. Then,
\[ \sup_{t \leq T_2} |A(t) - \tilde{A}(t)| = \sup_{t \leq T_2} |\tilde{S}(t) - S(t)| \leq 2d_{\max} = o(n). \]  

For every \( \varepsilon > 0 \), \( n^{-1} \tilde{A}(\varepsilon) \xrightarrow{p} H(e^{-\varepsilon}) < 0 \) by (4.2.27) and Lemma 4.9(ii), while \( A(\varepsilon) \geq 0 \), and it follows from (4.2.49) that whp \( T_2 < \varepsilon \). Hence, \( T_2 \xrightarrow{p} 0 \). We apply Proposition 4.10 (which holds in this case too, with \( \theta = 0 \)) and find that if \( C \) is the first component found, then \( |E(C)|/n \xrightarrow{p} 0 \).

Let \( \eta > 0 \). If \( |E(C''_{\max})| \geq \eta \ell_n \), then the probability that the first half-edge chosen by Step 1 belongs to \( C''_{\max} \), and thus \( C = C''_{\max} \), is \( 2|E(C''_{\max})|/(2\ell_n) \geq \eta \), and hence,
\[ \mathbb{P}(|E(C''_{\max})| \geq \eta \ell_n) \leq \eta^{-1} \mathbb{P}(|E(C)| \geq \eta \ell_n) \to 0. \]

The results follow since \( \ell_n = \Theta(n) \) by Condition 1.5(b) and \( |C_{\max}| \leq |E(C_{\max})| + 1 \). This completes the proof of Theorem 4.4(ii), and thus that of Theorem 4.4.

4.2.5 The giant component of related random graphs

In this section, we extend the results of Theorem 4.4 to some related models, such as uniform simple random graphs with a given degree sequence, as well as generalized random graphs.

Recall that \( UG_n(d) \) denotes a uniform simple random graph with degrees \( d \) (see [Volume 1, Section 7.5]). The results in Theorem 4.4 also hold for \( UG_n(d) \) when we assume that Conditions 1.5(a)-(c) hold:
4.2 Phase transition in the configuration model

**Theorem 4.11** (Phase transition in \( \text{UG}_n(d) \)) Let \( d \) satisfy Conditions 1.5(a)-(c). Then, the results in Theorem 4.4 also hold for a uniform simple graph with degree sequence \( d \).

**Proof** By [Volume 1, Corollary 7.17], and since \( d = (d_i)_{i \in [n]} \) satisfies Condition 1.5(a)-(c), any event \( E_n \) that occurs whp for CM\(_n(d)\), also occurs whp for UG\(_n(d)\). By Theorem 4.4, the event \( E_n \) that \( \{|C_{\text{max}}|/n - \zeta| \leq \varepsilon \} \) occurs whp for CM\(_n(d)\), so it also holds whp for UG\(_n(d)\). The proof for the other limits is identical. \( \square \)

Note that it is not obvious how to extend Theorem 4.11 to the case where \( \nu = \infty \), which we discuss now:

**Theorem 4.12** (Phase transition in \( \text{UG}_n(d) \) for \( \nu = \infty \)) Let \( d \) satisfy Conditions 1.5(a)-(b), and assume that there exists \( p > 1 \) such that

\[
\mathbb{E}[D_n^p] \to \mathbb{E}[D]^p. \tag{4.2.51}
\]

Then, the results in Theorem 4.4 also hold for a uniform simple graph with degree sequence \( d \).

**Proof** We do not present the entire proof, but rather sketch a route towards it following Bollobás and Riordan (2015), who show that for every \( \varepsilon > 0 \), there exists \( \delta = \delta(\varepsilon) > 0 \) such that

\[
\mathbb{P}(\{|C_{\text{max}}| - \zeta n| \geq \varepsilon n\}) \leq e^{-\delta n}, \tag{4.2.52}
\]

and

\[
\mathbb{P}(|\nu_k(C_{\text{max}}) - p_k(1 - \xi^k)n| \geq \varepsilon n) \leq e^{-\delta n}. \tag{4.2.53}
\]

This exponential concentration is quite convenient, as it allows us to extend the result to the setting of uniform random graphs by conditioning CM\(_n(d)\) to be simple. Indeed, since Conditions 1.5(a)-(b) imply that

\[
\mathbb{P}(\text{CM}_n(d) \text{ simple}) = e^{-o(n)}, \tag{4.2.54}
\]

it follows that the results also hold for the uniform simple random graph UG\(_n(d)\) when Conditions 1.5(a)-(b) hold. See Exercise 4.9 below.

We next prove Theorem 3.16 for rank-1 inhomogeneous random graphs, as already stated in Theorem 3.17, and restated here for convenience:

**Theorem 4.13** (Phase transition in GRG\(_n(w)\)) Let \( w \) satisfy Condition 1.1(a)-(c). Then, the results in Theorem 4.4 also hold for GRG\(_n(w)\), CL\(_n(w)\) and NR\(_n(w)\).

**Proof** Let \( d_i \) be the degree of vertex \( i \) in GRG\(_n(w)\) defined in [Volume 1, (1.3.18)], where we use a small letter to avoid confusion with \( D_n \), which is the degree of a uniform vertex in \([n]\). By [Volume 1, Theorem 7.18], the law of GRG\(_n(w)\) conditionally on the degrees \( d \) and CM\(_n(d)\) conditionally on being simple agree. Assume that \( (d_i)_{i \in [n]}\) satisfies that Conditions 1.5(a)-(c) hold in probability. Then, by [Volume 1, Theorem 7.18] and Theorem 4.4, the results in Theorem 4.4 also hold for GRG\(_n(w)\). By [Volume 1, Theorem 6.20], the same result applies to
The phase transition in the configuration model

CL$_n(w)$, and by [Volume 1, Exercise 6.39], also to NR$_n(w)$. The fact that Conditions 1.1(a)-(c) imply that Conditions 1.5(a)-(c) hold for GRG$_n(w)$ is stated in Theorem 1.7.

Unfortunately, when $\nu = \infty$, we cannot rely on the fact that by [Volume 1, Theorem 7.18], the law of GRG$_n(w)$ conditionally on the degrees $d$ and CM$_n(d)$ conditionally on being simple agree. Indeed, when $\nu = \infty$, the probability that CM$_n(d)$ is simple vanishes. Therefore, we instead rely on a truncation argument to extend Theorem 4.13 to the case where $\nu = \infty$. It is here that the monotonicity of GRG$_n(w)$ in terms of the edge probabilities can be used rather conveniently:

**Theorem 4.14 (Phase transition in GRG$_n(w)$)** Let $w$ satisfy Conditions 1.1(a)-(b). Then, the results in Theorem 4.4 also hold for GRG$_n(w)$, CL$_n(w)$ and NR$_n(w)$.

**Proof** We only prove that $|\mathcal{C}_{\max}/n \xrightarrow{\mathbb{P}} \zeta$, the other statements in Theorem 4.4 can be proved in a similar fashion (see Exercise 4.7 below). We prove Theorem 4.14 only for NR$_n(w)$, the proof for GRG$_n(w)$ and CL$_n(w)$ being similar. The required upper bound $|\mathcal{C}_{\max}/n \leq \zeta_n + o(1)$ follows by the local convergence in probability in Theorem 3.11 and (2.5.6). This provides the required upper bound on $|\mathcal{C}_{\max}|$.

For the lower bound, we bound NR$_n(w)$ from below by a random graph with edge probabilities

$$p_{ij}^{(K)} = 1 - e^{-(w_i \wedge K)(w_j \wedge K)/\ell_n}. \tag{4.2.55}$$

Therefore, also $|\mathcal{C}_{\max}| \geq |\mathcal{C}_{\max}^{(K)}|$, where $\mathcal{C}_{\max}^{(K)}$ is the largest connected component in the inhomogeneous random graph with edge probabilities $(p_{ij}^{(K)})_{i,j \in [n]}$. Let

$$w_i^{(K)} = (w_i \wedge K) \frac{1}{\ell_n} \sum_{j \in [n]} (w_j \wedge K), \tag{4.2.56}$$

so that the edge probabilities in (4.2.55) correspond to the Norros-Reittu model with weights $(w_i^{(K)})_{i \in [n]}$. It is not hard to see that when Condition 1.1(a) holds for $(w_i)_{i \in [n]}$, then Conditions 1.1(a)-(c) hold for $(w_i^{(K)})_{i \in [n]}$, where the limiting random variable equals $(W \wedge K)$. Therefore, Theorem 4.13 applies to $(w_i^{(K)})_{i \in [n]}$. We deduce that $|\mathcal{C}_{\max}^{(K)}/n \xrightarrow{\mathbb{P}} \zeta^{(K)}$, which is the survival probability of the two-stage mixed-Poisson branching process with mixing variable $(W \wedge K)$. Since $\zeta^{(K)} \rightarrow \zeta$ when $K \rightarrow \infty$, we conclude that $|\mathcal{C}_{\max}/n \xrightarrow{\mathbb{P}} \zeta$. \qed

4.3 Connectivity of CM$_n(d)$

Assume that $\mathbb{P}(D = 2) < 1$. By Theorem 4.4, we see that $|\mathcal{C}_{\max}/n \xrightarrow{\mathbb{P}} 1$ when $\mathbb{P}(D \geq 2) = 1$, as in this case the survival probability equals 1. In this section, we investigate conditions under which CM$_n(d)$ is whp connected, i.e., $\mathcal{C}_{\max} = [n]$ and $|\mathcal{C}_{\max}| = n$.

Our main result shows that CM$_n(d)$ is whp connected when $d_{\min} = \min_{i \in [n]} d_i \geq$
3. Interestingly enough, we need very few conditions for this result, and do not even need Conditions 1.5(a)-(b):

**Theorem 4.15 (Connectivity of CM\(_n(d)\))** Assume that Conditions 1.5(a)-(b) hold. Further, assume that \(d_i \geq 3\) for every \(i \in [n]\). Then CM\(_n(d)\) is connected whp. More precisely,

\[
P(\text{CM}_n(d) \text{ disconnected}) = o(1).
\]

When Condition 1.5(a) holds with \(p_1 = p_2 = 0\), then \(\nu \geq 2 > 1\) is immediate, so we are always in the supercritical regime. Also, \(\zeta = 1\) when \(p_1 = p_2 = 0\), since survival of the unimodular Galton-Watson tree occurs with probability 1. Therefore, Theorem 4.4 implies that the largest connected component has size \(n(1 + o_p(1))\) when Condition 1.5(a)-(b) hold. Theorem 4.15 extends this to the statement that CM\(_n(d)\) is with high probability connected. However, we do not assume that Condition 1.5 holds here.

We note that Theorem 4.15 yields an important difference between the generalized random graph and the configuration model, also from a practical point of view. Indeed, for the generalized random graph to be whp connected, the degrees must tend to infinity. This has already been observed for ER\(_n(p)\) in [Volume 1, Theorem 5.8]. For the configuration model, it is possible that the graph is connected while the average degree is bounded. Many real-world networks are connected, which makes the configuration model often more suitable than inhomogeneous random graphs.

**Proof** The proof is based on a simple counting argument. We recall that a configuration denotes a pairing of all the half-edges. We note that the probability of a configuration equals \(1/(\ell_n - 1)!!\). On the event that CM\(_n(d)\) is disconnected, there exists a set of vertices \(\mathcal{I} \subset [n]\) with \(|\mathcal{I}| \leq \lfloor n/2 \rfloor\) such that all half-edges incident to vertices in \(\mathcal{I}\) are only paired to other half-edges incident to other vertices in \(\mathcal{I}\). For \(\mathcal{I} \subseteq [n]\), we let

\[
\ell_n(\mathcal{I}) = \sum_{i \in \mathcal{I}} d_i.
\]

(4.3.2)

Since \(d_i \geq 3\), we can use Theorem 4.4 to conclude that most edges are in \(\mathcal{C}_{\text{max}}\), and \(\mathcal{I} \neq \mathcal{C}_{\text{max}}\). Therefore, \(\ell_n(\mathcal{I}) = o(\ell_n) = o(n)\), and we may, without loss of generality, assume that \(\ell_n(\mathcal{I}) \leq \ell_n/4\). We denote the event that there exists a connected component \(\mathcal{I}\) consisting of \(|\mathcal{I}| \leq \lfloor n/2 \rfloor\) vertices for which the sum of degrees is at most \(\ell_n(\mathcal{I}) \leq \ell_n/4\) by \(\mathcal{E}\).

Clearly, in order for the half-edges incident to vertices in \(\mathcal{I}\) to be paired only to other half-edges incident to vertices in \(\mathcal{I}\), \(\ell_n(\mathcal{I})\) needs to be even. The number of configurations for which this happens is bounded above by

\[
(\ell_n(\mathcal{I}) - 1)!!(\ell_n(\mathcal{I}^c) - 1)!!.
\]

(4.3.3)
As a result,
\[
\mathbb{P}(\text{CM}_n(d) \text{ disconnected}; \mathcal{E}) \leq \sum_{\mathcal{I} \subset [n]} \frac{(\ell_n(\mathcal{I}) - 1)!!(\ell_n(\mathcal{I}^c) - 1)!!}{(\ell_n - 1)!!} \quad (4.3.4)
\]
\[
= \sum_{\mathcal{I} \subset [n]} \ell_n(\mathcal{I})^2 \frac{\ell_n(\mathcal{I}) - 2j + 1}{\ell_n - 2j + 1},
\]
where the sum over \( \mathcal{I} \subset [n] \) is restricted to \( \mathcal{I} \) for which \( |\mathcal{I}| \leq \lfloor n/2 \rfloor \) and \( \ell_n(\mathcal{I}) \leq \ell_n/4 \). In Exercise 4.12, you will use (4.3.4) to give a bound on the probability of the existence of an isolated vertex (i.e., a vertex with only self-loops).

Define
\[
f(x) = \prod_{j=1}^{x} \frac{2x - 2j + 1}{\ell_n - 2j + 1}. \quad (4.3.5)
\]

We can rewrite
\[
f(x) = \prod_{j=1}^{x} \frac{2x - 2j + 1}{\ell_n - 2j + 1} = \prod_{i=0}^{x-1} \frac{2i + 1}{\ell_n - 2i - 1}, \quad (4.3.6)
\]
where we write \( i = x - j \) and \( k = j - 1 \) in the second equality. Thus, for \( x \leq \ell_n/4 \), \( x \mapsto f(x) \) is decreasing, since
\[
\frac{f(x+1)}{f(x)} = \frac{2x+1}{\ell_n - 2x - 1} \leq 1. \quad (4.3.7)
\]
Now, for every \( \mathcal{I} \), since \( d_i \geq 3 \) for every \( i \in [n] \) and since \( \ell_n(\mathcal{I}) \leq \ell_n/4 \) is even,
\[
\ell_n(\mathcal{I}) \geq 2 \lceil 3|\mathcal{I}|/2 \rceil, \quad (4.3.8)
\]
which only depends on the number of vertices in \( \mathcal{I} \). Since there are precisely \( \binom{n}{m} \) ways of choosing \( m \) vertices out of \([n]\), we conclude that
\[
\mathbb{P}(\text{CM}_n(d) \text{ disconnected}) \leq \sum_{\mathcal{I} \subset [n]} f(\lceil 3|\mathcal{I}|/2 \rceil) = \sum_{m=1}^{\lfloor n/2 \rfloor} \binom{n}{m} f(\lceil 3m/2 \rceil), \quad (4.3.9)
\]
with \( m = |\mathcal{I}| \).

Note that, for \( m \) odd,
\[
\frac{f(3m+1)/2)}{f(3m+1)} = \frac{f((3m+1)/2+1)}{f((3m+1)/2)} = \frac{3m+3}{\ell_n - 3m - 2}. \quad (4.3.10)
\]
while, for \( m \) even,
\[
\frac{f(3m+2)/2)}{f(3m+2)} = \frac{f(3m+2)}{f(3m+2)} = \frac{3m+5}{\ell_n - 3m - 5}. \quad (4.3.11)
\]
Define
\[
h_n(m) = \binom{n}{m} f(\lceil 3m/2 \rceil), \quad (4.3.12)
\]
so that
\[ P(\text{CM}_n(d) \text{ disconnected}) \leq \sum_{m=1}^{\lfloor n/2 \rfloor} h_n(m). \quad (4.3.13) \]

Then,
\[ \frac{h_n(m+1)}{h_n(m)} = \frac{n-m}{m+1} f\left(\left\lceil \frac{3(m+1)}{2} \right\rceil\right), \quad (4.3.14) \]
so that, for \( m \) odd and using \( \ell_n \geq 3n \),
\[ \frac{h_n(m+1)}{h_n(m)} = \frac{3(n-m)}{\ell_n - 3m - 3} \leq \frac{n-m}{n-m-1}. \quad (4.3.15) \]
while, for \( m \) even and using \( \ell_n \geq 3n \),
\[ \frac{h_n(m+1)}{h_n(m)} = \frac{3(n-m)}{\ell_n - 3m - 3} \leq \frac{n-m}{n-m-1} n - m - 2. \quad (4.3.16) \]
Thus, we obtain that, for \( m \leq n/2 \), there exists a \( c > 0 \) such that
\[ \frac{h_n(m+1)}{h_n(m)} \leq 1 + \frac{c}{n}. \quad (4.3.17) \]
We conclude that, for \( m \leq n/2 \) such that \( m \geq 3 \),
\[ h_n(m) = h_n(3) \prod_{j=3}^{m} \frac{h_n(j+1)}{h_n(j)} \leq h_n(3) \prod_{j=3}^{\lfloor n/2 \rfloor} \left(1 + \frac{c}{n}\right) \]
\[ \leq h_n(3) \left(1 + \frac{c}{n}\right)^{\lfloor n/2 \rfloor} \leq h_n(3)e^{c/2}, \quad (4.3.18) \]
so that
\[ P(\text{CM}_n(d) \text{ disconnected}; \mathcal{E}) \leq \sum_{m=1}^{\ell_n} h_n(m) \leq h_n(1) + h_n(2) + \sum_{m=3}^{\lfloor n/2 \rfloor} h_n(m) \]
\[ \leq h_n(1) + h_n(2) + nh_n(3)e^{c/2}/2. \quad (4.3.19) \]
By Exercises 4.12 and 4.14, \( h_n(1), h_n(2) = O(1/n) \), so we are left to compute \( h_n(3) \). For this, we note that \( \lfloor 3m/2 \rfloor = 5 \) when \( m = 3 \), so that
\[ h_n(3) = \binom{n}{3} f(5) = \frac{9!!n(n-1)(n-2)}{6(\ell_n-1)(\ell_n-3)(\ell_n-5)(\ell_n-7)(\ell_n-9)} = O(1/n^2), \quad (4.3.20) \]
so that \( nh_n(3) = O(1/n) \). We conclude that
\[ P(\text{CM}_n(d) \text{ disconnected}; \mathcal{E}) = O(1/n). \quad (4.3.21) \]
This is even stronger than required, and one would be tempted to believe that also \( P(\text{CM}_n(d) \text{ disconnected}) = O(1/n) \), but this proof does not show this since we start with the assumption that the non-giant (collection of) component(s) \( \mathcal{I} \) satisfies \( \ell_n(\mathcal{I}) \leq \ell_n/4 \).
The above proof is remarkably simple, and requires very little of the precise degree distribution except for $d_{\min} \geq 3$. In the sequel, we investigate what happens when this condition fails.

We continue by showing that $\text{CM}_n(d)$ is with positive probability disconnected when either $n_1 \gg n^{1/2}$:

**Proposition 4.16** (Disconnectivity of $\text{CM}_n(d)$ when $n_1 \gg n^{1/2}$) Let Condition 1.5(a)-(b) hold, and assume that $n_1 \gg n^{1/2}$. Then,

$$\lim_{n \to \infty} \mathbb{P}(\text{CM}_n(d) \text{ connected}) = 0.$$ (4.3.22)

**Proof** We note that $\text{CM}_n(d)$ is disconnected when there are two vertices of degree 1 whose half-edges are paired to each other. When the half-edges of two vertices of degree 1 are paired to each other, we say that a 2-pair is created. Then, since after $i$ pairings of degree-1 vertices to higher-degree vertices, there are $\ell_n - n_1 - i + 1$ half-edges incident to higher-degree vertices, out of a total of $\ell_n - 2i + 1$ unpaired half-edges,

$$\mathbb{P}(\text{CM}_n(d) \text{ contains no 2-pair}) = \prod_{i=1}^{n_1} \frac{\ell_n - n_1 - i + 1}{\ell_n - 2i + 1} = \prod_{i=1}^{n_1} \left(1 - \frac{n_1 - i}{\ell_n - 2i + 1}\right).$$ (4.3.23)

Since, for each $i \geq 1$,

$$1 - \frac{n_1 - i}{\ell_n - 2i + 1} \leq 1 - \frac{n_1 - i}{\ell_n} \leq e^{-(n_1-i)/\ell_n},$$ (4.3.24)

we arrive at

$$\mathbb{P}(\text{CM}_n(d) \text{ contains no 2-pair}) \leq \prod_{i=1}^{n_1} e^{-(n_1-i)/\ell_n} = e^{-n_1(n_1-1)/(2\ell_n)} = o(1),$$ (4.3.25)

since $\ell_n = \Theta(n)$ and $n_1 \gg n^{1/2}$. \hfill \Box

**Proposition 4.17** (Disconnectivity of $\text{CM}_n(d)$ when $p_2 > 0$) Let Conditions 1.5(a)-(b) hold, and assume that $p_2 > 0$. Then,

$$\limsup_{n \to \infty} \mathbb{P}(\text{CM}_n(d) \text{ connected}) < 1.$$ (4.3.26)

**Proof** We perform a second moment method on the number $P(2)$ of connected components consisting of two vertices of degree 2. The expected number of such components equals

$$\mathbb{E}[P(2)] = \frac{2n_2(n_2-1)}{2(\ell_n-1)(\ell_n-3)},$$ (4.3.27)

since there are $n_2(n_2-1)/2$ pairs of vertices of degree 2, and the probability
4.3 Connectivity of CM$_n(d)$

that a fixed pair forms a connected component is equal to $2/(\ell_n-1)(\ell_n-3)$. By Condition 1.5(a)-(b), which implies that $n_2/n \to p_2$,

$$E[P(2)] \to p_2^2/E[D]^2 \equiv \lambda_2.$$  

(4.3.28)

By assumption, $p_2 > 0$, so that also $\lambda_2 > 0$. We can use Theorem 2.6 to show that $P(2) \xrightarrow{d} \text{Poi}(\lambda_2)$, so that

$$P(CM_n(d) \text{ disconnected}) \geq P(P(2) > 0) \to 1 - e^{-\lambda_2} > 0,$$

(4.3.29)

as required. The proof that Theorem 2.6 can be applied is left as Exercise 4.11 below.

We close this section with a detailed result on the size of the giant component when $d_{\text{min}} \geq 2$:

**Theorem 4.18** (Connectivity of CM$_n(d)$ when $p_1 = 0$)  Let Conditions 1.5(a)-(b) hold, and assume that $p_2 \in (0,1)$. Assume that $d_i \geq 2$ for every $i \in [n]$. Then

$$n - |\mathcal{C}_{\text{max}}| \xrightarrow{d} \sum_{k \geq 2} kX_k,$$

(4.3.30)

where $(X_k)_{k \geq 2}$ are independent Poisson random variables with parameters $\lambda_2^{k}/(2k)$ with $\lambda_2 = 2p_2/E[D]$. Consequently,

$$P(CM_n(d) \text{ connected}) \to e^{-\sum_{k \geq 2} \lambda_2^{k}/(2k)} \in (0,1).$$

(4.3.31)

Rather than giving the complete proof of Theorem 4.18, we give a sketch of it:

**Sketch of proof of Theorem 4.18.** Let $P(k)$ denote the number of $k$-cycles consisting of degree 2 vertices, for $k \geq 2$. Obviously, every vertex in such a cycle is not part of the giant component, so that

$$n - |\mathcal{C}_{\text{max}}| \geq \sum_{k \geq 2} kP(k).$$

(4.3.32)

A multivariate moment method allows one to prove that $(P(k))_{k \geq 2} \xrightarrow{d} (X_k)_{k \geq 2}$, where $(X_k)_{k \geq 2}$ are independent Poisson random variables with parameters $\lambda_2^{k}/(2k) = \lim_n E[P(k)]$. See also Exercise 4.15, where you are asked to prove this.

In order to complete the argument, two approaches are possible (and have been used in the literature). First, Federico and van der Hofstad (2017) use counting arguments to show that as soon as a connected component has at least one vertex $v$ of degree $d_v \geq 3$, then it is whp part of the giant component $\mathcal{C}_{\text{max}}$. This then proves that (4.3.32) is whp an equality. See also Exercise 4.16.

Alternatively, and more in the style of Luczak (1992), one can pair up all the half-edges incident to vertices of degree 2, and realize that the graph, after pairing of all the vertices of degree 2 is again a configuration model with an changed degree distribution. The cycles consisting of only degree 2 vertices will be removed, so that we only need to consider the contribution of pairing strings
of degree 2 vertices to vertices of degrees at least 3. If both ends of the string are each connected to two distinct vertices of degrees $d_s, d_t$ at least 3, then we can imagine this string to correspond to a single vertex of degree $d_s + d_t - 2 \geq 4$, which is sufficiently large. Unfortunately, it is also possible that the string of degree 2 vertices is connected to the same vertex $u$ of degree $d_u \geq 3$. When $d_u \geq 5$, then there will still be at least 3 remaining half-edges of $u$, which is fine. Thus, we only need to care about the case where we create a cycle of vertices of degree 2 with one vertex $u$ in it of degree $d_u = 3$ or $d_u = 4$, which will then correspond to vertices of degree 1 and 2. In Exercise 4.17, you are asked to prove that there is a bounded number of such cycles. Thus, it suffices to extend the proof of Theorem 4.15 to the setting where there are a bounded number of vertices of degrees 1 and 2. We can deal with the degree 2 vertices in the same way. Pairing the degree 1 vertices again leads to vertices of (extra) degree at least $3-1=2$, which are fine when the extra degree is at least 4, and otherwise can be dealt with as with the other degree 2 vertices. We refrain from giving more details.

4.4 Related results for the configuration model

In this section, we discuss related results on connected components for the configuration model. We start by discussing the subcritical behavior of the configuration model.

The largest subcritical cluster
When $\nu < 1$, so that in particular $\mathbb{E}[D^2] < \infty$, the largest connected component for $\text{CM}_n(d)$ is closely related to the largest degree:

**Theorem 4.19** (Subcritical phase for $\text{CM}_n(d)$) Let $d$ satisfy Condition 1.5(a)-(c) with $\nu = \mathbb{E}[D(D-1)]/\mathbb{E}[D] < 1$. Suppose further that there exists $\tau > 3$ and $c_2 > 0$ such that

$$[1 - F_n](x) \leq c_2 x^{-(\tau - 1)}.$$  \hspace{1cm} (4.4.1)

Then, for $\text{CM}_n(d)$ with $d_{\text{max}} = \max_{j \in [n]} d_j$,

$$|C_{\text{max}}| = \frac{d_{\text{max}}}{1 - \nu} + o_n(n^{1/(\tau - 1)}).$$  \hspace{1cm} (4.4.2)

Theorem 4.19 is closely related to Theorem 3.19. In fact, we can use Theorem 4.19 to prove Theorem 3.19, see Exercise 4.18. Note that the result in (4.4.2) is most interesting when $d_{\text{max}} = \Theta(n^{1/(\tau - 1)})$, as it would be in the case when the degrees obey a power law with exponent $\tau$ (for example, in the case where the degrees are i.i.d.). When $d_{\text{max}} = o(n^{1/(\tau - 1)})$, instead, Theorem 4.19 implies that $|C_{\text{max}}| = o_n(n^{1/(\tau - 1)})$, a less precise result. Bear in mind that Theorem 4.19 only gives sharp asymptotics of $|C_{\text{max}}|$ when $d_{\text{max}} = \Theta(n^{1/(\tau - 1)})$ (see Exercise 4.19).

The intuition behind Theorem 4.19 is that from the vertex of maximal degree, there are $d_{\text{max}}$ half-edges that can reach more vertices. Since the random graph is
subcritical, one can use Theorem 4.1 to prove that the tree rooted at any half-edge incident to the vertex of maximal degree converges in distribution to a subcritical branching process. Further, the trees rooted at different half-edges are close to being independent. Thus, by the law of large numbers, one can expect that the total number of vertices in these \(d_{\text{max}}\) trees are close to \(d_{\text{max}}\) times the expected size of a single tree, which is \(1/(1 - \nu)\). This explains the result in Theorem 4.19. Part of this intuition is made precise in Exercises 4.21 and 4.22. Exercises 4.23 and 4.24 investigate conditions under which \(|C_{\text{max}}| = d_{\text{max}}/(1 - \nu)(1 + o_{\nu}(1))\) might or might not hold.

The near-critical supercritical behavior in the configuration model

Janson and Luczak (2009) also prove partial results on the near-critical behavior of \(CM_n(d)\), and these are further extended in van der Hofstad et al. (2016). We distinguish between the degrees having finite third-moment degrees, and the case where the degrees obey a power law with power-law exponent \(\tau \in (3, 4)\):

**Theorem 4.20** (Near-critical behavior \(CM_n(d)\) with finite third moments) Let \(d\) satisfy Condition 1.5(a)-(c) with \(\nu = E[D(D - 1)]/E[D] = 1\).

(a) Assume further that \(\alpha_n = \nu_n - 1 = E[D_n(D_n - 2)]/E[D_n] > 0\) is such that \(\alpha_n \gg n^{-1/3}\), and that

\[
E[D_{n}^{3+\varepsilon}] = O(1) \quad (4.4.3)
\]

for some \(\varepsilon > 0\). Let \(\beta = E[D(D - 1)(D - 2)]/E[D] > 0\). Then, \(CM_n(d)\) satisfies

\[
|C_{\text{max}}| = \frac{2}{\beta}E[D]n^\alpha_n + o(n^\alpha_n),
\]

\[
|v_k(C_{\text{max}})| = \frac{2E[D]}{\beta}k^\alpha_n + o(n^\alpha_n), \text{ for every } k \geq 0,
\]

\[
E(C_{\text{max}}) = \frac{2E[D]E[D^2]}{\beta}n^\alpha_n + o(n^\alpha_n),
\]

while \(|C_{(i)}| = o(n^\alpha_n)\) and \(E(C_{(i)}) = o(n^\alpha_n)\).

We next investigate the setting where the degrees do not have finite third-moments, which turns out to be quite different:

**Theorem 4.21** (Near-critical behavior \(CM_n(d)\) with infinite third moments) Let \(d\) satisfy Condition 1.5(a)-(c) with \(\nu = E[D(D - 1)]/E[D] = 1\). Let \(\zeta_n^*\) denote the survival probability of a branching process with offspring distribution \(D_n^* - 1\). Assume further that \(\alpha_n = \nu_n - 1 = E[D_n(D_n - 2)]/E[D_n] > 0\) is such that \(\alpha_n \gg n^{-1/3}(E[D_n]^2)^{2/3}\). Then, \(CM_n(d)\) satisfies

\[
|C_{\text{max}}| = E[D]\zeta_n^* n(1 + o_{\nu}(1)),
\]

\[
|v_k(C_{\text{max}})| = k^\alpha_n \zeta_n^* n(1 + o_{\nu}(1)), \text{ for every } k \geq 0,
\]

\[
E(C_{\text{max}}) = E[D]\zeta_n^* n(1 + o_{\nu}(1)),
\]

while \(|C_{(i)}| = o(n^\zeta_n^*)\) and \(E(C_{(i)}) = o(n^\zeta_n^*)\).
The asymptotics of $|C_{\text{max}}|$ in Theorem 4.20 can be understood by the fact that, for a branching process with offspring distribution $X$ having mean $E[X] = 1 + \varepsilon$ where $\varepsilon$ is small, the survival probability $\zeta$ satisfies $\zeta = 2\varepsilon/\text{Var}(X)(1 + o(1))$ (see Exercise 4.25). Therefore, the survival probability $\zeta^\ast$ of the branching process with offspring distribution $D^\ast - 1$ is close to $2\varepsilon/\beta$, where we note that $\beta = \text{Var}(D^\ast - 1) = \text{Var}(D^\ast)$. Since the limit of $|C_{\text{max}}|/n \zeta$ satisfies
\begin{equation}
\zeta = \sum_{k=1}^{\infty} p_k (1 - (1 - \zeta^\ast)^k),
\end{equation}
we further obtain that
\begin{equation}
\zeta = \zeta^\ast E[D](1 + o(1)).
\end{equation}
The results on $|v_k(C_{\text{max}})|$ and $|E(C_{\text{max}})|$ can be understood in a similar way. Exercises 4.25–4.26 investigate the asymptotics of survival probabilities under various tail and moment assumptions. The results in Theorem 4.21 are far more general, and Theorem 4.21 implies Theorem 4.20 (see Exercise 4.27).

In the case where the degrees obey a power law, for example when $1 - F_n(x) = \Theta(x^{-(\tau-1)})$ for all $x \leq n^{1/(\tau-1)}$, it can be seen that $\zeta^\ast n = \Theta(\alpha n^{1/(\tau-3)})$ (recall also Exercise 4.26), and the restriction on $\alpha_n$ becomes $\alpha_n \gg n^{-(\tau-3)/(\tau-1)}$ (see Exercise 4.28).

**The critical behavior in the configuration model**

In this section, we study the critical behavior of the configuration model. The critical phase can be characterized by the largest connected components being random, and often scaling limits are taken to identify their limiting distribution. We focus on the critical case of $\text{CM}_n(d)$ for i.i.d. degrees:

**Theorem 4.22** (Weak convergence of the ordered critical clusters: finite third moments) Let $d = (d_i)_{i \in [n]}$ be a sequence of i.i.d. random variables having the same distribution as $D$ satisfying $\nu = E[D(D-1)]/E[D] = 1$. Let $(|C_{(i)}|)_{i \geq 1}$ denote the clusters of $\text{CM}_n(d)$, ordered in size. Let $E[D^3] < \infty$. Then, as $n \to \infty$,
\begin{equation}
(n^{-2/3}|C_{(i)}|)_{i \geq 1} \xrightarrow{d} (\gamma_i)_{i \geq 1},
\end{equation}
in the product topology, for some non-degenerate limit $(\gamma_i)_{i \geq 1}$.

Theorem 4.22 is reminiscent of [Volume 1, Theorem 5.7] for the Erdős-Renyi random graph. In fact, it even turns out that the scaling limits $(\gamma_i)_{i \geq 1}$ are closely related to the scaling limits of critical connected components in Erdős-Renyi random graph. Thus, one can say that Theorem 4.22 describes the setting of weak inhomogeneity. We next study the case of strong inhomogeneity, which corresponds to degrees having infinite third moment:

**Theorem 4.23** (Weak convergence of the ordered critical clusters) Let $d = (d_i)_{i \in [n]}$ be a sequence of i.i.d. random variables having the same distribution as $D$ satisfying $\nu = E[D(D-1)]/E[D] = 1$. Let $(|C_{(i)}|)_{i \geq 1}$ denote the clusters of
CM\(_n(d)\), ordered in size. Let the distribution function \( F \) of \( D \) satisfy that there exists a \( \tau \in (3, 4) \) and \( 0 < c_\tau < \infty \) such that

\[
\lim_{x \to \infty} x^{\tau - 1} [1 - F(x)] = c_\tau.
\] (4.4.7)

Then, as \( n \to \infty \),

\[
(n^{-\tau/(\tau-1)} |C_i|)_{i \geq 1} \xrightarrow{d} (\gamma_i)_{i \geq 1},
\] (4.4.8)

in the product topology, for some non-degenerate limit \((\gamma_i)_{i \geq 1}\).

As mentioned before, Theorem 4.22 describes the setting where the effect of large degrees is negligible. When \( \mathbb{E}[D^3] = \infty \), on the other hand, the critical scaling changes rather dramatically, and the largest critical cluster has size \( n^\rho \), where \( \rho = (\tau - 2)/(\tau - 1) \in (1/2, 2/3) \). Thus, in the presence of high-degree vertices, critical connected components become smaller. This is similar to the fact that the near-critical branching process survival probability becomes smaller for heavy-tailed offspring distributions compared to light-tailed offspring distributions (compare Exercises 4.25 and 4.26).

There are many results extending Theorems 4.22–4.23 to fixed degrees, under similar (but stronger) assumptions as in Condition 1.5(a)-(c). We refer to Section 4.5 for an extensive overview of the literature.

4.5 NOTES AND DISCUSSION

**Notes on Section 4.1**

Theorem 4.1 is a classical result for the configuration model, and has appeared in various guises throughout the literature. For example, Dembo and Montanari (2010b) crucially rely on it in order to identify the limiting pressure for the Ising model on the configuration model.

**Notes on Section 4.2**

This section is adapted from Janson and Luczak (2009), which, in turn, generalizes the results by Molloy and Reed (1995, 1998). The results by Molloy and Reed (1995, 1998) are not phrased in terms of branching processes, which makes them a bit more difficult to grasp. We have chosen to reformulate the results using branching process terminology. We also refer to Bollabás and Riordan (2015), who give an alternative proof using branching process approximations on the exploration of the giant component. They also provide the extension of Theorem 4.11 that we explain there, by showing that the probability of a deviation of order \( \varepsilon n \) of \( v_k(C_{\text{max}}) \) is exponentially small for the configuration model. Since the probability of simplicity in CM\(_n(d)\) for power-law degrees with infinite variance is not exponentially small, this implies the stated extension in Theorem 4.12.
Notes on Section 4.3
These results are folklore. A version of Theorem 4.15 can be found in (Chatterjee and Durrett, 2009, Lemma 1.2). We could not find the precise version stated in Theorem 4.15. Theorem 4.18 is proved by Federico and van der Hofstad (2017). This paper also allows for a number of vertices $n_1$ of degree 1 satisfying $n_1 = \rho \sqrt{n}$. Earlier versions include the results by Luczak (1992) for $d_{\text{min}} \geq 2$ and Wormald (1981), who proved $r$-connectivity when $d_{\text{min}} = r$.

Notes on Section 4.4
Theorem 4.19 is proved by Janson (2008). (Janson, 2008, Theorem 1.1) shows that $|C_{\text{max}}| \leq An^{1/(\tau - 1)}$ when the power-law upper bound in (4.4.1) holds, while (Janson, 2008, Theorem 1.1) gives the asymptotic statement. Further, Janson (2008) remarks that the $j$th largest cluster has size $d_{(j)}/(1 - \nu) + o(n^{1/(\tau - 1)})$.

Theorem 4.20 was first proved under a finite $(4+\varepsilon)$-moment condition in Janson and Luczak (2009). It is improved to the case where $E[D_n^4] \to E[D^4]$ by van der Hofstad et al. (2016). Interestingly, the behavior is different when Conditions 1.5(a)-(b) hold, but $E[D_n^4] \to \infty$ sufficiently fast. Also this case is studied by van der Hofstad et al. (2016).

Theorems 4.22–4.23 are proved by Joseph (2014), who focuses on configuration models with i.i.d. degrees. For related (weaker but more robust) results in the case of fixed degrees satisfying an assumption as in Condition 1.5, see Hatami and Molloy (2010). There is a lot of work on scaling limits for critical configuration models (or the related problem of critical percolation on configuration models, which can be related to critical configuration models, see Janson (2009a)), and we now give some links to the literature. Nachmias and Peres (2010) studies critical percolation on random regular graphs. Riordan (2012) studies the critical behavior of configuration models with bounded degrees, while Dhara et al. (2017) extends these results to the (necessary) finite third-moment assumption. Dhara et al. (2016) shows that different scaling arises when the degrees have power-law tails with infinite third-moment degrees. Interestingly, these results are quite different from those in the setting of i.i.d. degrees studied by Joseph (2014). For an extensive overview of the literature, we refer to van der (Hofstad, 2018+, Chapter 4).

4.6 Exercises for Chapter 4

Exercise 4.1 (Convergence of $n$-dependent branching process) Assume that Conditions 1.5(a)-(b) hold. Prove that $D_n^* \xrightarrow{d} D^*$, and conclude that $\text{BP}_n(t) \xrightarrow{d} \text{BP}(t)$ for every $t$ finite.

Exercise 4.2 (Proof of no-overlap property in (4.1.19)) Assume that the conditions in Theorem 4.1 hold. Prove that $P(\ell_k^n(o_1) \approx t, o_2 \in B_k^{\infty}(o_1)) \to 0$, and conclude that (4.1.19) holds.

Exercise 4.3 (Cluster size of vertex 1 in a 2-regular graph) Let $n_2 = n$, and
4.6 Exercises for Chapter 4

let $C(1)$ denote the cluster size of vertex 1. Show that

$$|C(1)|/n \xrightarrow{d} T,$$

where $P(T \leq x) = 1 - \sqrt{1 - x}$.

**Exercise 4.4** (Cluster size in a 2-regular graph with some degree-1 vertices)

Let $n_1 \to \infty$ with $n_1/n \to 0$, and $n_2 = n - n_1$. Let $C(1)$ denote the cluster size of vertex 1. Show that

$$|C(1)|/n \xrightarrow{p} 0.$$ (4.6.2)

**Exercise 4.5** (Cluster size in a 2-regular graph with some degree-4 vertices)

Let $n_4 \to \infty$ with $n_4/n \to 0$, and $n_2 = n - n_4$. Let $C(1)$ denote the cluster size of vertex 1. Show that

$$|C(1)|/n \xrightarrow{p} 1.$$ (4.6.3)

**Exercise 4.6** (Limiting constants)  
Prove that for $t_1 = 0$ and $t_2 = \theta$, $e^{-kt_1} = 1$, $e^{-kt_2} = \xi$, $G_0(e^{-t_1}) = 1$, $G_0(e^{-t_2}) = 1 - \zeta$, $h(e^{-t_1}) = 2\mathbb{E}[D]$, $h(e^{-t_2}) = 2\mathbb{E}[D]\xi^2$.

**Exercise 4.7** (Theorem 4.4 for GRG$_n(w)$ with $\nu = \infty$) 
Prove that all statements in Theorem 4.4 hold for the GRG$_n(w)$ in Theorem 4.14 when $\nu = \infty$.

**Exercise 4.8** (Number of vertices with degree $k$) 
Let $w$ satisfy Conditions 1.1(a)-(b). Adapt the proof of $|C_{\text{max}}|/n \xrightarrow{p} \zeta$ to show that also $v_k(C_{\text{max}})/n \xrightarrow{p} p_k(1 - \xi^k)$ for NR$_n(w)$.

**Exercise 4.9** (Phase transition in UG$_n(d)$ for $\nu = \infty$) 
Combine (4.2.54) and (4.2.52)–(4.2.53) to complete the proof of Theorem 4.12.

**Exercise 4.10** (A lower bound on the probability of simplicity) 
Prove (4.2.54) under the assumption that Conditions 1.5(a)-(b) hold, or look the proof up in (Bollobás and Riordan, 2015, Lemma 21).

**Exercise 4.11** (Factorial moments of $P(2)$) 
Let Conditions 1.5(a)-(b) hold, and assume that $p_2 > 0$. Prove that, for every $k \geq 1$ and with $\lambda_2 = 2p_2^2/\mathbb{E}[D]^2$,

$$\mathbb{E}[(P(2))^k] \to \lambda_2^k.$$ (4.6.4)

Conclude that $P(2) \xrightarrow{d} \text{Poi}(\lambda_2)$.

**Exercise 4.12** (Isolated vertex) 
Use (4.3.4) to show that, when $d_i \geq 3$ for all $i \in [n]$,

$$P(\text{there exists an isolated vertex}) \leq \frac{3n}{(2\ell_n - 1)(2\ell_n - 3)}.$$ (4.6.5)

**Exercise 4.13** (Isolated vertex (Cont.)) 
Use (4.3.9) to reprove Exercise 4.12. Hence, the above bound is quite sharp.
Exercise 4.14 (A cluster of size two) Use (4.3.9) to prove that, when $d_i \geq 3$ for all $i \in [n], \Pr(\text{there exists a cluster of size 2}) \leq \frac{15n(n-1)}{(2\ell_n-1)(2\ell_n-3)(2\ell_n-5)}$. (4.6.6)

Exercise 4.15 (Cycles in CM$_n$(d)) Let $P(k)$ denote the number of $k$-cycles consisting of degree 2 vertices, for $k \geq 2$. Let $\lambda_j = 2p_j/\mathbb{E}[D]$. Use the multivariate moment method to prove that $(P(k))_{k \geq 2} \xrightarrow{d} (X_k)_{k \geq 2}$, where $(X_k)_{k \geq 2}$ are independent Poisson random variables with parameters $\lambda_j^2/(2k) = \lim_n \mathbb{E}[P(k)]$.

Exercise 4.16 ($C_{\max}$ when $d_{\min} = 2$) Consider CM$_n$(d) with $d_{\min} = 2$ and assume that $\Pr(D \geq 3) > 0$. Show that Theorem 4.18 holds if $\Pr(\exists v: d_v \geq 3$ and $v \notin C_{\max}) = o(1)$.

Exercise 4.17 (Cycles of degree 2 vertices with one other vertex) Let Conditions 1.5(a)-(b) hold, and suppose that $d_{\min} \geq 2$. Show that the expected number of cycles consisting of vertices of degree 2 except for one vertex of degree $k$ converges to $\frac{k(k-1)}{2\mathbb{E}[D]^2} \sum_{r \geq 1} (2p_2/\mathbb{E}[D])^r$.

Exercise 4.18 (Proof of Theorem 3.19) Use Theorem 4.19 and Theorem 1.7 to prove Theorem 3.19.

Exercise 4.19 (Sharp asymptotics in Theorem 4.19) Prove that $|C_{\max}| = d_{\max}/(1 - \nu)(1 + o_\nu(1))$ precisely when $d_{\max} = \Theta(n^{1/(\tau-1)})$.

Exercise 4.20 (Sub-polynomial subcritical clusters) Use Theorem 4.19 to prove that $|C_{\max}| = o_\nu(n^\epsilon)$ for every $\epsilon > 0$ when (4.4.1) holds for every $\tau > 1$ (where the constant $c_2$ is allowed to depend on $\tau$). Thus, when the maximal degree is sub-polynomial, then also the maximal connected component is.

Exercise 4.21 (Single tree asymptotics in Theorem 4.19) Assume that the conditions in Theorem 4.19 hold. Use Theorem 4.1 to prove that the tree rooted at any half-edge incident to the vertex of maximal degree converges in distribution to a subcritical branching process with expected total progeny $1/(1 - \nu)$.

Exercise 4.22 (Two-tree asymptotics in Theorem 4.19) Assume that the conditions in Theorem 4.19 hold. Use the local weak convergence in Theorem 4.1 to prove that the two trees rooted at any pair of half-edges incident to the vertex of maximal degree jointly converge in distribution to two independent subcritical branching processes with expected total progeny $1/(1 - \nu)$.

Exercise 4.23 (Theorem 4.19 when $d_{\max} = o(\log n)$) Assume that the conditions in Theorem 4.19 hold, so that $\nu < 1$. Suppose that $d_{\max} = o(\log n)$. Do you expect $|C_{\max}| = d_{\max}/(1 - \nu)(1 + o_\nu(1))$ to hold?

Exercise 4.24 (Theorem 4.19 when $d_{\max} \gg \log n$) Assume that the conditions in Theorem 4.19 hold, so that $\nu < 1$. Suppose that $d_{\max} \gg \log n$. Do you expect $|C_{\max}| = d_{\max}/(1 - \nu)(1 + o_\nu(1))$ to hold?
4.6 Exercises for Chapter 4

Exercise 4.25 (Survival probability of finite-variance branching process) Let $X$ be the offspring distribution of a branching process with finite variance. Show that its survival probability $\zeta = \zeta(\varepsilon)$ satisfies $\zeta(\varepsilon) = 2\varepsilon/\text{Var}(X)(1 + o(1))$. What does this imply for a unimodular branching process with finite third-moment root offspring distribution?

Exercise 4.26 (Survival probability of infinite-variance branching process) Let $X$ be the offspring distribution of a branching process with infinite variance and tail asymptotics $\mathbb{P}(X > x) = cx^{-(\tau-2)}(1 + o(1))$ for some $\tau \in (2, 3)$. Show that its survival probability $\zeta = \zeta(\varepsilon)$ satisfies $\zeta(\varepsilon) = \Theta(\varepsilon^{1/(\tau-3)})$. What does this imply for a unimodular branching process with root offspring distribution $X$ that satisfies $\mathbb{P}(X > x) = cx^{-(\tau-1)}(1 + o(1))$ for some $\tau \in (2, 3)$?

Exercise 4.27 (Relation Theorems 4.20 and 4.21) Show that Theorem 4.21 implies Theorem 4.20 when $\mathbb{E}[D_n^3] \to \mathbb{E}[D^3]$.

Exercise 4.28 (Near-critical configuration model with infinite third moment power-law degrees) Let the conditions in Theorem 4.21 hold. Assume that $1 - F_n(x) = \Theta(x^{-(\tau-1)})$ for all $x \leq n^{1/(\tau-1)}$. Recall that $\zeta_n^* = \Theta(\alpha_n^{1/(\tau-3)})$ and thus that $|C_{\max}| = \Theta(\alpha_n^{1/(\tau-3)}n)$. Show further that $\alpha_n \gg n^{-1/3}(\mathbb{E}[D_n^3])^{2/3}$ is equivalent to $\alpha_n \gg n^{-(\tau-3)/(\tau-1)}$. 

Chapter 5
CONNECTED COMPONENTS IN PREFERENTIAL ATTACHMENT MODELS

Abstract
In this chapter, we further investigate preferential attachment models. In Section 5.1 we start by discussing an important tool in this chapter: exchangeable random variables and their distribution described in De Finetti’s Theorem. We apply these results to Pólya urn schemes, which, in turn, we use to describe the distribution of the degrees in preferential attachment models. It turns out that Pólya urn schemes can also be used to describe the local limit of preferential attachment models.

Organization of this chapter
We start in Section 5.1 to discuss exchangeable random variables, and their fascinating properties. We continue in Section 5.2 with local weak convergence for preferential attachment models. In Section 5.4, we investigate the connectivity of PA\(^{(m, \delta)}\). Section 5.5 highlights some further results for preferential attachment models. We close this chapter in Section 5.6 with notes and discussion, and in Section 5.7 with exercises.

Throughout this chapter, we work with the preferential attachment model defined in Section 1.3.5 (and discussed in detail in [Volume 1, Chapter 8]) and denoted by (PA\(^{(m, \delta)}\))\(_{n \geq 1}\), unless stated otherwise. We recall that, for \(m = 1\), this model starts with a single vertex with 1 self-loops at time \(t = 1\) and at each time a vertex is added with 1 edges which are attached to the vertices in the graph with probabilities given in (1.3.55) for \(m = 1\). The model with \(m \geq 2\) is obtained by identifying blocks of \(m\) vertices in (PA\(^{1, \delta/m}\))\(_{n \geq 1}\). We sometimes also discuss other variants of the model, such as (PA\(^{m, \delta}\))\(_{n \geq 1}\), in which the \(m = 1\) model does not have any self-loops (recall (1.3.62)), so that the model is by default connected.

5.1 Exchangeable random variables and Pólya urn schemes

In this section, we discuss the distribution of infinite sequences of exchangeable random variables and their applications to Pólya urn schemes. We start by discussing De Finetti’s Theorem.

De Finetti’s Theorem for exchangeable random variables
We start by defining when sequences of random variables are exchangeable:

Definition 5.1 (Exchangeable random variables) A finite sequence of random variables \((X_i)_{i=1}^{n}\) is called exchangeable when the distribution of \((X_i)_{i=1}^{n}\) is the same as the one of \((X_{\sigma(i)})_{i=1}^{n}\) for any permutation \(\sigma: [n] \rightarrow [n]\). An infinite
sequence \((X_i)_{i \geq 1}\) is called exchangeable when \((X_i)_{i=1}^n\) is exchangeable for every \(n \geq 1\).

The notion of exchangeability is rather strong, and implies for example that the distribution of \(X_i\) is the same for every \(i\) (see Exercise 5.1), as well as that \((X_i, X_j)\) have the same distribution for every \(i \neq j\).

Clearly, when a sequence of random variables is i.i.d., then it is also exchangeable (see Exercise 5.2). A second example arises when we take a sequence of random variables that are i.i.d. conditionally on some random variables. An example could be a sequence of Bernoulli random variables that are i.i.d. conditionally on their success probability \(U\), but \(U\) itself is random. This is called a mixture of i.i.d. random variables. Remarkably, however, the distribution of an infinite sequence of exchangeable random variables is always such a mixture of i.i.d. random variables. This is the content of De Finetti’s Theorem, which we state and prove here in the case where \((X_i)_{i \geq 1}\) are indicator variables, which is the most relevant setting for our purposes:

**Theorem 5.2 (De Finetti’s Theorem)** Let \((X_i)_{i \geq 1}\) be an infinite sequence of exchangeable random variables, and assume that \(X_i \in \{0, 1\}\). Then there exists a random variable \(U\) with \(P(U \in [0, 1]) = 1\) such that, for all \(n \geq 1\) and \(1 \leq k \leq n\),

\[
P(X_1 = \cdots = X_k = 1, X_{k+1} = \cdots = X_n = 0) = E[U^k(1 - U)^{n-k}].
\]  

De Finetti’s Theorem (Theorem 5.2) states that an infinite exchangeable sequence of indicators has the same distribution as an independent Bernoulli sequence with a random success probability \(U\). Thus, the different elements of the sequence are not independent, but their dependence enters only through the random success probability \(U\). In particular, an infinite exchangeable sequence of indicators is a mixture of i.i.d. Bernoulli variables.

The proof of Theorem 5.2 can be relatively easily be extended to more general settings, for example, when \(X_i\) takes on at most a finite number of values. Since we only rely on Theorem 5.2 for indicator variables, we refrain from stating this more general version.

Define \(S_n\) to be the number of ones in \((X_i)_{i=1}^n\), i.e.,

\[
S_n = \sum_{k=1}^{n} X_k.
\]  

Then Theorem 5.2 is equivalent to the statement that

\[
P(S_n = k) = E\left[P(\text{Bin}(n, U) = k)\right].
\]  

You are asked to prove (5.1.3) in Exercise 5.4. Equation (5.1.3) also allows us to compute the distribution of \(U\). Indeed, when we would have that

\[
\lim_{n \to \infty} P(S_n \in (an, bn)) = \int_a^b f(u)du,
\]  

where
where \( f \) is a density, then (5.1.3) implies that \( f \) is in fact the density of the random variable \( U \). This is useful in applications of De Finetti’s Theorem (Theorem 5.2). Equation (5.1.4) follows by noting that \( S_n/n \overset{a.s.}{\rightarrow} U \) by the strong law of large numbers applied to the conditional law given \( U \). In Exercise 5.3, you can fill in the details.

**Proof of Theorem 5.2.** The proof makes use of Helly’s Theorem, which states that any sequence of bounded random variables has a weakly converging subsequence.

We fix \( m \geq n \) and condition on \( S_m \) to write

\[
P(X_1 = \cdots = X_k = 1, X_{k+1} = \cdots = X_n = 0) = \sum_{j=k}^{m} P(X_1 = \cdots = X_k = 1, X_{k+1} = \cdots = X_n = 0 | S_m = j) P(S_m = j).
\]  

(5.1.5)

By exchangeability and conditionally on \( S_m = j \), each sequence \((X_i)^{m}_{i=1}\) containing precisely \( j \) ones is equally likely. Since there are precisely \( \binom{m}{j} \) such sequences, and precisely \( \binom{m-n}{j-k} \) of them start with \( k \) ones and \( n-k \) zeros, we obtain

\[
P(X_1 = \cdots = X_k = 1, X_{k+1} = \cdots = X_n = 0 | S_m = j) = \frac{\binom{m-n}{j-k}}{\binom{m}{j}}.
\]  

(5.1.6)

Writing \((m)_k = m \cdot (m-1) \cdots (m-k+1)\) for the \( k \)th factorial moment of \( m \), we therefore arrive at

\[
P(X_1 = \cdots = X_k = 1, X_{k+1} = \cdots = X_n = 0) = \sum_{j=k}^{m} \frac{(j)_k (m-j)^{n-k}}{(m)_n} P(S_m = j).
\]  

(5.1.7)

When \( m \to \infty \) and for \( k \) and \( n \) with \( k \leq n \) fixed,

\[
\frac{(j)_k (m-j)^{n-k}}{(m)_n} = \left( \frac{j}{m} \right)^k \left( 1 - \frac{j}{m} \right)^{n-k} + o(1),
\]  

(5.1.8)

which can be seen by splitting between \( j > \varepsilon m \) and \( j \leq \varepsilon m \) for \( \varepsilon > 0 \) arbitrarily small. On the former, \( (j)_k = j^k (1 + o(1)) \), while on the latter, \( (j)_k \leq (\varepsilon m)^k \) and \((m-j)^{n-k}/(m)^n \leq m^{-k} \).

Recall that \( S_m = j \), so that

\[
P(X_1 = \cdots = X_k = 1, X_{k+1} = \cdots = X_n = 0) = \lim_{m \to \infty} \mathbb{E}[Y_m^k (1 - Y_m)^{n-k}],
\]  

(5.1.9)

where \( Y_m = S_m/m \). Note that it is here that we make use of the fact that \((X_i)^{\geq 1}\) is an infinite exchangeable sequence of random variables. Equation (5.1.9) is the point of departure for the completion of the proof.

We have that \( 0 \leq Y_m \leq 1 \) since \( 0 \leq S_m \leq m \), so that the sequence of random variables \((Y_m)^{m \geq 1}\) is a bounded sequence. By Helly’s Theorem, it contains a weakly converging subsequence, i.e., there exists a \((Y_{m_l})^{l \geq 1}\) with \( \lim_{l \to \infty} m_l = \infty \) and a random variable \( U \) such that \( Y_{m_l} \overset{d}{\rightarrow} U \). Since the random variable \( Y_m^k (1 -
Connected components in preferential attachment models

$Y_{m})^{n-k}$ is uniformly bounded for each $k,n$, Lebeque’s Dominated Convergence Theorem ([Volume 1, Theorem A.1]) gives that

$$
\lim_{m \to \infty} \mathbb{E}[Y_m^k (1 - Y_m)^{n-k}] = \lim_{l \to \infty} \mathbb{E}[Y_{m_l}^k (1 - Y_{m_l})^{n-k}] = \mathbb{E}[U^k (1 - U)^{n-k}].
$$

(5.1.10)

This completes the proof.

A careful reader may wonder about whether the above proof on the basis of subsequences indeed is enough. Indeed, it is possible that another subsequence $(Y_{m_l}')_{l \geq 1}$ with $\lim_{l \to \infty} m_l' = \infty$ has another limiting random variable $V$ such that $Y_{m_l}' \xrightarrow{d} V$. However, from (5.1.9) we then conclude that $\mathbb{E}[V^k (1 - V)^{n-k}] = \mathbb{E}[U^k (1 - U)^{n-k}]$ for every $k,n$. In particular, $\mathbb{E}[V^k] = \mathbb{E}[U^k]$ for every $k \geq 0$. Since the random variables $U, V$ are a.s. bounded by 1, and have the same moments, they also have the same distribution. We conclude that $Y_{m_l} \xrightarrow{d} U$ for every subsequence along which $Y_{m_l}$ converges, and this is equivalent to $Y_m \xrightarrow{d} U$. □

De Finetti’s Theorem implies that when $X_k$ and $X_n$ are coordinates of an infinite exchangeable sequence of indicators, then they are positively correlated, see Exercise 5.5. Thus, it is impossible for infinite exchangeable sequences of indicator variables to be negatively correlated, which is somewhat surprising.

In the proof of De Finetti’s Theorem, it is imperative that the sequence $(X_i)_{i \geq 1}$ is infinite. This is not mere a technicality of the proof. Rather, there are finite exchangeable sequences of random variables for which the equality (5.1.1) does not hold. Indeed, take an urn filled with $b$ blue and $r$ red balls and draw balls successively without replacement. Thus, the urn is sequentially being depleted, and it will be empty after the $(b+r)$th ball is drawn. Let $X_i$ denote the indicator that the $i$th ball drawn is blue. Then, clearly, the sequence $(X_i)_{i=1}^{r+b}$ is exchangeable. However,

$$
P(X_1 = X_2 = 1) = \frac{b(b-1)}{(b+r)(b+r-1)} > \left(\frac{b}{b+r}\right)^2 = P(X_1 = 1)P(X_2 = 1),
$$

(5.1.11)

so that $X_1$ and $X_2$ are negatively correlated. By Theorem 5.2, on the other hand, $P(X_1 = X_2 = 1) = \mathbb{E}[U^2] \geq \mathbb{E}[U]^2 = P(X_1 = 1)P(X_2 = 1)$, so in infinite exchangeable sequences, the random variables are positively correlated when $U$ is not constant. As a result, (5.1.1) fails.

**Pólya urn schemes**

An important application of De Finetti’s Theorem (Theorem 5.2) arises in so-called Pólya urn schemes. An urn consists of a number of balls, and we successively draw balls and replace them in the urn. We start with $B_0 = b_0$ blue balls and $R_0 = r_0$ red balls at time $n = 0$. Let $W_b, W_r : \mathbb{N} \to (0, \infty)$ be two weight
functions. Then, at time $n+1$, the probability of drawing a blue ball, conditionally on the number $B_n$ of blue balls at time $n$, is proportional to the weight of the blue balls at time $n$, i.e., the conditional probability of drawing a blue ball is equal to

$$\frac{W_b(B_n)}{W_b(B_n) + W_r(R_n)}. \quad (5.1.12)$$

After drawing a ball, it is replaced together with a second ball of the same color. We denote this Pólya urn scheme by $((B_n, R_n))_{n=1}^{\infty}$. Naturally, since we always replace one ball by two balls, the total number of balls is deterministic and satisfies $B_n + R_n = b_0 + r_0 + n$.

In this section, we restrict to the case where there exist $a_r, a_b > 0$ such that

$$W_b(k) = a_b + k, \quad W_r(k) = a_r + k, \quad (5.1.13)$$

i.e., both weight functions are linear with the same slope, but possibly a different intercept. Our main result concerning Pólya urn schemes is the following theorem:

**Theorem 5.3** (Limit theorem for linear Pólya urn schemes) Let $((B_n, R_n))_{n=1}^{\infty}$ be a Pólya urn scheme with linear weight functions $W_b$ and $W_r$ as in (5.1.13) for some $a_r, a_b > 0$. Then, as $n \to \infty$, $B_n \overset{a.s.}{\to} U$, where $U$ has a Beta-distribution with parameters $a = b_0 + a_b$ and $b = r_0 + a_r$, and

$$P(B_n = B_0 + k) = \mathbb{E}\left[P\left(\text{Bin}(n, U) = k\right)\right]. \quad (5.1.15)$$

Before proving Theorem 5.3, let us comment on its remarkable content. Clearly, the number of blue balls $B_n$ is not a binomial random variable, as early draws of blue balls reinforce the proportion of blue balls in the end. However, (5.1.15) states that we can first draw a random variable, and then conditionally on that random variable, the number of blue balls is binomial. Thus, we can instead think of the Pólya urn as first drawing the limiting random variable $U$, and, conditionally on its value, after this having a binomial experiment. This is an extremely useful perspective, as we will see later on. The urn conditioned on the limiting variable $U$ is sometimes called a Pólya urn with strength $U$, and Theorem 5.3 implies that this is a mere binomial experiment.

**Proof of Theorem 5.3.** Let $X_n$ denote the indicator that the $n$th ball drawn is blue. We first show that $(X_n)_{n \geq 1}$ is an infinite exchangeable sequence. For this, we note that

$$B_n = b_0 + \sum_{j=1}^{n} X_j, \quad R_n = r_0 + \sum_{j=1}^{n} (1 - X_j) = r_0 - b_0 + n - B_n. \quad (5.1.16)$$

Now, for any sequence $(x_t)_{t=1}^{n}$,

$$P((X_t)_{t=1}^{n} = (x_t)_{t=1}^{n}) = \prod_{t=1}^{n} \frac{W_b(b_{t-1})^{x_t}W_r(r_{t-1})^{1-x_t}}{W_b(b_{t-1}) + W_r(r_{t-1})}, \quad (5.1.17)$$
Thus, we arrive at (see Exercise 5.6).

\[ \prod_{t=1}^{n} (W_b(b_{t-1}) + W_r(r_{t-1})) = \prod_{t=1}^{n} (b_0 + r_0 + a_b + a_r + t - 1), \]  

(5.1.18)

while

\[ \prod_{t=1}^{n} W_b(b_{t-1})^{x_t} = \prod_{m=0}^{k-1} (b_0 + a_b + m), \quad \prod_{t=1}^{n} W_r(r_{t-1})^{1-x_t} = \prod_{j=0}^{n-k-1} (r_0 + a_r + j). \]  

(5.1.19)

Thus, we arrive at

\[ \mathbb{P}((X_t)_{t=1}^{n} = (x_t)_{t=1}^{n}) = \frac{\prod_{m=0}^{k-1} (b + m) \prod_{j=0}^{n-k-1} (r + j)}{\prod_{t=0}^{n-k-1} (b + r + t)}, \]  

(5.1.20)

where \( b = b_0 + a_b \) and \( r = r_0 + a_r \). In particular, (5.1.20) does not depend on the order in which the elements of \((x_t)_{t=1}^{n}\) appear, so that the sequence \((X_n)_{n \geq 1}\) is an infinite exchangeable sequence. Thus, by De Finetti’s Theorem (Theorem 5.2), the sequence \((X_n)_{n \geq 1}\) is a mixture of Bernoulli random variables with a random success probability \( U \), and we are left to compute the distribution of \( U \). We also observe that the distribution of depends only on \( b_0, r_0, a_b, a_r \) through \( b = b_0 + a_b \) and \( r = r_0 + a_r \).

We next verify (5.1.4). For fixed \( 0 \leq k \leq n \), there are \( \binom{n}{k} \) sequences of \( k \) ones and \( n-k \) zeros. Each sequence has the same probability given by (5.1.20). Thus,

\[ \mathbb{P}(S_n = k) = \frac{n \prod_{m=0}^{k-1} (b + m) \prod_{j=0}^{n-k-1} (r + j)}{\prod_{t=0}^{n-k-1} (b + r + t)} \]

\[ = \frac{\Gamma(n+1)}{\Gamma(k+1)\Gamma(n-k+1)} \times \frac{\Gamma(k+b)}{\Gamma(b)} \times \frac{\Gamma(n-k+r)}{\Gamma(r)} \times \frac{\Gamma(b+r)}{\Gamma(n+b+r)} \]

\[ = \frac{\Gamma(b+r)}{\Gamma(r)} \times \frac{\Gamma(k+b)}{\Gamma(k+1)} \times \frac{\Gamma(n-k+r)}{\Gamma(n-k+1)} \times \frac{\Gamma(n+1)}{\Gamma(n+b+r)}. \]  

(5.1.21)

For \( k \) and \( n-k \) large, by [Volume 1, (8.3.9)],

\[ \mathbb{P}(S_n = k) = \frac{\Gamma(b+r)}{\Gamma(r)} \frac{k^{b-1}(n-k)^{r-1}}{n^{b+r-1}} (1 + o(1)). \]  

(5.1.22)

Taking \( k = \lceil un \rceil \) (recall (5.1.4))

\[ \lim_{n \to \infty} n \mathbb{P}(S_n = \lceil un \rceil) = \frac{\Gamma(b+r)}{\Gamma(r)} \frac{k^{b-1}(1-u)^{r-1}}{n^{b+r-1}}, \]  

(5.1.23)

which is the density of a Beta-distribution with parameters \( b \) and \( r \). It is not hard to show from (5.1.23) that (5.1.4) holds with \( f(u) \) the right-hand side of (5.1.23) (see Exercise 5.6).
Multiple urn extensions

We next explain how Theorem 5.3 can be inductively extended to urns with several colors of balls. This will be essential in the analysis of preferential attachment models, where we will need a large number of urns. Assume that we have an urn with several colors \((C_i(n))_{i \in [\ell], n \geq 0}\). Again, we restrict to the setting where the weight functions in the urn are affine, i.e., there exist \((a_i)_{i \in [\ell]}\) such that

\[ W_i(k) = a_i + k. \tag{5.1.24} \]

We assume that the Pólya urn starts with \(k_i\) balls of color \(i\), and that a ball is drawn according to the weights \(W_i(C_i(n))\) for \(i \in [\ell]\), and then replaced by two balls of the same color. For \(j \in [\ell]\), we let

\[ a_{j,\ell} = \sum_{j=1}^{\ell} a_i \quad \text{and} \quad C_{j,\ell}(n) = \sum_{j=1}^{\ell} C_i(n). \]

We first view the balls of color 1 and the other colors as a two type urn. Thus,

\[ \frac{C_1(n)}{n} \overset{a.s.}{\to} U_1, \tag{5.1.25} \]

where \(U_1\) has a Beta distribution with parameters \(a = k_1 + a_1\) and \(b = k_{[2,\ell]} + a_{[2,\ell]}\). This highlights what the proportion of balls of color 1 is, but it groups all other balls together as one ‘combined’ color. This combined color takes a proportion \(1 - U_1\) of the balls. Now, the times that a ‘combined’ color ball is being drawn again forms a (multitype) Pólya urn scheme, now with the colors \(2, \ldots, \ell\). This implies that

\[ \frac{C_2(n)}{n} \overset{a.s.}{\to} U_2(1 - U_1), \tag{5.1.26} \]

where \(U_2\) is independent of \(U_1\) and has a Beta distribution with parameters \(a = k_2 + a_2\) and \(b = k_{[3,\ell]} + a_{[3,\ell]}\). Repeating gives that

\[ \frac{C_i(n)}{n} \overset{a.s.}{\to} U_i \prod_{j=1}^{i-1} (1 - U_j), \tag{5.1.27} \]

where \(U_i\) is independent of \((U_1, \ldots, U_{i-1})\) and has a Beta distribution with parameters \(a = k_i + a_i\) and \(b = k_{[i,\ell]} + a_{[i,\ell]}\). This not only gives an extension of Theorem 5.3 to urns with multiple colors, but also gives an appealing independence structure of the various limits.

Applications to relative sizes in scale-free trees

We close this section by discussing applications of Pólya urn schemes to scale-free trees. We start at time \(n = 2\) with an initial graph consisting of two vertices of which vertex 1 has degree \(d_1\) and vertex 2 has degree \(d_2\). Needless to say, in order for the initial graph to be possible, we need that \(d_1 + d_2\) to be even, and the graph may contain self-loops and multiple edges. After this, we successively attach vertices to older vertices with probability proportional to the degree plus \(\delta > -1\). We do not allow for self-loops in the growth of the trees, so that the structures connected to vertices 1 and 2 are trees (but the entire structure is not be when \(d_1 + d_2 > 2\)). This is a generalization of \((\text{PA}_n^{(\delta)}(b))_{n=2}^\infty\), in which we
are are more flexible in choosing the initial graph. The model for \((\text{PA}_n^{(1,\alpha)}(b))_{n \geq 1}\) arises when \(d_1 = d_2 = 2\) (see Exercise 5.8).

We decompose the growing tree in two trees. For \(i = 1, 2\), we let \(T_i(n)\) be the tree of vertices that are closer to \(i\) than to \(3-i\). Thus, the tree \(T_2(n)\) consists of those vertices for which the path in the tree from the vertex to the root passes through vertex 2, and \(T_1(n)\) consists of the remainder of the scale-free tree. Let \(S_1(n) = |T_1(n)|\) denote the number of vertices in \(T_1(n)\). Clearly, \(S_1(n) + S_2(n) = n\), which is the total number of vertices of the tree at time \(n\). We can apply Theorem 5.3 to describe the relative sizes of \(T_1(n)\) and \(T_2(n)\):

**Theorem 5.4** (Tree decomposition for scale-free trees) For scale-free trees with initial degrees \(d_1, d_2 \geq 1\), as \(n \to \infty\),

\[
\frac{S_1(n)}{n} \xrightarrow{a.s.} U,
\]

where \(U\) has a Beta-distribution with parameters \(a = (d_1 + \delta)/(2 + \delta)\) and \(b = (d_2 + \delta)/(2 + \delta)\), and

\[
\mathbb{P}(S_1(n) = k) = \mathbb{E}\left[\mathbb{P}(\text{Bin}(n-1, U) = k-1)\right].
\]

By Theorem 5.4, we can decompose a scale-free tree into two disjoint scale-free trees each of which contains a positive proportion of the vertices that converges almost surely to a Beta-distribution with parameters \(a = (d_1 + \delta)/(2 + \delta)\) and \(b = (d_2 + \delta)/(2 + \delta)\).

**Proof of Theorem 5.4.** The evolution of \((S_1(n))_{n \geq 2}\) can be viewed as a Pólya urn scheme. Indeed, when \(S_1(n) = s_1(n)\), then the probability of attaching the \((n+1)\)st vertex to \(T_1(n)\) is equal to

\[
\frac{(2s_1(n) + d_1 - 2) + \delta s_1(n)}{(2s_1(n) + d_1 - 2) + \delta s_1(n) + 2(s_2(n) + d_2) + \delta s_2(n)},
\]

since the number of vertices in \(T_1(n)\) equals \(S_1(n)\), while the total degree of \(T_1(n)\) equals \((2S_1(n) + d_1 - 2)\). We can rewrite this as

\[
\frac{s_1(n) + (d_1 - 2)/(2 + \delta)}{s_1(n) + s_2(n) + (d_1 + d_2 - 4)/(2 + \delta)},
\]

which is equal to (5.1.12) in the case (5.1.13) when \(r_0 = b_0 = 1\) and \(a_0 = (d_1 - 2)/(2 + \delta)\), \(a_r = (d_2 - 2)/(2 + \delta)\). Therefore, the proof of Theorem 5.4 follows directly from Theorem 5.3.

We continue by adapting the above argument to the size of the connected component of, or subtree containing, vertex 1 in \(\text{PA}_n^{(1,\alpha)}\), which we denote by \(S_1'(n)\).

**Theorem 5.5** (Tree decomposition for preferential attachment trees) For \((\text{PA}_n^{(1,\alpha)}))_{n \geq 1}\), as \(n \to \infty\),

\[
\frac{S_1'(n)}{n} \xrightarrow{a.s.} U',
\]
where $U'$ has a mixed Beta-distribution with parameters $a = I + 1$ and $b = 1 + (1 + \delta)/(2 + \delta)$, where, for $k \geq 2$,

$$\mathbb{P}(I = k) = \mathbb{P}(\text{first vertex that is not connected to vertex 1 is vertex } k).$$

(5.1.33)

Consequently,

$$\mathbb{P}(S_1(n) = k) = \mathbb{E}\left[\mathbb{P}(\text{Bin}(n - 1, U') = k - 1)\right].$$

(5.1.34)

Proof of Theorem 5.5. We note that $S_1'(n) = n$ for all $n < I$, and $S_1'(I) = I - 1, S_2'(I) = 1$. For $n \geq I + 1$, the evolution of $(S_1'(n))_{n \geq 2}$ can be viewed as a Pólya urn scheme. Indeed, when $S_1'(n) = s_1'(n)$, then the probability of attaching the $(n + 1)$st vertex to the tree rooted at vertex 1 is equal to

$$\frac{(2 + \delta)s_1'(n)}{(2 + \delta)n + 1 + \delta}. \quad (5.1.35)$$

We can rewrite this as

$$\frac{s_1'(n)}{n + (1 + \delta)/(2 + \delta)} = \frac{(s_1'(n) - I) + I}{(n - I) + I + (1 + \delta)/(2 + \delta)},$$

(5.1.36)

which is equal to (5.1.12) in the case (5.1.13) when $b_0 = r_0 = 1$ and $a_b = I, a_r = (1 + \delta)/(2 + \delta)$. Therefore, the proof of Theorem 5.5 follows directly from Theorem 5.3.

Applications to relative degrees in scale-free trees

We continue by discussing an application of Pólya urn schemes to the relative sizes of the initial degrees. For this, we fix an integer $k \geq 2$, and only regard times $n \geq k$ at which an edge is attached to one of the $k$ initial vertices. We work with $(PA_n^{(1, \delta)})_{n \geq 1}$, so that we start at time $n = 1$ with one vertex with one self-loop, after which we successively attach vertices to older vertices with probability proportional to the degree plus $\delta > -1$, allowing for self-loops. The main result is as follows:

**Theorem 5.6 (Relative degrees in scale-free trees)** For $(PA_n^{(1, \delta)})_{n \geq 1}$, as $n \to \infty$,

$$\frac{D_k(n)}{D_{[k]}(n)} \xrightarrow{a.s.} B_k,$$

(5.1.37)

where $D_{[k]}(n) = D_1(n) + \cdots + D_k(n)$ and $B_k$ has a Beta-distribution with parameters $a = 1 + \delta$ and $b = (k - 1)(2 + \delta)$.

By Theorem 1.11, $D_k(n)n^{-1/(2 + \delta)} \xrightarrow{a.s.} \xi_k$, where $\xi_k$ is positive almost surely by the argument in the proof of [Volume 1, Theorem 8.14]. It thus follows from Theorem 5.6 that $B_k = \xi_k/(\xi_1 + \cdots + \xi_k)$. We conclude that Theorem 5.6 allows to identify properties of the law of the limiting degrees.
Proof of Theorem 5.6. Denote the sequence of stopping times \( (\tau_k(n))_{n \geq 2k-1} \), by \( \tau_k(2k-1) = k - 1 \), and
\[
\tau_k(n) = \inf\{ t : D_{[k]}(t) = n \},
\] (5.1.38)
i.e., \( \tau_k(n) \) is the time where the total degree of vertices \( 1, \ldots, k \) equals \( n \). The initial condition \( \tau_k(2k-1) = k - 1 \) is chosen such that the half-edge incident to vertex \( k \) is already considered to be present at time \( k - 1 \), but the receiving end of that edge is not. This guarantees that also the attachment of the edge of vertex \( k \) is properly taken into account.

Since \( D_j(n) \xrightarrow{a.s.} \infty \) as \( n \to \infty \) for every \( j \), \( \tau_k(n) < \infty \) for every \( n \). Moreover, since \( \tau_k(n) \xrightarrow{n \to \infty} \infty \),
\[
\lim_{n \to \infty} \frac{D_k(n)}{D_{[k]}(n)} = \lim_{n \to \infty} \frac{D_k(\tau_k(n))}{D_{[k]}(\tau_k(n))} = \lim_{n \to \infty} \frac{D_k(\tau_k(n))}{n}.
\] (5.1.39)
Now, the random variables \( ((D_k(\tau_k(n)), D_{[k-1]}(\tau_k(n))))_{n \geq 2k-1} \) form a Pólya urn scheme, with \( D_k(\tau_k(2k-1)) = 1 \), and \( D_{[k-1]}(\tau_k(2k-1)) = 2k - 2 \). The edge at time \( \tau_k(n) \) is attached to vertex \( k \) with probability
\[
\frac{D_k(\tau_k(n)) + \delta}{n + k\delta},
\] (5.1.40)
which are the probabilities of a Pólya urn scheme in the linear weight case in (5.1.13) when \( a_0 = \delta, a_r = (k-1)\delta, b_0 = 1, r_0 = 2(k-1) \). Thus, the statement follows from Theorem 5.3.

Theorem 5.6 is easily extended to \( (PA_n^{(1,\ell)}(b))_{\ell \geq 1} \):

**Theorem 5.7 (Relative degrees in scale-free trees)** For \( (PA_n^{(1,\ell)}(b))_{\ell \geq 1} \), as \( t \to \infty \),
\[
\frac{D_k(n)}{D_{[k]}(n)} \xrightarrow{a.s.} B_k,
\] (5.1.41)
where \( B_k \) has a Beta-distribution with parameters \( a = 1 + \delta \) and \( b = (2k - 1) + (k - 1)\delta \).

Of course, the dynamics for \( (PA_n^{(1,\ell)}(b))_{\ell \geq 1} \) are slightly different. Indeed, now the random variables \( ((D_k(\tau_k(n)), D_{[k-1]}(\tau_k(n))))_{n \geq 2k} \) form a Pólya urn scheme, with \( D_k(\tau_k(2k)) = 1 \), and \( D_{[k-1]}(\tau_k(2k)) = 2k - 1 \). The edge at time \( \tau_k(n) \) is attached to vertex \( k \) with probability
\[
\frac{D_k(\tau_k(n)) + \delta}{n + k\delta},
\] (5.1.42)
which are the probabilities of a Pólya urn scheme in the linear weight case in (5.1.13) when \( a_0 = \delta, a_r = (k-1)\delta, b_0 = 1, r_0 = 2k - 1 \). Thus, again the statement follows from Theorem 5.3. See Exercise 5.10 for the complete proof. Remarkably,
even though \((\text{PA}^{(1,\delta)}_n)_{n \geq 1}\) and \((\text{PA}^{(1,\delta)}(b))_{n \geq 1}\) have the same asymptotic degree distribution, the limiting degrees in Theorems 5.6 and 5.7 are different.

In the following section, we bring the above discussion substantially forward and study the local weak limit of preferential attachment models.

5.2 Local Weak Convergence of Preferential Attachment Models

In this section, we study the local weak convergence of preferential attachment models, which is a much more difficult subject. Indeed, it turns out that the local weak limit is not described by a homogeneous Galton-Watson tree, but rather by an inhomogeneous one, leading to multi-type branching processes.

5.2.1 Local Weak Convergence of PAMs with Fixed Number of Edges

In this section, we study local weak limits of preferential attachment models, using Pólya urn schemes. We start with \((\text{PA}^m_1(d))_{t \geq 1}\), as this model turns out to be the simplest for local weak limits due to its close relation to Pólya urn schemes.

Recall that the graph starts at time \(t\) with two vertices with \(m\) edges between them. Let \(\tau_k\) be the \(k\)th time that an edge is added to either vertex 1 or 2. The relation to Pólya urn schemes can be informally explained by noting that the random variable \(k \mapsto D_1(\tau_k)/(D_1(\tau_k) + D_2(\tau_k))\) can be viewed as the proportion of type 1 vertices in a Pólya urn starting with \(m\) balls of type 1 and type 2, respectively. Application of a De Finetti Theorem, which relies on exchangeability, shows that \(D_1(\tau_k)/(D_1(\tau_k) + D_2(\tau_k))\) converges almost surely to a certain Beta-distribution, which we denote by \(\beta\). What is particularly nice about this description is that the random variable \(D_1(\tau_k)\) has exactly the same distribution as \(m\) plus a Bin\((k, \beta)\) distribution, i.e., conditionally on \(\beta\), \(D_1(\tau_k)\) is a sum of i.i.d. random variables.

In the graph context, the Pólya urn description becomes more daunting. However, the description again is in terms of Beta random variables, and the exchangeable version of the model \(\text{PA}^m_1(d)\) can again be given rather explicitly in terms of the arising random variables. Let us now give some details.

**Definition of the Pólya point graph**

We continue by introducing the limiting graph. Let

\[
    u = \frac{\delta}{2m},
\]

and write

\[
    \chi = \frac{1 + 2u}{2 + 2u} = \frac{m + \delta}{2m + \delta}, \quad \psi = \frac{1 - \chi}{\chi} = \frac{1}{1 + 2u} = \frac{m}{m + \delta}.
\]
The exponent \( \psi \) equals \( \psi = 1/(\tau - 2) \), where \( \tau = 3 + \delta/m \) equals the power-law exponent of the graph. We will show that asymptotically, the branching process obtained by exploring the neighborhood of a random vertex \( o_n \) in \( \text{PA}^{(m,\delta)}_n(d) \) is given by a multitype branching process, in which every vertex has a type that is closely related to the age of the vertex. To state our main theorem, we introduce this tree.

As for inhomogeneous random graphs and the configuration model, we will see that the local weak limit of \( \text{PA}^{(m,\delta)}_n(d) \) is again a tree, but it is not homogeneous. Further, each vertex has two different types of children, labelled \( L \) and \( R \). The children labeled with \( L \) are the older neighbors that one of the \( m \) edges that the vertex entered with is connected to, while children labeled \( R \) are younger vertices that used one of their \( m \) edges to connect to the vertex. The distinction between the different kind of vertices can be determined by giving each vertex an age variable. Since we are interested in the asymptotic neighborhood of a uniform vertex, the age of the root, which corresponds to the limit of \( o_n/n \), is a uniform random variable on \([0,1]\). In order to describe its immediate neighbors, we have to describe how many older vertices of type \( L \) the root \( o \) is connected to, as well as the number of younger vertices of type \( R \). After this, we again have to describe the number of \( R \) and \( L \) type children its children has, etc.

Let us now describe these constructs in detail. Let the random rooted tree \( T \) be defined by its vertices labeled by finite words, as in the Ulam-Harris labelling of trees,

\[
w = w_1w_2 \cdots w_l,
\]

(5.2.3)
each carrying a label \( R \) or \( L \), which are defined recursively as follows: The root \( \emptyset \) has an age \( U_\emptyset \), where \( U_\emptyset \) is chosen uniformly at random in \([0,1]\).

In the recursion step, we assume that \( w \) and the corresponding variable \( x_w \in [0,1] \) have been chosen in a previous step. For \( j \geq 1 \), let \( w_j \) be the \( j \)th child of \( w \), i.e., \( w_j = \emptyset w_1w_2 \cdots w_j \), and set

\[
m_-(w) = \begin{cases} m & \text{if } w \text{ is the root or of type } L, \\ m - 1 & \text{if } w \text{ is of type } R. \end{cases}
\]

(5.2.4)
Recall that a Gamma distribution \( Y \) with parameters \( r \) and \( \lambda \) has density

\[
f_Y(y) = y^{r-1}e^{-\lambda y}/\Gamma(r) \quad \text{for } y \geq 0,
\]

(5.2.5)
and 0 otherwise. Let \( Y \) have a Gamma distribution with parameters \( r = m + \delta \) and \( \lambda = 1 \), and let \( Y^* \) be the size-biased version of \( Y \), which has a Gamma distribution with parameters \( r = m + \delta + 1 \) and \( \lambda = 1 \) (see Exercise 5.11). We then take

\[
\Gamma_w \sim \begin{cases} Y & \text{if } w \text{ is the root or of type } R, \\ Y^* & \text{if } w \text{ is of type } L, \end{cases}
\]

(5.2.6)
independently of everything else.
5.2 Local weak convergence of preferential attachment models

Let \( w_1, \ldots, w_{m(w)} \) be the children of \( w \) type \( L \), and let their ages \( A_{w_1}, \ldots, A_{w_{m(w)}} \) be given by \( A_{w_j} = U_{w_j}^{1/(\tau-2)} A_w \), where \( (U_{w_j})_{j=1}^{m(w)} \) are i.i.d. uniform random variables on \([0,1]\) independent of everything else. Further, let \( (A_{w(m(w)+j)})_{j \geq 1} \) be the (ordered) points of a Poisson point process with intensity

\[
\rho_w(x) = \Gamma_w(\tau - 1) x^{1/(\tau-1)-1} A_w^{1/(\tau-1)} \tag{5.2.7}
\]

on \([0,1]\), and the vertices \( (w(m(w)+j))_{j \geq 1} \) have type \( R \). The children of \( w \) are the vertices \( w_j \) of type \( L \) and \( R \).

Let us discuss the degree structure of the above process. Obviously, there are finitely many children of type \( L \). Further, note that \( 1/(\tau-1) = m/(2m+\delta) > 0 \), so the intensity \( \rho_w \) in (5.2.7) of the Poisson process is integrable. Thus every vertex in the random tree has a.s. finitely many children. The above random tree is coined the \( \text{Pólya-point tree} \), and the point process \((a_w)_w\) the \( \text{Pólya-point process} \). The Pólya-point tree is a multitype discrete-time branching process, where the type of a vertex \( w \) is equal to the pair \((a_w, t_w)\), where \( a_w \in [0,1] \) corresponds to the age of the vertex, and \( t_w \in \{L, R\} \) is its type. Thus, the type-space of the multitype branching process is continuous.

With the above description in hand, we are ready to state our main result concerning local weak convergence of \( \text{PA}_{n}^{(m,\delta)}(d) \):

**Theorem 5.8** (Local weak convergence of preferential attachment models; fixed edges) Fix \( m \geq 1 \) and \( \delta > -m \). The preferential attachment model \( \text{PA}_{n}^{(m,\delta)}(d) \) converges locally in probability to the Pólya-point tree.

Theorem 5.8 does not include the case where self-loops are allowed. Given the robustness of the theorem (which applies to three quite related settings), we strongly believe that a similar result also applies to \( \text{PA}_{n}^{(m,\delta)} \). We refer to the discussion in Section 5.6 for more details, also on the history of Theorem 5.8.

**Degree structure of \( \text{PA}_{n}^{(m,\delta)}(d) \)**

Before turning to the proof of Theorem 5.8, we use it to describe some properties of the degrees of vertices in \( \text{PA}_{n}^{(m,\delta)}(d) \). This is done in the following lemma:

**Lemma 5.9** (Degree sequence of \( \text{PA}_{n}^{(m,\delta)}(d) \)) Let \( D_o(n) \) be the degree at time \( n \) of a vertex chosen uniformly at random from \([n]\) in \( \text{PA}_{n}^{(m,\delta)}(d) \). Then,

\[
\mathbb{P}(D_o(n) = k) \to p_k = \frac{2m + \delta \Gamma(m + 2 + \delta + \delta/m)}{m \Gamma(m + \delta)} \frac{\Gamma(k + \delta)}{\Gamma(k + 3 + \delta + \delta/m)}, \tag{5.2.8}
\]

and, with \( D'_o(n) \) be the degree at time \( n \) of one of the \( m \) older neighbors of \( o \),

\[
\mathbb{P}(D'_o(n) = k) \to p'_k = \frac{2m + \delta \Gamma(m + 2 + \delta + \delta/m)}{m^2 \Gamma(m + \delta)} \frac{(k - m + 1)\Gamma(k + 1 + \delta)}{\Gamma(k + 4 + \delta + \delta/m)} \tag{5.2.9}
\]

Further, let \( P_k(n) \) denote the proportion of vertices of degree \( k \) in \( \text{PA}_{n}^{(m,\delta)}(d) \), and
We now use the integral transform $u$ where we use that $\kappa$ for some constants $c$ conclude that there is a form of size-biasing on the limiting degree sequence. It is not hard to see from Lemma 5.9 that $p_k = c_{m, \delta} k^{-\tau}(1 + O(1/k))$, $p'_k = c'_{m, \delta} k^{-(\tau-1)}(1 + O(1/k))$, \hspace{1cm} (5.2.10)

for some constants $c_{m, \delta}$ and $c'_{m, \delta}$ and with $\tau = 3 + \delta/m$ (see Exercise 5.12). We conclude that there is a form of size-biasing in that older neighbors of a uniform vertex have a limiting degree distribution that again satisfies a power law (like the degree of the random vertex itself), but with an exponent that is one lower than that of the vertex itself. Exercises 5.13–5.15 study the joint distribution $(D, D')$ and various conditional power laws.

**Proof of Lemma 5.9.** We note that local weak convergence implies the convergence of the degree distribution. It thus suffices to study the distribution of the degree of the root in the Pólya graph. We first condition on the age $A_{\emptyset}$ of the root of the Pólya graph, where $U_{\emptyset}$ is standard uniform. Let $D$ be the degree of the root. Conditionally on $A_{\emptyset} = a$, the degree $D$ is $m$ plus a Poisson variable with parameter

$$
\frac{\Gamma_{\emptyset}}{a^{1/(\tau-1)}} \int_0^1 (\tau - 1)x^{1/(\tau-1)-1}dx = \Gamma_{\emptyset} \frac{1-a^{1/(\tau-1)}}{a^{1/(\tau-1)}} = \Gamma_{\emptyset} \kappa(a), \hspace{1cm} (5.2.11)
$$

where $\Gamma_{\emptyset}$ is a Gamma variable with parameters $r = m + \delta$ and $\lambda = 1$. Thus, taking expectations with respect to $\Gamma_{\emptyset}$, we obtain

$$
P(D = k \mid A_{\emptyset} = a) = \int_0^\infty P(D = k \mid A_{\emptyset} = a, \Gamma_{\emptyset} = y) \frac{y^{m+\delta-1}}{\Gamma(m+\delta)} e^{-y}dy
$$

$$
= \int_0^\infty e^{-y\kappa(a)} \frac{(y\kappa(a))^{k-m}}{(k-m)!} \frac{y^{m+\delta-1}}{\Gamma(m+\delta)} e^{-y}dy
$$

$$
= \frac{\kappa(a)^{k-m}}{(1 + \kappa(a))^{k-m+m+\delta}} \frac{\Gamma(k+\delta)}{(k-m)!\Gamma(m+\delta)}
$$

$$
= (1 - a^{1/(\tau-1)})^{k-m} a^{(m+\delta)/(2m+\delta)} \frac{\Gamma(k+\delta)}{(k-m)!\Gamma(m+\delta)}, \hspace{1cm} (5.2.12)
$$

where we use that $\kappa(a)/(1 + \kappa(a)) = 1 - a^{1/(\tau-1)}$. We thus conclude that

$$
P(D = k) = \int_0^1 P(D = k \mid A_{\emptyset} = a) da
$$

$$
= \int_0^1 (1 - a^{1/(\tau-1)})^{k-m} a^{(m+\delta)/(2m+\delta)} \frac{\Gamma(k+\delta)}{(k-m)!\Gamma(m+\delta)} da.
$$

We now use the integral transform $u = a^{1/(\tau-1)}$, for which $da = (\tau - 1)u^{2-\tau}du$, \hspace{1cm} (5.2.13)
to arrive at

\[ P(D = k) = (\tau - 1) \frac{\Gamma(k + \delta)}{(k - m) \Gamma(m + \delta)} \int_0^1 (1 - u)^{k-m} u^{m+\delta+1+\delta/m} du \]

\[ = (\tau - 1) \frac{\Gamma(k + \delta)}{(k - m) \Gamma(m + \delta)} \frac{\Gamma(k - m + 1) \Gamma(m + 2 + \delta + \delta/m)}{\Gamma(k + 3 + \delta + \delta/m)} \]

\[ = (\tau - 1) \frac{\Gamma(k + \delta) \Gamma(m + 2 + \delta + \delta/m)}{\Gamma(m + \delta) \Gamma(k + 3 + \delta + \delta/m)}, \tag{5.2.14} \]

as required.

We next extend this to convergence in distribution of \( D'_\varnothing(n) \), for which we again note that local weak convergence implies the convergence of the degree distribution of neighbors at distance 1 of the root, so in particular of \( D'_\varnothing(n) \). It thus suffices to study the distribution of the degree of a uniform neighbor of the root in the Pólya graph. We first condition on the age \( A_\varnothing = U_\varnothing \) of the root of the Pólya graph, where \( U_\varnothing \) is standard uniform, and recall that the age \( A_{\varnothing 1} \) of one of the \( m \) older vertices to which \( \varnothing \) is connected has distribution \( A_{\varnothing 1} = U_{\varnothing 1}^{(r-1)/(r-2)} A_\varnothing \), where \( U_{\varnothing 1} \) is uniform on \([0,1]\). Let \( D' \) be the degree of vertex \( \varnothing 1 \). Conditionally on \( A_{\varnothing 1} = b \), the degree \( D' \) is \( m \) plus a Poisson variable with parameter

\[ \frac{\Gamma_{\varnothing 1}}{b^{1/(r-1)}} \int_b^1 (\tau - 1) x^{1/(\tau-1)-1} dx = \Gamma_{\varnothing 1} \frac{1 - b^{1/(\tau-1)}}{b^{1/(\tau-1)}} \equiv \Gamma_{\varnothing 1} \kappa(b), \tag{5.2.15} \]

where \( \Gamma_{\varnothing 1} \) is a Gamma variable with parameters \( r = m + 1 + \delta \) and \( \lambda = 1 \).

Thus, taking expectations with respect to \( \Gamma_{\varnothing 1} \), we obtain as before

\[ P(D' = k \mid A_{\varnothing 1} = b) = \frac{1}{\Gamma_{\varnothing 1}} \int_0^\infty P(D' = k \mid A_{\varnothing 1} = b, \Gamma_{\varnothing 1} = y) y^{m+1+\delta} e^{-y} dy \]

\[ = \frac{\kappa(b)^{k-m}}{(k - m)! \Gamma(k + 1 + \delta)} \frac{y^{m+\delta}}{\Gamma(m + 1 + \delta)} e^{-y} dy \]

\[ = \frac{\Gamma(k + 1 + \delta)}{(k - m)! \Gamma(m + 1 + \delta)} (1 - b^{1/(\tau-1)})^{k-m} \lambda^{m+1+\delta}/(2m+\delta), \]

where we again use that \( \kappa(b)/(1 + \kappa(b)) = 1 - b^{1/(\tau-1)} \). We next use that \( A_{\varnothing 1} = U_{\varnothing 1}^{(r-2)/(r-1)} A_\varnothing \), where \( A_\varnothing \) is uniform on \([0,1]\).

Recall that the vector \( (A_\varnothing, U_{\varnothing 1}) \) has density 1 on \([0,1]^2\). Let \( (A_\varnothing, A_{\varnothing 1}) = (A_\varnothing, U_{\varnothing 1}^{(r-2)/(r-1)} A_\varnothing) \), then \( (A_\varnothing, A_{\varnothing 1}) \) has joint density, on \( \{(a,b) : b \leq a\} \) given by

\[ f_{(A_\varnothing, A_{\varnothing 1})}(a,b) = \frac{\tau - 2}{\tau - 1} a^{-(\tau-2)/(\tau-1)} b^{-1/(\tau-1)}. \tag{5.2.16} \]
Theorem 5.8 states local weak convergence in probability.

We thus conclude that
\[
P(D' = k) = \frac{\tau - 2}{\tau - 1} \int_0^1 a^{-(\tau - 1)/(\tau - 2)} \int_0^a b^{-1/\tau - 1} \, db \, \mathbb{P}(D = k \mid A_{\tau 1} = b) \, db da
\]
\begin{align*}
&= \frac{\tau - 2}{\tau - 1} \frac{\Gamma(k + 1 + \delta)}{(k - m)! \Gamma(m + 1 + \delta)} \\
&\quad \times \int_0^1 a^{-(\tau - 2)/(\tau - 1)} \int_0^a (1 - b^{1/(\tau - 1)})^{k - m} b^{m(1 + \delta) / (2m + \delta) - 1/\tau - 1} \, db \, da \\
&= \frac{\tau - 2}{\tau - 1} \frac{\Gamma(k + 1 + \delta)}{(k - m)! \Gamma(m + 1 + \delta)} \\
&\quad \times \int_0^1 a^{-(\tau - 2)/(\tau - 1)} \int_0^a (1 - b^{1/(\tau - 1)})^{k - m} b^{m(1 + \delta) / (2m + \delta)} \, db \, da \\
&= (\tau - 2)(\tau - 1) \frac{\Gamma(k + 1 + \delta)}{(k - m)! \Gamma(m + 1 + \delta)} \int_0^1 \int_0^1 (1 - v)^{k - m} v^{m + 1 + \delta/m} \, dv \, du,
\end{align*}

where we have now used the integral transform \( u = a^{\tau - 1} \) and \( v = b^{\tau - 1} \). Interchanging the integrals over \( u \) and \( v \) leads to
\[
\mathbb{P}(D' = k) = (\tau - 2)(\tau - 1) \frac{\Gamma(k + 1 + \delta)}{(k - m)! \Gamma(m + 1 + \delta)} \int_0^1 (1 - v)^{k - m + 1 + \delta/m} \, dv
\]
\begin{align*}
&= \frac{2m + \delta}{m^2} \frac{\Gamma(m + 2 + \delta + \delta/m)}{\Gamma(m + \delta)} \frac{(k - m + 1) \Gamma(k + 1 + \delta)}{\Gamma(k + 4 + \delta + \delta/m)},
\end{align*}

as required.

The statements that \( P_k(n) \overset{\tau}{\to} p_k \) and \( P_k'(n) \overset{\tau}{\to} p_k' \) follow from the fact that Theorem 5.8 states local weak convergence in probability.

The proof of Theorem 5.8 relies crucially on exchangeability and applications of De Finetti’s Theorem. The crucial observation is that De Finetti’s Theorem can be used to give an equivalent formulation of \( \text{PA}_n^{(m,\delta)}(d) \) that relies on independent random variables. We explain this now.

**Finite-graph Pólya version of \( \text{PA}_n^{(m,\delta)}(d) \)**

We now explain the finite-graph Pólya version of \( \text{PA}_n^{(m,\delta)}(d) \). We start by introducing the necessary notation. Let \((\psi_j)_{j \geq 1}\) be independent random variables with a Beta distribution with parameters \( \lambda = m + \delta, r = (2j - 3)m + \delta(j - 1) \), i.e.,
\[
\psi_j \sim \text{Beta}(m + \delta, (2j - 3)m + \delta(j - 1)).
\]

Here we recall that \( Y \) has a Beta distribution with parameters \((\alpha, \beta)\) when
\[
f_Y(y) = y^{\alpha - 1}(1 - y)^{\beta - 1} / B(\alpha, \beta) \quad \text{for} \quad y \in [0, 1],
\]

where \( B(\alpha, \beta) \) is the Beta function.
5.2 Local weak convergence of preferential attachment models

where \( B(\alpha, \beta) = \Gamma(\alpha)\Gamma(\beta)/\Gamma(\alpha + \beta) \) is the Beta-function. We denote this as \( Y \sim \text{Beta}(\alpha, \beta). \) Define

\[
\varphi_j^{(n)} = \psi_j \prod_{i=j+1}^n (1 - \psi_i), \quad S_k^{(n)} = \sum_{j=1}^k \varphi_j^{(n)} = \prod_{i=k+1}^n (1 - \psi_i). \tag{5.2.21}
\]

Here the latter equality follows simply by induction on \( k \geq 1 \) (see Exercise 5.16).

Finally, let \( I_k^{(n)} = [S_k^{(n)}, S_{k+1}^{(n)}]. \) We now construct a graph as follows:

\[\triangleright\] Conditioned on \( \psi_1, \ldots, \psi_n, \) choose \( (U_{k,i})_{k \in [n], i \in [m]} \) as a sequence of independent random variables, with \( U_{k,i} \) chosen uniformly at random from the (random) interval \([0, S_k^{(n)}]\);

\[\triangleright\] Join two vertices \( j \) and \( k \) if \( j < k \) and \( U_{k,i} \in I_j^{(n)} \) for some \( i \in [m] \) (with multiple edges between \( j \) and \( k \) if there are several such \( i \));

\[\triangleright\] Call the resulting random multi-graph the finite-size Pólya graph of size \( n. \)

The main result for \( \text{PA}^{(m, \delta)}_n(d) \) is as follows:

**Theorem 5.10** (Finite-graph Pólya version of \( \text{PA}^{(m, \delta)}_n(d) \)) Fix \( \delta > -m \) and \( m \geq 1. \) Then, the distribution of \( \text{PA}^{(m, \delta)}_n(d) \) is the same as that of the finite-size Pólya graph of size \( n. \)

The nice thing about Theorems 5.8–5.10 is that they also allow for an investigation of the degree distribution and various other quantities of interest in the preferential attachment model. Exercises 5.17–5.19 use Theorem 5.10 to derive properties of the number of multiple edges in \( \text{PA}^{(m, \delta)}_n(d) \) for \( m = 2. \)

In terms of the above Pólya-point tree, the proof will show that the Gamma variables \( Y \) that define the “strength” \( \Gamma_w \) are inherited from the Beta random variables \( (\psi_k)_{k \in [n]}, \) while the “position” variables \( X_w \) are inherited from the random variables \( (S_k^{(n)})_{k \in [n]} \) (see Lemmas 5.17 and 5.18 below).

Before moving to the proof of Theorem 5.10, we use it to identify the almost sure limit of degrees of fixed vertices. For this, we recall that \( \chi = (m+\delta)/(2m+\delta) \) and we define the following random variable:

\[
F_k = \lim_{n \to \infty} \left( \frac{n}{k} \right)^\chi \prod_{j=k+1}^n (1 - \psi_j). \tag{5.2.22}
\]

In Exercise 5.20 you are asked to prove that the almost sure limit in (5.2.22) exists. Then, the limiting distribution of fixed vertices is as follows:

**Lemma 5.11** (Degree structure of \( \text{PA}^{(m, \delta)}_n(d) \): fixed vertices) Consider \( \text{PA}^{(m, \delta)}_n(d) \) with \( m \geq 1 \) and \( \delta > -m. \) Then,

\[
n^{-1-\chi} D_k(n) \xrightarrow{d} (2m + \delta)k^\chi \psi_k F_k. \tag{5.2.23}
\]

We prove Lemma 5.11 below. We also note that \( 1 - \chi = m/(2m + \delta) = 1/(2 + \delta/m) = 1/\tau, \) where \( \tau = 3 + \delta/m \) denotes the degree power-law exponent of the preferential attachment model. Thus, Lemma 5.11 confirms the scaling
in [Volume 1, Theorem 8.2], but gives a different description for the limiting variables. Note further that these limiting random variables are slightly different in $\text{PA}_n^{(m, \delta)}(d)$ compared to $\text{PA}_n^{(m, \delta)}$.

Let us give some insight into the proof of Theorem 5.10, after which we give two full proofs. The first proof relies on Pólya urn methods, the second on a direct computation. For the Pólya urn proof, we rely on the fact that there is a close connection between the preferential attachment model and the Pólya urn model in the following sense: every new vertex that a vertex gains can be represented by a new ball added in the urn corresponding to that vertex. The initial number of balls in each urn is equal to $m$. Indeed, recall the discussion above about $D_1(\tau_k)/(D_1(\tau_k) + D_2(\tau_k))$ converging almost surely to a Beta-distribution. What is particularly nice about this description is that the random variable $D_1(\tau_k)$ has exactly the same distribution as $m$ plus a $\text{Bin}(k, \beta)$ distribution, i.e., conditionally on $\beta$, $D_1(\tau_k)$ is a sum of i.i.d. random variables, this equality in distribution being valid for all $k$. This observation can be extended to give a probabilistic description of $(D_1(\tau_k(n)), D_2(\tau_k(n)), \ldots, D_k(\tau_k(n)))_{n \geq 1}$ that is valid for all $k, n \geq 1$ fixed. Here $\tau_k(n) = \inf \{t: D_k(t) = n\}$ is the first time where the total degree of the vertices in $[k]$ is equal to $n$. Note that $\tau_{2mk} = k$. Further, the random variables $(D_1(\tau_k(n)), D_2(\tau_k(n)), \ldots, D_k(\tau_k(n)))_{n \geq 1}$ determine the law of $(\text{PA}_n^{(m, \delta)}(b))_{n \in [n]}$ uniquely. This explains why $\text{PA}_n^{(m, \delta)}(b)$ can be described in terms of independent random variables. The precise description in terms of $\psi_j$ in (5.2.19) follows from tracing back the graph construction obtained in this way. We next use this to complete the proof of Theorem 5.10:

Pólya urn proof of Theorem 5.10. Let us consider first a two-urn model, with the number of balls in one urn representing the degree of a particular vertex $k$, and the number of balls in the other representing the sum of the degrees of the vertices $1, \ldots, k - 1$. We will start this process at the point when $n = k$ and $k$ has connected to precisely $m$ vertices in $[k - 1]$. Note that at this point, the urn representing the degree of $k$ has $m$ balls, while the other one has $(2k - 3)m$ balls.

Consider a time in the evolution of the preferential attachment model when we have $n - 1 \geq k$ old vertices, and $i - 1$ edges between the new vertex $n$ and $[k - 1]$. Assume that at this point the degree of $k$ is $d_k$, and the sum of the degrees of vertices in $[k - 1]$ is $d_{<k}$. At this point, the probability that the $i$th edge from $n$ to $[n - 1]$ is attached to $k$ is

$$d_k + \delta$$

$$2m(n - 2) + (1 + \delta)(i - 1),$$

while the probability that it is connected to a vertex in $[k - 1]$ is equal to

$$d_{[k-1]} + \delta(k - 1)$$

$$2m(n - 2) + (1 + \delta)(i - 1).$$

Thus, conditioned on connecting to $[k]$, the probability that the $i$th edge from $n$ to $[n - 1]$ is attached to $k$ is $(d_k + \delta)/Z$, while the probability that the $i$th edge from
5.2 Local weak convergence of preferential attachment models

n to \([n-1]\) is attached to \([k-1]\) is \((d_{[k-1]} + \delta(k-1))/Z\), where \(Z = k\delta + d_{[k]}\) is the normalization constant. Taking into account that the two urns start with \(m\) and \((2k-3)m\) balls, respectively, we see that the evolution of the two bins is a Pólya urn with strengths \(\psi_k\) and \(1-\psi_k\), where \(\psi_k\) has the \(\beta(m+\delta,(2k-3)m+\delta(k-1))\) distribution. We next use this to complete the proof of Theorem 5.10, where we use induction. Indeed, using the two-urn process as an inductive input, we can now construct the Pólya graph defined in Theorem 5.10 in a similar way as it was done for Pólya urns with multiple colors in (5.1.27).

Let \(X_t \in \lceil t/m \rceil\) be the vertex receiving the \(t\)th edge in the sequential model (the other endpoint of this edge being the vertex \([t/m] + 1\)). For \(t \leq m\), \(X_t\) is deterministic (and equal to 1), but starting at \(t = m + 1\), we have a two-urn model, starting with \(m\) balls in each urn. As shown above, the two urns can be described as Pólya urns with strengths \(1 - \psi_2\) and \(\psi_2\). Once \(t > 2m\), \(X_t\) can take three values, but conditioned on \(X_t \leq 2\), the process continues to be a two-urn model with strengths \(1 - \psi_2\) and \(\psi_2\). To determine the probability of the event that \(X_t \leq 2\), we now use the above two-urn model with \(k = 3\), which gives that the probability of the event \(X_t \leq 2\) is \(1 - \psi_3\), at least as long as \(t \leq 3m\). Combining these two-urn models, we get a three-urn model with strengths \((1 - \psi_3)/(1 - \psi_3), \psi_2(1 - \psi_3)\) and \(\psi_3\). Again, this model remains valid for \(t > 3m\), as long as we condition on \(X_t \leq 3\). Continuing inductively, we see that the sequence \(X_t\) evolves in stages:

- For \(t \in [m]\), the variable \(X_t\) is deterministic: \(X_t = 1\).
- For \(t = m + 1, \ldots, 2m\), the distribution of \(X_t \in \{1, 2\}\) is described by a two-urn model with strengths \(1 - \psi_2\) and \(\psi_2\), where \(\psi_2 \sim B_2\).
- In general, for \(t = m(k-1) + 1, \ldots, km\), the distribution of \(X_t \in [k]\) is described by a \(k\)-urn model with strengths

\[
\varphi_{j}^{(k)} = \psi_j \prod_{i=j+1}^{k} (1 - \psi_i), \quad j = 1, \ldots, k. \tag{5.2.26}
\]

Here \(\psi_k\) is chosen at the beginning of the \(k\)th stage, independently of the previously chosen strengths \(\psi_1, \ldots, \psi_{k-1}\) (for convenience, we set \(\psi_1 = 1\)).

Note that the random variables \(\varphi_{j}^{(k)}\) can be expressed in terms of the random variables introduced in Theorem 5.10 as follows. By (5.2.21), \(S_k^{(n)} = \prod_{j=k+1}^{n}(1 - \psi_j)\). This implies that \(\phi_{j}^{(n)} = \psi_j/S_j^{(n)}\), which relates the strengths \(\phi_{j}^{(n)}\) to the random variables defined right before Theorem 5.10, and shows that the process derived above is indeed the process given in the theorem.

We next give a direct proof of Theorem 5.10, that is of independent interest:

**Direct proof of Theorem 5.10.** We continue with the direct proof of Theorem 5.10. In what follows, we let \(\text{PA}_n^{(\alpha)}\) denote the law of the finite-size Pólya graph of size \(n\). Our aim will be to show that \(\mathbb{P}(\text{PA}_n^{(\alpha)} = G) = \mathbb{P}(\text{PA}_n^{(m,\alpha)}(d) = G)\). Here, we think of \(G\) as being a directed graph, where every vertex has out-degree \(m\) and the
out-edges are labeled as \([m]\). Thus, the out-edges correspond to the edges from young to old. For this, on the one hand, we compute directly that

\[
\mathbb{P}(PA_{n}^{(m, \delta)}(d) = G) = \prod_{u \in [n], j \in [m]} \frac{d_{v_{j}}^{(u)}(u) + \delta}{2m(u - 2) + j - 1 + \delta(u - 1)}
\]

(5.2.27)

\[
= \prod_{i=0}^{d_{v_{j}}^{(u)}} (i + m + \delta) \prod_{u \in [n], j \in [m]} \frac{1}{2m(u - 2) + j - 1 + \delta(u - 1)}.
\]

To identify \(\mathbb{P}(PA_{n}' = G)\), it will be convenient to condition on the Beta variables \((\psi_{j})_{j \in [n]}\). We denote the conditional measure by \(\mathbb{P}_{n}\), i.e., for every event \(E\),

\[
\mathbb{P}_{n}(E) = \mathbb{P}(E \mid (\psi_{j})_{j \in [n]}).
\]

(5.2.28)

The advantage of this measure is that now edges are conditionally independent, which allows us to give exact formulas for the probability of a certain graph occurring. We start by computing the edge probabilities under \(\mathbb{P}_{n}\):

**Lemma 5.12** (Edge probabilities in \(PA_{n}'\) conditionally on Beta’s)  
Fix \(m \geq 2\) and \(\delta > -m\), and consider \(PA_{n}'\). For any \(u > v\) and \(j \in [m]\),

\[
\mathbb{P}_{n}(u \stackrel{j}{\rightarrow} v) = \psi_{v}(1 - \psi)(v, u),
\]

(5.2.29)

where, for \(A \subseteq [n]\),

\[
(1 - \psi)_{A} = \prod_{i \in A}(1 - \psi_{i}).
\]

(5.2.30)

**Proof**  
Recall the construction between (5.2.19) and Theorem 5.10. When we condition on \((\psi_{j})_{j \in [n]}\), the only randomness left is that in the uniform random variables \((U_{k, i})_{k \in [n], i \in [m]}\), where \(U_{k, i}\) is uniform on \([0, S_{k-1}(n)]\). Then, \(u \stackrel{j}{\rightarrow} v\) occurs precisely when \(U_{u, j} \in I^{(n)}_{v}\), which occurs with conditional \(\mathbb{P}_{n}\)-probability

\[
\frac{|I^{(n)}_{v}|}{S_{u-1}(n)}.
\]

(5.2.31)

Now we note that

\[
|I^{(n)}_{v}| = S^{(n)}_{v} - S^{(n)}_{v-1} = \psi_{v} \prod_{i=v+1}^{n}(1 - \psi_{i}) = \psi_{v}(1 - \psi)(v, n),
\]

(5.2.32)

while, from (5.2.21),

\[
S^{(n)}_{u-1} = \prod_{i=u}^{n}(1 - \psi_{i}) = (1 - \psi)_{[u, n]}.
\]

(5.2.33)

Taking the ratio yields, with \(u > v\),

\[
\mathbb{P}_{n}(u \stackrel{j}{\rightarrow} v) = \psi_{v}(1 - \psi)(v, u) = \psi_{v}(1 - \psi)(v, n),
\]

(5.2.34)

as required.
5.2 Local weak convergence of preferential attachment models

In the finite-graph Pólya graph PA′ₙ, different edges are conditionally independent, so that we obtain the following corollary:

**Corollary 5.13** (Graph probabilities in PA′ₙ conditionally on Beta’s) Fix m ≥ 2 and δ > −m, and consider PA′ₙ. For any G,

\[ \Pr_n(\text{PA}′ₙ = G) = \prod_{s=1}^{n} \psi_s^{a_s} \prod_{s=1}^{n} (1 - \psi_s)^{b_s}, \]

(5.2.35)

where \( a_s = a_s^{(G)} \) and \( b_s = b_s^{(G)} \) are given by

\[ a_s = d_s^{(G)} - m, \quad b_s = \sum_{u,j} \mathbb{1}_{\{s \in (v_j(u), u)\}} \]

(5.2.36)

where \( E(G) = \{(v_j(u), u) : u \in [n], j \in [m]\} \), so that \( v_j(u) < u \) is the vertex to which the \( j \)th edge of \( u \) is attached in \( G \).

**Proof** Multiply the factors \( \Pr_n(u \xrightarrow{j} v_j(u)) \) for every edge \( (u, v_j(u)) \in E(G) \), and collect the powers of \( \psi_s \) and \( 1 - \psi_s \).

We note that \( a_s \) equals the number of edges in the graph \( G \) that point towards \( s \). This is relevant, since in (5.2.29) in Lemma 5.12, every older vertex \( v \) in an edge receives a factor \( \psi_v \). Further, again by (5.2.29) in Lemma 5.12, there are factors \( 1 - \psi_s \) for every \( s \in (v,u) \) and all edges \( (u,v) \), so \( b_s \) counts how many factors \( 1 - \psi_s \) occur.

We see in Corollary 5.13 that we obtain expectations of the form \( \mathbb{E}[\psi^a(1 - \psi)^b] \), where \( \psi \sim \text{Beta}(\alpha, \beta) \) and \( a, b \geq 0 \). These are computed in the following lemma:

**Lemma 5.14** (Expectations of powers of Beta random variables) For all \( a, b \in \mathbb{N} \) and \( \psi \sim \text{Beta}(\alpha, \beta) \),

\[ \mathbb{E}[\psi^a(1 - \psi)^b] = \frac{(\alpha + a - 1)_a (\beta + b - 1)_b}{(\alpha + \beta + a + b - 1)_{a+b}}, \]

(5.2.37)

and where we recall that \( (x)_m = x(x-1) \cdots (x-m+1) \) denotes the \( m \)th falling factorial of \( x \).

**Proof** A direct computation based on the density of a Beta random variable in (5.2.20) yields that

\[ \mathbb{E}[\psi^a(1 - \psi)^b] = \frac{B(\alpha+a, \beta+b)}{B(\alpha, \beta)} = \frac{\Gamma(\alpha+\beta) \Gamma(\alpha+a) \Gamma(\beta+b)}{\Gamma(\alpha) \Gamma(\beta) \Gamma(\alpha+\beta+a+b)} \]

(5.2.38)

\[ = \frac{(\alpha + a - 1)_a (\beta + b - 1)_b}{(\alpha + \beta + a + b - 1)_{a+b}}. \]

The above computation, when applied to Corollary 5.13, leads to the following expression for the probability of observing a particular graph \( G \):
We next evaluate (5.2.40) in Corollary 5.15 explicitly. Since (5.2.36) is promising. We next identify the other terms, for which we start by analyzing $\alpha$ where $\delta > (5.2.39)$.

Corollary 5.15 (Graph probabilities in $\text{PA}'_n$) Fix $m \geq 2$ and $\delta > -m$, and consider $\text{PA}'_n$. For any $G$,

\[ \mathbb{P}(\text{PA}'_n = G) = \prod_{s=1}^{n-1} \frac{(\alpha + a_s - 1)_{a_s} (\beta_s + b_s - 1)_{b_s}}{\alpha + \beta_s + a_s + b_s - 1}_{a_s + b_s}, \]  

where $\alpha = m + \delta$, $\beta_s = (2s - 3)m + \delta(s - 1)$ and $a_s = a_s^{(G)}$ and $b_s = b_s^{(G)}$ are defined in (5.2.36).

Note that the contribution for $s = n$ equals 1, since $a_n^{(G)} = d_n^{(G)} - m = 0$ almost surely in $\text{PA}'_n$. Corollary 5.15 allows us to complete the direct proof of Theorem 5.10:

Corollary 5.16 (Graph probabilities in $\text{PA}'_n$ and $\text{PA}'_{m,\delta}(d)$) Fix $m \geq 2$ and $\delta > -m$, and consider $\text{PA}'_n$ and $\text{PA}'_{m,\delta}(d)$. For any $G$,

\[ \mathbb{P}(\text{PA}'_n = G) = \mathbb{P}(\text{PA}'_{m,\delta}(d) = G), \]  

where $\alpha = m + \delta$, $\beta_s = (2s - 3)m + \delta(s - 1)$ and $a_s = a_s^{(G)}$ and $b_s = b_s^{(G)}$ are defined in (5.2.36). Consequently, Corollaries 5.13 and 5.15 also hold for $\text{PA}'_{m,\delta}(d)$.

Proof We next evaluate (5.2.40) in Corollary 5.15 explicitly. Since $\alpha = m + \delta$ and $a_s = d_s^{(G)} - m$,

\[ (\alpha + a_s - 1)_{a_s} = \prod_{i=0}^{d_s^{(G)} - m} (i + m + \delta), \]  

which is promising. We next identify the other terms, for which we start by analyzing $b_s^{(G)}$ as

\[ b_s^{(G)} = \sum_{u,j} 1_{\{x \in (v_j(u),u)\}} = \sum_{u,j} 1_{\{x \in (v_j(u),n)\}} - 1_{\{x \in (u,n)\}} \]  

\[ = \sum_{u,j} 1_{\{v_j(u) \in [s-1]\}} - 1_{\{u \in [s-1]\}} = d_s^{(G)} - 2m(s-1), \]  

since $\sum_{u,j} 1_{\{v_j(u) \in [s-1]\}} = d_s^{(G)} - m(s-1)$. Thus,

\[ (\beta_s + b_s - 1)_{b_s} = ((2s - 3)m + \delta(s - 1) + d_s^{(G)} - 2m(s - 1) - 1)_{d_s^{(G)} - 2m(s-1)} \]  

\[ = (\delta(s - 1) + d_s^{(G)} - m - 1)_{d_s^{(G)} - 2m(s-1)}, \]  

while, since $a_s + b_s = d_s^{(G)} - m + d_s^{(G)} - 2m(s-1) = d_s^{(G)} - 2ms + m$ and $\alpha = m + \delta$,

\[ (\alpha + \beta_s + a_s + b_s - 1)_{a_s + b_s} = (m + \delta + (2s - 3)m + \delta(s - 1) + d_s^{(G)} - 2ms + m - 1)_{d_s^{(G)} - 2ms + m} \]  

\[ = (\delta s + d_s^{(G)} - m - 1)_{d_s^{(G)} - 2ms + m}, \]  

\[ = (\delta s + 2ms - m - 1)_{m(\delta s + d_s^{(G)} - m - 1)}\]  

\[ 2ms. \]
5.3 Proof local convergence for preferential attachment models

Therefore,

\[
\prod_{s=1}^{n-1} \frac{(\beta_s + b_s - 1)_{\beta_s}}{(a + \beta_s + a_s + b_s - 1)_{a_s + b_s}}^{n-1} \frac{(\delta(s - 1) + d^{(G)}_{s-1} - m - 1)_{d^{(G)}_{s-1} - 2m(s-1)}}{(\delta s + 2ms - m - 1)_{m}^{n-1}} \frac{(\delta s + d^{(G)}_{s} - m - 1)_{d^{(G)}_{s} - 2ms}}{(\delta s + 2ms - m - 1)_{m}^{n-1}}
\]

(5.2.45)

since the second factor is telescoping with starting and ending value equal to 1. The proof follows, with \( s = u - 1 \),

\[
\prod_{s=1}^{n-1} (\delta s + 2ms - m - 1)_{m} = \prod_{u \in [n]} \prod_{i \in [m]} (\delta(u - 1) + 2m(u - 1) - m - i - 1)_{m}
\]

\[
= \prod_{u \in [n], j \in [m]} (2m(u - 2) + j - 1 + \delta(u - 1)), \tag{5.2.46}
\]

where \( j = m - i \), as required. \(\square\)

5.3 Proof local convergence for preferential attachment models

Asymptotics of \((\psi_k)_{k \in [n]}\) and \((S_k^{(n)})_{k \in [n]}\)

We next continue to analyse the random variables in Theorem 5.10 to prepare us for proving local convergence in Theorem 5.8. We start by analyzing \(\psi_k\) for \( k \) large:

Lemma 5.17 (Gamma asymptotics of Beta variables) \( As k \to \infty, k\psi_k \overset{d}{\to} Y \), where \( Y \) has a Gamma distribution with \( r = m + \delta \) and \( \lambda = 2m + \delta \).

More precisely, take \( f_k(x) \) such that \( \Pr(\psi_k \leq f_k(x)) = \Pr(\chi_k \leq x) \), where \( \chi_k \) has a Gamma distribution with \( r = m + \delta \) and \( \lambda = 1 \) (so that \( Y \overset{d}{=} X/(2m + \delta) \)). For every \( \varepsilon > 0 \), there exists \( K = K_\varepsilon \geq 1 \) sufficiently large. Then, for all \( k \geq K \) and \( x \leq (\log k)^2 \),

\[
\frac{1 - \varepsilon}{k(2m + \delta)} x \leq f_k(x) \leq \frac{1 + \varepsilon}{k(2m + \delta)} x. \tag{5.3.1}
\]

Further, \( \chi_k \leq (\log k)^2 \) for all \( k \geq K \) and with probability at least \( 1 - \varepsilon \).
\textbf{Proof} Fix }x \geq 0. \text{ We compute that}

\begin{equation}
\mathbb{P}(k \psi_k \leq x) = \frac{\Gamma(m + \delta + (2k - 3)m + \delta(k - 1))}{\Gamma(m + \delta)\Gamma((2k - 3)m + \delta(k - 1))} \times \int_0^{x/k} u^{m+\delta-1}(1-u)^{(2k-3)m+\delta(k-1)-1} du
\end{equation}

\begin{equation}
= (1 + o(1)) \frac{[(2m + \delta)k]^m\varepsilon}{\Gamma(m + \delta)} k^{-m(\delta+1)} \int_0^x u^{m+\delta-1}(1 - u/k)^{(2k-3)m+\delta(k-1)-1} du.
\end{equation}

For every }u > 0, \((1 - u/k)^{(2k-3)m+\delta(k-1)-1} \rightarrow e^{-u(2m+\delta)}, \text{ so } \text{that dominated convergence implies that}

\begin{equation}
\mathbb{P}(k \psi_k \leq x) \rightarrow \int_0^x (2m + \delta)^{m+\delta} u^{m+\delta-1} e^{-u(2m+\delta)} du,
\end{equation}

\text{as required. We refrain from proving the other statements, that are more technical versions of the above argument.}

By Lemma 5.17, we see that indeed the Beta random variables \((\psi_k)_{k \in [n]} \) give rise to the Gamma random variables in (5.2.6). This explains the relevance of the Gamma random variables in the Pólya-point graph.

We continue by analyzing the asymptotics for the random variables \((S_k^{(n)})_{k \in [n]} \):

\textbf{Lemma 5.18 (Asymptotics of } S_k^{(n)} \text{)} Recall that } \chi = (m + \delta)/(2m + \delta). \text{ For every } \varepsilon > 0, \text{ there exists } \eta > 0 \text{ and } K < \infty \text{ such that for all } n \geq K \text{ and with probability at least } 1 - \varepsilon,

\begin{equation}
\max_{k \in [n]} \left| S_k^{(n)} - \left( \frac{k}{n} \right)^\chi \right| \leq \eta,
\end{equation}

\text{and}

\begin{equation}
\max_{k \in [n] \setminus [K]} \left| S_k^{(n)} - \left( \frac{k}{n} \right)^\chi \right| \leq \varepsilon \left( \frac{k}{n} \right)^\chi.
\end{equation}

\textbf{Proof} We give the intuition behind Lemma 5.18. We recall from (5.2.21) that } S_k^{(n)} = \prod_{i=k+1}^n (1 - \psi_i), \text{ where } (\psi_k)_{k \in [n]} \text{ are independent random variables. We start by investigating } S_n' = \prod_{i=1}^n (1 - \psi_i), \text{ so that } S_k^{(n)} = S_n'/S_n. \text{ We write}

\begin{equation}
\log S_n' = \sum_{i=1}^n \log(1 - \psi_i).
\end{equation}

Thus, using Lemma 5.17 and with } (\chi_i)_{i \geq 1} \text{ an i.i.d. sequence of Gamma random variables with parameters } r = m + \delta \text{ and using that } k \psi_k \to \chi_k/(2m + 1),

\begin{equation}
\log S_n' \approx - \sum_{i=1}^n \psi_i \approx \frac{1}{2m + \delta} \sum_{i=1}^n \chi_i/i
\end{equation}

\begin{equation}
\approx - \frac{1}{2m + \delta} \sum_{i=1}^n \frac{E[\chi_i]}{i} = \frac{m + \delta}{2m + \delta} \log n = \chi \log n.
\end{equation}
Thus, \( S_k = S_k' / S_n' \approx (k/n)^\chi \), as required. The proof can be completed using martingale techniques and is omitted here.

We use the above analysis to complete the proof of Lemma 5.11:

**Proof of Lemma 5.11.** We use a conditional second moment method on \( D_k(n) \), conditionally on \((\psi_k)_{k \in [n]}\). Denote

\[
F_k^{(n)} = \left( \frac{n}{k} \right)^\chi \prod_{j=k+1}^\ell (1 - \psi_j).
\]

Then,

\[
\mathbb{E}[D_k(n) \mid (\psi_k)_{k \in [n]}] = m \sum_{\ell=k+1}^n \varphi^{(\ell)}_k,
\]

and

\[
\text{Var}(D_k(n) \mid (\psi_k)_{k \in [n]}) \leq m \sum_{\ell=k+1}^n \varphi^{(\ell)}_k.
\]

It thus suffices to show that \( n^{1-\chi} m \sum_{\ell=k+1}^n \varphi^{(\ell)}_k \xrightarrow{a.s.} m \sum_{\ell=k+1}^n \ell^{-\chi} F_k^{(\ell)} \). For this, we note that since \( \varphi_k^{(\ell)} = \psi_k \prod_{j=k+1}^\ell (1 - \psi_j) \), we arrive at

\[
m \sum_{\ell=k+1}^n \varphi^{(\ell)}_k = m \psi_k \sum_{\ell=k+1}^n \left( \frac{k}{\ell} \right)^\chi F_k^{(\ell)} = m k^\chi \psi_k \sum_{\ell=k+1}^n \ell^{-\chi} F_k^{(\ell)}.
\]

Since \( \chi \in (0,1) \) and \( F_k^{(\ell)} \xrightarrow{a.s.} F_k \) as \( \ell \to \infty \), we obtain that

\[
n^{1-\chi} \sum_{\ell=k+1}^n \ell^{-\chi} F_k^{(\ell)} \xrightarrow{a.s.} F_k / (1 - \chi).
\]

This completes the proof.

### 5.3 Proof local convergence for preferential attachment models

Thus, \( S_k = S_k' / S_n' \approx (k/n)^\chi \), as required. The proof can be completed using martingale techniques and is omitted here.

We use the above analysis to complete the proof of Lemma 5.11:

**Proof of Lemma 5.11.** We use a conditional second moment method on \( D_k(n) \), conditionally on \((\psi_k)_{k \in [n]}\). Denote

\[
F_k^{(n)} = \left( \frac{n}{k} \right)^\chi \prod_{j=k+1}^\ell (1 - \psi_j).
\]

Then,

\[
\mathbb{E}[D_k(n) \mid (\psi_k)_{k \in [n]}] = m \sum_{\ell=k+1}^n \varphi^{(\ell)}_k,
\]

and

\[
\text{Var}(D_k(n) \mid (\psi_k)_{k \in [n]}) \leq m \sum_{\ell=k+1}^n \varphi^{(\ell)}_k.
\]

It thus suffices to show that \( n^{1-\chi} m \sum_{\ell=k+1}^n \varphi^{(\ell)}_k \xrightarrow{a.s.} m \sum_{\ell=k+1}^n \ell^{-\chi} F_k^{(\ell)} \). For this, we note that since \( \varphi_k^{(\ell)} = \psi_k \prod_{j=k+1}^\ell (1 - \psi_j) \), we arrive at

\[
m \sum_{\ell=k+1}^n \varphi^{(\ell)}_k = m \psi_k \sum_{\ell=k+1}^n \left( \frac{k}{\ell} \right)^\chi F_k^{(\ell)} = m k^\chi \psi_k \sum_{\ell=k+1}^n \ell^{-\chi} F_k^{(\ell)}.
\]

Since \( \chi \in (0,1) \) and \( F_k^{(\ell)} \xrightarrow{a.s.} F_k \) as \( \ell \to \infty \), we obtain that

\[
n^{1-\chi} \sum_{\ell=k+1}^n \ell^{-\chi} F_k^{(\ell)} \xrightarrow{a.s.} F_k / (1 - \chi).
\]

This completes the proof.

**Regularity properties of the Pólya-point tree and overview**

Due to Lemma 5.18, in the proof of Theorem 5.16 it will be convenient to deal with ‘locations’ rather than with ages of vertices, to indicate their type. For a vertex \( w \), let \( X_w = A_w \) denote its location. Then, we see in Lemma 5.18 that \( S_k^{(n)} \approx X_w / n \).

This will imply that the location of the vertex \( w_j \), for \( j \in [m - (w)] \), is uniform on \([0, X_w] \), which is a nice property. This makes the analysis a little easier at times. Since the type space in terms of ages can be recovered from that in terms of locations, the two multi-type branching processes can be retrieved from one another.

We next discuss some properties of the Pólya-point tree, showing that the rooted tree \( T \) is well defined and, within the \( r \)-neighborhood of the root, all random variables used in its description are uniformly bounded:

**Lemma 5.19 (Regularity Pólya-point tree)** Fix \( r \geq 1 \) and \( \epsilon > 0 \). Then there exist constants \( \eta > 0 \) and \( K < \infty \) such that, with probability at least \( 1 - \epsilon \),
$B_r(∅) \leq K$, $X_w \geq \eta$ for all $w \in B_r(∅)$, and further $\Gamma_w \leq K, \rho_w(\cdot) \leq K$ for all $w \in B_r(∅)$, while finally, $\min_{w,w' \in B_r(∅)} |X_w - X_{w'}| \geq \eta$.

Proof The proof of this lemma is standard, and can, for example, be obtained by induction on $r$. The last bound follows from the continuous nature of the random variables $X_w$, which implies that $X_w \neq X_{w'}$ for all distinct pairs $w, w'$, which implies that any finite number will be with probability at least $1 - \varepsilon$ separated by at least $\delta$.

Let us describe how the proof of Theorem 5.8 is organised. Similarly to the proofs of Theorems 3.11 and 4.1, we will investigate the number of $r$-neighborhoods of a specific shape $t$, using a second moment method. Let

$$N_{n,r}(t) = \sum_{v \in [n]} 1_{\{B^{(n)}_r(v) \simeq t\}}, \quad (5.3.13)$$

where $B^{(n)}_r(v)$ is the $r$-neighborhood of $v$ in $\text{PA}^{(m)\cup}(d)$. With $B_r(∅)$ the $r$-neighborhood of $∅$ in the Pólya-point graph $\mathcal{T}$, we will show that

$$\frac{N_{n,r}(t)}{n} \xrightarrow{\mathcal{P}} \mathbb{P}(B_r(∅) \simeq t). \quad (5.3.14)$$

Proving convergence of is much harder than for Theorems 3.11 and 4.1, as the type of a vertex is crucial in determining the number and types of its children, and the type space is continuous. Thus, we need to take the types of vertices into account, so as to make an inductive analysis in $r$ possible in which we couple both the neighborhoods as well as their types. We start with the first moment, for which we note that

$$\mathbb{E}[N_{n,r}(t)/n] = \mathbb{P}(B^{(n)}_r(o_n) \simeq t). \quad (5.3.15)$$

We do this using a coupling approach, by showing that we can couple $B^{(n)}_r(o_n)$, together with the types of the vertices involved, with $B_r(∅)$ such that the differences are quite small. After that, we use this as the starting point of a recursive analysis in $r$, which is the second step in the analysis. We complete the proof by studying the second moment of the number of $r$-neighborhoods of a specific shape ($t$). Let us now give the details of the convergence of the expected number of one-neighborhoods of a uniform point.

Integrate Proposition 5.20 more fluently in the text!

In what follows, it will be useful to regard $B_r(o)$ as a rooted graph, where the vertices receive labels in $[n]$, and the edges receive labels in $[m]$ corresponding to the label of the directed edge that gives rise to the edge (in either possible direction). Thus, the edge $\{u, v\}$ receives label $j$ when $u \xrightarrow{j} v$ or when $v \xleftarrow{j} u$.

Let $t$ be a rooted vertex- and edge-labelled tree of height exactly $r$, where the vertex labels are in $[n]$ and the edge labels in $[m]$. We write $B_r(o) = t$ to denote that the vertices, edges and edge labels in $B_r(o)$ are given by those in $t$. Note
that this is rather different from $B_r(o) \simeq t$ as defined in Definition 2.3, where $t$ is unlabeled and we investigate the isomorphism of $B_r(o)$ and $t$.

Let $V(t)$ denote the vertex labels in $t$. Also, let $\partial V(t)$ denote the vertices at distance exactly $r$ from the root of $t$, and let $V^*(t) = V(t) \setminus \partial V(t)$ denote the restriction of $t$ to all vertices at distance at most $r - 1$ from its root. With this notation in hand, we have the following characterization of the conditional law of $B_r(o)$:

**Proposition 5.20 (Law of neighborhood in $\text{PA}^{(m,\delta)}_n(d)$)** Fix $m \geq 2$ and $\delta > -m$, and consider $\text{PA}^{(m,\delta)}_n(d)$. Let $t$ be a rooted vertex- and edge-labelled tree with root $o$. Then, as $n \to \infty$,$$
P\left( B_r(o) = t \mid (\psi_v)_{v \in V^*(t)} \right) = (1 + o_v(1)) \prod_{v \in V^*(t)} \psi_v^{a'_v} e^{-(2m+\delta)\psi_v(1-(v/n)^{1-\delta})} (1 - \psi_v)^{b'_v} \times \prod_{s \in [n] \setminus V^*(t)} (\alpha + a'_s - 1)_{a'_s + b'_s - 1}(\alpha + \beta_s + a'_s + b'_s - 1)_{a'_s + b'_s}$$

(5.3.16)

where

$$a'_s = \mathbb{1}_{s \in V(t)} \sum_{u \in V(t)} \mathbb{1}_{u \sim s, u > s},$$

(5.3.17)

$$b'_s = \sum_{u, v \in V(t)} \mathbb{1}_{u \sim v} \mathbb{1}_{s \in (v, u)}.$$  

(5.3.18)

**Proof** We start by analyzing the conditional law of $B_r(o)$ given all $(\psi_v)_{v \in [n]}$. After this, we will take an expectation w.r.t. $\psi_v$ for $v \notin B_{r-1}(o)$ to get to the claim.

**Computing the conditional law of $B_r(o)$ given $(\psi_v)_{v \in [n]}$**

We start by introducing some useful notation. Define, for $u > v$, the edge probability

$$p(u, v) = \psi_v \prod_{s \in (v, u]} (1 - \psi_s).$$

(5.3.19)

We first condition on all $(\psi_v)_{v \in [n]}$ to obtain, for a trees $t$,

$$P_n(B_r(o_1) = t) = \prod_{v \in V(t)} \psi_v^{a'_v} \prod_{s=1}^n (1 - \psi_s)^{b'_v} \times \prod_{v \in V(t)} \prod_{u, j : u \sim v} [1 - p(u, v)],$$

(5.3.20)

where the first term is due to all the required edges in $B_r(o) = t$, while the second term is due to all other edges that are not allowed to be there.
The no-further-edge probability

We continue by analyzing the second line in (5.3.20), which for simplicity, we call the no-further-edge probability. First of all, since we are exploring the \( r \)-neighborhood of \( o \), the only edges that we need to require not to be there are of the form \( u \overset{j}{\not\rightarrow} v \) where \( v \in V^r(t) \) and \( u > v \), i.e., younger vertices than those in \( V^r(t) \) that do not form edges in \( t \). Since \( p(u, v) \) is small for all \( v \in V(t) \) and \( u > v \), and there are only finitely many elements in \( V^r(t) \), we can approximate

\[
\prod_{v \in V(t)} \prod_{u,j: u \not\overset{j}{\rightarrow} v} [1 - p(u, v)] = (1 + o_r(1)) \prod_{v \in V^r(t)} \prod_{u,j: u > v} [1 - p(u, v)].
\]

(5.3.21)

Recall (5.3.61). We can approximate

\[
\prod_{u,j: u > v} [1 - p(u, v)] = e^{\Theta(1)} \sum_{u,j: u \in (v,n]} p(u,v)^2 \exp(\sum_{u,j: u > v} p(u,v)).
\]

(5.3.22)

We compute

\[
\sum_{u,j: u \in (v,n]} p(u,v) = m\psi_v \sum_{u \in (v,n]} \prod_{s \in (v,u]} (1 - \psi_s)
\]

\[
= m\psi_v \sum_{u \in (v,n]} S_u^{(n)} \frac{S_v}{S_u^{(n)}}.
\]

We recall that \( n^{\chi}S_u^{(n)} \overset{a.s.}{\rightarrow} Ms^{-\chi} \). Further, \( u \mapsto S_u^{(n)} \) is decreasing, and \( s \mapsto Ms^{-\chi} \) is continuous, so that also

\[
\sum_{u \in (sn,n]} n^{\chi}S_u^{(n)} \overset{a.s.}{\rightarrow} \int_s^1 t^{-\chi} s^{\chi} ds.
\]

(5.3.24)

We conclude that

\[
\frac{m}{sn\psi_{sn}} \sum_{u \in (sn,n]} p(u,v) \overset{a.s.}{\rightarrow} \frac{m}{s} \int_s^1 t^{-\chi} s^{\chi} ds = \frac{m}{1 - \chi} [1 - s^{1-\chi}]
\]

\[
= (2m + \delta)[1 - s^{1-\chi}].
\]

(5.3.25)

Since \( (\psi_v)_{v \in V^r(t)} \) converges in distribution to a sequence of independent Gamma random variables,

\[
\sum_{u,j: u \in (v,n]} p(u,v)^2 \overset{p}{\rightarrow} 0.
\]

(5.3.26)

Therefore,

\[
\prod_{v \in V(t)} \prod_{u,j: u \not\overset{j}{\rightarrow} v} [1 - p(u, v)]
\]

\[
= (1 + o_r(1)) \prod_{v \in V^r(t)} e^{-(2m + \delta)\psi_v(1 - (v/n)^{\chi})}.
\]

(5.3.27)
5.3 Proof local convergence for preferential attachment models

Conclusion of the proof

We next take the expectation w.r.t. \( \psi_s \) for all \( s \not\in V^v(t) \) to obtain

\[
P(B_v(o) = t \mid (\psi_j)_{j \in [n] \setminus V^v(t)}) = (1 + o(1)) \prod_{v \in V^v(t)} \psi_v^{s'} e^{-(2m + \delta) \psi_v (1 - (v/n)^{1-\chi}) (1 - \psi_v) v'}
\]

\[
\times \prod_{s \in [n] \setminus V^v(t)} \frac{(\alpha + \alpha'_{s} - 1) \alpha'_{s} (\beta_{s} + b'_{s} - 1) \psi'_{s}}{(\alpha \beta_{s} + \alpha'_{s} + b'_{s} - 1) \alpha'_{s} + b'_{s}}
\]

as required. We further note that \( a'_{s} \in \{0, 1\} \) for all \( s \in [n] \) and that \( a'_{s} = 0 \) unless \( s \in V(t) \), so that \( a'_{s} = 1 \) can only occur for \( s \in \partial V(t) \).

Exploration of \( \text{PA}^{(m, d)}(d) \): one-neighborhood

In this step, we discuss how we can couple the one-neighborhood of a uniformly chosen vertex to that in \( \mathcal{T} \), so as to prove Theorem 5.8 for \( r = 1 \). This will serve as the starting point for an inductive analysis, and will give insight into the method of proof. The main result states that we can couple \( B_1^{(n)}(o_n) \) in \( \text{PA}^{(m, d)}(d) \) and \( B_1(\emptyset) \) in \( \mathcal{T} \) in such a way that all the variables involved are close:

**Lemma 5.21** (Coupling one-neighborhood of uniform vertex in \( \text{PA}^{(m, d)}(d) \) and Pólya-point tree) Fix \( \varepsilon > 0 \). Let \( o_n \) be chosen uniformly at random from \([n]\).

Let

\[
v_{o_n, 1}, \ldots, v_{o_n(m+q_{o_n})}
\]

be the neighbors of \( o_n \), ordered such that the edge \( \{o_n, v_{o_n(1)}\} \) was created before the edge \( \{o_n, v_{o_n(i+1)}\} \). Then, there exist \( \eta > 0 \) and \( N = N(\varepsilon) < \infty \) such that, with probability at least \( 1 - \varepsilon \), \( B_1^{(n)}(o_n) \), where \( o_n = v_{o_n} \), together with the variables \( (S_{v_{o_n}})_{i \geq 1} \) and \( B_1(\emptyset) \), together with its positions \( (X_{\emptyset})_{i \geq 1} \), can be coupled such that

(i) \( B_1^{(n)}(o_n) \simeq B_1(\emptyset) \) and \( |B_1(\emptyset)| \leq K \);

(ii) \( |X_{\emptyset} - S_{v_{o_n}}^{(n)}| \leq \eta \) for all \( k \) for which \( o_k \in B_1(\emptyset) \);

(iii) \( v_{o_n, 1}, \ldots, v_{o_n(m+q_{o_n})} \) are all distinct and \( v_{o_n,k} \geq \eta n \) for all \( v_{o_n,k} \in B_1^{(n)}(o_n) \);

(iv) \( \Gamma_{o_k} \leq K \) for all \( k \) such that \( o_k \in B_1(\emptyset) \).

Consequently, \( \mathbb{P}(B_1^{(n)}(o_n) \simeq t) \to \mathbb{P}(B_1(\emptyset) \simeq t) \).

We start with Lemma 5.21, as this proof will share many ingredients with the proof for the general \( r \geq 1 \), while its notation is relatively simple. Below, we will extend it to general \( r \geq 1 \).

**Proof** We start by proving (i)–(ii). Choose \( U_{\emptyset} \) uniform in \([0, 1]\), let \( X_{\emptyset} = U_{\emptyset}^X \), and let \( X_{\emptyset, 1}, \ldots, X_{\emptyset, (m+q_{\emptyset})} \) be the positions of the children of \( \emptyset \) in \((\mathcal{T}, \emptyset)\). Define \( o_n = \lceil nU_{\emptyset} \rceil \), and define \( v_{\emptyset, i} \) for \( i \in [m] \) by

\[
S_{v_{\emptyset}, i}^{(n)} \leq \frac{X_{\emptyset, i}}{X_{\emptyset}} S_{v_{\emptyset}}^{(n)} \leq S_{v_{\emptyset}}^{(n)}.
\]
By Theorem 5.10 and the observation that $U_{v\emptyset} = X_{\emptyset}/X_{\emptyset}$ is a collection of i.i.d. uniform random variables on $[0, 1]$, it follows that whp $\nu_{\emptyset} \in n[X_{\emptyset} - \delta, X_{\emptyset} + \delta]$, as required. In more detail, given $\varepsilon > 0$, choose $\delta > 0$ and $K > 0$ such that the regularity properties in Lemma 5.19 hold for $r = 1$. By Lemma 5.18, for $n$ large enough,

$$|\hat{S}^{(\nu)}_{v\emptyset} - X_{\emptyset}| \leq \eta, \quad \text{and} \quad |\hat{S}^{(\nu)}_{v\emptyset_i} - X_{\emptyset_i}| \leq \eta \forall i \in [m],$$

(5.3.30)

with probability at least $1 - 2\varepsilon$.

We continue by investigating the limiting distribution of the remaining neighbors $v_{\emptyset 1}, \ldots, v_{\emptyset (m+q(\nu))}$. Along the way, we will prove that $q^{(\nu)}_{\emptyset} = q_{\emptyset}$ whp. Note that by Theorem 5.10 and conditionally on $(\psi_k)_{k \in [n]}$, each vertex $v > v_{\emptyset}$ has $m$ independent chances of being connected to $v_{\emptyset}$. Let $X_{v,i}$ be the vertex to which the $i$th edge of $v$ is attached. Then, the events $\{X_{v,i} = v_{\emptyset}\}$ for $i \in [m]$ are, conditionally on $(\psi_k)_{k \in [n]}$, independent events with success probability

$$P_{v \rightarrow v_{\emptyset}} = \frac{(X_{v,i})}{\sum_{v_{\emptyset}}^{(r-1)}} = \frac{X_{v,i}}{\sum_{v_{\emptyset}}^{(r-1)}},$$

(5.3.31)

where the latter equality follows by (5.2.21).

Denote

$$N_{v\emptyset}(y) = \sum_{v=v_{\emptyset}+1}^{n} \sum_{i=1}^{m} 1_{\{X_{v,i} = v_{\emptyset}\}}.$$  

(5.3.32)

Our aim is to show that, conditionally on $U_{\emptyset}$, $(N_{v\emptyset}(y))_{y \in v_{\emptyset}/n}$ converges to a Poisson process on $[U_{\emptyset}, 1]$. For this, we first analyze the asymptotics of the success probabilities $P_{v \rightarrow v_{\emptyset}}$. By Lemma 5.19, $v_{\emptyset} \geq n\eta$ with probability at least $1 - \varepsilon$. Recall that $\psi_k = f_k(\alpha_k)$, where $\alpha_k$ has a Gamma distribution with parameters $\lambda = 1$ and $r = m + \delta$. Thus, we may apply Lemmas 5.17 and 5.18, which yield

$$(1 - \eta)\hat{P}_{v \rightarrow v_{\emptyset}} \leq P_{v \rightarrow v_{\emptyset}} \leq (1 + \eta)\hat{P}_{v \rightarrow v_{\emptyset}},$$

(5.3.33)

where

$$\hat{P}_{v \rightarrow v_{\emptyset}} = \frac{X_{v_{\emptyset}}}{(2m + \delta)n v_{\emptyset}} \frac{n}{v_{\emptyset}} \left(\frac{v_{\emptyset}}{v}\right)^x.$$

(5.3.34)

Letting $\tilde{N}_{v_{\emptyset}}(y) = \sum_{v=v_{\emptyset}+1}^{n} \sum_{i=1}^{m} 1_{\{X_{v,i} = v_{\emptyset}\}}$, where $(\tilde{X}_{v,i})_{v,i}$ are conditionally independent given $(\psi_k)_{k \geq 1}$ (or, equivalently, given $(\alpha_k)_{k \geq 1}$, we can couple $\tilde{N}_{v_{\emptyset}}(y)$ to a Poisson random variable with parameter

$$\frac{X_{v_{\emptyset}}}{2(1 + \delta/m)U_{\emptyset}} \int_0^y \left(\frac{U_{\emptyset}}{s}\right)^x \frac{x}{\psi s^{x-1}} \frac{U_{\emptyset}}{\psi s^{x-1}} ds,$$

(5.3.35)

on $[U_{\emptyset}, 1]$ such that the random variables are distinct with vanishing probability. Using that $\chi_{v_{\emptyset}} = \Gamma_{v_{\emptyset}}/(2m + \delta)$ proves that $\tilde{N}_{v_{\emptyset}}(x)$ converges to a Poisson random variable with parameter

$$\frac{X_{v_{\emptyset}}}{U_{\emptyset}} \int_0^x \psi s^{x-1} ds = \frac{\Gamma_{v_{\emptyset}}}{U_{\emptyset}} \int_0^x \psi s^{x-1} ds,$$

(5.3.36)
Let \( \rho \) be a distribution with density \( \gamma \). We assume that \( \Gamma_\omega = \chi_{\omega} \leq K \), that \( q_\omega^{(\omega)} = q_\omega \) and that \( (v_{\omega k})_{k \in \{m+q_\omega\}} \) are all distinct with \( \min_{k \in \{m+q_\omega\}} v_{\omega k} \geq \eta \). Let \( E \) denote the event that \( v_\omega \) is the uniform random vertex and that the neighbors of \( v_\omega \) are \( (v_{\omega k})_{k \in \{m+q_\omega\}} \). Let \( \chi_{\omega} \) denote the conditional distribution of \( (\chi_k)_{k \in [n]} \) conditioned on \( \Gamma_\omega = \gamma_\omega \) and \( E \). Further, let \( (\hat{\chi}_k)_{k \in [n]} \) denote an independent collection of random variables with distribution

\[
\hat{\chi}_{\omega k} \sim Y_{\omega k}^{*} \quad \forall k \in [m], \quad \hat{\chi}_k \sim Y_k \quad \forall k \in [n] \setminus [m].
\]  

(5.3.38)

Let \( \rho(\cdot \mid E, \chi_{\omega} = \gamma_\omega) \) denote the density of the (multi-dimensional) random variable \( (\hat{\chi}_k)_{k \in [n]} \), and let \( P \) be the joint distribution of \( PA_{m,\omega}^{d} \) and the random variables \( (\chi_k)_{k \in [n]} \). By Bayes' Theorem,

\[
\rho(\cdot \mid E, \chi_{\omega} = \gamma_\omega) = \frac{P(E \mid \cdot, \chi_{\omega} = \gamma_\omega)}{P(E \mid \chi_{\omega} = \gamma_\omega)} \rho_\omega(\cdot),
\]  

(5.3.39)

where \( \rho_\omega(\cdot) \) is the original density of the random variables \( (\chi_k)_{k \in [n] \setminus \{v_\omega\}} \). We denote the corresponding probability distributions and expectations as \( P_\omega \) and \( E_\omega \), respectively. We conclude that we have to compare \( P(\mathcal{E} \mid \cdot, \chi_{\omega}) \) to \( P(\mathcal{E} \mid \cdot, \chi_{\omega}) \). Since we are interested in the distribution of \( (\chi_k)_{k \in [n]} \), this amounts to comparing \( P(\mathcal{E} \mid (\chi_k)_{k \in [n]}) \) to \( P(\mathcal{E} \mid \chi_{\omega}) \). For this, we need to compute \( P(\mathcal{E} \mid (\chi_k)_{k \in [n]}) \) (and note that \( P(\mathcal{E} \mid \chi_{\omega}) = E_\omega[P(\mathcal{E} \mid (\chi_k)_{k \in [n]})] \)). This is our next step.
By Theorem 5.10, we can explicitly compute

$$P(E \mid (\chi_k)_{k \in [n]}) = m! \prod_{i=1}^{m} P_{v_{i}} \rightarrow v_{\varphi}, \prod_{j=1}^{q_{\varphi}} mP_{v_{\varphi}j} \rightarrow v_{\varphi} \left(1 - P_{v_{\varphi}j} \rightarrow v_{\varphi}\right)^{m-1} \quad (5.3.40)$$

$$\times \prod_{v > v_{\varphi} : v \notin \{v_{\varphi}(m+1), \ldots, v_{\varphi}(m+q)\}} \left(1 - P_{v \rightarrow v_{\varphi}}\right)^{m},$$

by the conditional independence given $(\psi_{k})_{k \in [n]}$, which is equivalent to conditional independence given $(\chi_{k})_{k \in [n]}$. We rewrite

$$P(E \mid (\chi_{k})_{k \in [n]}) = m! \prod_{i=1}^{m} P_{v_{i}} \rightarrow v_{\varphi}, \prod_{j=1}^{q_{\varphi}} mP_{v_{\varphi}j} \rightarrow v_{\varphi} \prod_{v > v_{\varphi}} \left(1 - P_{v \rightarrow v_{\varphi}}\right)^{m}, \quad (5.3.41)$$

where $P_{v \rightarrow v'}$ is given, and using (5.2.21), by

$$P_{v \rightarrow v'} = \frac{\psi_{v'}^{(v-1)}}{\psi_{v}^{(v)}} = \frac{S_{v'}^{(v)}}{S_{v}^{(v)}}, \quad (5.3.42)$$

We now perform asymptotics of this formula. We use Lemma 5.18 to show that we can for every $\eta > 0$, there exists $n_0$ sufficiently large, such that, with probability at least $1 - \varepsilon$,

$$(1 - \eta)P(E \mid (\chi_{k})_{k \in [n]}) \leq m! \prod_{i=1}^{m} \psi_{v_{i}}(\frac{v_{i}}{v_{\varphi}}) \prod_{j=1}^{q_{\varphi}} m\psi_{v_{\varphi}j}(\frac{v_{\varphi}j}{v_{\varphi}}) e^{-m\psi_{v_{\varphi}} \sum_{v > v_{\varphi}} (\frac{v_{\varphi}}{v})} \leq (1 + \eta)P(E \mid (\chi_{k})_{k \in [n]}). \quad (5.3.43)$$

To bound $P(E \mid \chi_{v_{\varphi}}) = E_{\varphi}[P(E \mid (\chi_{k})_{k \in [n]})]$, we also need a good bound on “bad” event, for which we use the deterministic upper bound

$$P(E \mid (\chi_{k})_{k \in [n]}) \leq m! \prod_{i=1}^{m} P_{v_{i}} \rightarrow v_{\varphi}, \prod_{j=1}^{q_{\varphi}} mP_{v_{\varphi}j} \rightarrow v_{\varphi} \leq m! (m\psi_{v_{\varphi}})^{\eta} \prod_{i=1}^{m} \psi_{v_{\varphi}i} \quad (5.3.44)$$

$$\leq C' m! \prod_{i=1}^{m} \psi_{v_{i}}(\frac{v_{i}}{v_{\varphi}}) \prod_{j=1}^{q_{\varphi}} m\psi_{v_{\varphi}j}(\frac{v_{\varphi}j}{v_{\varphi}}) e^{-m\psi_{v_{\varphi}} \sum_{v > v_{\varphi}} (\frac{v_{\varphi}}{v})},$$

where $C' = \eta^{-(m+N)} \sup_{n \geq 1} e^{mnf_{\varphi}(K)} < \infty$.

We conclude that with probability at least $1 - \varepsilon/2$ with respect to $P_{\varphi}$,

$$\sqrt{1 - \eta} \prod_{i=1}^{m} \frac{\psi_{v_{\varphi}i}}{E_{\varphi}[\psi_{v_{\varphi}i}]} \leq \frac{P(E \mid (\chi_{k})_{k \in [n]})}{P(E \mid \chi_{\varphi})} \leq \sqrt{1 + \eta} \prod_{i=1}^{m} \frac{\psi_{v_{\varphi}i}}{E_{\varphi}[\psi_{v_{\varphi}i}]} \quad (5.3.45)$$

With the help of Lemma 5.17, with probability at least $1 - \varepsilon$,

$$\left(1 - \eta\right) \prod_{i=1}^{m} \frac{\chi_{v_{\varphi}i}}{E_{\varphi}[\chi_{v_{\varphi}i}]} \leq \frac{P(E \mid (\chi_{k})_{k \in [n]})}{P(E \mid \chi_{\varphi})} \leq \left(1 + \eta\right) \prod_{i=1}^{m} \frac{\chi_{v_{\varphi}i}}{E_{\varphi}[\chi_{v_{\varphi}i}]} \quad (5.3.46)$$

The factors $\chi_{v_{\varphi}i}/E_{\varphi}[\chi_{v_{\varphi}i}]$ correspond to the size-biasing of $\hat{\chi}_{v_{\varphi}k}$ for $k \in [m]$ as
5.3 Proof local convergence for preferential attachment models

indicated in (5.3.38). Since this is of product structure, asymptotic independence of $(\hat{\chi}_{v_k})_{k \in [n] \setminus \{v_o\}}$ remains.

We conclude that with probability at least $1 - \varepsilon$ with respect to $\mathbb{P}_\omega$,

$$(1 - \eta) \hat{\rho}(\cdot) \leq \rho(\cdot | \mathcal{E}, \chi_{v_o} = \gamma_{\omega}) \leq (1 + \eta) \hat{\rho}(\cdot),$$

(5.3.47)

where $\hat{\rho}$ is the density of $(\hat{\chi}_{v_k})_{k \in [n] \setminus \{v_o\}}$, whose law we denote by $\hat{\mathbb{P}}$.

To complete the proof, we use [Volume 1, (2.2.8)] to compute that

$$d_{\text{TV}}(\rho(\cdot | \mathcal{E}, \chi_{v_o}), \hat{\rho}(\cdot)) = 1 - \int \rho(\vec{\chi} | \mathcal{E}, \chi_{v_o}) \wedge \hat{\rho}(\vec{\chi}) d\vec{\chi}. \tag{5.3.48}$$

Let $\Omega_\omega$ denote the where (5.3.47) holds, so that $\mathbb{P}_\omega(\Omega | \mathcal{E}, \chi_{v_o}) \geq 1 - \eta$. Then, we bound

$$d_{\text{TV}}(\rho(\cdot | \mathcal{E}, \chi_{v_o}), \hat{\rho}(\cdot)) \leq 1 - \int_{\Omega_\omega} \rho(\vec{\chi} | \mathcal{E}, \chi_{v_o}) \wedge \hat{\rho}(\vec{\chi}) d\vec{\chi}. \tag{5.3.49}$$

On $\Omega_\omega$, (5.3.47) holds, so that $\hat{\rho}(\vec{\chi}) \geq \rho(\vec{\chi} | \mathcal{E}, \chi_{v_o})/(1 + \eta)$. Thus,

$$d_{\text{TV}}(\rho(\cdot | \mathcal{E}, \chi_{v_o}), \hat{\rho}(\cdot)) \leq 1 - \frac{1}{1 + \eta} \int_{\Omega_\omega} \rho(\vec{\chi} | \mathcal{E}, \chi_{v_o}) \leq 1 - \frac{1 - \eta}{1 + \eta} = \frac{2\eta}{1 + \eta} \leq 2\eta. \tag{5.3.50}$$

We conclude that

$$d_{\text{TV}}(\rho(\cdot | A, \chi_{v_o}), \hat{\rho}(\cdot)) \leq 2\eta. \tag{5.3.51}$$

Choosing $\eta$ so small that $2\eta \leq \varepsilon$, this shows that we can couple $(\chi_k^{E, \gamma_{\omega}})_{k \in [n] \setminus \{v_o\}}$ and $(\hat{\chi}_k)_{k \in [n] \setminus \{v_o\}}$ such that they are equal with probability at least $1 - \varepsilon$, as required. It is straightforward to show that (iv) holds for the random variables $(\hat{\chi}_k)_{k \in [n] \setminus \{v_o\}}$ as these are independent Gamma distributions. This completes the proof of (iv).

The fact that $\mathbb{P}(B_{1}^{(\alpha)}(o_n) \simeq t) \rightarrow \mathbb{P}(B_{1}(\varnothing) \simeq t)$ follows directly from the fact that we have coupled such that $q_{\omega}^{(\alpha)} = q_{\omega}$, and this occurs with probability at least $1 - \varepsilon$. Since $\varepsilon > 0$ is arbitrary, the statement follows.

### Exploration of $\text{PA}^{(m, \delta)}(d)$: general neighborhood

In this step, we discuss how we can explore the one-neighborhood of a uniformly chosen vertex, so as to prove Theorem 5.8.

**Proposition 5.22** (Coupling general neighborhood of uniform vertex in $\text{PA}^{(m, \delta)}(d)$ and Polya-point tree) Fix $\varepsilon > 0$ and $r \geq 1$. Let $o_n$ be chosen uniformly at random from $[n]$. Let $v_w$ be the vertices that are within distance $r$ from $o_n$, ordered such that the edge $\{v_w, v_{w+1}\}$ was created before the edge $\{v_w, v_{w+i+1}\}$. Then, there exist $\eta > 0$ and $N = N(\varepsilon) < \infty$ such that, with probability at least $1 - \varepsilon$, $B_{r}^{(\alpha)}(o_n)$, where $o_n = v_o$, together with the variables $(S_{v_o}^{(\alpha)})_{|w| \leq r}$, and $B_r(\varnothing)$, together with its positions $(X_w)_{|w| \leq 1}$, can be coupled such that

1. $B_{r}^{(\alpha)}(o_n) \simeq B_r(\varnothing)$ and $|B_r(\varnothing)| \leq K$;
200 Connected components in preferential attachment models

\( \sum_{v_w \in B^{(r)}_{r - 1}(o_n)} \varphi^{(v_w)} \leq \frac{C}{n} \)  

for some \( C > 0 \). This implies that conditioning on \( v_w \not\in B^{(r - 1)}_{r - 1}(o_n) \) has only a negligible effect on the distribution of the children of \( v_w \). We thus proceed as in the proof of Lemma 5.21 to obtain a coupling between a sequence of i.i.d. values \( (X_w)_{i \in [m]} \) of type \( L \) and the children \( (v_w)_{i \in [m]} \). As before, we have that \( |S^{(r - 1)}_{v_w} - X_w| \leq \eta \), for \( n \) large enough, with probability at least \( 1 - \varepsilon \).

We can continue with this process for all \( v_w \in \partial B^{(r - 1)}_{r - 1}(o_n) = B^{(r - 1)}_{r - 1}(o_n) \setminus B^{(r - 2)}_{r - 2}(o_n) \), thus producing the set \( L_r \) consisting of all the type \( L \) children of vertex \( v_w \in B^{(r - 1)}_{r - 1}(o_n) \). It is not hard to see that with probability tending to 1 as \( n \to \infty \), the set \( L_r \) has no intersection with \( B^{(r - 1)}_{r - 1}(o_n) \), so we will assume this in the remainder of the proof.

Next, we continue with the children of type \( R \) of vertices in \( \partial B_r(\emptyset) \). Assume that we have already determined the children of type \( R \) of the vertices
5.3 Proof local convergence for preferential attachment models

in \( U_{r-1} \subseteq B_{r-1}^{(n)}(o_n) \setminus B_{r-1}^{(n)}(o_n) \). Denote the set of children obtained so far by \( R(U_{r-1}) \). We decompose this set as \( R(U_{r-1}) = \bigcup_{i=1}^{m} R^{(i)}(U_r) \), where \( R^{(i)}(U_r) = \{ v \in R_r : X_{v,i} \in U_{r-1} \} \). Now consider a vertex \( v_w \in \partial B_{r-1}^{(n)}(o_n) \setminus U_{r-1} \). Conditioning on the graph obtained so far is again not difficult, and now amounts to two conditions:

1. \( X_{v,i} \neq v_w \) if \( v \in B_{r-1}^{(n)}(o_n) \cup U_{r-1} \), since all the edges coming out of this set have already been determined;
2. For \( v \notin B_{r-1}^{(n)}(o_n) \cup U_{r-1} \), the probability that \( v_w \) receives the \( i \)-th edge from \( k \) is different from the probability given in (5.3.42), since the random variable \( X_{v,i} \) has been probed or checked before. Indeed, we know that \( X_{v,i} \notin B_{r-2}^{(n)}(o_n) \), since otherwise \( v \) would have sent out an edge to a vertex in \( B_{r-2}^{(n)}(o_n) \), which would mean that \( v \) would have been a child of type \( R \) in \( B_{r-1}^{(n)}(o_n) \), which gives a contradiction. We also know that \( X_{v,i} \notin U_{r-1} \), since otherwise \( v \in R^{(i)}(U_r) \).

Instead of (5.3.42), (2) means that we have to use the modified probability

\[
\tilde{P}_{v \rightarrow v'} = \frac{\varphi_v^{(v-1)}}{\tilde{S}_{v-1}^{(v-1)}},
\]

where

\[
\tilde{S}_{v-1}^{(v-1)} = \sum_{u > v-1: u \notin B_{r-1}^{(n)}(o_n) \cup U_r} \varphi_u^{(v-1)}.
\]

Since, by the induction hypothesis, \( S_{v-1}^{(v-1)} \leq \tilde{S}_{v-1}^{(v-1)} \leq \tilde{S}_{v-1}^{(v-1)} + C/n \), for some \( C > 0 \), we can still infer Lemma 5.18 to approximate \( \tilde{P}_{v \rightarrow v'} \) by \( \tilde{P}_{v \rightarrow v'} \), which equals

\[
\tilde{P}_{v \rightarrow v'} = \frac{1}{n} \frac{\chi_{v'}}{2m + \delta \sum_{v' \notin U_r} (v')^{-\gamma}}.
\]

From here onwards, the proof is a straightforward adaptation of the proof of Lemma 5.21, and we omit further details.

The fact that \( \mathbb{P}(B_r^{(n)}(o_n) \simeq (t)) \rightarrow \mathbb{P}(B_r(\emptyset) \simeq (t)) \) follows since we have coupled the neighborhoods such that with probability at least \( 1 - \varepsilon \), \( q_n^{(n)} = q_w \) for all \( |u| \leq r \). Since \( \varepsilon > 0 \) is arbitrary, the statement follows.

Exploration of PA\(^{(m,\delta)}(d)\): two neighborhoods

In this part, we argue that

\[
\mathbb{P}(B_r^{(n)}(o_1^{(n)}), B_r^{(n)}(o_2^{(n)}) \simeq t) \rightarrow \mathbb{P}(B_r(\emptyset) \simeq t)^2,
\]

where \( o_1^{(n)}, o_2^{(n)} \) are two vertices chosen uniformly at random and independently from \([n]\).

A second moment method then immediately shows that, for every rooted tree \( t \),

\[
\frac{1}{n} \sum_{v \in [n]} \mathbb{P}(B_r^{(n)}(o_v) \simeq t) \rightarrow \mathbb{P}(B_r(\emptyset) \simeq t),
\]
which completes the proof of the local weak convergence in probability in Theorem 5.8.

Incorporate Proposition 5.23 more fluently in the text!

Again, in what follows, we regard $B_r(o_i)$ as a rooted graph, where the vertices receive labels in $[n]$, and the edges receive labels in $[m]$ corresponding to the label of the directed edge that gives rise to the edge (in either possible direction). Thus, the edge $\{u,v\}$ receives label $j$ when $u \xrightarrow{j} v$ or when $v \xrightarrow{j} u$.

Let $t_1$ and $t_2$ be two rooted vertex- and edge-labelled trees of height exactly $r$, where the vertex labels are in $[n]$ and the edge labels in $[m]$. We write $B_r(o_i) = t_i$ to denote that the vertices, edges and edge labels in $B_r(o_i)$ are given by those in $t_i$. Note that this is rather different from $B_r(o_1) \simeq t_1$ as defined in Definition 2.3, where $t$ is unlabeled and we investigate the isomorphism of $B_r(o_i)$ and $t_i$.

Let $V(t_i)$ denote the vertex labels in $t_i$. Also, let $\partial V(t_i)$ denote the vertices at distance exactly $r$ from the root of $t_i$, and let $V^\circ(t_i) = V(t_i) \setminus \partial V(t_i)$ denote the restriction of $t_i$ to all vertices at distance at most $r - 1$ from its root. With this notation in hand, we have the following characterization of the conditional law of $B_r(o_1)$ and $B_r(o_2)$, which is a generalization of Proposition 5.20 to two neighborhoods:

**Proposition 5.23** (Law of neighborhoods in $\text{PA}^{(m,\delta)}_n(d)$) Fix $m \geq 2$ and $\delta > -m$, and consider $\text{PA}^{(m,\delta)}_n(d)$. Let $t_1$ and $t_2$ be two rooted vertex- and edge-labelled trees with disjoint vertex sets and roots $o_1$ and $o_2$, respectively. Then, as $n \to \infty$,

$$
P(B_r(o_1) = t_1, B_r(o_2) = t_2 \mid (\psi_v)_{v \in V^\circ(t_1) \cup V^\circ(t_2)}) = (1 + o(1)) \prod_{v \in V^\circ(t_1) \cup V^\circ(t_2)} \psi_v e^{-(2m+\delta)\psi_v(1-(v/n)^{1-\delta})}(1 - \psi_v)^v' \times \prod_{s \in [n] \setminus (V^\circ(t_1) \cup V^\circ(t_2))} \frac{(\alpha + a'_v - 1)\alpha'_v + b'_v - 1)}{(\alpha + \alpha'_v + a'_v + b'_v - 1)a'_v + b'_v},
$$

where

$$
a'_v = \mathbb{1}_{\{s \in V(t_1) \cup V(t_2)\}} \sum_{u \in V(t_1) \cup V(t_2)} \mathbb{1}_{\{u \sim s, u > s\}},
$$

$$
b'_v = \sum_{u, v \in V(t_1) \cup V(t_2)} \mathbb{1}_{\{u \sim v\}} \mathbb{1}_{\{s \in (v, u)\}}.
$$

**Proof** The proof follows the steps in the proof of Proposition 5.20. We start by analyzing the conditional law of $B_r(o_1)$ and $B_r(o_2)$ given all $(\psi_v)_{v \in [n]}$. After this, we will take an expectation w.r.t. $\psi_v$ for $v \notin B_{r-1}(o_1) \cup B_{r-1}(o_2)$ to get to the claim.
5.3 Proof local convergence for preferential attachment models

Computing the conditional law of \( B_r(o_1) \) and \( B_r(o_2) \) given \((\psi_v)_{v \in [n]}\)

We start by introducing some useful notation. Define, for \( u > v \), the edge probability

\[
p(u, v) = \psi_v \prod_{s \in (v, u]} (1 - \psi_s).
\]  
(5.3.61)

We first condition on all \((\psi_v)_{v \in [n]}\) to obtain, for two trees \( t_1 \) and \( t_2 \),

\[
\mathbb{P}_n(B_r(o_1) = t_1, B_r(o_2) = t_2) = \prod_{v \in V(t_1) \cup V(t_2)} \psi_v^n \prod_{s=1}^n (1 - \psi_s)^{v_s} \\
\times \prod_{u,v \in V(t_1) \cup V(t_2)} \prod_{u,j \notin u \to v} [1 - p(u, v)],
\]  
(5.3.62)

where the first term is due to all the required edges in \( B_r(o_1) = t_1, B_r(o_2) = t_2 \),

while the second term is due to all other edges that are not allowed to be there.

The no-further-edge probability

We continue by analyzing the second line in (5.3.62), which for simplicity, we call the no-further-edge probability. First of all, since we are exploring the \( r \)-neighborhoods of \( o_1 \) and \( o_2 \), the only edges that we need to require not to be there are of the form \( u \to v \) where \( v \in V^r(t_1) \cup V^r(t_2) \) and \( u > v \), i.e., younger vertices than those in \( V^r(t_1) \cup V^r(t_2) \) that do not form edges in \( t_1 \) and \( t_2 \). Since \( p(u, v) \) is small for all \( v \in V(t_1) \cup V(t_2) \) and \( u > v \), and there are only finitely many elements in \( V^r(t_1) \cup V^r(t_2) \), we can approximate

\[
\prod_{v \in V^r(t_1) \cup V^r(t_2)} \prod_{u,j : u \to v} [1 - p(u, v)] = (1 + o_r(1)) \prod_{v \in V^r(t_1) \cup V^r(t_2)} \prod_{u,j : u > v} [1 - p(u, v)].
\]  
(5.3.63)

Recall (5.3.61). We can approximate

\[
\prod_{u,j : u > v} [1 - p(u, v)] = e^{\Theta(1)} \sum_{u,j : u \in (v, n]} p(u, v)^2 \exp(\sum_{u,j : u > v} p(u, v)).
\]  
(5.3.64)

We compute

\[
\sum_{u,j : u \in (v, n]} p(u, v) = m\psi_v \sum_{u \in (v, n]} \sum_{s \in (v, u]} (1 - \psi_s)
\]  
(5.3.65)

\[
= m\psi_v \sum_{u \in (v, n]} \frac{S^{(n)}_u}{S^{(n)}/v},
\]

We recall that \( n^x S^{(n)}_u \xrightarrow{a.s.} M s^{-x} \). Further, \( u \mapsto S^{(n)}_u \) is decreasing, and \( s \mapsto M s^{-x} \) is continuous, so that also

\[
\sum_{u \in (s, n]} n^x S^{(n)}_u \xrightarrow{a.s.} \int_s^1 t^{-x} s^\lambda dt.
\]  
(5.3.66)
We conclude that
\[
\frac{m}{sn_2} \sum_{u \in \{sn_1, n\}} p(u, v) \xrightarrow{n \to \infty} \frac{m}{s} \int_1^s t^{-\gamma} ds = \frac{m}{1 - \gamma} [1 - s^{1 - \gamma}] \tag{5.3.67}
\]
\[
= (2m + \delta)[1 - s^{1 - \gamma}].
\]

Since \((v_1, s_1) \in V^\gamma(t_1) \cup V^\gamma(t_2)\) converges in distribution to a sequence of independent Gamma random variables,
\[
\sum_{u,j: u \in \{v, n\}} p(u, v)^2 \xrightarrow{a.s.} 0. \tag{5.3.68}
\]

Therefore,
\[
\prod_{v \in V(t_1) \cup V(t_2)} \prod_{u,j: u \in \{v, n\}} [1 - p(u, v)] = (1 + o_v(1)) \prod_{v \in V^\gamma(t_1) \cup V^\gamma(t_2)} e^{-(2m + \delta)v^{1-(v/n)^1-\gamma)}.
\tag{5.3.69}
\]

Conclusion of the proof
We next take the expectation w.r.t. \(\psi_s\) for all \(s \notin V^\gamma(t_1) \cup V^\gamma(t_2)\) to obtain
\[
P(B_r(o_1) = t_1, B_r(o_2) = t_2 \mid (\psi_j)_{j \in [n] \setminus (V^\gamma(t_1) \cup V^\gamma(t_2))}) \tag{5.3.70}
\]
\[
= (1 + o_v(1)) \prod_{v \in V^\gamma(t_1) \cup V^\gamma(t_2)} \psi_v^{b_v} e^{-(2m + \delta)v^{1-(v/n)^1-\gamma}} \prod_{s \in [n] \setminus (V^\gamma(t_1) \cup V^\gamma(t_2))} \frac{(\alpha + a'_s - 1)u'_s (\beta'_s + b'_s - 1)\delta'_s}{(\alpha + \beta'_s + a'_s + b'_s - 1)\alpha'_s + \beta'_s},
\]

as required. We further note that \(a'_s \in \{0,1\}\) for all \(s \in [n]\) and that \(a'_s = 0\) unless \(s \in V(t_1) \cup V(t_2)\), so that \(a'_s = 1\) can only occur for \(s \in \partial V(t_1) \cup \partial V(t_2)\) \(\square\)

Completion of proof
We note that
\[
P(B_r^{(\alpha)}(a^{(\alpha)}_n) \cap B_r^{(\alpha)}(a^{(\alpha)}_n) \neq \emptyset) = P(a^{(\alpha)}_n \notin B_{2r}^{(\alpha)}(a^{(\alpha)}_n) = 1 - o(1), \tag{5.3.71}
\]
since \(|B_{2r}^{(\alpha)}(a^{(\alpha)}_n)|\) is a tight sequence of random variables. Thus, it suffices to show that
\[
P(B_r^{(\alpha)}(a^{(\alpha)}_n), B_r^{(\alpha)}(a^{(\alpha)}_n) \simeq t, B_r^{(\alpha)}(a^{(\alpha)}_n) \cap B_r^{(\alpha)}(a^{(\alpha)}_n) = \emptyset) \to P(B_r(\emptyset) \simeq t)^2. \tag{5.3.72}
\]
To see (5.3.66), we condition on\(B_r^{(\alpha)}(a^{(\alpha)}_n), which is such that B_r(\emptyset^{(\alpha)}), (X_u)_{u \in B_r(\emptyset^{(\alpha)})} (\Gamma_u)_{u \in B_r(\emptyset^{(\alpha)})}, (\chi_u)_{u \in B_r(\emptyset^{(\alpha)})} and (\upsilon_u)_{u \in B_r(\emptyset^{(\alpha)})} such that Proposition 5.22 holds. This occurs whp, as Proposition 5.22. Consider
\[
P(B_r^{(\alpha)}(a^{(\alpha)}_n) \simeq t \mid B_r^{(\alpha)}(a^{(\alpha)}_n), a^{(\alpha)}_n \notin B_{2r}^{(\alpha)}(a^{(\alpha)}_n)). \tag{5.3.73}
\]
We adapt the proof of Proposition 5.22. Conditioning on $B^{(n)}(o^{(1)}_n), o^{(2)}_n \notin B^{(n)}_r(o^{(1)}_n)$ has the effect that (5.3.53) is now modified to

$$\tilde{P}_{v \rightarrow v'} = \frac{\tilde{\varphi}^{(s-1)}_{u \rightarrow v}}{\tilde{S}^{(s-1)}_{u-1}},$$

where now

$$\tilde{S}^{(s-1)}_{u-1} = \sum_{u > v-1: u \notin B^{(n)}_r(o^{(1)}_n) \cup U_r \cup B^{(n)}_r(o^{(1)}_n)} \varphi^{(s-1)}_{u}.$$

This again makes very little difference, so that indeed, for $B^{(n)}_r(o^{(1)}_n)$, which is such that $B^{(n)}_r(\emptyset) = \emptyset$, $(X_w)_{w \in B^{(n)}_r(\emptyset)}$, $(\Gamma_w)_{w \in B^{(n)}_r(\emptyset)}$, $(\chi_v)_{v \in B^{(n)}_r(\emptyset)}$ and $(v_w)_{w \in B^{(n)}_r(\emptyset)}$ such that Proposition 5.22 holds,

$$P(B^{(n)}_r(o^{(2)}_n) \simeq t \mid B^{(n)}_r(o^{(1)}_n), o^{(2)}_n \notin B^{(n)}_r(o^{(1)}_n)) \xrightarrow{p} P(B_r(\emptyset) \simeq t),$$

as required. \qed

### 5.3.1 Local weak convergence of related models

In this section, we discuss the local weak convergence of two related models. The main result is the following theorem:

**Theorem 5.24** (Local weak convergence of preferential attachment models: fixed edges) Fix $m \geq 1$ and $\delta > -m$. The preferential attachment models $PA^{(m,\delta)}_n$ and $PA^{(m,\delta)}(d)$ converge locally weakly in probability to the Pólya-point tree.

### 5.3.2 Local weak convergence Bernoulli preferential attachment

Recall the Bernoulli preferential attachment model ($BPA^{(f)}_t$) in Section 1.3.6. A special case is the ($BPA^{(f)}_t$) with an affine attachment function $f$, i.e., the setting where there exist $\gamma, \beta > 0$ such that

$$f(k) = \gamma k + \beta.$$  

Due to the attachment rules in (1.3.64), the model does not satisfy the rescaling property that the model with $cf$ has the same law as the model with $f$, for any $c > 0$. In fact, it turns out that the parameter $\gamma > 0$ (which is, by convention always taken to be 1 for ($PA^{(m,\delta)}_t$) is now the parameter that determines the tail behavior of the degree distribution (recall Exercise 1.12). In Exercises 5.21-5.22, you are asked to compute the average degree of this affine model, as well as the number of edges added at time $t$ for large $t$.

In this section, we investigate the local weak limit of Bernoulli preferential attachment models, as introduced in Section 1.3.6. The main result is as follows:

**Theorem 5.25** (Local weak convergence of preferential attachment models: conditionally independent edges) Fix $\delta \geq 0$. The preferential attachment model
Connected components in preferential attachment models

with conditionally independent edges converges locally weakly in probability to the Pólya-point graph with mixed Poisson distributions.

TO DO 5.1: Add local weak convergence for BPAM

5.4 Connectivity of preferential attachment models

In this section we investigate the connectivity of \((\text{PA}_t^{(m,\delta)})_{t \geq 1}\). We start by describing the connectivity when \(m = 1\), which is special. For \(m = 1\), the number of connected components of \(\text{PA}_t^{(1,\delta)}\) \(N_t\) has distribution given by

\[
N_t = I_1 + I_2 + \cdots + I_t,
\]

where \(I_i\) is the indicator that the \(i\)th edge connects to itself, so that \((I_i)_{i \geq 1}\) are independent indicator variables with

\[
P(I_i = 1) = \frac{1 + \delta}{(2 + \delta)(i - 1) + 1 + \delta}.
\]

It is not hard to see that this implies that \(N_t / \log t\) converges in probability to \((1 + \delta)/(2 + \delta) < 1\), so that whp there exists a largest connected component of size at least \(t/\log t\). As a result, whp \(\text{PA}_t^{(1,\delta)}\) is not connected, but has few connected components which are almost all quite large. We do not elaborate more on the connectivity properties for \(m = 1\) and instead leave the asymptotics of the number of connected components as Exercise 5.23.

For \(m \geq 2\), the situation is entirely different since then \(\text{PA}_t^{(m,\delta)}\) is connected whp at sufficiently large times:

**Theorem 5.26** (Connectivity of \(\text{PA}_t^{(m,\delta)}\) for \(m \geq 2\)) Fix \(m \geq 2\). Then, with high probability for \(T\) large, \(\text{PA}_t^{(m,\delta)}\) is connected for all \(t \geq T\).

**Proof of Theorem 5.26.** Again, we let \(N_t\) denote the number of connected components of \(\text{PA}_t^{(m,\delta)}\). We note that, \(I_t = N_t - N_{t-1} = 1\) precisely when all \(m\) edges of vertex \(t\) are attached to vertex \(t\). Thus

\[
P(I_t = 1) = \prod_{e=1}^{m} \frac{2e - 1 + \delta}{(2m + \delta)t + (2e - 1 + \delta)}.
\]

For \(m \geq 2\),

\[
\sum_{t=2}^{\infty} P(I_t = 1) < \infty,
\]

so that, almost surely, \(I_t = 1\) only occurs finitely often. As a result, \(\lim_{t \to \infty} N_t < \infty\) almost surely since \(N_t \leq 1 + \sum_{t=2}^{\infty} I_t\). This implies that, for \(m \geq 2\), \(\text{PA}_t^{(m,\delta)}\) almost surely contains only finitely many connected components. However, \(\text{PA}_t^{(m,\delta)}\) has a positive probability of being disconnected at a certain time \(t \geq 2\) (see Exercise 5.27 below). However, for \(m \geq 2\), \(I_t = N_t - N_{t-1}\) can also be negative, since
the edges of the vertex \( v_t \) can be attached to two distinct connected components. We will see that this happens with high probability, which explains why \( N_t = 1 \) whp for \( t \) large, as we next show.

We first fix \( K \geq 1 \) large. Then, with probability converging to 1 as \( K \to \infty \),
\[
\sum_{t=K}^{\infty} \mathbb{1}_{\{N_t > N_{t-1}\}} = 0.
\]
so that no new connected components are formed after time \( K \), and the number of connected components can only decrease in time. Let \( \mathcal{F}_s \) denote the \( \sigma \)-algebra generated by \( (\text{PA}_t^{(m,\delta)})_{s=1}^{t} \). We are left to prove that for \( t \) sufficiently large, the vertices \( 1, \ldots, K \) are whp all connected in \( \text{PA}_t^{(m,\delta)} \). This proof proceeds in two steps. We show that, if \( N_t \geq 2 \) and \( t \) is large, then
\[
P(\sum_{t=K}^{2t} \mathbb{1}_{\{N_t > N_{t-1}\}} = 0) \leq \varepsilon t,
\]
for some \( \varepsilon > 0 \). Thus, the probability that this happens for at least one \( s \in [2t] \setminus [t] \) is at least
\[
1 - \left( 1 - \frac{\varepsilon}{t} \right)^t,
\]
which is uniformly positive for every \( t \). Thus, \( \mathbb{P}(N_{2t} - N_t \leq -1 \mid \mathcal{F}_K, N_t \geq 2) \) is uniformly bounded from below. Indeed, conditionally on \( \mathcal{F}_K, N_t \geq 2 \), and using \( N_t \leq K \), \( \text{PA}_t^{(m,\delta)} \) must have one connected component of size at least \( t/K \). Every other component has at least one vertex in it, and its degree is at least \( m \). Fix \( s \in [2t] \setminus [t] \). This means that the probability that the first edge of \( v_s^{(m)} \) connects to the connected component of size at least \( t/K \), while the second connects to the connected component of size at least 1, conditionally on \( \mathcal{F}_t \) is at least
\[
\frac{m + \delta}{2(2m + \delta)t} \geq \frac{\varepsilon}{t},
\]
for some \( \varepsilon > 0 \). Thus, the probability that this happens for at least one \( s \in [2t] \setminus [t] \) is at least
\[
1 - \left( 1 - \frac{\varepsilon}{t} \right)^t,
\]
which is uniformly positive for every \( t \). Thus, \( \mathbb{P}(N_{2t} - N_t \leq -1 \mid \mathcal{F}_K, N_t \geq 2) \) is uniformly bounded from below. As a result, \( N_t \xrightarrow{a.s.} 1 \), so that \( N_T = 1 \) for some \( T < \infty \) a.s. Without loss of generality, we can take \( T \geq K \). When
\[
\sum_{t=K}^{\infty} \mathbb{1}_{\{N_t > N_{t-1}\}} = 0,
\]
if \( N_T = 1 \) for some \( T \), then \( N_t = 1 \) for all \( t \geq T \). This proves that \( \text{PA}_T^{(m,\delta)} \) is whp connected for all \( t \geq T \), where \( T \) is large, which implies Theorem 5.26.

### 5.4.1 Giant component for Bernoulli preferential attachment

The preferential attachment model \( \text{PA}_n^{(m,\delta)} \) turns out to be connected whp, as we discuss in more detail in the next section. This, however, is not true for the preferential attachment model with conditionally independent edges as defined in Section 1.3.6. Here, we describe the existence of the giant component in this model:

**Theorem 5.27** (Existence of a giant component: linear case) If \( f(k) = \gamma k + \beta \) for some \( 0 \leq \gamma < 1 \) and \( 0 < \beta \leq 1 \), then there exists a giant component if and only if
\[
\gamma \geq \frac{1}{2} \quad \text{or} \quad \beta > \frac{(1/2 - \gamma)^2}{1 - \gamma}.
\]

The notation used by Dereich and Mörters (2013) is slightly different from ours.
Dereich and Mörters (2009) prove that their model obeys an asymptotic power-

law with exponent \( \tau = 1 + 1/\gamma = 3 + \delta/m \), so that \( \gamma \) intuitively corresponds to \( \gamma = m/(2m + \delta) \). As a result, \( \gamma \geq \frac{1}{2} \) corresponds to \( \delta \leq 0 \), which is also precisely the setting where the configuration model always has a giant component (recall Theorem 4.4).

The more involved case of more general attachment functions \( k \mapsto f(k) \) is more delicate to describe. We start by introducing some notation, following Dereich and Mörters (2013). We call a preferential attachment function \( f: \{0, 1, 2, \ldots\} \to (0, \infty) \) concave when \( f(0) \leq 1 \) and \( \Delta f(k) := f(k + 1) - f(k) < 1 \) forall \( k \geq 0 \).

The operator \( A_\alpha \) should be thought of as describing the expected offspring of vertices of different types, as explained in more detail below. The main result on the existence of a giant component in the preferential attachment model with conditionally independent edges is the following theorem:

**Theorem 5.28** (Existence of a giant component) **No giant component exists if and only if there exists** \( 0 < \alpha < 1 \) **such that** \( A_\alpha \) **is a compact operator with spectral radius** \( \rho(A_\alpha) \leq 1 \).

It turns out that \( A_\alpha \) is a well-defined compact operator (Dereich and Mörters, 2013, Lemma 3.1) if and only if \( (A_\alpha 1)(0) < \infty \). When thinking of \( A_\alpha \) as the reproduction operator, the spectral radius \( \rho(A_\alpha) \) describes whether the multi-
type branching process has a positive survival probability. Thus, \( \rho(A_\alpha) \) should be thought of as the equivalent of the usual condition \( \mathbb{E}[X] \leq 1 \) for extinction of a discrete single-type branching process.
5.5 Further results for preferential attachment models

5.5 FURTHER RESULTS FOR PREFERENTIAL ATTACHMENT MODELS

TO DO 5.3: Add further results!

5.6 Notes and discussion

Notes on Section 5.1
The proof of Theorem 5.2 is adapted from Ross (1996). More recent discussions on exchangeable random variables and their properties can be found in Aldous (1985) and Pemantle (2007), the latter focussing on random walks with self-interaction, where exchangeability is a crucial tool. There is a lot of work on urn schemes, also in cases where the weight functions are not linear with equal slope, in which case the limits can be seen to obey rather different characteristics. See e.g., (Athreya and Ney, 1972, Chapter 9).

Notes on Section 5.2
The multitype branching process local weak limit in Theorem 5.8 has been established by Berger, Borgs, Chayes and Saberi (2014) for preferential attachment models with a fixed number of outgoing edges per vertex. Berger et al. (2014) only treat the case where \( \delta \geq 0 \), the more recent extension to \( \delta > -m \) is novel. This is due to the fact that Berger et al. (2014) view the attachment probabilities as a mixture between attaching uniformly and according to the degree. We, instead, rely on the Pólya urn description that works for all \( \delta > -m \). In this case, Theorem 5.8 is (Berger et al., 2014, Theorem 2.2). Theorem 5.8 states local weak convergence in probability to the Pólya-point tree, while (Berger et al., 2014, Theorem 2.2) states local weak convergence in distribution. Local weak convergence in probability can be deduced from the convergence in probability of subgraph counts in (Berger et al., 2014, Lemma 2.4). We refrain from discussing this issue further.

The proof of Theorem 5.10 follows (Berger et al., 2014, Section 3.1) closely, apart from the fact that we do not rely on the relation to a mixture of choosing a vertex uniformly and according to degree.

Berger et al. (2014) also study two related settings, one where the edges are attached independently (i.e., without the intermediate update of the degrees while attaching the \( m \) edges incident to the newest vertex), and the conditional model in which the edges are attached to distinct vertices. This shows that the result is quite robust, and hopefully also applies to the related settings of \( \text{PA}_n^{(m, \delta)} \) and \( \text{PA}_n^{(m, \delta)}(b) \).

A related version of Theorem 5.10 for \( \delta = 0 \) was proved by Bollobás and Riordan (2004a) in terms of a pairing representation. This applies to the model \( \text{PA}_n^{(m, \delta)} \) with \( \delta = 0 \). Another related version of Theorem 5.10 is proved by Rudas et al. (2007), which applies to general preferential attachment functions with \( m = 1 \) and relies on a continuous-time embedding in terms of continuous-time branching processes. We refer to Section 5.5 for more details.
Theorem 5.25 is proved by Dereich and Mörters (2009, 2011, 2013) for the Bernoulli preferential attachment model.

Notes on Section 5.4
The result on the giant component for preferential attachment models with conditionally independent edges in Theorem 5.28 is proved by Dereich and Mörters (2009, 2011, 2013).

Notes on Section 5.5
The embedding results in terms of continuous-time branching processes can be found in Rudas et al. (2007), as well as Athreya (2007); Athreya et al. (2008).

5.7 Exercises for Chapter 5

Exercise 5.1 (Stationarity of exchangeable sequences) Show that when \((X_i)_{i=1}^n\) are exchangeable, then the marginal distribution of \(X_i\) is the same as that of \(X_1\). Show also that the distribution of \((X_i, X_j)\), for \(j \neq i\), is the same as that of \((X_1, X_2)\).

Exercise 5.2 (I.i.d. sequences are exchangeable) Show that when \((X_i)_{i \geq 1}\) are i.i.d., they form an infinite sequence of exchangeable random variables.

Exercise 5.3 (Limiting density in De Finetti’s Theorem (Theorem 5.2)) Use De Finetti’s Theorem (Theorem 5.2) to prove that \(S_n/n \Rightarrow U\). Use this to prove (5.1.4).

Exercise 5.4 (The number of ones in \((X_i)_{i=1}^n\)) Prove (5.1.3).

Exercise 5.5 (Positive correlation of exchangeable random variables) Let \((X_i)_{i \geq 1}\) be an infinite sequence of exchangeable random variables. Prove that
\[
P(X_k = X_n = 1) \geq P(X_k = 1)P(X_n = 1). \tag{5.7.1}
\]
Prove that equality holds if and only if there exists a \(p\) such that \(P(U = p) = 1\).

Exercise 5.6 (Limiting density of mixing distribution for Pólya urn schemes) Prove that (5.1.23) proves (5.1.4).

Exercise 5.7 (Uniform recursive trees) A uniform recursive tree is obtained by starting with a single vertex, and successively attaching the \((n+1)\)st vertex to a uniformly chosen vertex in \([n]\). Prove that for uniform recursive trees the tree decomposition in Theorem 5.4 is such that
\[
\frac{S_1(n)}{S_1(n) + S_2(n)} \Rightarrow U, \tag{5.7.2}
\]
where \(U\) is uniform on \([0, 1]\). Use this to prove that \(P(S_1(n) = k) = 1/n\) for each \(k \in [n]\).
5.7 Exercises for Chapter 5

Exercise 5.8 (Scale-free trees) Recall the model studied in Theorem 5.4, where at time \( t = 2 \), we start with two vertices of which vertex 1 has degree \( d_1 \) and vertex 2 has degree \( d_2 \). After this, we successively attach vertices to older vertices with probability proportional to the degree plus \( \delta > -1 \) as in (1.3.62). Show that the model for \((\text{PA}_{(1,\delta)}(b))_{n \geq 1}\) for which the graph at time \( n = 2 \) consists of two vertices joined by two edges, arises when \( d_1 = d_2 = 2 \). What does Theorem 5.4 imply for \((\text{PA}_{(1,\delta)}(b))_{n \geq 1}\)?

Exercise 5.9 (Relative degrees of vertices 1 and 2) Use Theorem 5.6 to compute \( \lim_{t \to \infty} \mathbb{P}(D_2(n) \geq xD_1(n)) \) for \((\text{PA}_{(1,\delta)}(b))_{n \geq 1}\).

Exercise 5.10 (Proof of Theorem 5.7) Complete the proof of Theorem 5.7.

Exercise 5.11 (Size-biased version of Gamma) Let \( X \) have a Gamma distribution with parameter \( r \). Show that its size-biased version \( X^\star \) has a Gamma distribution with parameter \( r + 1 \).

Exercise 5.12 (Power-law exponents in \( \text{PA}_{(m,\delta)}(d) \)) Prove the power-law relations in (5.2.10) and identify \( c_m,\delta \) and \( c'_m,\delta \).

Exercise 5.13 (Joint law \((D,D')\) for \( \text{PA}_{(m,\delta)}(d) \) (Berger et al., 2014, Lemma 5.3))
Adapt the proof of Lemma 5.9 to show that, for \( j \geq m \) and \( k \geq m + 1 \),

\[
\mathbb{P}(D = j, D' = k) = \frac{2m + \delta}{m^2} \frac{\Gamma(k + 1 + \delta)}{k!\Gamma(m + 1 + \delta)} \frac{\Gamma(j + \delta)}{j!\Gamma(m + \delta)} \times \int_0^1 \int_0^1 (1 - u)^{k-1}u^{m+1+\delta/m}(1 - v)^{j-1}v^{m+\delta}dudv.
\]

Exercise 5.14 (Joint law \((D,D')\) for \( \text{PA}_{(m,\delta)}(d) \) (Berger et al., 2014, Lemma 5.3))
Use Lemma 5.9 and Exercise 5.13 to show that, for fixed \( j \geq m \) and as \( k \to \infty \),

\[
\mathbb{P}(D' = k \mid D = j) = C_j k^{-(2+\delta/m)}(1 + O(1/k)).
\]

Exercise 5.15 (Joint law \((D,D')\) for \( \text{PA}_{(m,\delta)}(d) \) (Berger et al., 2014, Lemma 5.3))
Use Lemma 5.9 and Exercise 5.13 to show that, for fixed \( k \geq m + 1 \) and as \( j \to \infty \),

\[
\mathbb{P}(D = j \mid D' = k) = \tilde{C}_j j^{-(m+4+\delta/m)}(1 + O(1/j)).
\]

Exercise 5.16 (Simple form of \( S_k \) in (5.2.21))
Prove, using induction on \( k \), that \( S_k = \prod_{i=k+1}^n (1 - \psi_i) \) in (5.2.21) holds.

Exercise 5.17 (Multiple edges and Theorem 5.10) Fix \( m = 2 \). Let \( M_t \) denote the number of multiple edges in \( \text{PA}_{(m,\delta)}(d) \). Use Theorem 5.10 to show that

\[
\mathbb{E}[M_{t+1} - M_t] = \sum_{k=1}^t \mathbb{E} \left[ \left( \frac{\varphi_k}{\psi_{k-1}} \right)^2 \right].
\]

Exercise 5.18 (Multiple edges and Theorem 5.10 (cont.)) Fix \( m = 2 \). Let \( M_t \)
denote the number of multiple edges in $\text{PA}_t^{(m, \delta)}(d)$. Compute $\mathbb{E} \left[ \left( \frac{\varphi_k}{S_{t-1}} \right)^2 \right]$ and use this to show that $\mathbb{E}[M_t]/\log t \to c$, and identify $c > 0$.

**Exercise 5.19** (Multiple edges and Theorem 5.10 (cont.)) Fix $m = 2$. Let $M_t$ denote the number of multiple edges in $\text{PA}_t^{(m, \delta)}(d)$. Use Theorem 5.10 to show that, conditionally on $(\psi_k)_{k \geq 1}$, the sequence $(M_{t+1} - M_t)_{t \geq 2}$ is an independent sequence with

$$
\mathbb{P}(M_{t+1} - M_t = 1 \mid (\psi_k)_{k \geq 1}) = \sum_{k=1}^{t-1} \left( \frac{\varphi_k}{S_{t-1}} \right)^2.
$$

(5.7.7)

**Exercise 5.20** (Almost sure limit of $F_k$ in (5.2.22)) Fix $k \geq 1$. Prove that $(M_n(k))_{n \geq k+1}$, where

$$
M_n(k) = \prod_{j=k+1}^{n} \frac{1 - \psi_j}{\mathbb{E}[1 - \psi_j]},
$$

is a multiplicative positive martingale and using the Martingale Convergence Theorem ([Volume 1, Theorem 2.24]) and thus converges. Conclude that (5.2.22) holds by showing that $\prod_{j=k+1}^{n} \mathbb{E}[1 - \psi_j] = (k/n)^\chi(1 + o(1))$.

**Exercise 5.21** (Recursion formula for total edges in affine $\text{BPA}_t^{(f)}$) Consider the affine $\text{BPA}_t^{(f)}$ with $f(k) = \gamma k + \beta$. Derive a recursion formula for $\mathbb{E}[|E(\text{BPA}_t^{(f)})|]$, where we recall that $|E(\text{BPA}_t^{(f)})|$ is the total number of edges in $\text{BPA}_t^{(f)}$. Identify $\mu$ such that $\mathbb{E}[|E(\text{BPA}_t^{(f)})|]/t \to \mu$.

**Exercise 5.22** (Number of edges per vertex in affine $\text{BPA}_t^{(f)}$) Consider the affine $\text{BPA}_t^{(f)}$ with $f(k) = \gamma k + \beta$. Assume that $|E(\text{BPA}_t^{(f)})|/t \xrightarrow{d} \mu$. Show that $D_t(n) \xrightarrow{d} \text{Poi}(\mu)$.

**Exercise 5.23** (CLT for number of connected components for $m = 1$) Show that the number of connected components $N_t$ in $\text{PA}_t^{(1, \delta)}$ satisfies a central limit theorem with equal asymptotic mean and variance given by

$$
\mathbb{E}[N_t] = \frac{1 + \delta}{2 + \delta} \log t (1 + o(1)), \quad \text{Var}(N_t) = \frac{1 + \delta}{2 + \delta} \log t (1 + o(1)).
$$

(5.7.9)

**Exercise 5.24** (Number of connected components for $m = 1$) Use Exercise 5.23 to show that the number of connected components $N_t$ in $\text{PA}_t^{(1, \delta)}$ satisfies $N_t/\log t \xrightarrow{p} (1 + \delta)/(2 + \delta)$.

**Exercise 5.25** (Number of self-loops in $\text{PA}_t^{(m, \delta)}$) Fix $m \geq 1$ and $\delta > -m$. Use a similar analysis as in Exercise 5.23 to show that the number of self-loops $S_t$ in $\text{PA}_t^{(m, \delta)}$ satisfies $S_t/\log t \xrightarrow{p} m(m + 1 + 2\delta)/[2(2m + \delta)]$.

**Exercise 5.26** (Number of self-loops in $\text{PA}_t^{(m, \delta)}(b)$) Fix $m \geq 1$ and $\delta > -m$. Use a similar analysis as in Exercise 5.25 to show that the number of self-loops $S_t$ in $\text{PA}_t^{(m, \delta)}(b)$ satisfies $S_t/\log t \xrightarrow{p} (m - 1)(m + 2\delta)/[2(2m + \delta)]$. 
Exercise 5.27 (All-time connectivity for $(\mathbb{PA}^{(m,d)}_t)_{t \geq 1}$) Fix $m \geq 2$. Compute the probability that $(\mathbb{PA}^{(m,d)}_t)_{t \geq 1}$ is connected for all times $t \geq 1$, and show that this probability is in $(0, 1)$. 


Part III

Small-world properties
of random graphs
Summary of Part II.
So far, we have considered the simplest connectivity properties possible. We have
focused on the degrees in Volume 1, and the existence and uniqueness of a macro-
scopic connected component in Part II of this book. We can summarize the results
obtained in the following meta theorem:

Meta Theorem A. (Existence and uniqueness of giant component) In a ran-
dom graph model with power-law degrees having power-law exponent $\tau$, there is a
unique giant component when $\tau \in (2,3)$ and there is a unique giant component
when the graph has average degree that exceeds a certain precise threshold when
$\tau > 3$.

The above means, informally, that the giant component is quite robust to
random removal of edges when $\tau \in (2,3)$, while it is not when $\tau > 3$. These
results make the general philosophy that ‘random graphs with similar degree
characteristics should behave alike’ precise, at least at the level of the existence
of a giant component.

Overview of Part III.
In this part, we aim to extend the discussion of similarity of random graphs
to their small-world characteristics, by investigating distances within the giant
component. We focus both on their typical distances, which means the graph
distances between most pair of vertices as characterized by the graph distance
between two uniformly chosen vertices conditioned on being connected, as well
as on their maximal distances as characterized by their diameters.

In more detail, this part is organised as follows. We study distances in gen-
eral inhomogeneous random graphs in Chapter 6, and those in the configuration
model, as well the closely related uniform random graph with prescribed degrees,
in Chapter 7. In the last chapter of this part, Chapter 8, we study distances in
the preferential attachment model.
Chapter 6

SMALL-WORLD PHENOMENA IN INHOMOGENEOUS RANDOM GRAPHS

Abstract

In this chapter, we investigate the small-world structure in rank-1 inhomogeneous random graphs. For this, we develop path-counting techniques that are interesting in their own right.

TO DO 6.1: Add case of distances in real-world network...

In this chapter, we investigate the small-world properties of inhomogeneous random graphs.

Organization of this chapter

We start in Section 6.1 by discussing results on the small-world phenomenon in inhomogeneous random graphs. We state results for general inhomogeneous random graphs, and then specialize to rank-1 inhomogeneous random graphs. Only for the rank-1 case, we provide complete proofs. For the general case, we instead explain the intuition and give proofs in special cases. In Section 6.2, we prove lower bounds on typical distances. In Section 6.3, we prove the corresponding upper bounds in the log log regime, and in Section 6.4, we discuss path-counting techniques to show the log upper bound for \( \tau > 3 \). In Section 6.5.1, we discuss the diameter of inhomogeneous random graphs. In Section 6.5, we discuss related results for distances in inhomogeneous random graphs. We close this chapter with notes and discussion in Section 6.6, and with exercises in Section 6.7.

6.1 Small-world effect in inhomogeneous random graphs

In this section, we consider the distances between vertices of IRG\(_n(\kappa_n)\) where, as usual, \((\kappa_n)\) is a graphical sequence of kernels with limit \(\kappa\).

Recall that we write dist\(_G(i, j)\) for the graph distance between the vertices \(i, j \in [n]\) in a graph \(G\), where the graph distance is the minimum number of edges in the graph \(G\) that form a path from \(i\) to \(j\), and where, by convention, we let dist\(_G(i, j) = \infty\) when \(i, j\) are in different connected components. We define the typical graph distance to be dist\(_G(o_1, o_2)\), where \(o_1, o_2\) are two vertices that are chosen uniformly at random from the vertex set \([n]\).

It is possible that no path connecting \(o_1\) and \(o_2\) exists, in which case we define dist\(_{IRG_n(\kappa_n)}(o_1, o_2) = \infty\). By Theorem 3.16, \(\mathbb{P}(\text{dist}_{IRG_n(\kappa_n)}(o_1, o_2) = \infty) \to 1 - \zeta^2 > 0\), since \(\zeta < 1\) (see Exercise 3.32). In particular, when \(\zeta = 0\), which is equivalent to \(\nu \leq 1\), \(\mathbb{P}(\text{dist}_{IRG_n(\kappa_n)}(o_1, o_2) = \infty) \to 1\). Therefore, in our main results, we shall condition on \(o_1\) and \(o_2\) to be connected, and only consider cases where \(\zeta > 0\).
Logarithmic asymptotics of typical graph distance in \( \text{IRG}_n(\kappa_n) \)

We start by discussing logarithmic asymptotics of the typical graph distance in the case where \( \nu = \|T_n\| \in (1, \infty) \). When \( \|T_n\| = \infty \), instead, then our results also prove that \( \text{dist}_{\text{IRG}_n(\kappa_n)}(o_1, o_2) = o_o(\log n) \), but they do not tell us much about their exact asymptotics.

The main result on typical graph distances in \( \text{IRG}_n(\kappa_n) \) is as follows:

**Theorem 6.1** (Typical distances in \( \text{IRG}_n(\kappa_n) \)) Let \( (\kappa_n) \) be graphical sequence of kernels with limit \( \kappa \), and with \( \nu = \|T_n\| \in (1, \infty) \). Let \( \varepsilon > 0 \) be fixed. Then, for \( \text{IRG}_n(\kappa_n) \),

(i) If \( \sup_{x,y,n} \kappa_n(x, y) < \infty \), then

\[
\mathbb{P}(\text{dist}_{\text{IRG}_n(\kappa_n)}(o_1, o_2) \leq (1 - \varepsilon) \log \nu, n) = o(1). \tag{6.1.1}
\]

(ii) If \( \kappa \) is irreducible, then

\[
\mathbb{P}(\text{dist}_{\text{IRG}_n(\kappa_n)}(o_1, o_2) \leq (1 + \varepsilon) \log \nu, n) = \zeta_n^2 + o(1). \tag{6.1.2}
\]

In the terminology of [Volume 1, Section 1.4], Theorem 6.1(ii) implies that \( \text{IRG}_n(\kappa_n) \) is a small world when \( \kappa \) is irreducible and \( \nu = \|T_n\| < \infty \). Theorem 6.1(i) shows that the graph distances are of order \( \Theta(\log n) \) when \( \sup_{x,y,n} \kappa_n(x, y) < \infty \), so that \( \text{IRG}_n(\kappa_n) \) is not an ultra-small world.

The intuition behind Theorem 6.1 is that, by (3.4.6) and (??), a Poisson multi-type branching process with kernel \( \kappa \) has neighborhoods that grow exponentially, i.e., the number of vertices at distance \( k \) grows like \( \|T_n\|^k \). Thus, if we examine the distance between two vertices \( o_1 \) and \( o_2 \) chosen uniformly at random from \( [n] \), then we need to explore the neighborhood of vertex \( o_1 \) up to the moment that it ‘catches’ vertex \( o_2 \). In this case, the neighborhood must be of size proportional to \( n \), so that we need that \( \|T_n\|^k = \nu^k \sim n \), i.e., \( k = k_n \sim \log \nu, n \). However, proving such a fact is quite tricky, since there are far fewer possible further vertices to explore when the neighborhood has size proportional to \( n \). The proof overcomes this fact by exploring from the two vertices \( o_1 \) and \( o_2 \) simultaneously up to the first moment that these neighborhoods share a common vertex. At this moment, we have found the shortest path.

We next specialize to rank-1 inhomogeneous random graphs, where we also investigate in more detail what happens when \( \nu = \infty \) in the case where the degree power-law exponent \( \tau \) satisfies \( \tau \in (2, 3) \).

**Distances in rank-1 IRGs with finite-variance weights**

We continue by investigating the behavior of \( \text{dist}_{\text{NR}_n(w)}(o_1, o_2) \) for \( \text{NR}_n(w) \) in the case where the weights have finite variance:

**Theorem 6.2** (Typical distances in \( \text{NR}_n(w) \) for finite-variance weights) In the Norros-Reittu model \( \text{NR}_n(w) \), where the weights \( w = (w_i)_{i \in [n]} \) satisfy Condition 1.1(a)-(c) and where \( \nu > 1 \), conditionally on \( \text{dist}_{\text{NR}_n(w)}(o_1, o_2) < \infty \),

\[
\frac{\text{dist}_{\text{NR}_n(w)}(o_1, o_2)}{\log n} \xrightarrow{\nu} 1 / \log \nu. \tag{6.1.3}
\]
The same result applies, under the same conditions, to $\text{GRG}_n(w)$ and $\text{CL}_n(w)$.

Theorem 6.2 can be seen as a special case of Theorem 6.1. However, in Theorem 6.1(i), we require that $\kappa_n$ is a bounded kernel. In the setting of Theorem 6.2, this would imply that $\max_i w_i$ is uniformly bounded in $n$. Theorem 6.2 does not require this.

We give a complete proof of Theorem 6.2 in Sections 6.2 and 6.4 below. There, we will also use the ideas in the proof of Theorem 6.2 to give a proof of Theorem 6.1.

The intuition behind Theorem 6.2 is as follows. In Section 3.5, we have argued that the neighborhood of a uniform vertex in $\text{NR}_n(w)$ is well-approximated by a two-stage branching process, where the second and all later generations have offspring distribution $(p^*_k)_{k \geq 0}$, which is the probability mass function of the size-biased version minus one of the degree distribution $D$, where $D$ has probability mass function $(p_k)_{k \geq 0}$ as in (3.5.37). When $\nu = \sum_{k \geq 0} kp^*_k < \infty$, the number of vertices at distance $k$ is close to $M\nu^k$, where $M$ is the martingale limit of $Z_k/\nu^k$.

To know what $\text{dist}_{\text{NR}_n(w)}(o_1, o_2)$ is, we need to grow the neighborhoods from the first uniform vertex until we find the second uniform vertex. The latter happens with reasonable probability when $Z_k \approx n$, which suggests that the relevant $k$ is such that $\nu^k \approx n$, so that $k \approx \log \nu n$.

While the above heuristic is quite convincing, the argument is fatally flawed. Indeed, as already argued in Section 3.5, the neighborhoods of a uniform vertex are well-approximated by a branching process as long as the number of vertices found is much smaller than $n$. When the number of vertices found becomes of order $n$, the depletion-of-points effect has already started to kick in. Therefore, the above approach is doomed to fail. Our proof instead, is divided in a lower and an upper bound on the typical distance $\text{dist}_{\text{NR}_n(w)}(o_1, o_2)$. For the proof of the lower bound in Section 6.2.1, we show that the expected number of paths of $k$ edges between two uniform vertices is approximately $\nu^k/\ell_n$, so that such a path whp does not exist when $k \leq (1 - \varepsilon) \log \nu n$. For the proof of the upper bound in Section 6.4, we use a second moment method to show that, conditionally on the two uniformly chosen vertices being in the giant component, whp there exists a path of $(1 + \varepsilon) \log \nu_n n$ edges. This requires novel path-counting techniques in random graphs that are interesting in their own right. Exercise 6.1 investigate typical distance for the Erdős-Rényi random graph, as well as in the case where $\nu = \infty$.

Theorem 6.2 leaves open what happens when $\nu = \infty$. We can use Theorem 6.2 to show that $\text{dist}_{\text{NR}_n(w)}(o_1, o_2) = o_n(\log n)$, see Exercise 6.2 below. We next study the case where the weights have an asymptotic power-law distribution with $\tau \in (2, 3)$.

### Distances in rank-1 IRGs with infinite-variance weights

We continue to study typical distances in the Norros-Reittu random graph $\text{NR}_n(w)$, in the case where the degrees obey a power-law with degree exponent $\tau$ satisfying $\tau \in (2, 3)$. 

Theorem 6.2 can be seen as a special case of Theorem 6.1. However, in Theorem 6.1(i), we require that $\kappa_n$ is a bounded kernel. In the setting of Theorem 6.2, this would imply that $\max_i w_i$ is uniformly bounded in $n$. Theorem 6.2 does not require this.

We give a complete proof of Theorem 6.2 in Sections 6.2 and 6.4 below. There, we will also use the ideas in the proof of Theorem 6.2 to give a proof of Theorem 6.1.

The intuition behind Theorem 6.2 is as follows. In Section 3.5, we have argued that the neighborhood of a uniform vertex in $\text{NR}_n(w)$ is well-approximated by a two-stage branching process, where the second and all later generations have offspring distribution $(p^*_k)_{k \geq 0}$, which is the probability mass function of the size-biased version minus one of the degree distribution $D$, where $D$ has probability mass function $(p_k)_{k \geq 0}$ as in (3.5.37). When $\nu = \sum_{k \geq 0} kp^*_k < \infty$, the number of vertices at distance $k$ is close to $M\nu^k$, where $M$ is the martingale limit of $Z_k/\nu^k$.

To know what $\text{dist}_{\text{NR}_n(w)}(o_1, o_2)$ is, we need to grow the neighborhoods from the first uniform vertex until we find the second uniform vertex. The latter happens with reasonable probability when $Z_k \approx n$, which suggests that the relevant $k$ is such that $\nu^k \approx n$, so that $k \approx \log \nu n$.

While the above heuristic is quite convincing, the argument is fatally flawed. Indeed, as already argued in Section 3.5, the neighborhoods of a uniform vertex are well-approximated by a branching process as long as the number of vertices found is much smaller than $n$. When the number of vertices found becomes of order $n$, the depletion-of-points effect has already started to kick in. Therefore, the above approach is doomed to fail. Our proof instead, is divided in a lower and an upper bound on the typical distance $\text{dist}_{\text{NR}_n(w)}(o_1, o_2)$. For the proof of the lower bound in Section 6.2.1, we show that the expected number of paths of $k$ edges between two uniform vertices is approximately $\nu^k/\ell_n$, so that such a path whp does not exist when $k \leq (1 - \varepsilon) \log \nu n$. For the proof of the upper bound in Section 6.4, we use a second moment method to show that, conditionally on the two uniformly chosen vertices being in the giant component, whp there exists a path of $(1 + \varepsilon) \log \nu_n n$ edges. This requires novel path-counting techniques in random graphs that are interesting in their own right. Exercise 6.1 investigate typical distance for the Erdős-Rényi random graph, as well as in the case where $\nu = \infty$.

Theorem 6.2 leaves open what happens when $\nu = \infty$. We can use Theorem 6.2 to show that $\text{dist}_{\text{NR}_n(w)}(o_1, o_2) = o_n(\log n)$, see Exercise 6.2 below. We next study the case where the weights have an asymptotic power-law distribution with $\tau \in (2, 3)$.

### Distances in rank-1 IRGs with infinite-variance weights

We continue to study typical distances in the Norros-Reittu random graph $\text{NR}_n(w)$, in the case where the degrees obey a power-law with degree exponent $\tau$ satisfying $\tau \in (2, 3)$.
that \( \tau \in (2, 3) \). In this case, \( \nu = \infty \), so that \( \text{dist}_{\text{NR}_{n}(w)}(o_1, o_2) = o_{\nu}(\log n) \) (recall Exercise 6.2). In turns out that the precise scaling of \( \text{dist}_{\text{NR}_{n}(w)}(o_1, o_2) \) depends sensitively on the precise way how \( \nu_{n} \to \infty \). Below, we assume that they obey a power-law with exponent \( \tau \) satisfying that \( \tau \in (2, 3) \). We will later discuss what happens in related settings, for example when \( \tau = 3 \).

Many of our arguments also apply to the generalized random graph \( \text{GRG}_{n}(w) \) and the Chung-Lu model \( \text{CL}_{n}(w) \). In this section, we discuss the setting where the weights \( w \) are heavy tailed. Recall that \( F_n(x) \) denotes the proportion of vertices \( i \) for which \( w_i \leq x \). Then, we assume that there exists a \( \tau \in (2, 3) \) such that for all \( \delta > 0 \), there exists \( c_1 = c_1(\delta) \) and \( c_2 = c_2(\delta) \) such that, uniformly in \( n \),

\[
c_1 x^{-(\tau - 1 + \delta)} \leq [1 - F_n](x) \leq c_2 x^{-(\tau - 1 - \delta)},
\]

(6.1.4)

where the upper bound is expected to hold for every \( x \geq 1 \), while the lower bound is only required to hold for \( 1 \leq x \leq n^{\alpha} \) for some \( \alpha > 1/2 \).

The assumption in (6.1.4) is what we need precisely, and it states that \( [1 - F_n](x) \) obeys power-law bounds for appropriate values of \( x \). Note that the lower bound in (6.1.4) cannot be valid for all \( x \), since \( F_n(x) > 0 \) implies that \( F_n(x) \geq 1/n \), so that the lower and upper bound in (6.1.4) are contradicting when \( x \gg n^{1/(\tau - 1)} \). Thus, the lower bound can hold only for \( x = O(n^{1/(\tau - 1)}) \). When \( \tau \in (2, 3) \), we have that \( 1/(\tau - 1) \in (1/2, 1) \), and we only need the lower bound to hold for \( x \leq n^{\alpha} \) for some \( \alpha \in (1/2, 1) \). Exercises 6.3 and 6.4 give simpler conditions for (6.1.4) in special cases, such as i.i.d. weights.

The main result on graph distances in the case of infinite-variance weights is as follows:

**Theorem 6.3** (Typical distances in \( \text{NR}_{n}(w) \) for \( \tau \in (2, 3) \)) Fix the Norros-Reittu model \( \text{NR}_{n}(w) \), where the weights \( w = (w_i)_{i \in [n]} \) satisfy Conditions 1.1(a)-(b) and (6.1.4). Then, conditionally on \( \text{dist}_{\text{NR}_{n}(w)}(o_1, o_2) < \infty \),

\[
\frac{\text{dist}_{\text{NR}_{n}(w)}(o_1, o_2)}{\log \log n} \xrightarrow{\tau} \frac{2}{|\log (\tau - 2)|}.
\]

(6.1.5)

The same results apply, under the same conditions, to \( \text{CL}_{n}(w) \) and \( \text{GRG}_{n}(w) \).

Theorem 6.3 implies that \( \text{NR}_{n}(w) \) with \( w \) as in (1.3.15), for \( \tau \in (2, 3) \), is an ultra-small world when (6.2.23) is satisfied.

The main tool to study distances in \( \text{NR}_{n}(w) \) is a comparison to branching processes, which is particularly pretty for \( \text{NR}_{n}(w) \). In the next two sections, we prove Theorems 6.2–6.3. When \( \tau > 3 \), then the branching process approximation has finite mean, and we can make use of the martingale limit results of the number of individuals in generation \( k \) as \( k \to \infty \). When \( \tau \in (2, 3) \), on the other hand, the branching process has infinite mean. In this case, the number of individuals in generation \( k \), conditionally on survival, grows super-exponentially, which explains why distances grow doubly logarithmically. See Section 7.3, where this is explained in more detail in the context of the configuration model.
6.2 Lower bounds on typical distances in IRGs

The super-exponential growth implies that a path between two vertices typically passes through vertices with larger and larger weights as we move away from the two vertices. Thus, starting from the first vertex \( o_1 \in [n] \), the path connecting \( o_1 \) to \( o_2 \) uses vertices that first grow until the midpoint of the path is reached, and then decrease again to reach \( o_2 \). This can be understood by noting that the probability that a vertex with weight \( w \) is not connected to any vertex with weight larger than \( y > w \) in \( NR_n(w) \) is

\[
e^{-\sum_{i: w_i > y} w_i / \ell_n} = e^{-w[1 - F_n^\star(y)]},
\]  

(6.1.6)

where \( F_n^\star(y) = \sum_{i: w_i \leq y} w_i / \ell_n \) is the distribution function of \( W_n^\star \) introduced in (6.2.33). When (6.1.4) holds, it follows that \( [1 - F_n^\star](y) \) is close to \( y^{-(\tau - 2)} \), the size-biasing increasing the power by one. Therefore, the probability that a vertex with weight \( w \) is not connected to any vertex with weight larger than \( y > w \) in \( NR_n(w) \) is approximately \( e^{-wy^{-(\tau - 2)}} \). Take \( w \) large, then this probability is small when \( y \gg w^{1/(\tau - 2)} \). Thus, a vertex of weight \( w \) is whp connected to a vertex of weight \( w^{1/(\tau - 2)} \). Since \( 1/(\tau - 2) > 1 \) when \( \tau \in (2, 3) \), we obtain that vertices with large weights \( w \) are whp connected to vertices with weight at least \( w^{1/(\tau - 2)} \).

The proof of Theorems 6.2–6.3 are organized as follows. In Section 6.2, we prove the lower bounds on the typical distance in \( NR_n(w) \), both when \( \tau > 3 \) and when \( \tau \in (2, 3) \). In Section 6.3, we prove the log log \( n \) upper bound for \( \tau \in (2, 3) \). In Section 6.4, we investigate the number of paths between sets of vertices in \( NR_n(w) \), and use this to prove the log \( n \) upper bound when \( \tau > 3 \). In each of our proofs, we formulate the precise results as separate theorems, and prove them under conditions that are slightly weaker than those in Theorems 6.2–6.3. Further, in Section 6.2 we also give the proof of Theorem 6.1(i), and in Section 6.4 we also give the proof of Theorem 6.1(ii).

6.2 Lower bounds on typical distances in IRGs

In this section, we prove lower bounds on typical graph distances. In Section 6.2.1, we prove the lower bound in Theorem 6.1(i) in the setting of Theorem 6.2.

6.2.1 Logarithmic lower bound distances for finite-variance degrees

In this section, we prove a logarithmic lower bound on the graph distance in \( NR_n(w) \). The main result is as follows:

**Theorem 6.4** (Logarithmic lower bound graph distances \( NR_n(w) \)) Assume that

\[
\limsup_{n \to \infty} \nu_n = \nu,
\]  

(6.2.1)

where \( \nu \in (1, \infty) \) and

\[
\nu_n = \frac{E[W_n^2]}{E[W_n]} = \sum_{i \in [n]} w_i^2 / \sum_{i \in [n]} w_i.
\]  

(6.2.2)
Small-world phenomena in inhomogeneous random graphs

Then, for any $\varepsilon > 0$,

$$\mathbb{P}(\text{dist}_{\text{NR}_n(w)}(o_1, o_2) \leq (1 - \varepsilon) \log_{\nu} n) = o(1). \quad (6.2.3)$$

The same results hold for $\text{CL}_n(w)$ and $\text{GRG}_n(w)$ under the same conditions.

Proof The idea behind the proof of Theorem 6.4 is that it is quite unlikely that a path exists that is much shorter than $\log_{\nu} n$ edges. In order to show this, we use a first moment bound and show that the expected number of occupied paths connecting the two vertices chosen uniformly at random from $[n]$ having length at most $k$ is $o(1)$. We now fill in the details.

We abbreviate $k_n = \lceil (1 - \varepsilon) \log_{\nu} n \rceil$. Then, conditioning on the uniform vertices chosen gives

$$\mathbb{P}(\text{dist}_{\text{NR}_n(w)}(i, j) \leq k_n) = \frac{1}{n^2} \sum_{i,j \in [n]} \mathbb{P}(\text{dist}_{\text{NR}_n(w)}(i, j) \leq k_n)$$

$$= \frac{1}{n^2} \sum_{i,j \in [n]} \sum_{k=0}^{k_n} \mathbb{P}(\text{dist}_{\text{NR}_n(w)}(i, j) = k). \quad (6.2.4)$$

In this section and in Section 6.4, we make use of path-counting techniques (see in particular Section 6.4.1). Here, we show that short paths are unlikely by giving upper bounds on the expected number of paths of various types. In Section 6.4.1, we give bounds on the variance of the number of paths of various types, so as to show that long paths are quite likely to exist. Such bounds on the variance of the number of paths are quite challenging, and here we give some basics to highlight the main ideas in a much simpler setting.

A path $\pi = (\pi_0, \ldots, \pi_k)$ of length $k$ between vertices $i$ and $j$ is a sequence of vertices connecting $\pi_0 = i$ to $\pi_k = j$. We call a path $\pi$ self-avoiding when it visits every vertex at most once, i.e., $\pi_i \neq \pi_j$ for every $i \neq j$. Let $\mathcal{P}_k(i, j)$ denote the set of $k$-step self-avoiding paths between vertices $i$ and $j$. See Figure 6.1 for an example of a 12-step self-avoiding path between $i$ and $j$.

When $\text{dist}_{\text{NR}_n(w)}(i, j) = k$, there must be path of length $k$ such that all edges $(\pi_l, \pi_{l+1})$ are occupied in $\text{NR}_n(w)$, for $l = 0, \ldots, k-1$. The probability in $\text{NR}_n(w)$ that the edge $(\pi_l, \pi_{l+1})$ is occupied is equal to

$$1 - e^{-w_{\pi_l, \pi_{l+1}}/\ell_n} \leq w_{\pi_l, \pi_{l+1}}/\ell_n. \quad (6.2.5)$$

Figure 6.1 A 12-step self-avoiding path connecting vertices $i$ and $j$. 
6.2 Lower bounds on typical distances in IRGs

For CL\(_n(w)\) and GRG\(_n(w)\), an identical upper bound holds, which explains why the proof of Theorem 6.4 for NR\(_n(w)\) applies verbatim to those models.

We say that \(\pi\) is occupied when all edges in \(\pi\) are occupied in NR\(_n(w)\). Then, by the union bound or Boole’s inequality,

\[
P(\text{dist}_{NR_n(w)}(i, j) = k) \leq P(\exists \pi \in \mathcal{P}_k(i, j) : \pi \text{ occupied}) 
\leq \sum_{\pi \in \mathcal{P}_k(i, j)} P(\pi \text{ occupied}).
\]

(6.2.6)

For any path \(\pi \in \mathcal{P}_k(i, j)\),

\[
P(\pi \text{ occupied}) = \frac{w_i w_j}{\ell_n} \prod_{l=1}^{k} \frac{w_{\pi_l}^2}{\ell_n} = \frac{w_i w_j}{\ell_n} \prod_{l=1}^{k} \frac{w_{\pi_l}^2}{\ell_n}.
\]

(6.2.7)

Therefore,

\[
P(\text{dist}_{NR_n(w)}(i, j) = k) \leq \frac{w_i w_j}{\ell_n} \frac{\ell_n}{n^2} \sum_{\pi \in \mathcal{P}_k(i, j)} \prod_{l=1}^{k} \frac{w_{\pi_l}^2}{\ell_n} = \frac{w_i w_j}{\ell_n} \prod_{l=1}^{k} \frac{w_{\pi_l}^2}{\ell_n} \nu_n^k,
\]

where \(\nu_n\) is defined in (6.2.2). We conclude that

\[
P(\text{dist}_{NR_n(w)}(i, j) \leq k_n) \leq \frac{1}{n^2} \sum_{i,j \in [n]} \sum_{k=0}^{k_n} \frac{w_i w_j}{\ell_n} \nu_n^k = \frac{\ell_n}{n^2} \sum_{k=0}^{k_n} \nu_n^k
\]

(6.2.9)

By (6.2.1), \(\limsup_{n \to \infty} \nu_n = \nu \in (1, \infty)\), so that, for \(n\) large enough, \(\nu_n \geq (\nu - \delta) > 1\), while \(\ell_n/n = E[W_n] \to E[W] < \infty\). Thus, since \(\nu \mapsto (\nu^{k+1} - 1)/(\nu - 1)\) is increasing for every integer \(k \geq 0\),

\[
P(\text{dist}_{NR_n(w)}(o_1, o_2) \leq k_n) \leq O((\nu - \delta)^{k_n}/n) = o(1),
\]

(6.2.10)

when \(\delta = \delta(\varepsilon) > 0\) is chosen such that \((1 - \varepsilon)/\log(\nu - \delta) < 1\), and since \(k_n = [(1 - \varepsilon) \log_n \nu_n]\). This completes the proof of Theorem 6.4.

The condition (6.2.1) is slightly weaker than Condition 1.1(c), which is assumed in Theorem 6.2, as shown in Exercise 6.6. Exercise 6.7 extends the proof of Theorem 6.4 to show that \((\text{dist}_{NR_n(w)}(o_1, o_2) \leq k_n) \leq \log_n \nu_n\) is tight.

We close this section by extending the above result to settings where \(\nu_n\) is not necessarily bounded, the most interesting case being \(\tau = 3\).
Corollary 6.5 (Lower bound graph distances $NR_n(w)$ for $\tau = 3$) Let $\nu_n$ be given in (6.2.2). Then, for any $\varepsilon > 0$,

$$P\left(\text{dist}_{NR_n(w)}(o_1, o_2) \leq (1 - \varepsilon) \log \nu_n n\right) = o(1). \quad (6.2.11)$$

The same results hold for $CL_n(w)$ and $GRG_n(w)$ under the same conditions.

The proof of Corollary 6.5 is left as Exercise 6.8. In the case where $\tau = 3$ and $[1 - F_n(x)]$ is, for a large range of $x$ values, of the order $x^{-2}$ (which is stronger than $\tau = 3$), it can be expected that $\nu_n = \Theta(\log n)$, so that in that case

$$P\left(\text{dist}_{NR_n(w)}(o_1, o_2) \leq (1 - \varepsilon) \frac{\log n}{\log \log n}\right) = o(1). \quad (6.2.12)$$

Exercise 6.9 investigates the situation where $\tau = 3$. Exercise 6.10 the case where $\tau \in (2, 3)$, where Corollary 6.5 unfortunately does not give highly interesting results.

Proof of Theorem 6.1(i)

The proof of the upper bound in Theorem 6.1(i) is closely related to that in Theorem 6.4. Note that

$$P(\text{dist}_{IRG_n(\kappa_n)}(i, j) = k) \leq \sum_{i_0, \ldots, i_k \in [n]} \prod_{l=0}^{k-1} \frac{\kappa_n(x_{i_l}, x_{i_{l+1}})}{n}, \quad (6.2.13)$$

where $i_0 = i, i_k = j$ and we can restrict the vertices to be distinct, so that

$$P(\text{dist}_{IRG_n(\kappa_n)}(i, j) = k) \leq \frac{1}{n^k} \sum_{i_0, i_1, \ldots, i_k \in [n]} \prod_{l=0}^{k-1} \kappa_n(x_{i_l}, x_{i_{l+1}}). \quad (6.2.14)$$

If the above $(k + 1)$-dimensional discrete integrals could be replaced by the continuous integral, then we would arrive at

$$\frac{1}{n} \int_S \cdots \int_S \prod_{l=0}^{k} \kappa(x_l, x_{l+1}) \prod_{l=0}^{k} \mu(dx_l) = \frac{1}{n} \|T_k^{k+1}1\|_1, \quad (6.2.15)$$

which is bounded from above by $\frac{1}{n} \|T_k\|^{k+1}$. Repeating the bound in (6.2.10) would then prove that, when $\nu = \|T_k\| > 1$,

$$P(\text{dist}_{IRG_n(\kappa_n)}(i, j) \leq (1 - \varepsilon) \log \nu n) = o(1). \quad (6.2.16)$$

However, in the general case, it is not so easy to replace the $(k + 1)$-fold discrete sum in (6.2.14) by a $(k + 1)$-fold integral. We next explain how this can be done, starting with the finite-types case.
In the finite-type case, (6.2.13) turns into

\[ P(\text{dist}_{\text{IRG}_n}(i, j) = k) \leq \frac{1}{n^2} \sum_{\nu_1, \ldots, \nu_k \in [\kappa]} \prod_{l=0}^{k-1} \frac{\kappa_n(x_{\nu_l}, x_{\nu_{l+1}})}{n} \]  

(6.2.17)

\[ \leq \frac{1}{n^2} \sum_{i_1, \ldots, i_{k-1} \in [\kappa]} \prod_{l=0}^{k-1} \kappa_n(i_{l}, i_{l+1}) \frac{n_{i_{l+1}}}{n}, \]

where the number of vertices of type \( i \) is denoted by \( n_i \), and where the probability that there exists an edge between vertices of type \( i \) and \( j \) is equal to \( \kappa_n(i, j)/n \).

Under the conditions in Theorem 6.1(i), we have that \( n_i/n \to p_i \) and that \( \kappa_n(i, j) \to \kappa_{ij} \). This also implies that \( \|T_{\kappa_n}\| \to \nu \), where \( \nu \) is largest eigenvalue of the matrix \( (m_{ij})_{i,j \in [\kappa]} \) with \( m_{ij} = \kappa_{ij}p_j \). Denoting \( m_i^{(n)} = \kappa_n(i, j)n_j/n \to m^{(n)}_{ij} \), we obtain

\[ P(\text{dist}_{\text{IRG}_n}(i, j) = k) \leq \frac{1}{n} \langle \mu^T, [M^{(n)}]^k \rangle, \]

(6.2.18)

where \( 1 \) is the all-one vector, \( \mu_i = n_i/n \to p_i \), and \( M_{ij}^{(n)} = m^{(n)}_{ij} \). Obviously, where there are \( N < \infty \) types,

\[ \langle \mu^T, [M^{(n)}]^k \rangle \leq \|M^{(n)}\|^k \|\mu\| \|1\| \leq \|M^{(n)}\|^k \sqrt{N}. \]

(6.2.19)

Thus,

\[ P(\text{dist}_{\text{IRG}_n}(i, j) = k) \leq \frac{\sqrt{N}}{n} \|M^{(n)}\|^k. \]

(6.2.20)

We conclude that

\[ P(d_{\text{IRG}_n}(\kappa_n)(o_1, o_2) \leq (1 - \varepsilon) \log_{\nu_n} n) = o(1), \]

(6.2.21)

where \( \nu_n = \|M^{(n)}\| \to \nu \). This proves the claim of Theorem 6.1(i) in the finite-type setting.

We next extend the proof of Theorem 6.1(i) to the infinite-type setting. Assume that the conditions in Theorem 6.1(i) hold. Recall the bound in (3.3.18), which bounds \( \kappa_n \) from above by \( \kappa_n^+ \), which is of finite-type. Then, use the fact that \( \|T_{\kappa_n}\| \to \nu > 1 \) to conclude that \( P(d_{\text{IRG}_n}(\kappa_n)(o_1, o_2) \leq (1 - \varepsilon) \log_{\nu_n} n) = o(1) \) holds under the conditions of Theorem 6.1(i). This completes the proof of Theorem 6.1(i).

Theorem 6.1 leaves open the case when \( \|T_{\kappa}\| = \infty \), which, for example for \( \text{CIR}_n(w) \), is the case when \( F \) has infinite second moment. (Bollobás et al., 2007, Theorem 3.14(iv)) states that when \( \|T_{\kappa}\| = \infty \), the typical graph distance is smaller than \( \log n \). More precisely, (Bollobás et al., 2007, Theorem 3.14(iv)) states
that if $\kappa$ is irreducible and $\|T_\kappa\| = \infty$, then there is a function $f(n) = o(\log n)$ such that

$$\mathbb{P}(\text{dist}_{\text{IRG}_n}(\kappa)(i,j) \leq f(n)) = \zeta_2^\kappa + o(1).$$

(6.2.22)

Exercise 6.12 shows that one can take $f(n) = o(\log n)$.

### 6.2.2 Log log lower bound on distances for infinite-variance degrees

In this section, we prove a log log-lower bound on the typical distances of $\text{NR}_n(w)$ for $\tau \in (2,3)$. The main result we prove is the following theorem:

**Theorem 6.6 (Loglog lower bound on typical distances in $\text{NR}_n(w)$)** Suppose that the weights $w = (w_i)_{i \in [n]}$ satisfy Condition 1.1(a) and that there exists a $\tau \in (2,3)$ and $c_2$ such that, for all $x \geq 1$,

$$[1 - F_n](x) \leq c_2 x^{-(\tau - 1)},$$

(6.2.23)

Then, for every $\varepsilon > 0$,

$$\mathbb{P}(\text{dist}_{\text{NR}_n}(w)(i,j) \leq (1 - \varepsilon) \frac{2 \log \log n}{|\log (\tau - 2)|}) = o(1).$$

(6.2.24)

The same results hold for $\text{CL}_n(w)$ and $\text{GRG}_n(w)$ under the same conditions.

We follow the proof of Theorem 6.4 as closely as possible. The problem with that proof is that, under the condition in (6.2.23), $\nu_n$ is too large. Indeed, Exercise 6.10 shows that the lower bound obtained in Corollary 6.5 is a constant, which is not very useful. What goes wrong in that argument is that there are too many vertices with too high weight, and they provide the main contribution to $\nu_n$ and hence to the upper bound as in (6.2.9). However, this argument completely ignores the fact that it is quite unlikely that a vertex with a high weight is chosen.

Indeed, as argued in (6.1.6), when starting from a vertex with weight $w$, say, the probability that it is directly connected to a vertex having weight $a_n$ is at most

$$\sum_{j : w_j \geq y} \frac{w w_j}{\ell_n} = w[1 - F_n^\ast](y),$$

(6.2.25)

which is small when $y$ is too large. On the other hand, the main contribution to $\nu_n$ comes from vertices having maximal weight of the order $n^{1/(\tau - 1)}$. This problem is resolved by a suitable truncation argument on the weights of the vertices in the occupied paths, which effectively removes these high-weight vertices. Therefore, instead of obtaining $\nu_n = \sum_{x \in [n]} w_x^2 / \ell_n$, we obtain a partial sum of this restricted to vertices having a relatively small weight. Effectively, this means that we split the space of all paths into good paths, i.e., paths that avoid vertices with too large weight, and bad paths, which are paths that jump to vertices with too high weight.
6.2 Lower bounds on typical distances in IRGs

\[ \pi_0 = i: w_{\pi_0} \leq b_0 \]

\[ \pi_1: w_{\pi_1} \leq b_1 \]

\[ \pi_2: w_{\pi_2} \leq b_2 \]

\[ \pi_3: w_{\pi_3} \leq b_3 \]

\[ \pi_4: w_{\pi_4} \leq b_4 \]

\[ \pi_5: w_{\pi_5} \leq b_5 \]

\[ \pi_6: w_{\pi_6} \leq b_6 \]

\[ \pi_7: w_{\pi_7} \leq b_7 \]

\[ \pi_8: w_{\pi_8} \leq b_8 \]

\[ \pi_9: w_{\pi_9} \leq b_9 \]

\[ \pi_{10} = j: w_{\pi_{10}} \leq b_0 \]

Figure 6.2 A 10-step good path connecting \( i \) and \( j \) and the upper bounds on the weight of its vertices. The height of a vertex is high for vertices with large weights.

We now present the details for this argument. We again start from

\[ \mathbb{P}(\text{dist}_{\text{NR}_n}(i,j) \leq k_n) = \frac{1}{n^2} \sum_{i,j \in [n]} \mathbb{P}(\text{dist}_{\text{NR}_n}(i,j) \leq k_n). \] (6.2.26)

When \( \text{dist}_{\text{NR}_n}(i,j) \leq k_n \), there exists an occupied path \( \pi \in \mathcal{P}_k(i,j) \) for some \( k \leq k_n \).

We fix an increasing sequence of numbers \((b_l)_{l=0}^\infty\) that serve as truncation values for the weights of vertices along our occupied path. We determine the precise values of \((b_l)_{l=0}^\infty\), which is quite delicate, below. We say that a path \( \pi \in \mathcal{P}_k(i,j) \) is good when \( w_{\pi_l} \leq b_l \wedge b_{k-l} \) for every \( l = 0, \ldots, k \), and bad otherwise. The condition \( w_{\pi_l} \leq b_l \wedge b_{k-l} \) for every \( l = 0, \ldots, k \) is equivalent to the statement that \( w_{\pi_l} \leq b_l \) for \( l \leq \lceil k/2 \rceil \), while \( w_{\pi_l} \leq b_{k-l} \) for \( \lceil k/2 \rceil < l \leq k \). Thus, \( b_l \) provides an upper bound on the weight of the \( l \)th vertex and the \((k-l)\)th vertex of the occupied path, ensuring that the weights occurring in the occupied path can not be too large. See Figure 6.2 for a description of a good path and the bounds on the weight of its vertices.

Let \( \mathcal{GP}_k(i,j) \) be the set of good paths in \( \mathcal{P}_k(i,j) \). Let

\[ \mathcal{E}_k(i,j) = \{ \exists \pi \in \mathcal{GP}_k(i,j): \pi \text{ occupied} \} \] (6.2.27)

denote the event that there exists a good path of length \( k \).

When \( \text{dist}_{\text{NR}_n}(i,j) \leq k_n \), but there does not exist a \( k \leq k_n \) and a good occupied path \( \pi \in \mathcal{GP}_k(i,j) \), then either there exists an \( l \leq \lceil k/2 \rceil \) such that \( w_{\pi_s} \leq b_s \) for every \( s < l \), while \( w_{\pi_l} > b_l \), or there exists an \( l \leq \lceil k/2 \rceil \) such that \( w_{\pi_{k-s}} \leq b_{k-s} \) for every \( s < l \), while \( w_{\pi_{k-l}} > b_{k-l} \). Let \( \mathcal{P}_k(i) = \cup_{l \in [n]} \mathcal{P}_k(i,l) \) denote
the set of all paths of length $k$ from $i$, and let
\[ \mathcal{B} \mathcal{P}_k(i) = \{ \pi \in \mathcal{P}_k(i) : w_{\pi_l} > b, w_{\pi_s} \leq b, \forall s < l \} \] (6.2.28)
denote the set of bad paths of length $k$, i.e., those $\pi \in \mathcal{P}_k(i)$ that are not in $\mathcal{G} \mathcal{P}_k(i, \pi_k)$. Let $\mathcal{F}_l(i)$ be the event that there exists a bad path of length $l$ starting from $i$, i.e.,
\[ \mathcal{F}_l(i) = \{ \exists \pi \in \mathcal{B} \mathcal{P}_l(i) : \pi \text{ occupied} \}. \] (6.2.29)

Then, since $\text{dist}_{NR_n}(\omega)(i,j) \leq k_n$ implies that there either is a good path or a bad path,
\[ \{ \text{dist}_{NR_n}(\omega)(i,j) \leq k_n \} \subseteq \bigcup_{k \leq k_n} (\mathcal{F}_k(i) \cup \mathcal{F}_k(j) \cup \mathcal{E}_k(i,j)), \] (6.2.30)
so that, by Boole’s inequality,
\[ \mathbb{P}(\text{dist}_{NR_n}(\omega)(i,j) \leq k_n) \leq k_n \sum_{k=0}^{k_n} \left[ \mathbb{P}(\mathcal{F}_k(i)) + \mathbb{P}(\mathcal{F}_k(j)) + \mathbb{P}(\mathcal{E}_k(i,j)) \right]. \] (6.2.31)

In order to estimate the probabilities $\mathbb{P}(\mathcal{F}_k(i))$ and $\mathbb{P}(\mathcal{E}_k(i,j))$, we introduce some notation. For $b \geq 0$, let
\[ \nu_n(b) = \frac{1}{\ell_n} \sum_{i \in [n]} w_i^2 \mathbb{1}_{\{w_i \leq b\}}, \] (6.2.32)
be the restriction of $\nu_n$ to vertices with weights at most $b$, and let
\[ F^*_n(x) = \frac{1}{\ell_n} \sum_{i \in [n]} w_i \mathbb{1}_{\{w_i \leq x\}} \] (6.2.33)
be the distribution function of $W^*_n$, the size-biased version of $W_n$. The following lemma gives bounds on $\mathbb{P}(\mathcal{F}_k(i))$ and $\mathbb{P}(\mathcal{E}_k(i,j))$ in terms of the tail distribution function $1 - F^*_n$ and the truncated second moment $\nu_n(b)$, which we will bound using Lemma:

**Lemma 6.7** (Truncated path probabilities) For every $k \geq 1$, $(b_l)_{l \geq 0}$ with $b_l \geq 0$ and $l \mapsto b_l$ non-decreasing,
\[ \mathbb{P}(\mathcal{F}_k(i)) \leq w_i [1 - F^*_n(b_k)] \prod_{l=1}^{k-1} \nu_n(b_l), \] (6.2.34)
and
\[ \mathbb{P}(\mathcal{E}_k(i,j)) \leq \frac{w_i w_j}{\ell_n} \prod_{l=1}^{k-1} \nu_n(b_l \wedge b_{k-l}), \] (6.2.35)
where
\[ \nu_n(b) \leq c_p b^{3-\tau}. \] (6.2.36)

When $b_l = \infty$ for each $l$, the bound in (6.2.35) equals that obtained in (6.2.8).
6.2 Lower bounds on typical distances in IRGs

Proof We start by proving (6.2.34). By Boole’s inequality,

\[ \Pr(F_k(i)) = \Pr(\exists \pi \in BP_l(i): \pi \text{ occupied}) \leq \sum_{\pi \in BP_l(i)} \Pr(\pi \text{ occupied}). \]  

(6.2.37)

By (6.2.7), (6.2.32) and (6.2.33),

\[ \Pr(F_k(i)) \leq \sum_{\pi \in BP_l(i)} w_i \prod_{l=1}^{k-1} \frac{w_{\pi_l}}{\ell_n} \leq w_i \sum_{\pi_\tau: \pi_\tau \geq b_k} \prod_{l=1}^{k-1} \sum_{\pi_l \leq b_{l-1}} w_{\pi_l}/\ell_n \]

(6.2.38)

\[ = w_i [1 - F^*_n](b_k) \prod_{l=1}^{k} \nu_n(b_l). \]

The same bound applies to CL \(_n(w)\) and GRG \(_n(w)\).

The proof of (6.2.35) is similar. Indeed, by (6.2.7),

\[ \Pr(E_k(i,j)) \leq \sum_{\pi \in GP_k(i,j)} w_i w_j \prod_{l=1}^{k-1} \frac{w_{\pi_l}}{\ell_n} \leq w_i w_j \prod_{l=1}^{k-1} \nu(b_l) \]

(6.2.39)

Now follow the steps in the proof of (6.2.34). Finally, (6.2.42) follows from (1.4.3) in Lemma 1.13, combined with \(\ell_n = \Theta(n)\) by Conditions 1.1(a)–(b). See also Exercise 6.5 below. Again the same bound applies to CL \(_n(w)\) and GRG \(_n(w)\).

In order to effectively apply Lemma 6.7, we use Lemmas 1.13 and 1.14 to derive bounds on \([1 - F^*_n](x)\) and \(\nu_n(b)\):

Lemma 6.8 (Bounds on sums) Suppose that the weights \(w = (w_i)_{i \in [n]}\) satisfy Conditions 1.1(a)–(b) and that there exist \(\tau \in (2, 3)\) and \(c_2\) such that, for all \(x \geq 1\),

\[ [1 - F_n](x) \leq c_2 x^{-(\tau - 1)}. \]  

(6.2.40)

Then, there exists a constant \(c^*_\nu > 0\) such that, for all \(x \geq 1\),

\[ [1 - F^*_n](x) \leq c^*_\nu x^{-(\tau - 2)}, \]  

(6.2.41)

and there exists a \(c_\nu > 0\) such that for all \(b \geq 1\),

\[ \nu_n(b) \leq c_\nu b^{3-\tau}. \]  

(6.2.42)

Proof The bound in (6.2.41) follows from Lemma 1.14, the bound in (6.2.42) from (1.4.3) in Lemma 1.13 with \(a = 2 > \tau - 1\) when \(\tau \in (2, 3)\). For both lemmas, the assumptions follow from (6.2.40) (which, in turn, equals (6.2.23)).

With Lemmas 6.7 and 6.8 in hand, we are ready to choose \((b_l)_{l \geq 0}\) and to complete the proof of Theorem 6.6:
Proof of Theorem 6.6. Take $k_n = 2(1 - \varepsilon) \log \log n / |\log (\tau - 2)|$. By (6.2.26) and (6.2.30),
\[
P(\text{dist}_{n_n(w)}(i, j) \leq k_n) \leq \frac{1}{n} + \sum_{k=1}^{k_n} \left[ \frac{2}{n} \sum_{x \in [n]} P(\mathcal{F}_k(i)) + \frac{1}{n^2} \sum_{i,j \in [n], i \neq j} P(\mathcal{E}_k(i, j)) \right],
\]
where the contribution $1/n$ is due to $i = j$ for which $\text{dist}_{n_n(w)}(i, i) = 0$. We use Lemmas 6.7 and 6.8 to provide bounds on $P(\mathcal{F}_k(i))$ and $P(\mathcal{E}_k(i, j))$. These bounds are quite similar.

We first describe how we choose the truncation values $(b_l)_{l=0}^\infty$ so that $[1 - F^*_n(b_k)]$ is so small that $P(\mathcal{F}_k(i))$ is small, and, for this choice of $(b_l)_{l=0}^\infty$, we show that $P(\mathcal{E}_k(i, j))$ is small. Intuitively, this means that it is quite unlikely that $i$ or $j$ is connected to a vertex at distance $k$ with too high weight, i.e., having weight at least $b_k$. At the same time, it is also unlikely that there is a path $\pi \in \mathcal{P}_k(i, j)$ whose weights are all small, i.e., for which $w_{\pi_k} \leq b_k$ for every $k \leq k_n$, because $b_k$ is too large.

By Lemma 6.7, we wish to choose $b_k$ so that $P(\mathcal{F}_k(i)) \leq [1 - F^*_n(b_k)] \prod_{l=0}^{k-1} \nu_n(b_l)$ is small. Below (6.1.6), it is argued that $b_k \approx b_k^1/(\tau - 2)$. In order to make this probability small, we will take $b_k$ somewhat larger. We now make this argument precise.

We take $\delta \in (0, \tau - 2)$ sufficiently small and let
\[
a = 1/(\tau - 2 - \delta) > 1.
\]
Take $b_l = e^A$ for some constant $A \geq 0$ sufficiently large and define $(b_l)_{l \geq 0}$ recursively by
\[
b_l = b_0^l, \quad \text{so that} \quad b_l = b_0^l = e^{A(\tau - 2 - \delta)^{-l}}.
\]
We start from (6.2.31). By Lemma 6.7, we obtain an upper bound on $P(\mathcal{F}_k(i))$ in terms of factors $\nu_n(b_l)$ and $[1 - F^*_n(b_k)]$, which are bounded in Lemma 6.8. We start by applying the bound on $\nu_n(b_l)$ to obtain
\[
\prod_{l=1}^{k-1} \nu_n(b_l) \leq \prod_{l=1}^{k-1} c_\nu b_l^{-3} = c_\nu^k e^{K(3-\tau) \sum_{l=1}^{k-1} a^l}
\]
\[
\leq c_\nu^{k-1} e^{K(3-\tau) a^k/(a-1) = c_\nu^* w^*_c c_k^{k-1} b_k^{-(3-\tau)/(a-1)}}.
\]
Combining (6.2.45) with the bound on $[1 - F^*_n(b_k)]$ in Lemma 6.8 yields
\[
P(\mathcal{F}_k(i)) \leq c^*_2 w^*_c c_k^{k-2} b_k^{-(2-\delta)/(3-\tau)/(a-1)}.
\]
Since $3 - \tau + \delta < 1$ when $\tau \in (2, 3)$ and $\delta \in (0, \tau - 2)$,
\[
(\tau - 2) - (3 - \tau)/(a - 1) = (\tau - 2) - (3 - \tau)(\tau - 2 - \delta)/(3 - \tau + \delta)
\]
\[
= \delta/(3 - \tau + \delta) > \delta,
\]
so that
\[
P(\mathcal{F}_k(i)) \leq c^*_2 w^*_c c_k^{k-\delta}.
\]
As a result, for each $\delta > 0$

$$
\frac{1}{n} \sum_{i \in [n]} \sum_{k=0}^{k_n} \mathbb{P}(F_k(i)) \leq \frac{c_2}{n} \sum_{i \in [n]} w_i \mathbb{1}_{\{w_i > k\}} + \frac{1}{n} \sum_{i \in [n]} c_2 w_i \sum_{k \geq 1} c_k^{\ell_n - \delta} \quad (6.2.49)
$$

\[= O(c_k^{b_k^{-\delta}}) \leq \varepsilon,\]

by (6.2.44) and when we take $A = A(\delta, \varepsilon)$ sufficiently large.

Similarly, since $b_i \geq 1$, by (6.2.45),

$$
\mathbb{P}(E_k(i, j)) \leq \frac{w_i w_j}{\ell_n} \prod_{l=1}^{k-1} n \leq \frac{w_i w_j}{\ell_n} c_\nu^{k-1} b_{[k/2]}^{2(3-\tau)/(\alpha-1)}, \quad (6.2.50)
$$

so that, using further that $l \mapsto b_l$ is increasing,

$$
\sum_{k=1}^{k_n} \frac{1}{n^2} \sum_{i, j \in [n]} \mathbb{P}(E_k(i, j)) \leq \frac{1}{n^2} \sum_{k=1}^{k_n} \sum_{i, j \in [n]} \frac{w_i w_j}{\ell_n} c_\nu^{k-1} b_{[k/2]}^{2(3-\tau)/(\alpha-1)} \quad (6.2.51)
$$

\[\leq \frac{\ell_n}{n^2} c_\nu c_{k_n} b_{[k_n/2]}^{2(3-\tau)/(\alpha-1)},\]

by (6.2.44) and the fact that $k \mapsto b_k$ is monotonically increasing. We complete the proof by analyzing this bound.

Recall that $k \leq k_n = 2(1 - \varepsilon) \log \log n / |\log (\tau - 2)|$. Take $\delta = \delta(\varepsilon) > 0$ so small that $(\tau - 2 - \delta)^{-1/2} \leq (\log n)^{1-\varepsilon/4}$. Then, by (6.2.44),

$$
b_{[k_n/2]} \leq e^{A(\tau - 2 - \delta)^{-1/2}} \leq e^{A(\log n)^{1-\varepsilon/4}}, \quad (6.2.52)
$$

and we conclude that

$$
\sum_{k=1}^{k_n} \frac{1}{n^2} \sum_{i, j \in [n]} \mathbb{P}(E_k(i, j)) \leq \frac{\ell_n}{n^2} c_\nu c_{k_n} \exp \left(2A(3 - \tau)(\log n)^{1-\varepsilon/4}\right) = o(1), \quad (6.2.53)
$$

since $k_n = O(\log \log n)$ and $\ell_n/n^2 = \Theta(1/n)$. This completes the proof of Theorem 6.6. \(\square\)

In Exercise 6.14, the above argument is extended to show that the sequence of random variables $\left(\text{dist}_{\text{NR}_n(w)}(o_1, o_2) \leq 2 \log \log n / \log (\tau^{2})\right)$ is tight.

### 6.3 The log log upper bound

In this section, we prove the log log upper bound on typical graph distances in the case where the asymptotic weight distribution has infinite variance. Recall the comparison neighborhoods of vertices in $\text{NR}_n(w)$ to branching processes discussed in Section 3.5.3. Throughout this section, we assume that there exist $\tau \in (2, 3)$, $\alpha > 1/2$ and $c_1$ such that, uniformly in $n$ and $x \leq n^\alpha$,

$$
[1 - F_n](x) \geq c_1 x^{-(\tau - 1)}. \quad (6.3.1)
$$
The bound in (6.3.1) corresponds to the lower bound in (6.1.4). The main result in this section is the following theorem:

**Theorem 6.9** (A log log upper bound on typical distance for $\tau \in (2, 3)$) Suppose that empirical distribution function $F_n$ of the weights $w = (w_i)_{i \in [n]}$ satisfies Conditions 1.1(a)-(b) and (6.3.1). Then, for every $\varepsilon > 0$, as $n \to \infty$,

$$P\left( \text{dist}_{\text{NR}}(w)(o_1, o_2) \leq 2 \left(1 + \varepsilon\right) \frac{\log \log n}{\log (\tau - 2)} \mid \text{dist}_{\text{NR}}(w)(o_1, o_2) < \infty \right) \to 1. \quad (6.3.2)$$

The same results hold for $\text{CL}_n(w)$ and $\text{GRG}_n(w)$ under the same conditions.

The proof Theorem 6.9 is organized as follows. We start by showing that the giant-weight vertices, i.e., the vertices with extremely high weight larger than $n^\alpha$, are all connected to one another. Thus, the giant-weight vertices form a complete graph. This is often referred to as a *clique* in the random graph community. In the second step, we show that connections from a vertex to the set of giant weight vertices occur at distance at most $\left(1 + \varepsilon\right) \frac{\log \log n}{\log (\tau - 2)}$. The latter is only true when the vertex is in the giant connected component, a fact we need to carefully into account. In the final step, we complete the proof of Theorem 6.9. We now start by defining the set of giant-weight vertices.

**The giant-weight vertices form a clique**

Recall the definition of $\alpha > 1/2$ in (6.3.1). Let

$$\text{Giant}_n = \{i : w_i \geq n^\alpha\} \quad (6.3.3)$$

denote the set of vertices with giant weights. Let $A \subseteq [n]$. We say that $A$ forms a *clique* when the edges $a_1a_2$ are occupied for all $a_1, a_2 \in A$. We continue by proving that, whp, $\text{Giant}_n$ forms a clique:

**Lemma 6.10** (High-weight vertices form clique) Under the conditions of Theorem 6.9,

$$P(\text{Giant}_n \text{ does not form clique}) \leq n^2 e^{-n^{2\alpha}/\ell_n} \quad (6.3.4)$$

The same results hold for $\text{CL}_n(w)$ under the same conditions, for $\text{GRG}_n(w)$ the diameter of $\text{Giant}_n$ is at most 2.

**Proof** Let $a_1, a_2 \in \text{Giant}_n$, so that $w_{a_1}, w_{a_2} \geq n^\alpha$. There are at most $|\text{Giant}_n|^2 \leq n^2$ pairs of vertices in $\text{Giant}_n$, so that

$$P(\text{Giant}_n \text{ does not form clique}) \leq n^2 \max_{a_1, a_2 \in \text{Giant}_n} P(a_1a_2 \text{ vacant}). \quad (6.3.5)$$

The edge $a_1a_2$ is vacant with probability

$$P(a_1a_2 \text{ vacant}) = e^{-w_{a_1}w_{a_2}/\ell_n} \leq e^{-n^{2\alpha}/\ell_n}, \quad (6.3.6)$$

since $w_a \geq n^\alpha$ for every $a \in \text{Giant}_n$. Multiplying out gives the result. For $\text{CL}_n(w)$, $P(a_1a_2 \text{ vacant}) = 0$, so the same proof applies.
For $\text{GRG}_n(w)$, we need to strengthen this analysis slightly. Indeed, for $\text{GRG}_n(w)$, for all $a_1, a_2 \in \text{Giant}_n$,
\[
\mathbb{P}(a_1a_2 \text{ present}) \geq \frac{n^{2a}}{\ell_n + n^{2a}} = 1 - \Theta(n^{1-2a}) \geq \frac{1}{2}.
\] (6.3.7)
Thus, the diameter of $\text{Giant}_n$ is bounded by the diameter of $\text{ER}_n(p)$ with $p = \frac{1}{2}$. Thus, it suffices to prove that the diameter of $\text{ER}_n(\frac{1}{2})$ is whp bounded by 2. For this, we note that
\[
\mathbb{P}(\text{diam}(\text{ER}_n(\frac{1}{2})) > 2) \leq n^2\mathbb{P}(\text{dist}_G(\frac{1}{2}, 1, 2) > 2).
\] (6.3.8)
The event $\{\text{dist}_{\text{ER}_n(\frac{1}{2})}(1, 2) > 2\}$ implies that all two-hop paths between 1 and 2 are not occupied, so that, by independence
\[
\mathbb{P}(\text{dist}_{\text{ER}_n(\frac{1}{2})}(1, 2) > 2) = (1 - \frac{1}{4})^{n-2},
\] (6.3.9)
which, combined with (6.3.8), completes the proof.

**Connections to Giant$_n$ occur at log log $n$ distances**

We next show that vertices that survive up to distance $m$ have a high probability of connecting to $\text{Giant}_n$ using a path of at most $(1 + \varepsilon)\frac{\log \log n}{\log (\tau - 2)}$ edges:

**Proposition 6.11** (Connecting to Giant$_n$) Let $i \in [n]$ be such that $w_i > 1$. Under the conditions of Theorem 6.9, there exist $c, c'_1 > 0$ and $\eta > 0$ such that
\[
\mathbb{P}\left(\text{dist}_{N_{\text{GRG}}(w)}(i, \text{Giant}_n) \geq (1 + \varepsilon)\frac{\log \log n}{\log (\tau - 2)} \right) \leq ce^{-c'_1w_i^\eta}.
\] (6.3.10)
Consequently, with $W_m(i) = \sum_{k \in \partial B_m(i)} w_k$ denoting the weight of vertices at graph distance $m$ from $i$,
\[
\mathbb{P}\left(\text{dist}_{N_{\text{GRG}}(w)}(N_m(i), \text{Giant}_n) \geq (1 + \varepsilon)\frac{\log \log n}{\log (\tau - 2)} \mid B_m(i) \right) \leq ce^{-c'_1W_m(i)^\eta}.
\] (6.3.11)

**Proof** We start by proving (6.3.10). The bound in (6.3.10) is trivial unless $w_i$ is large. We let $x_0 = i$, and define, recursively,
\[
x_\ell = \max\{j \in [n] : x_{\ell-1}j \text{ occupied}\}.
\] (6.3.12)
Thus, $x_\ell$ is the maximal-weight neighbor of $x_{\ell-1}$. We stop the above recursion when $w_{x_\ell} \geq n^{\alpha}$, since then $x_\ell \in \text{Giant}_n$. Recall the heuristic below (6.1.6), which shows that a vertex with weight $w$ is whp connected to a vertex with weight $w^{1/(\tau - 2)}$. We now make this precise.

We take $a = 1/(\tau - 2 + \delta)$, where we choose $\delta > 0$ so small that $a > 1$. By (6.2.33),
\[
\mathbb{P}(w_{x_{\ell+1}} < w_{x_\ell}^{a} \mid (x_s)_{s \leq \ell}) = e^{-w_{x_\ell} \sum_{i: w_{x_\ell} \geq w_{x_\ell}^{a} \mid (x_s)_{s \leq \ell}} w_i/\ell_n} = e^{-w_{x_\ell}^{1-a}F_a^*(w_{x_\ell}^{a})}.
\] (6.3.13)
We split the argument depending on whether \( w_{x_t}^a \leq n^\alpha \) or not. Firstly, when \( w_{x_t}^a \leq n^\alpha \), by (6.3.1) and uniformly for \( x \leq n^\alpha \),

\[
[1 - F_n^u](x) \geq \frac{xn}{\ell_n}[1 - F_n](x) \geq c_1^w x^{-(r-2)},
\]

(6.3.14)

where, for \( n \) large enough, we can take \( c_1^w = c_1/(2E[W]) \). Therefore,

\[
P(w_{x_{t+1}} < w_{x_t}^a \mid (x_s)_{s \leq t}) \leq e^{-c_1 w_{x_t}^{a-(r-2)a}} \leq e^{-c_1 w_{x_t}^a},
\]

(6.3.15)

since \( a = 1/(\tau - 2 + \delta) > 1 \) so that \( 1 - (\tau - 2)a = a\delta > \delta \).

Secondly, when \( w_{x_t}^a > n^\alpha \), but \( w_{x_t} < n^\alpha \), we can use (6.3.14) for \( x = n^\alpha \) to obtain

\[
P(w_{x_{t+1}} < n^\alpha \mid (x_s)_{s \leq t}) \leq e^{-c_1 w_{x_t}^{a-(r-2)a}} \leq e^{-c_1 w_{x_t}^{a}\alpha/a} \leq e^{-c_1 w_{x_t}^{\alpha/a}}.
\]

(6.3.16)

Therefore, in both cases, and with \( \eta = \alpha\delta/a \),

\[
P(w_{x_{t+1}} < (n^\alpha \wedge w_{x_t}^a) \mid (x_s)_{s \leq t}) \leq e^{-c_1 w_{x_t}^\eta}.
\]

(6.3.17)

As a result, when \( x_t \) is such that \( w_{x_t} \) is quite large, whp, \( w_{x_{t+1}} \geq w_{x_t} \). This produces, whp, a short path to Giant, \( \text{Giant}_n \). We now investigate the properties of this path.

Let the recursion stop at some integer time \( k \). The key observation is that when this occurs, we must have that \( w_{x_{t+1}} > w_{x_t}^a \) for each \( \ell \leq k - 1 \) where \( k \) is such that \( w_{x_{k-1}} \in [n^\alpha/a, n^\alpha] \), and at the same time \( w_{x_k} \geq n^\alpha \). Then, we conclude that the following facts are true:

1. \( w_{x_t} \geq w_{x_t}^a = w_{i}^a \) for every \( \ell \leq k - 1 \),
2. \( \text{dist}_{G_n(w)}(i, \text{Giant}_n) \leq k \).

By (1), \( w_{x_{k-1}} \geq w_{x_{k-1}}^a \), and \( w_{x_{k-1}} \in [n^\alpha/a, n^\alpha] \). Therefore, \( w_{i}^{a-1} \leq n^\alpha \), which, in turn, implies that

\[
a^{k-1} \leq \alpha \log n, \quad \text{or} \quad k - 1 \leq (\log \log n + \log \alpha)(\log a).
\]

(6.3.18)

Let \( k_n = (1 + \varepsilon)\log \log n / (\log (r-2)) \). By (1) and (2), when \( \text{dist}_{G_n(w)}(i, \text{Giant}_n) \geq k_n \) occurs, then there must exist an \( \ell \leq k_n \) such that \( w_{x_{t+1}} \leq n^\alpha \wedge w_{x_t}^a \). We conclude that

\[
P\left( \text{dist}_{G_n(w)}(i, \text{Giant}_n) \geq k_n \right) \leq \sum_{\ell=0}^{k_n} P(w_{x_{t+1}} \leq w_{x_t}^a)
\]

(6.3.19)

\[
\leq \sum_{\ell=0}^{k_n} EP\left( w_{x_{t+1}} \leq w_{x_t}^a \mid (x_s)_{s \leq t} \right)
\]

\[
\leq \sum_{\ell=0}^{k_n} E[e^{-c_1 w_{x_t}^a}] \leq \sum_{\ell=0}^{k_n} e^{-c_1 w_{x_t}^a} \leq ce^{-c_1 w_{x_t}^a}.
\]

This completes the proof of (6.3.10).

The proof of (6.3.11) is similar, by conditioning on \( B_m(i) \) and by noting that we can interpret \( B_m(i) \) as a single vertex having weight \( W_m(i) = \sum_{k \in B_m(i)} w_k \).
Completion of the proof of Theorem 6.9

To prove the upper bound in Theorem 6.9, for \( \varepsilon \in (0, 1) \), we take

\[
k_n = (1 + \varepsilon) \frac{\log \log n}{|\log (\tau - 2)|}.
\] (6.3.20)

so that it suffices to show, for every \( \varepsilon > 0 \),

\[
\lim_{n \to \infty} \mathbb{P}(\text{dist}_{NR_n(w)}(o_1, o_2) \leq 2k_n | \text{dist}_{NR_n(w)}(o_1, o_2) < \infty) = 1.
\] (6.3.21)

Since

\[
\frac{\mathbb{P}(\text{dist}_{NR_n(w)}(o_1, o_2) \leq 2k_n | \text{dist}_{NR_n(w)}(o_1, o_2) < \infty)}{\mathbb{P}(\text{dist}_{NR_n(w)}(o_1, o_2) < \infty)} = \frac{\mathbb{P}(\text{dist}_{NR_n(w)}(o_1, o_2) \leq 2k_n)}{\mathbb{P}(\text{dist}_{NR_n(w)}(o_1, o_2) < \infty)},
\]

this follows from the two bounds

\[
\liminf_{n \to \infty} \mathbb{P}(\text{dist}_{NR_n(w)}(o_1, o_2) < \infty) \leq \zeta^2,
\] (6.3.23)

\[
\limsup_{n \to \infty} \mathbb{P}(\text{dist}_{NR_n(w)}(o_1, o_2) \leq 2k_n) \geq \zeta^2,
\] (6.3.24)

with \( \zeta > 0 \) the survival probability of the underlying branching process approximation to the neighborhoods of \( NR_n(w) \) identified in Theorem 3.15. For (6.3.23), we split, for some \( m \geq 1 \),

\[
\mathbb{P}(\text{dist}_{NR_n(w)}(o_1, o_2) < \infty) \leq \mathbb{P}(\text{dist}_{NR_n(w)}(o_1, o_2) \leq 2m) + \mathbb{P}(|\partial B_m(o_1)| > 0, |\partial B_m(o_1)| > 0, \text{dist}_{NR_n(w)}(o_1, o_2) > 2m).
\] (6.3.25)

By Corollary 2.18, \( \mathbb{P}(\text{dist}_{NR_n(w)}(o_1, o_2) \leq 2m) = o(1) \), and, by (2.4.2) in Corollary 2.17,

\[
\mathbb{P}(|\partial B_m(o_1)| > 0, |\partial B_m(o_1)| > 0, \text{dist}_{NR_n(w)}(o_1, o_2) > 2m) = \mathbb{P}(|\partial B_m(o)| > 0) + o(1),
\] (6.3.26)

where \( \partial B_m(o) \) is the set of vertices at distance \( m \) in the local weak limit of \( NR_n(w) \) identified in Theorem 3.15. The right-hand side of (6.3.26) converges to \( \zeta^2 \) when \( m \to \infty \). This proves (6.3.23).

To prove (6.3.24), we fix \( m \geq 1 \) and write

\[
\mathbb{P}(2m < \text{dist}_{NR_n(w)}(o_1, o_2) \leq 2k_n) \geq \mathbb{P}(\text{dist}_{NR_n(w)}(o_1,\text{Giant}_n) \leq k_n, i = 1, 2, \text{dist}_{NR_n(w)}(o_1, o_2) > 2m) \geq \mathbb{P}(|\partial B_m(o_1)| > 0, |\partial B_m(o_1)| > 0, \text{dist}_{NR_n(w)}(o_1, o_2) > 2m) - 2\mathbb{P}(\text{dist}_{NR_n(w)}(o_1, \text{Giant}_n) < k_n, |\partial B_m(o_1)| > 0).
\] (6.3.27)

By (6.3.26), the first term converges to \( \zeta_m^2 \), which in turn converges to \( \zeta^2 \) when \( m \to \infty \).
For the second term, we condition on $B_m(o_1), B_m(o_2)$, and use that $\partial B_m(o_1)$ is measurable w.r.t. $B_m(o_1)$ to obtain

$$
P\left( \text{dist}_{\text{NR}_n}(w)(o_1, \text{Giant}_n) < k_n \mid |B_m(o_1)| > 0 \right) = \mathbb{E}\left[ P\left( \text{dist}_{\text{NR}_n}(w)(o_1, \text{Giant}_n) > k_n \mid B_m(o_1) \right) 1_{\{|\partial B_m(o_1)| > 0\}} \right].$$

By Proposition 6.11,

$$
P\left( \text{dist}_{\text{NR}_n}(w)(o_1, \text{Giant}_n) > k_n \mid B_m(o_1) \right) \leq c e^{-c^{\star} W_m(o_1)^\eta}.$$  \tag{6.3.29}

By Theorem 3.15 and Conditions 1.1(a)-(b), we obtain that

$$W_m(o_1)^\eta \overset{d}{\to} \sum_{i=1}^{\|\partial B_m(o_1)\|} W_i^{\star},$$ \tag{6.3.30}

where $(W_i^{\star})_{i \geq 1}$ are i.i.d. copies of random variables with distribution function $F^{\star}$. Therefore,

$$W_m(o_1) \overset{p}{\to} \infty \tag{6.3.31}$$

when first $n \to \infty$ followed by $m \to \infty$, and we use that $|\partial B_m(o)| \overset{p}{\to} \infty$ since $|\partial B_m(o)| > 0$. As a result,

$$P\left( \text{dist}_{\text{NR}_n}(w)(o_1, \text{Giant}_n) > k_n \mid B_m(o_1) \right) 1_{\{|\partial B_m(o_1)| > 0\}} \overset{p}{\to} 0, \tag{6.3.32}$$

which by Lebesgue Dominated Convergence Theorem [Volume 1, Theorem A.1] implies that

$$\mathbb{E}\left[ e^{-c^{\star} W_m(V_1)^\eta} 1_{\{|\partial B_m(o_1)| > 0\}} \right] \to 0, \tag{6.3.33}$$

when first $n \to \infty$ followed by $m \to \infty$. This proves (6.3.24), and thus completes the proof of the upper bound in Theorem 6.3.

### 6.4 Path counting and log upper bound for finite-variance weights

In this section, we give the proof of the log upper bound for finite-variance weights. For this, we use the second moment method to show that whp there exists a path of at most $(1 + \varepsilon) \log \log n$ edges between $o_1$ and $o_2$. To apply the second-moment method, we give a bound on the variance of the number of paths of given lengths using path-counting techniques. This section is organised as follows. In Section 6.4.1, we highlight the path-counting techniques. In Section 6.4.2, we apply these methods to give upper bounds on distances for finite-variance weights. We also investigate the case where $\tau = 3$, and prove that typical distances are bounded by $\log n / \log \log n$ under appropriate conditions as well. We close in Section 6.4.3 by proving Theorems 3.16 and 6.1(ii).
6.4 Path counting and log upper bound for finite-variance weights

6.4.1 Path-counting techniques

In this section, we study path-counting techniques in the context of inhomogeneous random graphs. We generalize the setting somewhat, and consider an IRG on the vertices \( I \) with edge probabilities \( p_{ij} = u_i u_j \), for some weights \((u_i)_{i \in I}\). We obtain \( CL_n(w) \) by taking \( u_i = w_i / \sqrt{\ell_n} \) and \( I = [n] \). Since the \( NR_n(w) \) random graph is closely related to \( CL_n(w) \), this suffices for our purposes.

For \( a, b \in I \) and \( k \geq 1 \), let

\[
N_k(a, b) = \# \{ \pi \in P_k(a, b): \pi \text{ occupied} \}
\]

(6.4.1)
denote the number of paths of length \( k \) between the vertices \( a \) and \( b \). Let

\[
n_k(a, b) = \mathbb{E}[N_k(a, b)]
\]

(6.4.2)
denote the expected number of occupied paths of length \( k \) connecting \( a \) and \( b \). Define

\[
\bar{n}_k(a, b) = u_a u_b \left( \sum_{i \in I \setminus \{a, b\}} u_i^2 \right)^{k-1}, \quad n_k(a, b) = u_a u_b \left( \sum_{i \in I_{a, b, k}} u_i^2 \right)^{k-1},
\]

(6.4.3)
where \( I_{a, b, k} \) is the subset of \( I \) in which \( a \) and \( b \), as well as the \( k + 2 \) vertices with highest weights have been removed. In Section 6.2, we have implicitly proved an upper bound on \( \mathbb{E}[N_k(a, b)] \) of the form (see also Exercise 6.15)

\[
n_k(a, b) \leq \bar{n}_k(a, b).
\]

(6.4.4)

In this section, we will prove that \( n_k(a, b) \) is a lower bound on \( n_k(a, b) \), and use these bounds to prove a variance bound on \( N_k(a, b) \).

Before stating our main result, we introduce some notation. Let

\[
\nu_z = \sum_{i \in I} u_i^2, \quad \gamma_z = \sum_{i \in I} u_i^3
\]

(6.4.5)
denote the sums of squares and third powers of \((u_i)_{i \in I}\), respectively. Our aim is to show that w.h.p paths of length \( k \) exist between the vertices \( a \) and \( b \) for an appropriate choice of \( k \). We do this by applying a second-moment method on \( N_k(a, b) \), for which we need a lower bound on \( \mathbb{E}[N_k(a, b)] \) and an upper bound on \( \text{Var}(N_k(a, b)) \), which are such that \( \text{Var}(N_k(a, b)) = o(\mathbb{E}[N_k(a, b)]^2) \) (recall [Volume 1, Theorem 2.18]). We prove lower bounds on \( \mathbb{E}[N_k(a, b)] \) and upper bounds on \( \text{Var}(N_k(a, b)) \) in the following proposition, which is interesting in its own right:

**Proposition 6.12 (Variance of numbers of paths)** For any \( k \geq 1 \), \( a, b \in I \) and \((u_i)_{i \in I}\),

\[
\mathbb{E}[N_k(a, b)] \geq n_k(a, b),
\]

(6.4.6)
while, assuming that \( \nu_z > 1 \),

\[
\text{Var}(N_k(a, b)) \leq n_k(a, b)
\]

(6.4.7)

\[
+ \bar{n}_k(a, b)^2 \left( \frac{\gamma_z \nu_z^2}{\nu_z - 1} \left( \frac{1}{u_a} + \frac{1}{u_b} \right) + \frac{\gamma_z^2 \nu_z}{u_a u_b (\nu_z - 1)^2} + e_k \right),
\]

where

\[
e_k = \frac{1}{\sqrt{\ell_n}} \sum_{i \in I \setminus \{a, b\}} \frac{1}{u_i} \left( \sum_{j \neq i} p_{ij} \right),
\]

and

\[
e = \frac{1}{\sqrt{\ell_n}} \sum_{i \in I \setminus \{a, b\}} \frac{1}{u_i} \left( \sum_{j \neq i} p_{ij} \right).
\]
where
\[ e_k = \left( 1 + \frac{\gamma_x}{u_a \nu_z} \right) \left( 1 + \frac{\gamma_x}{u_b \nu_z} \right) \frac{\nu_z}{\nu_z - 1} \left( e^{2k^3 \gamma_2^2 / \nu_z^2} - 1 \right). \]

The random variable \( N_k(a,b) \) is a sum of indicators. When all these indicators would be independent, then the upper bound \( n_k(a,b) \) would hold. The second term on the right-hand side of (6.4.6) accounts for the positive dependence between the indicators of two paths being occupied.

We apply Proposition 6.12 in cases where \( \mathbb{E}[N_k(a,b)] = n_k(a,b) \to \infty \), by taking \( \mathcal{I} \) is a large subset of \([n]\) and \( u_i = w_i / \sqrt{n} \). In this case, \( \nu_z \approx \nu_a \approx \nu \approx 1 \). In our applications of Proposition 6.12, the ratio \( \bar{n}_k(a,b) / n_k(a,b) \) will be bounded, and \( k^3 \gamma_2^2 / \nu_z^2 = o(1) \), so that the last term is an error term. The starting and end vertices \( a, b \in \mathcal{I} \) will correspond to a union of vertices in \([n]\) of quite large size. As a result, \( \gamma_x / u_a \) and \( \gamma_x / u_a \) are typically small, so that
\[ \text{Var}(N_k(a,b)) = \mathbb{E}[N_k(a,b)]^2 \approx \frac{\gamma_x \nu_z}{\nu_z - 1} \left( \frac{1}{u_a} + \frac{1}{u_b} \right) + \frac{\gamma^2 \nu_x}{u_a u_b (\nu_z - 1)^2}. \]

is small. The choice of \( a, b \) and \( \mathcal{I} \) is quite delicate, which explains why we formulate Proposition 6.12 in such generality.

**Proof** We note that \( N_k(a,b) \) is a sum of indicators
\[ N_k(a,b) = \sum_{\pi \in \mathcal{P}_k(a,b)} \mathbbm{1}(\pi \text{ occupied}). \]

As a result,
\[ \mathbb{E}[N_k(a,b)] = \sum_{\pi \in \mathcal{P}_k(a,b)} \mathbb{P}(\pi \text{ occupied}) = \sum_{\pi \in \mathcal{P}_k(a,b)} \prod_{l=0}^{k-1} u_{\pi_l} u_{\pi_{l+1}}. \]

For \( \pi \in \mathcal{P}_k(a,b) \), \( \pi_0 = a, \pi_k = b \). Further,
\[ \sum_{\pi \in \mathcal{P}_k(a,b)} \prod_{l=1}^{k-1} u_{\pi_l}^2 = \sum_{i_1, \ldots, i_{k-1} \in \mathcal{I} \setminus \{a,b\}} \prod_{l=1}^{k-1} u_{i_l}^2, \]

where we recall that \( \sum_{i_1, \ldots, i_{k-1} \in \mathcal{I}} \) denotes a sum over distinct indices. Each sum over \( i_j \) yields a factor that is at least \( \sum_{i \in \mathcal{I} \setminus \{a,b\}} u_i^2 \), which proves (6.4.6).

To compute \( \text{Var}(N_k(a,b)) \), we again start from (6.4.9), which yields
\[ \text{Var}(N_k(a,b)) = \sum_{\pi, \rho \in \mathcal{P}_k(a,b)} \left[ \mathbb{P}(\pi, \rho \text{ occupied}) - \mathbb{P}(\pi \text{ occupied}) \mathbb{P}(\rho \text{ occupied}) \right]. \]
For \( \pi, \rho \), we denote by \( \pi \cap \rho \) the edges the paths \( \pi \) and \( \rho \) have in common. The occupation statuses of \( \pi \) and \( \rho \) are independent precisely when \( \pi \cap \rho = \emptyset \), so that
\[
\text{Var}(N_k(a, b)) \leq \sum_{\pi, \rho \in \mathcal{P}_k(a, b) \atop \pi \cap \rho \neq \emptyset} \mathbb{P}(\pi, \rho \text{ occupied}). \tag{6.4.13}
\]

Define \( \rho \setminus \pi \) to be the edges in \( \rho \) that are not part of \( \pi \), so that
\[
\mathbb{P}(\pi, \rho \text{ occupied}) = \mathbb{P}(\pi \text{ occupied})\mathbb{P}(\rho \text{ occupied} \mid \pi \text{ occupied}) \tag{6.4.14}
\]
\[
= \prod_{l=0}^{k} u_{x_l} u_{x_{l+1}} \prod_{e \in \rho \setminus \pi} u_e u_{\bar{e}}.
\]
where, for an edge \( e = \{x, y\} \), we write \( \bar{e} = x, \bar{e} = y \). When \( \pi = \rho \), then
\[
\mathbb{P}(\pi, \rho \text{ occupied}) = \mathbb{P}(\pi \text{ occupied}), \tag{6.4.15}
\]
and this contributes \( n_k(a, b) \) to \( \text{Var}(N_k(a, b)) \). From now on, we consider \( \pi \neq \rho \).

The probability in (6.4.14) needs to be summed over all possible pairs of paths \((\pi, \rho)\) with \( \pi \neq \rho \) that share at least one edge. In order to do this effectively, we start by introducing some notation.

Let \( l = |\pi \cap \rho| \) denote the number of edges in \( \pi \cap \rho \), so that \( l \geq 1 \) precisely when \( \pi \cap \rho \neq \emptyset \). Since \( \pi \neq \rho \), \( l \leq k - 1 \). When \( \pi \neq \rho \), we have that \( l \leq k - 1 \), and since \( \pi \) and \( \rho \) are self-avoiding paths between \( a \) and \( b \), \( l \) cannot be equal to \( k - 1 \), so that we consider \( l \leq k - 2 \) from now on. Let \( k - l = |\rho \setminus \pi| \) be the number of edges in \( \rho \) that are not part of \( \pi \).

Let \( m \) denote the number of connected subpaths in \( \rho \setminus \pi \), so that \( m \geq 1 \) whenever \( \pi \neq \rho \). Since \( \pi_0 = \rho_0 = a \) and \( \pi_k = \rho_k = b \), these subpaths start and end in vertices along the path \( \pi \). We can view these subpaths as excursions of the path \( \rho \) from the walk \( \pi \). By construction, between two excursions, there is at least one edge that \( \pi \) and \( \rho \) have in common.

Fix \( m \). We define \( \text{Shape}(\pi, \rho) \), the shape of the pair \((\pi, \rho)\), by
\[
\text{Shape}(\pi, \rho) = (x_{m+1}, s_m, \bar{x}_m, \bar{s}_m, \bar{r}_m, \bar{r}_m), \tag{6.4.16}
\]
where
(1) \( x_{m+1} \in \mathbb{N}_0^{m+1} \) and \( x_j \geq 0 \) is the length of the subpath in \( \rho \cap \pi \) in between the \( (j - 1)\)th and \( j\)th subpath of \( \pi \setminus \rho \). Here \( x_1 \) is the number of common edges in the subpath of \( \rho \cap \pi \) that contains \( a \), while \( x_{m+1} \) is the number of common edges in the subpath of \( \rho \cap \pi \) that contains \( b \), so that \( x_1 \geq 0 \) and \( x_{m+1} \geq 0 \).
For \( j \in \{2, \ldots, m\} \), \( x_j \geq 1 \);
(2) \( s_m \in \mathbb{N}^m \) and \( s_j \geq 1 \) is the number of edges in the \( j\)th subpath of \( \pi \setminus \rho \);
(3) \( \bar{x}_m \in \mathbb{N}^m \) and \( \bar{x}_j \geq 1 \) is the number of edges in the \( j\)th subpath of \( \rho \setminus \pi \);
(4) \( \bar{s}_m \in [m+1]^{m+1} \) and \( \bar{s}_j \) is the order of the \( j\)th common subpath in \( \rho \cap \pi \) of the path \( \pi \) in \( \rho \), i.e., \( \bar{s}_1 = 5 \) means that the second subpath that \( \pi \) has in common with \( \rho \) is the 5th subpath that \( \rho \) has in common with \( \pi \). Note that \( o_1 = 1 \) and \( o_{m+1} = m + 1 \), since \( \pi \) and \( \rho \) start and end in \( a \) and \( b \), respectively;
Figure 6.3 An example of a pair of paths \((\pi, \rho)\) and its corresponding shape.

(5) \(\bar{r}_{m+1} \in \{0,1\}^{m+1}\) is such that \(r_j\) describes the direction in which the \(j\)th common subpath in \(\rho \cap \pi\) of the path \(\pi\) is traversed by \(\rho\), with \(r_j = 1\) when the direction is the same for \(\pi\) and \(\rho\) and 0 otherwise. Thus, \(r_1 = r_{m+1} = 1\).

The information in \(\text{Shape}(\pi, \rho)\) is precisely what is necessary to piece together the topology of the two paths, except for the information of the vertices involved in \(\pi\) and \(\rho\). See Figure 6.3 for an example of a pair of paths \((\pi, \rho)\) and its corresponding shape.

With \(l = |\pi \cap \rho|\), we have

\[
\sum_{j=1}^{m+1} x_j = l, \quad \sum_{j=1}^{m} s_j = \sum_{j=1}^{m} t_j = k - l. \tag{6.4.17}
\]

Let \(\text{Shape}_{m,l}\) denote the set of shapes corresponding to pairs of paths \((\pi, \rho)\) with \(m\) excursions and \(l\) common edges, so that (6.4.17) hold. Then,

\[
\text{Var}(N_k(a, b)) \leq n_k(a, b) + \sum_{l=1}^{k-2} \sum_{m=1}^{k-l} \sum_{\sigma \in \text{Shape}_{m,l}} \sum_{\pi, \rho \in \mathcal{P}_k(a, b) \cap \text{Shape}(\pi, \rho) = \sigma} \mathbb{P}(\pi, \rho \text{ occupied}). \tag{6.4.18}
\]

When \(\text{Shape}(\pi, \rho) = \sigma\) for some \(\sigma \in \text{Shape}_{m,l}\), and since \(\pi\) and \(\rho\) both start and end in \(a\) and \(b\), the union of paths \(\pi \cup \rho\) visits \(k + 1 + l - m\) distinct vertices. The vertex \(a\) is in \(1 + \delta_{x_1,0}\) edges, and \(b\) in \(1 + \delta_{x_{m+1},0}\) edges. Of the other \(k - 1\) vertices in \(\pi\), precisely \(2m - \delta_{x_1,0} - \delta_{x_{m+1},0}\) are part of three edges, and \(k - 1 - 2m + \delta_{x_1,0} + \delta_{x_{m+1},0}\) are part of two edges. The remaining \(k - l - m\) vertices in \(\rho\) that are not part of \(\pi\) are part of precisely 2 edges. By construction, the \(k + 1\) vertices of both \(\pi\) and \(\rho\) are disjoint, but the remaining \(k - l - m\) vertices in \(\rho\) may intersect those of \(\pi\). Therefore, denoting \(a_1 = \delta_{x_1,0}, a_{m+1} = \delta_{x_{m+1},0}\),

\[
\mathbb{P}(\pi, \rho \text{ occupied}) = u_{a_1}^1 u_{a_{m+1}}^1 \prod_{s=1}^{2m-a_1-a_{m+1}} u_{v_s}^3 \prod_{t=2m-a_1-a_{m+1}+1}^{2(k-1)-l-m} u_{v_t}^2. \tag{6.4.19}
\]

where \((v_1, \ldots, v_{k+1+l-m}) \in \mathcal{I}^{k-1+l-m}\).

For a fixed \(\sigma \in \text{Shape}_{m,l}\) now bound the sum over \(\pi, \rho \in \mathcal{P}_k(a, b)\) such that
6.4 Path counting and log upper bound for finite-variance weights 243

Shape(π, ρ) = σ from above by summing (6.4.19) over all \((v_1, \ldots, v_{k-1+l-m}) \in \mathcal{I}^{k-1+l-m}\), to obtain for any \(\sigma \in \text{Shape}_{m,l}\),

\[
\sum_{\pi, \rho \in \mathcal{P}_{l}(a,b) \atop \text{Shape}(\pi, \rho) = \sigma} \mathbb{P}(\pi, \rho \text{ occupied}) \leq u_a u_b \frac{2^m}{\gamma_x} \nu_x^{2k-1-3m-l} \left(\frac{u_a \nu_x^2}{\gamma_x}\right)^{\delta_{x_1,0}} \left(\frac{u_b \nu_x^2}{\gamma_x}\right)^{\delta_{x_{m+1},0}}
\]

\[
= \bar{n}_k(a, b)^2 \gamma_x^{2(m+1)} \nu_x^{-3(m-1)-l} \left(\frac{\gamma_x}{u_a \nu_x}\right)^{1-\delta_{x_1,0}} \left(\frac{\gamma_x}{u_b \nu_x}\right)^{1-\delta_{x_{m+1},0}}.
\]

Therefore, we arrive at

\[
\text{Var}(N_k(a, b)) \leq n_k(a, b) + \bar{n}_k(a, b)^2 \sum_{l=1}^{k-2} \sum_{m=1}^{k-l} \gamma_x^{2(m-1)} \nu_x^{-3(m-1)-l} \left(\frac{\gamma_x}{u_a \nu_x}\right)^{1-\delta_{x_1,0}} \left(\frac{\gamma_x}{u_b \nu_x}\right)^{1-\delta_{x_{m+1},0}}.
\]

We continue to compute the number of shapes in the following lemma:

**Lemma 6.13** (The number of shapes) Fix \(m \geq 1\) and \(l \leq k-2\). For \(m = 1\), the number of shapes in \(\text{Shape}_{m,l}\) with fixed \(a_1 = \delta_{x_1,0}, a_{m+1} = \delta_{x_{m+1},0}\) equals \(l\) when \(a_1 = a_{m+1} = 0\), \(1\) when \(a_1 + a_{m+1} = 1\) and \(0\) when \(a_1 = a_{m+1} = 1\). For \(m \geq 2\), the number of shapes in \(\text{Shape}_{m,l}\) with fixed \(a_1 = \delta_{x_1,0}, a_{m+1} = \delta_{x_{m+1},0}\) is bounded by

\[
2^{m-1}(m-1)! \left(\frac{k-l-1}{m-1}\right)^2 \left(\frac{l}{m-a_1-a_{m+1}}\right).
\]

Consequently, for all \(m \geq 2\),

\[
|\text{Shape}_{m,l}| \leq k \left(\frac{2k^3}{(m-1)!}\right)^{m-1}.
\]

**Proof** Since \(r_1 = r_{m+1} = 1\), there are \(2^{m-1}\) directions in which the common parts can be traversed. Since there are \(m\) distinct parts, there are \(m+1\) common parts. The first part contains vertex \(a\), the last part contains vertex \(b\). Thus, there are \((m-1)!\) orders \(\bar{o}_{m+1}\) of the common parts when we have fixed the directions in which the paths can be traversed.

In counting the number of \(\bar{x}_{m+1}, \bar{s}_m, \bar{t}_m\), we repeatedly use the fact that there are \((a+b-1)\) possible sequences \((y_1, \ldots, y_b) \in \mathbb{N}_0^b\) such that \(\sum_{j=1}^{b} y_j = a\). This can be seen by representing \(a\) as a sequence of \(a\) ones, separated by \(b-1\) zeros. We draw \(b\) zeros, which we can do in \((a+b-1)\) possible ways. Then, we note that a sequence \((y_1, \ldots, y_b) \in \mathbb{N}_0^b\) such that \(\sum_{j=1}^{b} y_j = a\) can be obtained uniquely by letting \(y_j\) be the number of ones in between the \((i-1)\)th and \(i\)th chosen zero. Similarly, there are \((a-b)\) possible sequences \((y_1, \ldots, y_b) \in \mathbb{N}_0^b\) such that \(\sum_{j=1}^{b} y_j = a\), since we can apply the previous equality to \((y_1-1, \ldots, y_b-1) \in \mathbb{N}_0^b\).
Using the above, we continue to count the number of shapes. The number of \((s_1, \ldots, s_m) \in \mathbb{N}^m\) such that \(s_j \geq 1\) and \(\sum_{j=1}^m s_j = k - l\) equals

\[
\binom{k - l - 1}{m - 1}.
\] (6.4.24)

The same applies to \((t_1, \ldots, t_m) \in \mathbb{N}^m\) such that \(t_j \geq 1\) and \(\sum_{j=1}^m t_j = k - l\). In counting the number of possible \(\vec{x}_{m+1}\) such that \(\sum_{j=1}^{m+1} x_j = l\), we need to count their numbers separately for \(x_1 = 0\) and \(x_1 \geq 1\), and for \(x_{m+1} = 0\) and \(x_{m+1} \geq 1\). When \(m = 1\), the number is zero when \(x_1 = 0\) and \(x_2 = 0\), since \(x_1 = x_2 = 0\) implies that the paths share no edges. Denote \(a_1 = \delta_{x_1,0}, a_{m+1} = \delta_{x_{m+1},0}\), and suppose that \(m - a_1 - a_{m+1} \geq 0\). Then, there are

\[
\binom{l}{m - a_1 - a_{m+1}}
\] (6.4.25)

possible choice of \(\vec{x}_{m+1}\) with fixed \(a_1 = \delta_{x_1,0}, a_{m+1} = \delta_{x_{m+1},0}\). The claim follows by multiplying these bounds on the number of choices for \(\vec{r}_{m+1}, \vec{o}_{m+1}, \vec{s}_{m}, \vec{t}_{m}\) and \(\vec{x}_{m+1}\).

To prove (6.4.23), we continue by bounding

\[
\binom{k - l - 1}{m - 1} \frac{1}{(m - 1)!} \left( \frac{(k - l - 1)!}{(k - l - m)!} \right)^2 \leq \frac{k^{2(m-1)}}{(m - 1)!}.
\] (6.4.26)

and, using that \(\binom{a}{b} \leq a^b/b!\) and \(l \leq k\),

\[
\frac{l^{m-a_1-a_{m+1}}}{(m - a_1 - a_{m+1})!} \leq k^m.
\] (6.4.27)

Therefore, the number of shapes in \(\text{Shape}_{m,l}\) is, for each \(l \geq 1\) and \(m \geq 2\), bounded by

\[
2^{m-1} \frac{k^{2(m-1)}}{(m - 1)!} k^m = k \frac{(2k^3)^{m-1}}{(m - 1)!}.
\] (6.4.28)

We are now ready to complete the proof of Proposition 6.12:

**Proof of Proposition 6.12.** By (6.4.21) and applying Lemma 6.13, it suffices to sum

\[
2^{m-1} (m - 1)! \binom{k - l - 1}{m - 1}^2 \frac{l}{m - a_1 - a_{m+1}}
\] (6.4.29)

\[
\times \left( \frac{2\gamma_x}{\nu_2} \right)^{m-1} \nu_x^{-l} \left( \frac{\gamma_x}{u_0 \nu_2} \right)^{1-a_1} \left( \frac{\gamma_x}{u_0 \nu_2} \right)^{1-a_{m+1}}
\]

over \(l \in [k-2], m \in [k-l]\) and \(a_1, a_{m+1} \in \{0, 1\}\), where, by convention, \(\binom{l}{-1} = 0\).
6.4 Path counting and log upper bound for finite-variance weights

We start with $m = 1$, for which we obtain that the sum of (6.4.29) over the other variables equals

$$
\gamma \left( \frac{1}{u_a} + \frac{1}{u_b} \right) \sum_{l=1}^{\infty} \nu_z^{-(l-1)} + \frac{\gamma_z^2}{u_a u_b \nu_z} \sum_{l=1}^{\infty} l \nu_z^{-(l-1)}
$$

$$
= \frac{\gamma_z \nu_z^2}{\nu_z - 1} \left( \frac{1}{u_a} + \frac{1}{u_b} \right) + \frac{\gamma_z^2 \nu_z}{u_a u_b (\nu_z - 1)^2}, \quad (6.4.30)
$$

where we use that, for $a \in [0, 1)$,

$$
\sum_{l=0}^{\infty} a^{-l} = a/(1 - a), \quad \sum_{l=0}^{\infty} l a^{-l} = a^2/(1 - a)^2. \quad (6.4.31)
$$

The terms in (6.4.30) are the first two terms appearing on the right-hand side of (6.4.7).

This leaves us to bound the contribution when $m \geq 2$. Since (6.4.23) is independent of $l$, we can start by summing (6.4.29) over $l \geq 1$, and over $a_1, a_{m+1} \in \{0, 1\}$ to obtain a bound of the form

$$
k(1 + \frac{\gamma_z}{u_a \nu_z}) (1 + \frac{\gamma_z}{u_b \nu_z}) \frac{\nu_z}{\nu_z - 1} \sum_{m \geq 2} (\frac{2k^3}{m-1})^{m-1} \frac{\gamma_z^2}{\nu_z^2}^{m-1} \quad (6.4.32)
$$

$$
= k(1 + \frac{\gamma_z}{u_a \nu_z}) (1 + \frac{\gamma_z}{u_b \nu_z}) \frac{\nu_z}{\nu_z - 1} (e^{2k^3 \gamma_z^2 / \nu_z^2} - 1).
$$

The term in (6.4.32) is the last term appearing on the right-hand side of (6.4.7).

Summing the bounds in (6.4.30) and (6.4.32) proves (6.4.7).

Exercises 6.16–6.18 study various consequences of our path-counting techniques. In the next section, we use Proposition 6.12 to prove lower bounds on graph distances.

6.4.2 Logarithmic distance bounds for finite-variance weights

In this section, we prove that two uniformly chosen vertices that are conditioned to be connected are with high probability within distance $(1 + \varepsilon) \log \nu n$, as formulated in the following theorem:

**Theorem 6.14** (Logarithmic upper bound graph distances $\mathcal{NR}_n(w)$) Assume that Conditions 1.1(a)-(c) hold, where $\nu = \mathbb{E}[W^2]/\mathbb{E}[W] \in (1, \infty)$. Then, for any $\varepsilon > 0$,

$$
\mathbb{P}(\text{dist}_{\mathcal{NR}_n(w)}(o_1, o_2) \leq (1 + \varepsilon) \log \nu n \mid \text{dist}_{\mathcal{NR}_n(w)}(o_1, o_2) < \infty) = 1 + o(1). \quad (6.4.33)
$$

The same results hold for $\mathcal{CL}_n(w)$ and $\mathcal{GR}_n(w)$ under the same conditions.

**Organization of the proof of Theorem 6.14**

We prove Theorem 6.14 by combining the branching process comparison to a second moment method using Proposition 6.12 on the number of paths of a given length. More precisely, we fix $m \geq 1$ large, and recall that $B_m(o_1)$ and $B_m(o_2)$
denote the vertices at distance at most $m$ from $o_1$ and $o_2$ respectively, and let $\partial B_m(o_1)$ and $\partial B_m(o_2)$ denote the vertices at distance precisely equal to $m$. We condition on $B_m(o_1)$ and $B_m(o_2)$ such that $\partial B_m(o_1) \neq \emptyset$ and $\partial B_m(o_2) \neq \emptyset$. By the local weak convergence in Theorem 3.11, the probabilities of the latter event is close to $\zeta_m$, where $\zeta_m = P(\partial B_m(o) \neq \emptyset)$ is the probability that the branching process that is the local weak limit of $\mathcal{N}_m(w)$ survives to generation $m$. See also Corollary 2.17 that proves the asymptotic independence of the neighborhoods of $o_1$ and $o_2$. Then, $\zeta_m \to \zeta$ when $m \to \infty$, and, conditionally on $|\partial B_m(o)| > 0$, $|\partial B_m(o)| \geq M$ whp, for any $M$ and as $m \to \infty$. This explains the branching-process approximation.

We now state the precise branching-process approximation result that we will rely upon. We take $u_i = w_i/\sqrt{t_n}$,

$$a = \partial B_m(o_1), \quad b = \partial B_m(o_2),$$

so that

$$u_a = \frac{1}{\sqrt{t_n}} \sum_{i \in \partial B_m(o_1)} w_i = W_m(o_1)/\sqrt{t_n},$$

$$u_b = \frac{1}{\sqrt{t_n}} \sum_{i \in \partial B_m(o_2)} w_i = W_m(o_2)/\sqrt{t_n}.$$

We formalize the above ideas in the following lemma:

**Lemma 6.15 (Branching process approximation)** As $n \to \infty$,

$$(W_m(o_1), W_m(o_2)) \stackrel{d}{\to} \left( \sum_{j=1}^{Z_m^{(1)}} W^{*1}(j), \sum_{j=1}^{Z_m^{(2)}} W^{*2}(j) \right),$$

where $(Z_m^{(1)}, Z_m^{(2)})_{m \geq 0}$ are the generation sizes of two independent branching processes, and $(W^{*1}(j))_{j \geq 1}$ and $(W^{*2}(j))_{j \geq 1}$ are two independent sequences of i.i.d. random variables with distribution $F^*$.

**Proof** By Corollary 2.17, $|\partial B_m(o_1)|$ and $|\partial B_m(o_2)|$ jointly converge in distribution to $(Z_m^{(1)}, Z_m^{(2)})$, which are independent generation sizes of the local weak limit of $\mathcal{N}_m(w)$ as in Theorem 3.15. Each of the individuals in $\partial B_m(o_1)$ and $\partial B_m(o_2)$ receives a mark $M_i$, and its weight is $w_{M_i}$. By Proposition 3.13, these marks are i.i.d. random variables conditioned to be unthinned. Whp no vertex in $B_m(o_1) \cup B_m(o_2)$ is thinned. Then, $W_m(o_1) = \sum_{j=1}^{Z_m^{(1)}} W^*_n(j)$, where $(W^*_n(j))_{j \geq 1}$ are i.i.d. copies of $W^*_n$. By Condition 1.1(a), $W^*_n \stackrel{d}{\to} W^*$, so that $W_m(o_1) \stackrel{d}{\to} \sum_{j=1}^{Z_m^{(1)}} W^{*1}(j)$. The joint convergence follows in a similar fashion, now using local weak convergence in probability. 

**Second moment method and path counting**

Fix $k = k_n = (1 + \varepsilon) \log n - 2m$. We next present the details of the second moment method that shows that whp, on the event that $\partial B_m(o_1) \neq \emptyset$ and
\[ \partial B_m(o_2) \neq \emptyset, \text{ there exist a path of length } k_n - 2m \text{ connecting } \partial B_m(o_1) \text{ and } \partial B_m(o_2). \] This ensures that, on the event that \( \partial B_m(o_1) \neq \emptyset \) and \( \partial B_m(o_2) \neq \emptyset, \) the event \( \text{dist}_{n;u}(o_1, o_2) \leq k_n - 2m \) occurs whp. For this, we take \( u_i = w_i/\sqrt{t_n}. \)

We aim to apply Proposition 6.12, for which we fix \( K \geq 1 \) sufficiently large and take \( a = \partial B_m(o_1), b = \partial B_m(o_2) \) and

\[ I_{a,b} = \{ i \in [n]: w_i \leq K \} \setminus (B_m(o_1) \cup B_m(o_2)). \quad (6.4.37) \]

In order to apply Proposition 6.12, we start by investigating the constants appearing in it:

**Lemma 6.16** (Parameters in path counting) Conditionally on \( B_m(o_1) \) and \( B_m(o_2), \) and with \( a = \partial B_m(o_1), b = \partial B_m(o_2), \) for \( k = (1 + \varepsilon) \log n, \)

\[ n_k(a,b) \xrightarrow{p} \infty, \quad \bar{n}_k(a,b) = (1 + o_1(1))n_k(a,b), \quad (6.4.38) \]

and, as \( n \to \infty, \)

\[ \frac{\text{Var}(N_k(a,b))}{\mathbb{E}[N_k(a,b)]^2} \leq \frac{K^2 \nu}{\nu - 1} \left( \frac{1}{\sqrt{t_n u_a}} + \frac{1}{\sqrt{t_n u_b}} \right) + \frac{K^2 \nu}{(\nu - 1) t_n u_a u_b} + o_1(1). \quad (6.4.39) \]

**Proof** By (6.4.3),

\[ n_k(a,b) = u_a u_b \nu_{a,b}^{k-1}, \quad (6.4.40) \]

and

\[ \bar{n}_k(a,b) = (\nu_{a,b}/\nu_{a,b,k})^{k-1}. \quad (6.4.41) \]

We start by investigating \( \nu_z. \) Denote

\[ \nu(K) = \frac{\mathbb{E}[W^2 1_{W \leq K}]}{\mathbb{E}[W]}. \quad (6.4.42) \]

Then, by (6.4.37) and the fact that \( B_m(o_1) \) and \( B_m(o_2) \) contain a finite number of vertices,

\[ \lim_{n \to \infty} \nu_{a,b} = \nu(K). \quad (6.4.43) \]

The same applies to \( \nu_{a,b,k}. \) Then, with \( K > 0 \) chosen so large that \( \nu(K) \geq \nu - \varepsilon/2 \) and with \( k = (1 + \varepsilon) \log n, \)

\[ n_k(a,b) = u_a u_b \nu_{a,b}^{k-1} = \frac{W_m(o_1) W_m(o_2)}{\ell_n} n^{(1 + \varepsilon) \log \nu \nu_{a,b}/\nu + 1} \xrightarrow{p} \infty, \quad (6.4.44) \]

where \( K \) and \( n \) are so large that \( (1 + \varepsilon) \nu(K)/\nu > 1. \) This proves the first property in (6.4.38).

To prove the second property in (6.4.38), we note that the set \( I_{a,b,k} \) is obtained from \( I_{a,b} \) by removing the \( k \) vertices with highest weight. Since \( w_i \leq K \) for all \( i \in I \) (recall (6.4.37)), \( \nu_{a,b} \leq \nu_{a,b,k} + k K/\ell_n. \) Since \( k \leq A \log n, \) we therefore arrive at

\[ \frac{\bar{n}_k(a,b)}{n_k(a,b)} \leq (1 + k K/\ell_n \nu_{a,b,k})^{k-1} = e^{k^2 K/(\ell_n \nu_{a,b,k})} \xrightarrow{p} 1, \quad (6.4.45) \]
as required.

To prove (6.4.39), we rely on Proposition 6.12. We have already shown that \( n_k(a, b) = \mathbb{E}[N_k(a, b)] \to \infty \), so that the first term on the right-hand side of (6.4.7) is \( o(\mathbb{E}[N_k(a, b)]^2) \). Further, by (6.4.37),

\[
\gamma_z \leq \gamma_z(\max_{i \in I} u_i) \leq \frac{\nu_z K}{\sqrt{\ell_n}},
\]

so that, for \( k \leq A \log n \) with \( A > 1 \) fixed,

\[
(1 + \frac{\gamma_z}{u_k \nu_z})(1 + \frac{\gamma_z}{u_b \nu_z})k(e^{2k \gamma_z^2 / \nu_z^2} - 1) = o(1).
\]

(6.4.47)

Substituting these bounds into (6.4.39) and using (6.4.38) yields the claim. \( \square \)

**Completion of the proof of Theorem 6.14**

Now we are ready to complete the proof of Theorem 6.14. We must show that

\[
\mathbb{P}(k_n < \text{dist}_{n_n(o_1)}(o_1, o_2) < \infty) = o(1).
\]

(6.4.48)

Indeed, then \( \mathbb{P}(< k_n \ | \text{dist}_{n_n(o_1)}(o_1, o_2) < \infty) = o(1) \) since, \( \mathbb{P}(\text{dist}_{n_n(o_1)}(o_1, o_2) < \infty) \to \zeta^2 > 0 \) by Theorem 3.17. We rewrite

\[
\mathbb{P}(k_n < \text{dist}_{n_n(o_1, o_2)} < \infty)
\]

(6.4.49)

\[
= \mathbb{P}(k_n < \text{dist}_{n_n(o_1, o_2)} < \infty, \partial B_m(o_1) \neq \emptyset, \partial B_m(o_2) \neq \emptyset)
\]

\[
\leq \mathbb{P}(N_{k_n - 2m}(\partial B_m(o_1, \partial B_m(o_2)) = 0, \partial B_m(o_1) \neq \emptyset, \partial B_m(o_2) \neq \emptyset).
\]

Recall that \( k = k_n = (1 + \varepsilon) \log n \). By the Chebyshev inequality [Volume 1, Theorem 2.18], the conditional probability of \( \{\text{dist}_{n_n(o_1)}(o_1, o_2) < k_n\} \) given \( B_m(o_1), B_m(o_2) \) is at most

\[
\text{Var}(N_{k_n - 2m}(a, b)) = \frac{K^2 \nu^2}{\nu - 1} \left( \frac{1}{\sqrt{\ell_n u_a}} + \frac{1}{\sqrt{\ell_n u_b}} \right) + \frac{K^2 \nu^2}{(\nu - 1) \ell_n u_a u_b} + o(1).
\]

(6.4.50)

When \( \partial B_m(o_1) \neq \emptyset \) and \( \partial B_m(o_2) \neq \emptyset \), by (6.4.35),

\[
\frac{1}{\sqrt{\ell_n u_a}} + \frac{1}{\sqrt{\ell_n u_b}} \xrightarrow{\nu \to \infty} \left( \sum_{j=1}^{Z^{(1)}_{n_2}} W^{(1)}(j) \right)^{-1} + \left( \sum_{j=1}^{Z^{(2)}_{n_2}} W^{(2)}(j) \right)^{-1} \xrightarrow{\nu \to \infty} 0,
\]

(6.4.51)

when \( m \to \infty \). Therefore,

\[
\mathbb{P}\left(N_{k_n - 2m}(a, b) = 0 \ | \partial B_m(o_1) \neq \emptyset, \partial B_m(o_2) \neq \emptyset \right) \xrightarrow{\nu \to \infty} 0,
\]

(6.4.52)

and, by Lebesgues’ Dominated Convergence Theorem [Volume 1, Theorem A.1],

\[
\mathbb{P}(\text{dist}_{n_n(o_1)}(o_1, o_2) > k_n, \partial B_m(o_1) \neq \emptyset, \partial B_m(o_2) \neq \emptyset) \to 0,
\]

(6.4.53)

when first \( n \to \infty \) followed by \( m \to \infty \), which completes the proof. \( \square \)

We close this section by describing what happens when \( \tau = 3 \), and there are no slowly-varying functions.
6.4 Path counting and log upper bound for finite-variance weights

Distances for the critical case $\tau = 3$

When $\tau = 3$, $w_i$ is approximately $c(n/i)^{1/2}$. It turns out that this changes the distances only by a doubly logarithmic factor:

**Theorem 6.17** (Graph distances NR$_n$(w) in critical $\tau = 3$ case) Assume that Condition 1.1(a)-(b) hold, and that there exists constants $c_2 > c_1 > 0$ and $\alpha > 0$ such that for all $x \leq n^\alpha$,

$$[1 - F_n(x)] \geq c_1/x^2, \quad (6.4.54)$$

and for all $x \geq 0$,

$$[1 - F_n(x)] \leq c_2/x^2. \quad (6.4.55)$$

Then, conditionally on $\text{dist}_{NR_n(w)}(o_1, o_2) < \infty$,

$$\frac{\text{dist}_{NR_n(w)}(o_1, o_2) \log \log n}{\log n} \xrightarrow{p} 1. \quad (6.4.56)$$

The same results hold for CL$_n$(w) and GRG$_n$(w) under the same conditions.

The lower bound in Theorem 6.17 is already stated in Corollary 6.5. The upper bound can be proved using the path-counting techniques in Proposition 6.12 and adaptations of it. We now sketch the proof.

Let $\eta > 0$ and let

$$\alpha_n = e^{\nu_1 - \eta}. \quad (6.4.57)$$

Define the core of NR$_n$(w) to be

$$\text{Core}_n = \{i : w_i \geq \alpha_n\}. \quad (6.4.58)$$

The proof of Theorem 6.17 follows from the following two propositions:

**Proposition 6.18** (Typical distances in the core) Under the conditions of Theorem 6.17, let $V_1', V_2' \in \text{Core}_n$ be chosen with probability proportional to their weight, i.e.,

$$P(V_i' = j) = \frac{w_j}{\sum_{v \in \text{Core}_n} w_v}, \quad (6.4.59)$$

and let $H_n'$ be the graph distance between $V_1', V_2'$ in Core$_n$. Then, for any $\varepsilon > 0$, there exists an $\eta > 0$ such that

$$P(H_n' \leq \frac{(1 + \varepsilon) \log n}{\log \log n}) \to 1. \quad (6.4.60)$$

**Proposition 6.19** (From the periphery to the core) Under the conditions of Theorem 6.17, let $o_1, o_2$ be two vertices chosen uniformly at random from $[n]$. Then, for any $\eta > 0$,

$$P(d_{NR_n(w)}(o_1, \text{Core}_n) \leq \nu_1^{1-\eta}, d_{NR_n(w)}(o_2, \text{Core}_n) \leq \nu_1^{1-\eta}) \to \zeta^2. \quad (6.4.61)$$
Proof of Theorem 6.17 subject to Propositions 6.18–6.19. To see that Propositions 6.18–6.19 imply Theorem 6.17, we note that

\[
d_{NR_n(w)}(o_1, o_2) \leq d_{NR_n(w)}(o_1, Core_n) + d_{NR_n(w)}(o_2, Core_n) + d_{NR_n(w)}(V'_1, V'_2),
\]

where \(V'_1, V'_2 \in Core_n\) are the vertices in \(Core_n\) found first in the breadth-first search from \(o_1\) and \(o_2\), respectively. Then, by Proposition 3.13, \(V'_1, V'_2 \in Core_n\) are chosen with probability proportional to their weight. Fix \(k_n = \log n / \log \log n\). We conclude that, when \(n\) is so large that \(\nu_n^{1-\eta} \leq \varepsilon k_n / 4\),

\[
P(d_{NR_n(w)}(o_1, o_2) \leq (1 + \varepsilon)k_n) \geq P(d_{NR_n(w)}(o_1, Core_n) \leq \nu_n^{1-\eta}, i = 1, 2)
\times P(d_{NR_n(w)}(V'_1, V'_2) \leq (1 + \varepsilon / 2)k_n \mid d_{NR_n(w)}(o_1, Core_n) \leq \nu_n^{1-\eta}, i = 1, 2).
\]

By Proposition 6.19, the first probability converges to \(\zeta^2\), and by Proposition 6.18, the second probability converges to 1. We conclude that

\[
P(d_{NR_n(w)}(o_1, o_2) \leq (1 + \varepsilon) \frac{\log n}{\log \log n}) \to \zeta^2.
\]

Since also \(P(d_{NR_n(w)}(o_1, o_2) < \infty) \to \zeta^2\), this completes the proof.

The proofs of Propositions 6.18–6.19 follow from path-counting techniques similar to the ones carried out above. We now sketch their proofs, starting with Proposition 6.18:

Proof of Proposition 6.18. We take

\[
a = V'_1, \quad b = V'_2, \quad \mathcal{I} = \{w_i : w_i \in [K, \sqrt{\alpha_n}]\}.
\]

The whole point is that there exists a constant \(c > 0\) such that

\[
\nu_z \geq c \log \alpha_n = c \alpha_n^{1-\eta},
\]

while \(u_a \geq \sqrt{\alpha_n}, u_b \geq \sqrt{\alpha_n}, \) so that

\[
\mathbb{E}[N_k(a, b)] \approx \alpha_n^2 c^k \nu_n^{k(1-\eta)} / \ell_n \to \infty
\]

for \(k = \log n / ((1 - \eta) \log \alpha_n) \leq 1 + \varepsilon / 2\) when \(\eta\) is such that \(1 / (1 - \eta) \leq 1 + \varepsilon / 2\). Further,

\[
\gamma_z \leq \sqrt{\alpha_n} / \sqrt{\ell_n},
\]

so that \(\text{Var}(N_k(a, b)) / \mathbb{E}[N_k(a, b)]^2 \to 0\) by Proposition 6.12. Since \(\nu_n = \Theta(\log n)\) when \(\tau = 3\) (see (6.4.54)–(6.4.55) in Theorem 6.17), this completes the proof.

See Exercises 6.19–6.21 for some further properties of paths within the core.

Proof of Proposition 6.19. We again condition on \(\partial B_m(o_1) \neq \emptyset, \partial B_m(o_2) \neq \emptyset\), the probability of which converges to \(\zeta^2\) when \(n \to \infty\) followed by \(m \to \infty\).
6.4 Path counting and log upper bound for finite-variance weights 251

Then, we perform a second moment method on the number of paths between \( \partial B_m(o_1) \) and Core\(_n\). For this, we take \( k = \nu_\eta n \) and
\[
a = \partial B_m(o_1), \quad b = \text{Core}_n, \quad I = \{i : w_i \leq K\} \setminus (B_m(o_1) \cup B_m(o_2)).
\]
(6.4.68)

Then we follow the proof in (6.4.50)–(6.4.53) to show that
\[
P(d_{NR}(w)(o_1, \text{Core}_n) > \nu_\eta n, \partial B_m(o_1) \neq \varnothing, \partial B_m(o_2) \neq \varnothing) \to 0,
\]
(6.4.69)
as required. Note, for this, that, conditionally on \( B_m(o_1), B_m(o_2) \)
\[
E[N_k(a,b)] \approx \nu(K)^k \mathcal{W}_m(o_1) \frac{1}{n} \sum_{i \in \text{Core}_n} w_i,
\]
(6.4.70)
where \( \nu(K) \to \infty \) as \( K \to \infty \), and where, by (6.4.54),
\[
\frac{1}{n} \sum_{i \in \text{Core}_n} w_i \geq \alpha_n [1 - F_n](\alpha_n) \geq c/\alpha_n.
\]
(6.4.71)
Therefore, \( E[N_k(a,b)] \to \infty \) as soon as \( k \geq 2 \log \alpha_n / \log \nu(K) \), which is satisfied for \( K \) sufficiently large and \( k = \nu_\eta n \). Exercise 6.22 asks you to complete the above proof.

6.4.3 Distances for IRG\(_n(\kappa)\): Proofs of Theorems 3.16 and 6.1(ii)

In this section, we use the path-counting techniques in Section 6.4.1 to give some missing proofs for general inhomogeneous random graphs. We assume that \( (\kappa_n) \) is graphical sequence of kernels with limit irreducible \( \kappa \), and with \( \nu = \|T_\kappa\| \in (1, \infty) \). We start by proving Theorem 6.1(ii), and then we use it to prove Theorem 3.16.

Logarithmic upper bound on distances in Theorem 6.1(ii)

Without loss of generality, we may assume that \( \kappa \) is bounded, i.e., \( \sup_{x,y} \kappa(x,y) < \infty \). Indeed, we can always stochastically dominate \( v \) with an unbounded \( \kappa_n \) from below by IRG\(_n(\kappa_n)\) with a bounded kernel that approximates it arbitrarily well. Since graph distances increase by decreasing \( \kappa_n \), if we prove Theorem 6.1(ii) in the bounded case, the unbounded case will follow immediately. Further, we can approximate a bounded \( \kappa_n \) from above by a finite-type kernel. Therefore, it now suffices to prove Theorem 6.1(ii) for a kernel of finite type.

Let us set up the notation for the finite-types case. For \( u, v \in [n] \), we let \( \kappa_n(u, v) = \kappa_{n,i_u,i_v} \), where \( i_u \in [k] \) denotes the type of vertex \( u \in [n] \). We assume that, for all \( i, j \in [k] \),
\[
\lim_{n \to \infty} \kappa_{n,i,j} = \kappa_{i,j},
\]
(6.4.72)
and
\[
\lim_{n \to \infty} \mu_n(i) = \frac{1}{n} \# \{v \in [n] : i_v = i\} = \mu(i).
\]
(6.4.73)
In this case, \( \|T_{\kappa_n}\| \) is the largest eigenvalue of the matrix \( M_i^{n,j} = \kappa_{n,i,j} \mu_n(j) \),
which converges to the largest eigenvalue of the matrix \( M_{i,j} = \kappa_{n,i,j} \mu(j) \), which equals \( \nu = \| T_n \| \in (1, \infty) \), by assumption. This factor arises due to the depletion-of-points effect, which guarantees that the paths used are self-avoiding. Without loss of generality, we assume that \( \mu(i) > 0 \) for all \( i \in [k] \). This sets the stage of our analysis.

We fix \( m \geq 1 \), and assume that \( \partial B_m(o_1), \partial B_m(o_2) \neq \emptyset \). We will prove that
\[
P(d_{\text{IRG}, \kappa_n}(o_1, o_2) \leq (1 + \varepsilon) \log \nu n \mid B_m(o_1), B_m(o_2)) = 1 + o_{\varepsilon}(1). \tag{6.4.74}
\]
We follow the proof of Theorem 6.14, and rely on path-counting techniques. We again take
\[
a = \partial B_m(o_1), \quad b = \partial B_m(o_2), \tag{6.4.75}
\]
and
\[
\mathcal{I}_{a,b} = [n] \setminus (B_m(o_1) \cup B_m(o_2)). \tag{6.4.76}
\]
Recall that
\[
N_k(a, b) = \#\{\pi \in \mathcal{P}_k(a, b) : \pi \text{ occupied}\}. \tag{6.4.77}
\]
We aim to use the second moment method for \( N_k(a, b) \), for which we investigate the mean and variance of \( N_k(a, b) \). We compute
\[
\mathbb{E}[N_k(a, b)] = \sum_{\pi \in \mathcal{P}_k(a, b)} \mathbb{P}(\pi \text{ occupied}) = \sum_{\pi \in \mathcal{P}_k(a, b)} \prod_{i=0}^{k} \frac{\kappa_n(\pi_i, \pi_{i+1})}{n} \tag{6.4.78}
\]
where \( x = (x_i)_{i=1}^k, y = (y_i)_{i=1}^k \) with \( x_i \) the number of type \( i \) vertices in \( \partial B_m(o_1) \) and \( y_i \) the number of type \( i \) vertices in \( \partial B_m(o_2) \), respectively. An identical lower bound holds with an extra factor \( (\mu_n - k)/\mu_n \), where \( \mu_n = \min_{j \in [k]} \mu_n(j) \to \min_{j \in [k]} \mu(j) > 0 \), by assumption.

Recall the notation and results in Section 3.4, and in particular Theorem 3.8(b). The types of \( o_1 \) and \( o_2 \) are asymptotically independent, and the probability that \( o_1 \) has type \( j \) is equal to \( \mu_n(j) \), which converges to \( \mu(j) \). On the event that the type of \( o_1 \) equals \( j_1 \), the vector of the numbers of individuals in \( \partial B_m(o_1) \) converges in distribution to \( (Z_{m,n}^{(1,j_1)}(i))_{i \in [k]} \), which, by Theorem 3.8(b), is close to \( M_n x_n(i) \) for some strictly positive random variable \( M_n \). We conclude that
\[
x \xrightarrow{d} Z_{m,n}^{(1,j_1)}(i), \quad x \xrightarrow{d} Z_{m,n}^{(2,j_2)}(i), \tag{6.4.79}
\]
where the limiting branching processes are independent. Equation (6.4.79) replaces the convergence in Lemma 6.15 for \( \text{GRG}_n(\mathbf{w}) \).

We conclude that, for \( k = k_n = (1 + \varepsilon) \log \nu n \), and conditionally on \( a = \partial B_m(o_1), b = \partial B_m(o_2) \) such that \( \partial B_m(o_1), \partial B_m(o_2) \neq \emptyset \),
\[
\mathbb{E}[N_k(a, b)] \xrightarrow{p} \infty. \tag{6.4.80}
\]
This completes the analysis of the first moment.
6.4 Path counting and log upper bound for finite-variance weights 253

We next analyse \( \Var(N_k(a, b)) \) and prove that \( \Var(N_k(a, b))/E[N_k(a, b)]^2 \to \infty \). Here our proof will be slightly more sketchy.

Recall (6.4.18). In this case, we compute, to replace (6.4.19),
\[
\mathbb{P}(\pi, \rho \text{ occupied}) = \mathbb{P}(\pi \text{ occupied})\mathbb{P}(\rho \text{ occupied} | \pi \text{ occupied}).
\]
(6.4.81)

Recall the definition of a shape in (6.4.16). Fix \( \sigma \in \text{Shape}_{n,t} \) and \( \rho \in \text{P}_k(a, b) \) with \( \text{Shape}(\pi, \rho) = \sigma \). The factor \( \mathbb{P}(\rho \text{ occupied} | \pi \text{ occupied}) \), summed over the free vertices of \( \rho \) (i.e., the ones that are not also vertices in \( \pi \)) gives rise to \( m \) factors of the form \( T^\pi_{\kappa_n}(i\pi_n, i\pi_{n+1})/n \) for \( i \in [m+1] \) and some vertices \( \pi_n \) and \( \pi_{n+1} \) in the path \( (\pi_i)_{i=0}^k \). We use that, uniformly in \( t \geq 1 \),
\[
\frac{k^2}{n} \max_{i,j \in [k]} T^\pi_{\kappa_n}(i, j) \leq \frac{Ck^2}{n} \|T_{\kappa_n}\|^t,
\]
(6.4.82)

where \( k^2 \) is an upper bound on the number of choices for \( \pi_n \) and \( \pi_{n+1} \). Thus, for every of the \( m \) subpaths of length \( t_i \) we obtain a factor \( C_{k^2}/n \|T_{\kappa_n}\|^t \). We arrive at
\[
\sum_{\pi, \rho \in \text{P}_k(a, b) \atop \text{Shape}(\pi, \rho) = \sigma} \mathbb{P}(\pi, \rho \text{ occupied}) \leq \mathbb{E}[N_k(a, b)] \prod_{i=1}^m \frac{Ck^2}{n} \|T_{\kappa_n}\|^t_i \leq \mathbb{E}[N_k(a, b)]\|T_{\kappa_n}\|^{k-1} \left( \frac{Ck^2}{n} \right)^m.
\]
(6.4.83)

This replaces (6.4.20), and the proof can now be completed in an identical way as the proof of (6.4.7) combined with the proof of (6.4.48) in the proof of Theorem 6.14. We omit further details.

Concentration of the giant in Theorem 3.16

By Theorem 3.11, we know that IRG\(_n(\kappa_n)\) converges in probability in the local weak sense. Therefore, we aim to apply Theorem 2.26, for which we need to verify that (2.5.7) holds.

We rewrite the expectation appearing in (2.5.7) as
\[
\frac{1}{n^2} \mathbb{E}\left[ \# \{ x, y \in [n] : |C(x)|, |C(y)| \geq k, x \leftrightarrow y \} \right] = \mathbb{P}(\|C(o_1)\|, |C(o_2)| \geq k, o_1 \leftrightarrow o_2).
\]
(6.4.84)

We condition on \( B_k(o_1) \) and \( B_k(o_2) \), and note that the events that \( \{|C(o_1)| \geq k\} \) and \( \{|C(o_2)| \geq k\} \) are measurable with respect to \( B_k(o_1) \) and \( B_k(o_2) \) to obtain
\[
\mathbb{P}(\|C(o_1)\|, |C(o_2)| \geq k, o_1 \leftrightarrow o_2) = \mathbb{E}\left[ \mathbb{1}_{\{|C(o_1)|, |C(o_2)| \geq k\}} \mathbb{P}(o_1 \leftrightarrow o_2 \mid B_k(o_1), B_k(o_2)) \right].
\]
(6.4.85)

By Exercise 6.23, together with Theorem 3.11,
\[
\lim_{k \to \infty} \lim_{n \to \infty} \mathbb{P}(\partial B_k(o_1) = \emptyset, |C(o_1)| \geq k) = 0.
\]
(6.4.86)
Therefore,
\[
\mathbb{P}(|\mathcal{C}(o_1)|,|\mathcal{C}(o_2)| \geq k, o_1 \leftrightarrow o_2) = \mathbb{E}\left[\mathbf{1}_{\{\mathcal{C}(o_1),|\mathcal{C}(o_2)| \geq k, \partial B_k(o_1), \partial B_k(o_2) \neq \emptyset\}} \mathbb{P}(o_1 \leftrightarrow o_2 | B_k(o_1), B_k(o_2))\right].
\] (6.4.87)

When \(\partial B_k(o_1), \partial B_k(o_2) \neq \emptyset\), it immediately follows that \(|\mathcal{C}(o_1)|,|\mathcal{C}(o_2)| \geq k\). Thus,
\[
\mathbb{P}(|\mathcal{C}(o_1)|,|\mathcal{C}(o_2)| \geq k, o_1 \leftrightarrow o_2) = \mathbb{E}\left[\mathbf{1}_{\{\partial B_k(o_1), \partial B_k(o_2) \neq \emptyset\}} \mathbb{P}(o_1 \leftrightarrow o_2 | B_k(o_1), B_k(o_2))\right] + o_k(1),
\] (6.4.88)

where \(o_k(1)\) denotes a quantity that tends to zero when first \(n \to \infty\) followed by \(k \to \infty\).

In the proof of Theorem 6.1(ii), we have shown that, on \(\{\partial B_k(o_1), \partial B_k(o_2) \neq \emptyset\}\),
\[
\mathbb{P}(d_{\text{IRG}_n}(\kappa_n)(o_1, o_2) \leq (1 + \varepsilon) \log_n n | B_k(o_1), B_k(o_2)) = 1 + o(1).\] (6.4.89)

In particular, on \(\{\partial B_k(o_1), \partial B_k(o_2) \neq \emptyset\}\),
\[
\mathbb{P}(o_1 \leftrightarrow o_2 | B_k(o_1), B_k(o_2)) = o(1).\] (6.4.90)

Since also \(\mathbb{P}(o_1 \leftrightarrow o_2 | B_k(o_1), B_k(o_2)) \leq 1\), the Dominated Convergence Theorem [Volume 1, Theorem A.1] completes the proof of (2.5.7) for \(\text{IRG}_n(\kappa_n)\), as required.

6.5 Related results on distances for inhomogeneous random graphs

In this section, we discuss some related results for inhomogeneous random graphs. While we give intuition about their proofs, we shall not include them in full detail.

6.5.1 The diameter in inhomogeneous random graphs

In this section, we investigate the diameter of \(\text{IRG}_n(\kappa_n)\), which is defined to be the maximal finite graph distance between any pair of vertices, i.e., the diameter \(\text{diam}(G)\) of the graph \(G\) equals
\[
\text{diam}(G) = \max_{u,v: \text{dist}_G(u,v) < \infty} \text{dist}_G(u,v).
\] (6.5.1)

We shall see that for \(\text{IRG}_n(\kappa)\), the diameter tends to be much larger than the typical graph distances, which is due to long thin lines which are distributed as subcritical \(\text{IRG}_n(\kappa)\) with a subcritical \(\kappa\) by a duality principle for \(\text{IRG}_n(\kappa)\). Before we state the results, we introduce the notion of the dual kernel:
6.5 Related results on distances for inhomogeneous random graphs

**Definition 6.20** (Dual kernel for IRG\(_n(\kappa)\)) Let (\(\kappa_n\)) be a sequence of super-critical kernels with limit \(\kappa\). The dual kernel is the kernel \(\hat{\kappa}\) defined by \(\hat{\kappa}(x,y) = \kappa(x,y), \) with reference measure \(d\hat{\mu}(x) = (1 - \zeta_n(x))\mu(dx)\).

The dual kernel shall describe the graph after the removal of the giant component. Here, the reference measure \(\hat{\mu}\) measures the size of the graph. In this case, a vertex \(x\) is in the giant component with probability \(1 - \zeta_n(x)\), in which case it must be removed. Thus, \(\hat{\mu}\) describes the proportion of vertices of the various types which are outside the giant component. As before, we define the operator \(\mathbf{T}_x\) by the equality

\[
(\mathbf{T}_x f)(x) = \int_S \hat{\kappa}(x,y)f(y)d\hat{\mu}(y) = \int_S \kappa(x,y)f(y)[1 - \zeta_n(x)]\mu(dy),
\]

and we write \(\|\mathbf{T}_x\|\) for

\[
\|\mathbf{T}_x\| = \sup \{\|\mathbf{T}_x f\|_2 : f \geq 0, \|f\|_{\hat{\mu},2} = 1\},
\]

where

\[
\|f\|_{\hat{\mu},2}^2 = \int_S f^2(x)d\hat{\mu}(dx).
\]

**Theorem 6.21** (The diameter of IRG\(_n(\kappa)\) in the finite-types case) Let (\(\kappa_n\)) be a sequence of kernels with limit \(\kappa\), which has finitely many types. If \(0 < \|\mathbf{T}_x\| < 1\), then

\[
\frac{\text{diam}(\text{IRG}_n(\kappa_n))}{\log n} \overset{\tau}{\to} \frac{1}{\log \|\mathbf{T}_x\|^{-1}}
\]

as \(n \to \infty\). If \(\|\mathbf{T}_x\| > 1\) and \(\kappa\) irreducible, then

\[
\frac{\text{diam}(\text{IRG}_n(\kappa_n))}{\log n} \overset{\tau}{\to} \frac{2}{\log \|\mathbf{T}_x\|^{-1}} + \frac{1}{\log \|\mathbf{T}_x\|},
\]

where \(\hat{\kappa}\) is the dual kernel to \(\kappa\).

If we compare Theorem 6.21 to Theorem 6.2, we see that the diameter has the same scaling as the typical graph distance, but that the limit in probability of \(\text{diam}(\text{IRG}_n(\kappa))/\log n\) is strictly larger than the one for \(\text{d}(\text{IRG}_n(\kappa)|_{\text{IRG}_n(\kappa)}))/\log n\) conditioned on \(d(\text{IRG}_n(\kappa)|_{\text{IRG}_n(\kappa)})/\log n\) converges to a finite limit, while \(\text{diam}(\text{IRG}_n(\kappa))/\log n\) converges to a non-zero limit. This can be explained by noticing that the diameter in IRG\(_n(\kappa)\) arises due to very this lines of length of order \(\log n\). Since these this lines involve only very few vertices, they will not contribute to \(d(\text{IRG}_n(\kappa)|_{\text{IRG}_n(\kappa)})\), but they do to \(\text{diam}(\text{IRG}_n(\kappa))\). This is another argument why we prefer to work with typical graph distances than with the diameter. Exercise 6.24 investigates the consequences for ER\(_n(\lambda/n)\).

We will not prove Theorem 6.21 here. For GRG\(_n(w)\), it will follow from Theorem 7.16, which proves a related result for the configuration model.
6.5.2 Distance fluctuations for GRG_n(w) with finite-variance degrees

We continue by studying the fluctuations of the typical graph distance when \( \mathbb{E}[W^2] < \infty \). We shall impose a slightly stronger condition on the distribution function \( F \) of \( W \), namely, that there exists a \( \tau > 3 \) and \( c > 0 \) such that

\[
1 - F(x) \leq cx^{-(\tau-1)}.
\]

Equation (6.5.7) implies that the degrees have finite variance, see Exercise 6.25.

**Theorem 6.22** (Limit law for the typical graph distance in CL_n(w)) Assume that (6.5.7) is satisfied, and let \( \nu = \mathbb{E}[W^2]/\mathbb{E}[W] \geq 1 \). For \( k \geq 1 \), define \( a_k = \lceil \log_\nu k \rceil - \log_\nu k \in (-1, 0) \). Then, for CL_n(w) with \( w \) as in (1.3.15), there exist random variables \( (R_a)_{a \in (-1, 0)} \) with \( \limsup_{K \to \infty} \sup_{a \in (-1, 0)} \mathbb{P}(|R_a| < K) = 1 \) such that, as \( n \to \infty \) and for all \( k \in \mathbb{Z} \),

\[
\mathbb{P}(d_{NR_n(w)}(o_1, o_2) - \lceil \log_\nu n \rceil = k \mid d_{NR_n(w)}(o_1, o_2) < \infty) = \mathbb{P}(R_a = k) + o(1).
\]

While Theorem 6.1 implies that, conditionally on \( o_1 \) and \( o_2 \) being connected, \( d_{NR_n(w)}(o_1, o_2)/\log n \overset{\mathbb{P}}{\to} 1/\log \nu \), Theorem 6.22 implies that the fluctuations of \( d_{NR_n(w)}(o_1, o_2) \) around \( \log_\nu n \) remain uniformly bounded in probability.

The random variables \( (R_a)_{a \in (-1, 0)} \) can be determined in terms of the limit law in a branching process approximation of the neighborhoods of CL_n(w), and depend sensitively on \( a \), which implies that although \( (d_{NR_n(w)}(o_1, o_2) - \lceil \log_\nu n \rceil)_{n=2}^\infty \) is a tight sequence of random variables, it does not weakly converges. See Exercises 6.26–6.28 for further properties of this sequence of random variables.

6.5.3 Distances GRG_n(w) for \( \tau = 3 \)

**Theorem 6.23** (Critical case: interpolation) Consider GRG_n(w), where the i.i.d. weights satisfy

\[
\mathbb{P}(W_1 > k) = k^{-2}\left(\log k\right)^{2\alpha + o(1)}
\]

for some \( \alpha > 0 \). Consider two vertices \( o_1, o_2 \) chosen independently and uniformly at random from the largest connected component \( C_{max} \) of GRG_n(w). Then

\[
\text{dist}_{GRG_n(w)}(o_1, o_2) = (1 + o(1)) \frac{1}{1 + 2\alpha \log \log n} \log n.
\]

6.6 Notes and discussion

**Notes on Section 6.1**

Theorem 6.1 is a simplified version of (Bollobás et al., 2007, Theorem 3.14). A first version of Theorem 6.1 was proved in Chung and Lu (2002a, 2003) for the expected degree random graph, in the case of admissible deterministic weights.
We refer to (Chung and Lu, 2003, p. 94) for the definition of admissible degree sequences.

Theorem 6.3 for the expected degree random graph or Chung-Lu model is first proved by Chung and Lu (2002a, 2003), in the case of deterministic weights \( w_i = c \cdot (i/n)^{-1/(\tau-1)} \), having average degree strictly greater than 1 and maximum degree \( m \) satisfying \( \log m \gg \log n / \log \log n \). These restrictions were lifted in (Durrett, 2007, Theorem 4.5.2). Indeed, the bound on the average distance is not necessary, since, for \( \tau \in (2, 3) \), \( \nu = \infty \) and therefore the IRG is always supercritical. An upper bound as in Theorem 6.3 for the Norros-Reittu model with i.i.d. weights is proved by Norros and Reittu (2006).

Theorem 6.2 has a long history, and many versions of it have been proven in the literature. We refer the reader to Chung and Lu (2002a, 2003) for the Chung-Lu model, and van den Esker et al. (2008) for its extensions to the Norros-Reittu model and the generalized random graph. Theorem 6.3 has also been proved in many versions, both fully as well as in partial forms, see Norros and Reittu (2006), Chung and Lu (2002a, 2003), as well as Dereich et al. (2012).

Notes on Section 6.2
As far as we are aware, the proof of Theorem 6.4 is new in the present context. Similar arguments have been used often though to prove lower bounds on distances in various situations.

The truncated first moment method in the proof of Theorem 6.6 is inspired by Dereich et al. (2012).

Notes on Section 6.3
Theorem 6.9 is novel in its precise form, and also its proof is different from the ones in the literature. See the notes of Section 6.1 for the relevant references.

Notes on Section 6.4
The path-counting techniques in Proposition 6.12 are novel. They are inspired by the path-counting techniques using by Eckhoff et al. (2013) for smallest-weight problems on the complete graph, where many of the counting arguments already appeared. Related proofs for the upper bound on \( \text{dist}_{\text{NR}}(w)(o_1, o_2) \) when \( \nu < \infty \) as in Theorem 6.14 often rely on branching process comparisons up to a generation \( m = m_n \to \infty \).

Notes on Section 6.5
Theorem 6.22 is proved by van den Esker et al. (2008), both in the case of i.i.d. degrees as well as for deterministic weights under a mild further condition on the distribution function.

Theorem 6.21 is a special case of (Bollobás et al., 2007, Theorem 3.16). In the special case of \( \text{ER}_n(\lambda/n) \), it is extends a previous result of Chung and Lu (2001) that proves logarithmic asymptotics of \( \text{diam} \left( \text{ER}_n(\lambda/n) \right) \), and it negatively answers a question of Chung and Lu (2001). Related results for the configuration model, which also imply results for the generalized random graph, can be found.
258 Small-world phenomena in inhomogeneous random graphs

in Fernholz and Ramachandran (2007). See also Theorem 7.16 below. See also Riordan and Wormald (2010) for additional results and branching-process proofs. There, also the case where \( \lambda > 1 \) with \( \lambda - 1 \gg n^{-1/3} \) are discussed. Similar results have been derived by Ding et al. (2011, 2010).

6.7 Exercises for Chapter 6

Exercise 6.1 (Typical distances in \( \text{ER}_n(\lambda/n) \)) Fix \( \lambda > 1 \). Use either Theorem 6.1 or Theorem 6.2 to prove that
\[
d_{\text{ER}_n(\lambda/n)}(o_1, o_2) / \log n \xrightarrow{p} 1 / \log \lambda
\]
conditionally on \( d_{\text{ER}_n(\lambda/n)}(o_1, o_2) < \infty \).

Exercise 6.2 (Typical distances when \( \nu = \infty \)) Prove that
\[
d_{\text{NR}_n(w)}(o_1, o_2) / \log n \xrightarrow{p} 0
\]
when \( \nu = \infty \), by using an appropriate truncation argument and monotonicity.

Exercise 6.3 (Power-law tails in key example of deterministic weights) Let \( w \) be defined as in (1.3.15), and assume that \( F \) satisfies
\[
1 - F(x) = x^{-(\tau - 1)}L(x),
\]
where the exponent satisfies \( \tau \in (2, 3) \), and where \( x \mapsto L(x) \) is slowly varying. Prove that (6.1.4) holds.

Exercise 6.4 (Power-law tails for i.i.d. weights) For i.i.d. weights \( w = (w_i)_{i \in [n]} \) with distribution \( F \) satisfying that (6.7.1) with \( \tau \in (2, 3) \), and where \( x \mapsto L(x) \) is slowly varying. Prove that (6.1.4) holds with probability converging to 1.

Exercise 6.5 (Bound on truncated forward degree \( \nu_n(b) \)) Prove (6.2.42) by combining (1.4.2) in Lemma 1.13 with \( \ell_n = \Theta(n) \) by Conditions 1.1(a)–(b).

Exercise 6.6 (Conditions (6.2.1) and Condition 6.4) Show that when there is precisely one vertex with weight \( w_1 = \sqrt{n} \), whereas \( w_i = \lambda > 1 \), then (6.2.1) holds, but Condition 6.4 does not. Argue that the upper bound derived in Theorem 6.4 is not sharp, since the vertex 1 can occur at most once in a self-avoiding path.

Exercise 6.7 (Lower bound on fluctuations) Adapt the proof of Theorem 6.4 to show that for every \( \varepsilon \), we can find a constant \( K = K(\varepsilon) > 0 \) such that
\[
P(d_{\text{NR}_n(w)}(o_1, o_2) \leq \log_{\nu_n}(n) - K) \leq \varepsilon.
\]
Conclude that if \( \log \nu_n = \log \nu + o(1 / \log n) \), then the same statement holds with \( \log \nu \) replacing \( \log \nu_n \).

Exercise 6.8 (Proof Corollary 6.5) Adapt the proof to Theorem 6.4 to prove Corollary 6.5.
Exercise 6.9 (Lower bound on typical distances for $\tau = 3$) Let $w_i = c\sqrt{(n/i)}$, so that $\tau = 3$. Prove that $\nu_n / \log n \to c$. Use Corollary 6.5 to obtain that for any $\varepsilon > 0$,
\[
P\left( d_{\text{NR}_n(w)}(o_1, o_2) \leq (1 - \varepsilon) \frac{\log n}{\log \log n} \right) = o(1). \tag{6.7.3}
\]

Exercise 6.10 (Lower bound on typical distances for $\tau \in (2, 3)$) Let $w_i = c / i^{1/(\tau - 1)}$ with $\tau \in (2, 3)$. Prove that there exists a constant $c' > 0$ such that $\nu_n \geq c'n^{(3-\tau)/(\tau-1)}$. Show that Corollary 6.5 implies that $d_{\text{NR}_n(w)}(o_1, o_2) \geq (\tau - 1)/(\tau - 3)$ in this case. How useful is this bound?

Exercise 6.11 (Convergence in probability of typical distance in IRG$_n(\kappa_n)$) Suppose that the graphical sequence of kernels $(\kappa_n)$ satisfies $\sup_{x, y, n} \kappa_n(x, y) < \infty$, where the limit $\kappa$ is irreducible and $\nu = \|T_n\| > 1$. Prove that Theorem 3.16 together with Theorem 6.1(i–ii) imply that, conditionally on $d_{\text{IRG}_n(\kappa_n)}(o_1, o_2) < \infty$,
\[
d_{\text{IRG}_n(\kappa_n)}(o_1, o_2) / \log n \xrightarrow{p} 1 / \log \nu. \tag{6.7.4}
\]

Exercise 6.12 (Convergence in probability of typical distance in IRG$_n(\kappa_n)$) Suppose that the graphical sequence of kernels $(\kappa_n)$ converges to $\kappa$, where $\kappa$ is irreducible and $\|T_n\| = \infty$. Prove that Theorem 3.16 together with Theorem 6.1(iii) imply that, conditionally on $d_{\text{IRG}_n(\kappa_n)}(o_1, o_2) < \infty$,
\[
d_{\text{IRG}_n(\kappa_n)}(o_1, o_2) / \log n \xrightarrow{p} 0. \tag{6.7.5}
\]

Exercise 6.13 (Distance between fixed vertices) Show that (6.2.31) and Lemma 6.7 imply that for all $a, b \in [n]$ with $a \neq b$,
\[
P(\text{dist}_{\text{NR}_n(w)}(a, b) \leq k_n) \leq \frac{w_a w_b}{\ell_n} \sum_{k=1}^{k_n} \prod_{l=1}^{k-1} \nu_n(b_l \wedge b_{k-l}) \tag{6.7.6}
+ \left( w_a + w_b \right) \sum_{k=1}^{k^*} [1 - F_n^*(b_k)] \prod_{l=1}^{k} \nu_n(b_l).
\]

Exercise 6.14 (Lower bound on fluctuations*) Adapt the proof of Theorem 6.6 to show that for every $\varepsilon$, we can find a constant $K = K(\varepsilon) > 0$ such that
\[
P\left( d_{\text{NR}_n(w)}(o_1, o_2) \leq \frac{2 \log \log n}{|\log (\tau - 2)|} - K \right) \leq \varepsilon. \tag{6.7.7}
\]

*Hint: choose $b_k = Lh_{k-1}^{1/(\tau-2)}$, where the constant $L$ is chosen sufficiently large.

Exercise 6.15 (Upper bound on the expected number of paths) Prove (6.4.4) for an inhomogeneous random graph with vertex set $\mathcal{I}$ and with edge probabilities $p_{ij} = u_i u_j$ for every $i, j \in \mathcal{I}$.

Exercise 6.16 (Variance of two paths) Prove that $\text{Var}(N_k(a, b)) \leq E[N_k(a, b)]$ for $k = 2$. 

6.7 Exercises for Chapter 6
Exercise 6.17 (Variance of three paths) Compute $\text{Var}(N_3(a,b))$ explicitly, and compare it to the bound in (6.4.7).

Exercise 6.18 (Variance on paths for ER$_n(\lambda/n)$) Let $A, B \subseteq [n]$, and let $N_k(A,B)$ denote the number of paths of length $k$ connecting $A$ to $B$ (where a path connecting $A$ and $B$ avoids $A$ and $B$ except for the starting and end point). Show that for $k \leq K \log n$,

$$E[N_k(A,B)] = \lambda^k|A||B|(1 - |A| + |B|/n)^k(1 + o(1)).$$

Use Proposition 6.12 to bound the variance of $N_k(A,B)$, and prove that $N_k(A,B)/E[N_k(A,B)] \xrightarrow{P} 1$ when $|A|, |B| \to \infty$ with $|A| + |B| = o(n/k)$.

Exercise 6.19 ($\nu_n$ bound for $\tau = 3$) Prove that (6.4.54) and (6.4.55) imply that $\nu \geq c \log \alpha_n$ by using

$$\frac{1}{n} \sum_{i \in I} w_i^2 = E[W_n^2 \mathbb{1}_{\{W_n \in [K,\sqrt{\alpha_n}]\}}] = 2 \int_K^{\sqrt{\alpha_n}} x[F_n(\sqrt{\alpha_n}) - F_n(x)]dx.$$ 

Exercise 6.20 (Expected number of paths within Core$_n$ diverges) Prove that $E[N_k(a,b)] \to \infty$ for $a = V_1', b = V_2'$ and $k = \log n/(1 - \eta) \log \nu_n$.

Exercise 6.21 (Concentration of number of paths within Core$_n$) Prove that $\text{Var}(N_k(a,b))/E[N_k(a,b)]^2 \to 0$ for $a = V_1', b = V_2'$ and $k = \log n/(1 - \eta) \log \nu_n$.

Exercise 6.22 (Completion proof Proposition 6.18) Complete the proof of Proposition 6.18 by adapting the arguments in (6.4.50)–(6.4.53).

Exercise 6.23 (Size versus boundary of large clusters) Let the graph $G = (V,E)$ converge locally weakly in probability. Prove that

$$\lim_{k \to \infty} \lim_{n \to \infty} P(\partial B_k(a_0) = \emptyset, |\mathcal{C}(a_0)| \geq k) = 0.$$ 

Exercise 6.24 (Diameter of ER$_n(\lambda/n)$) Recall Theorem 6.21. For ER$_n(\lambda/n)$, show that $\|T_\kappa\| = \lambda$ and $\|T_\kappa^*\| = \mu_\lambda$, where $\mu_\lambda$ is the dual parameter in [Volume 1, (3.6.6)], so that Theorem 6.21 becomes

$$\frac{\text{diam}(\text{ER}_n(\lambda/n))}{\log n} \xrightarrow{\text{d}} \frac{2}{\log \mu_\lambda^{-1}} + \frac{1}{\log \lambda}.$$ 

Exercise 6.25 (Finite variance degrees when (6.5.7) holds) Prove that (6.5.7) implies that $E[W^2] < \infty$. Use this to prove that the degrees have uniformly bounded variance when (6.5.7) holds.
Exercise 6.26 (Tightness of centered typical graph distances in $\text{CL}_n(w)$)  Prove that, under the conditions of Theorem 6.22, and conditionally on $d_{\text{NR}_n(w)}(o_1, o_2) < \infty$, the sequence $\left( d_{\text{NR}_n(w)}(o_1, o_2) - \lfloor \log_n n \rfloor \right)_{n=2}^{\infty}$ is tight.

Exercise 6.27 (Non-convergence of centered typical graph distances in $\text{CL}_n(w)$)  Prove that, under the conditions of Theorem 6.22, and conditionally on $d_{\text{NR}_n(w)}(o_1, o_2) < \infty$, the sequence $d_{\text{NR}_n(w)}(o_1, o_2) - \lfloor \log_n n \rfloor$ does not weakly converge when the distribution of $R_a$ depends continuously on $a$ and when there are $a, b \in (-1, 0]$ such that the distribution of $R_a$ is not equal to the one of $R_b$.

Exercise 6.28 (Extension Theorem 6.22 to $\text{GRG}_n(w)$ and $\text{NR}_n(w)$)  Use [Volume 1, Theorem 6.18] to prove that Theorem 6.22 holds verbatim for $\text{GRG}_n(w)$ and $\text{NR}_n(w)$ when (6.5.7) holds.
Chapter 7
SMALL-WORLD PHENOMENA IN THE CONFIGURATION MODEL

Abstract
In this chapter, we investigate the connectivity structure of the configuration model by investigating its typical distances and its diameter. We also show that the configuration model is whp connected when the minimal degree is at least three.

In this chapter, we investigate graph distances in the configuration model (CM). We start with a motivating example.

Motivating example
Recall Figure 1.4(a) in which graph distances in the Autonomous Systems (AS) graph in Internet are shown. Such distances are known under the name of the AS-count. A relevant question is whether such a histogram can be predicted by the graph distances in a random graph model having similar degree structure and size as the AS graph. Simulations indicating the properties of the typical graph distance for \( \tau \in (2, 3) \) can be seen in Figure 7.1. In it, the distances of the AS-graph are compared to the ones in \( \text{CM}_n(d) \) where \( n \) is equal to the number of AS, which is \( n = 10,940 \), and the best approximation to the exponent of the power-law for the degree sequence of the AS-graph, which is \( \tau = 2.25 \). We see that the simulations of graph distances in this \( \text{CM}_n(d) \) and the AS-counts are quite close.

![Figure 7.1](image-url)  
Figure 7.1 Number of AS traversed in hopcount data (blue) compared to the model (purple) with \( \tau = 2.25, n = 10,940 \).
This figure also again raises the question how graph distances depend on random graph properties, such as \( n \), but also the structure of the random graph in question. The configuration model is a highly flexible model, in the sense that it allows to choose the degree distribution in a highly general way. Thus, we can use the CM to single out the relation between the graph distances and the degree structure, in a similar way as that it allows to investigate the size of the giant component and the connectivity as a function of the degree distribution as in Chapter 4. Finally, we can verify whether the graph distances are closely related to those in inhomogeneous random graphs, as discussed in Chapter 6, so as to spot another sign of the wanted universality of structural properties of random graphs with similar degree distributions.

**Organization of this chapter**

This chapter is organized as follows. In Section 7.1, we study the typical graph distance in the configuration model. In Section 7.2, we prove these distance results, using path counting techniques and comparisons to branching processes. In Section 7.3, we study infinite-mean branching processes, as these arise in the configuration model with infinite-variance degrees. In Section 7.4, we study the diameter of the configuration model. In Section 7.5, we state further results in the configuration model. We close this chapter with notes and discussion in Section 7.6, and with exercises in Section 7.7.

### 7.1 The small-world phenomenon in \( \text{CM}_n(d) \)

In this section, we describe the main results on distances in the configuration model, both in the case of finite-variance degrees as well as in the case of infinite-variance degrees. These results will be proved in the following sections.

**Distances in configuration models with finite-variance degrees**

We start by analyzing the typical graph distance in the case where the configuration model \( \text{CM}_n(d) \) when Conditions 1.5(a)-(c) hold:

**Theorem 7.1** (Typical distances in \( \text{CM}_n(d) \) for finite-variance degrees) In the configuration model \( \text{CM}_n(d) \), where the degrees \( d = (d_i)_{i \in [n]} \) satisfy Conditions 1.5(a)-(c) and where \( \nu > 1 \), conditionally on \( \text{dist}_{\text{CM}_n(d)}(o_1, o_2) < \infty \),

\[
\text{dist}_{\text{CM}_n(d)}(o_1, o_2) / \log n \xrightarrow{p} 1 / \log \nu. \tag{7.1.1}
\]

Theorem 7.1 shows that the typical distances in \( \text{CM}_n(d) \) are of order \( \log \nu n \), and is thus similar in spirit as Theorem 6.2. We shall see that also its proof is quite similar.

**Finite mean, infinite variance degrees**

We next study the typical distance of the configuration model with degrees having finite mean and infinite variance. We start by formulating the precise condition on
the degrees that we shall work with. This condition is identical to the condition
on \( F_n \) for \( NR_n(w) \) formulated in (6.1.4). Recall that \( F_n(x) \) denotes the proportion of
vertices having degree at most \( x \). Then, we assume that there exists a \( \tau \in (2, 3) \)
and for all \( \delta > 0 \), there exist \( c_1 = c_1(\delta) \) and \( c_2 = c_2(\delta) \) such that, uniformly in \( n \),
\[
c_1 x^{-(\tau - 1 + \delta)} \leq [1 - F_n](x) \leq c_2 x^{-(\tau - 1 - \delta)},
\]
where the upper bound holds for every \( x \geq 1 \), while the lower bound is only re-
quired to hold for \( 1 \leq x \leq n^\alpha \) for some \( \alpha > 1/2 \). The typical distances of \( CM_n(d) \)
under the infinite-variance condition in (7.1.2) are identified in the following the-
orem:

**Theorem 7.2** (Typical distances in \( CM_n(d) \) for \( \tau \in (2, 3) \)) Let the degrees \( d = (d_i)_{i \in [n]} \) in the configuration model \( CM_n(d) \) satisfy Conditions 1.5(a)-(b) and (7.1.2). Then, conditionally on \( dist_{CM_n(d)}(o_1, o_2) < \infty \),
\[
dist_{CM_n(d)}(o_1, o_2) \quad \frac{2}{\log \log n} \xrightarrow{p} \frac{2}{|\log (\tau - 2)|}.
\]

Theorem 7.2 is similar in spirit to Theorem 6.3 for \( NR_n(w) \).

### 7.2 Proofs of small-world results \( CM_n(d) \)

In this section, we give the proofs of Theorems 7.1 and 7.2 describing the
small-world properties in \( CM_n(d) \). These proofs are adaptations of the proofs of
Theorems 6.2 and 6.3, and we focus on the differences in the proofs. This section is
organized as follows. In Section 7.2.1 we give a branching process approximation
for the neighborhoods of a pair of uniform vertices in \( CM_n(d) \). In Section 7.2.2
we perform similar path-counting techniques as in Section 6.4.1, adapted to the
configuration model where edges are formed by pairing half-edges.

#### 7.2.1 Branching process approximation

In this section, we summarize some links between the breadth-first exploration
in \( CM_n(d) \) and branching processes.

Let, by convention, \( Z_0(i) = 1 \) and, for \( m \geq 1 \), let \( Z_m(i) \) denote the number of
unpaired half-edges incident to vertices at graph distance \( m - 1 \) from vertex
\( i \), so that \( Z_1(i) = d_i = S_0 \). Thus, \( Z_2(i) \) is obtained after pairing all the \( Z_1(i) \)
half-edges at distance 1 from the root. Then, \( Z_m(i) \) equals \( S_{T_m(i)} \), where \( T_m(i) \)
is the time where all the \( Z_{m-1}(i) \) half-edges at distance \( m - 2 \) from vertex \( i \)
have been paired. The above describes the breadth-first exploration from a single
vertex \( i \). Let, as usual, \( V_1 \) and \( V_2 \) be two vertices chosen uniformly at random
from \([n]\), and denote \( Z_m^{(i)} = Z_m(V_i) \), so that \( Z_m^{(i)} \) is the number of unpaired half-
edges at distance \( m - 1 \) from vertex \( V_i \). The following proposition shows that,
for all \( m \geq 1 \), the processes \( (Z_1^{(i)}, Z_2^{(i)})_{k=0}^m \) are close to two independent two-stage
branching processes:
Corollary 7.3 (Coupling of neighborhoods of two vertices) Let the degrees \( \mathbf{d} = (d_i)_{i \in [n]} \) satisfy Conditions 1.5(a)-(b). Let \((Z^{(1)}_i, Z^{(2)}_i)_{i \geq 0}\) be two independent branching processes with offspring distribution \(D\) in the first generation, and offspring distribution \(D^* - 1\) in all further generations. Then, for every \(m \geq 1\),

\[
P(\text{dist}_{CM_n(d)}(o_1, o_2) \leq 2m) = o(1), \tag{7.2.1}
\]

and

\[
(Z^{(n,1)}_1, Z^{(n,2)}_1)_{i=0}^m \overset{d}{\rightarrow} (Z^{(1)}_1, Z^{(2)}_1)_{i=0}^m. \tag{7.2.2}
\]

Corollary 7.3 follows from the local-weak convergence in probability in Theorem 4.1, combined with Corollaries 2.17–2.18.

7.2.2 Path-counting techniques

In this section, we present path-counting techniques similar to those in Section 6.4.1. Since \(CM_n(\mathbf{d})\) is a multigraph, and not a simple graph as \(NR_n(\mathbf{u})\), we need to be precise what a path in \(CM_n(\mathbf{d})\) is. We start by introducing some notation.

A path \(\pi\) of length \(k\) in \(CM_n(\mathbf{d})\) means a sequence

\[
\pi = \{(\pi_0, s_0), (\pi_1, s_1, t_1), \ldots, (\pi_{k-1}, s_{k-1}, t_{k-1}), (\pi_k, t_k)\}, \tag{7.2.3}
\]

where \(\pi_i \in [n]\) denotes the \(i\)th vertex along the path, and \(s_i \in [d_{\pi_i}]\) denotes the label of the half-edge incident to \(\pi_i\), and \(t_{i+1} \in [d_{\pi_{i+1}}]\) denotes the label of the half-edge incident to \(\pi_{i+1}\). In particular, multiple edges between \(\pi_i\) and \(\pi_{i+1}\) give rise to distinct paths through the same vertices. For a path \(\pi\), we write \(\pi \subseteq CM_n(\mathbf{d})\) when the path \(\pi\) in (7.2.3) is present in \(CM_n(\mathbf{d})\), so that the half-edge corresponding to \(s_i\) is paired with the half-edge corresponding to \(t_{i+1}\) for \(i = 0, \ldots, k - 1\). Without loss of generality, we assume throughout that the path \(\pi\) is simple, i.e., \(\pi_0, \ldots, \pi_k\) are distinct vertices.

In this section, we perform first and second moment computations on the number of paths present in \(CM_n(\mathbf{d})\). We start by proving upper bounds on the expected number of paths.

Upper bounds on the expected number of paths in \(CM_n(\mathbf{d})\)

For \(a, b \in [n]\), \(I \subseteq [n]\) and \(k \geq 1\), we let \(\mathcal{P}_k(a, b) = \mathcal{P}_k(a, b; I)\) denote the set of \(k\)-paths that only use vertices in \(I\), and we let

\[
N_k(a, b) = N_k(a, b; I) = \#\{\pi \in \mathcal{P}_k(a, b; I) : \pi \subseteq CM_n(\mathbf{d})\} \tag{7.2.4}
\]

denote the number of paths of length \(k\) between the vertices \(a\) and \(b\). Then, we prove the following upper bound on the expected number of paths connecting \(a\) and \(b\): 

Proposition 7.4 (Expected numbers of paths) For any \(k \geq 1\), \(a, b \in [n]\) and \((d_i)_{i \in [n]}\),

\[
E[N_k(a, b)] \leq \frac{d_a d_b}{(\ell_n - 2k + 1)(\ell_n - 2k)} \nu^{k-1}_2, \tag{7.2.5}
\]
where
\[ \nu_x = \sum_{i \in I \setminus \{a,b\}} \frac{d_i(d_i - 1)}{\ell_n}. \] (7.2.6)

**Proof**

The probability that the path \( \pi \) in (7.2.3) is present in \( \mathcal{C} M_n(d) \) is equal to
\[ \prod_{i=1}^{k} \frac{1}{\ell_n - 2i + 1}, \] (7.2.7)
and the number of paths with fixed vertices \( \pi_0, \ldots, \pi_k \) is equal to
\[ d_{\pi_0} \left( \prod_{i=1}^{k-1} d_{\pi_i}(d_{\pi_i} - 1) \right) d_{\pi_k}. \] (7.2.8)

Substituting \( \pi_0 = a, \pi_k = b \), we arrive at
\[ E[N_k(a, b)] = \frac{d_a d_b}{\ell_n - 2k + 1} \sum_{\pi_1, \ldots, \pi_{k-1} \in R} \prod_{i=1}^{k-1} \frac{d_{\pi_i}(d_{\pi_i} - 1)}{\ell_n - 2i + 1}, \] (7.2.9)
where the sum is over distinct elements of \( I \setminus \{a, b\} \). Let \( R \) denote the subset of vertices of \( I \setminus \{a, b\} \) for which \( d_i \geq 2 \). Then,
\[ E[N_k(a, b)] = \frac{d_a d_b}{\ell_n - 2k + 1} \sum_{\pi_1, \ldots, \pi_{k-1} \in R} \prod_{i=1}^{k-1} \frac{d_{\pi_i}(d_{\pi_i} - 1)}{\ell_n - 2i + 1}. \] (7.2.10)

By an inequality of Maclaurin (Hardy et al., 1988, Theorem 52), for \( r = |R| \), \( 2 \leq k \leq r + 1 \) and any \( (a_i)_{i \in R} \) with \( a_i \geq 0 \),
\[ \frac{(r - k + 1)!}{r!} \sum_{\pi_1, \ldots, \pi_{k-1} \in R} \prod_{i=1}^{k-1} a_i \leq \left( \frac{1}{r} \sum_{i \in R} a_i \right)^{k-1}. \] (7.2.11)
Let \( a_i = d_i(d_i - 1) \), so that
\[ \sum_{i \in R} a_i = \ell_n \nu_x. \] (7.2.12)

We arrive at
\[ E[N_k(a, b)] = \frac{d_a d_b}{\ell_n - 2k + 1} \left( \ell_n \nu_x / r \right)^{k-1} \prod_{i=1}^{k-1} \left( \ell_n - 2i + 1 \right)^{-\frac{r}{i}} \] (7.2.13)
\[ \leq \frac{d_a d_b}{\ell_n - 2k + 1} \left( \ell_n - 2k \right)^{-\frac{r}{k}} \prod_{i=0}^{k-2} \left( 1 - \frac{i}{r} \right) \left( \ell_n - 2k \right)^{-\frac{r}{k}} \left( \ell_n - 2k - 2 \right)^{-\frac{r}{k}} \].

Further, \( \ell_n = \sum_{i \in [n]} d_i \geq 2r \), so that \( 1 - \frac{i}{r} \leq 1 - \frac{2i}{\ell_n} \). Substitution yields the required bound. \( \square \)
Logarithmic lower bound typical distances $\text{CM}_n(d)$

With Proposition 7.4 at hand, we can immediately prove the lower bound on the typical graph distance in the case where the degrees have finite second moment (as in Theorem 6.4):

**Theorem 7.5** (Logarithmic lower bound typical distances $\text{CM}_n(d)$)  

Assume that

$$\limsup_{n \to \infty} \nu_n > 1,$$

(7.2.14)

where

$$\nu_n = \frac{\mathbb{E}[D_n(D_n - 1)]}{\mathbb{E}[D_n]}.$$  (7.2.15)

Then, for any $\varepsilon > 0$,

$$\mathbb{P}(\text{dist}_{\text{CM}_n(d)}(o_1, o_2) \leq (1 - \varepsilon) \log \nu_n n) = o(1).$$  (7.2.16)

We leave the proof of Theorem 7.5, which is almost identical to that of Theorem 6.4 with (7.2.5) in hand, as Exercise 7.2.

**Truncated first moment method and log log lower bound for $\tau \in (2, 3)$**

We next extend the above upper bounds on the expected number of paths to deal with the case where $\tau \in (2, 3)$, where similarly to the setting in Section 6.2.2 where $\text{NR}_n(w)$ was investigated, we need to truncate the degrees occurring in the arising paths. Our main result is as follows:

**Theorem 7.6** (Loglog lower bound on typical distances in $\text{CM}_n(d)$)  

Suppose that the weights $d = (d_i)_{i \in [n]}$ satisfy Condition 1.5(a) and that there exists a $\tau \in (2, 3)$ and $c_2$ such that, for all $x \geq 1$,

$$[1 - F_n](x) \leq c_2 x^{-(\tau - 1)},$$

(7.2.17)

Then, for every $\varepsilon > 0$,

$$\mathbb{P}(\text{dist}_{\text{CM}_n(d)}(o_1, o_2) \leq (1 - \varepsilon) \frac{2 \log \log n}{\log(\tau - 2)}) = o(1).$$

(7.2.18)

The proof of Theorem 7.6 is identical to that of Theorem 6.6, and we discuss the changes only. For a fixed set of distinct vertices $(\pi_0, \ldots, \pi_k)$, (7.2.7)-(7.2.8) yield that the probability that there exists edges between $\pi_{i-1}$ and $\pi_i$ for all $i = 1, \ldots, k$ in $\text{CM}_n(d)$ is bounded above by

$$\frac{d_{\pi_0} d_{\pi_k}}{\ell_n - 2k + 1} \prod_{i=1}^{k-1} \frac{d_{\pi_i}(d_{\pi_i} - 1)}{\ell_n - 2i + 1}.$$  (7.2.19)

Equation (7.2.19) replaces the similar identity (6.2.7) for $\text{NR}_n(w)$. We see that $w_{\pi_0}$ and $w_{\pi_k}$ in (6.2.7) are replaced with $d_{\pi_0}$ and $d_{\pi_k}$ in (7.2.19), and, for $i = 1, \ldots, k - 1$, the factors $w_{\pi_i}^2$ in (6.2.7) are replaced with $d_{\pi_i}(d_{\pi_i} - 1)$ in (7.2.19), while the factors $\ell_n$ in (6.2.7) is replaced with $(\ell_n - 2i + 1)$ in (7.2.19).
7.2 Proofs of small-world results $CM_n(d)$

Define, as in (6.2.32),
\[
\nu_n(b) = \frac{1}{\ell_n} \sum_{i \in [n]} d_i (d_i - 1) \mathbb{1}_{\{d_i \leq b\}}.  \tag{7.2.20}
\]

Then, we can adapt the arguments in Section 6.2.2 to obtain that (see in particular Exercise 6.13),
\[
P(\text{dist}_{CM_n(d)}(a,b) \leq k_n) \leq \frac{d_a d_b}{\ell_n} \sum_{k=1}^{k_n} \frac{\ell^k_n (\ell_n - 2k - 1)!!}{(\ell_n - 1)!!} \prod_{l=1}^{k-1} \nu_n(b_l \wedge b_{k-l})
\]
\[
+ (d_a + d_b) \sum_{k=1}^{k'} \frac{\ell^k_n (\ell_n - 2k - 1)!!}{(\ell_n - 1)!!} [1 - F_n^a (b_k)] \prod_{l=1}^{k} \nu_n(b_l),
\]
i.e., the bound in (6.7.6) is changed by factors $\frac{\ell^k_n (\ell_n - 2k - 1)!!}{(\ell_n - 1)!!}$ in the sum. For $k = O(\log \log n)$ and when Conditions 1.5(a)-(b) hold,
\[
\frac{\ell^k_n (\ell_n - 2k - 1)!!}{(\ell_n - 1)!!} = \prod_{i=1}^{k} \frac{\ell_n - 2i + t}{\ell_n - 2i + 1} = 1 + O(k^2/\ell_n) = 1 + o(1),  \tag{7.2.22}
\]
so this change has only minor effect. Since (6.2.42) in Lemma 6.7 applies under the conditions of Theorem 7.6, we can follow the proof of Theorem 6.6 verbatim. This completes the proof of Theorem 7.6. \hfill \square

Second moment method for the number of paths in $CM_n(d)$

We next extend the above first moment bounds on the number of paths in $CM_n(d)$ to second moment methods. Define
\[
\bar{n}_k(a,b) = \frac{d_a d_b}{\ell_n} \sum_{i \in \mathcal{I} \setminus \{a,b\}} \frac{d_i (d_i - 1)}{\ell_n} \mathbb{1}_{\{d_i \leq b\}}^{k-1}, \tag{7.2.23}
\]
\[
\underline{n}_k(a,b) = \frac{d_a d_b}{\ell_n} \left( \sum_{i \in \mathcal{I}_{a,b,k}} \frac{d_i (d_i - 1)}{\ell_n} \right)^{k-1}, \tag{7.2.24}
\]
where $\mathcal{I}_{a,b,k}$ is the subset of $\mathcal{I}$ in which $a$ and $b$, as well as the $k - 1$ indices with highest degrees, have been removed. Let
\[
\nu_z = \frac{1}{\ell_n} \sum_{i \in \mathcal{I}} d_i (d_i - 1), \quad \gamma_z = \frac{1}{\ell_n^{3/2}} \sum_{i \in \mathcal{I}} d_i (d_i - 1) (d_i - 2). \tag{7.2.25}
\]
The following proposition replaces the similar Proposition 6.12, which was crucial in deriving lower bounds on typical distances:

**Proposition 7.7 (Variance of number of paths)** For any $k \geq 1$, $a, b \in \mathcal{I}$ and $(u_i), i \in \mathcal{I}$,
\[
\mathbb{E}[N_k(a,b)] \geq \underline{n}_k(a,b), \tag{7.2.26}
\]
while, assuming that \( \nu_z > 1 \),

\[
\text{Var}(N_k(a,b)) \leq n_k(a,b) + \bar{n}_k(a,b)^2 \left( \frac{\gamma_z \nu_z^2}{\nu_z - 1} \left( \frac{1}{d_a} + \frac{1}{d_b} \right) + \frac{\gamma_z^2 \nu_z}{d_a d_b (\nu_z - 1)^2} \right) + e'_k,
\]

(7.2.27)

where

\[
e'_k = \left( \prod_{i=1}^{k} \frac{\ell_n - 2i + 1}{\ell_n - 2i - 2k + 1} - 1 \right) + k \frac{\ell_n^2 (\ell_n - 4k - 1)!!}{(\ell_n - 1)!!} \left( 1 + \gamma_z d_a \nu_z \right) \left( 1 + \gamma_z d_b \nu_z \right) \frac{\nu_z}{\nu_x - 1} \left( e^{2 k \gamma_z^2 / \nu_x^2} - 1 \right).
\]

(7.2.28)

**Proof** The proof of (7.2.26) follows immediately from (7.2.9), together with the fact that \( 1/\left( \ell_n - 2i + 1 \right) \geq 1/\ell_n \).

For the proof of (7.2.27), we follow the proof of (6.4.7), and discuss the differences only. We recall that

\[
N_k(a,b) = \sum_{\pi \in P_k(a,b)} \mathbb{1}_{\{\pi \subseteq \text{CM}_n(d)\}}
\]

(7.2.29)

is the number of paths \( \pi \) of length \( k \) between the vertices \( a \) and \( b \), where a path is defined in (7.2.3). Since \( N_k(a,b) \) is a sum of indicators, its variance can be written as

\[
\text{Var}(N_k(a,b)) = \sum_{\pi,\rho \in P_k(a,b)} \left[ \mathbb{P}(\pi,\rho \subseteq \text{CM}_n(d)) - \mathbb{P}(\pi \subseteq \text{CM}_n(d)) \mathbb{P}(\rho \subseteq \text{CM}_n(d)) \right].
\]

(7.2.30)

Equation (7.2.30) replaces (6.4.12) for \( \text{NR}_n(w) \).

We say that two paths \( \pi \) and \( \rho \) are *disjoint* when they use distinct sets of half-edges. Thus, it is possible that the vertex sets \( \{\pi_1, \ldots, \pi_{k-1}\} \) and \( \{\rho_1, \ldots, \rho_{k-1}\} \) have a non-empty intersection, but then the half-edges leading in and out of the joint vertices for \( \pi \) and \( \rho \) must be distinct. For \( \text{NR}_n(w) \), pairs of paths using different edges are *independent*, so that these pairs do not contribute to \( \text{Var}(N_k(a,b)) \).

For \( \text{CM}_n(d) \), instead, for disjoint pairs \( \pi \) and \( \rho \),

\[
\mathbb{P}(\pi,\rho \subseteq \text{CM}_n(d)) = \prod_{i=1}^{k} \frac{\ell_n - 2i + 1}{\ell_n - 2i - 2k + 1} \mathbb{P}(\pi \subseteq \text{CM}_n(d)) \mathbb{P}(\rho \subseteq \text{CM}_n(d)),
\]

(7.2.31)

which explains the first contribution to \( e'_k \). For the other contributions, we follow the proof of (6.4.12) for \( \text{NR}_n(w) \), and omit further details.

With Proposition 7.7 at hand, we can adapt the proof of Theorem 6.14 to \( \text{CM}_n(d) \) to prove the following theorem:

**Theorem 7.8** (Logarithmic upper bound graph distances \( \text{CM}_n(d) \)) Assume
that Conditions 1.5(a)-(c) hold, where \( \nu = \mathbb{E}[D(D - 1)]/\mathbb{E}[D] \in (1, \infty) \). Then, for any \( \varepsilon > 0 \),
\[
\mathbb{P}(\text{dist}_{\text{CM}}(d)(o_1, o_2) \leq (1 + \varepsilon) \log \nu n | \text{dist}_{\text{CM}}(d)(o_1, o_2) < \infty) = 1 + o(1). \quad (7.2.32)
\]

We leave the proof of Theorem 7.8 as Exercise 7.3.

### 7.2.3 A \( \log \log \) upper bound on the diameter core \( \tau \in (2, 3) \)

In order to prove the upper bound on the typical distance for \( \text{CM}_n(d) \) in Theorem 7.2, we use a different approach compared to the one in the proof of Theorem 6.9. Our proof for the upper bound on the typical distance for \( \text{CM}_n(d) \) in Theorem 7.2 is organized as follows:

(a) We first prove an upper bound on the diameter of the core of \( \text{CM}_n(d) \), which consists of all vertices of degree at least \( (\log n)^\sigma \) for some \( \sigma > 0 \). This is the content of Theorem 7.9 below.

(b) Followed by the proof of Theorem 7.9, we use a second moment method to prove that any vertex that survives to sufficient large distance is whp quickly connected to the core.

Together, these two steps prove the upper bound on the typical distance for \( \text{CM}_n(d) \) in Theorem 7.2. The bound on the diameter of the core is also useful in studying the diameter of \( \text{CM}_n(d) \) when \( \tau \in (2, 3) \) and \( d_{\min} \geq 3 \), which we discuss in Section 7.4 below.

We take \( \sigma > 1/(3 - \tau) \) and define the core \( \text{Core}_n \) of the configuration model to be

\[
\text{Core}_n = \{i: d_i \geq (\log n)^\sigma\}, \quad (7.2.33)
\]
i.e., the set of vertices with degree at least \( (\log n)^\sigma \). Then, the diameter of the core is bounded in the following theorem, which is interesting in its own right:

**Theorem 7.9 (Diameter of the core)** Fix \( \tau \in (2, 3) \) and assume that (7.1.2) holds. For any \( \sigma > \frac{1}{3 - \tau} \), the diameter of \( \text{Core}_n \) is with high probability bounded above by

\[
2 \log \log n \left\lfloor \frac{\log (\tau - 2)}{\log (\tau - 1)} \right\rfloor + 1. \quad (7.2.34)
\]

**Proof** We note that (7.1.2) implies that, for some \( \alpha \in (1/2, 1/(\tau - 1)) \),
\[
\max_{i \in [n]} d_i \geq u_1, \quad \text{where} \quad u_1 = n^\alpha. \quad (7.2.35)
\]
Define
\[
\Gamma_1 = \{i: d_i \geq u_1\}, \quad (7.2.36)
\]
so that \( \Gamma_1 \neq \emptyset \). For some constant \( C > 0 \) to be determined later on, and for \( k \geq 2 \), we recursively define
\[
u_k = C \log n (u_{k-1})^{\tau - 2}.
\]

(7.2.37)
Then, we define
\[ \Gamma_k = \{ i : d_i \geq u_k \}. \] (7.2.38)

We identify \( u_k \) in the following lemma:

**Lemma 7.10** (Identification \((u_k)_{k \geq 1}\)) For every \( k \geq 1 \),
\[ u_k = \alpha^k (\log n)^b_n n^{c_k}, \] (7.2.39)
where
\[ c_k = \alpha (\tau - 2)^{k-1}, \quad a_k = b_k = \frac{1}{3} \left( 1 - (\tau - 2)^{k-1} \right). \] (7.2.40)

**Proof** We note that \( c_k, b_k, a_k \) satisfy the recursions, for \( k \geq 2 \),
\[ c_k = (\tau - 2)c_{k-1}, \quad b_k = 1 + (\tau - 2)b_{k-1}, \quad a_k = 1 + (\tau - 2)a_{k-1}, \] (7.2.41)
with initial conditions \( c_1 = \alpha, a_1 = b_1 = 0 \). Solving the recursions yields our claim.

In order to study connectivity of sets in \( CM_n(d) \), we rely on the following lemma, which is of independent interest:

**Lemma 7.11** (Connectivity sets in \( CM_n(d) \)) For any two sets of vertices \( A, B \subseteq [n] \),
\[ P( A \text{ not directly connected to } B) \leq e^{-d_A d_B / (2\ell_n)}, \] (7.2.42)
where, for any \( A \subseteq [n] \),
\[ d_A = \sum_{i \in A} d_i \] (7.2.43)
denotes the total degree of vertices in \( A \).

**Proof** There are \( d_A \) half-edges incident to the set \( A \), which we pair one by one. After having paired \( k \) half-edges, all to half-edges that are not incident to \( B \), the probability to pair the next half-edge to a half-edge that is not incident to \( B \) equals
\[ 1 - \frac{d_B}{\ell_n - 2k + 1} \leq 1 - \frac{d_B}{\ell_n}. \] (7.2.44)

Some half-edges incident to \( A \) may attach to other half-edges incident to \( A \), so that possibly fewer than \( d_A \) half-edges need to be paired to pair all half-edges incident to \( A \). However, since each pairing uses up at most 2 half-edges incident to \( A \), we need to pair at least \( d_A / 2 \) half-edges, so that
\[ P( A \text{ not directly connected to } B) \leq \left( 1 - \frac{d_B}{\ell_n} \right)^{d_A / 2} \leq e^{-d_A d_B / (2\ell_n)}, \] (7.2.45)
where we use that \( 1 - x \leq e^{-x} \).

The key step in the proof of Theorem 7.9 is the following proposition showing that whp every vertex in \( \Gamma_k \) is connected to a vertex in \( \Gamma_{k-1} \):
Proposition 7.12 (Connectivity between $\Gamma_{k-1}$ and $\Gamma_k$) Fix $\tau \in (2, 3)$ and assume that Conditions 1.5(a)-(b) and (7.1.2) hold. Fix $k \geq 2$, and take $C > 2E[D]/c$. Then, the probability that there exists an $i \in \Gamma_k$ that is not directly connected to $\Gamma_{k-1}$ is $o(n^{-\delta})$, for some $\delta > 0$ that is independent of $k$.

Proof. We note that, by definition,
\[
\sum_{i \in \Gamma_{k-1}} d_i \geq u_{k-1} |\Gamma_{k-1}| = u_{k-1} n [1 - F_n](u_{k-1}).
\] (7.2.46)
By (7.1.2), and since $k \mapsto u_k$ is decreasing with $u_1 = n^\alpha$, 
\[
[1 - F_n](u_{k-1}) \geq c(u_{k-1})^{1-\tau}.
\] (7.2.47)
As a result, we obtain that for every $k \geq 2$,
\[
\sum_{i \in \Gamma_{k-1}} d_i \geq cn(u_{k-1})^{2-\tau}.
\] (7.2.48)
By (7.2.48) and Lemma 7.11, using Boole’s inequality, the probability that there exists an $i \in \Gamma_k$ that is not directly connected to $\Gamma_{k-1}$ is bounded by 
\[
ne^{-\frac{u_{k-1}[1 - F(u_{k-1})]}{2n}} \leq ne^{-\frac{c(u_{k-1})^{2-\tau}}{2E[D]/c}} = n^{1 - \frac{C}{2E[D]/c}}.
\] (7.2.49)
By Conditions 1.5(a)-(b), $E[D_n] \to E[D]$, so that, as $n \to \infty$ and taking $C > 2E[D]/c$, we obtain the claim for any $\delta < \frac{C}{2E[D]/c} - 1$.

We now complete the proof of Theorem 7.9.

Proof of Theorem 7.9. Fix 
\[
k^* = \log \log n \frac{\log \tau}{|\log (\tau - 2)|}.
\] (7.2.50)
As a result of Proposition 7.12, whp, the diameter of $\Gamma_{k^*}$ is at most $2k^* + 1$, because the distance between any vertex in $\Gamma_{k^*}$ and $\Gamma_1$ is at most $k^*$, while, by Exercise 7.4, $\Gamma_1$ forms a complete graph. Therefore, it suffices to prove that 
\[
\text{Core}_n \subseteq \Gamma_{k^*}.
\] (7.2.51)
By (7.2.37), in turn, this is equivalent to 
\[
u_{k^*} \geq (\log n)^{\sigma},
\] for any $\sigma > 1/(3 - \tau)$. According to Lemma 7.10, 
\[
u_{k^*} = C^{\nu_{k^*}} (\log n)^{\nu_{k^*}} n^{e^{k^*}}.
\] (7.2.52)
We note that $n^{e^{k^*}} = e^{\log n^{(\tau - 2)k^*}}$. Since, for $2 < \tau < 3$, 
\[
x((\tau - 2)k^*) = x \cdot x^{-1} = 1,
\] (7.2.53)
we find with $x = \log n$ that $n^{e^{k^*}} = e$. Further, $b_k \to 1/(\tau - 3)$ as $k \to \infty$, so that $(\log n)^{b_k} = (\log n)^{1/((3 - \tau) + o(1))}$, and $a_k = b_k$, so that also $C^{a_k} = C^{1/((3 - \tau) + o(1))}$. We conclude that 
\[
u_{k^*} = (\log n)^{1/((3 - \tau) + o(1))},
\] (7.2.54)
so that, by picking \( n \) sufficiently large, we can make \( 1/(3 - \tau) + o(1) \leq \sigma \). This completes the proof of Theorem 7.9. \( \square \)

We continue to use Theorem 7.9 to prove a \( \log \log n \) upper bound on the typical distance \( \text{dist}_{CM_n(a)}(o_1, o_2) \) in the case where \( \tau \in (2, 3) \). We start by describing the setting. We assume that there exist \( \tau \in (2, 3), \alpha > 1/2 \) and \( c_1 \) such that, uniformly in \( n \) and \( x \leq n^\alpha \),

\[
[1 - F_n](x) \geq c_1 x^{-(\tau - 1)}. \tag{7.2.55}
\]

**Theorem 7.13 (A \( \log \log \) upper bound on typical distance for \( \tau \in (2, 3) \))** Suppose that the empirical distribution function \( F_n \) of the degrees \( d = (d_i)_{i \in [n]} \) satisfies Conditions 1.5(a)-(b) and (7.2.55). Then, for every \( \varepsilon > 0 \),

\[
\lim_{n \to \infty} \mathbb{P} \left( \text{dist}_{CM_n(a)}(o_1, o_2) \leq \frac{2(1 + \varepsilon) \log \log n}{| \log (\tau - 2) |} \mid \text{dist}_{CM_n(a)}(o_1, o_2) < \infty \right) = 1. \tag{7.2.56}
\]

**Proof** We make crucial use of the branching process approximation in Section 7.2.1. We let \( o_1, o_2 \) denote two vertices chosen uniformly at random from \( [n] \), and we recall that, for \( i \in \{1, 2\} \), \( Z_{m,i}^{(n)} \) denotes the number of unpaired or free half edges incident to vertices in \( B_m(o_i) \). By Corollary 7.3, \( (Z_{m,1}^{(1)}, Z_{m,2}^{(2)})_{m=0}^\infty \) can be whp perfectly coupled to \( (Z_{m,1}^{(1)}(Z_{m,2}^{(2)}))_{m=0}^\infty \), which are two independent unimodular or two-stage branching processes where the root has offspring distribution \( D \), and individuals in all further generations have offspring distribution \( D^* - 1 \).

We condition on \( (B_m(o_1), B_m(o_2)) \) are such that \( Z_{m,1}^{(n,1)} \geq 1, Z_{m,2}^{(n,2)} \geq 1 \). We couple these to \( (Z_{m,1}^{(1)}, Z_{m,2}^{(2)}) \). We note that, conditionally on \( Z_{m,1}^{(1)} \geq 1, Z_{m,2}^{(2)} \geq 1, Z_{m,1}^{(1)} \to \infty, Z_{m,2}^{(2)} \to \infty \) when \( m \to \infty \).

We condition on \( (B_m(o_1), B_m(o_2)) \), and denote the conditional distribution by \( \mathbb{P}_m \), and the expectation and variance under the measure \( \mathbb{P}_m \) by \( \mathbb{E}_m \) and \( \text{Var}_m \), respectively. We collapse the vertices in \( B_m(o_1) \) to a single vertex \( a_1 \) and \( B_m(o_2) \) to a single vertex \( a_2 \). The distribution of the resulting random graph is again a configuration model, with degrees \( d_{a_1} = Z_{m,1}^{(n,1)}, d_{a_2} = Z_{m,2}^{(n,2)} \) and vertex set \( R = [n] \cup \{(a_1, a_2) \} \setminus (B_m(o_1) \cup B_m(o_2)) \).

We apply Proposition 7.7 with \( k = \varepsilon \log \log n, a_1, b = \text{Core}_n \) and with \( I = \{i \in R: d_i \leq K\} \). Then, Proposition 7.7 gives that, conditionally on \( (B_m(o_1), B_m(o_2)) \) such that \( Z_{m,1}^{(n,1)} \geq 1, Z_{m,2}^{(n,2)} \geq 1 \),

\[
\mathbb{P}_m(N_k(a, b) = 0) \leq \text{Var}_m(N_k(a, b)) / \mathbb{E}_m[N_k(a, b)]^2 \leq O(K) \left( 1/Z_{m,1}^{(n,1)} + 1/Z_{m,2}^{(n,2)} \right) - \varepsilon, \tag{7.2.57}
\]

in the iterated limit where first \( n \to \infty \), followed by \( m \to \infty \). As a result, conditionally on \( (B_m(o_1), B_m(o_2)) \) such that \( Z_{m,1}^{(n,1)} \geq 1, Z_{m,2}^{(n,2)} \geq 1 \), with probability at least \( 1 - o(1) \), \( N_k(a, b) \geq 1 \), so that, on this event,

\[
\text{dist}_{CM_n(d)}(o_1, o_2) \leq \text{diam}_{CM_n(d)}(\text{Core}_n) + 2k \leq \frac{2(1 + \varepsilon) \log \log n}{| \log (\tau - 2) |}. \tag{7.2.58}
\]
We further use that, by Theorem 4.4,

\[ \mathbb{P}(\text{dist}_{\text{CM}_n(d)}(o_1, o_2) < \infty) \to \zeta^2, \quad (7.2.59) \]

while

\[ \mathbb{P}(Z_m^{(n_1)} \geq 1, Z_m^{(n_2)} \geq 1) \to \zeta^2 \quad (7.2.60) \]

in the iterated limit where first \( n \to \infty \), followed by \( m \to \infty \). As a result, with

\[ k_n = \frac{2(1+\varepsilon) \log \log n}{\log (\tau-2)}, \]

\[ \mathbb{P}(\text{dist}_{\text{CM}_n(d)}(o_1, o_2) \leq k_n, Z_n^{(1)} \geq 1, Z_n^{(2)} \geq 1) \]

\[ \geq \mathbb{P}(\text{dist}_{\text{CM}_n(d)}(o_1, o_2) < \infty) \]

\[ = \mathbb{P}(Z_n^{(1)} \geq 1, Z_n^{(2)} \geq 1) - o(1) \mathbb{P}(\text{dist}_{\text{CM}_n(d)}(o_1, o_2) < \infty) = 1 - o(1), \]

as required. This completes the proof of Theorem 7.13. \qed

Exercise 7.5 explores an alternative proof of Theorem 7.13.

### 7.3 Branching processes with infinite mean

When \( \tau \in (2, 3) \), the branching processes \((Z_j^{(1)})_{j \geq 0}\) and \((Z_j^{(2)})_{j \geq 0}\) are well-defined, but have infinite mean in generations 2, 3, etc. This leads us to consider branching processes with infinite mean. In this section, we give a scaling result for the generation sizes for such branching processes. This result will be crucial to describe the fluctuations of the typical distances in \( \text{CM}_n(d) \), and also allow us to understand how the ultrasmall distances of order \( \log \log n \) arise. The main result in this section is the following theorem:

**Theorem 7.14** (Branching processes with infinite mean) Let \((Z_n)_{n \geq 0}\) be a branching process with offspring distribution \( Z_1 = X \) having distribution function \( F_X \). Assume that there exist \( \alpha \in (0, 1) \) and a non-negative, non-increasing function \( x \mapsto \gamma(x) \), such that

\[ x^{-\alpha-\gamma(x)} \leq 1 - F_X(x) \leq x^{-\alpha+\gamma(x)}, \quad \text{for large } x, \quad (7.3.1) \]

where \( x \mapsto \gamma(x) \) satisfies

(i) \( x \mapsto x^{\gamma(x)} \) is non-decreasing,

(ii) \( \int_0^\infty \gamma(e^y) \, dy < \infty \), or, equivalently, \( \int_e^\infty \frac{\gamma(y)}{y \log y} \, dy < \infty \).

Then \( \alpha^n \log(Z_n \vee 1) \xrightarrow{a.s.} Y \), with \( \mathbb{P}(Y = 0) \) equal to the extinction probability of \((Z_n)_{n \geq 0}\).

In the analysis for the configuration model, \( \alpha = \tau - 2 \), as \( \alpha \) corresponds to the tail exponent of the size-biased random variable \( D^* \) (recall Lemma 1.14). Theorem 7.14 covers the case where the branching process has an offspring which
has very thick tails. Indeed, it is not hard to show that Theorem 7.14 implies that $E[X^s] = \infty$ for every $s > \alpha \in (0, 1)$ (see Exercise 7.7 below).

We do not prove Theorem 7.14 in full generality. Rather, we prove it in a simpler, yet still quite general case, in which $\gamma(x) = (\log x)^{\gamma-1}$ for some $\gamma \in [0, 1)$. See Exercise 7.6 to see that this case indeed satisfies the assumptions in Theorem 7.14.

**Proof of Theorem 7.14 for $\gamma(x) = (\log x)^{\gamma-1}$**. The proof is divided into four main steps. Define

$$M_n = \alpha^\gamma \log(Z_n \vee 1). \tag{7.3.2}$$

We shall first assume that $\mathbb{P}(Z_1 \geq 1) = 1$, so that $\eta = 1$. We start by splitting $M_n$ in a suitable way.

**The split**

For $i \geq 1$, we define

$$Y_i = \alpha^i \log\left(\frac{Z_i \vee 1}{(Z_{i-1} \vee 1)^{1/\alpha}}\right). \tag{7.3.3}$$

We can write

$$M_n = Y_1 + Y_2 + \cdots + Y_n. \tag{7.3.4}$$

From this split, it is clear that almost sure convergence of $M_n$ follows when the sum $\sum_{i=0}^\infty Y_i$ converges, which, in turn, is the case when

$$\sum_{i=1}^\infty E[|Y_i|] < \infty. \tag{7.3.5}$$

This is what we prove in the following three steps.

**Inserting normalization sequences**

We next investigate $E[|Y_i|]$. We prove by induction on $i$ that there exist constants $\kappa < 1$ and $C > 0$ such that

$$E[|Y_i|] \leq K\kappa^i. \tag{7.3.6}$$

For $i = 0$, this follows from the fact that, when (7.3.1) holds, the random variable $Y_1 = \alpha \log(Z_1 \vee 1)$ has a bounded absolute expectation. This initializes the induction hypothesis. We next turn to the advancement of the induction hypothesis. For this, we recall the definition of $u_n$ in [Volume 1, (2.6.7)], which states that

$$u_n = \inf\{x: 1 - F_x(x) \leq 1/n\}. \tag{7.3.7}$$

The interpretation of $u_n$ is that it indicates the order of magnitude of $\max_{i=1}^n X_i$, where $(X_i)_{i=1}^n$ are i.i.d. random variables with distribution function $F_x$.

Then we define

$$U_i = \alpha^i \log\left(\frac{u Z_{i-1} \vee 1}{u Z_{i-1} \vee 1^{1/\alpha}}\right), \quad V_i = \alpha^i \log\left(\frac{Z_i \vee 1}{u Z_{i-1} \vee 1}\right). \tag{7.3.8}$$
Then, $Y_i = U_i + V_i$, so that
\[
\mathbb{E}[|Y_i|] \leq \mathbb{E}[|U_i|] + \mathbb{E}[|V_i|].
\]  
(7.3.9)

We bound each of these terms separately.

Bounding the normalizing constants

In this step, we analyse the normalizing constants $n \mapsto u_n$, assuming (7.3.1), and use this, as well as the induction hypothesis, to bound $\mathbb{E}[|U_i|]$.

When (7.3.1) holds and since $\lim_{x \to \infty} \gamma(x) = 0$, there exists a constant $C \geq 1$ such that, for all $n \geq 1$,
\[
 u_n \leq C n^{1/\alpha + \varepsilon},
\]  
(7.3.10)

This gives a first bound on $n \mapsto u_n$. We next substitute this bound into (7.3.1) and use that $x \mapsto x^{\gamma(x)}$ is non-decreasing together with $\gamma(x) = (\log x)^{\gamma - 1}$, to obtain that
\[
1 + o(1) = n[1 - F_{X}(u_n)] \geq n[u_n^{-(\gamma - 1)} - (u_n)^{1/(\alpha + \varepsilon)}],
\]  
(7.3.11)

which, in turn, implies that there exists a constant $c > 0$ such that
\[
 u_n \leq n^{1/\alpha e^{(\log n)\gamma}}.
\]  
(7.3.12)

In a similar way, we can show the matching lower bound $u_n \geq n^{1/\alpha e^{-(\log n)\gamma}}$. As a result,
\[
\mathbb{E}[|U_i|] \leq c_\alpha \mathbb{E}[\log (Z_{i-1} \lor 1)^{\gamma}].
\]  
(7.3.13)

Using the concavity of $x \mapsto x^{\gamma}$ for $\gamma \in [0, 1)$, as well as Jensen’s Inequality, we arrive at
\[
\mathbb{E}[|U_i|] \leq c_\alpha \left( \mathbb{E}[\log (Z_{i-1} \lor 1)] \right)^{\gamma} = \alpha^{(1-\gamma)} \mathbb{E}[M_{i-1}]^{\gamma}.
\]  
(7.3.14)

By (7.3.4) and (7.3.6), which implies that $\mathbb{E}[M_{i-1}] \leq K\kappa/(1 - \kappa)$, we arrive at
\[
\mathbb{E}[|U_i|] \leq \alpha^{(1-\gamma)} e^{K\kappa/(1 - \kappa)}.
\]  
(7.3.15)

Logarithmic moment of an asymptotically stable random variable

In this step, we bound $\mathbb{E}[|V_i|]$. We note that by [Volume 1, Theorem 2.33] and for $Z_i$ quite large, the random variable $(Z_i \lor 1)/(u_{Z_i-1} \lor 1)$ should be close to a stable random variable. We make use of this fact by bounding
\[
\mathbb{E}[|V_i|] \leq \alpha \sup_{m \geq 1} \mathbb{E}[\log (S_m/u_m)],
\]  
(7.3.16)

where $S_m = X_1 + \cdots + X_m$, and $(X_i)_{i \geq 1}$ are i.i.d. copies of the offspring distribution $X$. We shall prove that there exists a constant $C > 0$ such that, for all $m \geq 1$,
\[
\mathbb{E}[\log (S_m/u_m)] \leq C.
\]  
(7.3.17)

In order to prove (7.3.17), we note that it suffices to bound
\[
\mathbb{E}\left[\log \frac{S_m}{u_m}\right] \leq C_+, \quad \mathbb{E}\left[\log \frac{S_m}{u_m}\right] \leq C_-, \quad \mathbb{E}\left[\log \frac{S_m}{u_m}\right] \leq C_0.
\]  
(7.3.18)
where, for \( x \in \mathbb{R}, x_+ = \max\{x, 0\} \) and \( x_- = \max\{-x, 0\} \). Since \( |x| = x_+ + x_- \), we then obtain \((7.3.17)\) with \( C = C_+ + C_- \). In order to prove \((7.3.17)\), we start by investigating \( \mathbb{E}\left[(\log (S_m/u_m))^-\right] \). We note that \((\log x)_- = \log(x^{-1} \vee 1)\), so that
\[
\mathbb{E}\left[(\log (S_m/u_m))^-\right] = \mathbb{E}\left[\log (u_m/(S_m \wedge u_m))\right],
\tag{7.3.19}
\]
where \( x \wedge y = \min\{x, y\} \). The function \( x \mapsto \log ((u_m/(x \wedge u_m)) \) is non-increasing, and, since \( S_m \geq X(m) \), where \( X(m) = \max_{1 \leq i \leq m} X_i \), we arrive at
\[
\mathbb{E}\left[\log (u_m/(S_m \wedge u_m))\right] \leq \mathbb{E}\left[\log (u_m/(X(m) \wedge u_m))\right],
\tag{7.3.20}
\]
We next use that, for \( x \geq 1, x \mapsto \log(x) \) is concave, so that, for every \( s, \)
\[
\mathbb{E}\left[\log (u_m/(X(m) \wedge u_m))\right] = \frac{1}{s} \mathbb{E}\left[\log ((u_m/(X(m) \wedge u_m))^s)\right] \leq \frac{1}{s} \log \left(\mathbb{E}[\left((u_m/(X(m) \wedge u_m))^s\right)]\right)
\leq \frac{1}{s} + \frac{1}{s} \log \left(u_m \mathbb{E}\left[X(m)^s\right]\right),
\tag{7.3.21}
\]
where, in the last step, we made use of the fact that \( u_m/(x \wedge u_m) \leq 1 + u_m/x \). Now rewrite \( X(m)^s = (\mathbb{Y}(m)^s)^s \), where \( \mathbb{Y}(m) \equiv X(1)^{-1} \) and \( \mathbb{Y}(m) = \max_{1 \leq j \leq m} \mathbb{Y}_j \). Clearly, \( \mathbb{Y}_j \in [-1, 0] \) since \( X_i \geq 1 \), so that \( \mathbb{E}[(-\mathbb{Y}(m)^s) < \infty \). Also, \( u_m \mathbb{Y}(m) = -u_m/X(m) \) converges in distribution to \( -E^{-1/\alpha} \), where \( E \) is exponential with mean 1, so it follows from \((\text{Pickands III, 1968, Theorem 2.1})\) that, as \( m \rightarrow \infty, \)
\[
\mathbb{E}\left[(u_m \mathbb{Y}(m)^s)\right] \rightarrow \mathbb{E}[E^{-1/\alpha}] < \infty,
\tag{7.3.22}
\]
We proceed with \( \mathbb{E}\left[(\log (S_m/u_m))^s\right] \), for which the proof is a slight adaptation of the above argument. Now we make use of the fact that \( (\log x)_+ = \log (x \vee 1) \leq 1 + x \) for \( x \geq 0 \), so that we must bound
\[
\mathbb{E}\left[\log (S_m \vee u_m/u_m)\right] = \frac{1}{s} \mathbb{E}\left[\log ((S_m \vee u_m/u_m))^s\right] \leq \frac{1}{s} + \log \left(\mathbb{E}\left[(S_m/u_m)^s\right]\right),
\tag{7.3.23}
\]
The discussion on \((\text{Hall, 1981, Page 565 and Corollary 1})\) yields, for \( s < \alpha, \)
\[
\mathbb{E}[S_m^s] = \mathbb{E}[|S_m|^s] \leq 2^{s/2} \lambda_s(t),
\]
for some function \( \lambda_s(t) \) depending on \( s, m \) and \( F_x \). Using the discussion on \((\text{Hall, 1981, Page 564})\), we have that \( \lambda_s(t) \leq C_s m^{s/\alpha} \ell(m^{1/\alpha})^s \), where \( \ell(\cdot) \) is a slowly-varying function. With some more effort, it can be shown that we can replace \( \ell(m^{1/\alpha}) \) by \( \ell(m) \), which gives
\[
\mathbb{E}[\log (S_m \vee u_m/u_m)] \leq \frac{1}{s} + \log \left(\mathbb{E}\left[(S_m/u_m)^s\right]\right) \leq \frac{1}{s} + \frac{C_s}{s} m^{s/\alpha} \ell(m)^s u_m^{-s} = \frac{1}{s} + 2^{s/2} \frac{C_s}{s},
\tag{7.3.24}
\]
and which proves the first bound in \((7.3.18)\) with \( C_+ = 1/s + 2^{s/2} \ell_s^s/s \).

**Completion of the proof of Theorem 7.14 when \( X \geq 1 \)**

Combining \((7.3.9)\) with \((7.3.15)\) and \((7.3.16)−(7.3.17)\), we arrive at
\[
\mathbb{E}[|Y_i|] \leq \alpha^{(1-\gamma)} c \left(\frac{K}{1-\gamma}\right)^\gamma + C \alpha^t \leq K \kappa^t,
\tag{7.3.25}
\]
7.3 Branching processes with infinite mean

when we take \( \kappa = \alpha^{1-\gamma} \) and we take \( K \) to be sufficiently large, for example \( K \geq 2C \) and \( K \geq 2c \left( \frac{1}{1 - \kappa} \right)^{\gamma} \). This completes the proof when the offspring distribution \( X \) satisfies \( X \geq 1 \).

**Completion of the proof of Theorem 7.14**

We finally extend the result to the setting where \( X = 0 \) with positive probability. Since \( E[X] = \infty \), the survival probability \( \zeta = P(Z_n \geq 1 \forall n \geq 0) \) satisfies \( \zeta > 0 \). Conditionally on extinction, clearly \( Z_n \xrightarrow{a.s.} 0 \), so that, on the survival event, \( \alpha^n \log(Z_n \vee 1) \xrightarrow{a.s.} Y \), where, conditionally on extinction, \( Y = 0 \).

It remains to prove that \( \alpha^n \log(Z_n \vee 1) \xrightarrow{a.s.} Y \) on the survival event. By [Volume 1, Theorem 3.12], we have that, conditionally on survival, \[
\frac{Z_n^{(\infty)}}{Z_n} \xrightarrow{a.s.} \xi > 0,
\]
where we recall that \( Z_n^{(\infty)} \) are the individuals in the \( n \)th generation which have an infinite line of descent. By [Volume 1, Theorem 3.11] and conditionally on survival, \( (Z_n^{(\infty)})_{n \geq 0} \) is again a branching process, now with offspring distribution \( p^{(\infty)} \) given in [Volume 1, (3.4.2)]. Note that, in particular, \( P(Z_1^{(\infty)} \geq 1) = 1 \), and we wish to apply Theorem 7.14 to \( Z_n^{(\infty)} \) instead of \( Z_n \). It is not hard to show that also \( p^{(\infty)} \) in [Volume 1, (3.4.2)] satisfies the conditions in Theorem 7.14 with the function \( x \mapsto \gamma^*(x) \), given by \( \gamma^*(x) = \gamma(x) + c/\log x \). Thus, conditionally on survival, \[
\alpha^n \log(Z_n^{(\infty)} \vee 1) \xrightarrow{a.s.} Y^{(\infty)},
\]
and combining (7.3.26) and (7.3.27), it immediately follows that, conditionally on survival, \[
\alpha^n \log(Z_n \vee 1) \xrightarrow{a.s.} Y^{(\infty)}.
\]
We conclude that Theorem 7.14 holds, where \( Y = 0 \) with probability \( \eta = 1 - \zeta \) and \( Y = Y^{(\infty)} \) with probability \( \zeta \).

We finally state some properties of the a.s. limit \( Y \) of \( (\alpha^n \log(Z_n \vee 1))_{n \geq 0} \), of which we omit a proof:

**Theorem 7.15** (Limiting variable for infinite-mean branching processes) *Under the conditions of Theorem 7.14,*

\[
\lim_{x \to \infty} \frac{\log P(Y > x)}{x} = -1,
\]
where is the a.s. limit of \( \alpha^n \log(Z_n \wedge 1) \).

Theorem 7.15 can be understood from the fact that, by (7.3.2)–(7.3.3),

\[
Y = \sum_{n=1}^{\infty} Y_i,
\]
which shows that $Y$ (7.3.3) suggests that the tails of $Y$ grow roughly at the same pace, and in particular, $|Y| \approx e^{x/(1+o(1))}$, which heuristically explains (7.3.29). The equality in (7.3.30) together with (7.3.3) suggests that the tails of $Y_1$ are equal to those of $Y$, which heuristically explains (7.3.29).

Intuition behind ultrasmall distances using Theorem 7.14

We now use the results in Theorem 7.14 to explain how we can understand the ultrasmall distances in the configuration model. First, note that

$$P(Y_1 > x) = P(Z_1 > e^{x/n}) = e^{-x(1+o(1))},$$

By (7.3.1),

$$E[Z_1] = e^{-x(1+o(1))},$$

which suggests (7.3.29). The equality in (7.3.30) together with (7.3.3) suggests that $Y_1$ satisfies (7.3.29). The equality in (7.3.30) together with (7.3.3) suggests that the tails of $Y_1$ are equal to those of $Y$, which heuristically explains (7.3.29).

So, let us use the branching process approximation from two sides. Now we rely on the statement that

$$P(\text{dist}_{CM,1}(o_1, o_2) \leq 2k) = P(B_2(o_1) \cap B_2(o_2) \neq \emptyset).$$

Again using (7.3.35), we see that $|B_2(o_1)| \approx e^{(\tau - 2) - k} Y_1(1 + o(1))$ and $|B_2(o_1)| \approx e^{(\tau - 2) - k} Y_2(1 + o(1))$, where $Y_1$ and $Y_2$ are independent. These grow roughly at the same pace, and in particular, $|B_2(o_1)| = n^{\Theta(1)}$ roughly at the same time, namely, when $k \approx \log \log n/|\log(\tau - 2)|$. Thus, we conclude that $\text{dist}_{CM,1}(o_1, o_2) \approx 2 \log \log n/|\log(\tau - 2)|$, as rigorously proved in Theorem 7.2. We will see that
7.4 The diameter of the configuration model

the above growth from two sides does allow for better branching-process approximations in more detail in Theorem 7.21 below.

7.4 The diameter of the configuration model

We continue the discussion of distances in the configuration model by investigating the diameter in the model.

7.4.1 The diameter of the configuration model: logarithmic case

Before stating the main result, we introduce some notation. Recall that $G^*_n(x)$ is the probability generating function of $p^* = (p^*_k)_{k \geq 0}$ defined in (4.1.1) (recall also (4.2.7)). We recall that $\xi$ is the extinction probability of the branching process with offspring distribution $p$ defined in (4.2.3) and further define

$$\mu = G^*_n(\xi) = \sum_{k=1}^{\infty} k^{\xi-1} p_k.$$  \hspace{1cm} (7.4.1)

When $\xi < 1$, we also have that $\mu \leq 1$. Then, the main result is as follows:

**Theorem 7.16** (Diameter of the configuration model)  Let Conditions 1.5(a)-(b) hold. Assume that $n_1 = 0$ when $p_1 = 0$, and that $n_2 = 0$ when $p_2 = 0$. Then,

$$\frac{\text{diam}(\text{CM}_n(d))}{\log n} \overset{p}{\rightarrow} \frac{1}{\log \nu} + \frac{\mathbb{1}_{\{p_1 > 0\}}}{|\log \mu|} + \frac{\mathbb{1}_{\{p_1 = 0, p_2 > 0\}}}{|\log p^*_2|}. \hspace{1cm} (7.4.2)$$

We note that, by Theorem 7.1 and Theorem 7.16, the diameter of the configuration model is strictly larger than the typical graph distance, except when $p_1 = p_2 = 0$. In the latter case, the degrees are at least three, so that thin lines are not possible, and the configuration model is whp connected (recall Theorem 4.15). We also remark that Theorem 7.16 applies not only to the finite-variance case, but also to the finite-mean and infinite-variance case. In the latter case, the diameter is of order $\log n$ unless $p_1 = p_2 = 0$, in which case Theorem 7.16 implies that the diameter is $o_p(\log n)$. By [Volume 1, Corollary 7.17], Theorem 7.16 also applies to uniform random graphs with a given degree sequence. This shall be used in the examples below:

**Random regular graphs**

Let $r$ be the degree of the random regular graph, where $r \geq 3$. By [Volume 1, Corollary 7.17], the diameter of a random regular $r$-graph has with high probability the same asymptotics as the diameter of $\text{CM}_n(d)$, where $d_i = r$ with probability 1. Thus, $p_r = 1$ and $p_i = 0$ for any $i \neq r$. We assume that $nr$ is even, so that the degree sequence is feasible. It is not hard to see that all assumptions of Theorem 7.16 are satisfied. Moreover, $\nu = r - 1$. When $r \geq 3$, we thus obtain
that
\[
\frac{\text{diam}(\text{CM}_n(d))}{\log n} \xrightarrow{p} \frac{1}{\log (r-1)}.
\] (7.4.3)

When \(r = 2\), on the other hand, the graph is critical, so that there is no giant component. Since \(\nu = 1\), we have that \(\mu = \nu = 1\), so that \(\text{diam}(\text{CM}_n(d)) \gg \log n\).

This is quite reasonable, since the graph will consist of a collection of cycles. The diameter of such a graph is equal to half the longest cycle. Exercises 7.10–7.11 explore the length of the longest cycle in a random 2-regular graph, and the consequence of having a long cycle on the diameter.

**Erdős-Rényi random graph**

We next study the diameter of \(\text{ER}_n(\lambda/n)\). We let \(\lambda > 1\). By [Volume 1, Theorem 5.12], Conditions 1.5(a)-(b) hold with \(p_k = e^{-\lambda} \lambda^k/k!\). Also, \(\mu = \mu_\lambda\), the dual parameter in [Volume 1, (3.6.6)] (see Exercise 7.12).

We again make essential use of [Volume 1, Theorem 7.18], which relates the configuration model and the generalized random graph. We note that \(\text{ER}_n(\lambda/n)\) is the same as \(\text{GRG}_n(w)\), where (recall [Volume 1, Exercise 6.1])
\[
w_i = \frac{n\lambda}{n - \lambda},
\] (7.4.4)

Clearly, \(w = (n\lambda/(n - \lambda))_{i \in [n]}\) satisfies Conditions 1.1(a)-(c), so that also the degree sequence of \(\text{ER}_n(\lambda/n)\) satisfies Conditions 1.5(a)-(c), where the convergence holds in probability (recall [Volume 1, Theorem 5.12]). From the above identifications and using [Volume 1, Theorem 7.18], we find that
\[
\frac{\text{diam}(\text{ER}_n(\lambda/n))}{\log n} \xrightarrow{p} \frac{1}{\log \lambda} + 2 \frac{1}{|\log \mu_\lambda|}.
\] (7.4.5)

This identifies the diameter of the Erdős-Rényi random graph. In particular, for this case, Theorem 7.16 agrees with Theorem 6.21.

**A sketch of the proof of Theorem 7.16**

We recognize the term \(\log n\) as corresponding to the typical distances in the graph (recall Theorem 7.1). Except when \(p_1 = p_2 = 0\), there is a correction to this term, which is due to long, yet very thin, neighborhoods.

When \(p_1 > 0\), yet \(\nu > 1\), the outside of the giant component is a subcritical graph, which locally looks like a subcritical branching process due to the duality principle. In the subcritical case, there exist trees of size up to \(\Theta(\log n)\), and the maximal diameter of such trees is close to \(\log n/|\log \mu|\) for some \(\mu < 1\). In the supercritical case, instead, in the complement of the giant, there exist trees of size up to \(\Theta(\log n)\), and the maximal diameter of such trees is close to \(\log n/|\log \mu|\), where \(\mu\) is the dual parameter. Of course, these will not be the components of maximal diameter in the supercritical case.

In the supercritical case, we can view the giant as consisting of the 2-core and all the trees that hang off it. The 2-core is the maximal subgraph of the
7.4 The diameter of the configuration model

The diameter of the configuration model giant for which every vertex has degree at least 2. It turns out that the trees that hang off the giant have a very similar law as the subcritical trees outside of the giant. Therefore, the maximal diameter of such trees is again of the order \(\log n/|\log \mu|\). Asymptotically, the diameter is formed between pairs of vertices for which the diameter of the tree that they are in when cutting away the 2-core is the largest, thus giving rise to the \(\log n/|\log \mu|\) contribution, whereas the distances between the two vertices closest to them in the 2-core is close to \(\log n\).

This can be understood by realizing that for most vertices, these trees have a bounded height, so that the typical distances in the 2-core and in the giant are closely related.

The above intuition does not give the right answer when \(p_1 = 0\), yet \(p_2 > 0\). Assume that \(n_1 = 0\). Then, the giant and the 2-core are close to each other, so that the above argument does not apply. Instead, there it turns out that the diameter is realized by the diameter of the 2-core, which is close to \(\log n[1/\log \nu + 1/|\log p_2^*|]\). Indeed, the 2-core will contain long paths of degree 2 vertices. The longest such paths will be is close to \(\log n/|\log p_2^*|\). Therefore, the longest distance between a vertex inside this long path and the ends of the path is close to \(\log n/2|\log p_2^*|\).

Now it turns out that pairs of such vertices realize the asymptotic diameter, which explains why the diameter is then close to \(\log n[1/\log \nu + 1/|\log p_2^*|]\).

Finally, we discuss what happens when \(p_1 = p_2 = 0\). In this case, the assumption in Theorem 7.16 implies that \(n_1 = n_2 = 0\), so that \(d_{\text{min}} \geq 3\). Then, \(\text{CM}_n(d)\) is whp connected, and the 2-core is the graph itself. Also, there cannot be any long lines, since every vertex has degree at least 3, so that local neighborhoods grow exponentially with overwhelming probability. Therefore, the graph distances and the diameter have the same asymptotics, as in Theorem 7.16.

The above case-distinctions explain the intuition behind Theorem 7.16. This intuition is far from a proof. The proof by Fernholz and Ramachandran (2007) involves a precise analysis of the trees that are augmented to the 2-core and their maximal diameter, as well as an analysis that the pairs determined above really do create shortest paths that are close to the diameter. We will not discuss this further.

7.4.2 Diameter of the configuration model for \(\tau \in (2, 3)\): \(\log \log\) case

We next use Theorem 7.9 to study the diameter of \(\text{CM}_n(d)\) when \(\tau \in (2, 3)\). Note that the diameter is equal to a positive constant times \(\log n\) by Theorem 7.16 when \(p_1 + p_2 > 0\). Therefore, we turn to the case where \(p_1 = p_2 = 0\). When \(d_{\text{min}} \geq 3\), we know by Theorem 4.15 that \(\text{CM}_n(d)\) is whp connected. The main result is as follows:

**Theorem 7.17** (Diameter of CM\(_n\)(d) for \(\tau \in (2, 3)\)) Suppose that the empirical distribution function \(F_n\) of the degrees \(d = (d_i)_{i \in [n]}\) satisfies Condition 1.5(a)-(b) and that (7.1.2) holds. Assume further that \(d_{\text{min}} = \min_{i \in [n]} d_i \geq 3\) and \(p_{d_{\text{min}} = 3}\).
\[ P(D = d_{\text{min}}) > 0. \] Then,
\[
\frac{\text{diam}(\text{CM}_n(d))}{\log \log n} \xrightarrow{\tau} \frac{2}{|\log (\tau - 2)|} + \frac{2}{\log (d_{\text{min}} - 1)}. \tag{7.4.6}
\]

When comparing Theorem 7.17 to Theorem 7.2, we see that for \(d_{\text{min}} \geq 3\), the diameter is of the same order \(\log \log n\) as the typical distance, but that the constant differs. As we have also seen in the sketch of proof of Theorem 7.16, the diameter is due to pairs of vertices that have small or thin local neighborhoods. When \(p_1 + p_2 > 0\), these thin parts are close to lines, and they can be of length that is logarithmic in \(n\). When \(d_{\text{min}} \geq 3\), however, we see that even the thinnest possible neighborhoods are bounded below by a binary tree. Indeed, by assumption, there is a positive proportion of vertices of degree \(d_{\text{min}}\). As a result, we will see that the expected number of vertices whose \((1 - \varepsilon) \log \log n / \log (d_{\text{min}} - 1)\) neighborhood only contains vertices with degree \(d_{\text{min}}\) tends to infinity. The minimal path between two such vertices then consists of three parts: the two paths from the two vertices to leave their minimally connected neighborhood, and then the path between the boundaries of these minimally connected neighborhoods. These minimal neighborhoods are at the typical distance \(2 \log \log n / |\log (\tau - 2)|\), as in Theorem 7.2. This explains why in Theorem 7.17 there is an extra term \(2 \log \log n / \log (d_{\text{min}} - 1)\).

We will not give the entire proof of Theorem 7.17, but give a rather detailed sketch of the proof. We sketch the upper and a lower bounds on the diameter. We start with the lower bound, which is the easier part:

**Lower bound on the diameter**

We call a vertex \(v\) minimally-\(k\)-connected when all \(i \in B_k(v)\) satisfy \(d_i = d_{\text{min}}\), so that all vertices at distance at most \(k\) have the minimal degree. Let \(M_k\) denote the number of minimally-\(k\)-connected vertices. To prove the lower bound in on the diameter, we show that \(M_k \xrightarrow{\tau} \infty\) for \(k = (1 - \varepsilon) \log \log n / \log (d_{\text{min}} - 1)\). Followed by this, we show that two minimally \(k\)-connected vertices \(o_1, o_2\) are such that whp the distance between \(\partial B_k(o_1)\) and \(\partial B_k(o_2)\) is at least \(2 \log \log n / |\log (\tau - 2)|\). We start by computing the first and second moment of \(M_k\) in the following lemma:

**Lemma 7.18 (Moments of number of minimally-\(k\)-connected vertices)** Let \(\text{CM}_n(d)\) satisfy that \(d_{\text{min}} \geq 3\), \(n d_{\text{min}} > d_{\text{min}} (d_{\text{min}} - 1)^{k-1}\). Then, for all \(k \geq 1\),
\[
E[M_k] = n d_{\text{min}} \prod_{i=1}^{d_{\text{min}} (d_{\text{min}} - 1)^{k-1}} \frac{d_{\text{min}} (n d_{\text{min}} - (i - 1))}{\ell_n - 2i + 1}, \tag{7.4.7}
\]
and, for \(k\) such that \(d_{\text{min}} (d_{\text{min}} - 1)^{k-1} \leq \ell_n / 8\),
\[
E[M_k^2] \leq E[M_k]^2 + E[M_k] \left[ \frac{d_{\text{min}}}{d_{\text{min}} - 2} (d_{\text{min}} - 1)^k \right] + \frac{2 n d_{\text{min}} d_{\text{min}}^2 (d_{\text{min}} - 1)^{2k}}{(d_{\text{min}} - 2) \ell_n}. \tag{7.4.8}
\]
Consequently, for \(k_n \leq (1 - \varepsilon)/ \log (d_{\text{min}} - 1)\),
\[
M_k \xrightarrow{\tau} \infty. \tag{7.4.9}
\]
Proof We start by proving (7.4.7). We note that each vertex of degree \(d_{\text{min}}\) has the same probability of being minimally-\(k\)-connected, and that there are precisely \(n_{d_{\text{min}}}\) vertices of degree \(d_{\text{min}}\), so that

\[
\mathbb{E}[M_k] = n_{d_{\text{min}}} \mathbb{P}(v \text{ is minimally-}k\text{-connected}).
\]  

(7.4.10)

Vertex \(v\) with \(d_v = d_{\text{min}}\) is minimally-\(k\)-connected when all its half-edges at distance at most \(k\) are paired to half-edges incident to a distinct vertex having minimal degree \(d_{\text{min}}\), and no cycles occur in \(B_k(v)\). When \(i - 1\) half-edges are paired to distinct vertices of degree \(d_{\text{min}}\), then the probability that the \(i\)th half-edge is again paired to a distinct vertex of degree \(d_{\text{min}}\) equals \(d_{\text{min}}(n_{d_{\text{min}}} - (i - 1)) / \ell_n - 2i + 1\). \hspace{1cm} (7.4.11)

Since for \(v\) to be minimally-\(k\)-connected, there are \(d_{\text{min}}(d_{\text{min}} - 1)^{k-1}\) half-edges that need to be paired to distinct vertices of degree \(d_{\text{min}}\), this proves (7.4.7).

To prove (7.4.8), we note that

\[
\mathbb{E}[M^2_k] = \sum_{v_1, v_2 \in [n]} \mathbb{P}(v_1, v_2 \text{ are minimally-}k\text{-connected}).
\]  

(7.4.12)

We split the above probability depending on whether \(B_k(o_1) \cap B_k(o_2) = \emptyset\) or not. The contribution to \(\mathbb{E}[M^2_k]\) due to \(B_k(o_1) \cap B_k(o_2) = \emptyset\) is, similarly to the proof of (7.4.7), equal to

\[
n_{d_{\text{min}}}(n_{d_{\text{min}}} - i_{k-1}) \prod_{i=1}^{2i} \frac{d_{\text{min}}(n_{d_{\text{min}}} - (i - 1))}{\ell_n - 2i + 1},
\]  

(7.4.13)

where we abbreviate \(i_{k} = d_{\text{min}}(d_{\text{min}} - 1)^{k-1}\) and note that \(n_{d_{\text{min}}} - i_{k-1} > 0\), since, by assumption, \(n_{d_{\text{min}}} > d_{\text{min}}(d_{\text{min}} - 1)^{k-1}\).

We use that \(i \mapsto d_{\text{min}}(n_{d_{\text{min}}} - (i - 1)) / (\ell_n - 2i + 1)\) is decreasing since

\[
\frac{(n_{d_{\text{min}}} - (i - 1))}{\ell_n - 2i + 1} \geq \frac{n_{d_{\text{min}}} - i}{\ell_n - 2i - 1},
\]  

(7.4.14)

precisely when \(\ell_n \geq 2n_{d_{\text{min}}} + 1\), and which is true since \(\ell_n \geq d_{\text{min}}n_{d_{\text{min}}} \geq 3n_{d_{\text{min}}}\). Therefore, the contribution to \(\mathbb{E}[M^2_k]\) from \(v_1\) and \(v_2\) satisfying \(B_k(v_1) \cap B_k(v_2) = \emptyset\) is at most

\[
n_{d_{\text{min}}}^2 \left( \prod_{i=1}^{i_{k}} \frac{d_{\text{min}}(n_{d_{\text{min}}} - (i - 1))}{\ell_n - 2i + 1} \right)^2 = \mathbb{E}[M^2_k],
\]  

(7.4.15)

which is the first contribution to the r.h.s. of (7.4.8).

We are left to deal with the contribution to \(\mathbb{E}[M^2_k]\) from \(v_1\) and \(v_2\) such that
When \( v_1 \) is \( k \)-minimally connected,

\[
|B_k(v_1)| = 1 + \sum_{l=1}^{k} d_{\text{min}}(d_{\text{min}} - 1)^{l-1} \quad (7.4.16)
\]

\[
= 1 + d_{\text{min}}(d_{\text{min}} - 1)^{-k} \leq d_{\text{min}}(d_{\text{min}} - 1)^{-1}.
\]

Therefore, the contribution due to \( v_2 \in B_k(v_1) \) is bounded by

\[
\mathbb{E}[M_{k}] \frac{d_{\text{min}}}{d_{\text{min}} - 2} (d_{\text{min}} - 1)^{-1}, \quad (7.4.17)
\]

which is the second contribution to the r.h.s. of (7.4.8). Finally, we study the case where \( B_k(v_1) \cap B_k(v_2) \neq \emptyset \), but \( v_2 \notin B_k(v_1) \). When \( B_k(v_1) \cap B_k(v_2) \neq \emptyset \), but \( v_2 \notin B_k(v_1) \), then one of the \( d_{\text{min}}(d_{\text{min}} - 1)^{k} \) half-edges in \( B_k(v_1) \) needs to be connected to one of the \( d_{\text{min}}(d_{\text{min}} - 1)^{l-k} \) half-edges in \( B_{l-k}(v_2) \), where \( l = \text{dist}_{CM}(d,v_1,v_2) \in [2k]/[k] \). Conditionally on \( v_1 \) being \( k \)-minimally connected and \( v_2 \) being \( l-k \)-minimally connected, the probability that this occurs is at most

\[
\mathbb{E}[M_{k} \sum_{l=k+1}^{2k} \mathbb{E}[M_{l-k}] \frac{2d_{\text{min}}(d_{\text{min}} - 1)^{k}d_{\text{min}}(d_{\text{min}} - 1)^{l-k-1}}{\ell_{n}}, \quad (7.4.19)
\]

We bound \( \mathbb{E}[M_{l-k}] \leq n_{d_{\text{min}}} \) and sum

\[
\sum_{l=k+1}^{2k} (d_{\text{min}} - 1)^{l-k-1} \leq \frac{(d_{\text{min}} - 1)^{2k}}{d_{\text{min}} - 2}, \quad (7.4.20)
\]

to arrive at the third and final contribution to the r.h.s. of (7.4.8).

To complete the proof of the lower bound on the diameter, we fix \( \varepsilon > 0 \) sufficiently small, and take \( k^{*} = \left(1 - \varepsilon\right)^{\frac{\log n}{\log (d_{\text{min}} - 1)}} \). Clearly,

\[
d_{\text{min}}(d_{\text{min}} - 1)^{k^{*} - 1} \leq (\log n)^{1 - \varepsilon} \leq \ell_{n}/8, \quad (7.4.21)
\]

so that, in particular, we may use Lemma 7.18.

We note that \( n_{d_{\text{min}}}/n \rightarrow p_{d_{\text{min}}} \) by Condition 1.5(a) and \( p_{d_{\text{min}}} > 0 \) by assumption. Therefore, by Conditions 1.5(a)-(b), \( d_{\text{min}}n_{d_{\text{min}}}/\ell_{n} \rightarrow \lambda_{d_{\text{min}}} \), where we define \( \lambda_{d_{\text{min}}} = d_{\text{min}}p_{d_{\text{min}}}/\mathbb{E}[D] \). By (7.4.7) in Lemma 7.18,

\[
\mathbb{E}[M_{k}] \geq n(p_{d_{\text{min}}} - \delta)(\lambda_{d_{\text{min}}} - \delta)d_{\text{min}}(d_{\text{min}} - 1)^{k^{*} - 1}. \quad (7.4.22)
\]

As a result, \( \mathbb{E}[M_{k^{*}}] \geq n(p_{d_{\text{min}}} - \delta)(\lambda_{d_{\text{min}}} - \delta)(\log n)^{1 - \varepsilon} \). Further, by (7.4.8) in Lemma 7.18,

\[
\text{Var}(M_{k^{*}}) = o(\mathbb{E}[M_{k^{*}}]^2), \quad (7.4.23)
\]
show that

\[ M_{k^*} / \mathbb{E}[M_{k^*}] \xrightarrow{p} 1. \]  

(7.4.24)

We conclude that, whp, \( M_{k^*} \geq n^{1-o(1)} \). Since each minimally-\( k^* \)-connected vertex uses up at most

\[ 1 + \sum_{i=1}^{k^*} d_{\min}(d_{\min} - 1)^{i-1} = n^{o(1)} \]  

(7.4.25)

vertices of degree \( d_{\min} \), whp there must be at least two minimally-\( k^* \)-connected vertices whose \( k^* \)-neighborhoods are disjoint. We fix two such vertices and denote them by \( v_1 \) and \( v_2 \). We note that \( v_1 \) and \( v_2 \) have precisely \( d_{\min}(d_{\min} - 1)^{k^*-1} \) unpaired half-edges in \( \partial B_{k^*}(v_1) \) and \( \partial B_{k^*}(v_2) \). Let \( A_{12} \) denote the event that \( v_1, v_2 \) are minimally-\( k^* \)-connected with their \( k^* \)-neighborhoods being disjoint.

Conditionally on \( A_{12} \), the random graph obtained by collapsing the half-edges in \( \partial B_{k^*}(v_1) \) to a single vertex \( a \) and the half-edges in \( \partial B_{k^*}(v_1) \) to a single vertex \( b \) is a configuration model on the vertex set \( \{a, b\} \cup [n] \setminus (B_{k^*}(v_1) \cup B_{k^*}(v_2)) \), having degrees \( \mathbf{d} \) given by \( \tilde{d}_a = \tilde{d}_b = d_{\min}(d_{\min} - 1)^{k^*-1} \) and \( \tilde{d}_i = d_i \) for every \( i \in [n] \setminus (B_{k^*}(v_1) \cup B_{k^*}(v_2)) \).

By the truncated first moment method on paths, performed in the proof of Theorem 7.6 (recall (7.2.21)), it follows that, for any \( \varepsilon > 0 \),

\[ \mathbb{P}\left( \text{dist}_{CM_n}(\mathbf{d})(\partial B_{k^*}(v_1), \partial B_{k^*}(v_2)) \leq (1 - \varepsilon) \frac{2 \log \log n}{\log (\tau - 2)} \mid A_{12} \right) = o(1). \]  

(7.4.26)

Therefore, whp,

\[ \text{diam}(CM_n(\mathbf{d})) \geq (1 - \varepsilon) \frac{2 \log \log n}{\log (\tau - 2)} + 2k^* \]  

(7.4.27)

\[ = (1 - \varepsilon) \log \log n \left[ \frac{2}{\log (\tau - 2)} + \frac{2}{\log (d_{\min} - 1)} \right]. \]

Since \( \varepsilon > 0 \) is arbitrary, this proves the lower bound on \( \text{diam}(CM_n(\mathbf{d})) \) in Theorem 7.17.

**Sketch of the upper bound on the diameter**

We now sketch the proof of the upper bound on the diameter. We aim to prove that, with \( k_n = \log \log n [2 / \log (\tau - 2)] + 2 / \log (d_{\min} - 1) \),

\[ \mathbb{P}(\exists v_1, v_2 \in [n]: \text{dist}_{CM_n}(\mathbf{d})(v_1, v_2) \geq (1 + \varepsilon)k_n) = o(1). \]  

(7.4.28)

We already know that whp \( \text{diam}(\text{Core}_n) \leq 2 \log \log n / \log (\tau - 2) + 1 \), and we will assume this from now on. Fix \( v_1, v_2 \in [n] \). Then (7.4.28) follows when we can show that

\[ \mathbb{P}(\exists v: \text{dist}_{CM_n}(\mathbf{d})(v, \text{Core}_n) \geq (1 + \varepsilon) \log \log n / \log (d_{\min} - 1)) = o(1), \]  

(7.4.29)

and we will argue that, uniformly in \( v \in [n] \),

\[ \mathbb{P}(\text{dist}_{CM_n}(\mathbf{d})(v, \text{Core}_n) \geq (1 + \varepsilon) \log \log n / \log (d_{\min} - 1)) = o(1/n), \]  

(7.4.30)
which would prove (7.4.28).

To prove (7.4.30), it is convenient to explore the neighborhood of the vertex \( v \) by only pairing up the first \( d_{\text{min}} \) half-edges incident to \( v \) and the \( d_{\text{min}} - 1 \) half-edges incident to any other vertex appearing in the neighborhood. We call this exploration graph the \( k \)-exploration graph. One can show that it is quite unlikely that there are many cycles within this exploration graph, so that it is actually close to a tree. Therefore, the number of vertices on the boundary of this \( k \)-exploration graph with \( k^* = (1 + \epsilon/2) \log \log n / \log (d_{\text{min}} - 1) \) is close to

\[
(d_{\text{min}} - 1)^k \approx (\log n)^{1+\epsilon/2}.
\]

(7.4.31)

This is large, but not extremely large. However, one of these vertices is bound to have quite large degree, and thus, from this vertex, it is quite likely that we connect to \( \text{Core}_n \) quickly, meaning in \( o(\log \log n) \) steps. More precisely, the main ingredient in the upper bound is the statement that, for \( k^* = (1 + \epsilon/2) \log \log n / \log (d_{\text{min}} - 1) \) and whp, the \( k \)-exploration tree connects to \( \text{Core}_n \) whp in less than \( k^* \). We omit further details.

7.5 Related results for the configuration model

In this section, we discuss related results for the configuration model. We start by discussing the distances in infinite-mean configuration models.

7.5.1 Distances for infinite-mean degrees

In this section, we assume that there exist \( \tau \in (1, 2) \) and \( c > 0 \) such that

\[
\lim_{x \to \infty} x^{\tau-1} \log [1 - F](x) = c.
\]

(7.5.1)

We study the configuration model \( \text{CM}_n(d) \) where the degrees \( d = (d_i)_{i \in [n]} \) are an i.i.d. sequence of random variables with distribution \( F \) satisfying (7.5.1).

We make heavy use of the notation used in [Volume 1, Theorem 7.23], which we first recall. Recall that the random probability distribution \( P = (P_i)_{i \geq 1} \) is given by

\[
P_i = Z_i / Z,
\]

(7.5.2)

where \( Z_i = \Gamma_i^{-1/(\tau-1)} \) and \( \Gamma_i = \sum_{j=1}^i E_i \) with \( (E_i)_{i \geq 1} \) an i.i.d. sequence of exponential random variables with parameter 1, and where \( Z = \sum_{i \geq 1} \Gamma_i \). Recall further that \( M_{p,k} \) is a multinomial distribution with parameters \( k \) and probabilities \( P = (P_i)_{i \geq 1} \). Thus, \( M_{p,k} = (B_1, B_2, \ldots) \), where, conditionally on \( P = (P_i)_{i \geq 1} \), \( B_i \) is the number of outcomes \( i \) in \( k \) independent trials such that each outcome is equal to \( i \) with probability \( P_i \). In [Volume 1, Theorem 7.23], the random variable \( M_{p,D_1} \) appears, where \( D_1 \) is independent of \( P = (P_i)_{i \geq 1} \). We let \( M_{p,D_1}^{(i)} \) and \( M_{p,D_2}^{(i)} \) be two random variables which are conditionally independent given \( P = (P_i)_{i \geq 1} \).
7.5 Related results for the configuration model

In terms of this notation, the main result on distances in \( CM_n(d) \) when the degrees have infinite mean is the following:

**Theorem 7.19** (Distances in \( CM_n(d) \) with i.i.d. infinite mean degrees) Fix \( \tau \in (1, 2) \) in (7.5.1) and let \((d_i)_{i \in [n]}\) be a sequence of i.i.d. copies of \( D \). Then, \( CM_n(d) \) satisfies

\[
\lim_{n \to \infty} P(\text{dist}_{CM_n(d)}(o_1, o_2) = 2) = 1 - \lim_{n \to \infty} P(\text{dist}_{CM_n(d)}(o_1, o_2) = 3) = p_F \in (0, 1).
\]  

(7.5.3)

In terms of this notation, the main result on distances in \( CM_n(d) \) for \( \tau = 1.8 \) and \( n = 10^3, 10^4, 10^5 \).

**Figure 7.2** Empirical probability mass function of the \( \text{dist}_{CM_n(d)}(o_1, o_2) \) for \( \tau = 1.8 \) and \( n = 10^3, 10^4, 10^5 \).

**Proof** We sketch the proof of Theorem 7.19. First, whp, both \( d_1 \leq \log n \) and \( d_2 \leq \log n \). The event that \( \text{dist}_{CM_n(d)}(o_1, o_2) = 1 \) occurs precisely when one of the \( d_1 \) half-edges of vertex 1 is attached to one of the \( d_2 \) half-edges of vertex 2. Also, with high probability, \( \ell_n \geq n^{1/(\tau - 1) - \epsilon} \). Therefore, on the event that \( \ell_n \geq n^{1/(\tau - 1) - \epsilon} \) and \( d_1 \leq \log n \) and \( d_2 \leq \log n \), the probability that \( \text{dist}_{CM_n(d)}(o_1, o_2) = 1 \) is bounded above by

\[
\frac{(\log n)^2}{n^{1/(\tau - 1) - \epsilon}} = o(1).
\]  

(7.5.4)

We note that the proof of [Volume 1, Theorem 7.23] implies that \( M_{P,D_1}^{(1)} \) denotes the number of edges between vertex 1 and the largest order statistics. Indeed, \( M_{P,D_1}^{(1)} = (B_1^{(1)}, B_2^{(1)}, \ldots) \), where \( B_i^{(1)} \) is the number of edges between vertex \( i \) and the vertex with degree \( d_{i+\ell_n} \). The same applies to vertex 2. As a result, when \( M_{P,D_1}^{(1)} \) and \( M_{P,D_2}^{(2)} \) have an identical outcome, then the typical graph distance equals 2. We are left to prove that the typical graph distance is bounded by 3 with high
probability. By [Volume 1, (2.6.17)], we have that \(\xi_k k^{1/(r-1)} \xrightarrow{p} 1\) as \(k \to \infty\). Thus, when \(K\) is large, the probability that vertex 1 is not connected to any of the vertices corresponding to \((d_{(n+1-i)})_{i=1}^{K}\) converges to 0 when \(K\) tends to infinity.

Let \(P_n\) denote the conditional probability given the degrees \((d_i)_{i \in [n]}\). For \(i \in [n]\), we let \(v_i\) be the vertex corresponding to the \(i\)th order statistic \(d_{(n+1-i)}\). By Lemma 7.11,

\[
P_n(v_i \text{ not directly connected to } v_j) \leq e^{-\frac{d_{(n+1-i)}d_{(n+1-j)}}{2n}}. \tag{7.5.5}
\]

Moreover, \(d_{(n+1-i)}, d_{(n+1-j)} \geq n^{1/(r-1)-\varepsilon}\) with high probability for \(n\) sufficiently large and any \(\varepsilon > 0\), while whp \(\ell_n \leq n^{1/(r-1)+\varepsilon}\). As a result, whp,

\[
P_n(v_i \text{ not directly connected to } v_j) \leq e^{-n^{1/(r-1)-3\varepsilon}}. \tag{7.5.6}
\]

Therefore, for fixed \(K\) and for every \(i, j \in [K]\), the vertices \(v_i\) and \(v_j\) are whp neighbors. This implies that the vertices corresponding to the high order statistics form a complete graph. We have already concluded that 1 is connected to \(v_i\) for some \(i \leq K\). In the same way, we conclude that vertex 2 is connected to \(v_j\) for some \(j \leq K\). Since \(v_i\) is whp connected to \(v_j\), we conclude that

\[
P_n(\text{dist}_{CM_n}(o_1, o_2) \leq 3) = 1 - o(1). \tag{7.5.7}
\]

This completes the proof. \(\square\)

### 7.5.2 Fluctuation of Distances for Finite-Variance Degrees

We continue to study the fluctuations of the distances in the configuration model, starting with the case where the degrees have finite variance. We need a limit result from branching process theory before we can identify the limiting random variables \((R_n)_{n \in (-1,0]}\).

Recall that \((Z_k)_{k \geq 0}\) denotes the unimodular branching process where in the first generation, the offspring has distribution \(D\) with distribution function \(F\) and in the second and further generations, the offspring has distribution \(D' = 1\), where \(D'\) is the size-biased distribution of \(D\). The process \((Z_k/E[D]^{\nu^{k-1}})_{k \geq 1}\) is a martingale with uniformly bounded expectation and consequently converges almost surely to a limit (see e.g., [Volume 1, Theorem 2.24 and Exercise 2.26]):

\[
\lim_{n \to \infty} \frac{Z_n}{E[D]^{\nu^{n-1}}} = W \quad a.s. \tag{7.5.8}
\]

In the theorem below we need two independent copies \(W_1\) and \(W_1\) of \(W\).

**Theorem 7.20** (Limit law for typical distance in \(CM_n(d)\)) Let \((d_i)_{i \in [n]}\) be a sequence of i.i.d. copies of a random variable \(D\), and assume that there exist \(\tau > 3\) and \(c < \infty\) such that, for all \(x \geq 1\),

\[
[1 - F(x)] \leq cx^{-(\tau-1)}, \tag{7.5.9}
\]

and let \(\nu > 1\). For \(k \geq 1\), let \(a_k = [\log_\nu k] - \log_\nu k \in (-1,0]\). Then, \(CM_n(d)\)
satisfies that there exist random variables \( (R_a)_{a \in (-1,0]} \) such that as \( n \to \infty \) and for all \( k \in \mathbb{Z} \),

\[
P(\text{dist}_{\text{CM}_n(d)}(o_1, o_2) - \lfloor \log_\nu n \rfloor = k | \text{dist}_{\text{CM}_n(d)}(o_1, o_2) < \infty) = P(R_n = k) + o(1).
\]

(7.5.10)

The random variables \( (R_a)_{a \in (-1,0]} \) can be identified as

\[
P(\text{dist}_{\text{CM}_n(d)}(o_1, o_2) - \lfloor \log_\nu n \rfloor = k | \text{dist}_{\text{CM}_n(d)}(o_1, o_2) < \infty) = P(R_a_n = k) + o(1).
\]

(7.5.11)

where \( W_1 \) and \( W_2 \) are independent limit copies of \( W \) in (7.5.8) and where \( \kappa = E[D] (\nu - 1)^{-1} \).

In words, Theorem 7.20 states that for \( \tau > 3 \), the graph distance \( \text{dist}_{\text{CM}_n(d)}(o_1, o_2) \) between two randomly chosen connected vertices grows like the \( \log_\nu n \), where \( n \) is the size of the graph, and that the fluctuations around this mean remain uniformly bounded in \( n \).

The law of \( R_a \) is involved, and can in most cases not be computed exactly. The reason for this is the fact that the random variables \( W \) that appear in its statement are hard to compute explicitly (see also [Volume 1, Chapter 3]).

There are two examples where the law of \( W \) is known. The first example is when all degrees in the graph are equal to some \( r \geq 3 \), and we obtain the \( r \)-regular graph. In this case, \( E[D] = r, \nu = r - 1 \), and \( W = 1 \) a.s. In particular, \( P(\text{dist}_{\text{CM}_n(d)}(o_1, o_2) < \infty) = 1 + o(1) \). Therefore,

\[
P(R_a > k) = \exp\{- \frac{r}{r - 2} (r - 1)^{a+k} \},
\]

(7.5.12)

and \( \text{dist}_{\text{CM}_n(d)}(o_1, o_2) \) is asymptotically equal to \( \log_{r-1} n \). The second example is when \( p^* \) is the probability mass function of a geometric random variable, in which case the branching process with offspring \( p^* \) conditioned to be positive converges to an exponential random variable with parameter 1. This example corresponds to

\[
p_j^* = p(1 - p)^{j-1}, \quad \text{so that} \quad p_j = \frac{1}{jc_p} p(1 - p)^{j-2}, \quad \forall j \geq 1,
\]

(7.5.13)

and \( c_p \) is a normalization constant. For \( p > \frac{1}{2} \), the law of \( W \) has the same law as the sum of \( D_1 \) copies of a random variable \( \mathcal{Y} \), where \( \mathcal{Y} = 0 \) with probability \( \frac{1-p}{p} \) and equal to an exponential random variable with parameter 1 with probability \( \frac{2p-1}{p} \). Even in this simple case, the computation of the exact law of \( R_a \) is non-trivial.

7.5.3 FLUCTUATION OF DISTANCES FOR INFINITE-VARIANCE DEGREES

We next study the fluctuations of typical distances in \( \text{CM}_n(d) \) in the setting where the degrees are i.i.d. and satisfy that there exist \( \tau \in (2,3), \gamma \in [0,1) \) and \( C < \infty \) such that

\[
x^{-\tau+1-C(\log x)^{\gamma-1}} \leq 1 - F(x) \leq x^{-\tau+1+C(\log x)^{\gamma-1}}, \quad \text{for large } x.
\]

(7.5.14)
Theorem 7.21 (Fluctuations graph distance $CM_n(d)$ for infinite-variance degrees) Let $(d_i)_{i \in [n]}$ be a sequence of i.i.d. copies of a random variable $D$. Fix $\tau \in (2,3)$ and assume that (7.5.14) holds. Then, $CM_n(d)$ satisfies that there exist random variables $(R_{a})_{a \in (-1,0]}$ such that, as $n \to \infty$ and for all $l \in \mathbb{Z}$,
\[
P\left( \text{dist}_{CM_n(d)}(o_1, o_2) = 2 \left\lfloor \frac{\log \log n}{\log(\tau - 2)} \right\rfloor + l \mid \text{dist}_{CM_n(d)}(o_1, o_2) < \infty \right) = P(R_{a} = l) + o(1),
\]
where $a_n = \left\lfloor \frac{\log \log n}{\log(\tau - 2)} \right\rfloor - \frac{\log \log n}{\log(\tau - 2)} \in (-1,0]$. Here, the random variables $(R_{a})_{a \in (-1,0]}$ are given by
\[
P(R_{a} > l) = P\left( \min_{x \in \mathbb{Z}} \left[ (\tau - 2)^{-s}Y_1 + (\tau - 2)^{\epsilon - a}Y_2 \right] \leq (\tau - 2)^{1/2}a_{1}Y_1Y_2 > 0 \right),
\]
where $c_{1} = 1$ if $l$ is even, and zero otherwise, and $Y_1, Y_2$ are two independent copies of the limit random variable in Theorem 7.14.

In words, Theorem 7.2 states that for $\tau \in (2,3)$, the graph distance $\text{dist}_{CM_n(d)}(o_1, o_2)$ between two randomly chosen connected vertices grows proportional to $\log \log$ of the size of the graph, and that the fluctuations around this mean remain uniformly bounded in $n$.

We next discuss an extension, obtained by possibly truncating the degree distribution. In order to state the result, we make the following assumption:

Condition 7.22 (Truncated infinite-variance degrees) There exists a $\beta_n \in (0,1/(\tau - 1)]$ such that $F_n(x) = 1$ for $x \geq n^{\beta_n(1+\varepsilon)}$ for all $\varepsilon > 0$, while for all $x \leq n^{\beta_n(1-\varepsilon)}$,
\[
1 - F_n(x) = \frac{L_n(x)}{x^{\tau-1}},
\]
with $\tau \in (2,3)$, and a function $L_n(x)$ that satisfies that, for some constant $C_1 > 0$ and $\eta \in (0,1)$, that
\[
e^{-C_1(\log x)^{\eta}} \leq L_n(x) \leq e^{C_1(\log x)^{\eta}}.
\]

Theorem 7.23 (Fluctuations distances $CM_n(d)$ for truncated infinite-variance degrees) Let $(d_i)_{i \in [n]}$ satisfy Condition 7.22 for some $\tau \in (2,3)$. Assume that $d_{\min} \geq 2$, and that there exists $\kappa > 0$ such that
\[
\max\{d_{TV}(F_n, F), d_{TV}(F^{*}_n, F^{*})\} \leq n^{-\kappa\beta_n}.
\]
When $\beta_n \to 1/(\tau - 1)$, we further require that the limit random variable $Y$ in Theorem 7.14 has no pointmass on $(0, \infty)$. Then,
\[
\text{dist}_{CM_n(d)}(o_1, o_2) - 2 \left\lfloor \frac{\log \log(n^{\beta_n})}{\log(\tau - 2)} \right\rfloor - \frac{1}{\beta_n(\tau - 3)}
\]
is a tight sequence of random variables.
Which of the two terms in (7.5.19) dominates depends sensitively on the choice of $\beta_n$. When $\beta_n \to \beta \in (0, 1/(\tau - 1)]$, the first term dominates. When $\beta_n = (\log n)^{-\gamma}$ for some $\gamma \in (0, 1)$, the second term dominates. Both terms are of the same order of magnitude when $\beta_n = \Theta(1/\log \log n)$, in which case $n^{\beta_n} = \Theta(1)$ and instead Theorem 7.1 applies. Thus, even after the truncation of the degrees, in the infinite-variance case, distances are always ultra-small.

Much more precise results are known, as we explain now. For this, we need some extra notation. Define

$$Y^{(n)}_i = (\tau - 2)^{t_n} \log (Z^{(n, i)}_{t_n})$$

for an appropriate $t_n$, which is the first $t$ for which $\max\{Z^{(n, 1)}_{t_n}, Z^{(n, 2)}_{t_n}\} \geq n^{\rho_n}$ for some appropriate $\rho_n < \beta_n$. As it turns out, $(Y^{(n)}_1, Y^{(n)}_2) \overset{d}{\to} (Y_1, Y_2)$. Further, let $b^{(i)}_n$ be the fractional part of

$$\frac{\log \log (n^{\beta_n}) - \log Y^{(n)}_i}{|\log (\tau - 2)|}.$$ 

Then the most precise result available for graph distances in $CM_n(d)$ for infinite-variance degrees is the following theorem:

**Theorem 7.24** (Fluctuations distances $CM_n(d)$ for truncated infinite-variance degrees) **Assume that the assumptions in Theorem 7.23 hold.** Then,

$$\text{dist}_{CM_n(d)}(o_1, o_2) - 2 \frac{\log \log (n^{\beta_n})}{|\log (\tau - 2)|} - b^{(1)}_n - b^{(2)}_n - \left[ \frac{1}{3} \frac{(\tau - 2)b^{(1)}_n - (\tau - 2)b^{(2)}_n}{3 - \tau} \right] \overset{d}{\to} -1 - \frac{Y_1 Y_2}{|\log (\tau - 2)|}. \quad (7.5.21)$$

Theorem 7.24 not only describes the precise fluctuations of the graph distances in the infinite-variance case, but also identifies how this depends on the precise truncation of the degrees. While the weak limit in Theorem 7.24 looks different from that in Theorem 7.21, it turns out that they are closely related. As Exercise 7.17 shows, Condition 7.22 applies for example in the case of degrees that have an exponential truncation. Exercise 7.18 investigates paths in $CM_n(d)$ where, instead of truncating the degrees, we only allow for paths through vertices of maximal degree $n^{\beta_n}$. This gives insight into the length of paths that must avoid the highest-degree vertices.

### 7.6 Notes and discussion

**Notes on Section 7.1**

Distances in the configuration model were first obtained in a non-rigorous way in Newman et al. (2000b, 2002), see also Cohen and Havlin (2003) for results on ultrasmall distances. Theorem 7.1 is proved by van der Hofstad et al. (2005). Theorem 7.2 is proved by van der Hofstad et al. (2007a).
Notes on Section 7.2
Proposition 7.4 is adapted from (Janson, 2010b, Lemma 5.1). The path counting techniques used in Section 7.2 are novel, but the combinatorial arguments are following Eckhoff et al. (2013), where shortest weighted paths in the complete graph are investigated. Comparisons to branching processes appear in many papers on the configuration model (see, in particular, Bhamidi et al. (2010a), van der Hofstad et al. (2005), van der Hofstad et al. (2007a)). We have strived for a construction that is most transparent and complete. The proof of Theorem 7.9 is close in spirit to the analysis in Reittu and Norros (2004), the only difference being that we have simplified the argument slightly.

Notes on Section 7.3
Theorem 7.14 is proved by Davies (1978), whose proof we follow. A related result, under stronger conditions, appeared by Darling (1970). Branching processes with infinite mean have attracted considerable attention, see e.g., Schuh and Barbour (1977); Seneta (1973) and the references therein. There is a balance between the generality of the results and the conditions on the offspring distribution, and in our opinion Theorem 7.14 strikes a nice balance in that the result is relatively simple and the conditions fairly general.

Notes on Section 7.4
Theorem 7.16 is (Fernholz and Ramachandran, 2007, Theorem 4.1). Theorem 7.17 is proved by Caravenna et al. (2019). A log $n$ lower bound on the diameter when $\tau \in (2, 3)$ is also proved by van der Hofstad et al. (2007b), but is substantially weaker than Theorem 7.16. A sharper result on the diameter of random regular graphs was obtained by Bollobás and Fernandez de la Vega (1982). For a nice discussion and results about the existence of a large $k$-core, we refer to Janson and Luczak (2007).

Notes on Section 7.5
Theorem 7.19 is proved in van den Esker et al. (2006). The explicit identification of $\mathbb{P}(\text{dist}_{\text{CM}}(o_1, o_2) = 2)$ is novel. One might argue that including degrees larger than $n - 1$ is artificial in a network with $n$ vertices. In fact, in many real-world networks, the degree is bounded by a physical constant. Therefore, in van den Esker et al. (2006), also the case where the degrees are conditioned to be smaller than $n^\alpha$ is considered, where $\alpha$ is an arbitrary positive number. Of course, we cannot condition on the degrees to be at most $M$, where $M$ is fixed and independent of $n$, since in this case, the degrees are uniformly bounded, and this case is treated in van der Hofstad et al. (2005) instead. Therefore, van den Esker et al. (2006) considers cases where the degrees are conditioned to be at most a given power of $n$. In this setting, it turns out that the average distance is equal to $k + 3$ with high probability, whenever $\alpha \in (1/(\tau + k), 1/(\tau + k - 1))$. It can be expected that a much more detailed picture can be derived when instead conditioning on the degrees being at most $n^{\beta_n}$ as in van der Hofstad and Komjáthy (2017) for the $\tau \in (2, 3)$ case.
7.7 Exercises for Chapter 7

Theorem 7.20 is proved by van der Hofstad et al. (2005). Theorem 7.21 is proved by van der Hofstad et al. (2007a). The assumption that the limit $Y$ in Theorem 7.14, conditionally on $Y > 0$, has a density was implicitly made by van der Hofstad et al. (2007a). However, this assumption is not necessary for Theorem 7.21. Instead, it suffices to have that $P(0 < Y \leq \varepsilon)$ vanishes as $\varepsilon \searrow 0$ (see e.g., (Hofstad et al., 2007a, Lemma 4.6)), which is generally true.

Theorem 7.23 is proved by van der Hofstad and Komjáthy (2017). There also the assumption on the random variable $Y$ not having point mass in $(0, \infty)$ was explicitly made and discussed in detail. Several extensions have been proved by van der Hofstad and Komjáthy (2017). Theorem 7.24 is (Hofstad and Komjáthy, 2017, Corollary 1.9). We refer to van der Hofstad and Komjáthy (2017) for more details.

7.7 Exercises for Chapter 7

**Exercise 7.1** (Random regular graph) Fix $r \geq 2$ and consider the $r$-regular graph on $n$ vertices, where $nr$ is even. Show that $d_{TV}(p^{\ast(n)}, p^{\ast}) = 0$.

**Exercise 7.2** (Proof Theorem 7.5) Let $o_1, o_2$ be two independent vertices chosen uniformly at random from $[n]$. Use Proposition 7.4 with $a = o_1, b = o_2, \mathcal{I} = [n]$ to prove Theorem 7.5.

**Exercise 7.3** (Proof Theorem 7.8) Use Proposition 7.7 to prove Theorem 7.8 by adapting the proof of Theorem 6.14.

**Exercise 7.4** (\(\Gamma_1\) is a complete graph) Use Lemma 7.11 and $\alpha > 1/2$ to show that, whp, $\Gamma_1$ in (7.2.36) forms a complete graph, i.e., whp, every $i, j \in \Gamma_1$ are direct neighbors in $\text{CM}_n(d)$.

**Exercise 7.5** (Alternative proof Theorem 7.13) Give an alternative proof of Theorem 7.13 by adapting the proof of Theorem 6.9.

**Exercise 7.6** (Example of infinite-mean branching process) Prove that $\gamma(x) = (\log x)^{\gamma - 1}$ for some $\gamma \in [0, 1)$ satisfies the assumptions in Theorem 7.14.

**Exercise 7.7** (Infinite mean under conditions Theorem 7.14) Prove that $E[X] = \infty$ when the conditions in Theorem 7.14 are satisfied. Extend this to show that $E[X^s] = \infty$ for every $s > \alpha \in (0, 1)$.

**Exercise 7.8** (Conditions in Theorem 7.14 for individuals with infinite line of descent) Prove that $p^{\ast(n)}$ in [Volume 1, (3.4.2)] satisfies the conditions in Theorem 7.14 with the function $x \mapsto \gamma^{\ast}(x)$, given by $\gamma^{\ast}(x) = \gamma(x) + c/\log x$.

**Exercise 7.9** (Convergence for $Z_n + 1$) Show that, under the conditions of Theorem 7.14, also $\alpha^s \log(Z_n + 1)$ converges to $Y$ almost surely.

**Exercise 7.10** (Diameter of soup of cycles) Prove that in a graph consisting solely of cycles, the diameter is equal to the longest cycle divided by 2.
Exercise 7.11 (Longest cycle 2-regular graph)  What is the limit law of the size of the longest cycle of the 2-regular graph?

Exercise 7.12 (Parameters for ERₙ(λ/n))  Fix λ > 1. Prove that ν = λ and μ = μₗ, where μₗ ∈ (0, 1) is the dual parameter, i.e., the unique μ satisfying me⁻μ = λe⁻λ.  \( (7.7.1) \)

Exercise 7.13 (Typical distance is at least 2 whp)  Complete the argument that \( P(\text{dist}_{\text{CM}_n(d)}(o_1, o_2) = 1) = o(1) \) in the proof of Theorem 7.19.

Exercise 7.14 (Typical distance equals 2 whp for \( τ = 1 \))  Let the \( (d_i)_{i \in [n]} \) be a sequence of i.i.d. copies of \( D \) with distribution function \( F \) satisfying that \( x \mapsto \lfloor 1 - F(x) \rfloor \) is slowly varying at \( ∞ \). Prove that \( \text{CM}_n(d) \) satisfies that \( \text{dist}_{\text{CM}_n(d)}(o_1, o_2) \xrightarrow{\text{Law}} 2 \).

Exercise 7.15 (Convergence along subsequences van der Hofstad et al. (2005))  Fix an integer \( n_1 \). Prove that, under the assumptions in Theorem 7.20, and conditionally on \( \text{dist}_{\text{CM}_n(d)}(o_1, o_2) < ∞ \), along the subsequence \( n_k = \lfloor n_1 λ^{k-1} \rfloor \), the sequence of random variables \( \text{dist}_{\text{CM}_n(d)}(o_1, o_2) - \lfloor \log_γ n_k \rfloor \) converges in distribution to \( R_{d,n_k} \) as \( k \to ∞ \).

Exercise 7.16 (Tightness of the hopcount van der Hofstad et al. (2005))  Prove that, under the assumptions in Theorem 7.20,

(i) with probability \( 1 - o(1) \) and conditionally on \( \text{dist}_{\text{CM}_n(d)}(o_1, o_2) < ∞ \), the random variable \( \text{dist}_{\text{CM}_n(d)}(o_1, o_2) \) is in between \( (1 + ε) \log_γ n \) for any \( ε > 0 \);

(ii) conditionally on \( \text{dist}_{\text{CM}_n(d)}(o_1, o_2) < ∞ \), the random variables \( \text{dist}_{\text{CM}_n(d)}(o_1, o_2) - \log_γ n \) form a tight sequence, i.e.,

\[
\lim_{K \to ∞} \limsup_{n \to ∞} P( |\text{dist}_{\text{CM}_n(d)}(o_1, o_2) - \log_γ n| \leq K | \text{dist}_{\text{CM}_n(d)}(o_1, o_2) < ∞) = 1.
\]

(7.7.2)

As a consequence, prove that the same result applies to a uniform random graph with degrees \( (d_i)_{i \in [n]} \). Hint: Make use of [Volume 1, Theorem 7.21].

Exercise 7.17 (Exponential truncation for degrees)  Suppose that \( A_n = Θ(n^{β_n}) \). Show that Condition 7.22 holds when

\[
1 - F_n(x) = e^{-x/A_n} x^{β_n - 1},
\]

(7.7.3)

with \( β \in (2, 3) \).

Exercise 7.18 (Paths through vertices with degree constraints)  Suppose that Condition 7.22 holds for \( β_n = 1/(τ - 1) \). Show that Theorem 7.24 can be used to show that minimal number of edges in paths connecting \( o_1 \) and \( o_2 \), for which all the degrees on the path are at most \( n^{α_n} \), where \( α_n \gg 1/\log n \), scales as in (7.5.21) with \( β_n \) in it replaced by \( α_n \).
Chapter 8
SMALL-WORLD PHENOMENA IN PREFERENTIAL ATTACHMENT MODELS

Abstract
In this chapter, we investigate graph distances in preferential attachment models. We focus on typical distances as well as the diameter of preferential attachment models. We again rely on path-counting techniques.

Motivation: local structure versus small-world properties
In Chapters 6–7, we have seen that random graphs with infinite-variance degrees are ultra-small worlds. This can be understood informally by two effects. First, we note that such random graph models contain super-hubs, whose degrees are much larger than $n^{1/2}$ and that form a complete graph of connections, as well as that these super-hubs are often part of shortest paths between two typical vertices; and (b) vertices of large degree $d \gg 1$ are typically connected to vertices of much larger degree, more precisely, of degree roughly $d^{1/(\tau-2)}$. Combined, these two effects mean that it takes roughly $\log \log n / |\log(\tau-2)|$ steps from a typical vertex to reach one of the super-hubs, and thus roughly $2\log \log n / |\log(\tau-2)|$ steps to connect two typical vertices to each other. Of course, the proofs are more technical, but this is the bottom line.

The above explanation depends crucially on the local structure of the generalized random graph as well as the configuration model, in which indeed vertices of high degree are whp connected to vertices of much high degree. In this chapter, we will see that for the preferential attachment model, such consideration need to be subtly adapted. Indeed, fix $\delta \in (-m,0)$ and $m \geq 2$, so that the degree distribution has infinite variance. Then, in preferential attachment models, vertices of large degree $d$ tend to be the old vertices, but old vertices are not necessarily connected to much older vertices, which would be necessary to increase their degree from $d$ to $d^{1/(\tau-2)}$. However, vertices of degree $d \gg 1$ do tend to be connected to typical vertices that in turn tend to be connected to vertices of degree roughly $d^{1/(\tau-2)}$. We conclude that distances seem about twice as high in preferential attachment models as they are in the corresponding generalized random graphs or configuration models, and that this can be explained by the differences in the local connectivity structure. Unfortunately, due to its dynamic nature, the results for preferential attachment models are harder to prove, and they are somewhat less complete.

Throughout this chapter, we work with the preferential attachment model defined in [Volume 1, Section 8.2], which we denote by $(PA_n^{(m,\delta)})_{n \geq 1}$, unless stated otherwise. We recall that this model starts with a single vertex with $m$ self-loops.
at time $t = 1$ and at each time a vertex is added with $m$ edges which are attached to the vertices in the graph with probabilities given in [Volume 1, (8.2.1)] for $m = 1$, and as described on [Volume 1, page 282] for $m \geq 2$. This model can also be obtained by identifying blocks of $m$ vertices in $(\text{PA}_n^{(1,\delta/m)})_{n \geq 1}$. We sometimes also discuss other variants of the model, such as $(\text{PA}_n^{(m,\delta)}(b))_{n \geq 1}$, in which the $m = 1$ model does not have any self-loops, or $(\text{PA}_n^{(m,\delta)}(d))_{n \geq 1}$, which is particularly nice due to its relation to Pólya urn schemes (recall Theorems 5.8 and 5.10).

**Organization of this chapter**

In Section 8.1, we start by showing that preferential attachment trees (where every vertex comes in with one edge to earlier vertices and $m = 1$) have logarithmic height. In Section 8.2 we investigate graph distances in $\text{PA}_n^{(m,\delta)}$ and formulate our main results. In Section 8.3, we investigate path counting techniques in preferential attachment models, which we use in Section 8.4 to prove logarithmic lower bounds on distances, and in Section 8.5 to prove doubly logarithmic lower bounds. In Section 8.6, we prove the matching logarithmic upper bounds on graph distances for $\delta \geq 0$, and in Section 8.7 the matching doubly logarithmic upper bound for $\delta < 0$. In Section 8.8, we discuss the diameter of preferential attachment models for $m \geq 2$. In Section 8.9, we discuss further results about distances in preferential attachment models. We close this chapter with notes and discussion in Section 8.10, and with exercises in Section 8.11.

**Connectivity notation preferential attachment models**

In this chapter, we write that $u \xrightarrow{j} v$ when the $j$th edge from $u$ is connected to $v$, where $u \geq v$. We further write $u \leftrightarrow v$ when there exists a path of edges connecting $u$ and $v$, so that $v \in C(u)$, where $C(u)$ is the connected component of $u$. Finally, we write $u \sim v$ when $\{u, v\}$ is an edge in the preferential attachment model.

8.1 LOGARITHMIC DISTANCES IN PREFERENTIAL ATTACHMENT TREES

In this section, we investigate distances in scale-free trees, arising for $m = 1$. We start by studying the *typical distance from the root* and the *height* of the tree $\text{PA}_n^{(1,\delta)}$, where the height of a tree $T$ is defined as

$$\text{height}(T) = \max_{v \in [n]} \text{dist}_T(1, v),$$

followed by its *typical distances between pairs* and its *diameter*. The conclusion is that the two-vertex objects are twice as large as the one-vertex objects, as can perhaps be expected on trees:

**Theorem 8.1** (Typical distance from root in scale-free trees) Fix $m = 1$ and $\delta > -1$. Let $\theta$ be the non-negative solution of

$$\theta + (1 + \delta)(1 + \log \theta) = 0.$$
Then, as \( n \to \infty \),
\[
\frac{\text{dist}_{PA_n^{(1, \delta)}}(1, o)}{\log n} \xrightarrow{p} \frac{1 + \delta}{2 + \delta}, \quad \frac{\text{height}(PA_n^{(1, \delta)})}{\log n} \xrightarrow{p} \frac{1 + \delta}{(2 + \delta)\theta}. \tag{8.1.3}
\]

**Theorem 8.2** (Typical distance and diameter of scale-free trees) Fix \( m = 1 \) and \( \delta > -1 \). Let \( \theta \) be the non-negative solution of (8.1.2). Then
\[
\frac{\text{dist}_{PA_n^{(1, \delta)}}(1, o_1, o_2)}{\log n} \xrightarrow{p} \frac{2(1 + \delta)}{2 + \delta}, \quad \frac{\text{diam}(PA_n^{(1, \delta)})}{\log n} \xrightarrow{a.s.} \frac{2(1 + \delta)}{(2 + \delta)\theta}. \tag{8.1.4}
\]

The proofs of Theorems 8.1–8.2 rely on the fact that \( PA_n^{(1, \delta)} \) consists of a collection of trees with precisely one self-loop. We shall perform explicit path-counting techniques to prove Theorem 8.2. Exercises 8.1–8.2 explore properties of the constant \( \theta \).

In the proof of Theorems 8.1-8.2 it will be useful to work with \( PA_n^{(1, \delta)}(b) \) instead of \( PA_n^{(1, \delta)} \), for which a similar result holds:

**Theorem 8.3** (Distances in scale-free trees \( PA_n^{(1, \delta)}(b) \)) Fix \( m = 1 \) and \( \delta > -1 \), and let \( \theta \) be the non-negative solution of (8.1.2). Then
\[
\frac{\text{dist}_{PA_n^{(1, \delta)}(b)}}{\log n} \xrightarrow{p} \frac{1 + \delta}{2 + \delta}, \quad \frac{\text{height}(PA_n^{(1, \delta)}(b))}{\log n} \xrightarrow{a.s.} \frac{1 + \delta}{(2 + \delta)\theta}. \tag{8.1.5}
\]
and
\[
\frac{\text{dist}_{PA_n^{(1, \delta)}(b)}}{\log n} \xrightarrow{p} \frac{2(1 + \delta)}{2 + \delta}, \quad \frac{\text{diam}(PA_n^{(1, \delta)}(b))}{\log n} \xrightarrow{a.s.} \frac{2(1 + \delta)}{(2 + \delta)\theta}. \tag{8.1.6}
\]

We start by proving the upper bound in Theorem 8.3. We remark that the proof will indicate that the almost sure limit of the height in Theorem 8.3 does not depend on the precise starting configuration of the graph \( PA_2^{(1, \delta)}(b) \), which will be useful in extending the results in Theorem 8.3 to Theorems 8.1–8.2.

In the proof of the upper bound, we make use of the following result which computes the probability mass function of the distance between vertex \( v \) and the root 1. Before stating the result, we need some more notation. We write \( u \leftrightarrow v = u \overset{1}{\leftarrow} v \) when in \((PA_n^{(1, \delta)}(b))_{n \geq 1}\) the edge of vertex \( u \) is connected to vertex \( v \). Note that for this to happen, we need that \( u > v \). For \( u = \pi_0 > \pi_2 > \cdots > \pi_k = 1 \), and denoting \( \vec{\pi} = (\pi_0, \pi_2, \ldots, \pi_k) \), we write the event that \( \vec{\pi} \) is present in the tree \( PA_n^{(1, \delta)}(b) \) as
\[
\{ \vec{\pi} \subseteq PA_n^{(1, \delta)}(b) \} = \bigcap_{i=0}^{k-1} \{ \pi_i \sim \pi_{i+1} \}. \tag{8.1.7}
\]

For a configuration of \( PA_n^{(1, \delta)}(b) \), we let \( \text{dist}(u, v) \) denote the unique value of \( k \) such that there exists a \( \vec{\pi} \) satisfying \( u = \pi_0 \sim \pi_1 \sim \cdots \sim \pi_{k-1} \sim \pi_k = v \). Then the probability mass function of \( \text{dist}(u, v) \) can be explicitly identified as follows:
Proposition 8.4 (Distribution of \( \text{dist}(u,v) \) in \( \text{PA}_{n}^{(1,\delta)}(b) \)) \hspace{1em} \text{Fix} \ m = 1 \text{ and } \delta > -1. \text{ Then, for all } u > v,

\[
\mathbb{P}(\text{dist}(u,v) = k) = \left( \frac{1 + \delta}{2 + \delta} \right)^{k} \frac{\Gamma(u + \frac{1}{2 + \delta})\Gamma(v)}{\Gamma(v + \frac{1}{2 + \delta})\Gamma(u + 1)} \sum_{i=1}^{k-1} \prod_{\pi} \frac{1}{\pi},
\]

where the sum is over ordered vectors \( \vec{\pi} = (\pi_0, \ldots, \pi_k) \) of length \( k+1 \) with \( \pi_0 = u \) and \( \pi_k = v \). Further,

\[
\mathbb{P}(\vec{\pi} \subseteq \text{PA}_{n}^{(1,\delta)}(b)) = \left( \frac{1 + \delta}{2 + \delta} \right)^{k} \frac{\Gamma(u + \frac{1}{2 + \delta})\Gamma(v)}{\Gamma(v + \frac{1}{2 + \delta})\Gamma(u + 1)} \prod_{\pi} \frac{1}{\pi},
\]

Proof of Proposition 8.4. Since the path between vertex \( u \) and \( v \) is unique, we obtain

\[
\mathbb{P}(\text{dist}(u,v) = k) = \sum_{\vec{\pi}} \mathbb{P}\left( \bigcap_{i=0}^{k-1} \{ \pi_i \Rightarrow \pi_{i+1} \} \right),
\]

where again the sum is over all ordered vectors \( \vec{\pi} = (\pi_0, \ldots, \pi_k) \) of length \( k+1 \) with \( \pi_0 = u \) and \( \pi_k = v \). Therefore, (8.1.8) follows immediately from (8.1.9).

We claim that the events \( \{ \pi_i \Rightarrow \pi_{i+1} \} \) are independent, i.e., for every sequence \( \vec{\pi} = (\pi_0, \ldots, \pi_k) \)

\[
\mathbb{P}\left( \bigcap_{i=0}^{k-1} \{ \pi_i \Rightarrow \pi_{i+1} \} \right) = \prod_{i=0}^{k-1} \mathbb{P}(\pi_i \Rightarrow \pi_{i+1}).
\]

We prove the independence in (8.1.11) by induction on \( k \geq 1 \). For \( k = 1 \), there is nothing to prove, and this initializes the induction hypothesis. To advance the induction hypothesis in (8.1.11), we condition on \( \text{PA}_{n-1}^{(1,\delta)}(b) \) to obtain

\[
\mathbb{P}(\vec{\pi} \subseteq \text{PA}_{n}^{(1,\delta)}(b)) = \mathbb{E}\left[ \mathbb{P}\left( \bigcap_{i=0}^{k-1} \{ \pi_i \Rightarrow \pi_{i+1} \} \left| \text{PA}_{n-1}^{(1,\delta)}(b) \right. \right) \right]
\]

\[
= \mathbb{E}\left[ \mathbb{E}\left[ \bigcap_{i=0}^{k-1} \{ \pi_i \Rightarrow \pi_{i+1} \} \left| \pi_0 \Rightarrow \pi_1 \left| \text{PA}_{n-1}^{(1,\delta)}(b) \right. \right. \right) \right],
\]

since the event \( \bigcap_{i=0}^{k-1} \{ \pi_i \Rightarrow \pi_{i+1} \} \) is measurable with respect to \( \text{PA}_{n}^{(1,\delta)}(b) \) because \( \pi_0 - 1 \geq \pi_i \) for all \( i \in [k-1] \). Furthermore, from [Volume 1, (8.2.2)],

\[
\mathbb{P}(\pi_0 \Rightarrow \pi_1 \left| \text{PA}_{n-1}^{(1,\delta)}(b) \right. \right) = \frac{D_{\delta}(\pi_0 - 1) + \delta}{(2 + \delta)(\pi_0 - 1)}.
\]

In particular,

\[
\mathbb{P}(\pi_0 \Rightarrow \pi_1) = \mathbb{E}\left[ \frac{D_{\delta}(\pi_0 - 1) + \delta}{(2 + \delta)(\pi_0 - 1)} \right].
\]

Therefore,
\[
\mathbb{P}(\vec{\pi} \subseteq \text{PA}^{(1,\delta)}_n(b)) = \mathbb{E}\left[\prod_{i=1}^{k-1} \frac{D_{\pi_i}(\pi_i - 1) + \delta}{(2 + \delta)(\pi_i - 1)}\right]
\]
\[
= \mathbb{P}\left(\bigcap_{i=1}^{k-1} \{\pi_i \sim \pi_{i+1}\}\right) \mathbb{E}\left[\frac{D_{\pi_0}(\pi_0 - 1) + \delta}{(2 + \delta)(\pi_0 - 1)}\right],
\]  
(8.1.15)
since the random variable \(D_{\pi_1}(\pi_0 - 1)\) only depends on how many edges are connected to \(\pi_1\) after time \(\pi_1\), and is thus independent of the event \(\bigcap_{i=1}^{k-1} \{\pi_i \sim \pi_{i+1}\}\), which only depends on the attachment of the edges up to and including time \(\pi_1\). We conclude that
\[
\mathbb{P}(\vec{\pi} \subseteq \text{PA}^{(1,\delta)}_n(b)) = \mathbb{P}(\pi_0 \sim \pi_1)^{k-1} \prod_{i=1}^{k-1} \mathbb{E}\left[\frac{D_{\pi_i}(\pi_i - 1) + \delta}{(2 + \delta)(\pi_i - 1)}\right].
\]  
(8.1.16)
The claim in (8.1.11) for \(k\) follows from the induction hypothesis. Combining (8.1.10) with (8.1.11) and (8.1.14), we obtain that
\[
\mathbb{P}(\vec{\pi} \subseteq \text{PA}^{(1,\delta)}_n(b)) = \prod_{i=0}^{k-1} \mathbb{E}\left[\frac{D_{\pi_{i+1}}(\pi_i + 1) + \delta}{(2 + \delta)(\pi_i - 1)}\right].
\]  
(8.1.17)
By [Volume 1, (8.11.3)], for all \(n \geq s\),
\[
\mathbb{E}\left[\frac{D_s(n) + \delta}{(2 + \delta)n}\right] = (1 + \delta) \frac{\Gamma(s + \frac{1}{2+\delta})\Gamma(s)}{(2 + \delta)s\Gamma(s + \frac{1}{2+\delta})}
\]  
\[
= \frac{1 + \delta}{2 + \delta} \frac{\Gamma(s + \frac{1}{2+\delta})\Gamma(s + \frac{1}{2+\delta})}{\Gamma(s + 1)\Gamma(s + \frac{1}{2+\delta})},
\]  
(8.1.18)
so that
\[
\mathbb{P}(\vec{\pi} \subseteq \text{PA}^{(1,\delta)}_n(b))
\]
\[
= \left(\frac{1 + \delta}{2 + \delta}\right)^k \prod_{i=0}^{k-1} \frac{\Gamma(\pi_i + \frac{1}{2+\delta})\Gamma(\pi_{i+1} + \frac{1}{2+\delta})}{\Gamma(\pi_i + 1)\Gamma(\pi_{i+1} + \frac{1}{2+\delta})}
\]
\[
= \left(\frac{1 + \delta}{2 + \delta}\right)^k \frac{\Gamma(\pi_0 + \frac{1}{2+\delta})\Gamma(v)\prod_{i=1}^{k-1} \frac{1}{\pi_i}}{\Gamma(\pi_k + \frac{1}{2+\delta})\Gamma(\pi_0)}\prod_{i=1}^{k-1} \frac{1}{\pi_i},
\]  
(8.1.19)
This completes the proof of Proposition 8.4.

**Proof of the upper bounds in Theorem 8.3.** We first use Proposition 8.3 to prove that, for \(\omega\) chosen uniformly at random from \([n]\),
\[
\frac{\text{dist}(1,\omega)}{\log n} \leq (1 + \omega(1)) \frac{1 + \delta}{2 + \delta}.
\]  
(8.1.21)
and, almost surely for large enough $n$,}
\[
\frac{\text{dist}(1,n)}{\log n} \leq (1 + \varepsilon) \frac{(1 + \delta)}{(2 + \delta)\theta},
\]  
(8.1.22)

where $\theta$ is the non-negative solution of (8.1.2). We start by noting that (8.1.21)–(8.1.22) immediately prove (8.1.5). Further, (8.1.22) implies that, almost surely for large enough $n$,

\[
\text{height}(\text{PA}_n^{(1,\delta)}(b)) \leq (1 + \varepsilon) \frac{(1 + \delta)}{(2 + \delta)\theta}.
\]  
(8.1.23)

By the triangle inequality,
\[
\text{dist}_{\text{PA}_n^{(1,\delta)}}(o_1, o_2) \leq \text{dist}(1, o_1) + \text{dist}(1, o_2),
\]  
(8.1.24)
\[
\text{diam}(\text{PA}_n^{(1,\delta)}(b)) \leq 2 \text{height}(\text{PA}_n^{(1,\delta)}(b)),
\]  
(8.1.25)

so that (8.1.21)–(8.1.22) imply the upper bounds in Theorem 8.3.

We proceed to prove (8.1.21) and (8.1.22), and start with some preparations. We use (8.1.8) and symmetry to obtain
\[
\mathbb{P}(\text{dist}(1,n) = k) = \left(1 + \frac{\delta}{2 + \delta}\right)^k \frac{\Gamma(n + \frac{1}{2 + \delta})}{\Gamma(1 + \frac{1}{2 + \delta})\Gamma(n + 1)} \sum_{\vec{t}_{k-1}}^{k-1} \frac{1}{(k-1)!} \prod_{i=1}^{k-1} \frac{1}{t_i},
\]  
(8.1.26)

where the sum now is over all vectors $\vec{t}_{k-1} = (t_1, \ldots, t_{k-1})$ with $1 < t_i < n$ with distinct coordinates. We can upper bound this sum by leaving out the restriction that the coordinates of $\vec{t}_{k-1}$ are distinct, so that
\[
\mathbb{P}(\text{dist}(1,n) = k) \leq \left(1 + \frac{\delta}{2 + \delta}\right)^k \frac{\Gamma(n + \frac{1}{2 + \delta})}{\Gamma(1 + \frac{1}{2 + \delta})\Gamma(n + 1)} \frac{1}{(k-1)!} \left(\sum_{s=2}^{n-1} \frac{1}{s}\right)^{k-1}.
\]  
(8.1.27)

Since $x \mapsto 1/x$ is monotonically decreasing
\[
\sum_{s=2}^{n-1} \frac{1}{s} \leq \int_1^n \frac{1}{x} dx = \log n.
\]  
(8.1.28)

Also, we use [Volume 1, (8.3.9)] to bound, for some constant $K_\delta > 0$,
\[
\mathbb{P}\left(\text{dist}(1,n) = k\right) \leq K_\delta n^{-\frac{1 + \delta}{2 + \delta}} \frac{(1 + \frac{\delta}{2 + \delta} \log n)^{k-1}}{(k-1)!} = K_\delta \mathbb{P}\left(\text{Poi}\left(\frac{1 + \delta}{2 + \delta} \log n\right) = k - 1\right).
\]  
(8.1.29)

Now we are ready to prove (8.1.21). We note that $o$ is chosen uniformly in $[n]$,.
8.1 Logarithmic distances in preferential attachment trees

so that, with \( C \) denoting a generic constant that may change from line to line,

\[
P(\text{dist}(1, o) = k) = \frac{1}{n} \sum_{s=1}^{n} P(\text{dist}(v_s, v_1) = k) \leq \frac{1}{n} \sum_{s=1}^{n} K_\delta s^{-\frac{s}{2+\delta}} \left( \frac{1+\delta}{2+\delta} \log n \right)^{k-1} (k-1)!
\]

\[
\leq \left( \frac{1+\delta}{2+\delta} \log n \right)^{k-1} \sum_{s=1}^{n} K_\delta s^{-\frac{s}{2+\delta}} \leq C \left( \frac{1+\delta}{2+\delta} \log n \right)^{k-1} (k-1)!
\]

\[
= CP \left( \text{Poi} \left( \frac{1+\delta}{2+\delta} \log n \right) = k-1 \right).
\]  

(8.1.30)

Therefore,

\[
P(\text{dist}(1, o) > k) \leq CP \left( \text{Poi} \left( \frac{1+\delta}{2+\delta} \log n \right) \geq k \right).
\]  

(8.1.31)

Now we fix \( \varepsilon > 0 \) and take \( k_n = \frac{(1+\varepsilon)(1+\delta)}{2+\delta} \log n \), to arrive at

\[
P(\text{dist}(1, o) > k_n) \leq CP \left( \text{Poi} \left( \frac{1+\delta}{2+\delta} \log n \right) \geq \frac{(1+\varepsilon)(1+\delta)}{2+\delta} \log n \right)
\]

\[
= o(1),
\]

by the law of large numbers and for any \( \varepsilon > 0 \), as required.

We continue to prove (8.1.22). By (8.1.29),

\[
P(\text{dist}(1, o) > k_n) \leq K_\delta P \left( \text{Poi} \left( \frac{1+\delta}{2+\delta} \log n \right) \geq \frac{(1+\varepsilon)(1+\delta)}{2+\delta} \log n \right)
\]

\[
(8.1.32)
\]

Let \( x \) be the solution of

\[
x \left( \log \left( x(2+\delta)/(1+\delta) \right) - 1 \right) + \frac{1+\delta}{2+\delta} = 1,
\]

(8.1.35)

so that \( x = \frac{(1+\delta)}{(2+\delta)p} \). Then, for every \( a > x \),

\[
P(\text{dist}(1, n) > a \log n) = O(n^{-p}),
\]

(8.1.36)

where

\[
p = [a \log \left( a(2+\delta)/(1+\delta) \right) - 1] + \left( 1 + \delta \right) \left( 2 + \delta \right) > 1.
\]

As a result, by the Borel-Cantelli Lemma, the event \{dist(1, n) > k_n\} occurs only finitely often, and we conclude that (8.1.22) holds.
Proof of the lower bound on \( \text{dist}_{PA_n(t, \delta)}(1, \theta) \) in Theorem 8.3. We use (8.1.30) to obtain that
\[
P(\text{dist}_{PA_n(t, \delta)}(1, \theta) \leq k) \leq C \mathbb{P}\left( \text{Poi}\left( \frac{1 + \delta}{2 + \theta} \log n \right) \leq k \right). \tag{8.1.38}
\]
Fix \( k_n = \frac{(1 + \delta)(1 - \epsilon)}{(2 + \delta)} \log n \), and note that \( P(\text{dist}_{PA_n(t, \delta)}(1, \theta) \leq k_n) = o(1) \) by the law of large numbers.

To complete the proof of \( \text{height}(PA_n(t, \delta))(b) \) converges in probability, we use the second moment method to prove that \( \text{height}(PA_n(t, \delta))(b) \leq \frac{(1 + \delta)(1 - \epsilon)}{(2 + \delta)\theta} \log n \) has vanishing probability. Together with (8.1.23), this certainly proves that
\[
\text{height}(PA_n(t, \delta))(b) / \log n \xrightarrow{\mathbb{P}} \frac{(1 + \delta)(1 - \epsilon)}{(2 + \delta)\theta}.
\]
However, since \( \text{height}(PA_n(t, \delta))(b) \) is a non-decreasing sequence of random variables, this also implies that this convergence even holds almost surely, as we argue in more detail below Proposition 8.5.

We formalize the statement that \( \text{height}(PA_n(t, \delta))(b) \leq \frac{(1 - \epsilon)(1 + \delta)}{(2 + \delta)\theta} \log n \) in the following proposition:

**Proposition 8.5** (Height of \( PA_n(t, \delta)(b) \) converges in probability) For every \( \epsilon > 0 \) there exists a \( \eta = \eta(\epsilon) > 0 \) such that
\[
P\left( \text{height}(PA_n(t, \delta))(b) \leq \frac{(1 - \epsilon)(1 + \delta)}{(2 + \delta)\theta} \log n \right) \leq O(n^{-\eta}). \tag{8.1.39}
\]

Proof of lower bound on \( \text{height}(PA_n(t, \delta))(b) \) in Theorem 8.3 subject to Proposition 8.5. Fix \( \alpha > 0 \), and take \( t_k = t_k(\alpha) = e^{\alpha k} \). For any \( \alpha > 0 \), by Proposition 8.5 and the fact that \( t_k^{-\eta} \) is summable, almost surely, \( \text{height}(PA_n(t, \delta))(b) \geq \frac{(1 - \epsilon)(1 + \delta)}{(2 + \delta)\theta} \log t_k \). This proves the almost sure lower bound on \( \text{height}(PA_n(t, \delta))(b) \) along the subsequence \( (t_k)_{k \geq 0} \). To extend this to an almost sure lower bound when \( n \to \infty \), we use that \( t \mapsto \text{height}(PA_n(t, \delta))(b) \) is non-decreasing, so that, for every \( t \in [t_{k-1}, t_k] \),
\[
\text{height}(PA_n(t, \delta))(b) \geq \text{height}(PA_{t_{k-1}}(t, \delta))(b) \tag{8.1.40}
\]
\[
\geq \frac{(1 - \epsilon)(1 + \delta)}{(2 + \delta)\theta} \log t_{k-1}
\]
\[
\geq (1 - \epsilon)(1 - \alpha) \frac{(1 + \delta)}{(2 + \delta)\theta} \log n,
\]
where the third inequality follows from the almost sure lower bound on \( \text{height}(PA_{t_{k-1}}(t, \delta))(b) \).

The above bound holds for all \( \epsilon,\alpha > 0 \), so that letting \( \epsilon,\alpha \downarrow 0 \) proves our claim.

Proof of Proposition 8.5. We perform a path counting argument. We fix \( T \in [n] \) and \( k \in \mathbb{N} \). Recall that a path \( \pi = (\pi_0, \ldots, \pi_k) \) is a sequence of vertices. In this section we assume that \( \pi_i > \pi_{i+1} \), since our paths will be part of the scale-free
Using [Volume 1, (8.3.9)], we therefore arrive at

\[ N_k(n) = \#\{\pi \subseteq PA_n^{(1,\delta)}(b) : \pi_k \in [T]\} \quad (8.1.41) \]

denote the number of \( k \)-step paths in \( PA_n^{(1,\delta)}(b) \) with an endpoint in \([T]\). By Proposition 8.4,

\[
E[N_k(n)] = \left(\frac{1+\delta}{2+\delta}\right)^k \sum_{\pi_k=1}^{n} \sum_{\pi_0=\pi_k}^{n} \frac{\Gamma(\pi_0 + \frac{1}{2+\delta})\Gamma(\pi_k)}{\Gamma(\pi_k + \frac{1}{2+\delta})\Gamma(\pi_0 + 1)} \sum_{i=1}^{k-1} \frac{1}{\pi_i} \quad (8.1.42)
\]

\[
\geq \left(\frac{1+\delta}{2+\delta}\right)^k \sum_{\pi_k=1}^{n} \frac{\Gamma(\pi_k)}{\Gamma(\pi_k + \frac{1}{2+\delta})} \sum_{\pi_0=\pi_k}^{n} \frac{\Gamma(\pi_0 + \frac{1}{2+\delta})}{\Gamma(\pi_0 + 1)} \frac{1}{(k-1)!} \sum_{i=1}^{k-1} \frac{1}{t_i},
\]

where again the sum is over all ordered \( \vec{\pi} = (\pi_0, \ldots, \pi_k) \) with \( \pi_k \in [T] \) and \( \pi_0 \in [n] \). We can bound this from below by

\[
E[N_k(n)] \geq \left(\frac{1+\delta}{2+\delta}\right)^k \sum_{\pi_k=1}^{T} \frac{\Gamma(\pi_k)}{\Gamma(\pi_k + \frac{1}{2+\delta})} \sum_{\pi_0=\pi_k}^{n} \frac{\Gamma(\pi_0 + \frac{1}{2+\delta})}{\Gamma(\pi_0 + 1)} \frac{1}{(k-1)!} \sum_{i=1}^{k-1} \frac{1}{t_i},
\]

(8.1.43)

where now the sum is over all vectors \( \vec{t}_k = (t_1, \ldots, t_{k-1}) \) with distinct coordinates with \( t_i \in [\pi_k + 1, \pi_0 - 1] \). For fixed \( \pi_0, \pi_k, \)

\[
\sum_{i=1}^{k-1} \frac{1}{t_i} \geq \left(\frac{\pi_0-1}{s}\right)^{k-1},
\]

(8.1.44)

We can lower bound

\[
\sum_{s=\pi_k+1}^{\pi_0} \frac{1}{s} \geq \int_{\pi_k+1}^{\pi_0} \frac{1}{x} \, dx = \log(\pi_0/(\pi_k + k)) \geq (1 - \varepsilon) \log n,
\]

(8.1.45)

when \( \pi_0 \geq n/2, \pi_k \leq T \) and \( \log[2(T + k)] \leq \varepsilon \log n \). Thus, we conclude that

\[
E[N_k(n)] \geq \left(\frac{1+\delta}{2+\delta}\right)^k \sum_{\pi_k=1}^{T} \frac{\Gamma(\pi_k)}{\Gamma(\pi_k + \frac{1}{2+\delta})} \sum_{\pi_0=\pi_k}^{n} \frac{\Gamma(\pi_0 + \frac{1}{2+\delta})}{\Gamma(\pi_0 + 1)} \frac{1}{(k-1)!} \left[1 - \varepsilon \right] \log n]^{k-1}.
\]

(8.1.46)

Using [Volume 1, (8.3.9)], we therefore arrive at

\[
E[N_k(n)] \geq (1 + o(1)) \left(\frac{1+\delta}{2+\delta}\right)^k \sum_{\pi_k=1}^{T} \sum_{\pi_0=\pi_k}^{n} \pi_0^{-1/(2+\delta)} \sum_{\pi_0=\pi_k}^{n} \pi_0^{-1/(2+\delta)} \frac{1}{(k-1)!} \left[1 - \varepsilon \right] \log n]^{k-1}
\]

\[
\geq c \left(\frac{1+\delta}{2+\delta}\right)^k T^{1/(2+\delta)} n^{1/(2+\delta)} \frac{1}{(k-1)!} \left[1 - \varepsilon \right] \log n]^{k-1}
\]

\[
= c T^{1/(2+\delta)} n^{1/(1-\varepsilon)(1+\delta)/(2+\delta)} \mathbb{P}\left(Poi\left(\frac{1+\delta}{2+\delta}(1 - \varepsilon) \log n\right) = k - 1\right).
\]

(8.1.47)
When we take \( k = k_n = \frac{(1+\varepsilon)(1+\delta)}{2+\delta} \log n \), then
\[
\Pr\left( \text{Poi}\left( \frac{1+\delta}{2+\delta} (1-\varepsilon) \log n \right) = k - 1 \right) \geq cn^{-1/(1-\varepsilon)}/\sqrt{\log n},
\]
so that
\[
E[N_k(n)] \geq T^{(1+\delta)/(2+\delta) + \epsilon/(2+\delta) + o(1)}.
\]
We next take \( T = n^\varepsilon \) to arrive at
\[
E[N_k(n)] \geq n^\varepsilon + o(1).
\]
This provides the required lower bound on \( E[N_k(n)] \). We defer the proof of an upper bound on \( \text{Var}(N_k(n)) \) to Section 8.3, where we prove that there exist \( C > 0 \) and \( \eta = \eta(\varepsilon) > 0 \) such that \( \text{Var}(N_k(n)) \leq C E[N_k(n)]^2 n^{-\eta} \) (see Lemma 8.12 below). As a result, by the Chebychev inequality ([Volume 1, Theorem 2.18])
\[
\Pr(N_k(n) = 0) \leq \frac{\text{Var}(N_k(n))}{E[N_k(n)]^2} \leq Cn^{-\eta}.
\]
Since \( \text{height}(PA_n^{(1,\delta)}(b)) \geq k \) when \( N_k(n) \geq 1 \), this proves the claim in Proposition 8.5.

We complete this section by proving Theorems 8.2–8.3:

**Proof of Theorems 8.2 and 8.3.** We first prove the upper bound on the diameter of \( PA_n^{(1,\delta)}(b) \) in Theorem 8.3, for which we use that
\[
\text{diam}(PA_n^{(1,\delta)}(b)) \leq 2 \cdot \text{height}(PA_n^{(1,\delta)}(b)).
\]
Equation (8.1.52) together with the upper bound in Theorem 8.3 imply that
\[
\limsup_{n \to \infty} \frac{\text{diam}(PA_n^{(1,\delta)}(b))}{\log n} \leq \frac{2(1+\delta)}{(2+\delta)\theta}.
\]
For the lower bound, we use the lower bound on \( \text{diam}(PA_n^{(1,\delta)}(b)) \) in Theorem 8.3 and the decomposition of scale-free trees in Theorem 5.4. Theorem 5.4 states that the scale-free tree \( PA_n^{(1,\delta)}(b) \) can be decomposed into two scale-free trees, having a similar distribution as copies \( PA_{S_1(n)}(1, \delta) \) and \( PA_{n-S_1(n)}(b2) \), where \( (PA_{n-S_1(n)}(b1))_{n \geq 1} \) and \( (PA_{n}(b2))_{n \geq 1} \) are independent scale-free tree processes, and the law of \( S_1(t) \) is described in (5.1.29). By this tree decomposition
\[
\text{diam}(PA_n^{(1,\delta)}(b)) \geq \text{height}(PA_{S_1(n)}^{(1,\delta)}(b1)) + \text{height}(PA_{n-S_1(n)}^{(1,\delta)}(b2)).
\]
The two trees \( (PA_{n-S_1(n)}^{(1,\delta)}(b1))_{n \geq 1} \) and \( (PA_{n}^{(1,\delta)}(b2))_{n \geq 1} \) are not exactly equal in distribution to \( (PA_{n-S_1(n)}^{(1,\delta)}(b1))_{n \geq 1} \), because the initial degree of the starting vertices at time \( n = 2 \) is different. However, the precise almost sure scaling in Theorem 5.4 does not depend in a sensitive way on \( d_1 \) and \( d_2 \), and also the height of the scale-free tree in Theorem 8.3 does not depend on the starting graphs \( PA_1^{(1,\delta)}(b1) \) and \( PA_2^{(1,\delta)}(b1) \) (see the remark below Theorem 8.3). Since \( S_1(n)/n \xrightarrow{a.s.} U \), with \( U \) having a Beta-distribution with parameters \( a = (3+\delta)/(2+\delta) \) and \( b = (1+\delta)/(2+\delta) \), we obtain that \( \text{height}(PA_{S_1(n)}^{(1,\delta)}(b1))/\log n \xrightarrow{a.s.} (1+\delta)/(2+\delta) \) and
8.2 Small-world phenomena in preferential attachment models

height \( \left( \text{PA}_{n-S_t(n)}(b2) \right) / \log n \xrightarrow{a.s.} \frac{(1+\delta)}{(2+\delta)} \). Thus, we conclude that, almost surely for large enough \( n \),

\[
\frac{\text{diam}(\text{PA}_{n}(b))}{\log n} \geq (1 - \varepsilon) \frac{2(1+\delta)}{(2+\delta)^2}.
\]

(8.1.55)

Combining (8.1.53) and (8.1.55) proves the convergence of \( \text{diam}(\text{PA}_{n}(b)) / \log n \) in Theorem 8.3.

We proceed with the proof of the convergence of \( \text{dist}(\text{PA}_{n}(b)) / \log n \) in Theorem 8.3. We write

\[
\text{dist}(\text{PA}_{n}(b)) = \text{dist}(\text{PA}_{n}(1), 1) + \text{dist}(\text{PA}_{n}(1), 0)
\]

− \( \text{dist}(\text{PA}_{n}(1), V) \),

where \( V \) is the last vertex that is on both paths from \( 1 \rightarrow o_1 \) as well as from \( 1 \rightarrow o_2 \). The asymptotics of the first two terms are identified in (8.1.5) in Theorem 8.3, so that it suffices to show that \( \text{dist}(\text{PA}_{n}(1), V) = o_p(\log n) \). We defer this result to Lemma 8.13 below, after we have discussed path-counting techniques for preferential attachment models.

To prove Theorem 8.2, we note that the connected components of \( \text{PA}_{n}(b) \) are similar in distribution to single scale-free tree \( \text{PA}_{t_1}(b1), \ldots, \text{PA}_{t_N}(bN_n) \), apart from the initial degree of the root. Here \( t_i \) denotes the size of the \( i \)th tree at time \( n \), and we recall that \( N_n \) denotes the total number of trees at time \( n \). Since \( N_n / \log n \xrightarrow{d} (1+\delta)/(2+\delta) \) (recall Exercise 5.23), whp the largest connected component has size at least \( \varepsilon n / \log n \).

\[
\log (\varepsilon n / \log n) = \log n(1 + o(1)),
\]

(8.1.57)

Theorem 8.2 follows along the same lines as in the proof of Theorem 8.3.

8.2 Small-world phenomena in preferential attachment models

In the next sections we investigate distances in preferential attachment models for \( m \geq 2 \). These results are not as complete as those for inhomogeneous random graphs or the configuration model as discussed in Chapters 6 and 7, respectively. This is partly due to the fact that the dynamic preferential attachment models are substantially harder to analyze than these static models. We investigate both the diameter as well as typical distances.

By Theorem 5.26, \( \text{PA}_{n}^{(m,b)} \) is whp connected when \( m \geq 2 \). Recall that in a connected graph, the typical distance \( \text{dist}^{(m,b)}(o_1, o_2) \) is the graph distance between two vertices chosen uniformly at random from the vertex set \([n]\). Recall further that the power-law degree exponent for \( \text{PA}_{n}^{(m,b)} \) is equal to \( \tau = 3 + \delta/m \). Therefore, \( \tau > 3 \) precisely when \( \delta > 0 \). For the generalized random graph and the configuration model, we have seen that distances are logarithmic in the size of the graph when \( \tau > 3 \), and doubly logarithmic when \( \tau \in (2,3) \). We will see similar results for \( \text{PA}_{n}^{(m,b)} \).
Logarithmic distances in PA models with $m \geq 2$ and $\delta > 0$.

We start by investigating the case where $\delta > 0$ so that also the power-law degree exponent $\tau = 3 + \delta/m$ satisfies $\tau > 3$. In this case, the typical distance is logarithmic in the size of the graph:

**Theorem 8.6** (Logarithmic bounds for typical distances of $\text{PA}^{(m,\delta)}_n$ for $\delta > 0$) Fix $m \geq 2$ and $\delta > 0$. There exists $\nu \in (1, \infty)$ such that, as $n \to \infty$,  
\[
\frac{\text{dist}_{\text{PA}^{(m,\delta)}_n}(o_1, o_2)}{\log(n)} \xrightarrow{p} 1.
\]

(8.2.1)

We will see that $\nu$ again refers to the exponential growth of an appropriate multi-type branching process, given by the Pólya point tree.

Distances in PA models with $m \geq 2$ and $\delta = 0$.

For $\delta = 0$, $\tau = 3$. For $\text{NR}_n(w)$, distances grow as $\log(n)/\log \log n$ in this case (recall Theorem 6.17). The same turns out to be true for $\text{PA}^{(m,\delta)}_n$:

**Theorem 8.7** (Typical distances of $\text{PA}^{(m,\delta)}_n$ for $\delta = 0$) Fix $m \geq 2$ and $\delta = 0$. As $n \to \infty$,  
\[
\frac{\text{dist}_{\text{PA}^{(m,\delta)}_n}(o_1, o_2) \log \log n}{\log n} \xrightarrow{p} 1.
\]

(8.2.2)

Theorem 8.7 shows that distances for $\tau = 3$ are similar in $\text{PA}^{(m,\delta)}_n$ as in $\text{NR}_n(w)$.

Doubly logarithmic distances in PA models with $m \geq 2$ and $\delta < 0$.

We close this section by discussing the case where $\delta \in (-m, 0)$, so that $\tau \in (2, 3)$. In this case, it turns out that distances again grow doubly logarithmically in the size of the graph:

**Theorem 8.8** (Doubly logarithmic asymptotics for the diameter for $\delta < 0$) Fix $m \geq 2$ and assume that $\delta \in (-m, 0)$. As $n \to \infty$,  
\[
\frac{\text{dist}_{\text{PA}^{(m,\delta)}_n}(o_1, o_2) \log \log n}{\log n} \xrightarrow{p} \frac{4}{|\log(\tau - 2)|}.
\]

(8.2.3)

Interestingly, the term $4/|\log(\tau - 2)|$ appearing in Theorem 8.8 replaces the term $2/|\log(\tau - 2)|$ in Theorem 6.3 for the Norros-Reittu model $\text{NR}_n(w)$ and in Theorem 7.2 for the configuration model $\text{CM}_n(d)$ when the power-law exponent $\tau$ satisfies $\tau \in (2, 3)$. Thus, typical distances are twice as big for $\text{PA}^{(m,\delta)}_n$ compared to $\text{CM}_n(d)$ with the same power-law exponent. This can be intuitively explained as follows. For the configuration model $\text{CM}_n(d)$, vertices with high degrees are likely to be directly connected (see e.g. Lemma 7.11). For $\text{PA}^{(m,\delta)}_n$, this is not the case. However, vertices with high degrees are likely to be at distance two, as whp there is a young vertex that will connect to both of the old vertices of high degree. This makes distances in $\text{PA}^{(m,\delta)}_n$ effectively twice as big as those for $\text{CM}_n(d)$ with the same degree sequence. This effect is special for $\delta < 0$ and is studied in more detail in Exercise 8.3.
Universality in distances for scale-free graphs
The available results are all consistent with the prediction that distances in preferential attachment models have the same asymptotics as distances in the configuration model with the same degree sequence. This suggest a strong form of universality, which is interesting in its own right. However, certain local effects of the graph may change graph distances a little bit, as exemplified by the fact that distances in Theorem 8.8 are twice as big as for the Norros-Reittu model $\text{NR}_n(w)$ in Theorem 6.3, and for the configuration model $\text{CM}_n(d)$ in Theorem 7.2. This shows that the details of the model are relevant.

Organisation of the proof of small-world properties of the PAM
We prove these results in the following four sections. This proof is organized as follows. We start in Section 8.3 by discussing path-counting techniques for preferential attachment models. While preferential attachment models lack the kind of independence or weak dependence between the edge statuses of the graph present in inhomogeneous random graphs and configuration models, it turns out that the probability of the existence of paths can still be bounded from above by products of probabilities, due to an inherent negative correlation between edge statuses. This allows us to still obtain upper bounds for the expected number of paths of given lengths connecting several vertices. Therefore, the lower bounds on typical graph distances can be performed in a highly similar way as for rank-1 inhomogeneous random graphs as in Theorem 6.4 or for the configuration model in Theorem 7.5. These bounds will also apply to $\delta = 0$. For $\delta < 0$, instead, we again use a truncated path-counting argument alike in the proof of Theorems 6.6 and 7.6. The resulting lower bounds are performed in Sections 8.4 and 8.5 for $\delta \geq 0$ and $\delta < 0$ respectively.

For the upper bounds instead, we can rely on the weak upper bounds for $m = 1$ in Section 8.1 for $\delta > 0$, which are obviously not sharp. For $\delta < 0$, instead, we again define a notion of a core as for the configuration model in Theorem 7.9. However, due to the dynamics, the precise definition will be more involved and incorporate the dynamics in an appropriate way. This upper bound in performed in Sections 8.6 and 8.7 for $\delta \geq 0$ and $\delta < 0$ respectively, completing the proofs of the typical distances in $\text{PA}_n^{(m,\delta)}$. Sub-results will be stated explicitly, as these are sometimes sharper than the results stated in this section and are thus interesting in their own right.

8.3 Path counting in preferential attachment models

In this section we study the probability that a certain path is present in $\text{PA}_n^{(m,\delta)}$. Recall that we call a path $\vec{\pi} = (\pi_0, \pi_1, \ldots, \pi_l)$ self-avoiding when $\pi_i \neq \pi_j$ for all $1 \leq i < j \leq l$. The following proposition studies the probability that a path $\vec{\pi}$ is present in $\text{PA}_n^{(m,\delta)}$:

**Proposition 8.9** (Path counting in $\text{PA}_n^{(m,\delta)}$) Denote $\gamma = m/(2m + \delta)$. Fix $l \geq 0$
and let \( \bar{\pi} = (\pi_0, \pi_1, \ldots, \pi_l) \) be an \( l \)-step self-avoiding path consisting of the \( l + 1 \) unordered vertices \( \pi_0, \pi_1, \ldots, \pi_l \). Then, there exists a constant \( C > 0 \) such that, for all \( l \geq 1 \),
\[
\mathbb{P}(\bar{\pi} \subseteq \text{PA}_n^{(m,\delta)}) \leq \left(C m^2\right)^l \prod_{i=0}^{l-1} \frac{1}{(\pi_i \wedge \pi_{i+1})^\gamma (\pi_i \vee \pi_{i+1})^{1-\gamma}}. \tag{8.3.1}
\]

Paths are formed by repeatedly forming edges. When \( m = 1 \), paths always go from younger to older vertices. When \( m \geq 2 \), this monotonicity property of paths is lost, which generally makes proofs harder. We start by investigating intersections of events that specify which edges are present in \( \text{PA}_n^{(m,\delta)} \).

We recall some notation. The event that the \( j \)-th edge of vertex \( u \) is attached to the earlier vertex \( v \), where \( u, v \in [n] \), is denoted by
\[
\{u \xrightarrow{j} v\}, \quad j \in [m]. \tag{8.3.2}
\]
It will often be convenient to translate statements from \( \text{PA}_n^{(m,\delta)} \) to \( \text{PA}_n^{(1,\delta/m)} \). Indeed, for \( \text{PA}_n^{(m,\delta)} \), the event \( \{u \xleftarrow{j} v\} \) means that in \( \text{PA}_n^{(1,\delta/m)} \), the edge from vertex \( m(u-1) + j \) is attached to one of the vertices \([mv] \setminus [m(v-1)] \).

It is a direct consequence of the definition of PA-models that the event (8.3.2) increases the preference for vertex \( v \), and hence decreases (in a relative way) the preference for the other vertices in \([n] \setminus \{v\}\). It should be intuitively clear that another way of expressing this effect is to say that, for different \( v_1 \neq v_2 \), the events \( \{u_1 \xleftarrow{j_1} v_1\} \) and \( \{u_2 \xleftarrow{j_2} v_2\} \) are negatively correlated. We now formalize this result. For an integer \( n_v \geq 1 \), we denote the event that the \( j \)-th edge of vertex \( u_i^{(v)} \) is attached to the earlier vertex \( v \), for \( i \in [n_v] \), by
\[
\mathcal{E}_{n_v,v} = \bigcap_{i=1}^{n_v} \{u_i^{(v)} \xleftarrow{j} v\}. \tag{8.3.3}
\]
We start by proving that the events \( \mathcal{E}_{n_v,v} \), for different \( s \), are negatively correlated for each choice of \( k \geq 1 \) and all possible choices of \( u_i^{(v)}, j_i^{(v)} \):

**Lemma 8.10** (Negative correlation for connection of edges) For distinct \( v_1, v_2, \ldots, v_k \in [n] \) and all \( n_{v_1}, \ldots, n_{v_k} \geq 1 \), both for \( \text{PA}_n^{(m,\delta)} \) as well as for \( \text{PA}_n^{(m,\delta)}(b) \),
\[
\mathbb{P}\left(\bigcap_{s=1}^{k} \mathcal{E}_{n_{v_s},v_s}\right) \leq \prod_{s=1}^{k} \mathbb{P}(\mathcal{E}_{n_{v_s},v_s}). \tag{8.3.4}
\]

**Proof** We only prove the statement for \( \text{PA}_n^{(m,\delta)} \), the proof for \( \text{PA}_n^{(m,\delta)}(b) \) is identical. We define the edge number of the event \( \{u \xleftarrow{j} v\} \) to be \( m(u-1) + j \), which is the order of the edge when we consider the edges as being attached in sequence in \( \text{PA}_n^{(1,\delta/m)} \).

We use induction on the largest edge number present in the events \( \mathcal{E}_{n_{v_1},v_1}, \ldots, \mathcal{E}_{n_{v_k},v_k} \). The induction hypothesis is that (8.3.4) holds for all \( k \), all distinct \( v_1, v_2, \ldots, v_k \in [n] \), all \( n_{v_1}, \ldots, n_{v_k} \geq 1 \), and all choices of \( u_i^{(v)}, j_i^{(v)} \) such that \( \max_{i,s} m(u_i^{(v)} - 1) + j_i^{(v)} \leq c \), where induction is performed with respect to \( c \).
To initialize the induction, we note that for \( e = 1 \), the induction hypothesis holds trivially, since \( \bigcap_{k=1}^{e} E_{n_{v_{s}}, v_{s}} \) can be empty or consist of exactly one event, and in the latter case there is nothing to prove. This initializes the induction.

To advance the induction, we assume that (8.3.4) holds for all \( k \), all distinct \( v_{1}, v_{2}, \ldots, v_{k} \in [n] \), all \( n_{v_{1}}, \ldots, n_{v_{k}} \geq 1 \), and all choices of \( u_{1}^{(v_{s})}, j_{1}^{(v_{s})} \) such that \( \max_{i,s} m(u_{1}^{(v_{s})}) - 1) + j_{1}^{(v_{s})} \leq e - 1 \), and we extend it to all \( k \), all distinct \( v_{1}, v_{2}, \ldots, v_{k} \in [n] \), all \( n_{v_{1}}, \ldots, n_{v_{k}} \geq 1 \), and all choices of \( u_{1}^{(v_{s})}, j_{1}^{(v_{s})} \) such that \( \max_{i,s} m(u_{1}^{(v_{s})}) - 1) + j_{1}^{(v_{s})} \leq e - 1 \), the bound follows from the induction hypothesis, so we may restrict attention to the case that \( \max_{i,s} m(u_{1}^{(v_{s})}) - 1) + j_{1}^{(v_{s})} = e \).

We note that there is a unique choice of \( u, j \) such that \( m(u - 1) + j = e \). There are two possibilities: (1) Either there is exactly one choice of \( s \) and \( u_{1}^{(v_{s})}, j_{1}^{(v_{s})} \) such that \( u_{1}^{(v_{s})} = t, j_{1}^{(v_{s})} = j \), or (2) there are at least two such choices. In the latter case, \( \bigcap_{k=1}^{e} E_{n_{v_{s}}, v_{s}} = \emptyset \), since the \( e \)th edge is connected to a unique vertex. Hence, there is nothing to prove.

We are left to investigate the case where there exists a unique \( s \) and \( u_{1}^{(v_{s})}, j_{1}^{(v_{s})} \) such that \( u_{1}^{(v_{s})} = u, j_{1}^{(v_{s})} = j \). Denote the restriction of \( E_{n_{v_{s}}, v_{s}} \) to all other edges by

\[
E'_{n_{v_{s}}, v_{s}} = \bigcap_{i=1: (u_{1}^{(v_{s})}, j_{1}^{(v_{s})}) \neq (u, j)}^{n_{v_{s}}} \{ u_{i}^{(v_{s})}, j_{i}^{(v_{s})} \} = \bigcap_{i=1: (u_{1}^{(v_{s})}, j_{1}^{(v_{s})}) \neq (u, j)}^{n_{v_{s}}} \{ u_{i}^{(v_{s})}, j_{i}^{(v_{s})} \} \tag{8.3.5}
\]

Then we can write

\[
\bigcap_{i=1}^{k} E_{n_{v_{s}}, v_{s}} = \{ u \xymatrix{ \rightarrow \, v_{s} } \} \cap E'_{n_{v_{s}}, v_{s}} \cap \bigcap_{i=1: v_{i} \neq v_{s}}^{k} E_{n_{v_{s}}, v_{s}}. \tag{8.3.6}
\]

By construction, all the edge numbers of the events in \( E'_{n_{v_{s}}, v_{s}} \cap \bigcap_{i=1: v_{i} \neq v_{s}}^{k} E_{n_{v_{s}}, v_{s}} \) are at most \( e - 1 \).

By conditioning, we obtain

\[
P\left( \bigcap_{i=1}^{k} E_{n_{v_{s}}, v_{s}} \right) \leq E\left[ \mathbb{1}_{E_{n_{v_{s}}, v_{s}} \cap \bigcap_{i=1: v_{i} \neq v_{s}}^{k} E_{n_{v_{s}}, v_{s}}} P_{e-1}(u \xymatrix{ \rightarrow \, v_{s})} \right], \tag{8.3.7}
\]

where \( P_{e-1} \) denotes the conditional probability given the edge attachments up to the \((e - 1)\)st edge, or, equivalently, given \( PA_{e-1}^{(1,k/m)} \), and we have used that the event \( E'_{n_{v_{s}}, v_{s}} \cap \bigcap_{i=1: v_{i} \neq v_{s}}^{k} E_{n_{v_{s}}, v_{s}} \) is measurable with respect to \( PA_{e-1}^{(1,k/m)} \). We compute

\[
P_{e-1}(u \xymatrix{ \rightarrow \, v_{s}) = \frac{D_{vs}(u - 1, j - 1) + \delta}{z_{u,j}}, \tag{8.3.8}
\]

where we recall that \( D_{vs}(u - 1, j - 1) \) is the degree of vertex \( s \) after \( j - 1 \) edges of vertex \( u \) have been attached and we write the normalization in (8.3.8) as

\[
z_{u,j} = z_{u,j}(\delta, m) = (2m + \delta)(u - 1) + (j - 1)(2 + \delta/m) + 1 + \delta. \tag{8.3.9}
\]
We wish to use the induction hypothesis. For this, we note that
\[ D_{e_i}(u-1, j-1) = m + \sum_{(u', j') : m u' + j' \leq e-1} 1 \]  
(8.3.10)
where we recall that \( e-1 = m(u-1) + j - 1 \).

Each of the events \( \{u' \overset{\sim}{\rightleftharpoons} v_s \} \) in (8.3.10) has edge number strictly smaller than \( e \) and occurs with a non-negative multiplicative constant. As a result, we may use the induction hypothesis for each of these terms. Thus, we obtain, using also \( m \geq 0 \), that,
\[
P \left( \bigcap_{s=1}^{k} E_{u,v_s} \right) \leq \frac{m + \delta}{z_{u,j}} P \left( E_{u,v_s}^e \right) \prod_{t=1 : v_t \neq v_s}^{k} P \left( E_{u,v_t} \right) + \sum_{(u', j') : m u' + j' \leq e-1} \frac{P \left( E_{u,v_s}^e \right) \prod_{t=1 : v_t \neq v_s}^{k} P \left( E_{u,v_t} \right) \prod_{t=1 : v_t \neq v_s}^{k} P \left( E_{u,v_t} \right)}{z_{u,j}} P \left( E_{u,v_s} \right).
\]  
(8.3.11)
We use (8.3.10) to recombine the above as
\[
P \left( \bigcap_{s=1}^{k} E_{u,v_s} \right) \leq \mathbb{E} \left[ \prod_{t=1 : v_t \neq v_s}^{k} P \left( E_{u,v_t} \right) \right] \prod_{t=1 : v_t \neq v_s}^{k} P \left( E_{u,v_t} \right) \]  
(8.3.12)
and the advancement is completed when we note that
\[
\mathbb{E} \left[ \prod_{t=1 : v_t \neq v_s}^{k} P \left( E_{u,v_t} \right) \right] = P \left( E_{u,v_s} \right).
\]  
(8.3.13)
This advances the induction hypothesis. The claim in Lemma 8.10 then follows by induction.

We next study the probabilities of the events \( E_{n_v,v} \) when \( n_v \leq 2 \):

**Lemma 8.11** (Edge connection events for at most two edges) Consider \( PA_{n}^{(m,\delta)} \) and denote \( \gamma = m/(2m+\delta) \). There exist absolute constants \( M_1 = M_1(\delta, m) \), \( M_2 = M_2(\delta, m) \), such that the following bounds hold:

(i) For \( m = 1 \) and any \( u > v \),
\[
P (u \overset{\sim}{\rightleftharpoons} v) = \frac{1 + \delta \Gamma(u) \Gamma(v - \frac{1}{2m})}{2 + \delta \Gamma(u + \frac{1+\gamma}{2m}) \Gamma(v)}.
\]  
(8.3.14)
Consequently, for \( m \geq 2 \) and any \( 1 \leq j \leq m \) and \( u > v \),
\[
P (u \overset{\sim}{\rightleftharpoons} v) = \frac{m + \delta}{2m + \delta} \frac{1}{u^{1-\gamma} v^{1-\gamma}} (1 + o(1)) \leq \frac{M_1}{u^{1-\gamma} v^{1-\gamma}}.
\]  
(8.3.15)
the asymptotics in the first equality in (8.3.15) referring to the limit when \( v \) grows large.

(ii) For \( m = 1 \) and any \( u_2 > u_1 > v \),
\[
P (u_1 \overset{\sim}{\rightleftharpoons} v, u_2 \overset{\sim}{\rightleftharpoons} v) = \frac{1 + \delta}{2 + \delta} \frac{\Gamma(u_2) \Gamma(u_1 + \frac{1+\gamma}{2m}) \Gamma(v + \frac{1+\gamma}{2m})}{\Gamma(u_2 + \frac{1+\gamma}{2m}) \Gamma(u_1 + 1) \Gamma(v + \frac{3+\gamma}{2m})}.
\]  
(8.3.16)
8.3 Path counting in preferential attachment models

Consequently, for \( m \geq 2 \) and any \( 1 \leq j_1, j_2 \leq m \) and \( u_2 > u_1 > v \),

\[
P(u_1 \xrightarrow{j_1} v, u_2 \xrightarrow{j_2} v) \leq \frac{m + \delta}{2m + \delta} \cdot \frac{m + 1 + \delta}{2m + \delta} \cdot \frac{1}{(u_1 u_2)^{1-\gamma} u_2^{2\gamma}} \cdot (1 + o(1)) \tag{8.3.17}
\]

the asymptotics in the first equality in (8.3.17) referring to the limit when \( v \) grows large.

**Proof** We only prove (8.3.14) and (8.3.16), the bounds in (8.3.15) and (8.3.17) follow immediately from the Stirling-type formula in [Volume 1, (8.3.9)].

By the definition of \( \text{PA}(m, \delta)_n \geq 1 \) in terms of \( \text{PA}(1, \delta/m)_n \), this implies the result for general \( m \geq 1 \), where the factors of \( m \) follow from the fact that vertex \( u \) in \( \text{PA}(m, \delta)_n \) corresponds to vertices \( mu, \ldots, m(u + 1) - 1 \) in \( \text{PA}(1, \delta/m)_{mn} \), which are all at least \( mu \). Note, in particular, that \( u \xrightarrow{j_1} v \) for \( m \geq 2 \) in \( \text{PA}(m, \delta)_n \) is equivalent to \( m(u-1) + j_1 \xrightarrow{\overline{m}} [ms] \setminus [m(s-1)] \) in \( \text{PA}(1, \delta/m)_{mn} \). We will thus start the proofs of both parts (i) and (ii) with \( m = 1 \), and then use the results obtained to deduce the results for \( m \geq 2 \).

Fix \( m = 1 \) and consider part (i). For (8.3.14), we use [Volume 1, Theorem 8.2] to compute, for \( u > v \),

\[
P(u \xrightarrow{j_1} v) = \mathbb{E}[\mathbb{1}_{\{u \xrightarrow{j_1} v\}} | \text{PA}(1, \delta)_u] = \mathbb{E}\left[ \frac{D_v(u-1) + \delta}{(2 + \delta)(u-1) + 1 + \delta} \frac{\Gamma(u)\Gamma(v - \frac{1}{2\delta})}{\Gamma(u - 1 + \frac{1}{2\delta})\Gamma(v)} \right]
\]

\[
= 1 + \frac{\delta \Gamma(u)\Gamma(v - \frac{1}{2\delta})}{2 + \delta \Gamma(u + \frac{1+z}{2\delta})\Gamma(v)},
\]

as required. The proof of (8.3.15) follows by recalling that \( u \xrightarrow{j_1} v \) occurs when \((m-1)u + j \xrightarrow{\overline{m}} [mv] \setminus [m(v-1)]\). Since \( j \in [m] \) and \( m \geq 2 \) is fixed, we obtain (8.3.15).

We proceed with the proof of part (ii) and start with (8.3.16). Take \( u_2 > u_1 \). We compute

\[
P(u_1 \xrightarrow{j_1} v, u_2 \xrightarrow{j_2} v) = \mathbb{E}\left[ \mathbb{1}_{\{u_1 \xrightarrow{j_1} v, u_2 \xrightarrow{j_2} v\}} | \text{PA}(1, \delta)_{u_2-1} \right] = \mathbb{E}\left[ \mathbb{1}_{\{u_1 \xrightarrow{j_1} v\}} \frac{D_v(u_2 - 1) + \delta}{u_2 - 1}(2 + \delta) + 1 + \delta \right].
\]

(8.3.19)
We use the iteration, for $u_2 - 1 \geq u,$

$$
\mathbb{E} \left[ 1_{\{u_1 \sim v\}} (D_v(u) + \delta) \right] \\
= \left( 1 + \frac{1}{(2 + \delta)(u_1 - 1) + 1 + \delta} \right) \mathbb{E} \left[ 1_{\{u_1 \sim v\}} (D_v(u_1 - 1) + \delta) \right] \\
= \frac{u_1}{u_1 - 1 + \frac{1 + \delta}{2 + \delta}} \mathbb{E} \left[ 1_{\{u_1 \sim v\}} (D_v(u_1) + \delta) \right] \\
= \frac{\Gamma(u + 1)\Gamma(u_1 + \frac{1 + \delta}{2 + \delta})}{\Gamma(u_1 + 1)} \mathbb{E} \left[ 1_{\{u_1 \sim v\}} (D_v(u_1) + \delta) \right].
$$

Therefore,

$$
\mathbb{P} \left( u_1 \xrightarrow{\delta} v, u_2 \xrightarrow{\delta} v \right) \\
= \frac{1}{(u_2 - 1)(2 + \delta) + 1 + \delta} \frac{\Gamma(u_2)\Gamma(u_1 + \frac{1 + \delta}{2 + \delta})}{\Gamma(u_2 - \frac{1 + \delta}{2 + \delta})\Gamma(u_1 + 1)} \mathbb{E} \left[ 1_{\{u_1 \sim v\}} (D_v(u_1) + \delta) \right] \\
= \frac{1}{2 + \delta} \frac{\Gamma(u_2)\Gamma(u_1 + \frac{1 + \delta}{2 + \delta})}{\Gamma(u_2 + \frac{1 + \delta}{2 + \delta})\Gamma(u_1 + 1)} \mathbb{E} \left[ 1_{\{u_1 \sim v\}} (D_v(u_1) + \delta) \right].
$$

We are lead to compute $\mathbb{E} \left[ 1_{\{u_1 \sim v\}} (D_v(u_1) + \delta) \right].$ We use recursion to obtain

$$
\mathbb{E} \left[ 1_{\{u_1 \sim v\}} (D_v(u_1) + \delta) \right] \mid \text{PA}^{(m, \delta)}_{\{u_1 \sim v\}} \\
= \mathbb{E} \left[ 1_{\{u_1 \sim v\}} (D_v(u_1) - D_v(u_1 - 1)) \right] \mid \text{PA}^{(m, \delta)}_{\{u_1 \sim v\}} \\
+ \mathbb{E} \left[ 1_{\{u_1 \sim v\}} (D_v(u_1) - 1 + \delta) \right] \mid \text{PA}^{(m, \delta)}_{\{u_1 \sim v\}} \\
= \frac{(D_v(u_1) - 1 + \delta)(D_v(u_1 - 1) + 1 + \delta)}{(u_1 - 1)(2 + \delta) + 1 + \delta}.
$$

By [Volume 1, Proposition 8.15],

$$
\mathbb{E}[(D_v(u) + \delta)(D_v(u) + 1 + \delta)] = \frac{1}{c_2(u)} \mathbb{E}[Z_{v, 2}(u)] = \frac{c_2(v)}{c_2(u)} (2 + \delta)(1 + \delta). \quad (8.3.23)
$$

Recalling that $c_k(j) = \Gamma(j + \frac{1 + \delta}{2 + \delta}) \Gamma(j + \frac{k + 1 + \delta}{2 + \delta}),$ this brings us to

$$
\mathbb{E}[(D_v(u) + \delta)(D_v(u) + 1 + \delta)] = \frac{\Gamma(u + \frac{3 + \delta}{2 + \delta})\Gamma(v + \frac{1 + \delta}{2 + \delta})}{\Gamma(u + \frac{1 + \delta}{2 + \delta})\Gamma(v + \frac{3 + \delta}{2 + \delta})} (2 + \delta)(1 + \delta). \quad (8.3.24)
$$
Consequently,
\[
\mathbb{E} \left[ 1_{\{u_1 \xrightarrow{(j)} v\}} (D_v(u_1 - 1) + \delta) \right] = \frac{\Gamma(u_1 + \frac{1}{2+\delta})\Gamma(v + \frac{1+\delta}{2+\delta})}{[(u_1 - 1)(2 + \delta) + 1 + \delta]\Gamma(u_1 - \frac{1}{2+\delta})\Gamma(v + \frac{1+\delta}{2+\delta})} (2 + \delta)(1 + \delta) \\
= (1 + \delta) \frac{\Gamma(u_1 + \frac{1}{2+\delta})\Gamma(v + \frac{1+\delta}{2+\delta})}{\Gamma(u_1 + \frac{1+\delta}{2+\delta})\Gamma(v + \frac{1+\delta}{2+\delta})}.
\]

Combining (8.3.21)–(8.3.25), we arrive at
\[
P(u_1 \xrightarrow{(j)} v, u_2 \xrightarrow{(j)} v),
\]

(8.3.26)

as claimed in (8.3.16).

The proof of (8.3.17) follows again by recalling that \( u \xrightarrow{(j)} v \) occurs when \( (m - 1)u + j \xrightarrow{(m)} [mv] \setminus [m(v - 1)] \). Now there are two possibilities, depending on whether \( m(u_1 - 1) + j \xrightarrow{(m)} v_1 \) and \( m(u_2 - 1) + j \xrightarrow{(m)} v_2 \) for the same \( v_1 = v_2 \in [mv] \setminus [m(v - 1)] \), or for two different \( v_1, v_2 \in [mv] \setminus [m(v - 1)] \).

In the case where \( v_1 = v_2 \), we use (8.3.16) to obtain a contribution that is asymptotically equal to
\[
\frac{m}{2m + \delta} \frac{1}{(u_1 u_2)^{1-\gamma v^2\gamma}} (1 + o(1)),
\]

(8.3.27)

where the factor \( m \) comes from the \( m \) distinct choices for \( v_1 \), and the factor \( 1/m^2 \) originates since we should multiply \( u_1, u_2 \) and \( v \) in (8.3.16) by \( m \).

In the case where \( v_1 \neq v_2 \), we use the negative correlation in Lemma 8.10 to bound this contribution from above by the product of the probabilities in (8.3.15), so that this contribution is asymptotically bounded by
\[
\frac{m(m - 1)}{m^2} \frac{1}{(2m + \delta)^2} \frac{1}{(u_1 u_2)^{1-\gamma v^2\gamma}} (1 + o(1)).
\]

(8.3.28)

Summing (8.3.27) and (8.3.28) completes the proof of (8.3.17).

With Lemmas 8.10 and 8.11 in hand, we are ready to prove Proposition 8.9:

Proof of Proposition 8.9. Since \( \vec{\pi} \) is self-avoiding, we can write
\[
\{ \vec{\pi} \subseteq \text{PA}_n^{(m,j)} \} = \bigcap_{s=1}^{k} \mathcal{E}_{\pi_{vs},v_s},
\]

(8.3.29)
where either
\[ \mathcal{E}_{n,v_s} = \{ u \overset{j}{\rightarrow} v_s \} \] (8.3.30)
for some \( u > v \) and some \( j \in [m] \), or
\[ \mathcal{E}_{n,v_s} = \{ u_1 \overset{j_1}{\rightarrow} v_s, u_2 \overset{j_2}{\rightarrow} v_s \} \] (8.3.31)
for some \( u_1, u_2 > v \) and some \( j_1, j_2 \in [m] \).

In the first case, by (8.3.14),
\[ P(\mathcal{E}_{n,v_s}) = P(u \overset{j}{\rightarrow} v_s) \leq M_1 \frac{u^\gamma}{u^1} \] (8.3.32)
whereas in the second case, according to (8.3.16),
\[ P(\mathcal{E}_{n,v_s}) = P(u_1 \overset{j_1}{\rightarrow} v_s, u_2 \overset{j_2}{\rightarrow} v_s) \leq M_2 \frac{u_1^{1-\gamma} u_2^{1-\gamma}}{u_1^{1} u_2^{1}}. \] (8.3.33)

In both cases \( M_i, i = 1, 2 \), is an absolute constant. Lemma 8.10 then yields (8.3.1),
where the factor \( m^2 \) originates from the number of possible choices of \( j_i \in [m] \) for \( i \in [k] \) and the possible \( s_i \) that are collapsed to the same vertex.

We close this section by proving some results that have been used earlier. The following variance estimate was used in the proof of Proposition 8.5:

**Lemma 8.12** (A variance estimate on \( N_k(n) \)) \( \text{Recall the definition of } N_k(n) \text{ in (8.1.41) and let } T = n^\varepsilon \text{ for some } \varepsilon > 0. \text{ Then there exist constants } C > 0 \text{ and } \eta = \eta(\varepsilon) > 0 \text{ such that} \)
\[ \text{Var}(N_k(n)) \leq C \mathbb{E}[N_k(n)]^2 n^{-\eta}. \] (8.3.34)

**Proof** By Exercise 8.4, when the path \( \varepsilon \pi = (\pi_0, \ldots, \pi_k) \) is completely disjoint from \( \varepsilon \rho = (\rho_0, \ldots, \rho_k) \),
\[ P(\bigcap_{i=0}^{k-1} \{ \pi_i \rightsquigarrow \pi_{i+1} \} \cap \bigcap_{i=0}^{k-1} \{ \rho_i \rightsquigarrow \rho_{i+1} \}) \leq P(\bigcap_{i=0}^{k-1} \{ \pi_i \rightsquigarrow \pi_{i+1} \}) P(\bigcap_{i=0}^{k-1} \{ \rho_i \rightsquigarrow \rho_{i+1} \}). \] (8.3.35)

Therefore, the indicators of disjoint paths are negatively correlated. As a result, we can bound
\[ \text{Var}(N_k(n)) \leq \sum_{\varepsilon \pi, \varepsilon \rho \neq \varepsilon \pi} P(\varepsilon \pi, \varepsilon \rho \subseteq \text{PA}\varepsilon_1, \varepsilon_2(n)) (b)). \] (8.3.36)

Since \( m = 1 \), the paths \( \varepsilon \pi, \varepsilon \rho \) of length \( k \) must merge at some point, before moving off to their common end point in \([n]\). When \( \varepsilon \rho = \varepsilon \pi \), then we obtain a contribution \( \mathbb{E}[N_k(n)] \), so that from now on we assume that \( \varepsilon \pi \neq \varepsilon \rho \).

Write \( \varepsilon \pi = (\pi_0, \ldots, \pi_k) \) and \( \varepsilon \rho = (\rho_0, \ldots, \rho_k) \) and \( \varepsilon \pi \neq \varepsilon \rho \). Then there must be an
8.3 Path counting in preferential attachment models

$l \in [k-1]$ such that $\pi_j = \rho_j$ for all $j = l, \ldots, k$. Fix two fixed paths $\vec{\pi}$ and $\vec{\rho}$ for which $\pi_j = \rho_j$ for all $j = l, \ldots, k$, while $\pi_{l-1} \neq \rho_{l-1}$. Then, by Lemma 8.10,

$$P(\vec{\pi}, \vec{\rho} \subseteq PA_n^{(l,\delta)}(\rho)) \leq \prod_{i=1}^{l-1} P(\pi_{i-1} \leadsto \pi_i) P(\rho_{i-1} \leadsto \rho_i)$$

$$\times P(\pi_{l-1}, \rho_{l-1} \leadsto \pi_l) \prod_{j=l+1}^{k} P(\pi_{j-1} \leadsto \pi_j).$$

By (8.1.9) in Proposition 8.4,

$$\prod_{i=1}^{l-1} P(\pi_{i-1} \leadsto \pi_i) = \left(\frac{1 + \delta}{2 + \delta}\right)^{l-1} \frac{\Gamma(\pi_0 + \frac{1}{2+\delta})\Gamma(\pi_{l-1})}{\Gamma(\rho_{l-1} + \frac{1}{2+\delta})\Gamma(\rho_0 + 1)} \prod_{i=1}^{l-2} \frac{1}{\pi_i}$$

By symmetry, we may without loss of generality that assume $\pi_{l-1} > \rho_{l-1}$. Then, by (8.3.16),

$$P(\pi_{l-1}, \rho_{l-1} \leadsto \pi_l) = \left(1 + \delta\right)^{\Gamma(\rho_{l-1} - \frac{\delta}{2+\delta})\Gamma(\pi_{l-1} - \frac{1+k+\delta}{2+\delta})\Gamma(\pi_l)}$$

As a result,

$$P(\vec{\pi}, \vec{\rho} \subseteq PA_n^{(l,\delta)}(\rho))$$

$$\leq \left(\frac{1 + \delta}{2 + \delta}\right)^{l+k-3} \frac{\Gamma(\pi_0 + \frac{1}{2+\delta})\Gamma(\pi_{l-1})}{\Gamma(\rho_{l-1} + \frac{1}{2+\delta})\Gamma(\rho_0 + 1)} \prod_{i=1}^{l-2} \frac{1}{\pi_i}$$

$$\times \frac{\Gamma(\pi_l + \frac{1}{2+\delta})\Gamma(\pi_k)}{\Gamma(\pi_k + \frac{1}{2+\delta})\Gamma(\pi_l + 1)} \prod_{i=l+1}^{k-1} \frac{1}{\pi_i}$$

$$= \left(1 + \delta\right)^{l+k-3} \frac{\Gamma(\pi_0 + \frac{1}{2+\delta})\Gamma(\rho_0 + \frac{1}{2+\delta}) \Gamma(\rho_{l-1} - \frac{\delta}{2+\delta})\Gamma(\pi_l)}{\Gamma(\rho_{l-1} + \frac{1}{2+\delta})\Gamma(\rho_0 + 1) \Gamma(\rho_{l-1} + \frac{1}{2+\delta})\Gamma(\pi_l + \frac{1}{2+\delta})} \prod_{i=1}^{l-2} \frac{1}{\pi_i} \prod_{i=l+1}^{k-1} \frac{1}{\pi_i}.$$

By [Volume 1, (8.3.9)], this can be bounded by

$$C \left(\frac{1 + \delta}{2 + \delta}\right)^{l+k} (\pi_0 \rho_0)^{-1+\delta/(2+\delta)} (\pi_l \pi_k)^{-1/(2+\delta)} \prod_{i=1}^{l-1} \frac{1}{\rho_i} \prod_{i=l+1}^{k-1} \frac{1}{\pi_i},$$

where $C$ is a uniform constant. We need to sum the above over $l < k$, all decreasing $\vec{\pi} = (\pi_0, \ldots, \pi_k)$ with $\pi_k \in [n]$ and all decreasing $(\rho_0, \ldots, \rho_{l-1})$ with $\rho_{l-1} > \pi_l$. The sum can be bounded from above by summing over all decreasing $(\rho_0, \ldots, \rho_{l-1})$,........
and bounding \( \pi_i^{-1/(2+\delta)} \leq 1 \), to obtain an upper bound as in (8.1.43) of the form
\[
\mathbb{E}[N_k(n)] \leq \frac{(1 + \delta)}{2 + \delta} \left( \frac{\log n}{l!} \right)^t \sum_{s \in [T]} s^{-(1+\delta)/(2+\delta)}.
\] (8.3.42)

Therefore,
\[
\text{Var}(N_k(n)) \leq \mathbb{E}[N_k(n)] + \mathbb{E}[N_k(n)] \sum_{l=1}^{k-1} \left( \frac{1 + \delta}{2 + \delta} \right) \left( \frac{\log n}{l!} \right)^t \sum_{s \in [T]} s^{-1}.
\] (8.3.43)

When \( T = n^\epsilon \), it is not hard to adapt the arguments in (8.1.43)–(8.1.49) to show that this is at most \( \mathbb{E}[N_k(n)]^2 n^{-\eta} \) for some \( \eta = \eta(\epsilon) > 0 \), as required.

We use similar means as in the above argument to show the missing ingredient
\[
\text{dist}_{\text{PA}(1, \delta)}(1, V) = o(\log n)
\] in the proof of Theorem 8.3:

**Lemma 8.13** (Distance to most-recent common ancestor) Fix \( o_1, o_2 \) to be two vertices in \([n]\) chosen independently and uniformly at random, and let \( V \) be the oldest vertex that the path from 1 to \( o_1 \) and that from 1 to \( o_2 \) have in common in \( \text{PA}_n(1, \delta) \). Then,
\[
\frac{\text{dist}_{\text{PA}(1, \delta)}(1, V)}{\log n} \xrightarrow{p} 0.
\] (8.3.44)

**Proof** We will show that, for every \( \varepsilon > 0 \),
\[
\mathbb{P}(\text{dist}_{\text{PA}(1, \delta)}(1, V) > \varepsilon \log n) = o(1).
\] (8.3.45)

By definition,
\[
\mathbb{P}(\text{dist}_{\text{PA}(1, \delta)}(1, V) > \varepsilon \log n) = \frac{1}{n^2} \sum_{k > \varepsilon \log n} \sum_{u_1, u_2, s_1, s_2} \mathbb{P}(\text{dist}_{\text{PA}(1, \delta)}(1, v) = k, s_1, s_2 \leadsto v, s_1 \leftrightarrow u_1, s_2 \leftrightarrow u_2).
\] (8.3.46)

We use Lemma 8.10 to bound this by
\[
\mathbb{P}(\text{dist}_{\text{PA}(1, \delta)}(1, V) > \varepsilon \log n) = \frac{1}{n^2} \sum_{k > \varepsilon \log n} \sum_{u_1, u_2, s_1, s_2} \mathbb{P}(\text{dist}(1, v) = k) \mathbb{P}(s_1, s_2 \leadsto v) \mathbb{P}(s_1 \leftrightarrow u_1) \mathbb{P}(s_2 \leftrightarrow u_2),
\] (8.3.47)

where we also rely on (8.1.11) to note that
\[
\mathbb{P}(\text{dist}(s_1, u_1) = l) = \sum_{\pi_1, \ldots, \pi_{l-1}} \mathbb{P}(s_1 \leadsto \pi_1 \leadsto \cdots \leadsto \pi_{l-1} \leadsto u_1) = \sum_{\pi_1, \ldots, \pi_{l-1}} \mathbb{P}(s_1 \leadsto \pi_1) \mathbb{P}(\pi_{l-1} \leadsto u_1) \prod_{i=1}^{l-2} \mathbb{P}(\pi_i \leadsto \pi_{i+1}).
\] (8.3.48)

This allows us to recombine products of probabilities into connection events.
By (8.1.9) in Proposition 8.4,
\[ \mathbb{P}(\text{dist}(s_1, u_1) = l) = \frac{(1 + \delta)^l}{2 + \delta} \frac{\Gamma(u_1 + \frac{1}{2 + \delta})\Gamma(s_1)}{\Gamma(s_1 + \frac{1}{2 + \delta})\Gamma(u_1 + 1)} \sum_{\delta_i \in \mathbb{F}_l} \frac{1}{\tau_i} \tag{8.3.49} \]
Summing this over \( l \geq 1 \) leads to
\[ \mathbb{P}(s_1 \leftrightarrow u_1) \leq C \left( \frac{u_1}{s_1} \right)^{(1+\delta)/(2+\delta)} \frac{\Gamma(u_1 + \frac{1}{2 + \delta})\Gamma(s_1)}{\Gamma(s_1 + \frac{1}{2 + \delta})\Gamma(u_1 + 1)} \tag{8.3.50} \]
Thus, we obtain
\[ \mathbb{P}(\text{dist}_{\text{PA}}(1, V) > \varepsilon \log n) \leq C n^2 \sum_{k > \varepsilon \log n} \sum_{v, s_1, s_2} \mathbb{P}(\text{dist}(1, v) = k) \mathbb{P}(s_1, s_2 \hookrightarrow v) \frac{(n-s_1)(n-s_2)}{s_1 s_2}, \tag{8.3.51} \]
By (8.3.17) in Lemma 8.11, we can further bound this by
\[ \mathbb{P}(\text{dist}_{\text{PA}}(1, V) > \varepsilon \log n) \leq C \sum_{k > \varepsilon \log n} \sum_{v, s_1, s_2} \mathbb{P}(\text{dist}(1, v) = k) \frac{1}{(s_1 s_2)^{2-\gamma} v^{2\gamma}}, \tag{8.3.52} \]
with the restriction that \( v < s_1, s_2 \). Summing out over \( s_1, s_2 > v \) and using that \( 2 - \gamma > 1 \) leads to
\[ \mathbb{P}(\text{dist}_{\text{PA}}(1, V) > \varepsilon \log n) \leq C \sum_{k > \varepsilon \log n} \sum_{v} \mathbb{P}(\text{dist}(1, v) = k) \frac{1}{v^2}. \tag{8.3.53} \]
Again using (8.3.49), we are led to
\[ \mathbb{P}(\text{dist}_{\text{PA}}(1, V) > \varepsilon \log n) \leq C \sum_{k > \varepsilon \log n} \sum_{v} \frac{1 + \delta}{2 + \delta} k \frac{(\log v)^{k-1}}{(k-1)!} \frac{1}{v^2} \tag{8.3.54} \]
\[ \leq C \sum_{k > \varepsilon \log n} \sum_{v} \frac{1 + \delta}{2 + \delta} \frac{(\log v)^{k-1}}{(k-1)!} \frac{1}{v^{2+(1+\delta)/(2+\delta)}} \]
\[ \leq C \mathbb{P}\left( \text{Poi}\left( (1 + \delta)(\log W)/(2 + \delta) \right) \geq \varepsilon \log n \right), \]
where \( \mathbb{P}(W = v) = A/v^{2+(1+\delta)/(2+\delta)} \) for an appropriate \( A > 0 \) and for all \( v \geq 1 \). Since \( (1 + \delta)(\log W)/(2 + \delta) \) is a finite random variable, this probability vanishes, as required. \( \Box \)

In Exercise 8.8, you are asked to conclude from the above proof that \( \text{dist}_{\text{PA}}(1, V) \) is a tight sequence of random variables.
8.4 Small-world effect in PAMs: logarithmic lower bounds for $\delta \geq 0$

In this section we prove lower bounds on distances in PA$_n^{(m,\delta)}$ with $m \geq 2$, using the bounds derived in Section 8.3 on the probability that a path exists in PA$_n^{(m,\delta)}$, which is our main tool in this section. We prove the lower bounds on distances for $\delta > 0$ in Section 8.4.1 and for $\delta = 0$ in Section 8.4.2. The doubly logarithmic upper bounds on distances for $\delta < 0$ are deferred to Section 8.5.

8.4.1 Logarithmic lower bounds on distances for $\delta > 0$

In this section we investigate logarithmic lower bounds on the distances when $\delta > 0$, in which case $\gamma = m/(2m + \delta) < 1/2$, as stated in the lower bound in Theorem 8.6.

We start by proving a lower bound with the wrong constant. By Proposition 8.9,

$$\mathbb{P}\left( \text{dist}_{PA_n^{(m,\delta)}}(1,n) = k \right) \leq c^k \sum_{\pi} \prod_{j=0}^{k-1} \frac{1}{(\pi_j \land \pi_{j+1})^\gamma (\pi_j \lor \pi_{j+1})^{1-\gamma}},$$

(8.4.1)

where $c = m^2C$, and where the sum is over all self-avoiding paths $\pi = (\pi_0, \ldots, \pi_k)$ with $\pi_k = n, \pi_0 = 1$. Define

$$f_k(i,n) = \sum_{\pi} \prod_{j=0}^{k-1} \frac{1}{(\pi_j \land \pi_{j+1})^\gamma (\pi_j \lor \pi_{j+1})^{1-\gamma}},$$

(8.4.2)

where now the sum is over all self-avoiding $\pi = (\pi_0, \ldots, \pi_k)$ with $\pi_k = n, \pi_0 = i$, so that

$$\mathbb{P}\left( \text{dist}_{PA_n^{(m,\delta)}}(i,n) = k \right) \leq c^k f_k(i,n).$$

(8.4.3)

We study the function $f_k(i,n)$ in the following lemma, which shows that $k \mapsto f_k(i,n)$ grows at most exponentially, and gives uniform control over the dependence on $i$ and $n$ as well:

**Lemma 8.14** (A bound on $f_k$) Fix $\gamma < 1/2$. Then, for every $b > \gamma$ such that $\gamma + b < 1$, there exists a $C_{\gamma,b} > 0$ such that, for every $1 \leq i < n$ and all $k \geq 1$,

$$f_k(i,n) \leq \frac{C_{\gamma,b}^b}{i^b n^{1-b}}.$$  \hspace{1cm} (8.4.4)

We will prove sharper asymptotics on the graph distances in PA$_n^{(m,\delta)}$ in the sequel. However, also in that proof, Lemma 8.14 turns out to be convenient.

**Proof** We prove the lemma using induction on $k \geq 1$. To initialize the induction hypothesis, we note that, for $1 \leq i < n$ and every $b \geq \gamma$,

$$f_1(i,n) = \frac{1}{(i \land n)^{1-\gamma}} = \frac{1}{i^{1-\gamma}} \leq \frac{1}{i^{1-\gamma}} \leq \frac{1}{n^{1-\gamma}} = \frac{1}{n^{1-\gamma}}.$$

(8.4.5)

This initializes the induction hypothesis whenever $C_{\gamma,b} \geq 1$. 

8.4 Small-world effect in PAMs: logarithmic lower bounds for $\delta \geq 0$

To advance the induction hypothesis, note that

$$f_k(i, n) \leq \sum_{s=1}^{i-1} \frac{1}{s^{\gamma} i^{1-\gamma}} f_{k-1}(s, n) + \sum_{s=i+1}^{\infty} \frac{1}{s^{\gamma} i^{1-\gamma}} f_{k-1}(s, n).$$  \hspace{1cm} (8.4.6)

We now bound each of these two contributions in (8.4.6), making use of the induction hypothesis. We bound the first sum in (8.4.6) by

$$\sum_{s=1}^{i-1} \frac{1}{s^{\gamma} i^{1-\gamma}} f_{k-1}(s, n) \leq C_{\gamma, b}^{k-1} \sum_{s=1}^{i-1} \frac{1}{s^{\gamma} n^{1-b}} \sum_{s=i+1}^{\infty} \frac{1}{s^{\gamma} i^{1-\gamma} n^{1-b}} = C_{\gamma, b}^{k-1} \sum_{s=1}^{i-1} \frac{1}{s^{\gamma} i^{1-\gamma} n^{1-b}} \sum_{s=i+1}^{\infty} \frac{1}{s^{\gamma} i^{1-\gamma} n^{1-b}} \leq \frac{C_{\gamma, b}^{k-1}}{1 - \gamma - b i^{1-b}}.$$  \hspace{1cm} (8.4.7)

since $\gamma + b < 1$. We bound the second sum in (8.4.6) by

$$\sum_{s=i+1}^{\infty} \frac{1}{s^{\gamma} i^{1-\gamma}} f_{k-1}(s, n) \leq C_{\gamma, b}^{k-1} \sum_{s=i+1}^{\infty} \frac{1}{s^{\gamma} n^{1-b}} + C_{\gamma, b}^{k-1} \sum_{s=n+1}^{\infty} \frac{1}{s^{\gamma} i^{1-\gamma} n^{1-b}} \leq \frac{C_{\gamma, b}^{k-1}}{b - \gamma i^{1-b}} + \frac{C_{\gamma, b}^{k-1}}{1 - \gamma - b i^{1-b}},$$  \hspace{1cm} (8.4.8)

since $1 + b - \gamma > 1$, $2 - \gamma - b > 1$, $b > \gamma$ and $(n/i)^{\gamma} \leq (n/i)^{b}$. We conclude that

$$f_k(i, n) \leq C_{\gamma, b}^{k-1} \left( \frac{1}{b - \gamma} + \frac{1}{1 - \gamma - b} \right) \leq \frac{C_{\gamma, b}^{k}}{b^{1-b}},$$  \hspace{1cm} (8.4.9)

whenever

$$C_{\gamma, b} = \frac{1}{b - \gamma} + \frac{2}{1 - \gamma - b} \geq 1.$$  \hspace{1cm} (8.4.10)

This advances the induction hypothesis, and completes the proof of Lemma 8.14. \hfill $\Box$

We next extend the above discussion to typical distances:

**Lemma 8.15** (Typical distances for $\delta > 0$) Fix $m \geq 1$ and $\delta > 0$. Let dist$_{\text{PAM}}(n, a, b, \delta)$ be the distance between two uniformly chosen vertices in $[n]$. Then, whp, dist$_{\text{PAM}}(n, a, b, \delta) \geq c_2 \log n$ for $c_2 = c_2(m, \delta) > 0$ sufficiently small.

**Proof** By Lemma 8.14, with $K = \log (eC_{\gamma, b} \vee 2)$ and $\gamma < b < 1 - \gamma$, and for all $1 \leq i < j \leq t$,

$$P(\text{dist}_{\text{PAM}}(n, a, b, \delta) = k) \leq e^k f_k(i, j) \leq \frac{e^{Kk}}{i^b j^{1-b}}.$$  \hspace{1cm} (8.4.11)

As a result,

$$P(\text{dist}_{\text{PAM}}(n, a, b, \delta) \leq c_2 \log n) \leq \frac{n^{Kc_2}}{i^b j^{1-b} e^{K} - 1},$$  \hspace{1cm} (8.4.12)
and thus, using also $\sum_{i=1}^{j-1} i^{-b} \leq j^{1-b}/(1 - b)$,
\[
\mathbb{P}(\text{dist}_{\mathbf{PA}(m,\delta)}(o_1, o_2) \leq c_2 \log n) = \frac{2}{n^2} \sum_{1 \leq i < j \leq n} \mathbb{P}(\text{dist}_{\mathbf{PA}(m,\delta)}(i, j) \leq c_2 \log n) \leq 2 \sum_{1 \leq i < j \leq n} \frac{n^{Kc_2}}{j^{b-1}} = O(n^{Kc_2-1}) = o(1),
\]
for every $c_2 > 0$ such that $Kc_2 + 1 < 2$.

\[\square\]

8.4.2 Extension to lower bounds on distances for $\delta = 0$ and $m \geq 2$

In this section we investigate lower bounds on the distances in $\mathbf{PA}_n^{(m,\delta)}$ when $m \geq 2$ and $\delta = 0$ and prove the lower bound in Theorem 8.7.

We again start from Proposition 8.9, which as we will show implies that for $\delta = 0$,
\[
k^*_n = \frac{\log n}{\log(2Cm^2 \log n)}
\]
is a lower bound for the typical distances in $\mathbf{PA}_n^{(m,\delta)}$. Consider a path $\vec{\pi} = (\pi_0, \pi_1, \ldots, \pi_k)$, then (8.3.1) in Proposition 8.9 implies that
\[
\mathbb{P}(\vec{\pi} \subseteq \mathbf{PA}_n^{(m,\delta)}) \leq (Cm^2)^{k} \prod_{j=0}^{k} \frac{1}{\sqrt{j} \pi_j^{j+1}} = (Cm^2)^{k} \prod_{j=1}^{k} \frac{1}{\pi_j^{j}}.
\]
We use $\sum_{a=1}^{n} 1/\sqrt{a} \leq \int_{0}^{n} 1/\sqrt{x} dx = 2\sqrt{n}$ to compute that
\[
\mathbb{P}(\text{dist}_{\mathbf{PA}(m,\delta)}(o_1, o_2) = k) \leq \frac{4(Cm^2)^k}{n^2} \sum_{i} \frac{1}{\sqrt{\pi_i \pi_{i+1}}} \prod_{j=1}^{k-1} \frac{1}{\pi_j} \leq \frac{4(Cm^2)^k}{n} \sum_{1 \leq \pi_1 \leq \ldots \leq \pi_{k-1} \leq \pi_k} \prod_{j=1}^{k-1} \frac{1}{\pi_j}.
\]
Thus,
\[
\mathbb{P}(\text{dist}_{\mathbf{PA}(m,\delta)}(o_1, o_2) = k) \leq \frac{4(Cm^2)^k}{n} \left( \sum_{s=1}^{n} \frac{1}{s} \right)^{k-1} \leq \frac{4(Cm^2)^k}{n} (\log n)^{k-1} \leq 4\left(\frac{1}{2}\right)^k (\log n)^{-1} \to 0,
\]
precisely when $(2Cm^2 \log n)^k \leq n$, or, equivalently,
\[
k \leq \frac{\log n}{\log(2Cm^2 \log n)}.
\]
8.5 Typical distance in PA-models: log log-lower bound for $\delta < 0$

Equality in (8.4.18) holds for $k = k_n^*$ in (8.4.14). This implies that the typical distances are whp at least $k_n^*$ in (8.4.14), and completes the proof of the lower bound on the graph distances for $\delta = 0$ in Theorem 8.7.

In Exercise 8.6, you are asked to prove that also the distance between vertices $n-1$ and $n$ is also whp at least $k_n^*$ in (8.4.14). In Exercise 8.7, you are asked to check whether the above proof also implies that the distance between vertices 1 and 2 is also whp at least $k_n^*$ in (8.4.14).

8.5 Typical distance in PA-models: log log-lower bound for $\delta < 0$

In this section we prove the lower bound in Theorem 8.8. We do so in a more general setting, by assuming an upper bound on the existence of paths in the model that is inspired by Proposition 8.9:

**Assumption 8.16 (Path probabilities)** There exist $\kappa$ and $\gamma$ such that, for all $n$ and self-avoiding paths $\vec{\pi} = (\pi_0, \ldots, \pi_k) \in [n]^l$,

$$P(\vec{\pi} \subseteq PA_n) \leq \prod_{i=1}^{k} \kappa^{(\pi_i - 1) \wedge (\pi_i - 1)} (\pi_i \vee (\pi_{i-1})^{-1}). \quad (8.5.1)$$

By Proposition 8.9, Assumption 8.16 is satisfied for $PA_n^{(m, \delta)}$ with $\gamma = m/(2m + \delta)$. We expect log log-distances in such networks if and only if $\delta \in (-m, 0)$, so that $1/2 < \gamma < 1$. Theorem 8.17, which is the main result in this section, gives a lower bound on the typical distance in this case:

**Theorem 8.17 (Doubly logarithmic lower bound on distances PAMs)** Let $(PA_n)_{n \geq 1}$ be a random graph model that satisfies Assumption 8.16 for some $\gamma$ satisfying $1/2 < \gamma < 1$. Fix random vertices $o_1$ and $o_2$ chosen independently and uniformly from $[n]$, and define

$$\tau = \frac{2 - \gamma}{1 - \gamma} \in (2, 3). \quad (8.5.2)$$

Then, for $K$ large,

$$P(\text{dist}_{PA_n}(o_1, o_2) \geq \frac{4 \log \log n}{|\log(\tau - 2)|} - K) = 1 - o(1). \quad (8.5.3)$$

Here, $o(1)$ refers to a term that vanishes as first $n \to \infty$, followed by $K \to \infty$.

We will prove Theorem 8.17 in the form where $|\log(\tau - 2)|$ is replaced with $\log(\gamma/(1-\gamma))$. For $PA_n^{(m, \delta)}$, $\gamma = m/(2m + \delta)$, so that

$$\frac{\gamma}{1 - \gamma} = \frac{m}{m + \delta} = \frac{1}{\tau - 2}, \quad (8.5.4)$$

where we recall that $\tau = 3 + \delta/m$. Therefore, Theorem 8.17 proves the lower bound in Theorem 8.8, and even extends this to lower tightness for the distances.
We will prove a slightly different version of Theorem 8.17, namely, that, uniformly in \( u, v \geq \varepsilon n \), we can choose \( K = K_\varepsilon > 0 \) sufficiently large, so that, uniformly in \( n \geq 1 \),

\[
\mathbb{P}(\text{dist}_{PA_n}(u,v) \leq \frac{4 \log \log n}{\log(n) - K} - K) \leq \varepsilon. \tag{8.5.5}
\]

Since \( \mathbb{P}(o_1 \leq \varepsilon n) \leq \varepsilon \), this clearly implies Theorem 8.17.

The proof of Theorem 8.17 is based on a constrained or truncated first moment method, similar to the ones used for \( NR_n(w) \) in Theorem 6.6 and for \( CM_n(d) \) in Theorem 7.6. Due to the fact that the probability for existence of paths in Assumption 8.16 satisfies a rather different bound compared to those for \( NR_n(w) \) and \( CM_n(d) \), this truncated first order method looks rather different compared to those presented in the proof of Theorems 6.6 and 7.6.

Let us now briefly explain the truncated first moment method. We start with an explanation of the (unconstrained) first moment bound and its shortcomings. Let \( v, w \) be distinct vertices of \( PA_n \). We think of \( v, w \geq \varepsilon n \) for \( \varepsilon > 0 \) sufficiently small. Then, for \( k_n \in \mathbb{N} \),

\[
\mathbb{P}(\text{dist}_{PA_n}(u,v) \leq 2k_n) = \mathbb{P}\left( \bigcup_{k=1}^{2k_n} \bigcup_{\vec{p}\subseteq PA_n} \{\vec{p} \subseteq PA_n\} \right) \leq \sum_{k=1}^{2k_n} \sum_{\vec{p}} \prod_{j=1}^{k} p(\pi_{j-1}, \pi_j), \tag{8.5.6}
\]

where \( \vec{p} = (\pi_0, \ldots, \pi_k) \) is a self-avoiding path in \( PA_n \) with \( \pi_0 = u \) and \( \pi_k = v \) and, for \( a, b \in \mathbb{N} \), we define

\[
p(a,b) = \kappa(a \wedge b)^{-\gamma}(a \vee b)^{\gamma-1}. \tag{8.5.7}
\]

To each path \( \vec{p} = (\pi_0, \ldots, \pi_k) \), we assign the weight

\[
p(\vec{p}) = \prod_{j=1}^{k} p(\pi_{j-1}, \pi_j), \tag{8.5.8}
\]

so that the upper bound is just the sum over the weights of all paths from \( u \) to \( v \) of length no more than \( 2k_n \).

The shortcoming of the above bound is that the paths that contribute most to the total weight are those that connect \( v \), respectively \( w \), quickly to vertices with extremely small indices, and consequently extremely high degrees. Since on the other hand, such paths are quite unlikely to be present, it is actually quite unlikely that this occurs. This explain why they have to be removed in order to get a reasonable estimate, and why this leads to small errors when we do so.

To this end we define a decreasing sequence \( g = (g_k)_{k=0, \ldots, k} \) of positive integers and consider a path \( \vec{p} = (\pi_0, \ldots, \pi_k) \) to be good when \( \pi_l \wedge \pi_{k-l} \geq g_l \) for all \( l \in \{0, \ldots, k\} \). We denote the event that there exists a good path of length \( k \)
8.5 Typical distance in PA-models: log log-lower bound for $\delta < 0$

between $v$ and $w$ by $E_k(v, w)$. We further denote the event that there exists a bad path of length $k$ in $\mathcal{P}A_n$ starting at $v$ by $\mathcal{F}_k(v)$. This means that there exists a path $\vec{\pi}$ with $u = \pi_0 \leftrightarrow \cdots \leftrightarrow \pi_l$, such that $\pi_0 \geq g_0, \ldots, \pi_{l-1} \geq g_{l-1}$, but $\pi_l < g_l$, i.e., a path that traverses the threshold after exactly $l$ steps. For fixed vertices $u, v \geq g_0$, we thus obtain

$$
\mathbb{P}(\text{dist}_{\mathcal{P}A_n}(u, v) \leq 2k) \leq \sum_{l=1}^{k} \mathbb{P}(\mathcal{F}_l(u)) + \sum_{l=1}^{k} \mathbb{P}(\mathcal{F}_l(v)) + \sum_{k=1}^{2k} \mathbb{P}(E_l(u, v)). \quad (8.5.9)
$$

The truncated first moment estimate arises when bounding the events of the existence of certain good or bad paths by their expected number. Due to the split in good and bad paths, these sums will now behave better than without this split.

Equation (8.5.9) is identical to the inequality (6.2.31) used in the proof of Theorem 6.6. However, the notion of good has changed due to the fact that vertices no longer have a weight, but rather an age, and vertices that have appeared early in $\mathcal{P}A_n$ are the most likely to have large degrees. This explains why good vertices have high indices for $\mathcal{P}A_n$, while good vertices have high weights for $\mathcal{N}R_n(w)$.

By assumption,

$$
\mathbb{P}(\vec{\pi} \subseteq \mathcal{P}A_n) \leq p(\vec{\pi}) \quad (8.5.10)
$$

so that for $v \geq \ell_0$ and $l \in [k]$, and with $\vec{\pi} = (\pi_0, \ldots, \pi_l)$ with $\pi_0 = u$,

$$
\mathbb{P}(\mathcal{F}_l(u)) \leq \sum_{\pi_1 = g_1}^{n} \cdots \sum_{\pi_{l-1} = g_{l-1}}^{n} \sum_{\pi_l = 1}^{g_l-1} p(\vec{\pi}). \quad (8.5.11)
$$

Given $\varepsilon > 0$ we choose $g_0 = \lfloor \varepsilon n \rfloor$ and $(g_j)_{j=0,\ldots,k}$ decreasing fast enough so that the first two summands on the right hand side of (8.5.9) together are no larger than $2\varepsilon$. For $l \in [k]$ and $u, v \in [n]$, we set

$$
f_{l,n}(u, v) := \mathbb{I}_{\{v \geq g_l\}} \sum_{\pi_1 = g_1}^{n} \cdots \sum_{\pi_{l-1} = g_{l-1}}^{n} p(u, \pi_1, \ldots, \pi_{l-1}, v), \quad (8.5.12)
$$

and set $f_{0,n}(u, v) = \mathbb{I}_{\{v = u\}} \mathbb{I}_{\{u \leq g_1\}}$. To rephrase the truncated first moment estimate in terms of $f_{l,n}$, note that $p$ is symmetric so that, for all $l \leq 2k$,

$$
\mathbb{P}(E_l(u, v)) \leq \sum_{\pi_1 = g_1}^{n} \cdots \sum_{\pi_{l-1} = g_{l-1}}^{n} \sum_{\pi_{l} = 1}^{g_l-1} p(u, \pi_1, \ldots, \pi_{l-1}, \pi_l) = \sum_{\pi_{l} = 1}^{n} f_{l,n}(u, \pi_{l}) f_{l,n}(v, \pi_{l}). \quad (8.5.13)
$$

Using the recursive representation

$$
f_{k+1,n}(u, v) = \sum_{w = g_k}^{n} f_{k,n}(u, w)p(w, v), \quad (8.5.14)
$$
we establish upper bounds on $f_{k,n}(u,v)$ and use these to show that the rightmost term in (8.5.9) remains small when $k = k_n$ is chosen appropriately. This leads to the lower bounds for the typical distance in Theorem 8.17. Let us now make these ideas precise, starting with the derivation of some intermediate results, as well as bounds on the growth of $k \mapsto f_{k,n}(u,v)$.

We assume that Assumption 8.16 holds for a $\gamma \in (\frac{1}{2}, 1)$ with a fixed constant $\kappa$. Recall the definition of $f_{k,n}$ and the key estimates (8.5.9), (8.5.11) and (8.5.13), which combined give the truncated first moment bound

$$
P(\text{dist}_{PA_n}(u,v) \leq 2k_n) \leq \sum_{k=1}^{k_n} \sum_{w=1}^{g_{k-1}} f_{k,n}(u,w) + \sum_{k=1}^{k_n} \sum_{w=1}^{g_{k-1}} f_{k,n}(v,w)$$

(8.5.15)

$$+ \sum_{k=1}^{2k_n} \sum_{\pi([k/2]) = g_{[k/2]}} f_{[k/2],n}(u,\pi([k/2])) f_{[k/2],n}(v,\pi([k/2])).$$

The remaining task of the proof is to choose $k_n \in \mathbb{N}$, as well as a decreasing sequence $(g_k)_{k=0}^\infty$ such that $2 \leq g_k \leq \cdots \leq g_0 \leq n$, that allow us to bound the right-hand side of (8.5.15).

Our aim is to provide an upper bound of the form

$$f_{k,n}(u,v) \leq \alpha k v^{-\gamma} + 1_{\{v > g_k - 1\}} \beta_k v^{\gamma - 1},$$

(8.5.16)

for suitably chosen parameters $\alpha_k, \beta_k \geq 0$. Key to this choice is the following lemma:

**Lemma 8.18** (A recursive bound on $f_{k,n}(v,m)$ for $\gamma \in (\frac{1}{2}, 1)$) Let $\gamma \in (\frac{1}{2}, 1)$ and suppose that $2 \leq \ell \leq n$, $\alpha, \beta \geq 0$ and $q: [n] \rightarrow [0, \infty)$ satisfies

$$q(w) \leq 1_{\{w > \ell\}} (\alpha w^{-\gamma} + \beta w^{\gamma - 1}) \quad \text{for all } w \in [n].$$

(8.5.17)

Then there exists a constant $c = c(\gamma, \kappa) > 1$ such that, for all $u \in [n]$,

$$\sum_{w=1}^n q(w)p(w,u) \leq c (\alpha \log(n/\ell) + \beta n^{2\gamma - 1}) u^{-\gamma}$$

(8.5.18)

$$+ c 1_{\{u > \ell\}} (\alpha \ell^{1-2\gamma} + \beta \log(n/\ell)) u^{\gamma - 1}.$$

**Proof** We use (8.5.7) to rewrite

$$\sum_{w=1}^n q(w)p(w,u) = \sum_{w=u \wedge \ell}^n q(w)p(w,u) + 1_{\{u > \ell\}} \sum_{w=\ell}^{u-1} q(w)p(w,u)$$

(8.5.19)

$$= \sum_{w=u \wedge \ell}^n \kappa (\alpha w^{-\gamma} + \beta w^{\gamma - 1}) w^{\gamma - 1} u^{-\gamma}$$

$$+ 1_{\{u > \ell\}} \sum_{w=\ell}^{u-1} \kappa (\alpha w^{-\gamma} + \beta w^{\gamma - 1}) w^{-\gamma} u^{\gamma - 1}.$$
Simplifying the sums leads to

\[ \sum_{w=1}^{n} q(w)p(w, u) \leq \kappa \left( \alpha \sum_{w=\mathbb{V} \cup \ell}^{n} w^{-1} + \beta \sum_{w=\mathbb{V} \cup \ell}^{n-1} w^{2\gamma-2} \right) u^{-\gamma} \]

\[ + \kappa \mathbb{1}_{\{u > \ell\}} \left( \alpha \sum_{w=\ell}^{u-1} w^{-2\gamma} + \beta \sum_{w=\ell}^{u-1} w^{-1} \right) u^{\gamma-1} \]

\[ \leq \kappa \left( \alpha \log \left( \frac{u}{\ell - 1} \right) + \beta \frac{2\gamma - 1}{n^{2\gamma-1}} \right) u^{-\gamma} \]

\[ + \kappa \mathbb{1}_{\{u > \ell\}} \left( \frac{\alpha}{1 - 2\gamma} (\ell - 1)^{1-2\gamma} + \beta \log \left( \frac{u}{\ell - 1} \right) \right) u^{\gamma-1}. \]

This immediately implies the assertion since \( \ell \geq 2 \) and \( u \in [n] \) by assumption.

We aim to apply Lemma 8.18 iteratively. We use induction in \( k \) to prove that there exist \( (g_k)_{k \geq 0}, (\alpha_k)_{k \geq 1} \) and \( (\beta_k)_{k \geq 1} \) such that

\[ f_{k,n}(u, v) \leq \alpha_k v^{-\gamma} + \beta_k v^{\gamma-1} \quad \text{for all } u, v \in [n]. \]  

(8.5.21)

The sequences \( (g_k)_{k \geq 0}, (\alpha_k)_{k \geq 1} \) and \( (\beta_k)_{k \geq 1} \) are chosen as follows:

**Definition 8.19** (Choices of parameters \( (g_k)_{k \geq 0}, (\alpha_k)_{k \geq 1} \) and \( (\beta_k)_{k \geq 1} \)) Fix \( R > 1 \) and \( \varepsilon \in (0, 1) \). We define

\[ g_0 = \lceil \varepsilon n \rceil, \quad \alpha_1 = Rg_0^{2\gamma-1}, \quad \beta_1 = Rg_0^{-\gamma}, \]  

(8.5.22)

and recursively, for \( k \geq 1, \)

1. \( g_k \) is the smallest integer such that

\[ \frac{1}{1 - \gamma} \alpha_k g_k^{1-\gamma} \geq \frac{6 \varepsilon}{n^2 k^2}; \]  

(8.5.23)

2. \( \alpha_{k+1} = c(\alpha_k \log(n/g_k) + \beta_k n^{2\gamma-1}); \)  

(8.5.24)

3. \( \beta_{k+1} = c(\alpha_k g_k^{1-2\gamma} + \beta_k \log(n/g_k)) \),  

(8.5.25)

where \( c = c(R, \gamma) \) is the constant appearing in Lemma 8.18.

One can check that \( k \mapsto g_k \) is non-increasing, while \( k \mapsto \alpha_k, \beta_k \) are non-decreasing.

We recall that \( f_{k,n}(u, v) \) was introduced in (8.5.12), with \( p(z, w) \) defined in (8.5.7). As a consequence, \( f_{k,n} \) satisfies the recursion, for all \( k \geq 1, \)

\[ f_{k+1,n}(u, v) = \sum_{z=g_k}^{n} f_{k,n}(u, z)p(z, v). \]  

(8.5.26)

The following lemma derives recursive bounds on \( f_{k,n}: \)
Lemma 8.20 (Recursive bound on $f_{k,n}$) For the sequences in Definition 8.19, for every $l \in [n]$ and $k \in \mathbb{N}$,
\[ f_{k,n}(u, v) \leq \alpha_k v^{-\gamma} + 1_{\{v > g_{k-1}\}} \beta_k v^{\gamma-1}. \quad (8.5.27) \]

Proof We prove (8.5.27) by induction on $k$. For $k = 1$, using $\alpha_1 = R_{g_0}^{-\gamma}$ and $\beta_1 = R_{g_0}^{-\gamma}$,
\[ f_{1,n}(u, v) = p(u, v) 1_{\{u \geq g_0\}} \leq R_{g_0}^{-\gamma} v^{-\gamma} + 1_{\{v > g_0\}} R_{g_0}^{-\gamma} v^{\gamma-1} \]
\[ = \alpha_1 v^{-\gamma} + 1_{\{l > g_0\}} \beta_1 v^{\gamma-1}, \]
as required. This initiates the induction hypothesis.

We now proceed to advance the induction: suppose that $g_{k-1}$, $\alpha_k$ and $\beta_k$ are such that
\[ f_{k-1,n}(u, v) \leq \alpha_k v^{-\gamma} + 1_{\{v > g_{k-1}\}} \beta_k v^{\gamma-1}. \quad (8.5.29) \]
We use the recursive property of $f_{k,n}$ in (8.5.26). We apply Lemma 8.18, with $g = g_k$ and $q(m) = f_{k,n}(x, m) 1_{\{m \geq g_k\}}$, so, by Definition 8.19,
\[ f_{k+1,n}(u, v) \leq c \left[ \alpha_k \log(n/g_k) + \beta_k n^{2\gamma-1} \right] v^{-\gamma} + \epsilon 1_{\{v > g_k\}} \left[ \alpha_k g_k^{\gamma} + \beta_k \log(n/g_k) \right] v^{\gamma-1} \]
\[ = \alpha_{k+1} v^{-\gamma} + 1_{\{v > g_k\}} \beta_{k+1} v^{\gamma-1}, \]
as required. This advances the induction hypothesis, and thus completes the proof.

We next use (8.5.21) to prove Theorem 8.17. Summing over (8.5.27) in Lemma 8.20, and using (8.5.21) and (8.5.23) we obtain
\[ \sum_{w=1}^{g_{k-1}} f_{k,n}(v, w) \leq \frac{1}{1-\gamma} \alpha_k g_k^{1-\gamma} \leq \frac{6\epsilon}{n^2 k^2}, \quad (8.5.31) \]
which, when summed over all $k \geq 1$ is bounded by $\epsilon$. Hence the first two summands on the right-hand side in (8.5.15) together are smaller than $2\epsilon$. This shows that the probability that there exists a bad path from either $u$ or $v$ is small, uniformly in $u, v \geq \varepsilon n$. It thus remains to deal with the good paths.

For this, it remains to choose $k_\alpha$ as large as possible, while ensuring that $g_{k_\alpha} \geq 2$ and at the same time
\[ \sum_{k=1}^{2k_\alpha} \sum_{\pi_{\lfloor k/2 \rfloor} \leq g_{\lfloor k/2 \rfloor}} f_{\lfloor k/2 \rfloor,n}(u, \pi_{\lfloor k/2 \rfloor}) f_{\lfloor k/2 \rfloor,n}(v, \pi_{\lfloor k/2 \rfloor}) = o(1). \quad (8.5.32) \]
Proving (8.5.56) will be the main content of the remainder of this section.

Recall from Definition 8.19 that $g_k$ is the largest integer satisfying (8.5.23) and that the parameters $\alpha_k$, $\beta_k$ are defined via equalities in (8.5.24)–(8.5.25). To establish lower bounds for the decay of $g_k$, we instead investigate the growth of $\eta_k = n/g_k > 0$ for large $k$. For this, we first derive a recursive bound on $\eta_k$:
8.5 Typical distance in PA-models: log log-lower bound for $\delta < 0$

Lemma 8.21 (Recursive bound on $\eta_k$) Recall Definition 8.19, and let $\eta_k = n/g_k$. Then there exists a constant $C > 0$ such that

$$\eta_{k+2}^{1-\gamma} \leq C \left[ \eta_k^{1-\gamma} + \eta_{k+1}^{1-\gamma} \log \eta_{k+1} \right].$$  \hfill (8.5.33)

Proof By definition of $g_k$ in (8.5.23),

$$\eta_{k+2}^{1-\gamma} = n^{1-\gamma} g_{k+2}^{\gamma-1} \leq n^{1-\gamma} \frac{\pi^2}{1-\gamma} \frac{(k+2)^2}{6\epsilon} \alpha_{k+2}. \hfill (8.5.34)$$

By definition of $\alpha_k$ in (8.5.24),

$$n^{1-\gamma} \frac{\pi^2}{1-\gamma} \frac{c}{6\epsilon} \frac{(k+2)^2}{\pi^2(k+1)^2} \left[ \alpha_{k+1} \log \eta_{k+1} + \beta_{k+1} n^{2\gamma-1} \right]. \hfill (8.5.35)$$

We bound each of the two terms in (8.5.35) separately.

By definition of $g_k$, the relation in (8.5.23) holds with the opposite inequality if we replace $g_k$ by $g_k - 1$ in the left hand side. This, with $k + 1$ instead of $k$, yields

$$\alpha_{k+1} \leq \frac{6(1-\gamma)\epsilon}{\pi^2(k+1)^2} (g_{k+1} - 1)^{\gamma-1}. \hfill (8.5.36)$$

Since $\alpha_{k+1} \geq 2$, we must have that $g_{k+1} \geq 2$, so that

$$(g_{k+1} - 1)^{\gamma-1} \leq 2^{1-\gamma} g_{k+1}^{\gamma-1}. \hfill (8.5.37)$$

We conclude that the first term in (8.5.35) is bounded by

$$n^{1-\gamma} \frac{c}{1-\gamma} \frac{\pi^2}{6\epsilon} (k+2)^2 \alpha_{k+1} \log \eta_{k+1} \hfill (8.5.38) \leq n^{1-\gamma} \frac{c}{1-\gamma} \frac{\pi^2}{6\epsilon} (k+2)^2 \frac{6(1-\gamma)\epsilon}{\pi^2(k+1)^2} g_{k+1}^{\gamma-1} \log \eta_{k+1} \hfill (8.5.39) \leq c 2^{1-\gamma} \frac{(k+2)^2}{(k+1)^2} \eta_{k+1}^{1-\gamma} \log \eta_{k+1},$$

which is part of the second term in (8.5.33).

We now have to bound the second term in (8.5.35) and show that it is bounded by the right hand side of (8.5.33). This term equals

$$n^{1-\gamma} \frac{c}{1-\gamma} \frac{\pi^2}{6\epsilon} (k+2)^2 \beta_{k+1} n^{2\gamma-1} = \frac{c}{1-\gamma} \frac{\pi^2}{6\epsilon} \frac{(k+2)^2}{\pi^2} \beta_{k+1} n^\gamma. \hfill (8.5.40)$$

We use the definition of $\beta_k$ in (8.5.25) to write

$$c 2^{1-\gamma} \frac{(k+2)^2}{(k+1)^2} \eta_{k+1}^{1-\gamma} \log \eta_{k+1},$$

which again leads to two terms that we bound separately. For the first term in
(8.5.40), we again use the fact that \( \alpha_k \leq 2^{1-\gamma} \frac{6(1-\gamma)}{\pi k^2} g_k^{1-2\gamma} \), we arrive at that
\[
\frac{c}{1-\gamma} \frac{\pi^2}{6\epsilon} (k+2)^2 n^\gamma c\alpha_k g_k^{1-2\gamma} 
\leq \frac{c}{1-\gamma} \frac{\pi^2}{6\epsilon} (k+2)^2 n^\gamma \frac{6(1-\gamma)}{\pi k^2} g_k^{\gamma-1} g_k^{1-2\gamma} = c \frac{2^{1-\gamma} (k+2)^2}{k^2} \eta_k,
\]
which contributes to the first term on the right hand side of (8.5.33).

By Definition 8.19, we have \( c\beta_k n^{2\gamma-1} \leq \alpha_{k+1} \), so that, using (8.5.36) and (8.5.37), the second term in (8.5.40) is bounded by
\[
\frac{c}{1-\gamma} \frac{\pi^2}{6\epsilon} (k+2)^2 n^\gamma c\beta_k \log \eta_k
\leq \frac{c}{1-\gamma} \frac{\pi^2}{6\epsilon} (k+2)^2 \alpha_{k+1} n^{1-\gamma} \log \eta_k
\leq c \frac{2^{1-\gamma} (k+2)^2}{(k+1)^2} g_{k+1}^{\gamma-1} n^{1-\gamma} \log \eta_k
= c \frac{2^{1-\gamma} (k+2)^2}{(k+1)^2} \eta_{k+1}^{\gamma-1} \log \eta_k.
\]

Since \( k \mapsto g_k \) is decreasing, \( k \mapsto \eta_k \) is increasing, so that
\[
c \frac{2^{1-\gamma} (k+2)^2}{(k+1)^2} \eta_{k+1}^{\gamma-1} \log \eta_k \leq c \frac{2^{1-\gamma} (k+2)^2}{(k+1)^2} \eta_{k+1}^{1-\gamma} \log \eta_{k+1},
\]
which again contributes to the second term in (8.5.33). This proves that both terms in (8.5.40), and thus the second term in (8.5.40), are bounded by the right hand side of (8.5.33).

Putting together all the bounds and taking a sufficiently large constant \( C = C(\gamma) \), we obtain (8.5.33).

We can now obtain a useful bound on the growth of \( \eta_k \):

**Proposition 8.22** (Inductive bound on \( \eta_k \)) *Recall Definition 8.19, and let \( \eta_k = n/g_k \). Assume that \( \epsilon \in (0, 1) \). Then, there exists a constant \( B = B_\epsilon \) such that, for any \( k = O(\log \log n) \),
\[
\eta_k \leq e^{B(\tau-2)^{-k/2}},
\]
where we recall \( \tau \) from (8.5.2).

Recall that we sum over \( \pi_k = g_k \), which is equivalent to \( n/\pi_k \leq \eta_k \). The sums in (8.5.15) are such that the summands obey this bound for appropriate values of \( k \). Compare this to (6.2.31), where, instead, the weights obey the bounds \( w_{\pi_k} \leq b_k \). We see that \( \eta_k \) plays a similar role as \( b_k \). Recall indeed that \( w_{\pi_k} \) is close to the degree of \( \pi_k \) in GRG\(_m\)(w), while the degree of vertex \( \pi_k \) in PA\(_m\)(w) is close to \( (n/\pi_k)^{1/(\tau-1)} \). Thus, the truncation \( n/\pi_k \leq \eta_k \) can be interpreted as a bound \( e^{(\tau-2)^{-k/2}} \) on the degree of \( \pi_k \). Note, however, that \( b_k \approx e^{(\tau-2)^{-k}} \) by (6.2.44),
8.5 Typical distance in PA-models: log log-lower bound for $\delta < 0$

which grows roughly twice as slowly as $\eta_k$. This is again a sign that distances in $\text{PA}_n^{(m, \delta)}$ for $\delta < 0$ are twice as large as those in $\text{GRG}_n(w)$.

**Proof.** We prove the proposition by induction on $k$, and start by initializing the induction. For $k = 0$,

$$\eta_0 = n/g_0 = \frac{n}{\lfloor \varepsilon n \rfloor} \leq \varepsilon^{-1} \leq e^B, \quad (8.5.45)$$

when $B \geq \log(1/\varepsilon)$, which initializes the induction.

We next advance the induction hypothesis. Suppose that the statement is true for $l \in [k-1]$, and we will advance it to $k$. We use that

$$(z+w)^{1/(1-\gamma)} \leq 2^{1/(1-\gamma)} \left(z^{1/(1-\gamma)} + w^{1/(1-\gamma)}\right). \quad (8.5.46)$$

By Lemma 8.21, we can then write, for a different constant $C$,

$$\eta_k \leq C \left[ \frac{\eta_{k-2}^{1/(1-\gamma)}}{1-\gamma} + \eta_{k-1} (\log \eta_{k-1})^{1/(1-\gamma)} \right]$$

$$= C \left[ \frac{\eta_{k-2}^{1/(1-\gamma)}}{1-\gamma} + \eta_{k-1} (\log \eta_{k-1})^{1/(1-\gamma)} \right]. \quad (8.5.47)$$

Using this inequality, we can write

$$\eta_{k-2} \leq C \left[ \frac{\eta_{k-4}^{1/(1-\gamma)}}{1-\gamma} + \eta_{k-3} (\log \eta_{k-3})^{1/(1-\gamma)} \right], \quad (8.5.48)$$

so that, by $(z+w)^{1/(1-\gamma)} \leq 2^{1/(1-\gamma)} (z^{1/(1-\gamma)} + w^{1/(1-\gamma)})$,

$$\eta_k \leq C(2C)^{1/(1-\gamma)} \left[ \frac{\eta_{k-4}^{1/(1-\gamma)}}{1-\gamma} + \eta_{k-3} (\log \eta_{k-3})^{1/(1-\gamma)} \right]$$

$$+ C \eta_{k-1} (\log \eta_{k-1})^{1/(1-\gamma)}. \quad (8.5.49)$$

Renaming $2C$ as $C$ for simplicity, and iterating these bounds, we obtain

$$\eta_k \leq C^{k/2} \eta_0 (\tau-2)^{-k/2}$$

$$+ \sum_{i=1}^{k/2} C^{k/2} (\tau-2)^{-i} \eta_{k-2i+1} (\log \eta_{k-2i+1})^{(\tau-2)^{-i-1}}. \quad (8.5.50)$$

For the first term in (8.5.50), we use the upper bound $\eta_0 \leq 1/\varepsilon$ to obtain

$$C^{k/2} \eta_0 (\tau-2)^{-k/2} \leq C^{k/2} (\tau-2)^{-k/2} e^{B(\tau-2)^{-k/2}}$$

$$\leq \frac{1}{2} e^{B(\tau-2)^{-k/2}}, \quad (8.5.51)$$

again when $B \geq \log(1/\varepsilon)$.

For the second term in (8.5.50), we use the induction hypothesis to obtain

$$\sum_{i=1}^{k/2} C^{k/2} (\tau-2)^{-i} \eta_{k-2i+1} (\log \eta_{k-2i+1})^{(\tau-2)^{-i-1}} \leq \sum_{i=1}^{k/2} C^{k/2} (\tau-2)^{-i} e^{B(\tau-2)^{-i-1/2}}$$

$$\leq \sum_{i=1}^{k/2} C^{k/2} (\tau-2)^{-i} e^{B(\tau-2)^{-i-1/2}} \left( B(\tau-2)^{-i-1/2} \right)^{(\tau-2)^{-i-1} / (i-1)}. \quad (8.5.52)$$
We can write
\[ e^{B(\tau - 2)^{-k/2}} = e^{B(\tau - 2)^{-k/2}} e^{B(\tau - 2)^{-k/2}(\sqrt{\tau - 2} - 1)}. \]  
(8.5.53)

Since \( \sqrt{\tau - 2} - 1 < 0 \), for \( k = O(\log \log n) \), we can take \( B \) large enough such that, uniformly in \( k \geq 1 \),
\[ \sum_{i=1}^{k/2} C^{\sum_{i=1}^{n} (\tau - 2)^{-i}} e^{B(\tau - 2)^{-k/2}(\sqrt{\tau - 2} - 1)} \left[ B(\tau - 2)^{-(k-2i+1)/2} \right]^{(\tau - 2)^{-(i-1)}} < \frac{1}{2}. \]  
(8.5.54)

We can now sum the bounds in (8.5.51) and (8.5.52)–(8.5.54) to obtain
\[ \eta_k \leq (\frac{1}{2} + \frac{1}{2}) e^{B(\tau - 2)^{-k/2}} = e^{B(\tau - 2)^{-k/2}}, \]  
(8.5.55)
as required. This advances the induction hypothesis, and thus completes the proof of Lemma 8.22.

We are now ready to complete the proof of Theorem 8.17:

Completion of the proof of Theorem 8.17. Recall that we were left to prove (8.5.56), i.e., uniformly in \( u, v \geq \varepsilon n \),
\[ \sum_{k=1}^{2k_n} \sum_{\pi_{[k/2]} = g_{[k/2]}}^{n} f_{[k/2], n}(u, \pi_{[k/2]}) f_{[k/2], n}(v, \pi_{[k/2]}) = o(1). \]  
(8.5.56)

We note that, by Lemma 8.22,
\[ g_k \geq n/\eta_k \geq ne^{B(\tau - 2)^{-k/2}}. \]  
(8.5.57)

We now use (8.20) to estimate, writing \( u = \pi_{[k/2]} \),
\[ \sum_{k=1}^{2k_n} \sum_{\pi_{[k/2]} = g_{[k/2]}}^{n} f_{[k/2], n}(u, \pi_{[k/2]}) f_{[k/2], n}(v, \pi_{[k/2]}) \]  
(8.5.58)
\[ \leq \sum_{k=1}^{2k_n} \sum_{w = g_{[k/2]}}^{n} \left( \alpha_{[k/2]} w^{\gamma} + 1_{w > g_{[k/2] - 1}} \beta_{[k/2]} w^{\gamma - 1} \right) \left( \alpha_{[k/2]} w^{-\gamma} + 1_{w > g_{[k/2] - 1}} \beta_{[k/2]} w^{-\gamma - 1} \right). \]

We use that \( k \mapsto \alpha_k \) and \( k \mapsto \beta_k \) are non-decreasing, while \( k \mapsto g_k \) is non-increasing, to estimate the above as
\[ \sum_{k=1}^{2k_n} \sum_{\pi_{[k/2]} = g_{[k/2]}}^{n} f_{[k/2], n}(u, \pi_{[k/2]}) f_{[k/2], n}(v, \pi_{[k/2]}) \]  
(8.5.59)
\[ \leq \sum_{k=1}^{2k_n} \sum_{w = g_{[k/2]}}^{n} \left( \alpha_{[k/2]} w^{\gamma} + 1_{w > g_{[k/2] - 1}} \beta_{[k/2]} w^{\gamma - 1} \right)^2 \]
\[ \leq 2 \sum_{k=1}^{2k_n} \sum_{w = g_{[k/2]}}^{n} \alpha_{[k/2]}^2 w^{-2\gamma} + 1_{w > g_{[k/2] - 1}} \beta_{[k/2]}^2 w^{2(\gamma - 1)}. \]
This gives two terms that we estimate one by one. We start with, using that \( \gamma > \frac{1}{2} \) and using that \( k \mapsto \alpha_k \) is non-decreasing, while \( k \mapsto g_k \) is non-increasing,

\[
2 \sum_{k=1}^{2k_n} \sum_{w=\lfloor k/2 \rfloor}^{n} \alpha_{\lfloor k/2 \rfloor}^2 w^{-2\gamma} \leq \frac{1}{2\gamma - 1} \sum_{k=1}^{2k_n} \alpha_{\lfloor k/2 \rfloor}^2 g_{\lfloor k/2 \rfloor}^{1-2\gamma}
\]

(8.5.60)

\[
\leq \sum_{k=1}^{2k_n} \alpha_{\lfloor k/2 \rfloor}^2 g_{\lfloor k/2 \rfloor}^{1-2\gamma} = 2\frac{\eta_k}{n} \sum_{k=1}^{k_n} \alpha_{k}^2 g_k^{-2-2\gamma}.
\]

Using (8.5.23), we obtain

\[
\alpha_{k}^2 g_k^{-2-2\gamma} \leq \left( k^{-2} \frac{6\varepsilon}{\pi^2} (1 - \gamma) \right)^2,
\]

(8.5.61)

so that

\[
2 \sum_{k=1}^{2k_n} \sum_{w=\lfloor k/2 \rfloor}^{n} \alpha_{\lfloor k/2 \rfloor}^2 w^{-2\gamma} \leq C\frac{\eta_{k+1}}{n} \sum_{k\geq 1} \varepsilon^2 k^{-2} \leq C\frac{\eta_{k+1} \varepsilon^2}{n}.
\]

(8.5.62)

This bounds the first term on the right hand side of (8.5.59).

For the second term on the right hand side of (8.5.59), we again use that \( k \mapsto g_k \) is non-increasing, to obtain

\[
2 \sum_{k=1}^{2k_n} \sum_{w=\lfloor k/2 \rfloor+1}^{n} \beta_{\lfloor k/2 \rfloor}^2 w^{-2(\gamma-1)} \leq \frac{4}{2\gamma - 1} \sum_{k=1}^{k_n} \beta_k n^{2\gamma-1}.
\]

(8.5.63)

By the definition of \( \alpha_k \) in (8.5.24), we get \( \beta_k n^{2\gamma-1} \leq \alpha_{k+1} \). Thus,

\[
\sum_{k=1}^{k_n} \beta_k^2 n^{2\gamma-1} \leq \sum_{k=1}^{k_n} \alpha_{k+1}^2 n^{1-2\gamma} = \frac{1}{n} \eta_{k+1} \sum_{k=1}^{k_n} \alpha_{k+1}^2 g_{k+1}^{-2-2\gamma} \leq C\frac{\eta_{k+1} \varepsilon^2}{n},
\]

(8.5.64)

as in (8.5.62).

We conclude that, using (8.5.44) in Proposition 8.22,

\[
2 \sum_{k=1}^{2k_n} \sum_{\pi_{\lfloor k/2 \rfloor} = \pi_{\lfloor k/2 \rfloor}}^{n} f_{\lfloor k/2 \rfloor, n}(u, \pi_{\lfloor k/2 \rfloor}) f_{\lfloor k/2 \rfloor, n}(v, \pi_{\lfloor k/2 \rfloor})
\]

(8.5.65)

\[
\leq C\frac{\varepsilon^2 \eta_{k+1}}{n} \leq C\frac{\varepsilon^2}{n} u^{B(\tau - 2) - k_{n}/2}.
\]

Hence, choosing

\[
k_n \leq \frac{2 \log \log n}{\log (\tau - 2)} - K,
\]

(8.5.66)

we obtain that, for \( \varepsilon > 0 \) and \( K = K_\varepsilon \) sufficiently large,

\[
P(\text{dist}_{PA_n}(u, v) \leq 2k_n) \leq 2\varepsilon + \varepsilon,
\]

(8.5.67)

whenever \( u, v \geq g_0 = \lceil \varepsilon n \rceil \). Note from (8.5.44) in Proposition 8.22 that this choice also ensures that \( g_{k_n} = n/\eta_{k_n} \geq 2 \) when \( K \) is taken sufficiently large. This implies the statement of Theorem 8.17.
8.6 Small-world PAMs: logarithmic upper bounds for $\delta \geq 0$

In this section, we derive the upper bounds on the typical distances as stated in Section 8.2 for $\delta \geq 0$. This section is organised as follows. In Section 8.6.1, we start by investigating the logarithmic upper bounds in Theorem 8.6. We close in Section 8.6.2 by studying the upper bound on typical distances for $\delta = 0$ in Theorem 8.7.

8.6.1 Logarithmic upper bounds for $\delta > 0$

In this section we prove lower bounds on distances in $\text{PA}_n^{(m,\delta)}$. We start by proving a relatively simple logarithmic bound on the typical distances for $(\text{PA}_n^{(m,\delta)}(b))_{n \geq 1}$, after which we extend the result to the optimal constant, which requires more work:

Proof of a logarithmic upper bound in Theorem 8.6. We start by proving a logarithmic bound on the diameter of $(\text{PA}_n^{(m,\delta)}(b))_{n \geq 1}$. Since $(\text{PA}_n^{(m,\delta)}(b))_{n \geq 1}$ is obtained from $(\text{PA}_n^{(1,\delta/m)(b)})_{n \geq 1}$ by collapsing $m$ successive vertices, $\text{diam}(\text{PA}_n^{(m,\delta)}(b)) \preceq \text{diam}(\text{PA}_n^{(1,\delta/m)(b)})$, and the result follows from Theorem 8.3.

We next extend this result to a logarithmic bound on typical distances for $(\text{PA}_n^{(m,\delta)}(b))_{n \geq 1}$. We make use of various results from Chapter 5. We start by using the local weak convergence result in Theorem 5.24. This implies that $|B_r(n,\delta^1)|$ and $|B_r(n,\delta^2)|$ converge to two independent sizes of the $r$th generations of the Pólya Point Tree, which is a multi-type branching process. Denote the limits by $Z_r^{(1)}$ and $Z_r^{(2)}$. Since this multi-type branching process is supercritical, each will contain a large number of vertices when $r$ is large. We follow the first edge of the vertices in the boundaries of $B_r(n,\delta^1)$ and $B_r(n,\delta^2)$, and subsequently follow these edges leading to older vertices. By Theorem 5.5, these paths will end in vertex 1 with strictly positive probability, conditionally independently given $U'$. As a result, whp, one of these paths will indeed end in vertex 1 whp. We conclude that, for $r > 1$ large, the diameter of $(\text{PA}_n^{(m,\delta)}(b))_{n \geq 1}$ is at most $2r$ plus the diameter of the tree containing vertex 1 in $\text{PA}_n^{(1,\delta/m)(b)}$, which is of logarithmic order again by Theorem 8.3 and its extension to $(\text{PA}_n^{(1,\delta)}(b))_{n \geq 1}$. This completes the proof.

8.7 Small-world effect in PAMs: loglog upper bounds for $\delta < 0$

In this section we investigate $\text{PA}_n^{(m,\delta)}$ with $m \geq 2$ and $\delta \in (-m, 0)$ and prove the a first step towards the upper bound in Theorem 8.8.

The proof of Theorem 8.8 is divided into two key steps. In the first, in Theorem 8.23 in Section 8.7.1, we give a bound on the diameter of the core which consists of the vertices with degree at least a certain power of $\log n$. This argument is close in spirit to the argument in Reittu and Norros (2004) used to prove bounds on typical distances in the configuration model, but substantial adaptations are necessary to deal with preferential attachment, an argument that was already adapted to the configuration model in Theorem 7.9. After this, in Theorem 8.28,
we derive a bound on the distance between vertices with a small degree and the core. We start by defining and investigating the core of the preferential attachment model. In the sequel, it will be convenient to prove Theorem 8.8 for $2n$ rather than for $n$. Clearly, this does not make any difference for the results.

8.7.1 The diameter of the core for $\delta < 0$

We adapt the proof of Theorem 7.9 to $\text{PA}_{2n}^{(m, \delta)}$. We recall that

$$\tau = 3 + \frac{\delta}{m},$$

so that $-m < \delta < 0$ corresponds to $\tau \in (2, 3)$. Throughout this section, we fix $m \geq 2$.

We take $\sigma = \frac{1}{3 - \tau} = -\frac{m}{\tau} > 1$ and define the core $\text{Core}_n$ of the PA-model $\text{PA}_{2n}^{(m, \delta)}$ to be

$$\text{Core}_n = \{ i \in [n] : D_i(n) \geq (\log n)\sigma \},$$

i.e., all the vertices in $[n]$ which at time $n$ have degree at least $(\log n)\sigma$.

Let us explain the philosophy of the proof. Note that $\text{Core}_n$ only requires information of $\text{PA}_{2n}^{(m, \delta)}$, while we will study its diameter in $\text{PA}_{2n}^{(m, \delta)}$. This will allow us to use the edges originating from vertices in $[2n] \setminus [n]$ as a sprinkling of the graph that will create shortcuts in $\text{PA}_{2n}^{(m, \delta)}$. Such shortcuts will tremendously shorten the distances. We will call the vertices that create such shortcuts $n$-connectors.

Basically, this argument will show that a vertex $v \in [n]$ of large degree $D_v(n) \gg 1$, will likely have an $n$-connector to a vertex $u \in [n]$ satisfying that $D_u(n) \geq D_v(n)^{1/(\tau - 2)}$. Thus, it takes two steps to link a vertex of large degree to another vertex of even larger degree. In the proof for the configuration model in Theorem 7.9, this happened in only one step. Therefore, distances in $\text{PA}_{2n}^{(m, \delta)}$ are (at least in terms of upper bounds) twice as large as the corresponding ones for a configuration model with similar degree structure. Let us now state our main result, for which we need to define some notation.

For a graph $G$ with vertex set $[n]$ and a given edge set, recall that we write $\text{dist}_{i,j}$ for the shortest-path distance between $i$ and $j$ in the graph $G$. Also, for $A \subseteq [n]$, we write

$$\text{diam}_n(A) = \max_{i,j \in A} \text{dist}_{i,j}(A).$$

Then, the diameter of the core in the graph $\text{PA}_{2n}^{(m, \delta)}$, which we denote by $\text{diam}_{2n}(\text{Core}_n)$, is bounded in the following theorem:

**Theorem 8.23 (Diameter of the core)** Fix $m \geq 2$ and $\delta \in (-m, 0)$. For every $\sigma > \frac{1}{3 - \tau}$, whp there exists a $K = K_{\delta} > 0$ such that

$$\text{diam}_{2n}(\text{Core}_n) \leq \frac{4\log \log n}{|\log (\tau - 2)|} + K.$$
The proof of Theorem 8.23 is divided into several smaller steps. We start by proving that the diameter of the inner core $\text{Inner}_n$, which is defined by

$$\text{Inner}_n = \{i \in [n]: D_i(n) \geq n^{\frac{1}{1+m}} (\log n)^{-\frac{1}{2}}\}, \quad (8.7.5)$$

is whp bounded by some finite constant $K_\delta < \infty$. After this, we will show that the distance from the outer core, which is $\text{Outer}_n = \text{Core}_n \setminus \text{Inner}_n$, to the inner core can be bounded by a fixed constant times $\log \log n$. This also shows that the diameter of the outer core is bounded by a different constant times $\log \log n$. We now give the details, starting with the diameter of the inner core:

**Proposition 8.24 (Diameter of the inner core)** Fix $m \geq 2$ and $\delta \in (-m, 0)$. Then, whp,

$$\text{diam}_{2n}(\text{Inner}_n) < K_\delta. \quad (8.7.6)$$

Before proving Proposition 8.24, we first introduce the important notion of an $n$-connector between a vertex $i \in [n]$ and a set of vertices $A \subseteq [n]$, which plays a crucial role throughout the proof. Fix two sets of vertices $A$ and $B$. We say that the vertex $j \in [2n] \setminus [n]$ is an $n$-connector between $A$ and $B$ if one of the edges incident to $j$ connects to $A$ and another edge incident to $j$ connects to a vertex in $B$. Thus, when there exists an $n$-connector between $A$ and $B$, the distance between $A$ and $B$ in $\text{PA}_{2n}^{(m, \delta)}$ is at most 2. The following lemma gives bounds on the probability of an $n$-connector not existing:

**Lemma 8.25 (Connectivity sets in $\text{PA}_{2n}^{(m, \delta)}$)** Fix $m \geq 2$ and $\delta \in (-m, 0)$. For any two sets of vertices $A, B \subseteq [n]$, there exists $\eta > 0$ such that

$$\mathbb{P}(\text{no } n\text{-connector for } A \text{ and } B \mid \text{PA}_{n}^{(m, \delta)}) \leq e^{-\eta D_A(n) D_B(n)/n}, \quad (8.7.7)$$

where, for any $A \subseteq [n],

$$D_A(n) = \sum_{i \in A} D_i(n), \quad (8.7.8)$$

denotes the total degree of vertices in $A$ at time $n$.

**Proof** We note that for two sets of vertices $A$ and $B$, conditionally on $\text{PA}_n^{(m, \delta)}$, the probability that $j \in [2n] \setminus [n]$ is an $n$-connector for $A$ and $B$ is at least

$$\frac{(D_A(n) + \delta|A|)(D_B(n) + \delta|B|)}{[2n(2m + \delta)]^2}, \quad (8.7.9)$$

independently of the fact whether the other vertices are $n$-connectors or not.

Since $D_i(n) + \delta \geq m + \delta > 0$ for every $i \leq n$, and $\delta < 0$, for every $i \in B,$

$$D_i(n) + \delta = D_i(n) \left(1 + \frac{\delta}{D_i(n)}\right) \geq D_i(n) \left(1 + \frac{\delta}{m}\right) = D_i(n) \frac{m + \delta}{m}, \quad (8.7.10)$$

and, thus, also $D_A(n) + \delta|A| \geq D_A(n) \frac{m + \delta}{m}$. As a result, for $\eta = (m + \delta)^2/(2m(2m + \delta)) > 0$, the probability that $j \in [2n] \setminus [n]$ is an $n$-connector for $A$ and $B$ is at least $\frac{\eta D_A(n) D_B(n)}{n^2}$, independently of the fact whether the other vertices are $n$-connectors.
or not. Therefore, the probability that there is no $n$-connector for $A$ and $B$ is, conditionally on $\text{PA}_n^{(m,i)}$, bounded above by
\[
\left(1 - \frac{\eta D_A(n)D_B(n)}{n^2}\right)^n \leq e^{-\eta D_A(n)D_B(n)/n},
\]
(8.7.11)
as required.

We make use of Lemma 8.25 in several places throughout the proof, particularly with $B = \{i\}$. The bound in Lemma 8.25 replaces the related bound in Lemma 7.11 for the configuration model.

We now give the proof of Proposition 8.24:

Proof of Proposition 8.24. From [Volume 1, Theorem 8.3] whp, $\text{Inner}_n$ contains at least $\sqrt{n}$ vertices and denote the first $\sqrt{n}$ vertices of $\text{Inner}_n$ by $I$. Observe that $n^{(r-1)^{-1}} \downarrow 0$ for $r > 2$, so that, for any $i, j \in I$, the probability that there exists an $n$-connector for $i$ and $j$ is bounded below by
\[
1 - \exp\{-\eta n^{(r-1)^{-1}(\log n)^{-1}}\} \geq p_n \equiv n^{(r-1)^{-1}(\log n)^{-2}},
\]
(8.7.12)
for $n$ sufficiently large.

We wish to couple $\text{Inner}_n$ with an Erdős-Rényi random graph with $N_n = \sqrt{n}$ vertices and edge probability $p_n$, which we denote by $\text{ER}_{N_n}(p_n)$. For this, for $i, j \in [N_n]$, we say that an edge between $i$ and $j$ is present when there exists an $n$-connector connecting the $i$th and $j$th vertex in $I$. We now prove that this graph is bounded below by $\text{ER}_{N_n}(p_n)$. Note that (8.7.12) does not guarantee this coupling, instead we should prove that the lower bound holds uniformly, when $i$ and $j$ belong to $I$. For this, we order the $N_n(N_n - 1)/2$ edges in an arbitrary way, and bound the conditional probability that the $l$th edge is present conditionally on the previous edges from below by $p_n$, for every $l$. This would prove the claimed stochastic domination by $\text{ER}_{N_n}(p_n)$.

Indeed, the $l$th edge is present precisely when there exists an $n$-connector connecting the corresponding vertices which we call $i$ and $j$ in $I$. Moreover, we shall not make use of the first vertices which were used to $n$-connect the previous edges. This removes at most $N_n(N_n - 1)/2 \leq n/2$ possible $n$-connectors, after which at least another $n/2$ remain. The probability that one of them is an $n$-connector for the $i$th and $j$th vertex in $I$ is, for $n$ sufficiently large, bounded below by
\[
1 - \exp\{-\eta n^{(r-2)(\log n)^{-1}}\} = 1 - \exp\{-\eta n^{(r-1)(\log n)^{-1}/2}\}
\]
(8.7.13)
\[
\geq p_n \equiv n^{(r-1)/(\log n)^{-2}},
\]
using $1 - e^{-x} \geq x/2$ for $x \in [0, 1]$ and $\eta/2 \geq \log n^{-1}$ for $n$ sufficiently large.

This proves the claimed stochastic domination of the random graph on the vertices $I$ and $\text{ER}(N_n, p_n)$. Next, we show that $\text{diam}(\text{ER}_{N_n}(p_n))$ is, whp, bounded by a uniform constant.

For this we use the result in (Bollobás, 2001, Corollary 10.12), which gives sharp bounds on the diameter of an Erdős-Rényi random graph. Indeed, this result implies that if $p^d n^{d-1} - 2 \log n \to \infty$, while $p^{d-1} n^{d-2} - 2 \log n \to -\infty$,
then \( \text{diam}(\text{ER}_n(p)) = d \), whp. In our case, \( n = N_n = n^{1/2} \) and \( p = p_n = n^{\frac{1}{2} - \frac{1}{3}} (\log n)^{-2} \), which implies that, whp, \( \frac{\tau - 1}{3 - \tau} < d \leq \frac{\tau - 1}{3 - \tau} + 1 \). Thus, we obtain that the diameter of \( I \) in \( \text{PA}_{2n}^{(m, \delta)} \) is whp bounded by \( 2d \leq 2(\frac{\tau - 1}{3 - \tau} + 1) \).

We finally show that for any \( i \in \text{Inner}_n \setminus I \), the probability that there does not exist an \( n \)-connector between \( i \) and \( I \) is small. Indeed, this probability is, since \( D_i(n) \geq \sqrt{n \frac{\tau - 1}{3 - \tau}} (\log n)^{-1/2} \), and \( D_i(n) \geq n^{\frac{1}{2} - \frac{1}{3}} (\log n)^{-1/2} \), the probability of there not existing an \( n \)-connector between \( i \) and \( I \) is bounded above by \( e^{-n^{1/(\tau - 1) - 1/2}(\log n)^{-1}} \), which is tiny since \( \tau < 3 \). This proves that whp the distance between any vertex \( i \in \text{Inner}_n \setminus I \) and \( I \) is bounded by 2, and, together with the fact that \( \text{diam}_{2n}(I) \leq 2(\frac{\tau - 1}{3 - \tau} + 1) \) thus implies that \( \text{diam}_{2n}(\text{Inner}_n) \leq 2(\frac{\tau - 1}{3 - \tau} + 2) \equiv K_4 \).

We proceed by studying the distances between the outer core \( \text{Core}_n \setminus \text{Inner}_n \) and \( \text{Inner}_n \) in the following proposition, which is the main ingredient in the proof:

**Proposition 8.26** (Distance between outer and inner core) Fix \( m \geq 2 \) and \( \delta \in (−m, 0) \). The inner core \( \text{Inner}_n \) can whp be reached from any vertex in the outer core \( \text{Outer}_n \) using no more than \( \frac{2 \log \log n}{|\log (\tau - 2)|} \) edges in \( \text{PA}_{2n}^{(m, \delta)} \), i.e., whp,

\[
\max_i \min_j \text{dist}_{\text{PA}_{2n}^{(m, \delta)}}(i, j) \leq \frac{2 \log \log n}{|\log (\tau - 2)|}.
\]

**Proof** Recall that

\[
\text{Outer}_n = \text{Core}_n \setminus \text{Inner}_n.
\]

and define

\[
\Gamma_1 = \text{Inner}_n = \{i: D_i(n) \geq u_1\},
\]

where

\[
u_1 = n^{\frac{1}{2} - \frac{1}{3}} (\log n)^{-1/2}.
\]

We now recursively define a sequence \( u_k \), for \( k \geq 2 \), so that for any vertex \( i \in [n] \) with degree at least \( u_k \), the probability that there is no \( n \)-connector for the vertex \( i \) and the set

\[
\Gamma_{k-1} = \{j: D_j(n) \geq u_{k-1}\},
\]

conditionally on \( \text{PA}_{n}^{(m, \delta)} \), is tiny, as we will show in Lemma 8.27 below. According to Lemma 8.25 and [Volume 1, Exercise 8.20], this probability is at most

\[
\exp \left\{ - \frac{n \eta B [u_{k-1}]^{2 - \tau} u_k}{n} \right\}.
\]

To make this sufficiently small, we define

\[
u_k = D \log n (u_{k-1})^{\tau - 2},
\]

with \( D \) exceeding \( (\eta B)^{-1} \). Note that the recursion in (8.7.20) is identical to that in (7.2.37). Therefore, by Lemma 7.10,

\[
u_k = D^{\alpha_k} (\log n)^{b_k} n^{c_k},
\]
8.7 Small-world effect in PAMs: loglog upper bounds for $\delta < 0$

where
\[
c_k = \frac{(\tau - 2)^{k-1}}{2(\tau - 1)}, \quad b_k = \frac{1 - (\tau - 2)^{k-1}}{3 - \tau} - \frac{1}{2}(\tau - 2)^{k-1}, \quad a_k = \frac{1 - (\tau - 2)^{k-1}}{3 - \tau}.
\] (8.7.22)

The key step in the proof of Proposition 8.26 is the following lemma:

**Lemma 8.27** (Connectivity between $\Gamma_{k-1}$ and $\Gamma_k$)
Fix $m, k \geq 2$ and $\delta \in (-m, 0)$. Then the probability that there exists an $i \in \Gamma_k$ that is not at distance two from $\Gamma_{k-1}$ in $PA_{2n}^{m, \delta}$ is $o(n^{-1})$.

**Proof.** We note that, by [Volume 1, Exercise 8.20], with probability exceeding $1 - o(n^{-1})$, for all $k$,
\[
\sum_{i \in \Gamma_{k-1}} D_i(n) \geq B \alpha \left( u_{k-1} \right)^{2-\tau} - \tau.
\] (8.7.23)

On the event that the bounds in (8.7.23) hold, we obtain by Lemma 8.25 that the conditional probability, given $PA_{2n}^{m, \delta}$, that there exists an $i \in \Gamma_k$ such that there is no $n$-connector between $i$ and $\Gamma_{k-1}$ is bounded, using Boole’s inequality, by
\[
n e^{-\eta B \alpha u_{k-1}} = n e^{-\eta B D \log n} = o(n^{-1}),
\] (8.7.24)

where we have used (8.7.20) and we have taken $D > 2(\eta B)^{-1}$.

We now complete the proof of Proposition 8.26. Fix
\[
k^\star = \left\lfloor \frac{\log \log n}{\log (\tau - 2)} \right\rfloor.
\] (8.7.25)

As a result of Lemma 8.27, we have that the distance between $\Gamma_{k^\star}$ and $\text{Inner}_n$ is at most $2k^\star$. Therefore, we are done when we can show that
\[
\text{Outer}_n \subseteq \{ i : D_i(n) \geq (\log n)^\sigma \} \subseteq \Gamma_{k^\star} = \{ i : D_i(n) \geq u_{k^\star} \},
\] (8.7.26)

so that it suffices to prove that $(\log n)^\sigma \geq u_{k^\star}$, for any $\sigma > \frac{1}{3-\tau}$. For this, we note that, by Lemma 7.10 as noted in (8.7.21),
\[
u_{k^\star} = D_{\nu_{k^\star}} \cdot (\log n)^{\nu_{k^\star}} n^{o(1)}.
\] (8.7.27)

We compute that $n^{\nu_{k^\star}} = O(1) = (\log n)^{o(1)}$, $(\log n)^{\nu_{k^\star}} = (\log n)^{1/3-\tau + o(1)}$, and $D_{\nu_{k^\star}} = (\log n)^{o(1)}$. Thus,
\[
u_{k^\star} = (\log n)^{1/3-\tau + o(1)},
\] (8.7.28)

so that, by picking $n$ sufficiently large, we can make $\sigma \geq \frac{1}{3-\tau} + o(1)$. This completes the proof of Proposition 8.26.

**Proof of Theorem 8.23.** We note that whp $\text{diam}_{2n}(\text{Core}_n) \leq K_\delta + 2k^\star$, where $k^\star$ is the upper bound on $\max_{i \in \text{Outer}_n} \min_{j \in \text{Inner}_n} d_{PA_{2n}^{m, \delta}}(i, j)$ in Proposition 8.26, and we have made use of Proposition 8.24. This proves Theorem 8.23. □
8.7.2 Connecting the periphery to the core for $\delta < 0$

In this section, we extend the results of the previous section and, in particular, study the distance between the vertices not in the core $\text{Core}_n$ and the core. The main result in this section is the following theorem:

**Theorem 8.28** (Connecting the periphery to the core)  
Fix $m \geq 2$. For every $\sigma > \frac{1}{3-\tau}$, whp, the distance between a uniformly chosen vertex $o_1 \in [2n]$ and $\text{Core}_n$ in $\text{PA}_{2n}^{(m,\delta)}$ is whp bounded from above by $C \log \log \log n$ for some $C > 0$.

Together with Theorem 8.23, Theorem 8.28 proves the upper bound in Theorem 8.8:

**Proof of the upper bound in Theorem 8.8.** Choose $o_1, o_2 \in [2n]$ uniformly at random. We bound, using the triangle inequality,

$$\text{dist}_{\text{PA}_{2n}^{(m,\delta)}}(o_1, o_2) \leq \text{dist}_{\text{PA}_{2n}^{(m,\delta)}}(o_1, \text{Core}_n) + \text{dist}_{\text{PA}_{2n}^{(m,\delta)}}(o_2, \text{Core}_n) + \text{diam}(\text{Core}_n).$$

By Theorem 8.28, the first two terms are each whp bounded by $C \log \log \log n$. Further, by Theorem 8.23, the third term is bounded by $(1 + o(1)) \frac{4 \log \log n}{|\log(\tau-2)|}$, which completes the proof of the upper bound in Theorem 8.8.

Exercise 8.9 shows that $\text{dist}_{\text{PA}_{2n}^{(m,\delta)}}(o_1, o_2) - 2 \log \log n / |\log(\tau-2)|$ is upper tight when $\text{dist}_{\text{PA}_{2n}^{(m,\delta)}}(o_1, \text{Core}_n)$ is tight. Exercise 8.10 investigates the tightness of $\text{dist}_{\text{PA}_{2n}^{(m,\delta)}}(o_1', \text{Core}_n)$ when $o_1$ is chosen uniformly at random in $[n]$.

**Proof of Theorem 8.28.** We use the same ideas as in the proof of Theorem 8.23, but now start from a vertex of large degree at time $n$ instead. We need to show that, for fixed $\varepsilon > 0$, a uniformly chosen vertex $o \in [(2-\varepsilon)n]$ can be connected to $\text{Core}_n$ using no more than $C \log \log \log n$ edges in $\text{PA}_{2n}^{(m,\delta)}$. This is done in two steps.

In the first step, we explore the neighborhood of $o$ in $\text{PA}_{2n}^{(m,\delta)}$ until we find a vertex $v_0$ with degree $D_{v_0}(n) \geq u_0$, where $u_0$ will be determined below. Denote the set of all vertices in $\text{PA}_{2n}^{(m,\delta)}$ that can be reached from $o$ using exactly $k$ different edges from $\text{PA}_{2n}^{(m,\delta)}$ by $S_k$. Denote the first $k$ for which there is a vertex in $S_k$ with arbitrarily small location, and thus also with arbitrarily large degree at time $n$ by

$$T_u^{(o)} = \inf \{ k : S_k \cap \{ v : D_v(n) \geq u \} \neq \emptyset \}.$$  

Recall the local weak convergence in Theorem 5.24, as well as the fact that each vertex $v$ has $m$ older neighbors $v_1, \ldots, v_m$, whose locations $x_{v_1}, \ldots, x_{v_m}$ are uniform on $[0, x_v]$, where $x_v$ is the location of $v$. Therefore, whp, there will be a vertex in $S_k$ with arbitrarily small location, and thus also with arbitrarily large degree at time $n$. As a result, there exists a $C = C_{u,\varepsilon}$ such that, for sufficiently large $n$,

$$\mathbb{P}(T_u^{(o)} \geq C_{u,\varepsilon}) \leq \varepsilon.$$  

(8.7.31)
8.8 Diameters in preferential attachment models

The second step is to show that a vertex $v_0$ satisfying $D_{v_0}(n) \geq u_0$ for sufficiently large $u_0$ can be joined to the core by using $O(\log \log \log n)$ edges. To this end, we apply Lemma 8.25, to obtain, for any vertex $a$ with $D_a(n) \geq w_b$, the probability that there does not exist a vertex $b$ with $D_b(n) \geq w_b$, conditionally on $PA_{m,a}^{(m,a)}$, is at most

$$\exp \left\{ -\eta D_a(n) D_B(n) / n \right\},$$

(8.7.32)

where $B = \{b: D_b(n) \geq w_b\}$. Since, as in (8.7.23),

$$D_B(n) \geq w_b \# \{b: D_b(n) \geq w_b\} \equiv w_b N_{\geq w_b}(n) \geq c n w_b^{2-\tau},$$

(8.7.33)

we thus obtain that the probability that such a $b$ does not exists is at most

$$\exp \left\{ -\eta’ w_a w_b^{2-\tau} \right\},$$

(8.7.34)

where $\eta’ = \eta c$. Fix $\varepsilon > 0$ such that $(1 - \varepsilon)/(\tau - 2) > 1$. We then iteratively take $u_k = u_{k-1}^{(1-\varepsilon)/(\tau-2)}$, to see that the probability that there exists a $k$ for which there does not exist a $v_k$ with $D_{v_k}(n) \geq u_k$ is at most

$$\sum_{l=1}^{k} \exp \left\{ -\eta’ u_{l-1} \right\}.$$  

(8.7.35)

Since

$$u_k = u_0^{\kappa}, \quad \text{where} \quad \kappa = (1 - \varepsilon)/(\tau - 2) > 0,$$  

(8.7.36)

we obtain that the probability that there exists a $k$ for which there does not exist a $v_k$ with $D_{v_k}(n) \geq u_k$ is at most

$$\sum_{l=1}^{k} \exp \left\{ -\eta’ u_0^{\kappa l-1} \right\}.$$  

(8.7.37)

Now fix $k = k_n = C \log \log \log n$ and choose $u_0$ so large that

$$\sum_{l=1}^{k_n} \exp \left\{ -\eta’ u_0^{\kappa l-1} \right\} \leq \varepsilon/2.$$  

(8.7.38)

Then we obtain that, with probability at least $\varepsilon/2$, $v_0$ is connected in $k$ steps to a vertex $v_{k_n}$ with $D_{v_{k_n}}(n) \geq u_0^{\kappa n}$ is at most $\varepsilon/2$. Since, for $C \geq 1/\log \kappa$,

$$u_0^{\kappa n} \geq u_0^{\log \log n} \geq (\log n)^\sigma,$$  

(8.7.39)

when $\log u_0 \geq \sigma$, we obtain that $v_{k_n} \in \text{Core}_{n}$ whp.

8.8 DIAMETERS IN PREFERENTIAL ATTACHMENT MODELS

In this section, we investigate the diameter in preferential attachment models. We start by discussing logarithmic bounds for $\delta > 0$, continue with the doubly-logarithmic bounds for $\delta < 0$, and finally discuss the case where $\delta = 0$. 

\[\square\]
The following theorem shows that for $\delta > 0$, the diameter grows logarithmically in the size of the graph:

**Theorem 8.29** (A $\log n$ bound for the diameter in PAMs) Fix $m \geq 1$ and $\delta > 0$. For $\text{PA}^{(m, \delta)}_n$, there exist $0 < b_1 < b_2 < \infty$ such that, as $n \to \infty$,

$$
\mathbb{P}(b_1 \log n \leq \text{diam}(\text{PA}^{(m, \delta)}_n) \leq b_2 \log n) = 1 - o(1) \quad (8.8.1)
$$

Similar bounds hold for $\text{PA}^{(m, \delta)}_n(b)$.  

**Proof** The lower bound on the diameter of $\text{PA}^{(m, \delta)}_n$ follows from the lower bound in Theorem 8.6. The upper bound for $\text{PA}^{(m, \delta)}_n(b)$ has already been proved in Section 8.6.1. We omit the proof for $\text{PA}^{(m, \delta)}_n$, which can be completed along similar lines as in Section 8.6.1. 

**Doubly-logarithmic bounds on the diameter**

When $\delta < 0$, we know that the typical distances grow like $\log \log n$. The following theorem shows that this extends to the diameter of $\text{PA}^{(m, \delta)}_n$:

**Theorem 8.30** (Diameter of $\text{PA}^{(m, \delta)}_n$ for $\delta < 0$) Fix $m \geq 2$ and $\delta \in (-m, 0)$. For $\text{PA}^{(m, \delta)}_n$, as $n \to \infty$,

$$
\frac{\text{diam}(\text{PA}^{(m, \delta)}_n)}{\log \log n} \to \frac{4}{|\log (\tau - 2)|} + \frac{2}{\log m} \quad (8.8.2)
$$

Theorem 8.30 is an adaptation of Theorem 7.17 to $\text{PA}^{(m, \delta)}_n$. Indeed, the first term on the right-hand side of corresponds to the typical distances as in Theorem 8.17, just like the first term on the right-hand side of (7.4.6) in Theorem 7.17 corresponds to the typical distances in Theorem 7.2. Further, the additional term $2/\log m$ can be interpreted as $2/\log (d_{\text{min}} - 1)$, where $d_{\text{min}}$ is the minimal degree of an *internal* vertex in the neighborhood of a vertex in $\text{PA}^{(m, \delta)}_n$, as in Theorem 7.17. In turn, this can be interpreted as twice $1/\log m$, where $\log \log n/\log m$ can be viewed as the depth of the worst trap.

For the upper bound, we explore the $r$-neighborhoods of vertices with $r = (1 + \varepsilon) \log \log n/\log m$. Then, these boundaries are so large, that with probability of order $1 - o(1/\varepsilon^2)$, pairs of such boundaries are quickly connected to the core $\text{Core}_n$. An application of Theorem 8.23 then completes the upper bound.

We give some more details on the lower bound. The proof of the lower bound proceeds by defining the notion of a vertex being $k$-minimally connected, which means that the $k$-neighborhood of the vertex is as small as possible. In this case, this means that it has size $m^k$ (rather than $d_{\text{min}}(d_{\text{min}} - 1)^{k-1}$ as it is in $\text{CM}_n(d)$). It is then shown that there are plenty of such $k$-minimally connected vertices as long as $k \leq (1 - \varepsilon) \log \log n/\log m$, similarly to the analysis in Lemma 7.18 for $\text{CM}_n(d)$. However, the analysis itself is quite a bit harder, due to the more involved nature of $\text{PA}^{(m, \delta)}_n$ compared to $\text{CM}_n(d)$. We provide a few details, and
show that the number $M_k$ of $k$-minimally connected vertices tends to infinity in probability.

We define

$$\mathcal{M}_k = \{ v \in [n] : D_v(n) = m, D_u(n) = m + 1 \; \forall u \in B_{k-1}(v) \setminus \{v\}, \quad (8.8.3)$$

$$B_k(v) \cap [n/2] = \emptyset \},$$

and

$$M_k = |\mathcal{M}_k|. \quad (8.8.4)$$

In words, $\mathcal{M}_k$ consists of those vertices whose $k$-neighborhood is minimally connected at time $n$, and only contains vertices in $[n/2]$. The following lemma shows that, for $k \leq M_k \to \infty$:

**Lemma 8.31** (Many minimally $k$-connected vertices in $PA_n^{(m,\delta)}$) Consider $PA_n^{(m,\delta)}$ and $PA_n^{(m,\delta)}(d)$ with $m \geq 2$ and $\delta \in (-m,0)$. For $k \leq (1 - \varepsilon) \log \log n / \log m$,

$$M_k \to \infty. \quad (8.8.5)$$

**Proof** We prove the lemma for $PA_n^{(m,\delta)}(d)$, and rely on the arguments in the proof of Proposition 5.23.

As in the proof of Lemma 7.18, we use a second moment method. We start by analyzing $\mathbb{E}[M_k]$. Note that

$$\mathbb{E}[M_k] = n \mathbb{P}(o \in \mathcal{M}_k) = n \sum_{t : V(t) \subseteq [n/2]} \mathbb{P}(B_k(o) = t), \quad (8.8.6)$$

where the sum is over all rooted trees $t$ with root $o$, $V(t) \subseteq [n/2]$, depth $k$, and having root degree $m$ and degree $m+1$ for all other non-leaf vertices. By the analysis in the proof of Proposition 5.23, see also (5.3.62),

$$\mathbb{P}_n(B_k(o) = t) = \prod_{v \in V(t)} \psi_{v}^{a_v'} - \prod_{u \in V(t)} (1 - \psi_{u})^{b_u'}, \quad (8.8.7)$$

where we recall $p(u,v)$ in (5.3.61), and we now let

$$a_v' = \sum_{u \in V(t)} 1_{\{v \sim s, u > s\}}, \quad b_u' = \sum_{u \in V(t)} 1_{\{u \sim v, 1 \{u \in V(t)\}}. \quad (8.8.8)$$

Using (5.3.63)-(5.3.64), this simplifies to

$$\mathbb{P}_n(B_k(o) = t) = \frac{1 + a_v(1)}{n} \prod_{v \in V(t)} \psi_{v}^{a_v'} e^{-m \sum_{u \in V(t)} p(u,v) \prod_{s=1}^{n} (1 - \psi_{s})^{b_u'}}, \quad (8.8.9)$$

where the factor $1/n$ comes from the probability that $o$ equals the root of $t$. By (5.3.65)-(5.3.66), this further simplifies to

$$\mathbb{P}_n(B_k(o) = t) = \frac{1 + a_v(1)}{n} \prod_{v \in V(t) \setminus \{o\}} \psi_{v}^{a_v'} e^{-(2m+\delta)v_s(1-\psi_s(1-x))} \prod_{s=1}^{n} (1 - \psi_{s})^{b_u'}. \quad (8.8.10)$$
Therefore, also
\[
\mathbb{P}(B_k(o) = t) = \frac{1 + o(1)}{n} \prod_{v \in V(t)} \mathbb{E}\left[\psi_{v}^{a_{v}'} e^{-(2m+d)\psi_v(1-(v/n)^{1-x})(1 - \psi_v)^{b_{v}'}}\right] \quad (8.8.11)
\]
\[
\times \prod_{s \in [n] \setminus V(t)} \frac{(\beta_s + b'_s - 1)\psi_s}{(\alpha + \beta_s + b'_s - 1)\psi_s},
\]
where, for \( v \in V(t) \), \( a_{v}' = 1 \) unless \( v \) is the root of \( t \).

It is not hard to see that there exists a \( q > 0 \) such that, uniformly in \( v \in [n] \setminus [n/2] \),
\[
n\mathbb{E}\left[\psi_{v} e^{-(2m+d)\psi_v(1-(v/n)^{1-x})(1 - \psi_v)^{b_{v}'}}\right] \geq q. \quad (8.8.12)
\]
Denote
\[
i_k = |V(t)| = \frac{n^{k+1} - 1}{m - 1}. \quad (8.8.13)
\]
Then, since \( b'_s = 0 \) for all \( s \in [n/2] \), using the argument leading to (??), there exists an \( \eta > 0 \) such that
\[
\prod_{s \in [n] \setminus V(t)} \frac{(\beta_s + b'_s - 1)\psi_s}{(\alpha + \beta_s + b'_s - 1)\psi_s} \geq \eta^{i_k}. \quad (8.8.14)
\]
Therefore, uniformly in \( t \) such that \( V(t) \subseteq [n/2] \),
\[
\mathbb{P}(B_k(o) = t) \geq n^{-i_k} (q\eta)^{i_k}. \quad (8.8.15)
\]
For every collection of vertices \( V \subseteq [n] \setminus [n/2] \), there is at least one tree \( t \) such that \( V(t) = V \). Since the total number of such sets equals \( \binom{n/2}{i_k} \), we arrive at
\[
\mathbb{E}[M_k] \geq n \left(\frac{n/2}{i_k}\right) n^{-i_k} (q\eta)^{i_k}. \quad (8.8.16)
\]
Note that, for \( k \leq (1 - \varepsilon) \log \log n / \log m \),
\[
i_k \leq \frac{m}{m - 1} \left( m^{k} \right) \leq \frac{m}{m - 1} (\log n)^{1 - \varepsilon}. \quad (8.8.17)
\]
Therefore, we can further bound from below
\[
\left( \frac{n/2}{i_k} \right) \geq \frac{(n/2 - i_k)^{i_k}}{i_k!} \geq 2^{-i_k} \frac{n^{i_k}}{i_k!}(1 + o(1)), \quad (8.8.18)
\]
to arrive at
\[
\mathbb{E}[M_k] \geq n \frac{(q\eta/2)^{i_k}}{i_k!} \geq n^{1 - \varepsilon} \to \infty. \quad (8.8.19)
\]
This proves that \( \mathbb{E}[M_k] \to \infty \).
8.8 Diameters in preferential attachment models

We continue with $E[M_k^2]$, which we write as

$$E[M_k^2] = n^2P(o_1, o_2 \in M_k) = E[M_k] + n^2P(o_1, o_2 \in M_k, o_1 \neq o_2) \quad (8.8.20)$$

We note that if $o_1 \in M_k$, then $o_2 \in B_{k-1}(o_1) \setminus \{o_1\}$ is not possible. Therefore, only $t_1$ and $t_2$ for which $V(t_1)$ and $V(t_2)$ intersect at the boundaries contribute. We will refrain from giving all details, but the proof of (8.8.11) can then be followed for the above event, to show that

$$E[M_k^2] = E[M_k]^2(1 + o(1)), \quad (8.8.21)$$

so that, for $k \leq (1 - \varepsilon) \log \log n / \log m$,

$$E[M_k] \xrightarrow{p} 1. \quad (8.8.22)$$

Together with (8.8.19), this completes the proof that $M_k \xrightarrow{p} \infty$ for $k \leq (1 - \varepsilon) \log \log n / \log m$. \qed

To complete the lower bound on $\text{diam}(\text{PA}^n_{m,\delta})$ in Theorem 8.30, we take two vertices $u, v \in M_k$ with $k = (1 - \varepsilon) \log \log n / \log m$. Then, by definition, $\partial B_k(u), \partial B_k(v) \subset [n] \setminus [n/2]$. We can then adapt the proof of Theorem 8.17 is adapted to show that whp, the distance between $\partial B_k(u)$ and $\partial B_k(v)$ is whp still bounded from below by $4 \log \log n / \log (\tau - 2)$. Therefore, the lower bound on the diameter is then whp given by

$$\text{diam}(\text{PA}^n_{m,\delta}) \geq \text{dist}_{\text{PA}^n_{m,\delta}}(u, v) \quad (8.8.23)$$

$$= 2k + \text{dist}_{\text{PA}^n_{m,\delta}}(\partial B_k(u), \partial B_k(v)) \geq 2(1 - \varepsilon) \log \log n / \log m + 4 \log \log n / \log (\tau - 2).$$

This gives an informal proof of the lower bound.

The critical case of $\delta = 0$

We complete this section, by discussing the diameter for $\delta = 0$:

**Theorem 8.32** (Diameter of $\text{PA}^n_{m,\delta}$ for $\delta = 0$) \textit{Fix $m \geq 2$. For $\text{PA}^n_{m,\delta}$, as $n \to \infty$,}

$$\text{diam}(\text{PA}^n_{m,\delta}) \frac{\log \log n}{\log n} \xrightarrow{p} 1. \quad (8.8.24)$$

As we see, the diameter in Theorem 8.32 and the typical distances in Theorem 8.7 behave similarly. This is rather unique in the critical $\delta = 0$ setting.

The proof of Theorem 8.32 relies on the linear cord diagram (LCD) description of the $\delta = 0$ model by Bollobás and Riordan (2004a).
8.9 Further results on distances in preferential attachment models

8.9.1 Distance evolution for $\delta < 0$

In this section, we survey results about the evolution of the graphs' distances. With this, we mean that we study $t \mapsto \text{dist}_t(i,j) = \text{dist}_{PA_t(m,\delta)}(i,j)$. In Exercise 8.12, you are asked to show that $t \mapsto \text{dist}_t(i,j)$ is non-increasing. Exercises 8.13–8.17 study various aspects of the evolution of distances.

The main result below shows how distances between uniform vertices in $[n]$ decrease as time progresses. This is done by investigating what the distance between these vertices is in $PA_t(m,\delta)$ for all $t \geq n$:

**Theorem 8.33 (Evolution of distances)** Consider $(PA_t(m,\delta))_{t \geq 1}$ with $m \geq 2$ and $\delta \in (-m,0)$. Choose $o_1(n), o_2(n)$ uniformly at random from $[n]$. Then, for all $t \geq n$,

$$\sup_{t \geq n} \left| \text{dist}_{PA_t(m,\delta)}(o_1(n), o_2(n)) - 4 \left\lfloor \frac{\log \log n - \log (1 \vee \log(t/n))}{\log(\tau - 2)} \right\rfloor \right| \log(\tau - 2) \right| \right.$$  \quad (8.9.1)

is a tight sequence of random variables.

Theorem 8.33 is very strong, as it describes very precisely how the distances decrease exactly when the graph $PA_t(m,\delta)$ grows. Further, Theorem 8.33 also proves that $\text{dist}_{PA_t(m,\delta)}(o_1(n), o_2(n)) - 4 \log \log n/|\log(\tau - 2)|$ is a tight sequence of random variables. While the lower tightness (i.e., the tightness of $|\text{dist}_{PA_t(m,\delta)}(o_1(n), o_2(n)) - 4 \log \log n/|\log(\tau - 2)|$) follows from Theorem 8.17, we have not proved the upper tightness (i.e., the tightness of $|\text{dist}_{PA_t(m,\delta)}(o_1(n), o_2(n)) - 4 \log \log n/|\log(\tau - 2)|$) in Theorem 8.8 (recall (8.7.29)). Exercises 8.19–8.19 investigate what happens when $t = t_n = ne^{(\log n)\alpha}$ when $\alpha \in (0,1)$ and $\alpha > 1$, respectively. This leaves open the interesting case where $\alpha = 1$.

The proof of Theorem 8.33 by Jorritsma and Komjáthy (Prepr.(2020)) uses related ideas as used in the present chapter, but significantly extends them in order to prove the uniformity in $t \geq n$. While we do not present the full proof here, we do explain why $\text{dist}_{PA_t(m,\delta)}(o_1(n), o_2(n)) = 2$ whp when $t = t_n \gg n^{-2m/\delta}$, which is clearly an interesting case and explains why the supremum in (8.9.1) can basically be restricted to $t \in [n,n^{-2m/\delta+o(1)}]$. This sheds light on what happens precisely when $t = t_n = ne^{(\log n)\alpha}$ for $\alpha = 1$, a case that is left open in the above cases.

The probability that one of the $m$ edges of vertex $n + t + 1$ connects to $u$ and one to $v$ (which certainly makes the distance between $u$ and $v$ equal to 2) is close
8.9 Further results on distances in preferential attachment models

347
to

\[ m(m - 1)\mathbb{E}\left[ \frac{D_v(n + t) + \delta}{2m(n + t) + (n + t)\delta} \frac{D_v(n + t) + \delta}{2m(n + t) + (n + t)\delta} \mid \text{PA}^{(m,\delta)}_n \right] \tag{8.9.2} \]

\[
= (1 + o(1)) \frac{m(m - 1)}{(2m + \delta)^2}\mathbb{E}\left[ (D_v(n + t) + \delta)(D_v(n + t) + \delta) \mid \text{PA}^{(m,\delta)}_n \right] \]

\[
= (1 + o(1)) \frac{m(m - 1)}{(2m + \delta)^2}2^{(2+\delta/m)}(n+\delta)(\frac{t}{n})^{2/(2+\delta/m)} \]

\[= (1 + o(1))\frac{m(m - 1)}{(2m + \delta)^2}t^{-2(m+\delta)/(2m+\delta)}n^{-2/(2+\delta/m)}.\]

When taking \( u = o_1^{(n)} \), \( v = o_2^{(n)} \), we have that \( D_v(n) \overset{d}{\to} D_1 \), \( D_u(n) \overset{d}{\to} D_2 \), where \( (D_1, D_2) \) are two i.i.d. copies of a random variable with asymptotic degree distribution \( \mathbb{P}(D = k) = p_k \). Thus, the conditional expectation of the total number of double attachments to both \( o_1^{(n)} \) and \( o_2^{(n)} \) up to time \( n + t \) is close to

\[
\sum_{s=1}^{t} \frac{m(m - 1)D_1D_2}{(2m + \delta)^2}s^{-(2(m+\delta)/(2m+\delta))n^{-2/(2+\delta/m)}} \approx (1 + o(1))\frac{m(m - 1)D_1D_2}{(2m + \delta)(-\delta)}n^{-2m/(2m+\delta)}t^{-\delta/(2m+\delta)}, \tag{8.9.3}
\]

which becomes \( \Theta(1) \) when \( t = Kn^{-2m/\delta} \). The above events, for different \( t \), are close to being independent. This suggests that the process of attaching to both \( o_1^{(n)} \) and \( o_2^{(n)} \) is, conditionally on their degrees \( (D_1, D_2) \), Poisson with some random intensity.

8.9.2 Distances in the Bernoulli PAM

Recall the Bernoulli preferential attachment model introduced in Section 1.3.6. The nice thing about this model is that its degree structure is understood much more generally than for preferential attachment models with a fixed number of edges. This also allows one to zoom in into particular instances of the preferential attachment function. While this is quite a different model, for example due to the fact that the graph is not connected whp, unlike \( \text{PA}^{(m,\delta)} \) for \( m \geq 2 \) and \( n \) large, in terms of distances it behaves rather similarly to the fixed-degree models such as \( \text{PA}^{(m,\delta)}_n \). This is exemplified by the following theorem, which applies in the infinite-variance degree setting:

**Theorem 8.34** (Evolution of distances for scale-free \( \text{BPA}^{(f)}_n \)) Let \( \text{BPA}^{(f)}_n \) be the sublinear preferential attachment model obtained from a concave attachment rule \( f \) satisfying

\[ f(k) = \gamma k + \beta, \tag{8.9.4} \]

where \( \gamma \in (\frac{1}{2}, 1) \), so that \( \tau = 1 + 1/\gamma \in (2, 3) \). Then Theorem 8.33 applies to this setting when we restrict to \( t \geq n \) such that \( o_1^{(n)} \) and \( o_2^{(n)} \) are connected in \( \text{BPA}^{(f)}_n \).
In particular, conditionally on \( o_1^{(n)} \) and \( o_2^{(n)} \) being connected in \( \text{BPA}^{(f)}_n \),

\[
\text{dist}_{\text{PA}(m, \delta)}(o_1^{(n)}, o_2^{(n)}) - 4 \frac{\log \log n}{|\log(\tau - 2)|} \tag{8.9.5}
\]

is a tight sequence of random variables.

The situation of affine preferential attachment functions \( f \) in (8.9.4) where \( \gamma \in (0, \frac{1}{2}) \), so that \( \tau = 1 + 1/\gamma > 3 \) is not so well understood, but one can conjecture that again, the distance between \( o_1^{(n)} \) and \( o_2^{(n)} \) is whp logarithmic at some base related to the operator norm of the multi-type branching process that describes the local weak limit.

The following theorem describes nicely how the addition of an extra power of a logarithm in the degree distribution affects the distances:

**Theorem 8.35** (Critical case: interpolation) Let \( \text{BPA}^{(f)}_n \) be the sublinear preferential attachment model obtained from a concave attachment rule \( f \) satisfying

\[
f(k) = \frac{1}{2}k + \frac{\alpha}{2 \log k} + o\left(\frac{k}{\log k}\right), \tag{8.9.6}
\]

for some \( \alpha > 0 \). Consider two vertices \( o_1, o_2 \) chosen independently and uniformly at random from the largest connected component \( C_{\text{max}} \) of \( \text{BPA}^{(f)}_n \). Then

\[
\text{dist}_{\text{BPA}^{(f)}_n}(o_1, o_2) = (1 + o_P(1)) \frac{1}{1 + \alpha \log \log n} \tag{8.9.7}
\]

Comparing Theorem 8.35 to Theorem 6.23, we see that, for large \( \alpha \), the distances in \( \text{BPA}^{(f)}_n \) are about twice as large as those in \( \text{GRG}_n(W) \) for a degree distribution with the same asymptotic form. This can be seen as an explanation of the occurrence of the extra factor 2 in Theorem 8.8 compared to Theorem 6.3 for the Norros-Reittu model \( \text{NR}_n(w) \) and Theorem 7.2 for the configuration model \( \text{CM}_n(d) \) when the power-law exponent \( \tau \) satisfies \( \tau \in (2, 3) \). Note that this extra factor is absent precisely when \( \alpha = 0 \).

**8.10 Notes and discussion**

Notes on Section 8.1

Scale-free trees have received substantial attention in the literature, we refer to Bollobás and Riordan (2004b); Pittel (1994) and the references therein. Theorem 8.2 is (Pittel, 1994, Theorem 1).

There is a beautiful result on the height of trees using branching processes due to Kingman (1975), which Pittel (1994) makes crucial use of. This approach is based on exponential martingales, and allows for a relatively short proof of the lower bound on the height of the tree.

There is a close analogy between \( \text{PA}^{(1, \delta)}_n \) and so-called uniform recursive trees. In uniform recursive trees, we grow a tree such that at time 1, we have a unique
8.10 Notes and discussion

vertex called the root, with label 1. At time \( n \), we add a vertex and connect it to a uniformly chosen vertex in the tree. See Smythe and Mahmoud (1994) for a survey of recursive trees.

A variant of a uniform recursive tree is the case where the probability that a newly added vertex is attached to a vertex is proportional to the degree of the vertices (and, for the root, the degree of the root plus one). This process is called a \textit{random plane-oriented recursive tree}. For a uniform recursive tree of size \( n \), it is proved by Pittel (1994) that the maximal distance between the root and any other vertex is with high probability equal to
\[
\frac{1}{\gamma} \log n(1 + o(1)),
\]
where \( \gamma \) satisfies (8.1.2) with \( \delta = 0 \). It is not hard to see that this implies that the maximal graph distance between any two vertices in the uniform recursive tree is equal to
\[
\frac{1}{\gamma} \log n(1 + o(1)).
\]

Notes on Section 8.2.
Theorem 8.6 is proved by Dommers et al. (2010).

Theorem 8.7 is proved by Bollobás and Riordan (2004a). More precisely, the result for the diameter of \( PA_{m,0}^n \) in Theorem 8.32 is proved by Bollobás and Riordan (2004a). This proof can be rather easily be extended to the proof for the typical distances.

A weaker version of Theorem 8.8 is proved by Dommers et al. (2010). The current theorem is inspired by Dereich et al. (2012).

Notes on Section 8.3.
The bound in Proposition 8.9 for \( \delta = 0 \) was proved in (Bollobás and Riordan, 2004a, Lemma 3) in a rather different way. The current version for all \( \delta \) is taken from Dommers et al. (2010), and also its proof is adapted from there.

Notes on Section 8.4.
The proofs in this section for \( \delta = 0 \) first appeared in (Bollobás and Riordan, 2004a, Section 4).

Notes on Section 8.5.
The proof of Theorem 8.17 is adapted from Dereich et al. (2012).

Notes on Section 8.6.

Notes on Section 8.7.
Theorem 8.33 is proved by Jorritsma and Komjáthy (Prepr.(2020)), who study the more general problem of \textit{first-passage percolation} on the preferential attachment model. In first-passage percolation, the edges are weighted. These weights can be interpreted as the traversal time of an edge in a rumor spread model. Then, Jorritsma and Komjáthy (Prepr.(2020)) study the time it takes the rumor to go from a random source to a random destination. They obtain sharp results, concerning the leading order asymptotics of this traversal time, as well as for these quantities in a dynamical perspective.
Theorem 8.23 is proved by Dommers et al. (2010), who also first proved a weaker upper bound than in Theorem 8.8. Theorem 8.28 was proved by Dereich et al. (2012).

**Notes on Section 8.8.**

Theorem 8.29 is proved by Dommers et al. (2010). Theorem 8.32 is proved by Bollobás and Riordan (2004a). Theorem 8.30 is proved by Caravenna et al. (2019).

**Notes on Section 8.9.**

Theorem 8.35 is proved by Dereich et al. (2017). Theorem 8.34 is proved by Jorritsma and Komjáthy (Prepr. (2020)).

8.11 Exercises for Chapter 8

**Exercise 8.1** (Bound on $\theta$) Prove that the solution $\theta$ of (8.1.2) satisfies $\theta < 1$. What does this imply for the diameter and typical distances in scale-free trees?

**Exercise 8.2** (Bound on $\theta$) Prove that the solution $\theta$ of (8.1.2) satisfies $\theta \in (0, e^{-1})$.

**Exercise 8.3** (Early vertices are whp at distance 2 for $\delta < 0$) Let $\delta \in (-m, 0)$ and $m \geq 2$. Show that for $i, j$ fixed
\[
\lim_{n \to \infty} \mathbb{P}(\text{dist}_{PA_n^m(\delta)}(n-1, n) \leq 2) = 1.
\] (8.11.1)

**Exercise 8.4** (Negative correlations for $m = 1$) Show that when $m = 1$, Lemma 8.10 implies that when $(\pi_0, \ldots, \pi_k)$ contains different coordinates as $(\rho_0, \ldots, \rho_k)$, then
\[
\mathbb{P}\left(\bigcap_{i=0}^{k-1} \{\pi_i \leftrightarrow \pi_{i+1}\} \cap \bigcap_{i=0}^{k-1} \{\rho_i \leftrightarrow \rho_{i+1}\}\right) \leq \mathbb{P}\left(\bigcap_{i=0}^{k-1} \{\pi_i \leftrightarrow \pi_{i+1}\}\right) \mathbb{P}\left(\bigcap_{i=0}^{k-1} \{\rho_i \leftrightarrow \rho_{i+1}\}\right).
\] (8.11.2)

**Exercise 8.5** (Extension of (8.3.16) to $PA_n^{\delta,1}(b)$) Prove that for $PA_n^{\delta,1}(b)$, (8.3.16) is replaced with
\[
\mathbb{P}(u_1 \downarrow v, u_2 \downarrow v) = (1 + \delta) \frac{\Gamma(u_1 - \delta/(2 + \delta))\Gamma(u_2 - (1 + \delta)/(2 + \delta))\Gamma(v)}{\Gamma(u_1 + 1/(2 + \delta))\Gamma(u_2)\Gamma(v + 2/(2 + \delta))}.
\] (8.11.3)

**Exercise 8.6** (Distance between $n-1$ and $n$ in $PA_n^{m,0}$) Show that whp,
\[
dist_{PA_n^{m,0}}(n-1, n) \geq k^*_n,
\] (8.11.4)
where $k^*_n$ is defined in (8.4.14).
Exercise 8.7 (Distance between vertices 1 and 2 in $\text{PA}_n^{(m,0)}$) Check what the analysis in Section 8.4.2 implies for $\text{dist}_{\text{PA}_n^{(m,0)}}(1,2)$. In particular, where does the proof that $\text{dist}_{\text{PA}_n^{(m,0)}}(1,2) \geq k_n^0$, where $k_n^0$ is defined in (8.4.14), fail, and why should it?

Exercise 8.8 (Most recent common ancestor in $\text{PA}_n^{(1,\delta)}$) Fix $o_1, o_2$ to be two vertices in $[n]$ chosen at random, and let $V$ be the oldest vertex that the path from 1 to $o_1$ and that from 1 to $o_2$ have in common in $\text{PA}_n^{(1,\delta)}$. Use the proof of Lemma 8.13 to show that $\text{dist}_{\text{PA}_n^{(1,\delta)}}(1,V)$ is tight.

Exercise 8.9 (Upper tightness criterion for centered $\text{dist}_{\text{PA}_n^{(m,\delta)}}(o_1, o_2))$ Fix $\delta \in (-m,0)$. Use (8.7.29) and Theorem 8.23 to show that $\text{dist}_{\text{PA}_n^{(m,\delta)}}(o_1, o_2) - 2\log \log n/|\log(\tau - 2)|$ is upper tight when $\text{dist}_{\text{PA}_n^{(m,\delta)}}(o_1, \text{Core}_n)$ is tight.

Exercise 8.10 (Tightness of $\text{dist}_{\text{PA}_n^{(m,\delta)}}(o', \text{Core}_n)$) Fix $\delta \in (-m,0)$ and let $o'$ be chosen uniformly at random from $[n]$. Show that $\text{dist}_{\text{PA}_n^{(m,\delta)}}(o', \text{Core}_n)$ is a tight sequence of random variables, by using $n$-connectors.

Exercise 8.11 (Path to upper tightness criterion for centered $\text{dist}_{\text{PA}_n^{(m,\delta)}}(o_1, o_2)$) How can the ideas in the previous two exercises be combined to show that $\text{dist}_{\text{PA}_n^{(m,\delta)}}(o_1, o_2) - 2\log \log n/|\log(\tau - 2)|$ is upper tight? You do not have to give the full proof, but rather convince yourself that this is indeed true, by adapting the proof of Theorem 8.28. [Hint: Use that $o_1 \in [2(1-\varepsilon)n]$ with probability at least $\varepsilon$ when $o_1 \in [2n]$ is chosen uniformly. Then only use vertices in $[2n] \setminus [2(1-\varepsilon)n]$ as possible $n$-connectors.]

Exercise 8.12 (Monotonicity of distances in $\text{PA}_n^{(m,\delta)}$) Fix $m \geq 1$ and $\delta > -m$. Show that $t \mapsto \text{dist}_t(i,j) = \text{dist}_{\text{PA}_n^{(m,\delta)}}(i,j)$ is non-decreasing.

Exercise 8.13 (Distance evolution in $\text{PA}_n^{(1,\delta)}$) Fix $m = 1$ and $\delta > -1$. Show that $t \mapsto \text{dist}_t(i,j) = \text{dist}_{\text{PA}_n^{(1,\delta)}}(i,j)$ is constant for $t > i \wedge j$.

Exercise 8.14 (Distance structure on $\mathbb{N}$ due to $\text{PA}_n^{(m,\delta)}$) Fix $m \geq 1$ and $\delta > -m$. Use Exercise 8.12 to show that $\text{dist}_t(i,j) \xrightarrow{\text{a.s.}} \text{dist}_\infty(i,j) < \infty$ for all $i,j \geq 1$. Thus, $\text{dist}_\infty$ is a distance function on $\mathbb{N}$.

Exercise 8.15 (Nearest-neighbors on $\mathbb{N}$ due to $\text{PA}_n^{(m,\delta)}$) Fix $m \geq 2$ and $\delta > -m$. Compute $\mathbb{P}(\text{dist}_\infty(i,j) = 1)$.

Exercise 8.16 (Infinitely many neighbors) Fix $m \geq 2$ and $\delta > -m$. Show that $\#\{j: \text{dist}_\infty(i,j) = 1\} = \infty$ a.s. for all $i \geq 1$.

Exercise 8.17 (Eventually distances at most 2 in $\text{PA}_n^{(m,\delta)}$) Fix $m \geq 2$ and $\delta \in (-m,0)$. Show that $d_\infty(i,j) \leq 2$ for all $i,j$.

Exercise 8.18 (Evolution of distances in $\text{PA}_n^{(m,\delta)}$; critical parametric choice) Fix $m \geq 2$ and $\delta \in (-m,0)$. Choose $o_1^{(n)}, o_2^{(n)}$ uniformly at random from $[n]$, and take
$t = t_n = ne^{(\log n)\alpha}$ for some $\alpha \in (0, 1)$. Use Theorem 8.33 to identify $\theta_\alpha$ such that
\[
\frac{\text{dist}_{\text{PA}^{(m,\delta)}}(o_1^{(n)}, o_2^{(n)})}{\log \log n} \xrightarrow{p} \theta_\alpha. \tag{8.11.5}
\]

**Exercise 8.19 (Tight distances in $\text{PA}^{(m,\delta)}$ for $\delta < 0$)** Fix $m \geq 2$ and $\delta \in (-m, 0)$. Choose $o_1^{(n)}, o_2^{(n)}$ uniformly at random from $[n]$, and take $t = t_n = ne^{(\log n)\alpha}$ for some $\alpha > 1$. Use Theorem 8.33 to show that $\text{dist}_{\text{PA}^{(m,\delta)}}(o_1^{(n)}, o_2^{(n)})$ is a tight sequence of random variables.

**Exercise 8.20 (Degree evolution $\text{PA}^{(m,\delta)}$)** Fix $m \geq 1$ and $\delta > -m$. Take a vertex $v$ such that $D_v(n) = k$. Show that, for $t \geq 1$, the random process $t \mapsto D_v(n + t)$ evolves like a Pólya urn starting with $k$ red balls and weight $k + \delta$, while the number and weight of the blue balls are $n - k$ and $(2m + \delta)n - k - \delta$. Identify the Beta random variable $U$ for which, for all $t \geq 1$,
\[
P(D_v(n + t) = k + l \mid D_v(n) = k) = E[P(\text{Bin}(t, U) = l)]. \tag{8.11.6}
\]
Part IV

Related models and problems
In Part III, we have investigated the small-world behavior of random graphs, extending the results on the existence and uniqueness of the giant component as informally described in Meta Theorem A on Page III. It turns out that the results are all quite similar, even though the details of the description of the models are substantially different. We can summarize the results obtained in the following meta theorem:

**Meta Theorem B.** (Small- and ultra-small-world characteristics) In a random graph model with power-law degrees having power-law exponent $\tau$, the typical distances of the giant component in a graph of size $n$ are order $\log \log n$ when $\tau \in (2,3)$, while they are of order $\log n$ when $\tau > 3$. Further, these typical distances are highly concentrated.

Informally, these results quantify the ‘six-degrees of separation’ paradigm in random graphs, where we see that random graphs with very heavy-tailed degrees have ultra-small typical distances, as could perhaps be expected. The lines of proof of these results are even often similar, relying on clever path-counting techniques. In particular, the results show that in generalized random graphs and configuration models alike, in the $\tau \in (2,3)$ regime, vertices of high degrees, say $k$, are typically connected to vertices of even higher degree or order $k^{1/(\tau - 2)}$. In the preferential attachment model, on the other hand, this is not true, yet vertices of degree $k$ tend to be connected to vertices of degree $k^{1/(\tau - 2)}$ in two steps, making typical distances roughly twice as large.

**Overview of Part IV.**

In this part, we study several related random graph models that can be seen as extensions of the simple models studied so far. They incorporate features that we have not seen yet, such as direction of the edges, existence of clustering, communities and/or geometry. The red line through this part will be the study to which extent the main results informally described in Meta Theorems A (see page 217) and B (see above) remain valid, and if not, to which extent they need to be adapted. We will not give complete proofs, but instead informally explain why results are similar as in these Meta Theorems A and B, or why instead they are different.
Chapter 9
RELATED MODELS

Abstract
In this chapter, we discuss some related random graph models that have been studied in the literature. We explain their relevance, as well as some of the properties in them. We discuss directed random graphs, random graphs with community structure, as well as spatial random graphs.

Organization of this chapter
We start in Section 9.1 by extensively discussing two examples of real-world networks. The aim there is to show that the models as discussed so far, tend not to be highly appropriate for real-world examples. We choose the example of citation networks, which are directed, have substantial clustering, have a hierarchical community structure and possibly even a spatial component to them. While we focus on citation networks, we also highlight that we could have chosen other real-world examples as well. Section 9.2 by discussing directed versions of the random graphs studied in this book. In Section 9.3 and 9.4, we introduce several random graph models that have a community structure in them, so as to model the communities occurring in many (maybe even most?) real-world networks. Section 9.3 studies the setting where communities are macroscopic, in the sense that there are a bounded number of communities even when the network size tends to infinity. In Section 9.4, instead, we look at the setting where the communities have a bounded average size. We will argue that both settings are relevant. In Section 9.5, we discuss random graph models that have a spatial component to them, and explain how the spatial structure gives rise to high clustering. We close this chapter with notes and discussion in Section 9.6, and with exercises in Section 9.7.

9.1 Two real-world network examples
In this section, we discuss two real-world examples of networks, citation networks in Section 9.1.1 and terrorist networks in Section 9.1.2. These examples show that many real-world networks are different from the stylized network models that we have discussed to far. In fact, many real-world networks are directed, in that edges (or arcs) are from one vertex to the other, and not necessarily in a symmetric way between pairs of vertices. Also, real-world networks often display a community structure, in that certain parts are more densely connected that between those parts and the rest of the network. Sometimes such communities are relatively small, between few vertices, and sometimes they are quite large, and even communities with sizes of the order of the entire network exist.
9.1.1 Citation networks

In this section, we discuss citation networks, where vertices denote scientific papers and the directed edges correspond to citations of one paper to another. Obviously, such citations are directed, since it makes a difference whether your paper cites mine, or my paper cites yours, leading to a directed network. In such a network, a directed edge is sometimes called an arc.

Obviously, citation networks grow in time. Indeed, papers do not disappear, so a citation, once made in a published paper, does not disappear either. Further, their dynamics is by time, and growth is enormous. See Figure 9.1(a) to see that the number of papers in various fields grows exponentially with time, meaning that more and more papers are being written. If you ever wondered why scientists seem to become ever more busy, then this may be an apparent explanation. In Figure 9.1(a), we have displayed the number of papers in three different domains, namely, Probability and Statistics (PS), Electrical Engineering (EE) and Biotechnology and Applied Microbiology (BT). The data comes from the Web Of Science database. While the exponential growth is quite prominent in the data, it is unclear how this exponential growth arises. This could either be due to the fact that the number of journals that are listed in Web Of Science grows over time, or that journals contain more and more papers. However, the exponential growth has been observed already as early as the 80’s, see the book by Derek De Solla Price (1986), appropriately called ‘Little science, big science’.

As you can see, we have already restricted to certain subfields in science, the reason being that the publication and citation cultures in different fields are vastly different. Thus, we have attempted to go to a situation in which the networks that we investigate are a little more homogeneous. For this, it is relevant to be able to distinguish such fields, and to decide which papers (or journals) contribute to which field. This is a fairly daunting task, as you can imagine. However, it is also an ill-defined task, as no subdomain is truly homogeneous. Let me restrict myself to probability and statistics, as I happen to know this area best. In probability and statistics, there are subdomains that are very pure, as well as areas that are highly applied, such as applied statistics. These areas do indeed have different publication and citation cultures. Thus, science as a whole is probably hierarchical, where large scientific disciplines can be identified, that can, in turn, be subdivided into smaller subdomains, etc. However, one should stop somewhere, and the current three scientific disciplines are homogeneous enough to make our point.

Figure 9.1(b) shows the loglog plot for the in-degree distribution in these 3 citation networks. We notice that these datasets have empirical power-law citation distributions. Thus, on average, papers attract few citations, but the amount of variability in the number of citations is rather substantial. We are also interested in the dynamics of the citation distribution of the papers published in a given year, as time proceeds. This can be observed in Figure 9.2. We see a dynamical power law, meaning that at any time the degree distribution of a cohort of papers
from a given time period (in this case 1984) is close to a power law, but the exponent changes over time (and in fact decreases, which corresponds to heavier tails). When time grows quite large, the power law approaches a fixed value.

Interestingly, the existence of power-law in-degrees in citation networks also has a long history. Already in 1965, Derek De Solla Price (1965) observed it, and even proposed a model for it that relied on a preferential attachment mechanism, more than two decades before Barabási and Albert (1999) proposed the first preferential attachment model.

We wish to discuss two further properties of citation networks and their dynamics. In Figure 9.3, we see that the majority of papers stop receiving citations after some time, while few others keep being cited for longer times. This inhomogeneity in the evolution of node degrees is not present in classical PAMs, where the degree of every fixed vertex grows as a positive power of the graph size. Figure 9.3 shows that the number of citations of papers published in the same year can be rather different, and the majority of papers actually stop receiving citations quite soon. In particular, after a first increase, the average increment of citations decreases over time (see Figure 9.4). We observe a difference in this aging effect between the PS dataset and the other two datasets, due to the fact that in PS,
scientists tend to cite older papers than in EE or BT, again exemplifying the differences in citation and publication patterns in different fields. Nevertheless, the average increment of citations received by papers in different years tends to decrease over time for all three datasets.

A last characteristic that we observe is the lognormal distribution of the age of cited papers. In Figure 9.5, we plot the distribution of cited papers, looking at references made by papers in different years. We have used a 20 years time window in order to compare different citing years. Notice that this lognormal distribution seems to be very similar within different years, and the shape is similar over different fields.

Let us summarize the differences between citation networks and the random graph models that form the basis of network science. Citation networks are directed, which is different from the typical undirected models that we have discussed so far. However, it is not hard to adapt our models to become directed, and we will explain this in Section 9.2. Secondly, citation networks have a substantial community structure, in that parts of the network exist that are much more densely connected than the whole network. We can argue that both communities exist on a macroscopic scale, for example in terms of the various scientific disciplines that science consist of, as well as on a microscopic scale, where research networks of small groups of scientists create subnetworks that are more densely connected. One could even argue that geography plays an important role in citation networks, since many collaborations between scientists are within their own university or country, even though we all work with various researchers around the globe.

Citation networks are dynamic, like preferential attachment models (PAMs), but their time evolution is quite different to PAMs, as the linear growth in PAMs is replaced by an exponential growth in citation networks. Further, papers in citation networks seem to age, as seen both in Figures 9.3 and 9.4, in that citation rates become smaller for large times, in such a way that typical papers even completely stop receiving citations at some (random) point in time.

In conclusion, finding an appropriate model for citation networks is quite a

![Figure 9.3](image-url) Time evolution for the number of citations of samples of 20 randomly chosen papers from 1980 for PS and EE, and from 1982 for BT.
9.1 Two real-world network examples

Figure 9.4 Average degree increment over a 20-years time window for papers published in different years. PS presents an aging effect different from EE and BT, showing that papers in PS receive citations longer than papers in EE and BT.

Figure 9.5 Distribution of the age of cited papers for different citing years.

challenge, and one should be quite humble in one’s expectation that the standard models are anywhere near to the complexity of real-world networks.

9.1.2 Terrorist networks

9.1.3 General real-world network models

We conclude that real-world networks tend to have many properties that our different from the models that we have discussed so far in this book. Indeed, they often have a pronounced community structure, which can act at a microscopic scale or at a macroscopic scale, or anything in between, i.e., such networks are hierarchically organised. Real-world networks can be directed, with all the particulars that this gives rise to. The spatial positioning of the vertices can have a profound effect on the likelihood of edges being present. Real-world networks are often changing over time, and should therefore be considered as temporal networks. In some cases, even several networks work together, and can thus not be seen as separate networks. For example, in transporting people within a country, the railroad network and the road network are both highly relevant. At larger
distances, also airline networks become involved. Thus, to study how people move around the globe, we cannot study each of these networks in isolation. In science, the collaboration networks of authors and the citation networks of papers together give a much clearer picture of how science works than any single one in isolation, even though these more restricted views can offer useful insight.

One may become rather overwhelmed by the complexity that real-world networks provide. Indeed, high-level complex network science is part of complexity theory, the science of complex systems. However, the past decades have given rise to a bulk of insights, often based on relatively simple models such as the ones discussed so far. Indeed, often Box (1976) is quoted as saying that “All models are wrong but some are useful”. In Box (1979), a more elaborate version of this quote is as follows:

> Now it would be very remarkable if any system existing in the real world could be exactly represented by any simple model. However, cunningly chosen parsimonious models often do provide remarkably useful approximations. For example, the law $PV = RT$ relating pressure $P$, volume $V$ and temperature $T$ of an “ideal” gas via a constant $R$ is not exactly true for any real gas, but it frequently provides a useful approximation and furthermore its structure is informative since it springs from a physical view of the behavior of gas molecules. For such a model there is no need to ask the question “Is the model true?” If “truth” is to be the “whole truth” the answer must be “No”. The only question of interest is “Is the model illuminating and useful?”.

Thus, we should not feel discouraged at all! In particular, it is important to know when to include an extra features into a model, so that the model at hand becomes more “useful”. For this, the first step is to come up with models that do incorporate these extra features. It turns out that many of the models discussed so far can easily be adapted so as to include features such as directedness, community structure and geometry. Further, it is straightforward to combine such properties, so as to include more of them. The simpler models that do not have such features than serve as a useful model to compare to, and can thus act as a “benchmark” for more complex situation. The understanding of such simple models often helps us in understanding the more complex models, since many properties, tools and ideas can be easily extended to the more complex settings. In some cases, the extra feature gives rise to a richer behavior, which then merits being studied in full detail. In these steps, network science has moved significantly forward compared to the models described so far. The aim of this chapter is to highlight some of the lessons learned.

Below, we will discuss directed random graphs in Section 9.2, random graphs
9.2 Directed random graphs

Many real-world networks are directed, in the sense that edges are oriented. For example, in the World-Wide Web, the vertices are web pages, and the edges are the hyperlinks between them, which are clearly oriented. One could naturally forget about these directions, but that would discard a wealth of information. For example, in citation networks, it makes a substantial difference whether my paper cites a paper, or that paper cites mine.

This section or organised as follows. A directed graph is often called a digraph. We start by defining digraphs. After this, we discuss various models invented for directed graphs. We start by discussing directed inhomogeneous random graphs in Section 9.2.1, then discuss directed configuration models in Section 9.2.2, and close with directed preferential attachment models in Section 9.2.3.

A digraph $D = (V(D), E(D))$ on the vertex set $V(D) = [n]$ has an edge set that is a subset of the set $E(D) \subseteq [n]^2 = \{(u, v): u, v \in [n]\}$ of all ordered pairs of elements of $[n]$. Elements of $D$ are called directed edges or arcs. In many cases, a direction in the edges is natural. For example, citation networks and the World-Wide Web are naturally directed.

The connectivity structure of digraphs.

The orientation in the edges causes substantial differences in the connectivity structure of the graphs involved. Indeed, a given vertex $v$ has both a forward connected component consisting of all the vertices that it is connected to, as well as a backward connected component. Finally, every vertex has a strongly connected component (SCC), which consists of those vertices to which there exists a forward and a backward path. See Exercise 9.1 for a description of the topology of digraphs that ensures that the SCC of a vertex is well defined, in that it does not depend on the vertex chosen.

The different notions of connectivity divide the graph up into several disjoint parts. Often, there is a unique largest strongly connected component (SSC) that contains a positive proportion of the graph. This is the part of the graph that is most strongly connected. There are collections of vertices that are forward connected to the SCC, but not backward, these are the IN parts of the graph. Further, there are the collections of vertices that are backward connected to the SCC, but not forward, these are the OUT parts of the graph. And finally, there are the parts of the graph that are neither, and consist of their own SCC and IN and OUT parts. See Figure 9.6 (which is [Volume 1, Figure 1.19] repeated) for a description of these parts for the WWW, as well as their relative sizes.
Local weak convergence of digraphs.

It turns out that there are several ways to define local weak convergence for digraphs. This is due to the fact that LWC is defined in terms of neighborhoods, which in turn depend on the connectivity that one wishes to make use of. For the neighborhood $B^G_r(v)$, do we wish to explore all vertices that can be reached from $v$ (which is relevant when $v$ is the source of an infection), or all the vertices from which we can reach $v$ (which is relevant when investigating whether $v$ can be infected by others, or in the case of PageRank). These are relevant when exploring a graph from or towards a vertex, and we will call then exploration neighborhoods. One might also consider all edges at the same time, thus ignoring the direction of edges for the local neighborhoods.

Let $\text{dist}_G(u, v)$ be the (directed) graph distance from $u$ to $v$, i.e., the minimal number of directed edges needed to connect $u$ to $v$. Note that, for digraphs, $\text{dist}_G(u, v)$ and $\text{dist}_G(v, u)$ may be different. We will consider the forward exploration neighborhood $B^{G^f}_r(u) = (V(B^{G^f}_r(u)), E(B^{G^f}_r(u)), u)$ defined by

$$V(B^{G^f}_r(u)) = \{ v \in V(G): \text{dist}_G(u, v) \leq r \}, \quad (9.2.1)$$
$$E(B^{G^f}_r(u)) = \{ (x, y) \in E(G): \text{dist}_G(u, x), \text{dist}_G(u, y) \leq r \}. \quad (9.2.2)$$

as in (2.2.1). For the backward exploration neighborhood, $\text{dist}_G(u, v) \leq r$ in the definition of $V(B^{G^b}_r(u))$ is replaced by $\text{dist}_G(v, u) \leq r$, and $\text{dist}_G(u, x), \text{dist}_G(u, y)$ in the definition of $E(B^{G^b}_r(u))$ is replaced by $\text{dist}_G(x, u), \text{dist}_G(y, u)$. One could even consider the forward-backward exploration neighborhood, which is the union of the forward and the backward exploration neighborhoods.

The notion of isomorphisms in Definition 2.3, and of the metric on directed rooted graphs are straightforwardly adapted from Definition 2.4. From this moment on, the definitions of forward local weak convergence of deterministic di-
9.2 Directed random graphs

graphs, and that of forward local weak convergence of random digraphs in their various settings are straightforwardly adapted as well.

There is a minor catch though. While the notion of forward exploration neighborhood keeps track of the out-degrees \(d^{(\text{out})}_v\) for all \(v \in B^{(G)}_{r-1}(u)\), it does not keep track of the in-degrees \(d^{(\text{in})}_v\) for \(v \in B^{(G)}_{r-1}(u)\). Similarly, the notion of backward exploration neighborhood keeps track of the out-degrees \(d^{(\text{in})}_v\) for all \(v \in B^{(G)}_{r-1}(u)\), but not of the out-degrees \(d^{(\text{out})}_v\) for \(v \in B^{(G)}_{r-1}(u)\). Below, we need these as well.

For this, we add a degree-mark to the vertices that indicates the in-degrees in the forward exploration neighborhood, and the out-degrees of the vertices in the backward exploration neighborhood. Particularly in random graphs that are locally tree-like, the edges in the other direction than the one explored, will often go to vertices that are far away, and are thus often not in the explored neighborhood. Thus, to use this information, we need to explicitly keep track of it.

Let us explain how this marking is defined. For a vertex \(v\), let \(m(v)\) denote its mark. We then define an isomorphism \(\phi: V(G_1) \to V(G_2)\) between two labeled rooted graphs \((G_1,o_1)\) and \((G_2,o_2)\) to be an isomorphism between \((G_1,o_1)\) and \((G_2,o_2)\) that respects the marks, i.e., for which \(m_1(v) = m_2(\phi(v))\) for every \(v \in V(G_1)\), where \(m_1\) and \(m_2\) denote the degree-mark functions on \(G_1\) and \(G_2\) respectively. We then define \(R^*\) as in (2.1.2), and the metric on rooted degree-marked graphs as in (2.1.3). We call the resulting notion of LWC marked forward LWC, and marked backward LWC, respectively.

Even when considering forward-backward neighborhoods, the addition of marks is necessary. Indeed, while for the root of this graph, we do know its in- and out-degree by construction, for the other vertices in the forward and backward neighborhoods this information is still not available. While the above discussion may not be relevant for all questions that one may wish to investigate using local weak convergence techniques, it is useful for the discussion of PageRank, as we discuss next.

Exercises 9.3–9.5 investigate the various notions of local weak convergence for some directed random graphs that are naturally derived from undirected graphs.

Weak convergence of PageRank.

Recall the discussion in Section 2.4.4 about local weak convergence of PageRank on undirected graphs. Here we will extend this discussion to digraphs, which are far more relevant in practice.

Recall the definition of PageRank from [Volume 1, Section 1.5]. Let us first explain the solution in the absence of dangling ends, so that \(d^{\text{(out)}}_i \geq 1\) for all \(i \in [n]\). Let \(G = ([n],E)\) denote the web graph, so that \([n]\) corresponds to the web pages and a directed edge \((i,j) \in E\) is present precisely when page \(i\) refers to \(j\). Denote the out-degree of vertex \(i\) by \(d^{\text{(out)}}_i\). Fix \(\alpha \in (0,1)\). Then, we let the
vector of PageRanks \( (R_i)_{i\in[n]} \) be the unique solution to the equation

\[
R_i = \alpha \sum_{j \to i} \frac{R_j}{d^\text{(out)}_j} + 1 - \alpha,
\]

(9.2.3)
satisfying the normalization \( \sum_{i \in [n]} R_i = n \) (recall (2.4.51)).

The parameter \( \alpha \in (0, 1) \) is called the damping factor, and guarantees that (2.4.50) has a unique solution. This solution can be understood in terms of the stationary distribution of a bored surfer. Indeed, denote \( \pi_i = R_i/n \), so that \( (\pi_i)_{i\in[n]} \) is a probability distribution that satisfies a similar relation as \( (R_i)_{i\in[n]} \) in (2.4.50), namely

\[
\pi_i = \alpha \sum_{j \to i} \frac{\pi_j}{d^\text{(out)}_j} + \frac{1 - \alpha}{n},
\]

(9.2.4)
Therefore, \( (\pi_i)_{i\in[n]} \) is the stationary distribution of a random walker that, with probability \( \alpha \) jumps according to a simple random walk, i.e., it chooses any of the out-edges with equal probability, while with probability \( 1 - \alpha \), the walker jumps to a uniform location.

In the presence of dangling ends, we can just redistribute their mass equally over all vertices that are not dangling ends, so that (2.4.50) becomes

\[
R_i = \alpha \sum_{j \to i} \frac{R_j}{d^\text{(out)}_j} + \frac{(1 - \alpha)\sum_{j \in D} R_j + 1}{n} - \alpha,
\]

(9.2.5)
where \( D \subseteq [n] \) denotes the collection of dangling nodes (and, when \( i \in D \), the first term is zero by convention, and (9.2.5) reduces to (9.2.3)).

The damping factor \( \alpha \) is quite crucial. When \( \alpha = 0 \), the stationary distribution is just \( \pi_i = 1/n \) for every \( i \), so that all pages have PageRank 1. This is not very informative. On the other hand, PageRank concentrates on dangling ends when \( \alpha \) is close to one, and this is also not what we want. Experimentally, \( \alpha = 0.85 \) seems to work well and strikes a nice balance between these two extremes.

We next investigate the convergence of the PageRank distribution on a directed graph sequence \( (G_n)_{n \geq 1} \) that converges locally weakly:

**Theorem 9.1** (Existence of asymptotic PageRank distribution) Consider a sequence of directed random graphs \( (G_n)_{n \in \mathbb{N}} \). Then, the following hold:

(i) If \( G_n \) converges in distribution in the marked backward local weak sense, then there exists a limiting distribution \( R_{\emptyset} \), with \( \mathbb{E}[R_{\emptyset}] \leq 1 \), such that

\[
R_n \xrightarrow{d} R_{\emptyset}.
\]

(ii) If \( G_n \) converges in probability in the marked backward local weak sense, then there exists a limiting distribution \( R_{\emptyset} \), with \( \mathbb{E}[R_{\emptyset}] \leq 1 \), such that, for every \( r > 0 \),

\[
\frac{1}{n} \sum_{v \in [n]} 1_{\{R_n > r\}} \xrightarrow{p} \mathbb{P}(R_{\emptyset} > r).
\]

(9.2.7)
9.2 Directed random graphs

Theorem 9.1 is the directed version of Theorem 2.24. Interestingly, the positivity of the damping factor also allows us to give a power-iteration formula for $R^{(G_n)}$, and thus for $R_\emptyset$. Indeed, let

$$A^{(G_n)}_{i,j} = \frac{1_{(j \rightarrow i)}}{d_j^{(\text{out})}}, \quad i, j \in V(G_n)$$

(9.2.8)

denote the (normalized) adjacency matrix of the graph $G_n$. Then, $R^{(G_n)}$ can be computed as

$$R^{(G_n)}_v = (1 - \alpha) \sum_{k=0}^{\infty} \alpha^k \sum_{i \in V(G_n)} (A^{(G_n)})^k_{v,i},$$

(9.2.9)

As a result, when $G_n$ converges in distribution in the marked backward local weak sense with limit $(G, \emptyset)$, then also $R_\emptyset$ can be computed as

$$R_\emptyset = (1 - \alpha) \sum_{k=0}^{\infty} \alpha^k \sum_{i \in V(G)} (A^{(G)})^k_{\emptyset,i},$$

(9.2.10)

where $A^{(G)}_{i,j}$ is the normalized adjacency matrix of the backwards local weak limit $G$.

The power-law hypothesis for PageRank.

Recall from [Volume 1, Section 1.5] that the PageRank power-law hypothesis states that the PageRank distribution satisfies a power law with the same exponent as that of the in-degree. By Theorem 9.1, this can be rephrased by stating that $P(R_\emptyset > r) \asymp r^{-\tau_{\text{in}}}$ when $P(D^{(\text{in})}_\emptyset > r) \asymp r^{-\tau_{\text{in}}}$ (we are on purpose being vague about what $\asymp$ means in this context).

Exercises 9.6–9.7 investigate the implications of (9.2.10) for the power-law hypothesis for random graphs having bounded out-degrees.

We continue by discussing some models of directed random graphs or random di-graphs. We start in Section 9.2.1 by discussing inhomogeneous random di-graphs, in Section 9.2.2 we continue with the directed configuration model, and in Section 9.2.3, we close with directed preferential attachment models.

9.2.1 Directed inhomogeneous random graphs

Let $(x_i)_{i \in [n]}$ be a sequence of variables with values in $\mathcal{S}$ such that the empirical distribution of $(x_i)_{i \in [n]}$ approximates a measure $\mu$ as $n \to \infty$. That is, we assume that, for each $\mu$-continuous Borel set $\mathcal{A} \subseteq \mathcal{S}$, as $n \to \infty$,

$$\frac{1}{n} |\{i \in [n] : x_i \in \mathcal{A}\}| \to \mu(\mathcal{A}).$$

(9.2.11)

Here we say that a Borel set $\mathcal{A}$ is $\mu$-continuous whenever its boundary $\partial \mathcal{A}$ has zero probability, i.e., $\mu(\partial \mathcal{A}) = 0$. 

Given \( n \), let \( G_n \) be the random digraph on the vertex set \((x_i)_{i \in [n]}\) with independent arcs having probabilities

\[
p_{ij} = \mathbb{P}((x_i, x_j) \in E(G_n)) = 1 \land (\kappa(x_i, x_j)/n), \quad i, j \in [n]. \tag{9.2.12}
\]

We denote the resulting graph by \( \text{DIRG}_n(\kappa) \). Combining \( \mathcal{S}, \mu, \) and \( \kappa \), we obtain a large class of inhomogeneous digraphs with independent arcs. Obviously, the model will include digraphs with in-degree and out-degree distributions which have power laws. This general model was studied in detail in Chapter 3. We are a little less general than there, since we assume that \( \kappa \) in (9.2.12) not to depend on \( n \). This simplifies the exposition considerably. Note that in the case of random graphs it is necessary to assume, in addition, that the kernel \( \kappa \) is symmetric. In the definition of digraphs \( G_n \) for \( n \geq 2 \), we do not require the symmetry of the kernel.

### Assumptions on the kernel.

We need to impose further conditions on the kernel \( \kappa \), like those in Chapter 3. Namely, we need to assume that the kernel \( \kappa \) is irreducible \((\mu \times \mu)\)-almost everywhere. That is, for any measurable \( \mathcal{A} \subseteq \mathcal{S} \) with \( \mu(\mathcal{A}) \neq 1 \) or \( \mu(\mathcal{A}) \neq 0 \), the identity \((\mu \times \mu)((s, t) \in \mathcal{A} \times (\mathcal{S} \setminus \mathcal{A}) \mid \kappa(s, t) = 0) = 0\) implies that either \( \mu(\mathcal{A}) = 0 \) or \( \mu(\mathcal{S} \setminus \mathcal{A}) = 0 \). In addition, we assume that \( \kappa \) is continuous almost everywhere on \( (\mathcal{S} \times \mathcal{S}, \mu \times \mu) \), and the number of arcs in \( \text{DIRG}_n(\kappa) \), denoted by \(|E(\text{DIRG}_n(\kappa))|\), satisfies that, as \( n \to \infty \),

\[
\frac{1}{n} |E(\text{DIRG}_n(\kappa))| \xrightarrow{\text{a.s.}} \int_{\mathcal{S} \times \mathcal{S}} \kappa(s, t) \mu(ds) \mu(dt) < \infty. \tag{9.2.13}
\]

Note that here we implicitly assume that \( \kappa \) is integrable.

### Examples of directed inhomogeneous random graphs

We next discuss some examples of \( \text{DIRG}_n(\kappa) \).

**Directed Erdős-Rényi random graph.** The most basic example is the directed Erdős-Rényi random graph, in which \( p_{ij} = p_{ji} = \lambda/n \). In this case, \( \kappa(x_i, x_j) = \lambda \).

**Finite-type directed inhomogeneous random graphs.** Slightly more involved are kernels of finite type, in which case \( (s, t) \mapsto \kappa(s, t) \) takes on finitely many values. Such kernels are also highly convenient to approximate more general models with, as has been exemplified in the undirected setting in Chapter 3.

**Directed rank-1 inhomogeneous random graphs.** We next generalize random-1 inhomogeneous random graphs. For \( v \in [n] \), let \( w^{(\text{in})}_v \) and \( w^{(\text{out})}_v \) be its respective in- and out-weights that will have the interpretation of the asymptotic average in- and out-degrees, respectively, under a summation-symmetry condition on the weights. Then, the directed generalized random graph \( \text{DGRG}_n(w) \) has edge probabilities given by

\[
p_{ij} = p_{ij}^{(\text{DGRG})} = \frac{w^{(\text{out})}_i w^{(\text{in})}_j}{\ell_n + w^{(\text{out})}_i w^{(\text{in})}_j}, \tag{9.2.14}
\]
9.2 Directed random graphs

where

$$\ell_n = \frac{1}{2} \sum_{i \in [n]} (w_i^{(\text{out})} + w_i^{(\text{in})}).$$

(9.2.15)

Let \((W_n^{(\text{out})}, W_n^{(\text{in})}) = (w_o^{(\text{out})}, w_o^{(\text{in})})\) denote the in- and out-weights of a uniformly chosen vertex \(o \in [n]\). Similarly to Condition 1.1, we assume that

\[(W_n^{(\text{out})}, W_n^{(\text{in})}) \xrightarrow{d} (W^{(\text{out})}, W^{(\text{in})}), \quad \mathbb{E}[W_n^{(\text{out})}] \to \mathbb{E}[W^{(\text{out})}], \quad \mathbb{E}[W_n^{(\text{in})}] \to \mathbb{E}[W^{(\text{in})}],\]

(9.2.16)

where \((W^{(\text{out})}, W^{(\text{in})})\) is the limiting in- and out-weight distribution. Exercises 9.8 investigates the expected number of edges in this setting.

When thinking of \(w_i^{(\text{out})}\) and \(w_i^{(\text{in})}\) as corresponding to the approximate out- and in-degrees of vertex \(i \in [n]\), it is reasonable to assume that

$$\mathbb{E}[W^{(\text{in})}] = \mathbb{E}[W^{(\text{out})}].$$

(9.2.17)

Indeed, we know that, with \(D_i^{(\text{out})}\) and \(D_i^{(\text{in})}\) denoting the out- and in-degrees of vertex \(i \in [n]\), that (recall Exercise 9.2)

$$\sum_{i \in [n]} D_i^{(\text{out})} = \sum_{i \in [n]} D_i^{(\text{in})}.$$  (9.2.18)

Thus, if indeed \(w_i^{(\text{out})}\) and \(w_i^{(\text{in})}\) are the approximate out- and in-degrees of vertex \(i \in [n]\), then also

$$\sum_{i \in [n]} w_i^{(\text{out})} \approx \sum_{i \in [n]} w_i^{(\text{in})},$$

(9.2.19)

which, assuming (9.2.16), proves (9.2.17).

As discussed in more detail in Section 1.3.2 (see in particular (1.3.8) and (1.3.9)), many related versions of the rank-1 inhomogeneous random graph exist. We refrain from giving more details here.

**Multi-type marked branching processes for \(\text{DIRG}_n(\kappa)\).**

For large \(n\), the local weak convergence and phase transition in the digraph \(G_n\) can be described in terms of the survival probabilities of the related multi-type Galton-Watson branching processes with type space \(S\). Let us introduce the necessary mixed-Poisson branching processes now. Given \(s \in S\), let \(X(s)\) and \(Y(s)\) denote the Galton-Watson processes starting at a particle of type \(s \in S\) such that the number of children of types in a subset \(A \subseteq S\) of a particle of type \(t \in S\) has a Poisson distribution with means

$$\int_A \kappa(t, u) \mu(du), \quad \text{and} \quad \int_A \kappa(u, t) \mu(du),$$

(9.2.20)

respectively. These numbers are independent for disjoint subsets \(A\) and for different particles. These two branching processes correspond to the forward and backward limits of \(\text{DIRG}_n(\kappa)\).

We now extend the discussion by also defining the marks. When we consider
the branching process \(X(s)\), to each individual of type \(t\), we associate an independent mark having a Poisson distribution with mean \(\int_S \kappa(t,u)\mu(du)\). When we consider the branching process \(X(s)\), to each individual of type \(t\), we associate an independent mark having a Poisson distribution with mean \(\int_S \kappa(u,t)\mu(du)\), instead. These random variables correspond to the ‘in-degrees’ for the forward exploration process \(X(s)\), and the ‘out-degrees’ for the backward exploration process \(Y(s)\). Finally, for the forward-backward setting, we let the marked branching processes \(X(s)\) and \(Y(s)\) be independent. We call these objects \textit{Poisson marked branching processes with kernel} \(\kappa\).

As in Section 3.4.3, we let \(T_\kappa\) be defined as in (3.4.13), i.e., for \(f: S \rightarrow \mathbb{R}\), we let \((T_\kappa f)(x) = \int_S \kappa(x,y) f(y)\mu(dy)\).

We now come to the main results on \(\text{DIRG}_n(\kappa)\), which involve its local weak convergence of and its phase transition.

**Local weak convergence for \(\text{DIRG}_n(\kappa)\)**

Recall that \(C_{\text{max}}\) denotes the maximal SCC, and \(C_{(2)}\) the second largest SCC, in \(\text{DIRG}_n(\kappa)\). Here ties are broken arbitrarily when needed. The following theorem describes the local weak convergence of \(\text{DIRG}_n(\kappa)\):

**Theorem 9.2 (Local weak convergence of \(\text{DIRG}_n(\kappa)\))** Suppose that \(\kappa\) is irreducible, continuous almost everywhere on \((S \times S, \mu \times \mu)\), and that (9.2.13) holds. Then, \(\text{DIRG}_n(\kappa)\) converges in probability in the marked forward and backward and forward-backward local weak convergence sense to the above Poisson marked branching processes with kernel \(\kappa\), where the law of the type of the root \(\emptyset\) is \(\mu\).

We do not give a proof of Theorem 9.2, and refer to Section 9.6 for its history. In Exercise 9.10, you are asked to determine the local weak limit of the Erdős-Rényi random graph. Exercise 9.11 proves Theorem 9.2 in the case of finite-type kernels, while Exercise 9.12 investigates local weak convergence of the directed generalized random graph.

**The phase transition in \(\text{DIRG}_n(\kappa)\)**

The critical point of the emergence of the giant SCC is determined by the averaged joint survival probability

\[
\zeta = \int_S \zeta_X(s)\zeta_Y(s)\mu(ds)
\]

being positive. Here \(\zeta_X(s)\) and \(\zeta_Y(s)\) denote the non-extinction probabilities of \(X(s)\) and \(Y(s)\), respectively. The following theorem describes the phase transition on \(\text{DIRG}_n(\kappa)\):

**Theorem 9.3 (Phase transition in \(\text{DIRG}_n(\kappa)\))** Suppose that \(\kappa\) is irreducible, continuous almost everywhere on \((S \times S, \mu \times \mu)\), and that (9.2.13) holds.

(a) When \(\nu = \|T_\kappa\| > 1\), \(\zeta\) in (9.2.21) satisfies \(\zeta \in (0,1]\) and

\[
|C_{\text{max}}|/n \xrightarrow{p} \zeta,
\]

(9.2.22)
9.2 Directed random graphs

while $|\mathcal{E}(2)|/n \xrightarrow{p} 0$ and $|E(\mathcal{E}(2))|/n \xrightarrow{p} 0$.

(b) When $\nu = \|T_k\| < 1$, $\zeta$ in (9.2.21) satisfies $\zeta = 0$ and $|\mathcal{C}(2)|/n \xrightarrow{p} 0$ and $|E(\mathcal{C}(2))|/n \xrightarrow{p} 0$.

Theorem 9.3 is the directed version of Theorem 3.16. Exercises 9.13 and 9.14 investigate the conditions for a giant component to exist for the directed Erdős-Rényi random graph and generalized random graphs.

9.2.2 The directed configuration model

One way to obtain a directed version of $\text{CM}_n(d)$ is to give each edge a direction, chosen with probability $\frac{1}{2}$, independently of all other edges. In this model, however, the correlation coefficient between the in- and out-degree of vertices is close to one, particularly when the degrees are large (see Exercise 9.15). In real-world applications, correlations between in- and out-degrees can be positive or negative, depending on the precise application, so we aim to formulate a model that is more general. Therefore, we formulate a general model of directed graphs, where we can prescribe both the in- and out-degrees of vertices.

Let $d^{(\text{in})}$ be a sequence of in-degrees, where $d^{(\text{in})}_i$ denotes the in-degree of vertex $i$. Similarly, we let $d^{(\text{out})}$ be a sequence of out-degrees. Naturally, we need that

$$\sum_{i \in [n]} d^{(\text{in})}_i = \sum_{i \in [n]} d^{(\text{out})}_i$$  \hspace{1cm} (9.2.23)

in order for a graph with in- and out-degree sequence $d = (d^{(\text{in})}, d^{(\text{out})})$ to exist (recall Exercise 9.2).

We think of $d^{(\text{in})}_i$ as the number of in-half-edges incident to vertex $i$ and $d^{(\text{out})}_i$ as the number of out-half-edges incident to vertex $i$. The directed configuration model $\text{DCM}_n(d)$ is obtained by pairing each in-half-edge to a uniformly chosen out-half-edge. The resulting graph is a random multigraph, where each vertex $i$ has in-degree $d^{(\text{in})}_i$ and out-degree $d^{(\text{out})}_i$. Similarly to $\text{CM}_n(d)$, $\text{DCM}_n(d)$ can have self-loops as well as multiple edges. A self-loop arises at vertex $i$ when one of its in-half-edges pairs to one of its out-half-edges. Let $(D^{(\text{in})}_n, D^{(\text{out})}_n)$ denote the in- and out-degree of a vertex chosen uniformly at random from $[n]$.

We continue to investigate the strongly connected component of $\text{DCM}_n(d)$. Assume, similarly to Condition 1.5(a)-(b), that

$$(D^{(\text{in})}_n, D^{(\text{out})}_n) \xrightarrow{d} (D^{(\text{in})}, D^{(\text{out})}),$$  \hspace{1cm} (9.2.24)

and that

$$E[D^{(\text{in})}_n] \rightarrow E[D^{(\text{in})}], \quad \text{and} \quad E[D^{(\text{out})}_n] \rightarrow E[D^{(\text{out})}].$$  \hspace{1cm} (9.2.25)

Naturally, by (9.2.23), this implies that $E[D^{(\text{out})}] = E[D^{(\text{in})}]$.

Let

$$p_{k,l} = P(D^{(\text{in})} = k, D^{(\text{out})} = l)$$  \hspace{1cm} (9.2.26)
denote the asymptotic joint in- and out-degree distribution. We refer to \((p_{k,l})_{k,l \geq 0}\) simply as the asymptotic degree distribution of \(DCM_n(d)\). The distribution \((p_{k,l})_{k,l \geq 0}\) plays a similar role for \(DCM_n(d)\) as \((p_k)_{k \geq 0}\) does for \(CM_n(d)\). We further define

\[
\begin{align*}
p^{\text{(in)}}_k &= \sum_l l p_{k,l} / \mathbb{E}[D^{\text{(out)}}], & p^{\text{(out)}}_l &= \sum_k k p_{k,l} / \mathbb{E}[D^{\text{(in)}}].
\end{align*}
\] (9.2.27)

The distributions \((p^{\text{(in)}}_k)_{k \geq 0}\) and \((p^{\text{(out)}}_l)_{l \geq 0}\) correspond to the asymptotic forward in- and out-degree of a uniformly chosen edge in \(CM_n(d)\).

### The local weak limit of the directed configuration model

Let us now formulate the construction of the appropriate marked forward and backward branching processes that will arise as the local weak limit of \(DCM_n(d)\).

For the forward branching process, we let the root have out-degree with distribution \(p^{\text{(out)}}_l = \mathbb{P}(D^{\text{(out)}} = l) = \sum_{l \geq 0} p_{k,l}\), whereas every other vertex except the root has independent out-degree with law \((p^{\text{(out)}}_l)_{l \geq 0}\). Further, for a vertex of out-degree \(l\), we let the mark (that will correspond to its asymptotic in-degree) be \(k\) with probability \(p_{k,l} / p^{\text{(out)}}_l\). For the marked backward branching process, we reverse the role of in and out. For the marked forward-backward branching process, we let the root have joint out- and in-degree distribution \((p_{k,l})_{k,l \geq 0}\), and define the forward and backward processes and marks as before. We call the above branching process the *marked modular branching process with degree distribution* \((p_{k,l})_{k,l \geq 0}\).

Then, we have the following local weak limit result:

**Theorem 9.4 (Local weak convergence of \(DCM_n(d)\))** Suppose that the out- and in-degrees in directed configuration model \(DCM_n(d)\) satisfy (9.2.24) and (9.2.25). Then, \(DCM_n(d)\) converges in probability in the marked forward and backward and forward-backward local weak convergence sense to the above marked modular branching process with degree distribution \((p_{k,l})_{k,l \geq 0}\).

It will not come as a surprise that Theorem 9.4 is the directed version of Theorem 4.1. Exercise 9.16 asks you to give a proof of Theorem 9.4.

### The giant component in the directed configuration model

Recall that the *strongly connected component* of a vertex \(v\) is the set of \(u\) for which there are directed paths from \(v\) to \(u\) and from \(u\) to \(v\), so that \(u\) is both in the forward and backward cluster of \(v\). We let \(C_{\text{max}}\) denote the size of the largest strongly connected component in \(DCM_n(d)\).

Let \(\theta^{\text{(in)}}\) and \(\theta^{\text{(out)}}\) be the survival probabilities of the branching processes with offspring distributions \((p^{\text{(in)}}_k)_{k \geq 0}\) and \((p^{\text{(out)}}_k)_{k \geq 0}\), respectively, and define

\[
\zeta^{\text{(in)}} = 1 - \sum_{k,l} p_{k,l} (1 - \theta^{\text{(in)}})^l, & \zeta^{\text{(out)}} = 1 - \sum_{k,l} p_{k,l} (1 - \theta^{\text{(out)}})^k.
\] (9.2.28)

Then, \(\zeta^{\text{(out)}}\) has the interpretation of the asymptotic probability that a uniform
9.2 Directed random graphs

vertex has a large forward cluster, while \( \zeta^{(\text{in})} \) has the interpretation of the asymptotic probability that a uniform vertex has a large backward cluster. Here, the backward cluster of a vertex \( v \) consists of all vertices \( u \) that are connected to \( v \), and the forward cluster of \( v \) consists of all vertices \( u \) for which \( v \) is connected to \( u \). Further, let

\[
\psi = \sum_{k,l} p_{k,l} (1 - \theta^{(\text{in})})^l (1 - \theta^{(\text{out})})^k
\]

(9.2.29)

so that \( \psi \) has the interpretation of the asymptotic probability that a uniform vertex has finite forward and backward cluster. We conclude that \( 1 - \psi \) is the probability that a uniform vertex has either a large forward or backward cluster, and thus

\[
\zeta = \zeta^{(\text{out})} + \zeta^{(\text{in})} - (1 - \psi)
\]

(9.2.30)

has the interpretation of the asymptotic probability that a uniform vertex has both a large forward and backward cluster.

Finally, we let

\[
\nu = \sum_{k=0}^{\infty} k \sum_{l} p_{k,l} = \sum_{k,l} k l p_{k,l} / \mathbb{E}[D^{(\text{out})}] = \frac{\mathbb{E}[D^{(\text{in})}] D^{(\text{out})}}{\mathbb{E}[D^{(\text{out})}]},
\]

(9.2.31)

Alternatively, \( \nu = \sum_{k=0}^{\infty} k p_k^{(\text{out})} \).

Then, the main result concerning the size of the giant is as follows:

**Theorem 9.5** (Phase transition in \( \text{DCM}_n(d) \)) Suppose that the out- and in-degrees in directed configuration model \( \text{DCM}_n(d) \) satisfy (9.2.24) and (9.2.25).

(a) When \( \nu > 1 \), \( \zeta \) in (9.2.30) satisfies \( \zeta \in (0, 1] \) and

\[
|\mathcal{C}_{\text{max}}|/n \overset{p}{\to} \zeta,
\]

(9.2.32)

while \( |\mathcal{C}_2|/n \overset{p}{\to} 0 \) and \( |\mathbb{E}(\mathcal{C}_2)|/n \overset{p}{\to} 0 \).

(b) When \( \nu < 1 \), \( \zeta \) in (9.2.30) satisfies \( \zeta = 0 \) and \( |\mathcal{C}_{\text{max}}|/n \overset{p}{\to} 0 \) and \( |\mathbb{E}(\mathcal{C}_{\text{max}})|/n \overset{p}{\to} 0 \).

Theorem 9.5 is the adaptation to \( \text{DCM}_n(d) \) of Theorem 4.4 for \( \text{CM}_n(d) \).

In Exercises 9.17, you are asked to prove that the probability that the size of \( |\mathcal{C}_{\text{max}}|/n \) exceeds \( \zeta + \varepsilon \) vanishes, whenever a graph sequence converges in the marked forward-backward sense locally weakly in probability. In Exercise 9.18, this is used to prove Theorem 9.5(b).

**Logarithmic typical distances in the directed configuration model**

We continue by studying the small-world nature of typical distances in the directed configuration model. Let \( u, v \in [n] \), and let \( \text{dist}_{\text{DCM}_n(d)}(u, v) \) denote the graph distance between \( u \) and \( v \), i.e., the minimal number of directed edges needed to connect \( u \) to \( v \). Our main result is as follows:
Theorem 9.6 (Logarithmic typical distances in DCM\(_n(d)\))  Suppose that the out- and in-degrees in directed configuration model DCM\(_n(d)\) satisfy (9.2.24) and (9.2.25), and assume that

\[ \nu = \frac{\mathbb{E}[D^{(\text{in})}]D^{(\text{out})}}{\mathbb{E}[D^{(\text{out})}]} > 1. \]  \hspace{1cm} (9.2.33)

Further, assume that

\[ \mathbb{E}[(D_n^{(\text{in})})^2] \to \mathbb{E}[(D^{(\text{in})})^2] < \infty, \quad \mathbb{E}[(D_n^{(\text{out})})^2] \to \mathbb{E}[(D^{(\text{out})})^2] < \infty. \]  \hspace{1cm} (9.2.34)

Then, conditionally on \( o_1 \to o_2 \),

\[ \frac{\text{dist}_{\text{DCM}_n(d)}(o_1, o_2)}{\log(n)} \xrightarrow{p} \frac{1}{\log \nu}. \]  \hspace{1cm} (9.2.35)

Theorem 9.6 is the directed version of Theorem 7.1. The philosophy behind the proof is quite similar: by using a breadth-first exploration process, we see that \( |\partial B^{(\text{out})}(r, o_1)| \) grows roughly like \( \nu^r \), so in order to ‘catch’ \( o_2 \), one would need \( r \approx \log \nu(n) \). Of course, up to this stage, the branching process approximation starts to fail, which is why one needs to grow the neighborhoods from two sides.

It would be tempting to believe that (9.2.34) is not precisely needed, and this is the content to Exercise \( ?? \).

There are no results as of yet that investigate the typical distances \( \text{dist}_{\text{DCM}_n(d)}(o_1, o_2) \) when \( \nu = \infty \). However, as you are asked to prove in Exercise \( ?? \), it is possible to show that \( \text{dist}_{\text{DCM}_n(d)}(o_1, o_2) = o_r(\log n) \) in this case.

**Logarithmic diameter in the directed configuration model**

We next investigate the logarithmic asymptotics of the diameter of DCM\(_n(d)\). Here, we define the diameter of the digraph DCM\(_n(d)\) to be equal to

\[ \text{diam}(\text{DCM}_n(d)) = \max_{u,v \in V: u \to v} \text{dist}_{\text{DCM}_n(d)}(u,v). \]  \hspace{1cm} (9.2.36)

In order to state the result, we introduce some notation. Let

\[ f(s,t) = \mathbb{E}[s^{D^{(\text{in})}}t^{D^{(\text{out})}}] \]  \hspace{1cm} (9.2.37)

be the bivariate generating function of \( (D^{(\text{in})}, D^{(\text{out})}) \). Recall that \( \theta^{(\text{in})} \) and \( \theta^{(\text{out})} \) are the survival probabilities of a branching process with offspring distributions \( (p_k^{(\text{in})})_{k \geq 0} \) and \( (p_k^{(\text{out})})_{k \geq 0} \), respectively. Write

\[ \nu^{(\text{in})} = \frac{\partial}{\partial s} f(1 - \theta^{(\text{in})}, 1), \quad \text{and} \quad \nu^{(\text{out})} = \frac{\partial}{\partial s} f(1, 1 - \theta^{(\text{out})}). \]  \hspace{1cm} (9.2.38)

Then, the diameter in the directed configuration model DCM\(_n(d)\) behaves as follows:

**Theorem 9.7 (Logarithmic diameter in DCM\(_n(d)\))**  Suppose that the out- and
9.2 Directed random graphs 375

in-degrees in directed configuration model $\text{DCM}_n (d)$ satisfy (9.2.24) and (9.2.25). Further, assume that (9.2.34) holds. Then, when $\nu = \frac{\mathbb{E}[D^{(\text{in})}]}{\mathbb{E}[D^{(\text{out})}]} > 1$,

$$\frac{\text{diam}(\text{DCM}_n (d))}{\log(n)} \xrightarrow{\nu} \frac{1}{\log \nu^{(\text{in})}} + \frac{1}{\log \nu^{(\text{out})}}, \quad (9.2.39)$$

Theorem 9.7 is the directed version of Theorem 7.16. The interpretation of the different terms is similar to that in Theorem 7.16: the terms involving $\nu^{(\text{in})}$ and $\nu^{(\text{out})}$ indicate the depths of the deepest traps, where a trap indicates that the neighborhood lives for a long time without gaining substantial mass. The term involving $\nu^{(\text{in})}$ is the largest in-trap, and that involving $\nu^{(\text{out})}$ is the largest in-trap. These numbers are determined by first taking $r$ such

$$\mathbb{P}(\partial B_r (o) \in [1, K]) \approx \Theta \left( \frac{1}{n} \right), \quad (9.2.40)$$

where $\partial B_r (o)$ corresponds to the ball of the backward $r$-neighborhood for $\nu^{(\text{in})}$, and to the forward $r$-neighborhood for $\nu^{(\text{out})}$, while $K$ is arbitrary and large. Due to large deviations for supercritical branching processes, one can expect that

$$\mathbb{P}(\partial B_r (o) \in [1, K]) \approx (\nu^{(\text{in/out})})^r. \quad (9.2.41)$$

Then, we can identify $r^{(\text{in})} = \log_{\nu^{(\text{in})}} (n)$ and $r^{(\text{out})} = \log_{\nu^{(\text{out})}} (n)$. The solutions to (9.2.41) are given by (9.2.38). For those special vertices $u, v$ for which $|\partial B_r^{(\text{in})} (u)| \in [1, K]$ and $|\partial B_r^{(\text{out})} (v)| \in [1, K]$, it then takes around $\log_{\nu^{(\text{in})}} (n)$ steps to connect $\partial B_r^{(\text{in})} (u)$ to $\partial B_r^{(\text{out})} (v)$, explaining the asymptotics in Theorem 9.7. Of course, proving that this heuristic is correct is quite a bit harder.

It is tempting to conjecture that Theorem 9.7 remains valid under weaker assumptions that in (9.2.34), however, this has not been shown and it may be hard.

9.2.3 Directed preferential attachment models

Bollobás, Borgs, Chayes and Riordan (2003) investigate a directed preferential attachment model and prove that the degrees obey a power law similar to the one in [Volume 1, Theorem 8.3]. For its definition and the available results on degree structure, we refer the reader to [Volume 1, Section 8.9]. Unfortunately, the type of properties investigated in this book have so far not been analyzed for this random graph model. In particular, there is no description of the strong connected component, nor of typical distances and diameters of this model.

One can also interpret normal preferential attachment models as directed graphs by orienting edges from young to old. This can be a useful perspective, for example when modeling temporal networks in which younger vertices can only connect to older vertices, such as for citation networks. We refrain from discussing this in more detail, as the connectivity structure of such directed versions is not so interesting. For example, the strongly connected component is always small (see Exercise 9.19).
9.3 Random graphs with community structure: global communities

Many real-world networks have communities that are global in their size. For example, when partitioning science up in its core fields, citation networks have such a global community structure. Also, in Belgian telecommunication networks of who calls who, the division into the French and the Flemish speaking parts is clearly visible, while in U.S. politics, the partition into Republicans and Democrats plays a pronounced effect on the network structure of social interactions between politicians. In this section, we discuss random graph models for networks with a global community structure. This section is organised as follows. In Section 9.3.1, we discuss stochastic block models, which are the models of choice for networks with community structures. In Section 9.3.2, we discuss degree-corrected stochastic blockmodels, which are similar to stochastic blockmodels, but allow for more inhomogeneity in the degree structure. In Section 9.3.3, we discuss configuration models with global communities, and we close in Section 9.3.4 with preferential attachment models with global communities. We introduce the models, state the most important results in them, and also discuss the topic of community detection in such models, a topic that has attracted considerable attention due to its practical importance.

9.3.1 Stochastic blockmodel

We have already encountered stochastic blockmodels as inhomogeneous random graphs with finitely many types in Chapter 3. Here, we repeat the definition, aster which we focus on the (highly interesting and challenging) community detection results.

Fix \( r \geq 2 \) and suppose we have a graph with \( r \) different types of vertices. Let \( \mathcal{S} = \{1, \ldots, r\} \). Let \( n_i \) denote the number of vertices of type \( i \), and let \( \mu_n(i) = n_i/n \). Let \( \text{IRG}_n(\kappa) \) be the random graph where two vertices of types \( i \) and \( j \), respectively, joined by an edge with probability \( n^{-1} \kappa(i, j) \) (for \( n \geq \max \kappa \)). Then \( \kappa \) is equivalent to an \( r \times r \) matrix, and the random graph \( \text{IRG}_n(\kappa) \) has vertices of \( r \) different types (or colors). We will assume that the type distribution \( \mu_n \) satisfies

\[
\lim_{n \to \infty} n_i/n = \mu_i. \tag{9.3.1}
\]

Exercise 3.11 then shows that the resulting graph is graphical, so that the results in Chapters 3 and 6 apply. As a result, we will not spend much time on the degree distribution, giant and graph distances in this model, as they have been addressed there. Exercise 9.20 studies the degree structure, while Exercise 9.21 investigates when a giant exists in this model.

Let us mention that for the stochastic blockmodel to be a good model for networks with a global community structure, one would expect that the edge probabilities of internal edges, i.e., edges between vertices of the same type, are larger than those of the external types, i.e., edges between vertices of different types. In formulas, this means that \( \kappa_{i,i} > \kappa_{i,j} \) for all \( i, j \in \mathcal{S} \). For example, the
bipartite Erdős-Rényi random graph, which has a structure that is quite opposite to a random graph with global communities (as vertices are only neighbors of vertices of a different type) is not considered a stochastic blockmodel.

Community detection in stochastic blockmodels

We next discuss the topic of community detection in stochastic blockmodels. Before we can say anything about when it is possible to detect communities, we must first define what this means. A community detection algorithm is an assignment \( \pi : [n] \mapsto [r] \) where \( \pi(i) = s \) means that the algorithm assigns type \( s \) to vertex \( i \). In what follows, we will assume that the communities have equal size. Then, in a random guess for the group memberships, two vertices will be guessed to be of the same type with probability \( 1/r \). As a result, we are only impressed with the performance of a community detection algorithm when it does far better than random guessing. This explains the following definition:

**Definition 9.8** (Solvable community detection) Consider a stochastic blockmodel where there are the same number of vertices of each of the \( r \) types, and where \( \sigma(i) \) denotes the type of vertex \( i \in [n] \). We call a community detection problem solvable when there exists an algorithm \( \hat{\sigma} : [n] \mapsto [r] \) and an \( \varepsilon > 0 \) such that, whp as \( n \to \infty \),

\[
\frac{1}{n^2} \sum_{i,j \in [n]} \left[ \mathbb{1}_{\{\sigma(i) = \sigma(j), \sigma(i) = \hat{\sigma}(j)\}} - \frac{1}{r} \right] \geq \varepsilon, \tag{9.3.2}
\]

and otherwise we call the problem unsolvable.

The problem is the most difficult when the degrees of all the different types are the same. This is not so surprising, as otherwise one may aim to classify based on the degrees of the graph. As a result, from now on, we will assume that the degrees of all types of vertices are the same. Some ideas about how one can prove that the problem is solvable can be obtained from Exercises 9.22 and 9.23.

We also assume that there are just two types, so that we can take \( p_{ij} = a/n \) for vertices of the same type, and \( p_{ij} = b/n \) for vertices of opposite type. Here we think of \( a > b \). The question whether the community detection is solvable is answered in the following theorem:

**Theorem 9.9** (Stochastic blockmodel threshold) Take \( n \) to be even. Consider a stochastic blockmodel of two types, each having \( n/2 \) vertices, where the edge probability is \( p_{ij} = a/n \) for vertices of the same type, and \( p_{ij} = b/n \) for vertices of opposite type. Then, the community detection problem is solvable as in Definition 9.8 when

\[
\frac{(a - b)^2}{2(a + b)} > 1, \tag{9.3.3}
\]

while it is unsolvable when

\[
\frac{(a - b)^2}{2(a + b)} < 1. \tag{9.3.4}
\]
Theorem 9.9 is quite surprising. Indeed, it shows that not only should \( a > b \) in order to have a chance to perform community detection, but it should be sufficiently large compared to \( a + b \). Further, the transition in Theorem 9.9 is sharp, in the sense that (9.3.3) and (9.3.4) really complement each other. It is unclear what happens in the critical case when \((a - b)^2 = 2(a + b)\). The solvable case in (9.3.3) is sometimes called an ‘achievability result’, the unsolvable case in (9.3.4) an ‘impossibility result’.

We will not give the proof of Theorem 9.9, as this is quite involved. The proof of the solvable case also shows that the correlation in tends to 1 when \((a - b)^2/[2(a + b)]\) grows large.

In Exercise 9.24, you are asked to show that (9.3.3) implies that \( a - b > 2 \) and \( a + b > 2 \), and to conclude that a giant exists in this setting.

While the results for general number of types \( r \) are less complete, there is an achievability result when \( p_{i,i} = a/n \) and \( p_{i,j} = b/n \) for all \( i, j \in [r] \), in which (9.3.3) is replaced by

\[
\frac{(a - b)^2}{r(a + (k - 1)b)} > 1,
\]

which indeed reduced to (9.3.3) for \( r = 2 \). Also, many results are known about whether efficient algorithms for community detection exist. In general, this means that not only an algorithm should exist, but it should also be computable in reasonable time (say \( \Theta(n \log n) \)). We refer to Section 9.6 for a more elaborate discussion on such results.

Let us close this section by explaining how thresholds such as (9.3.3) and (9.3.5) can be interpreted. Interestingly, there is a close connection to multi-type branching processes. Consider a branching process with finitely many types. Kesten and Stigum (1966) asked in this context when it would be possible to estimate the type of the root when observing the type of the vertices in generation \( k \) for very large \( k \). Offspring matrix \( M_{i,j} = \kappa_{i,j} \mu_i \), which is \( r \times r \). Let \( \lambda_1 > \lambda_2 \) be the two largest eigenvalues of \( M \). Then, the Kesten-Stigum criterion is that this is possible with probability strictly larger than \( 1/r \) when

\[
\frac{\lambda_2^2}{\lambda_1} > 1. \tag{9.3.6}
\]

Next, consider a general finite type inhomogeneous random graph, with limiting types \( \mu_i \) and expected-neighbor matrix \( M_{i,j} = \kappa_{i,j} \mu_i \), which is \( r \times r \). Obviously, the local weak limit of the stochastic blockmodel is the above multi-type branching process, so the link can indeed be expected. Under the condition in (9.3.6), it is believed that the community detection problem is solvable, and that communities can even be detected in polynomial time. For \( r = 2 \), this is sharp, as we have seen above. For \( r \geq 3 \), the picture is much more involved. It is believed that for \( r \geq 4 \), a double phase-transition occurs: Detection should be possible in polynomial time when \( \lambda_2^2/\lambda_1 > 1 \), much harder but still possible (i.e., the best algorithms take
exponentially long) when \(\lambda_2^2/\lambda_1 > c^*\) for some \(0 < c^* < 1\), and information-theoretically impossible when \(\lambda_2^2/\lambda_1 < c^*\). However, this is not yet known in the general case.

The way how to get from a condition like (9.3.6) to an algorithm for community detection is by using the two largest eigenvalues of the random graph, and obtain an estimate for the partition by using the eigenvectors corresponding to the non-backtracking random walk on the graph. The leading eigenvalue converges to \(\lambda_1\) in probability, while the second is bounded by \(|\lambda_2|\). This, together with a good approximation of the corresponding eigenvectors, suggests a specific estimation procedure that we explain now.

Let \(B\) be the non-backtracking matrix of the graph \(G\). This means that \(B\) is indexed by the oriented edges \(\vec{E}(G) = \{(u,v): \{u,v\} \in E(G)\}\), so that \(B\) is \(E(G)\)-valued. For an edge \(e \in \vec{E}(G)\), denote \(e = (e_1, e_2)\), and write

\[
B_{e,f} = \begin{cases} 1 & (e_2 = f_1, e_1 \neq f_2) \\ 0 & \text{otherwise} \end{cases},
\]

(9.3.7)

which indicates that \(e\) ends in the vertex in which \(f\) starts, but \(e\) is not the reversal of \(f\). The latter property explains the name non-backtracking matrix.

Now we come to eigenvalues. We restrict ourselves to the case where \(r = 2\), even though some of the results extend with modifications to higher values of \(r\). Let \(\lambda_1(B)\) and \(\lambda_2(B)\) denote the two leading eigenvalues of \(B\). Then, for the stochastic blockmodel,

\[
\lambda_1(B) \xrightarrow{p} \lambda_1, \quad \lambda_2(B) \xrightarrow{p} \lambda_2,
\]

(9.3.8)

where we recall that \(\lambda_1 > \lambda_2\) are the two largest eigenvalues of \(M\). It turns out that for the Erdős-Rényi random graph with edge probability \((a + b)/[2n] = \alpha/n\), the first eigenvalue \(\lambda_1(B) \xrightarrow{p} \lambda_1 = \alpha\), while the second eigenvalue \(\lambda_2(B)\) satisfies \(\lambda_2(B) \leq \sqrt{\alpha + o(1)}\). Note that this does not follow from (9.3.8), since \(M\) is a one by one matrix. Now, for the stochastic blockmodel with \(r = 2\), we instead have that \(\lambda_2(B) \xrightarrow{p} \lambda_2 = (a - b)/2\). Thus, when

\[
\frac{\lambda_2(B)^2}{\lambda_1(B)} > 1,
\]

(9.3.9)

then we can expect that the graph is a stochastic blockmodel, while if the reverse inequality holds, then we are not even sure whether the model is an Erdős-Rényi random graph or a stochastic blockmodel instead. In the latter case, the graph is so random and homogeneously distributed that we will not be able to make a good estimate for the types of the vertices, which strongly suggests that this case is unsolvable. This at least informally explains (9.3.6).

We next explain how the above analysis of eigenvalues can be used to estimate the types. Assume that \(\lambda_2^2/\lambda_1 > 1\). Let \(\xi_2(B): \vec{E}(G) \to \mathbb{R}\) denote the normalized eigenvector corresponding to \(\lambda_2(B)\). We fix a constant \(\theta > 0\). Then, we estimate \(\hat{\sigma}(v) = 1\) when

\[
\sum_{e: e_2 = v} \xi_2(e) \geq \frac{\theta}{\sqrt{n}}.
\]

(9.3.10)
and otherwise we estimate that \( \hat{\sigma}(v) = 2 \), for some deterministic threshold \( \theta \). This estimation can then be shown to achieve (9.3.2) due to the sufficient separation of the eigenvalues.

9.3.2 Degree-corrected stochastic blockmodel

While the stochastic blockmodel is a nice model to model communities, it has degrees that have Poisson tails (recall Theorem 3.4). Thus, in order to account for the abundant inhomogeneity, an adaptation of the stochastic blockmodel has been proposed, where the degrees are more flexible. This is called the degree-corrected blockmodel. The degree-corrected blockmodel is an inhomogeneous random graph that takes features of both the rank-1 inhomogeneous random graphs as well as of the stochastic blockmodel. Just like the rank-1 setting, there are various possible versions of the model. Here we stick to a version for which the strongest community detection results have been proved. Let us now define the model.

For each vertex \( v \), we sample a random variable \( X_v \), where we assume that \( (X_v)_{v \in [n]} \) are i.i.d. These will be the vertex weights. Conditionally on \( (X_v)_{v \in [n]} \), we then assume that the edge between vertices \( u \) and \( v \) is present with probability

\[
p_{uv} = \left( \kappa_{\sigma(u),\sigma(v)} \frac{X_u X_v}{n} \right) \wedge 1, \tag{9.3.11}
\]

where \( \sigma(u) \in [r] \) denotes the type of vertex \( u \).

Let us first discuss this setting in the simplest case where \( r = 1 \). When the weights \( (x_v)_{v \in [n]} \) in (9.3.11) would be fixed, then we could take them as \( x_v = w_v / \ell_n \), where \( \ell_n = \sum_{v \in [n]} w_v \), in order to obtain the Chung-Lu model CL\(_n\)(\( w \)). This intuition will be helpful in what follows. Unfortunately, when \( (w_v)_{v \in [n]} \) are i.i.d., and \( x_v = w_v / \ell_n \), then \( (x_v)_{v \in [n]} \) are not i.i.d., so it is not obvious how to transfer between the settings. As a result, we stick to the setting in (9.3.11).

From now on, we will assume that there are \( r \) types of vertices, each of which occurs roughly (or precisely, depending on the setting) equally often. We also assume that \( \kappa_{\sigma,\sigma'} \) takes on two values: \( \kappa_{i,i} = a \) and \( \kappa_{i,j} = b \) when \( i \neq j \). This leads us to a very similar model as the stochastic blockmodel, except that the weight structure \( (x_v)_{v \in [n]} \) adds some additional inhomogeneity in the vertex roles, where vertices with high weights generally have larger degrees than those with small weights.

Exercise 9.25–9.26 study the degree structure of the degree-corrected blockmodel, while Exercise 9.27 investigates when a giant exists in this model.

Community detection in degree-corrected stochastic blockmodels

We now come to the main result of this section, which involves the solvability of the estimation in degree-corrected stochastic blockmodels:

**Theorem 9.10** (Degree-corrected stochastic blockmodel threshold) **Take** \( n \) **to be even. Consider a stochastic blockmodel of two types, each having** \( n/2 \) **vertices,**
9.3 Random graphs with community structure: global communities

where the edge probabilities are given by (9.3.11), with $\kappa_{i,i} = a$, and $\kappa_{i,j} = b$ for $i \neq j$. The community detection problem is unsolvable when

$$\frac{(a-b)^2 \mathbb{E}[X^2]}{2(a+b)} < 1.$$  \hspace{1cm} (9.3.12)

Assume further that there exists $\beta > 8$ such that

$$\mathbb{P}(X > x) \leq \frac{1}{x^\beta}.$$  \hspace{1cm} (9.3.13)

Then, the community detection problem is solvable as in Definition 9.8 when

$$\frac{(a-b)^2 \mathbb{E}[X^2]}{2(a+b)} > 1.$$  \hspace{1cm} (9.3.14)

The impossibility result in (9.3.12) in Theorem 9.10 is extended to all $r \geq 2$ under the condition that $(a-b)^2 \mathbb{E}[X^2] < r(a+b)$. The crux in the proof is to show that, for two vertices $o_1, o_2$ chosen uniformly at random, and with $G_n = ([n], E(G_n))$ the realization of the graph of the degree-corrected stochastic blockmodel,

$$\mathbb{P}(\sigma(o_1) = s | \sigma(o_2), G_n) \xrightarrow{\text{r}} \frac{1}{r},$$  \hspace{1cm} (9.3.15)

for every $s \in [r]$. Thus, the type of $o_2$ gives one, asymptotically, no information about the type of $o_1$. This would make detection quite hard, and explains (9.3.12). The proof for the achievability result follows a spectral argument similar to that of the ordinary stochastic blockmodel discussed in Section 9.3.1, and we refrain from discussing it further here. We can expect the power-law bound in (9.3.13) to be too strict and the results to extend under slightly milder assumptions.

9.3.3 Configuration models with global communities

The stochastic blockmodel is an adaptation of the Erdős-Rényi random graph to incorporate global communities, and the degree-corrected stochastic blockmodel is an adaptation of the Chung-Lu model. In a similar way, one can adapt the configuration model to incorporate global communities. Surprisingly, this has not attracted substantial attention in the literature, which is why this section is relatively short. Let us discuss the obvious setting though. Let every vertex $v \in [n]$ have a type $\sigma(v) \in [r]$.

For a vertex $v \in [n]$ and a type $s \in [r]$, we then let $d_v^s$ denote the number of half-edges to be connected from vertex $v$ to vertices of type $s$. We first specify the structure of the graph between vertices of the same type. Let $\sum_{v: \sigma(v) = s} d_v^s$ be the total number of half-edges between vertices of type $s$, and assume that this number is even. We assume that the graph on $\{v: \sigma(v) = s\}$ is a configuration model with $n_s = \#\{v \in [n]: \sigma(v) = s\}$ vertices and degrees $(d_v^s)_{v: \sigma(v) = s}$. This specifies the degrees within communities.

For the structure of the edges between vertices of different types, we recall
what the \textit{bipartite configuration model} is. We suppose that we have vertices of two types, say type 1 and type 2, and we have $n_1$ and $n_2$ vertices of these two types respectively. Let the type of a vertex $v$ again be given by $\sigma(v) \in \{1, 2\}$. Vertices of type 1 have degrees $(d_v)_{v: \sigma(v)=1}$ and vertices of type 2 have degrees $(d_v)_{v: \sigma(v)=2}$. We assume that
\[
\sum_{v: \sigma(v)=1} d_v = \sum_{v: \sigma(v)=1} d_v, \tag{9.3.16}
\]
and we pair the half-edges incident to vertices of type 1 uniformly at random to those incident to vertices of type 2. As a result, there are only edges between vertices of types 1 and 2, and the total number of edges is given in (9.3.16).

Using the above definition, we let the edges between vertices of types $s, t$ be given by a bipartite configuration model between the vertices in $\{ v: \sigma(v) = s \}$ and $\{ v: \sigma(v) = t \}$, where the former have degrees $(d_v)_{v: \sigma(v)=s}$ and the latter degrees $(d_v)_{v: \sigma(v)=t}$. To make the construction feasible, we assume that
\[
\sum_{v: \sigma(v)=s} d_v = \sum_{v: \sigma(v)=t} d_v, \tag{9.3.17}
\]
Special cases of this model are the configuration model for which $r = 1$, and the bipartite configuration model itself, for which $r = 2$ and $d_v$ for every $v$ with $\sigma(v) = t$.

Let $\mu_n(s)$ denote the number of vertices of type $s$. We again assume, as in (9.3.1) that the type distribution $\mu_n(s) = n_s/n$ satisfies that for all $s \in [r],$
\[
\lim_{n \to \infty} n_s/n = \mu_s. \tag{9.3.18}
\]
Also, in order to describe the local and global properties of the configuration models with global communities, one should make assumptions similar to those for the original configuration model in Condition 1.5, but now for the \textit{matrix} of degree distributions. For example, it is natural to assume that, for all $s \in [r]$, the joint distribution function of all the type degrees to satisfy
\[
F_n^{(s)}(x_1, \ldots, x_r) = \frac{1}{n_s} \sum_{v: \sigma(v)=s} 1_{\{d_v^{(1)} \leq x_1, \ldots, d_v^{(r)} \leq x_r\}} \to F^{(s)}(x_1, \ldots, x_r), \tag{9.3.19}
\]
for all $x_1, \ldots, x_r \in \mathbb{R}$ and some limiting joint distribution $F^{(s)}: \mathbb{R}^r \to [0, 1]$. Further, it is natural to assume that an adaptation of Condition 1.5(a) holds for all these degrees, such as that for all $s, t \in [r],$
\[
\frac{1}{n_s} \sum_{v: \sigma(v)=s} d_v^{(t)} \to \mathbb{E}[D^{(s,t)}]. \tag{9.3.20}
\]
While the configuration model, as well as its bipartite version, have attracted some attention, the above extension has so far remained unexplored. Exercises 9.28–9.30 informally investigate some of its properties.
9.3.4 Preferential attachment models with global communities

We next investigate preferential attachment models with global communities. Again assume that there are $r$ communities. The preferential attachment models with global communities is naturally a dynamic random graph model, where now every vertex $n$ has a type $\sigma(n) \in [r]$. The graph is considered to be directed where all edges go from young to old. Each vertex comes in with out-degree equal to $m$, as in the normal preferential attachment model. We will see that the extension studied in the literature is most similar to $(\text{PA}^{(m,0)}_n(b))_{n \geq 0}$, but an extension to $(\text{PA}^{(m,\delta)}_n(b))_{n \geq 0}$ will be discussed as well later on. In a similar way, extensions to $(\text{PA}^{(m,\delta)}_n(a))_{n \geq 0}$ can be formulated.

We start with an initial graph at time $n_0$ given by $G_{n_0}$, in which we assume that every vertex has out-degree $m$, and every vertex has a label $\sigma(v)$ for all $v \in [n_0]$. The graph then evolves as follows. At time $n + 1$, let vertex $n + 1$ have a type $\sigma(n + 1)$ that is chosen in an i.i.d. way from $[r]$, where

$$\mu_s = P(\sigma(n + 1) = s)$$

is the type distribution. We consider $m$ edges incident to vertex $n + 1$ to be half-edges, similarly to how the configuration model is constructed. We give all the out-half-edges incident to a vertex $v$ the label $\sigma(v)$. Further, let the matrix $\kappa: [r] \times [r] \to [0, \infty)$ be the affinity matrix. If $\sigma(n + 1) = s$, then give each half-edge of label $t$ a weight $\kappa_{s,t}$. Choose a half-edge according to these weights, meaning that a half-edge $x$ with label $t$ has a probability proportional to $\kappa_{s,t}$ to be chosen as the pair of any of the $m$ half-edges incident to vertex $n + 1$. We do this for all $m$ half-edges incident to vertex $n + 1$ independently. This creates the graph $G_{n+1}$ at time $n + 1$. The dynamics is iterated indefinitely.

Degree distribution of PAMs with global communities

We start by investigating the degree distribution in the preferential attachment models with global communities. We start by introducing some notation.

Let $\eta_n(s)$ denote the fraction of half-edges with label $s$, and let $\eta_n = (\eta_n(s))_{s \in [r]}$ denote the empirical distribution of the labels of half-edges. For a probability distribution $\eta$ on $[r]$, and a type $s \in [r]$, let

$$h_s(\eta) = \mu_s + \sum_{t \in [r]} \mu_t \frac{\kappa_{s,t} \eta_t}{\sum_{t' \in [r]} \kappa_{s,t'} \eta_{t'}} - 2 \eta_s.$$  

Then, the half-edge label distribution satisfies

$$\eta_n(s) \xrightarrow{a.s.} \eta^*(s), \quad \text{where} \quad h_s(\eta^*) = 0 \quad \forall s \in [r].$$

The probability distribution $\eta^*$ can be shown to be the unique probability distribution that solves $h_s(\eta^*) = 0$ for all $s \in [r]$. We next define the crucial parameters in the model.
For $s, t \in [r]$, let

$$\theta_{s,t} = \frac{\kappa_{s,t}}{\sum_{t' \in [r]} \kappa_{s,t'} \eta^*(t')}.$$  \hfill (9.3.24)

and write

$$\theta_s = \sum_{t \in [r]} \mu_t \theta_{s,t}^*.$$  \hfill (9.3.25)

For a vertex $v \in [n]$, we let $\sigma(v) \in [r]$ be its type, and $n_s = \# \{ v : \sigma(v) = s \}$ denote the type counts. We next study the degree distribution in the above preferential attachment models with global communities. For $s \in [r]$, define

$$P_s(n) = \frac{1}{n} \sum_{v \in [n]} 1_{\{ D_v(n) = k, \sigma(v) = s \}}$$  \hfill (9.3.26)

where $\theta_s$ is defined in (9.3.25), and where, for $\theta > 0$,

$$p_k(\theta) = \frac{\Gamma(m + 1/\theta)}{\theta \Gamma(m)} \frac{\Gamma(k)}{\Gamma(k + 1 + 1/\theta)}.$$  \hfill (9.3.28)

Exercise 9.31 shows that the limiting distribution in (9.3.28) is indeed a probability distribution. Theorem 9.11 shows that the degree distribution has a power-law tail, as can be expected from the fact that the preferential attachment mechanism is profoundly present. Moreover, (9.3.27) also shows that the degree of vertices of type $s$ satisfies a power-law with exponent that depends sensitively on the type through the key parameters $(\theta_s^*)_{s \in [r]}$. Exercise 9.32 shows that the global degree distribution also converges, as can be expected by the convergence of the per-type degree distribution. Further, Exercise 9.33 shows that the global degree distribution has a power-law tail with exponent $\tau = 1 + 1/\max_{s \in [r]} \theta_s^*$, provided that $\mu_s > 0$ for all $s \in [r]$.

Related properties of the preferential attachment models with global communities seem not to have been investigated, so we move to the community detection problem.
9.4 Random graphs with community structure: local communities

Community detection in PAMs with global communities

It is important to discuss what we assume to be known. We think of the graph as being directed, edges being directed from young to old. Further, we assume that \( m \) and the probability distribution \((\mu_s)_{s \in [r]}\) are known. Finally, and probably most importantly, we assume that the labels or age of the vertices in the graph are known. Some of these parameters can be estimated \((m \) being the easiest one).

While for stochastic blockmodels, it is clear that the case where the communities are all equally large and have the same inter- and intra-community edge probabilities, this is not obvious here. However, to mimic the stochastic blockmodel setting, you can take the case where \( \mu_s = 1/r \) in mind as a key example. Also, the setting where \( \kappa_{s,s} = a \) for all \( s \in [r] \) and \( \kappa_{s,t} = b \) for all \( s,t \in [r] \) with \( s \neq t \) is particularly interesting, where, to model community structure, \( a > b \) is natural. By scaling invariance, we may assume that \( b = 1 \). In this case, \( \eta^*_s = 1/r \) for all \( s \in [r] \), \( \theta^*_{s,t} = ar/[2(a + r - 1)] \) for all \( s,t \in [r] \) with \( s \neq t \), while \( \theta^*_{s,s} = a/[2(a + r - 1)] \).

Also, \( \theta^*_{s} = 1/2 \) for all \( s \in [r] \). This case is probably the most difficult.

the aim is to estimate the type \( \sigma(v) \) for all \( v \in [n] \), based on the above information. This can be done by several algorithms. One algorithm performs the estimation of the label of \( v \) on the basis of the neighbors of \( v \). A second algorithm does it on the basis of the degree of \( v \) at time \( n \). Let \( \text{Err}_n \) denote the fraction of errors in the above algorithms. In the above setting, it can be shown that

\[
\text{Err}_n \xrightarrow{\mathcal{P}} \text{Err},
\]

for some limiting constant \( \text{Err} \) depending on the algorithm. The precise form of \( \text{Err} \) is known, but difficult to obtain rigorously as it relies on a continuous-time branching process approximation of the graph evolution. In particular, it is not obvious that, in the setting where \( \mu_s = 1/r, \kappa_{s,s} = a > 1 \) for all \( s \in [r] \) and \( \kappa_{s,t} = 1 \) for all \( s,t \in [r] \), it is unclear whether \( \text{Err} < (r - 1)/r \).

Many more detailed result are proved, for example that the probability that a vertex label of vertex \( t \) is estimated wrongly converges uniformly for all \( t \in [n] \setminus [\delta n] \). Also, there exists an algorithm that estimates \( \sigma(v) \) correctly whp provided that \( v = o(n) \). We refrain from discussing this further.

9.4 Random graphs with community structure: local communities

In the previous section, we investigated settings where the models have a finite number of communities, making the communities global. This setting is realistic when we would like to partition a network of choice into a finite number of parts, for example corresponding to the main scientific fields in citation or collaboration networks, or the continents in the Internet. However, in many other settings this is not realistic. Indeed, most communities of social networks correspond to smaller entities, such as school classes, families, sports teams, etc. In most real-world settings, it is not even clear what communities look like. As a result, community
detection has become an art. Needless to say, such techniques are highly relevant, to find and use the available structure in real-world networks.

The topic is relevant, since most models (including the models with global community structure from Section 9.4) have rather low clustering. For example, consider a general inhomogeneous random graph $IRG_n(\kappa_n)$ with kernel $\kappa_n$. Assume that $\kappa_n(x, y) \leq n$. Then, we can compute that the expected number of triangles in an $IRG_n(\kappa_n)$ is equal to

$$E[\# \text{ triangles in } IRG_n(\kappa)] = \frac{1}{6} \sum_{i,j,k \in [n]} \kappa_n(x_i, x_j)\kappa_n(x_j, x_k)\kappa_n(x_k, x_i).$$

(9.4.1)

Under relatively weak conditions on the kernel $\kappa_n$, it follows that

$$E[\# \text{ triangles in } IRG_n(\kappa_n)] \rightarrow \frac{1}{6} \int_{S^3} \kappa(x_1, x_2)\kappa(x_2, x_3)\kappa(x_3, x_1)\mu(dx_1)\mu(dx_2)\mu(dx_3).$$

(9.4.2)

Therefore, the clustering coefficient converges to zero as $1/n$. In many real-world networks, particularly in social networks, the clustering coefficient is strictly positive.

In this section, we discuss random graph models with a community structure in which most communities are quite small. Indeed, in most cases it is assumed that the average community size is bounded. However, since the community sizes can also be quite large, we might call them mesoscopic rather than microscopic. See Figures 9.7 and 9.8 for some empirical data on real-world networks. Figure 9.7 shows that there is enormous variability in the tail probabilities of the degree distribution, community-size distribution and inter-community degrees in real-world networks. Figure 9.8 shows that the edge-densities of communities generally decreases with their sizes. Here, the edge density of a community of size $s$ is given by $2e_in/[s(s-1)]$, where $e_in$ is the number of edges inside the community. Here, the communities were extracted (or detected) using the so-called Louvain method. See Stegehuis et al. (2016b) for details.

We conclude that we need to be flexible in our community structure to be able to realistically model the community structure in real-world networks. Below, we discuss several models that attempt to do so. We start in Section 9.4.1 by discussing inhomogeneous random graphs with community structures. We continue in Section 9.4.2 by discussing the hierarchical configuration model as well as some close cousins, followed by a discussion of random intersection graphs in Section 9.4.3 and exponential random graphs in Section 9.4.4.

9.4.1 INHOMOGENEOUS RANDOM GRAPHS WITH COMMUNITIES

In this section, we discuss a model similar to the inhomogeneous random graph $IRG_n(\kappa)$ that incorporates clustering. The idea behind this model is that instead of only adding edges independently, we can also add other graphs on $r$ vertices in an independent way. For example, we could study a graph where each pair of vertices is independently connected with probability $\lambda/n$, as for $ER_n(\lambda/n)$, but also
Figure 9.7 Tail probabilities of the degrees, community sizes $s$ and the inter-community degrees $k$ in real-world networks. a) Amazon co-purchasing network, b) Gowalla social network, c) English word relations, d) Google web graph. Pictures taken from Stegehuis et al. (2016b).

each collection of triples forms a triangle with probability $\mu/n^2$, independently for all triplets and independently of the status of the edges. Here the exponent $n^{-2}$ is chosen as to make the expected number of triangles containing a vertex bounded. See Exercise 9.34 for the clustering in such random graphs. In social networks, also complete graphs of size four, five, etc., are present more often than in usual random graph. Therefore, we also wish to add those independently.

We start by introducing the model, followed by an examination of some of its properties.

Model introduction

In order to formulate this general version of the model, we start by introducing some notation. We repeatedly make use of similar notation as in Chapter 3.

Let $\mathcal{F}$ consist of one representative of each isomorphism class of finite connected graphs, chosen so that if $F \in \mathcal{F}$ has $r$ vertices then $V(F) = \{1, 2, \ldots, r\}$. Simple examples of such $F$ are the complete graphs on $r$ vertices, but also other examples are possible. Recall that $\mathcal{S}$ denotes the type space. Given $F \in \mathcal{F}$ with $r$ vertices, let $\kappa_F$ be a measurable function from $\mathcal{S}^r$ to $[0, \infty)$. The function $\kappa_F$ is called the kernel corresponding to $F$. A sequence $\vec{\kappa} = \{\kappa_F\}_{F \in \mathcal{F}}$ is a kernel family.

Let $\vec{\kappa}$ be a kernel family and $n$ an integer. We define a random graph $\text{IRG}_n(\vec{\kappa})$ with vertex set $[n]$. First let $x_1, x_2, \ldots, x_n \in \mathcal{S}$ be i.i.d. (independent and identically distributed) with the distribution $\mu$. Given $\mathbf{x} = (x_1, \ldots, x_n)$, construct $\text{IRG}_n(\vec{\kappa})$ as follows, starting with the empty graph. For each $r$ and each $F \in \mathcal{F}$
Figure 9.8 The relation between the denseness of a community $2c_m/(s^2-s)$ and the community size $s$ can be approximated by a power law. a) Amazon co-purchasing network, b) Gowalla social network, c) English word relations, d) Google web graph. Pictures taken from Stegehuis et al. (2016b).

with $|F| = r$, and for every $r$-tuple of distinct vertices $(v_1, \ldots, v_r) \in [n]^r$, add a copy of $F$ on the vertices $v_1, \ldots, v_r$ (with vertex $i$ of $F$ mapped to $v_i$) with probability

$$p(v_1, \ldots, v_r; F) = \left( \frac{\kappa_F(x_{v_1}, \ldots, x_{v_r})}{n^{r-1}} \right) \wedge 1,$$

(9.4.3)

all these choices being independent. We shall often call the added copies of the various $F$ that together form $\text{IRG}_n(\hat{\kappa})$ atoms as they may be viewed as indivisible building blocks. Sometimes we refer to them as small graphs, although there is in general no bound on their sizes.

The reason for dividing by $n^{r-1}$ in (9.4.3) is that we wish to consider sparse graphs; indeed, our main interest is the case when $\text{IRG}_n(\hat{\kappa})$ has $O(n)$ edges. As it turns out, we can be slightly more general; however, when $\kappa_F$ is integrable (which we shall always assume), the expected number of added copies of each graph $F$ is $O(n)$. Note that all incompletely specified integrals are with respect to the appropriate $r$-fold product measure $\mu^r$ on $S^r$.

There are several plausible choices for the normalization in (9.4.3). The one we have chosen means that if $\kappa_F = c$ is constant, then (asymptotically) there are on average $cn$ copies of $F$ in total, and each vertex is on average in $rc$ copies of $F$. An alternative is to divide the expression in (9.4.3) by $r$; then (asymptotically)
each vertex would on average be in \( c \) copies of \( F \). Another alternative, natural when adding cliques only but less so in the general case, would be to divide by \( r! \); this is equivalent to considering unordered sets of \( r \) vertices instead of ordered \( r \)-tuples. When there is only one kernel, corresponding to adding edges, this would correspond to the normalization used in Bollobás et al. (2007), and in particular to that of the classical model \( \text{ER}_n(\lambda/n) \); the normalization we use here differs from this by a factor of 2.

In the special case where all \( \kappa_F \) are zero apart from \( \kappa_{K_2} \), the kernel corresponding to an edge, we recover (essentially) a special case of the model of Bollobás et al. (2007); we call this the edge-only case, since we add only edges, not larger graphs. We write \( \kappa_2 \) for \( \kappa_{K_2} \). Note that in the edge-only case, given \( x \), two vertices \( i \) and \( j \) are joined with probability

\[
\frac{\kappa_2(x_i, x_j) + \kappa_2(x_j, x_i)}{n} + O\left(\frac{\left(\kappa_2(x_i, x_j) + \kappa_2(x_j, x_i)\right)^2}{n^2}\right).
\]

The correction term will never matter, so we may as well replace \( \kappa_2 \) by its symmetrized version. In fact, we shall always assume that \( \kappa_F \) is invariant under permutations of the vertices of the graph \( F \).

For any kernel family \( \tilde{\kappa} \), let \( \kappa_e \) be the corresponding edge kernel, defined by

\[
\kappa_e(x, y) = \sum_{F} \sum_{ij \in E(F)} \int_{S^{|F| \setminus \{i, j\}}} \kappa_F(x_1, \ldots, x_{i-1}, x, x_{i+1}, \ldots, x_{j-1}, y, x_{j+1}, \ldots, x_{|F|}),
\]

where the second sum runs over all \( 2E(F) \) ordered pairs \((i, j)\) with \( ij \in E(F) \), and we integrate over all variables apart from \( x \) and \( y \). Note that the sum need not always converge; since every term is positive this causes no problems: we simply allow \( \kappa_e(x, y) = \infty \) for some \( x, y \). Given \( x_i \) and \( x_j \), the probability that \( i \) and \( j \) are joined in \( \text{IRG}_n(\tilde{\kappa}) \) is at most \( \kappa_e(x_i, x_j)/n + O(1/n^2) \). In other words, \( \kappa_e \) captures the edge probabilities in \( \text{IRG}_n(\tilde{\kappa}) \), but not the correlations.

**The number of edges**

Before proceeding to deeper properties, let us note that the expected number of added copies of \( F \) is \((1 + O(n^{-1}))n \int_{S^{|F|}} \kappa_F \). Unsurprisingly, the actual number turns out to be concentrated about this mean. Let

\[
\xi(\tilde{\kappa}) = \sum_{F \in \mathcal{F}} E(F) \int_{S^{|F|}} \kappa_F = \frac{1}{2} \int_{S^2} \kappa_e \leq \infty
\]

be the asymptotic edge density of \( \tilde{\kappa} \). Since every copy of \( F \) contributes \( E(F) \) edges, the following theorem is almost obvious, provided that we can ignore overlapping edges:

**Theorem 9.12** (Edge density in \( \text{IRG}_n(\tilde{\kappa}) \))  
As \( n \to \infty \),

\[
E[E(\text{IRG}_n(\tilde{\kappa}))/n] \to \xi(\tilde{\kappa}) \leq \infty.
\]

Moreover, \( \xi(\tilde{\kappa}) < \infty \), then

\[
E(\text{IRG}_n(\tilde{\kappa}))/n \xrightarrow{p} \xi(\tilde{\kappa}).
\]
In other words, Theorem 9.12 states that if $\xi(\kappa) < \infty$ then $E(\text{IRG}_n(\kappa)) = \xi(\kappa)n + o(n)$, and if $\xi(\kappa) = \infty$ then $E(\text{IRG}_n(\kappa)) > Cn$ whp for every constant $C$. We conclude that the model is sparse when $\xi(\kappa)$. This is certainly true when $\kappa_F$ is uniformly bounded, but may also be true more generally under some integrability assumptions.

The giant

We next consider the emergence of the giant component. When studying the component structure of $\text{IRG}_n(\kappa)$, the model can be simplified somewhat. Recalling that the atoms $F \in \mathcal{F}$ are connected by definition, when we add an atom $F$ to a graph $G$, the effect on the component structure is simply to unite all components of $G$ that meet the vertex set of $F$, so only the vertex set of $F$ matters, not its graph structure. We say that $\kappa$ is a clique kernel family if the only non-zero kernels are those corresponding to complete graphs; the corresponding random graph model $\text{IRG}_n(\kappa)$ is a clique model. For questions corresponding to component structure, it suffices to study clique models. For clique kernels we write $\kappa_r$ for $\kappa_K_r$; as above, we always assume that $\kappa_r$ is symmetric, here meaning invariant under all permutations of the coordinates of $S^r$. Given a general kernel family $\kappa$, the corresponding (symmetrized) clique kernel family is given by $\tilde{\kappa} = (\kappa_r)_{r \geq 2}$, with

$$\kappa_r(x_1, \ldots, x_r) = \sum_{F \in \mathcal{F} : |F| = r} \frac{1}{r!} \sum_{\pi \in \Theta_r} \kappa_F(x_{\pi(1)}, \ldots, x_{\pi(r)}), \quad (9.4.9)$$

where $\Theta_r$ denotes the symmetric group of permutations of $[r]$. (This is consistent with our notation $\kappa_2 = \kappa_{K_2}$.) When considering the size (meaning number of vertices) of the giant component in $\text{IRG}_n(\kappa)$, we may always replace $\kappa$ by the corresponding clique kernel family.

In our analysis we also consider the linear operator $T_{\kappa_c}$ defined by

$$T_{\kappa_c}(f)(x) = \int_{S^r} \kappa_c(x, y)f(y)d\mu(y), \quad (9.4.10)$$

where $\kappa_c$ is defined by (9.4.5). We need to impose some sort of integrability condition on our kernel family:

**Definition 9.13** (i) A kernel family $\kappa = (\kappa_F)_{F \in \mathcal{F}}$ is integrable if

$$\int \kappa = \sum_{F \in \mathcal{F}} |F| \int_{S^{|F|}} \kappa_F < \infty. \quad (9.4.11)$$

This means that the expected number of atoms containing a given vertex is bounded.

(ii) A kernel family $\kappa = (\kappa_F)_{F \in \mathcal{F}}$ is edge integrable if

$$\sum_{F \in \mathcal{F}} E(F) \int_{S^{|F|}} \kappa_F < \infty; \quad (9.4.12)$$

equivalently, $\xi(\kappa) < \infty$. This means that the expected number of edges in
9.4 Random graphs with community structure: local communities 391

IRG\(_n(\tilde{\kappa})\) is \(O(n)\), see Theorem 9.12, and thus the expected degree of a given vertex is bounded.

Note that a clique kernel \((\kappa_r)\) is integrable if and only if \(\sum r \geq 2^r \int S_r \kappa_r < \infty\), and edge integrable if and only if \(\sum r(r-1) \int S_r \kappa_r < \infty\).

The main results concerning the phase transition on IRG\(_n(\tilde{\kappa})\) is that if \(\tilde{\kappa}\) is an integrable kernel family satisfying a certain extra assumption, then the normalized size of the giant component in IRG\(_n(\tilde{\kappa})\) is simply \(\zeta(\tilde{\kappa}) + o_P(1)\).

We are now ready to formulate the main result in this section involving the phase transition in IRG\(_n(\tilde{\kappa})\). Recall that \(|C_{\text{max}}|\) denotes the number of vertices in the largest connected component of the graph under consideration, and \(|C_{(2)}|\) the number of vertices in its second largest component.

Theorem 9.14 (The giant in clustered inhomogeneous random graphs)  Let \(\tilde{\kappa}' = (\kappa_F')_{F \in \mathcal{F}}\) be an irreducible, integrable kernel family, and let \(\tilde{\kappa} = (\kappa_r)_{r \geq 2}\) be the corresponding clique kernel, given by (9.4.9). Then, there exists a \(\zeta(\tilde{\kappa}) \in [0,1)\) such that

\[
|C_{\text{max}}| = \zeta(\tilde{\kappa}) n + o_P(n),
\]

and \(|C_{(2)}| = o_P(n)\).

Theorem 9.14 is proved by showing that (in the clique kernel case) the branching process that captures the ‘local structure’ of IRG\(_n(\tilde{\kappa})\) is a good approximation, and indeed describes the existence of the giant. Along the way, it is also proved that, in the clique kernel setting, IRG\(_n(\tilde{\kappa})\) converges in the local weak sense to this ‘branching process’. This is however not a branching process in the usual sense, but instead a branching process describes the connections between the cliques.

For Theorem 9.14 to be useful we would like to know something about \(\zeta(\tilde{\kappa})\), which can be calculated from \(x \mapsto \zeta(x)\), which is in turn the largest solution to the functional equation

\[
f(x) = 1 - e^{-S(\kappa_f)(x)}.
\]

We can think of \(\zeta(\tilde{\kappa})\) as the probability that a vertex of type \(x \in \mathcal{S}\) has a ‘large’ connected component. The question when \(\zeta(\tilde{\kappa}) > 0\) is settled in the following theorem:

Theorem 9.15 (Condition for existence giant component)  Let \(\tilde{\kappa}\) be an integrable clique kernel. Then, \(\zeta(\tilde{\kappa}) > 0\) if and only if \(|T_{\kappa_c}| > 1\). Furthermore, if \(\tilde{\kappa}\)
is irreducible and \( \| \mathbf{T}_{\kappa_r} \| > 1 \), then \( \zeta_{\kappa}(x) \) is the unique non-zero solution to the functional equation (9.4.14), and \( \zeta_{\kappa}(x) > 0 \) holds for a.e. \( x \).

In general, \( \| \mathbf{T}_{\kappa_e} \| \) may be rather hard to calculate. When we suppose that each \( \kappa_r \) is constant, however, this can be done. Indeed, say that \( \kappa_r = c_r \). Then \( \kappa_e(x, y) = \sum_r r(r-1)c_r = 2\zeta(\kappa) \) for all \( x \) and \( y \), so that

\[
\| \mathbf{T}_{\kappa_e} \| = 2\zeta(\kappa). \tag{9.4.15}
\]

This is perhaps surprising: it tells us that for such uniform clique kernels, the critical point where a giant component emerges is determined only by the total number of edges added; it does not matter what size cliques they lie in, even though, for example, the third edge in every triangle is ‘wasted’. This turns out not to be true for arbitrary kernel families, where, rather each atom needs to be replaced by a clique.

### 9.4.2 Configuration models with community structure

In this section, we investigate several models that are related to configuration model, yet have a pronounced community structure. We start by discussing the hierarchical configuration model and its properties. We then discuss a particular version that goes under the name of the household model, and we close by discussing a model where triangles are explicitly added to the model.

#### Hierarchical configuration model: model introduction

Also the configuration model has low clustering, which often makes it inappropriate in applied contexts. A possible solution to overcome this low clustering is by introducing a community or household structure. Consider the configuration model \( \text{CM}_N(\mathbf{d}) \) with a degree sequence \( \mathbf{d} = (d_i)_{i \in [N]} \) satisfying Condition 1.5(a)-(b), now with \( N \) taking the role of \( n \). Now we replace each of the vertices by a small graph. Thus, vertex \( i \) is replaced by a local graph \( G_i \). We assign each of the \( d_i \) half-edges incident to vertex \( i \) to a vertex in \( G_i \) in an arbitrary way. Thus, vertex \( i \) is replaced by the pair of the community graph \( G_i = (V(G_i), E(G_i)) \) and the inter-community degrees \( \mathbf{d}^{(b)} = (d^{(b)}_u)_{u \in V(G_i)} \) satisfying that \( \sum_{u \in V(G_i)} d^{(b)}_u = d_i \).

Naturally, the size of the graph becomes \( n = \sum_{i \in [N]} |V(G_i)| \).

As a result, we obtain a graph with two levels of hierarchy, whose local structure is described by the local graphs \( (G_i)_{i \in [N]} \), whereas its global structure is described by the configuration model \( \text{CM}_N(\mathbf{d}) \). This model is called the hierarchical configuration model. A natural assumption is that the degree sequence \( \mathbf{d} = (d_i)_{i \in [N]} \) satisfies Condition 1.5(a)-(b), while the empirical distribution of the graphs satisfies that, as \( N \to \infty \),

\[
\mu_n(H, \bar{d}) = \frac{1}{N} \sum_{i \in [N]} 1_{\{G_i = H, (d^{(b)}_u)_{u \in V_i} = \bar{d}\}} \to \mu(H, \bar{d}), \tag{9.4.16}
\]

for some probability distribution on graphs with integer marks associated to the vertices. We assume that \( \mu_n(H, \bar{d}) = 0 \) for all \( H \) that are disconnected. Indeed,
we think of the graphs \((G_i)_{i \in [N]}\) as describing the community structure of the graph, so it makes sense to assume that \((G_i)_{i \in [N]}\) are connected. In particular, (9.4.16) shows that a typical community has bounded size. We will often also make assumptions on the average size of the community of a random vertex. For this, it will be necessary to impose that, with \(\mu_n(H) = \sum d \mu_n(H, d)\) and \(\mu(H) = \sum d \mu(H, d)\) the community distribution,

\[
\sum_H |V(H)| \mu_n(H) = \frac{1}{n} \sum_{i \in [N]} |V(G_i)| \to \sum_H |V(H)| \mu(H) < \infty, 
\tag{9.4.17}
\]

for every connected graph \(H\) of size \(|V(H)|\) and degree vector \(d = (d_h)_{h \in |V(H)|}\).

Equation (9.4.17) indicates that the community of a random individual has a tight size, since the community size of a random vertex has the size-biased community distribution (see Exercise 9.36).

The degree structure of the hierarchical configuration model is built into the model, and is thus of less interest. We next discuss the giant and distances in the hierarchical configuration model.

The giant in the hierarchical configuration model

We now investigate the existence and size of the giant component in the hierarchical configuration model. Let \(C_{\text{max}}\) be the largest connected component in the hierarchical configuration model, and let \(C_{(2)}\) denote the second largest cluster (breaking ties arbitrarily when needed). The main result concerning the size of the giant is the following theorem:

**Theorem 9.16** (Giant in the hierarchical configuration model) Assume that the inter-community degree sequence \(d = (d_i)_{i \in [N]}\) satisfies Conditions 1.5(a)-(b), while the communities satisfy (9.4.16) and (9.4.17). Then, there exists \(\zeta \in [0, 1]\) such that

\[
\frac{1}{n} |C_{\text{max}}| \xrightarrow{\mathbb{P}} \zeta, \quad \frac{1}{n} |C_{(2)}| \xrightarrow{\mathbb{P}} 0. \quad (9.4.18)
\]

Let \(D\) denote the asymptotic degree distribution \(d = (d_i)_{i \in [N]}\) in Conditions 1.5(a)-(b), and write \(\nu = \mathbb{E}[D(D - 1)]/\mathbb{E}[D]\). Then, \(\zeta > 0\) precisely when \(\nu > 1\).

Since the communities \((G_i)_{i \in [N]}\) are connected, the sizes of the clusters in the hierarchical configuration are closely related to those in \(\text{CM}_N(d)\). Indeed, for \(v \in [n]\), let \(i_v\) denote the vertex for which \(v \in V(G_{i_v})\). Then,

\[
|C(v)| = \sum_{i \in C'(i_v)} |V(G_i)|, \tag{9.4.19}
\]

where \(C'(i)\) denotes the connected component of \(i\) in \(\text{CM}_N(d)\). This allows one to move back and forth between the hierarchical configuration model and its corresponding configuration model \(\text{CM}_N(d)\) that describes the inter-community connections.
It also allows us to identify the limit \( \zeta \). Let \( \xi \in [0,1] \) be the extinction probability of the local weak limit of \( \text{CM}_N(d) \) of a vertex of degree 1, so that a vertex of degree \( d \) survives with probability \( 1 - \xi^d \).

Then,

\[
\zeta = \sum_H \sum_d |V(H)| \mu(H)[1 - \xi^d],
\]

where \( d = \sum_{h \in V(H)} d_h \). Further, \( \xi = 1 \) precisely when \( \nu \leq 1 \), see e.g., Theorem 4.4. This explains the result in Theorem 9.16. In Exercise 9.38, you are asked to fill in the details.

Before moving to graph distances, we discuss an example of the hierarchical configuration model that has attracted attention under the same configuration model with household structure.

### The configuration model with household structure

In the configuration model with household structure, each \( G_i \) is a complete graph of size \( |V(G_i)| \), while each inter-community degree equals 1, i.e., \( d_u = 1 \) for all \( u \in V(G_i) \). As a result, for \( v \in [n] \), its degree is equal to \( |V(G_{i_v})| \), where we recall that \( i_v \) is such that \( v \in V(G_{i_v}) \). This is a rather interesting example, where every vertex is in a ‘household’, and every household member has precisely one connection to the outside, which connects it to another household. In this case, the degree distribution is equal to

\[
F_n(x) = \frac{1}{n} \sum_{v \in [n]} \mathbb{1}_{|V(G_{i_v})| \leq x} = \frac{1}{n} \sum_{i \in [N]} \mathbb{1}_{|V(G_i)| \leq x},
\]

where we recall that \( i_v \) is such that \( v \in G_{i_v} \). As a result, the degree distribution in the household model is the size-biased degree in the configuration model that describes its inter-community structure. In particular, this implies that if the limiting degree distribution \( D \) in \( \text{CM}_N(d) \) is a power-law with exponent \( \tau' \), then that in the household model is a power-law with exponent \( \tau = \tau' - 1 \). This is sometimes called a power-law shift, and is clearly visible in Figure 9.9.

### Graph distances in the hierarchical configuration model

We next turn ourselves to graph distances in the hierarchical configuration model. Again, a link can be made to the distances in the configuration model, particularly when the diameters of the community graphs \( (G_i)_{i \in [N]} \) are uniformly bounded. Indeed, in this case, one can expect the graph distances in the hierarchical configuration model to be of the same order of magnitude as in the configuration model \( \text{CM}_N(d) \).

This link is the easiest to formulate for the household model. Indeed, there the diameter of the community graphs is 1 since all community graphs are complete graphs. Denote the hierarchical configuration model by \( \text{HCM}_n(G) \). Then, unless \( u = v \) or the half-edge incident to \( u \) being paired to that incident to \( v \),

\[
\text{dist}_{\text{HCM}_n(G)}(u,v) = 2\text{dist}_{\text{CM}_N(d)}(i_u,i_v) - 1,
\]

where we recall that \( i_v \) is such that \( v \in V(G_{i_v}) \). Thus, distances in \( \text{HCM}_n(G) \) are asymptotically twice as large as
Figure 9.9 The degree distribution of a household model follows a power law with a smaller exponent than the community size distribution and outside degree distribution.

Those in CM_{N}(d). The reason is that paths in the household model alternative between intra-community edges and inter-community edges. This is because the inter-community degrees are all equal to 1, so there is no way to jump to a vertex using an inter-community edge and leave through one again. This is in general different.

Configuration model with clustering

We close this section on adaptations of the configuration model with local communities by introducing a model that has not attracted a lot of attention yet in the mathematical community.

The low clustering of CM_{n}(d) can be resolved by introducing households as described above. Alternatively, and in the spirit of clustered inhomogeneous random graphs as described in Section 9.4.1, we can also introduce clustering directly. In the configuration model with clustering, we assign two types of ‘degrees’ to a vertex \( i \in [n] \). We let \( d_i^{(o)} \) denote the number of simple half-edges incident to vertex \( i \), and we let \( d_i^{(v)} \) denote the number of triangles that vertex \( i \) is part of. In this terminology, the degree \( d_v \) of a vertex is equal to \( d_v = d_i^{(o)} + 2d_i^{(v)} \). We can then say that there are \( d_i^{(o)} \) half-edges incident to vertex \( i \), and \( d_i^{(v)} \) third-triangles.

The graph is built by (a) recursively choosing two half-edges uniformly at random without replacement, and pairing them into edges (as for CM_{n}(d)); and (b) choosing triples of third-triangles uniformly at random and without replacement, and drawing edges between the three vertices incident to the third-triangles that are chosen.

Let \((D_n^{(o)}, D_n^{(v)})\) denote the number of simple edges and triangles incident to
a uniform vertex in \([n]\), and assume that \((D_n^{(si)}, D_n^{(tr)}) \xrightarrow{d} (D^{(si)}, D^{(tr)})\) for some limiting distribution \((D^{(si)}, D^{(tr)})\). Newman (2009) performs a generating function analysis of when a giant component is expected to exist. The criterion Newman finds is that

\[
\left( \frac{\mathbb{E}[(D^{(si)})^2]}{\mathbb{E}[D^{(si)}]} - 2 \right) \left( \frac{2\mathbb{E}[(D^{(tr)})^2]}{\mathbb{E}[D^{(tr)}]} - 3 \right) < \frac{2\mathbb{E}[D^{(si)} D^{(tr)}]}{\mathbb{E}[D^{(si)}] \mathbb{E}[D^{(tr)}]},
\]

(9.4.22)

When \(D^{(tr)} = 0\) a.s., so that there are no triangles, this reduces to

\[
\frac{\mathbb{E}[(D^{(si)})^2]}{\mathbb{E}[D^{(si)}]} - 2 > 0,
\]

(9.4.23)

which is equivalent to \(\nu = \mathbb{E}[D^{(si)}(D^{(si)} - 1)]/\mathbb{E}[D^{(si)}] > 1\).

It would be of interest to analyze this model mathematically. While the extra triangles do created extra clustering in the graph, in that the graph is no longer locally treelike, it is less clear what the community structure of the graph is. Of course, the above setting can be generalized to arbitrary cliques and possible other community structure, but this will make the mathematical analysis substantially more involved.

### 9.4.3 Random intersection graphs

In most of the above models, the local communities that vertices are part of partition the vertex space. However, in most real-world applications, particularly in social networks, the vertices are not just part of one community, but often of several. We now present a model, the random intersection graph, where the group memberships are rather general. In a random intersection graph, vertices are members of groups. The group memberships arise in a random way. Once the group memberships are chosen, the random intersection graph is constructed by giving an edge to two vertices precisely when they are both members of the same group. Formally, let \(V(G_n)\) denote the vertex set, and \(A(G_n)\) the collection of groups. Let \(M(G_n) = \{(v, a) : v \text{ is in group } a \subseteq V(G_n) \times A(G_n)\}\) denote the group memberships. Then, the edge set of the random intersection graph with these group memberships is

\[
E(G_n) = \{\{u, v\} : (u, a), (v, a) \in M(G_n) \text{ for some } a \in A(G_n)\}.
\]

(9.4.24)

Thus, the random intersection graph is a deterministic function of the group memberships. In turn, the groups give rise to a community structure in the resulting network. Since vertices can be in several groups, the groups are no longer a partition. It is even possible that pairs of vertices are both in several groups, even though we will see that this is rare.

The group memberships occur through some random process. There are many possibilities for this. We will discuss several different versions of the model, as they have been introduced in the literature, and then will focus on one particular type of model to state the main results. In general, we do wish that our models are sparse. Suppose that the probability that vertex \(v\) is part of group \(a\) equals
9.4 Random graphs with community structure: local communities

$p_{ia}$, and that group $a$ has on average $m_a$ group elements. Then the average degree of vertex $i$ equals

$$\mathbb{E}[D_i] = \sum_{a \in A(G_n)} p_{va}(m_a - 1),$$  \hspace{1cm} (9.4.25)

which we aim to keep bounded or converge. We now discuss several different choices.

**Random intersection graphs with independent group-memberships**

The most studied model is when there are $n$ vertices, $m = m_n$ groups, often with $m = \beta n^\alpha$ for some $\alpha > 0$, and each vertex is independently connected to each group with probability $p_n$. In this case, $p_{va} = p_n$ and $m_a = np_a$, so that (9.4.25) turns into

$$\mathbb{E}[D_i] \approx n p^2_n m = \beta n^{1+\alpha} p_n^2,$$  \hspace{1cm} (9.4.26)

and choosing $p = \gamma n^{-(1+\alpha)/2}$ yields expected degree $\beta \gamma^2$.

A more flexible version is obtained by giving a weight $w_v$ to each of the vertices, for example in an i.i.d. way, and then taking

$$p_{va} = (\gamma w_v p_n) \wedge 1,$$  \hspace{1cm} (9.4.27)

and making all the edges between vertices and groups conditionally independent given the weights $(w_i)_{i \in [n]}$. In the theorem below, we assume that $(w_i)_{i \in [n]}$ is a sequence of i.i.d. random variables with finite mean:

**Theorem 9.17** (Degrees in random intersection graph with i.i.d. vertex weights)

Consider the above random intersection graph, where the number of groups equal $m = \beta n^\alpha$, the vertices have i.i.d. weights $(w_v)_{v \in [n]}$ with law $W \sim F$ and finite mean, and where the edge probability equals $p_{va} = (\gamma w_v n^{-(1+\alpha)/2}) \wedge 1$. Then,

(a) $D_v \xrightarrow{p} 0$ when $\alpha < 1$;
(b) $D_v \xrightarrow{d} \sum_{i \geq 1} Y_i$ when $\alpha = 1$, where $(Y_i)_{i \geq 1}$ are i.i.d. Poisson $(\gamma)$ random variables and $X \sim \text{Poi}(\beta \gamma W)$;
(c) $D_v \xrightarrow{d} X$ where $X \sim \text{Poi}(\beta \gamma^2 W)$ when $\alpha > 1$.

Theorem 9.17 can be understood as follows: The expected number of groups that individual $v$ belongs to is roughly $\beta \gamma w_v n^{-(1+\alpha)/2}$. When $\alpha < 1$, this is close to zero, so that $D_v = 0$ whp. For $\alpha = 1$, the number of groups that $v$ is part of is close to Poisson with parameter $\beta \gamma w_v$, and the number of other individuals in each of these groups is approximately Poisson$(\gamma)$ distributed. For $\alpha > 1$, individual $v$ belongs to a number of groups that tends to infinity as $p_n m \approx \beta \gamma w_v n^{(\alpha-1)/2}$ when $n \to \infty$, while each group has expected size of vanishing size $n^{(1-\alpha)/2}$. This means that group sizes are generally zero or one, and these choice are asymptotically independent, giving rise to the Poisson distribution specified in part (c).
Random intersection graphs with prescribed groups

We next discuss a setting in which the number of groups per vertex and the group sizes are deterministic, and are obtained by randomly pairing vertices to groups. As such, the random intersection graph is obtained by (9.4.24), where now the edges between vertices in $V(G_n)$ and groups in $A(G_n)$ are modeled as a bipartite configuration model.

Above we have studied random intersection graphs, where connections are randomly and independently formed between individuals and groups. We now describe a model in which vertex $v \in [n]$ belongs to $d^{(ve)}_v$ groups, while group $g \in [m]$ has size $d^{(gr)}_g$. Here $n$ is the number of individuals, while $m$ is the number of groups.

Naturally, in order for the model to be well defined, we need that

$$\sum_{v \in [n]} d^{(ve)}_v = \sum_{a \in [m]} d^{(gr)}_a. \tag{9.4.28}$$

As in (9.4.24), we call two vertices $v_1$ and $v_2$ neighbors when they are connected to the same group, so that the degree of a vertex $v$ is the total number of other vertices $u$ for which there exists a group of which both $u$ and $v$ are members.

We will now focus on this particular setting, for concreteness. We refer to the extensive discussion in Section 9.6 for more details and references to the literature.

Local limit of random intersection graphs with prescribed groups

We now investigate the local limit of the giant component in the random intersection graph with prescribed degrees and groups:

**Theorem 9.18** (Local limit of random intersection graphs with prescribed groups)

Consider the random intersection graphs with prescribed groups, where the number of groups satisfies $m_n = \beta n$. Assume that the group membership sequence $d^{(ve)} = (d^{(ve)}_v)_{v \in [n]}$ and the group size sequence $d^{(gr)} = (d^{(gr)}_a)_{a \in [m]}$ both satisfy Conditions 1.5(a)-(b), where also (9.4.28) holds. Then, the model converges locally in probability to a so-called clique tree, where

- the number of groups that the root participate in has law $D^{(ve)}$, which is the limiting law of $d^{(ve)}$;
- the number of groups that every other vertex participates in has law $Y^* - 1$, where $Y^* - 1$ is the size-biased version of $D^{(ve)}$;
- the number of vertices per group are i.i.d. random variables with law $X^*$, where $X^*$ is the size-biased version of the limiting law of $d^{(gr)}$.

As a result of Theorem 9.18, the degree distribution of the random intersection graphs with prescribed groups is equal to

$$D = \sum_{i=1}^{d^{(ve)}} (X^*_i - 1), \tag{9.4.29}$$

which can be compared to Theorem 9.17(b). The intuition behind Theorem 9.18 is that the random intersection graph can easily be obtained from the bipartite
9.4 Random graphs with community structure: local communities

configuration model by making all group members direct neighbors. By construction, the local limit of the bipartite configuration model can be described by an alternating branching process of the size-biased vertex and group distributions.

The giant in random intersection graphs with prescribed degrees

We now investigate the existence and size of the giant component in the random intersection graph with prescribed degrees and groups. Let $C_{\text{max}}$ be its largest connected component, and $C_{(2)}$ denote its second largest cluster (breaking ties arbitrarily when needed). The main result concerning the size of the giant is the following theorem:

**Theorem 9.19** (Giant in random intersection graphs with prescribed degrees)

Consider the random intersection graphs with prescribed degrees and groups, where the number of groups satisfies $m_n = \beta n$. Assume that the group membership sequence $d^{(ve)}(v) = (d^{(ve)}(v))_{v \in [n]}$ and the group size sequence $d^{(gr)}(a) = (d^{(gr)}(a))_{v \in [m]}$ both satisfy Conditions 1.5(a)-(b), where also (9.4.28) holds. Then, there exists $\zeta \in [0, 1]$ such that

$$
\frac{1}{n} |C_{\text{max}}| \xrightarrow{P} \zeta, \quad \frac{1}{n} |C_{(2)}| \xrightarrow{P} 0, \quad (9.4.30)
$$

where $\zeta$ is the survival probability of the local limit in Theorem 9.18. Further, $\zeta > 0$ precisely when $\nu > 1$, where

$$
\nu = \frac{\mathbb{E}[(D^{(ve)} - 1)D^{(ve)}]}{\mathbb{E}[D^{(ve)}]} \frac{\mathbb{E}[(D^{(gr)} - 1)D^{(gr)}]}{\mathbb{E}[D^{(gr)}]}, \quad (9.4.31)
$$

In terms of the intuition right below Theorem 9.18, the survival probability of the random intersection graph is identical to that of the bipartite configuration model. Due to its alternating nature in odd and even generations, the bipartite configuration model needs to be treated with care. However, the number of individuals in odd or in even generations is a normal branching process with offspring distribution $\sum_{i=1}^{Y^{(ve)}_i} (X^{(ve)}_i - 1)$ for the even generations, and $\sum_{i=1}^{X^{(gr)}_i} (Y^{(gr)}_i - 1)$ for the odd generations. These branching processes have a positive survival probability when either of the two expected values $\nu$ satisfy $\nu > 1$. Obviously, the expected offsprings in the branching processes describing even and odd generations agree, by Wald’s identity. This leads us to $\nu$ in (9.4.31).

9.4.4 Exponential random graphs and maximal entropy

Suppose we have a real-world network, for which we observe a large amount of occurrences of certain subgraphs $F \in \mathcal{F}$. For example, while the model is sparse, we do see a linear amount of triangles. How can we model such a network? This is particularly relevant in the area of social science, where it was early on observed that social networks have much more clustering, i.e., many more triangles, than one might expect based on many of the classical models. This raises the question how to devise models that have similar features.

One solution may be to take the subgraph counts for granted, and use as a
model a random graph with precisely these subgraph counts. For example, consider-
ing the subgraphs to be the starts of any order, this would fix the degrees of all the vertices, which would leave us to the uniform graph with prescribed degrees. However, this model is notoriously difficult to work with, and even to simulate. It only becomes more difficult when taking more involved quantities, such as the number of triangles or the number of triangles per vertex, into account. Thus, this solution may be practically impossible. Also, it may be that the numbers we observe are merely the end product of a random process, so that we should see the realizations of a description of the mean of the values, so a soft constraint rather than a hard constraint.

The exponential random graph is a way to leverage the randomness, and still get a model that one can write down. Indeed, let $\mathcal{F}$ be a collection of subgraphs, and suppose that we observe that, on our favorite real-world network, the number of occurrences of subgraph $F$ equals $\alpha_F$ for every $F \in \mathcal{F}$. Let us now write out what this might mean. Let $F$ be a graph on $|V(F)| = m$ vertices. For a graph $G$ on $n$ vertices and $m$ vertices $v_1, \ldots, v_m$, let $G|_{(v_i), i \in [m]}$ be the subgraph spanned by $(v_i)_{i \in [m]}$. This means that the vertex set of $G|_{(v_i), i \in [m]}$ equals $[m]$, while its edge set equals $\{i, j \mid \{v_i, v_j\} \in E(G)\}$. The number of occurrences of $F$ in $G$ can then be written as

$$N_F(G) = \sum_{v_1, \ldots, v_m \in V(G)} \mathbb{1}_{G|_{(v_i), i \in [m]} = F}.$$ \hfill (9.4.32)

Here, it will be convenient to recall that we may equivalently write $G = (x_{i,j})_{1 \leq i < j \leq n}$, where $x_{i,j} \in \{0, 1\}$ and $x_{i,j} = 1$ if and only if $\{i, j\} \in E(G)$. Then, we can write $N_F(G) = N_F(x)$.

In order to define a measure, we can take a so-called exponential family of the form

$$p_{\beta}(x) = \frac{1}{Z_n(\beta)} e^{\sum_{F \in \mathcal{F}} \beta_F N_F(x)},$$ \hfill (9.4.33)

where $Z = Z_n(\beta)$ is the normalization constant

$$Z_n(\beta) = \sum_x e^{\sum_{F \in \mathcal{F}} \beta_F N_F(x)}.$$ \hfill (9.4.34)

In order to make sure that $p_{\beta}$ has the correct mean values for $N_F$, we choose $\beta = (\beta_F)_{F \in \mathcal{F}}$ as the solution to

$$\sum_x N_F(x) p_{\beta}(x) = \alpha_F \quad \text{for all } F \in \mathcal{F}.$$ \hfill (9.4.35)

In this case, the mean of $N_F(X)$, when $\mathbb{P}(X = x) = p_{\beta}(x)$, is exactly such that $\mathbb{E}[N_F(X)] = \alpha_F$ for all $F \in \mathcal{F}$. Further, when conditioning on $N_F(x) = q_F$ for some parameters $(q_F)_{F \in \mathcal{F}}$, the conditional exponential random graph is uniform over the set of graphs with this property. This is a conditioning property of exponential random graphs.
We next discuss two examples that we know quite well, and that arise as exponential random graphs with certain specific subgraph counts:

Example 9.20 (Example: ER$_n(\lambda/n)$) Take $N_F(x) = \sum_{i,j \in [n]} x_{ij} = |E(G_n)|$, so that we put a restriction on the expected number of edges in the graph. In this case, we see that, with $G_n = (x_{i,j})_{1 \leq i < j \leq n}$,

$$Z_n(\vec{\beta}) = \sum_x e^{\beta|E(G_n)|} = (1 + e^\beta)^\binom{n}{2},$$

(9.4.36)

and

$$p_{\vec{\beta}}(x) = \frac{1}{Z_n} e^{\sum_{v \in [n]} \beta_v d_v(G_n)} = \prod_{1 \leq i < j \leq n} \frac{e^{\beta x_{i,j}}}{1 + e^\beta}.$$  

(9.4.37)

Thus, the different edges are independent, and an edge is present with probability $e^\beta/(1 + e^\beta)$, and absent with probability $1/(1 + e^\beta)$. In the sparse setting, we aim that

$$\mathbb{E}[|E(G_n)|] = \frac{\lambda}{2} (n-1),$$

(9.4.38)

so that the average degree per vertex is precisely equal to $\lambda$. The constraint in (9.4.35) thus reduces to

$$\left(\frac{n}{2}\right) \frac{e^\beta}{1 + e^\beta} = \frac{\lambda}{2} (n-1).$$

(9.4.39)

This leads to ER$_n(\lambda/n)$, where

$$\frac{e^\beta}{1 + e^\beta} = \frac{\lambda}{n},$$

(9.4.40)

that is, $e^\beta = \lambda/(n - \lambda)$. This shows that the ER$_n(\lambda/n)$ is an example of an exponential random graph with a constraint on the expected number of edges in the graph. Further, by the conditioning property of exponential random graphs, conditionally on ER$_n(\lambda/n) = m$, the distribution is uniform over all graphs with $m$ edges.

Example 9.21 (Example: GRG$_n(w)$) The second example arises when we fix on the expected degrees of all the vertices. This arises when we take $N_v(x) = \sum_{j \in [n]} x_{vj} = d_v^{(G_n)}$ for every $v \in [n]$, so that we put a restriction on the expected degree of all the vertices in the graph. In this case, we see that, with $G_n = (x_{i,j})_{1 \leq i < j \leq n}$,

$$Z_n(\vec{\beta}) = \sum_x e^{\sum_{v \in [n]} \beta_v d_v^{(G_n)}} = \sum_x e^{\sum_{i,j \in [n]} (\beta_i + \beta_j) x_{i,j}} = \prod_{1 \leq i < j \leq n} (1 + e^{\beta_i + \beta_j}),$$

(9.4.41)

and

$$p_{\vec{\beta}}(x) = \frac{1}{Z_n} e^{\sum_{v \in [n]} \beta_v d_v^{(G_n)}} = \prod_{1 \leq i < j \leq n} \frac{e^{(\beta_i + \beta_j) x_{i,j}}}{1 + e^{\beta_i + \beta_j}}.$$  

(9.4.42)
Thus, the different edges are independent, and edge \( \{i, j\} \) is present with probability \( e^{\beta_i + \beta_j} / (1 + e^{\beta_i + \beta_j}) \), and absent with probability \( 1 / (1 + e^{\beta_i + \beta_j}) \). In the sparse setting, we aim that

\[
\mathbb{E}[d_v^{(G_n)}] = \alpha_v,
\]

so that the average degree of vertex \( v \) is precisely equal to \( \alpha_v \). The constraint in (9.4.35) thus reduces to

\[
\sum_{j \neq v} e^{\beta_i + \beta_j} / (1 + e^{\beta_i + \beta_j}) = \alpha_v.
\]

This leads to \( \text{GRG}_n(w) \), where

\[
\frac{w_v}{\sqrt{\sum_{u \in [n]} w_u}} = e^{\beta_v},
\]

that is, \( \beta = \lambda / (n - \lambda) \). This shows that the \( \text{GRG}_n(w) \) is an example of an exponential random graph with a constraint on the expected number of edges in the graph. Further, by the conditioning property of exponential random graphs, conditionally on \( d_v^{(G_n)} = \alpha_v \) for all \( v \in [n] \), the distribution is uniform over all graphs with these degrees. This gives an alternative proof of Theorem 1.4.

Now that we have discussed two quite nice examples of exponential random graphs, let us also discuss its intricacies. The above choices, in Examples 9.20 and 9.21, and quite special in the sense that the exponent in (9.4.33) is linear in the edge occupation statuses \( (x_{i,j})_{1 \leq i < j \leq n} \). This gives rise to exponential random graphs that have independent edges. However, when investigating more intricate subgraph counts, such as triangles, this linearity is no longer true. Indeed, the number of triangles is a cubic function of \( (x_{i,j})_{1 \leq i < j \leq n} \). In such cases, the edges will no longer be independent, making the exponential random graph very hard to study.

Indeed, the exponential form in (9.4.33) naturally leads to large deviations of random graphs, a topic that is much better understood in the dense setting where the number of edges grown proportionally to \( n^2 \). In the sparse setting, such problems are hard, and sometimes ill-defined. We refer to the notes and discussion in Section 9.6 for more background and references.

9.5 Spatial random graphs

The models described so far do not incorporate geometry at all. Yet, geometry may be relevant. Indeed, in many networks, the vertices are located somewhere in space, and their locations may be relevant. People who live closer to one another are more likely to know each other, even though we all know people who live far away from us. This is a very direct link to the geometric properties of networks. However, the geometry may also be much more indirect or latent. This could arise, for example, since people who have similar interests are also more likely to
know one another. This, when associating a whole bunch of attributes to vertices in the network, vertices with more similar attributes (age, interests, profession, music preference, etc.) may be more likely to know each other. In any case, we are rather directly led to studying networks where the vertices are embedded in some general geometric space. This is what we refer to as spatial networks.

One further aspect of spatial random graphs deserves to be mentioned. Due to the fact that nearby vertices are more likely to be neighbors, it is also true that two neighbors of a vertex are more likely to be connected. Therefore, geometry leads to clustering, and it may be one of the most natural ways to do so.

9.5.1 Small-world model

The small-world model was the first spatial model to be proposed by Watts and Strogatz (1998). We again refer to the notes and discussion in Section 9.6 for more background and references, including the history of the model. The aim was to describe how the small-world effect can arise in a simple and natural way through the addition of long-range edges. Here, long-range refers to the underlying geometry, where the vertices are located in some geometric space with a natural intrinsic distance on it. Long-range edges refers to pair of vertices that are far away in the geometry, yet are neighbors in the network. Such long-range edges can lead to substantial shortcuts, and thus significantly decrease graph distances.

Let us describe a first version of the small-world network.

We start with a finite torus, and add random long-range connections to them, independently for each pair of vertices. This gives rise to a graph that is a small perturbation of the original lattice, but has occasional long-range connections that are crucial in order to shrink graph distances. From a practical point of view, we can think of the original graph as being the local description of acquaintances in a social network, while the shortcuts describe the occasional far apart acquaintances. The main idea is that, even though the shortcuts only form a small part of the connections in the graph, they are crucial in order to make it a small world.

Small-world behavior in the continuous circle model

The simplest version of the model studied by Barbour and Reinert (2001) is obtained by taking the circle of circumference $n$, and adding a Poisson number of shortcuts with parameter $n\rho/2$, where the starting and endpoints of the shortcuts are chosen uniformly at random independently of each other. This model is called the continuous circle model. Distance is measured as usual along the circle, and the shortcuts have, by convention, length zero. Thus, one can think of this model as the circle where the points along the random shortcut are identified, thus creating a puncture in the circle. Multiple shortcuts then lead to multiple punctures of the circle, and the distance is then the usual distance along the punctured graph. Bear in mind that this is something different from the usual graph distances, which count the number of edges along the shortest path between pairs of vertices. The following result describes this punctured graph distance:
**Theorem 9.22** (Distance in continuous circle model)  

Let $D_n$ the distance between two uniformly chosen points along the punctured circle in the continuous circle model. Then, for every $\rho > 0$, as $n \to \infty$,  

$$D_n(2\rho)/\log(\rho n) \xrightarrow{p} 1.$$  

(9.5.1)

More precisely,  

$$\rho(D_n - \log(\rho n)/2) \xrightarrow{d} T,$$  

(9.5.2)

where $T$ is a random variable satisfying  

$$P(T > t) = \int_0^\infty \frac{e^{-y}dy}{1 + e^{2y}}.$$  

(9.5.3)

The random variable $T$ can also be described by  

$$P(T > t) = E[e^{-2W^{(1)}W^{(2)}}],$$  

(9.5.4)

where $W^{(1)}, W^{(2)}$ are two independent exponential random variables with parameter 1. Alternatively, it can be see that $T = (G_1 + G_2 - G_3)/2$, where $G_1, G_2, G_3$ are three independent Gumbel distributions (see (Barbour and Reinert, 2006, Page 1242)).

Interestingly, the method of proof of Theorem 9.22 (see Barbour and Reinert (2001)) is quite close to the method of proof for Theorem 7.20. Indeed, again the parts of the graph that can be reached in distance at most $t$ are analyzed. Let $P_1$ and $P_2$ be two uniform points along the circle, so that $D_n$ has the same distribution as the distance between $P_1$ and $P_2$. Denote by $R^{(1)}(t)$ and $R^{(2)}(t)$ the parts of the graph that can be reached within distance $t$. Then, $D_n = 2T_n$, where $T_n$ is the first time that $R^{(1)}(t)$ and $R^{(2)}(t)$ have a non-zero intersection. The proof then consists of showing that, up to time $T_n$, the processes $R^{(1)}(t)$ and $R^{(2)}(t)$ are close to certain continuous-time branching processes, primarily due to the fact that the probability that there are two intervals that are overlapping is quite small. Then, $W^{(1)}$ and $W^{(2)}$ can be viewed as appropriate martingale limits of these branching processes.

Compared to Theorem 7.20, we see that now the rescaled distance $D_n$, after subtraction of the right multiple of $\log n$, converges in distribution, while in Theorem 7.20, convergence is at best along subsequences. This is due to the fact that $D_n$ is a continuous random variable, while the graph distance is integer-valued. Therefore, the latter suffers from discretization effects. In the next paragraph, we will see that the graph distances in the small-world model suffer similar issues.

**Small-world behavior in the small-world model**

We now extend the analysis to the graph distances in the discrete model, which is highly similar to the original small-world model. In the discrete model, the vertices are located on a discrete torus containing $n$ vertices, and an extra Poisson($n\rho n/2$) edges is added uniformly at random. Again, these edges identify the points at the two sides of the edge, rather than being a proper edge between them. Now
9.5 Spatial random graphs

dist_{G_n}(o_1, o_2) denotes the graph distances in this (connected) graph between two vertices o_1, o_2 chosen uniformly at random from [n]. The following theorem describes the asymptotics of this random variable:

**Theorem 9.23 (Distance in small-world model)** Let G_n be the discrete small-world model, where vertices are connected to their immediate neighbors along the circle of size n, and a Poisson number with parameter \( \rho n/2 \) edges are added uniformly at random. Let dist_{G_n}(o_1, o_2) denote the graph distance between two uniformly chosen vertices o_1, o_2 \( \in [n] \). Then,

\[
dist_{G_n}(o_1, o_2)/\log n \xrightarrow{\mathcal{L}} \frac{1}{\log (1 + 2\rho)}.
\]

(9.5.5)

More precisely,

\[
P\left( \text{dist}_{G_n}(o_1, o_2) - \left\lfloor \frac{\log n}{\log (1 + 2\rho)} \right\rfloor + x \right) = E\left[ e^{-\phi_0(1+2\rho)W_1W_2} \right],
\]

(9.5.6)

where W_1, W_2 are two i.i.d. copies of an appropriate martingale limit, for a certain \( \phi_0 = \phi_0(n, \rho) \).

It would be of interest to extend this to the small-world model where the long-range edges are also being counted as having distance 1, rather than distance zero.

**Explain the 1 + 2\rho!**

9.5.2 Hyperbolic random graphs

Here we consider the hyperbolic random graph where vertices are in a disk of radius R, and connected if their hyperbolic distance is at most R Krioukov et al. (2010). These graphs are very different from general inhomogeneous random graphs, because their geometry creates more clustered random graphs. The hyperbolic random graph has two key parameters, \( \nu \) and \( \alpha \). The model samples \( n \) vertices on a disk of radius \( R = 2\log(n/\nu) \), where the density of the radial coordinate \( r \) of a vertex \( p = (r, \phi) \) is

\[
\rho(r) = \alpha \frac{\sinh(\alpha r)}{\cosh(\alpha R) - 1}.
\]

(9.5.7)

Here \( \nu \) parametrizes the average degree of the generated networks and \( -\alpha \) the so-called negative curvature of the space. The angle \( \phi \) of \( p \) is sampled uniformly from \([0, 2\pi]\), so that the points have a spherically symmetric distribution. Then, two vertices are connected when their hyperbolic distance is at most R. Here, the hyperbolic distance \( x = d_{\mathbb{H}}(u, v) \) between two points at polar coordinates \( u = (r, \phi) \) and \( v = (r', \phi') \) is given by the hyperbolic law of cosines

\[
cosh(x) = \cosh(r) \cosh(r') - \sinh(r) \sinh(r') \cos(\|\phi - \phi'\|),
\]

(9.5.8)

where \( \|\phi - \phi'\| = \pi - |\phi - \phi'| \) is the difference between the two angles (which is the Euclidean distance on the circle).
For a point \( i \) with radial coordinate \( r_i \), we define its type \( t_i \) as \[ t_i = e^{(R-r_i)/2}. \] (9.5.9)

Then, the degree of vertex \( i \) can be approximated by a Poisson random variable with mean \( t_i \). Thus, with high probability,

\[ D_i = \Theta(t_i), \] (9.5.10)

where \( D_i \) denotes the degree of vertex \( i \). Furthermore, the random variables \( (t_i)_{i \geq 1} \) are distributed as a power-law with exponent \( \tau \), so that the degrees have a power-law distribution as well. Let us now explain the degree structure in more detail.

### Degree structure of hyperbolic random graphs

Let \[ P_{k}^{(n)} = \frac{1}{n} \sum_{v \in [n]} \mathbf{1}_{\{D_v = k\}} \] (9.5.11)
denote the degree distribution in the hyperbolic random graph. As explained informally in the previous paragraph, we may expect that the degree distribution obeys a power law. This is the content of the following theorem:

**Theorem 9.24** (Power-law degrees in hyperbolic random graphs) As \( n \to \infty \), there exists a probability distribution \((p_k)_{k \geq 0}\) such that

\[ P_{k}^{(n)} \xrightarrow{p} p_k, \] (9.5.12)

where \((p_k)_{k \geq 0}\) obeys an asymptotic power law, i.e., there exists a \( c > 0 \) such that

\[ p_k = ck^{-\tau}(1 + o(1)), \] (9.5.13)

where \( \tau = 2\alpha + 1 \).

We deduce that the model is scale-free, meaning that the asymptotic degree distribution has infinite variance, precisely when \( \alpha \in \left(\frac{1}{2}, 1\right) \), and otherwise the degree distribution obeys a power law with a larger degree exponent.

The exact form of \( p_k \) is identified by Fountoulakis et al. (2020), and involves several special functions. This result is quite impressive, and the proof is quite involved. For the purpose of this book, the exact shape of \( p_k \) is not so relevant.

Much more is known about the local structure of the hyperbolic random graph, for example, its local weak limit has been identified. We postpone this discussion to the next section, in which we discuss the local weak limit in geometric inhomogeneous random graphs. It turns out that we can interpret the hyperbolic random graph as a special case, which is very interesting in its

### The giant in hyperbolic random graphs

We next study the giant in hyperbolic random graphs. The main result, which is somewhat surprising, is as follows:
Theorem 9.25 (Giant in hyperbolic random graphs) Let $C_{\text{max}}$ and $C_{(2)}$ be the maximal and second largest cluster in the hyperbolic random graph with parameters $\alpha > \frac{1}{2}$ and $\nu > 0$. Then, the largest connected components satisfy the following properties:

(a) For $\alpha > 1$, $|C_{\text{max}}|/n \xrightarrow{\mathbb{P}} 0$ for all $\nu > 0$.

(b) For $\alpha \in \left(\frac{1}{2}, 1\right)$, there exists a $\zeta$ such that $|C_{\text{max}}|/n \geq \zeta > 0$ whp, and $|C_{(2)}|/n \xrightarrow{\mathbb{P}} 0$ for all $\nu > 0$.

(c) For $\alpha = 1$, there exist $\pi/8 \leq \nu_0 \leq \nu_1 \leq 20\pi$ such that $|C_{\text{max}}| \geq n/610$ and $|C_{(2)}|/n \xrightarrow{\mathbb{P}} 0$ for all $\nu > \nu_1$, while $|C_{\text{max}}|/n \xrightarrow{\mathbb{P}} 0$ for $\nu \leq \nu_0$.

We see that a giant component only exists in the scale-free regime, which is quite surprising. In various other random graphs, also a giant component exists in settings where the degrees have finite variance, particularly when there are sufficiently many edges. This turns out not to be the case for hyperbolic random graphs, and we next explain the cause of this surprising result. The fact that no giant exists when $\tau > 3$ is explained in more detail in Section 9.5.3, see below Theorem 9.32. There, it is explained that the hyperbolic random graph is a special case of a one-dimensional geometric inhomogeneous random graph. Giants in one dimension only exist when the degrees have infinite variance. We postpone a further discussion to that section.

More precise results are known in some of these cases. For $\alpha > 1$, Bode et al. (2015) showed that $|C_{\text{max}}| = \Theta_\nu(R^2(\log \log R)^3n^{1/\alpha})$, while for $\alpha \in \left(\frac{1}{2}, 1\right)$, Kiwi and Mitsche (2019) proved that the second largest component is at most poly-logarithmic (i.e., at most a power of $\log n$).

Ultra-small distances in hyperbolic random graphs

Obviously, there is little point in studying typical distances in the hyperbolic random graph when there is no giant component, even though such results do shed light on the component structure of smaller components. Thus, we will not restrict to the setting where $\alpha \in \left(\frac{1}{2}, 1\right)$, where typical distances are ultra-small:

Theorem 9.26 (Ultra-small distances in hyperbolic random graphs) For $\alpha \in \left(\frac{1}{2}, 1\right)$, conditionally on $o_1, o_2$ being connected, with $G_n$ the hyperbolic random graph with parameters $\alpha > \frac{1}{2}$ and $\nu > 0$,

$$
\frac{d_{G_n}(o_1, o_2)}{\log \log n} \xrightarrow{\mathbb{P}} \frac{2}{|\log(2\alpha - 1)|} \quad \text{ (9.5.14)}
$$

Note that $2\alpha - 1 = \tau - 2$, so that Theorem 9.26 is very similar to the ultra-small nature of the random graphs studied in this book.

Clustering in hyperbolic random graphs

We now study $c(k)$ for the hyperbolic random graph. The main result is as follows:
Figure 9.10 A hyperbolic embedding of the Internet at Autonomous Systems level by Boguná et al. (2010) (see (Boguná et al., 2010, Figure 3))

Theorem 9.27 (Clustering spectrum of hyperbolic random graphs) The local clustering coefficient in the hyperbolic random graph satisfies with high probability

\[ c(k) \propto \begin{cases} 
  k^{-1} & \tau > \frac{5}{2}, k \gg 1 \\
  k^{4-2\tau} & \tau < \frac{5}{2}, 1 \ll k \ll \sqrt{n} \\
  k^{2\tau-6}n^{5-2\tau} & \tau < \frac{5}{2}, k \gg \sqrt{n}.
\end{cases} \tag{9.5.15} \]

In the hyperbolic random graph, typical triangles that contribute most to \( c(k) \) are given by vertices of very typical degrees, as we next explain. The typical triangle for \( \tau > 5/2 \) is a triangle where one vertex has degree \( k \), and the other two have constant degree. When \( \tau < 5/2 \) and \( k < \sqrt{n} \), the typical triangle has three vertices of degree \( k \). When \( k > \sqrt{n} \), a typical triangle has one vertex of degree \( k \) and two of degree \( n/k \).
Hyperbolic embeddings of complex networks

In recent years, the question whether hyperbolic geometries can be used to efficiently map real-world complex networks has attracted considerable attention. Indeed, hyperbolic random graphs have very small distances on the one hand, while on the other hand also have the necessary clustering to make them appropriate for many complex networks. Of course, the question of how to embed them precisely is highly relevant and also quite difficult.

The example that has attracted the most attention is the Internet. In Figure 9.10, you can see a hyperbolic embedding of the Internet, as performed by Boguná et al. (2010). We see that the regions on the boundary of the outer circle can be grouped in a fairly natural way, where the countries in which the autonomous systems reside seem to be grouped according to their geography, with some exceptions (for example, it is not so clear why Kenya is almost next to the Netherlands). This hyperbolic geometry is first of all quite interesting, but also be helpful in sustaining the ever growing Internet traffic (see Boguná et al. (2010) and the references therein).

9.5.3 Geometric inhomogeneous random graphs

We start by defining the geometric inhomogeneous random graph. Here vertices are in a general metric space \( X \subseteq \mathbb{R}^d \) for some dimension \( d \geq 1 \). We let \((X_i)_{i \in [n]}\) denote their locations, and assume that the \((X_i)_{i \in [n]}\) are chosen in an i.i.d. way from some measure \( \mu \) on \( X \). Often, we will assume that \( X = [0,1]^d \) is the cube of width 1, with periodic boundary conditions. Further, we assume that each vertex \( i \) has a weight \( W_i \) associated to it, where \((W_i)_{i \in [n]}\) are assumed to be i.i.d. Often, we assume that \( W_i \) are power-law random variables, even though this is not necessary for the definition.

The edges are conditionally independent given \((x_i)_{i \in [n]}\) and \((w_i)_{i \in [n]}\), where the conditional probability that the edge between \( u \) and \( v \) is present is equal to

\[
p_{u,v} = p(x_u, x_v, w_u, w_v, (w_s)_{s \in [n] \setminus \{u,v\}}) = \Theta \left( 1 \wedge \left( \frac{w_u w_v}{\sum_{i \in [n]} w_i} \right)^{\max(\alpha,1)} \|x_u - x_v\|^{-d \alpha} \right),
\]

where \( \alpha > 0 \) is an appropriate parameter. We will assume that the vertex weights obey a power law, i.e.,

\[
P(W > w) = L(w)w^{-(\tau - 1)}
\]

for some slowly varying function \( L: [0, \infty) \to (0, \infty) \). As is usual, the literature treats a variety of models and settings, and we refer to the notes and discussion in Section 9.6 for more details. To avoid trivialities, we assume that there exists a \( \varepsilon > 0 \) such that

\[
P(W < \varepsilon) = 1.
\]
Assumption 9.28 (Limiting connection probabilities exist) Set $X[-\frac{1}{2}, \frac{1}{2}]^d$ with $d \geq 1$. Assume the following:

(a) There exists a measurable event $E_n$ such that $\mathbb{P}(E_n) \to 1$;
(b) There exist intervals $I_\Delta(n) \subseteq \mathbb{R}^+$, $I_W(n) \subseteq [1, \infty)$ and a sequence $\varepsilon(n)$ such that $I_\Delta(n) \to \mathbb{R}^+$, $I_W(n) \to [1, \infty)$ and $\varepsilon(n) = o(1)$ as $n \to \infty$;
(c) There exists $h : \mathbb{R}^d \times \mathbb{R} \times \mathbb{R}$ such that, on the event $E_n$, for almost every $x \in [-\frac{1}{2}, \frac{1}{2}]^d$, all $u, v \in [n]$

\[
\begin{align*}
    \frac{\left| p(x, x + \Delta/n^{1/d}, w_u, w_v, (W_s)_{s \in [n] \setminus \{u, v\}}) - h(\Delta, w_u, w_v) \right|}{h(\Delta, w_u, w_v)} & \leq \varepsilon(n). \tag{9.5.19}
\end{align*}
\]

Having introduced the model, we will first explain how the hyperbolic random graph can be cast into this format, followed by a discussion of the properties of the geometric inhomogeneous random graph (GIRG).

**Hyperbolic random graph as GIRG**

We now explain that the hyperbolic random graph can be interpreted as a special case of the geometric inhomogeneous random graph. To this end, we embed the disk of the native hyperbolic model into our model with dimension 1, hence we reduce the geometry of the hyperbolic disk to the geometry of a circle, but gain additional freedom as we can choose the weights of vertices. Notice that a single point on the hyperbolic disk has measure zero, so we can assume that no vertex has radius $r_v = 0$. Recall that $d_H(u, v)$ denotes the hyperbolic distance defined in (9.5.8). Here, we also explain this relation for the soft hyperbolic graph, for which the probability that there exists an edge between the vertices $u = (r_u, \phi_u)$ and $v = (r_v, \phi_v)$ equals

\[
p_{H}(d_H(u, v)) = \left(1 + e^{(d_H(u, v) - R_n)/(2T)}\right)^{-1}. \tag{9.5.20}
\]

In the limit $T \searrow 0$, this gives rise to the (hard) hyperbolic graph as studied in the previous section. The identification is given in the following theorem:

**Theorem 9.29 (Hyperbolic random graphs as GIRGs)** The soft hyperbolic random graph is a geometric inhomogeneous random graph that satisfies Assumption 9.28 with parameters

\[
    d := 1, \quad \tau := 2\alpha + 1, \quad \alpha := 1/T. \tag{9.5.21}
\]

and limiting connection probabilities given by

\[
h(\Delta, w_u, w_v) = \frac{1}{1 + \left(\frac{\Delta}{w_u w_v}\right)^c}, \quad \tag{9.5.22}
\]

where $c = \sqrt{2}\pi/\nu$.

The hard hyperbolic random graph can be mapped to a threshold GIRG, where vertices $u, v \in [n]$ are connected with probability one when $\|x_u - x_v\| \leq \nu w_u w_v/\pi$. 

Sketch of proof of Theorem 9.29. Here we give a sketch of the proof. We define the mapping

\[ w_v := e^{(R-r_v)/2} \quad \text{and} \quad x_v := \phi_v/2\pi, \]  

(9.5.23)

where, for \( v \in [n] \), its radial coordinate equals \( r_v \) and its angular coordinate \( \phi_v \). The map \((r, \phi) \mapsto (e^{(R-r)/2}, \phi/2\pi)\) is a bijection from the hyperbolic space to \([0, e^{R/2}] \times \mathbb{T}_{1,1}\), where \( \mathbb{T}_{1,1} = [-\pi, \pi] \) with periodic boundary conditions. Therefore, it also has an inverse, which we will denote by \( g(w_v, x_v) = (r_v, \phi_v) \). Then, for \( i, j \in [n] \), we let

\[ p_{i,j} = p_{i,j}((w_i, x_i), (w_j, x_j)) = p_n(d_u(g(w_i, x_i), g(w_j, x_j))). \]

(9.5.24)

This maps the soft hyperbolic random graph to an explicit geometric inhomogeneous random graph.

We next verify the conditions on the edge probabilities that are required for geometric inhomogeneous random graphs. We start by computing from (9.5.7) that

\[ \mathbb{P}(r_v \leq r) = \int_0^r \rho(r) dr = \int_0^r \frac{\sinh(\alpha r)}{\cosh(\alpha R) - 1} dr = \frac{\cosh(\alpha r) - 1}{\cosh(\alpha R) - 1}. \]

(9.5.25)

Since \( R_n = 2 \log(n/\nu) \) is large,

\[ \mathbb{P}(r_v \leq r) = \frac{e^{\alpha(r-R)} + e^{-\alpha(R+r)}}{1 + e^{-2R}} = e^{\alpha(r-R)(1 + o(1))}. \]

(9.5.26)

This has the following consequence for the distribution of \( w_v \), since

\[ \mathbb{P}(w_v > w) = \mathbb{P}(r_v < R - 2 \log w) = e^{-2\alpha(1 + o(1))} = w^{-(r-1)(1 + o(1))}, \]

(9.5.27)

as required by (9.5.17). In particular, this implies that most vertices in \([n]\) will have \( r_v \in [R, R - A] \) for \( A \) large. Thus, most vertices will be close to the boundary of the hyperbolic disc.

We next investigate the role of the radial coordinate. For this, we assume that \( r_v \in [0, R] \), which is true almost surely, and \( r_u + r_v \geq R \), which occurs with high probability for \( R \) large. We use \( \cosh(x \pm y) = \cosh(x) \cosh(y) \pm \sinh(x) \sinh(y) \) to rewrite (9.5.8) for \( u = (r_u, \phi_u) \) and \( v = (r_v, \phi_v) \) as

\[ \cosh(d_u(u, v)) = \cosh(r_u - r_v) + \sinh(r_u) \sinh(r_v) [1 - \cos(\|\phi_u - \phi_v\|)]. \]

(9.5.28)

When \( \|\phi_u - \phi_v\| \) is small, as in Assumption 9.28, we thus obtain that

\[ \cosh(d_u(u, v)) = \cosh(r_u - r_v) + \sinh(r_u) \sinh(r_v) \|\phi_u - \phi_v\|^2 / 2(1 + o(1)). \]

(9.5.29)

Denote \( s_n = d_n(u, v) - R_n \), and note that

\[ e^{s_n} = 2 \cosh(d_u(u, v)) e^{-R_n} - e^{-s_n - 2R_n}. \]

(9.5.30)
Further, multiplying (9.5.29) by $e^{-R_n}$ leads to, with $R_n = 2 \log(n/\nu)$,
\[
\cosh(d_u(u,v))e^{-R_n} = e^{\nu u - R_n} e^{-R_n} e^{\| \phi_u - \phi_v \|^2} 2/(1 + o(1)) + \cosh(r_u - r_v)e^{-R_n}
\]
\[
= \frac{2\pi^2 n^2 (x_u - x_v)^2}{\nu^2 (w_u w_v)^2} (1 + o(1)) + \nu^2 (w_u/w_v)^2 + (w_u/w_v)^2.
\]
Combining (9.5.30) and (9.5.31), and substituting this into (9.5.20) for $x_v = x + \Delta/n$, leads to
\[
p_H(d_H(u,v)) = \frac{1}{1 + \left( \frac{2\pi^2 \nu^2 \Delta^2}{|w_u w_v|^2} \right)^{1/(2\tau)}} (1 + o(1)) \quad (9.5.32)
\]
where $c = \sqrt{2\pi}/\nu$.

Since hyperbolic random graphs are special cases of the GIRG, we will see that some results can simply be deduced from their analysis for the GIRG. Also, the fact that hyperbolic graphs arise as a one-dimensional GIRG helps us to explain why there is no giant for $\alpha > 1$, for which the degree power-law exponent satisfies $\tau > 3$.

**Degree structure of geometric inhomogeneous random graphs**

We start by analysing the degree structure of the graph. Let
\[
P_k^{(n)} = \frac{1}{n} \sum_{v \in [n]} 1_{\{D_v = k\}}
\]
denote the degree distribution in the hyperbolic random graph. As explained informally in the previous paragraph, we may expect that the degree distribution obeys a power law. This is the content of the following theorem:

**Theorem 9.30** (Power-law degrees in geometric inhomogeneous random graphs)

Let the edge probabilities in the geometric inhomogeneous random graph by given by
\[
p_{u,v} = \lambda \left( 1 \wedge \left( \frac{W_u W_v}{n} \right)^{\max\{\alpha,1\}} \| X_u - X_v \|^{-d\alpha} \right),
\]
where $(W_u)_{u \in [n]}$ are i.i.d. random variables satisfying $P(W > w) = c_w w^{-(\tau - 1)}(1 + o(1))$ for some $c_w > 0$ and $w \to \infty$, while $(X_u)_{u \in [n]}$ are i.i.d. uniform random variables on $[-\frac{1}{2}, \frac{1}{2}]^d$.

As $n \to \infty$, there exists a probability distribution $(p_k)_{k \geq 0}$ such that
\[
P_k^{(n)} \Rightarrow p_k,
\]
where $(p_k)_{k \geq 0}$ obeys an asymptotic power law, i.e., there exists a $c > 0$ such that
\[
p_k = ck^{-\tau}(1 + o(1)).
\]
This theorem has not been proved in the current form anywhere in the literature, so we prove it explicitly here:

**Proof of Theorem 9.30.** We use a second moment method, and start by analysing $\mathbb{E}[P_k^{(n)}]$. Note that

\[ \mathbb{E}[P_k^{(n)}] = \mathbb{P}(D_{o_n} = k), \quad (9.5.37) \]

where $o_n \in [n]$ is chosen uniformly at random. Further, note that

\[ D_v = \sum_{u \in [n]} I_{u,v}, \quad (9.5.38) \]

where

\[ \mathbb{P}(I_{u,v} = 1 \mid (x_i)_{i \in [n]}, (w_i)_{i \in [n]}) = p(x_u, x_v, w_u, w_v, (w_s)_{s \in [n] \setminus \{u,v\}}). \quad (9.5.39) \]

\[ \square \]

**Local limit of geometric inhomogeneous random graphs**

We next study the local limit of geometric inhomogeneous random graphs. Before stating the result, we introduce the local limit. We fix a Poisson process with intensity 1 on $\mathbb{R}^d$, with an additional point on the origin that will serve as our point of reference. To each point $x \in \mathbb{R}^d$ in this process, we associate an independent copy of the limiting weight $W_x$. Then, we draw an edge between two vertices $x, y$ with probability $h(x, y, W_x, W_y)$, where these variables are conditionally independent given the points of the Poisson process and the weights of its points. Call the above process the **Poisson infinite GIRG with edge-probabilities given by $h$**. When $h \in \{0, 1\}$, we are dealing with a threshold Poisson infinite GIRG with edge statuses given by $h$.

The main result is as follows:

**Theorem 9.31** (Local limit of geometric inhomogeneous random graphs) Consider a geometric inhomogeneous random graph under Assumption 9.28. Then, it converges locally weakly to the Poisson infinite GIRG with edge-probabilities given by $h$. The same result is true for a threshold geometric inhomogeneous random graph.

Together with Theorem 9.29, Theorem 9.31 also identifies the local weak limit of both the soft as well as the hard hyperbolic random graph.

**The giant in geometric inhomogeneous random graphs**

We next study the giant in hyperbolic random graphs. The main result, which is somewhat surprising, is as follows:

**Theorem 9.32** (Giant in geometric inhomogeneous random graphs) Let $C_{\max}$
and $C_{(2)}$ be the maximal and second largest cluster in the geometric inhomogeneous random graph with parameters $\tau \in (2,3)$. Assume that the edge probabilities satisfy (9.5.16). Then, the largest connected components satisfy that $|C_{\text{max}}|/n \geq \zeta > 0$, while $|C_{(2)}|/n \rightarrow 0$ for all $\alpha > 0$.

Thus, a giant component always exists in the scale-free regime. A general result for $\tau > 3$ has not been proved (yet?), and is much more delicate. We will return to this in Section 9.5.5, where we investigate scale-free percolation, which lives on $\mathbb{Z}^d$. There, the existence of an infinite component containing a positive proportion of the vertices should correspond to the existence of a giant in our finite random graphs.

Let us complete this part by discussing the situation when $d = 1$. Here, we can show that when $\tau > 3$, there is no infinite component in the local limit, which suggests that also no giant exists in the pre-limit. Recall Corollary 2.25, but beware that we do not know that the local limit exists in probability. We call a vertex $u$ an isolated vertex when there is no edges $\{v_1, v_2\}$ with $v_1 \leq u \leq v_2$ present in the graph. In particular, this means that the connected component of $u$ is finite. We will prove that the expected number of edges $\{v_1, v_2\}$ with $v_1 \leq u \leq v_2$ that are present is bounded. Indeed, for the local limit of the hard hyperbolic graph, this number is equal to

$$\sum_{v_1 \leq u \leq v_2} \mathbb{E}\left[\frac{1}{{1 + \left(\|v_1 - v_2\|^{\tau} / cW_{v_1}W_{v_2}\right)^{\alpha}}}\right].$$

(9.5.40)

We bound this by

$$C \sum_{v_1 \leq u \leq v_2} \mathbb{E}\left[\frac{(W_{v_1}W_{v_2})^{\alpha}}{1 + \|v_1 - v_2\|^\alpha} \wedge 1\right] \leq C \sum_{k \geq 1} k \mathbb{E}\left[\frac{X^{\alpha}}{1 + k^{\alpha}} \wedge 1\right],$$

(9.5.41)

where $X = W_1W_2$ is the product of two independent $W$ variables. Now, when $\mathbb{P}(W > w) = w^{-(\tau - 1)}$, it is not hard to see that

$$\mathbb{P}(X^{\alpha} > x) \leq C \frac{\log x}{x^{-(\tau - 1)/\alpha}}.$$  

(9.5.42)

In turn, this implies that

$$\mathbb{E}\left[\frac{X^{\alpha}}{1 + k^{\alpha}} \wedge 1\right] \leq Ck^{-(\tau - 1)},$$

(9.5.43)

which is summable and thus the expected number of edges $\{v_1, v_2\}$ with $v_1 \leq u \leq v_2$ that are present is bounded. In turn, this suggests that the number equals zero with strictly positive probability (beware, these numbers are not independent for different $u$), and this in turn suggests that there is a positive proportion of them. However, when there is a positive proportion of them, then there cannot be an infinite component. This intuitively explains why the existence of the giant component is restricted to $\tau \in (2,3)$. Thus, the absence of a giant in hyperbolic graphs with power-law exponent $\tau$ with $\tau > 3$ is intimately related to this model being inherently one-dimensional.
Ultra-small distances in geometric inhomogeneous random graphs

Obviously, there is little point in studying typical distances in the hyperbolic random graph when there is no giant component, even though such results do shed light on the component structure of smaller components. Thus, we will not restrict to the setting where \( \alpha \in (\frac{1}{2}, 1) \), where typical distances are ultra-small:

**Theorem 9.33** (Ultra-small distances in geometric inhomogeneous random graphs) For \( \tau \in (2, 3) \), conditionally on \( o_1, o_2 \) being connected, with \( G_n \) the generalized inhomogeneous random graph as in

\[
\frac{d_{G_n}(o_1, o_2)}{\log \log n} \xrightarrow{\tau} \frac{2}{|\log (\tau - 2)|}.
\]

**Clustering in geometric inhomogeneous random graphs**

We now study the clustering properties of geometric inhomogeneous random graph. The main result is as follows:

**Theorem 9.34** (Clustering in geometric inhomogenous random graphs) The global clustering coefficient in the geometric inhomogeneous random graph is \( \Theta(1) \) whp.

### 9.5.4 Spatial preferential attachment models

In the past years, several spatial preferential attachment models have been considered. We shall now discuss three of such models.

**Geometric preferential attachment model.**

We next discuss a class of geometric preferential attachment models that combines aspects of random geometric graphs and preferential attachment graphs. Let \( G_t = (V_t, E_t) \) denote the graph at time \( t \). Let \( S \) be the sphere \( S \) in \( \mathbb{R}^3 \) with area equal to 1. Then, we let \( V_t \) be a subset of \( S \) of size \( t \).

The process \((G_t)_{t \geq 0}\) evolves as follows. At time \( t = 0 \), \( G_0 \) is the empty graph. At time \( t + 1 \), given \( G_t \), we obtain \( G_{t+1} \) as follows. Let \( x_{t+1} \) be chosen uniformly at random from \( S \), and denote \( V_{t+1} = V_t \cup \{x_{t+1}\} \). We assign \( m \) edges to the vertex \( x_{t+1} \), which we shall connect independently of each other to vertices in \( V_t(x_{t+1}) \equiv V_t \cap B_r(x_{t+1}) \), where \( B_r(u) = \{x \in S: ||x - u|| \leq r\} \) denotes the spherical cap of radius \( r \) around \( u \). Let

\[
D_t(x_{t+1}) = \sum_{v \in V_t(x_{t+1})} D_v^{(m)},
\]

where \( D_v^{(m)} \) denotes the degree of vertex \( v \in V_t \) in \( G_t \). The \( m \) edges are connected to vertices \((y_1, \ldots, y_m)\) conditionally independently given \((G_t, x_{t+1})\), so that, for all \( v \in V_t(x_{t+1}) \),

\[
P(y_t = v \mid G_t) = \frac{D_v^{(m)}}{\max(D_t(x_{t+1}), \alpha m A_r t)},
\]

where \( A_r \) is the area of the spherical cap of radius \( r \) around \( u \).
while
\[ P(y_i = x_{t+1} \mid G_t) = 1 - \frac{D_t(x_{t+1})}{\max(D_t(x_{t+1}), \alpha m A_t)}, \]  
(9.5.47)

where \( A_r \) is the area of \( B_r(u) \), \( \alpha \geq 0 \) is a parameter, and \( r \) is a radius which shall be chosen appropriately. Similarly to the situation of geometric random graphs, the parameter \( r \) shall depend on the size of the graph, i.e., we shall be interested in the properties of \( G_n \) when \( r = r_n \) is chosen appropriately. The main result of Flaxman et al. (2006) is the study of the degree sequence of the arising model. Take \( r_n = n^{\beta-1/2} \log n \), where \( \beta \in (0, 1/2) \) is a constant. Finally, let \( \alpha > 2 \). Then, there exists a probability distribution \( \{p_k\}_{k=0}^{\infty} \) such that, whp,
\[ P_k = p_k(1 + o(1)), \]  
(9.5.48)

where \( (p_k)_{k \geq m} \) satisfies (1.1.7) with \( \tau = 1 + \alpha \in (3, \infty) \). The precise result is in (Flaxman et al., 2006, Theorem 1(a)) and is quite a bit sharper, as detailed concentration results are proved as well. Further results involve the proof of connectivity of \( G_n \) and an upper bound on the diameter when \( r \geq n^{-1/2} \log n \), \( m \geq K \log n \) for some large enough \( K \) and \( \alpha \geq 0 \) of order \( O(\log (n/r)) \). In Flaxman et al. (2007), these results were generalized to the setting where, instead of a unit ball, a smoother version is used.

**Spatial preferential attachment as influence.**

We next consider a spatial preferential attachment model with local influence regions, as a model for the Web graph. The model is directed, but it can be easily adapted to an undirected setting. The idea behind this model is that for normal preferential attachment models, new vertices should be aware of the degrees of the already present vertices. In reality, it is quite hard to observe the degrees of vertices, and, therefore, we let vertices instead have a region of influence in some metric space, for example the torus \([0, 1]^m\) for some dimension \( m \), for which the metric equals
\[ d(x, y) = \min\{\|x - y + u\|_\infty : u \in \{0, 1, -1\}^m\}. \]  
(9.5.49)

When the new vertex arrives, it is uniformly located somewhere in the unit cube, and it connects to each of the older vertices in which region of influence they land independently and with fixed probability \( p \). These regions of influence evolve as time proceeds, in such a way that the volume of the influence region of the vertex \( i \) at time \( t \) is equal to
\[ R(i, t) = \frac{A_1 D_i^{(t)} + A_2}{t + A_3}, \]  
(9.5.50)

where now \( D_i^{(t)} \) is the in-degree of vertex \( i \) at time \( t \), and \( A_1, A_2, A_3 \) are parameters which are chosen such that \( p A_1 \leq 1 \). One of the main results of the paper is that this model is a scale-free graph process. Indeed, denote
\[ p_k = \frac{p^k}{1 + kpA_1 + pA_2 \prod_{j=0}^{k-1} (1 + A_2 + pA_2)}, \]  
(9.5.51)
then (Aiello et al., 2008, Theorem 1.1) shows that whp, for $k \leq (n^{1/8}/\log n)^{6pA_1/(2pA_1+1)}$, the degree sequence of the graph of size $n$ satisfies

$$F_k^{(n)} = p_k(1 + o(1)),$$

(9.5.52)

and $(p_k)_{k \geq 0}$ satisfies (1.1.7) with $\tau = 1 + 1/(pA_1) \in [2, \infty)$. Further results involve the study of maximal in-degrees and the total number of edges.

### 9.5.5 Complex network models on the hypercubic lattice

In this section, we define a percolation model that interpolates between long-range percolation and the scale-free rank-1 inhomogeneous random graphs as discussed in Chapter 3. This model, termed *scale-free percolation* in Deijfen et al. (2013), provides a percolation model in which the degree of a vertex can have finite mean but infinite variance. Mind that this phenomenon is impossible for independent percolation models, since the independence of the edge variables implies that the variance of the degrees is always bounded by their mean.

**Scale-free percolation**

Scale-free percolation is defined on the lattice $\mathbb{Z}^d$. Let each vertex $x \in \mathbb{Z}^d$ be equipped with an i.i.d. weight $W_x$. Conditionally on the weights $(W_x)_{x \in \mathbb{Z}^d}$, the edges in the graph are independent and the probability that there is an edge between $x$ and $y$ is defined by

$$p_{xy} = 1 - e^{-\lambda W_x W_y/|x-y|^\alpha},$$

(9.5.53)

for $\alpha, \lambda \in (0, \infty)$. We say that the edge $\{x, y\}$ is *occupied* with probability $p_{xy}$ and *vacant* otherwise.

Let us discuss the role of the different parameters in the scale-free percolation model. The parameter $\alpha > 0$ describes the *long-range nature* of the model, while we think of $\lambda > 0$ as the *percolation parameter*. The *weight distribution* is the last parameter that describes the model. We are mainly interested in settings where the $W_x$ have unbounded support in $[0, \infty)$, and then particularly when they vary substantially.

Naturally, the model for fixed $\lambda > 0$ and weights $(W_x)_{x \in \mathbb{Z}^d}$ is the same as the one for $\lambda = 1$ and weights $(\sqrt{\lambda}W_x)_{x \in \mathbb{Z}^d}$, so there is some redundancy in the parameters of the model. However, we view the weights $(W_x)_{x \in \mathbb{Z}^d}$ as creating a *random environment* in which we study the percolative properties of the model. Thus, we think of the random variables $(W_x)_{x \in \mathbb{Z}^d}$ as fixed once and for all and we change the percolation configuration by varying $\lambda$. We can thus view our model as percolation in a random environment given by the weights $(W_x)_{x \in \mathbb{Z}^d}$. The downside of scale-free percolation is that the edge statuses are no longer independent random variables, but are rather positively correlated (see Exercise 9.40).

Scale-free percolation interpolates between long-range percolation and rank-1 inhomogeneous random graphs. Indeed, we retrieve long-range percolation when we take $W_x \equiv 1$. We retrieve the Norros-Reittu model in (3.2.8) with i.i.d. edge
weights when we take \( \alpha = 0, \lambda = 1/\sum_{i \in [n]} W_i \) and consider the model on \([n]\) instead of \(\mathbb{Z}^d\). Thus, this model can be considered to be an interpolation between long-range percolation and rank-1 inhomogeneous random graphs.

**Choice of edge weights**

We assume that the distribution \( F_w \) of the weights \((W_x)_{x \in \mathbb{Z}^d}\) has a regularly-varying tail with exponent \(\tau - 1\), that is, denoting by \(W\) a random variable with the same distribution as \(W_0\) and by \(F_w\) its distribution function, we assume that

\[
1 - F_w(w) = \mathbb{P}(W > w) = w^{-(\tau-1)}L(w),
\]

where \(w \mapsto L(w)\) is a function that varies slowly at infinity. Here we recall that a function \(L\) varies slowly at infinity when, for every \(x > 0\),

\[
\lim_{t \to \infty} \frac{L(tx)}{L(x)} = 1.
\]

Examples of slowly-varying functions are powers of logarithms. See the classical work by Bingham, Goldie and Teugels Bingham et al. (1989) for more information about regularly-varying functions. We interpret \(\tau > 1\) as the final parameter of our model, next to \(\alpha, \lambda\) (and the dimension \(d\)). Of course, there may be many vertex-weight distributions having the asymptotics in (9.5.54) with the same \(\tau\), but the role of \(\tau\) is so important in the sequel that we separate it out.

Write \(D_x\) for the degree of \(x \in \mathbb{Z}^d\) and note that, by translation invariance, \(D_x\) has the same distribution as \(D_0\). The name **scale-free** percolation is justified by the following theorem:

**Theorem 9.35 (Power-law degrees for power-law weights Deijfen et al. (2013))**

Fix \(d \geq 1\).

(a) Assume that the weight distribution satisfies (9.5.54) with \(\alpha \leq d\) or \(\gamma = \alpha(\tau - 1)/d \leq 1\). Then \(\mathbb{P}(D_0 = \infty | W_0 > 0) = 1\).

(b) Assume that the weight distribution satisfies (9.5.54) with \(\alpha > d\) and \(\gamma = \alpha(\tau - 1)/d > 1\). Then there exists \(s \mapsto \ell(s)\) that is slowly varying at infinity such that

\[
\mathbb{P}(D_0 > s) = s^{-\gamma}\ell(s).
\]

The fact that the degrees have a power-law distribution is why this model is called **scale-free percolation**. The parameter \(\gamma\) measures how many moments of the degree distribution are finite. In general, when edges are present independently, but not with the same probability, it is impossible to have infinite-variance degrees in the long-range setting (see Exercise 9.39). We continue by studying the percolative properties of scale-free percolation. As usual, denote by \(\mathcal{C}(x) = \{y: x \leftrightarrow y\}\) the connected component or cluster of \(x\), and by \(|\mathcal{C}(x)|\) the number of vertices in \(\mathcal{C}(x)\). Further, define the **percolation probability** as

\[
\theta(\lambda) = \mathbb{P}(|\mathcal{C}(0)| = \infty),
\]

(9.5.57)
and the critical percolation threshold $\lambda_c$ as

$$\lambda_c = \inf\{\lambda : \theta(\lambda) > 0\}. \tag{9.5.58}$$

As before, we denote by $\lambda_c$ the infimum of all $\lambda \geq 0$ with the property $\mathbb{P}(|C(0)| = \infty) > 0$. It is a priori unclear whether $\lambda_c < \infty$ or not. Deijfen et al. (Deijfen et al., 2013, Theorem 3.1) prove that $\lambda_c < \infty$ holds in most cases. Indeed, if $\mathbb{P}(W = 0) < 1$, then $\lambda_c < \infty$ in all $d \geq 2$. Naturally, $d = 1$ again is special and the results in (Deijfen et al., 2013, Theorem 3.1) are not optimal. It is shown that if $\alpha \in (1, 2]$ and $\mathbb{P}(W = 0) < 1$, then $\lambda_c < \infty$ in $d = 1$, while if $\alpha > 2$ and $\tau > 1$ is such that $\gamma = \alpha(\tau - 1)/d > 2$, then $\lambda_c = \infty$ in $d = 1$.

More interesting is whether $\lambda_c = 0$ or not. The following theorem shows that this depends on whether the degrees have infinite variance or not:

**Theorem 9.36 (Positivity of the critical value Deijfen et al. (2013))** Assume that the weight distribution satisfies (9.5.54) with $\tau > 1$ and that $\alpha > d$.

(a) Assume that $\gamma = \alpha(\tau - 1)/d > 2$. Then, $\theta(\lambda) = 0$ for small $\lambda > 0$, that is, $\lambda_c > 0$.

(b) Assume that $\gamma = \alpha(\tau - 1)/d < 2$. Then, $\theta(\lambda) > 0$ for every $\lambda > 0$, that is, $\lambda_c = 0$.

In ordinary percolation, instantaneous percolation in the form $\lambda_c = 0$ can only occur when the degree of the graph is infinite. The randomness in the vertex weights facilitates instantaneous percolation in scale-free percolation. We see a similar phenomenon for rank-1 inhomogeneous random graphs, such as the Norros-Reittu model. The instantaneous percolation is related to robustness of the random network under consideration. Again we see that graph distances are rather small if $\gamma \in (1, 2)$, whereas graph distances are much larger for $\gamma > 2$.

We see that $\gamma \in (1, 2)$, where the variance of the degrees is infinite, is special in the sense that instantaneous percolation occurs as for rank-1 random graphs. This raises the questions to which extent the analogy extends. For example, in rank-1 random graphs, the scaling limits within the scaling window are different for random graphs having infinite third moments of the degrees than for those for which the third moment is finite. This indicates that the critical behavior of scale-free percolation might be different for $\gamma \in (2, 3)$, where the degrees have infinite third moment, compared to $\gamma > 3$ where the degrees have finite third moment, particularly in high dimensions. Indeed, we can think of the Norros-Reittu as a kind of mean-field model for this setting, certainly when we restrict scale-free percolation to the torus. It would be of great interest to investigate these models in more detail.

**Another scale-free percolation model by Yukich**

In this section, we discuss the results in Yukich (2006) on an infinite scale-free percolation model. Note that, for a transitive graph with fixed degree $r$ and percolation with a fixed percolation parameter $p$, the degree of each vertex has
a binomial distribution with parameters \( r \) and \( p \). Since \( r \) is fixed, this does not allow for a power-law degree sequence. As a result, it is impossible to have a scale-free random graph when dealing with independent percolation, so that we shall abandon the assumption of independence of the different edges, while keeping the assumption of translation invariance.

The model considered in Yukich (2006) is on \( \mathbb{Z}^d \), and, thus, the definition of a scale-free graph process does not apply so literally. We adapt the definition slightly by saying that an infinite random graph is scale-free when

\[
P_k = \mathbb{P}(D_o = k),
\]

where \( D_o \) is the degree of vertex \( x \in \mathbb{Z}^d \) and \( o \in \mathbb{Z}^d \) is the origin, satisfies (1.4.4). This is a reasonable definition, since if let \( B_r = [-r,r]^d \cap \mathbb{Z}^d \) be a cube of width \( r \) around the origin, and denote \( n = (2r+1)^d \), then, for each \( k \geq 0 \),

\[
P^n_k = \frac{1}{n} \sum_{x \in B_r} \mathbb{1}_{(D_o = k)},
\]

which, assuming translation invariance and ergodicity, converges to \( p_k \).

We next describe the model in Yukich (2006). We start by taking an i.i.d. sequence \( \{U_x \}_{x \in \mathbb{Z}^d} \) of uniform random variables on \([0,1]\). Fix \( \delta \in (0,1] \) and \( q \in (1/d, \infty) \). The edge \( \{x,y\} \in \mathbb{Z}^d \times \mathbb{Z}^d \) appears in the random graph precisely when

\[
|x - y| \leq \delta \min\{U_x^{-q}, U_y^{-q}\}.
\]

We can think of the ball of radius \( \delta U_x^{-q} \) as being the region of influence of \( x \), and two vertices are connected precisely when each of them lies into the region of influence of the other. This motivates the choice in (9.5.61). The parameter \( \delta \) can be interpreted as the probability that nearest-neighbors are connected, and in the sequel we shall restrict ourselves to \( \delta = 1 \), in which case the infinite connected component equals \( \mathbb{Z}^d \). We denote the resulting (infinite) random graph by \( G_q \).

We next discuss the properties of this model, starting with its scale-free nature. In (Yukich, 2006, Theorem 1.1), it is shown that, with \( \tau = qd/(qd - 1) \in (1, \infty) \), the limit

\[
\lim_{k \to \infty} k^{\tau-1} \mathbb{P}(D_o \geq k)
\]

exists, so that the model is scale free with degree power-law exponent \( \tau \) (recall (1.4.3)). The intuitive explanation of (9.5.62) is as follows. Suppose we condition on the value of \( U_o = u \). Then, the conditional distribution of \( D_o \) given that \( U_o = u \) is equal to

\[
D_o = \sum_{x \in \mathbb{Z}^d} \mathbb{1}_{\{|x| \leq \min\{U_x^{-q}, U_o^{-q}\}\}} = \sum_{x: |x| \leq u^{-q}} \mathbb{1}_{\{|x| \leq u^{-q}\}}.
\]

Note that the random variables \( \{\mathbb{1}_{\{|x| \leq u^{-q}\}}\}_{x \in \mathbb{Z}^d} \) are independent Bernoulli random variables with probability of success equal to

\[
\mathbb{P}(\mathbb{1}_{\{|x| \leq u^{-q}\}} = 1) = \mathbb{P}(U \leq |x|^{-1/q}) = |x|^{-1/q}.
\]
In order for $D_o \geq k$ to occur, for $k$ large, we must have that $U_o = u$ is quite small, and, in this case, a central limit theorem should hold for $D_o$, with mean equal to

$$\mathbb{E}[D_o \mid U_o = u] = \sum_{x: |x| \leq u^{-q}} |x|^{-1/q} = cu^{-(qd-1)}(1 + o(1)), \quad (9.5.65)$$

for some explicit constant $c = c(q,d)$. Furthermore, the conditional variance of $D_o$ given that $U_o = u$ is bounded above by its conditional expectation, so that the conditional distribution of $D_o$ given that $U_o = u$ is highly concentrated. We omit the detail, and merely note that this can be made precise by using standard large deviations result. Assuming sufficient concentration, we obtain that the probability that $D_o \geq k$ is asymptotically equal to the probability that $U \leq u_k$, where $u_k$ is determined by the equation that

$$\mathbb{E}[D_o \mid U_o = u_k] = cu_k^{-(qd-1)}(1 + o(1)) = k, \quad (9.5.66)$$

so that $u_k = (k/c)^{-1/(qd-1)}$. This suggests that

$$P(D_o \geq k) = P(U \leq u_k)(1 + o(1)) = (ck)^{-1/(qd-1)}(1 + o(1)), \quad (9.5.67)$$

which explains (9.5.62).

We next turn to distances in this scale-free percolation model. For $x,y \in \mathbb{Z}^d$, we denote by $d_G(x,y)$ the graph distance (or chemical distance) between the vertices $x$ and $y$, i.e., the minimal number of edges in $G_q$ connecting $x$ and $y$. The main result in Yukich (2006) is the following theorem:

**Theorem 9.37** (Ultra-small distances for scale-free percolation) *For all $d \geq 1$ and all $q \in (1/d, \infty)$, whp as $|x| \to \infty$,*

$$d_{G_q}(o,x) \leq 8 + 4 \log \log |x|. \quad (9.5.68)$$

The result in Theorem 9.37 shows that distances in the scale-free percolation model are much smaller than those in normal percolation models. It would be of interest to investigate whether the limit $d_{G_q}(o,x)/\log \log |x|$ exists, and, if so, what this limit is.

While Theorem 9.37 resembles the results in Theorem 7.21, there are a few essential differences. First of all, $G_q$ is an infinite graph, whereas the models considered in Theorem 7.21 are all finite. It would be of interest to extend Theorem 9.37 to the setting on finite tori, where the Euclidean norm $|x - y|$ in (9.5.61) is replaced by the Euclidean norm on the torus, and the typical distance $H_n$ is considered. This result is not immediate from the proof of Theorem 9.37. Secondly, in Theorems 7.20 and 7.21, it is apparent that the behavior for $\tau > 3$ is rather different compared to the behavior for $\tau \in (2,3)$. This feature is missing in Theorem 9.37. It would be of interest to find a geometric random graph model where the difference in behavior between $\tau > 3$ and $\tau \in (2,3)$ also appears.

The result in Theorem 9.37 can be compared to similar results for long-range percolation, where edges are present independently, and the probability that the edge $\{x,y\}$ is present equals $|x - y|^{-s+o(1)}$ for some $s > 0$. In this case, detailed results exist for the limiting behavior of $d(o,x)$ depending on the value of $s$. For
example, in Benjamini et al. (2004), it is shown that the diameter of this infinite percolation model is equal to \( \lceil d/(d - s) \rceil \) a.s. See also Biskup (2004) and the references therein.

**Spatial configuration models on the lattice**

**TO DO 9.14:** Introduce the spatial configuration model and results of Deijfen et al.

### 9.6 Notes and discussion

**Notes on Section 9.1.**

*Citation networks.* Our discussion of citation networks is inspired by, and follows, Garavaglia et al. (2017). For citation networks, there is a rich literature to propose models for them using preferential attachment schemes and adaptations of them, mainly in the physics literature. Aging effects, i.e., considering the *age of a vertex* in its likelihood to obtain children, have been extensively considered as the starting point to investigate their dynamics, see Wang et al. (2009, 2008); Hajra and Sen (2005, 2006); Csárdi (2006). Here the idea is that old papers are less likely to be cited than new papers. Such aging has been observed in many citation network datasets and makes PAMs with weight functions depending only on the degree ill-suited for them. As mentioned above, such models could more aptly be called *old-get-richer* models, i.e., in general old vertices have the highest degrees. In citation networks, instead, papers with many citations appear all the time. Wang et al. (2013) investigate a model that incorporates these effects, see also Wang et al. (2014) for a comment on the methods in this paper. On the basis of empirical data, they suggest a model where the aging function follows a lognormal distribution with paper-dependent parameters, and the preferential attachment function is the identity. Wang et al. (2013) *estimate* the fitness function rather than the more classical approach where it is taken to be an i.i.d. sample of random variables.

**Notes on Section 9.2.**

The local weak convergence for PageRank was proved by Garavaglia et al. (2020), where also the notion of the marked backward local weak limit was introduced. We have extended the discussion on local weak convergence for directed graphs here, since these other notions are useful in other contexts as well, for example in studying the strongly connected component.

*Directed inhomogeneous random graphs.* Bloznelis et al. (2012) study the general directed inhomogeneous random graph studied in this section, and prove Theorem 9.3. Cao and Olvera-Cravioto (2020) continue this analysis, and generalize it substantially. While the local weak convergence result in Theorem 9.2 has not been proved anywhere explicitly, it is the leading idea in the identification of the phase transition, as well as the description of the limiting branching processes and joint degree distribution. Garavaglia et al. (2020) does investigate the marked di-
rected forward local weak convergence for directed rank-1 inhomogeneous random graphs.

Cao and Olvera-Cravioto (2020) specifically investigate general rank-1 inhomogeneous digraphs. Our choice of edge probabilities in (9.2.14)–(9.2.15) is slightly different from that by Cao and Olvera-Cravioto (2020), particularly since the factor $\frac{1}{2}$ in (9.2.15) is missing in Cao and Olvera-Cravioto (2020). We have added it so as to make the average in-degree of a vertex of in-weight $w_i^{(in)}$ approximately $w_i^{(in)}$. If this were to be true for every $i$, then we would also need that

$$\sum_{i \in [n]} w_i^{(in)} \approx \sum_{i \in [n]} w_i^{(out)},$$

which would imply the limiting statement in (9.2.17). Lee and Olvera-Cravioto (2017) use the results in Cao and Olvera-Cravioto (2020) to prove that the limiting PageRank of such directed generalized random graphs exists, and that the solution obeys the same recurrence relation as on a Galton-Watson tree. In particular, under certain independence assumptions, this implies that the PageRank power-law hypothesis is valid for such models.

**Directed configuration models.** The directed configuration model was investigated by Cooper and Frieze (2004), where the results discussed here are proved. In fact, the results in Cooper and Frieze (2004) are much more detailed than the one in Theorem 9.5, and also include detailed bounds on the strongly connected component in the subcritical regime, as well as precise bounds on the number of vertices whose forward and backward clusters are large and the asymptotic size of forward and backward clusters. A substantially simpler proof was given by Cai and Perarnau (2020b).

Both Cooper and Frieze (2004) as well as Cai and Perarnau (2020b) make additional assumptions on the degree distribution. In particular, they assume that $E[D_{n}^{(in)}D_{n}^{(out)}] \rightarrow E[D_{\infty}^{(in)}D_{\infty}^{(out)}] < \infty$, which we do not assume. Further, Cooper and Frieze (2004) assume that $d$ is proper, which is a technical requirement on the degree sequence stating that (a) $E[(D_{n}^{(in)})^2] = O(1)$, $E[(D_{n}^{(out)})^2] = O(1)$; (b) $E[D_{n}^{(in)}(D_{n}^{(out)})^2] = o(n^{4/5} \log n)$. In view of the fact that such conditions do not appear in Theorem 4.4, these conditions can be expected to be suboptimal for Theorem 9.5 to hold, and we next explain how they can be avoided by a suitable truncation argument.

Assume that the out- and in-degrees in directed configuration model $DCM_n(d)$ satisfy (9.2.24) and (9.2.25). By Exercise 9.17 below, $|\mathcal{C}_\max| \leq n(\zeta + \varepsilon)$ whp for $n$ large and any $\varepsilon > 0$. Exercise 9.17 is proved by an adaptation of the proof of Corollary 2.25 in the undirected setting. Thus, we only need to show that $|\mathcal{C}_\max| \leq n(\zeta - \varepsilon)$ whp for $n$ large and any $\varepsilon > 0$.

Fix $K > 1$ very large. We now construct a lower bounding directed configuration model where all the degrees are bounded by $K$. This is similar to the construction for the undirected configuration model in Section 4.2.1. When $v$ is such that $d_v = d_v^{(out)} + d_v^{(in)} \geq K$, we split $v$ into $n_v = \lceil d_v/K \rceil$ vertices, and we deterministically redistribute all out- and in-half-edges over the $n_v$ vertices in an
arbitrary way such that the out- and in-degrees of all the vertices that used to correspond to \( v \) now have both out- and in-degree bounded by \( K \). Denote the corresponding random graph by \( \text{DCM}^n \).

The resulting degree sequence again satisfies (9.2.24) and (9.2.25). Moreover, for \( K > 1 \) large and by (9.2.24) and (9.2.25), the limits in (9.2.24) and (9.2.25) for the new degree sequence are quite close to the original limits of the old degree sequence, while at the same time now having bounded degrees. As a result, we can apply the original result in Cooper and Frieze (2004) or Cai and Perarnau (2020b) to the new setting.

Denote the size of the SCC by \( |C'| \). Obviously, since we split vertices, \( |C'| \leq |C| + \sum_{v \in [n]} (n_v - 1) \). Therefore, \( |C'| \geq |C| - \sum_{v \in [n]} (n_v - 1) \). Take \( K > 0 \) so large that \( \sum_{v \in [n]} (n_v - 1) \leq \varepsilon n / 3 \) and that \( \zeta' \geq \zeta - \varepsilon / 3 \), where \( \zeta' \) is the backward-forward survival probability of the limiting directed \( \text{DCM}' \) and \( \zeta \) that of \( \text{DCM}_n(d) \). Finally, for every \( \varepsilon > 0 \), whp \( |C'| \geq n(\zeta' - \varepsilon / 3)n \). As a result, we obtain that, again whp,

\[
|C'| \geq |C'| - n\varepsilon / 3 \geq n(\zeta' - 2\varepsilon / 3) \geq n(\zeta - \varepsilon),
\]

(9.6.1)
as required.

Chen and Olvera-Cravioto (2013) study a way to obtain nearly i.i.d. in-and out-degrees in the directed configuration model. Here, the problem is that when \( ((d_i^{\text{in}}, d_i^{\text{out}}))_{i \in [n]} \) is an i.i.d. bivariate distribution with equal means, then \( \sum_{i \in [n]} (d_i^{\text{in}} - d_i^{\text{out}}) \) has Gaussian fluctuations at best, so that it will not be zero. Chen and Olvera-Cravioto (2013) indicate how the excess in- or out-half-edges can be removed so as to keep the degrees close to i.i.d. Further, they also show that the removal of self-loops and multiple directed edges does not significantly change the degree distribution (so that in particular, one would assume that the local weak limits are the same, but Chen and Olvera-Cravioto (2013) stick to the degree distribution).

Theorem 9.7 follows from (Cai and Perarnau, 2020a, Proposition 7.7). Theorem 9.6 was first proved by van der Hoorn and Olvera-Cravioto (2018) under stronger assumptions, but then also the claim proved is much stronger. Indeed, van der Hoorn and Olvera-Cravioto (2018) not only identify the first order asymptotics as in Theorem 9.7, but also the fluctuations like those stated for the undirected configuration model in Theorem 7.20. This proof is substantially harder that that of (Cai and Perarnau, 2020a, Proposition 7.7). Theorem 9.7 was proved by Cai and Perarnau (2020a).

Notes on Section 9.3.

Stochastic blockmodel. Stochastic blockmodels were introduced by Holland et al. (1983) in the context of group structures in social networks. Their introduction was inspired by the work of Fienberg and Wasserman (1981). There, also the problem of community detection was discussed. Of course, we have already seen them in the context of inhomogeneous random graphs, so our focus here is on the detection problem.
The definition of solvable in Definition 9.8, in particular (9.3.2), has appeared in different forms in the literature. For example, Bordenave et al. (2018) instead use the definition

$$\min_{p: [r] \to [r]} \frac{1}{n} \sum_{v \in [n]} \left[ \mathbb{1}_{\{\sigma(v) = (p \circ \sigma)(v)\}} - \frac{1}{r} \right],$$

(9.6.2)

where the minimum is over all possible permutations $p$ from $[r]$ to $[r]$.

The precise threshold in Theorem 9.9 was conjectured by Decelle et al. (2011), based on a non-rigorous belief propagation method. The proof of the impossibility result of the block model threshold conjecture was first given by Mossel et al. (2015). The proof of the solvable part of the block model threshold conjecture was first given by Mossel et al. (2018), and independently by Massoulié (2014). Mossel et al. (2016) give an algorithm that maximizes the fraction of vertices labeled correctly. The result were announced in a high-impact format by Krzakala et al. (2013). The achievability result for general number of types $r$ in (9.3.5) was proved by Abbe and Sandon (2018) and Bordenave et al. (2018). We follow the presentation in Bordenave et al. (2018). The convergence of the eigenvalues in (9.3.8) is (Bordenave et al., 2018, Theorem 4), the estimation of the types in (9.3.10) is investigated in (Bordenave et al., 2018, Theorem 5). There, it is also shown that this choice has the positive overlap property in (9.6.2), which implies the solvability in Definition 9.8. We have simplified the presentation substantially by considering $r = 2$ and an equal size of the groups of the two types.

**Degree-corrected stochastic blockmodel.** The degree-corrected stochastic block model was first introduced in Karrer and Newman (2011). Consistent estimation of communities in degree-corrected stochastic block model was investigated by Zhao et al. (2012), under the assumption that the average degree tends to infinity for weak consistency, and grows faster than $\log n$ for strong consistency. Sparse settings suffer from a similar threshold phenomenon as for the original stochastic blockmodel as derived in Mossel et al. (2018); Massoulié (2014). The impossibility result in Theorem 9.10 was proved by Gulikers et al. (2018). The proof of the solvable parts were proved in Gulikers et al. (2017a,b).

**Preferential attachment models with community structure.** Jordan (2013) investigates preferential attachment models in general metric spaces. When one takes these metric spaces as discrete sets, one can interpret the geometric location as a type or community label. This interpretation was proposed by Hajek and Sankagiri (2018), where also our results on community detection are proved. We will return to the geometric interpretation of the model in Section 9.5.4. The result in (9.3.23) is stated in (Jordan, 2013, Theorem 2.1). The asymptotics of the degree distribution in Theorem 9.11 is (Jordan, 2013, Theorem 2.2). (Jordan, 2013, Theorem 2.3) further contains some estimates of the number of vertices in given regions and with given degrees. We come back to such issues below. The convergence of the proportion of errors in (9.3.29) is stated in (Hajek and Sankagiri, 2018, Proposition 8), to which we refer for the formula for Err.

A preferential attachment model with community structure, phrased as a coex-
istence model, was introduced by Antunović et al. (2016). In their model, contrary to the setting of Jordan (2013), the vertices choose their type based on the types of their neighbors, thus creating the possibility of denser connectivity between vertices of the same type and thus community structure. Again, this may lead to community structure. The focus of Antunović et al. (2016) is the coexistence of all the different types or rather a winner-takes-it-all phenomenon, depending on the precise probability of choosing a type depending on the number of neighbors of all types. Even with two types, the behavior is quite involved and depends sensitively on the type choosing distribution. For example, for the majority rule (where the type of the new vertex is the majority of types of its older neighbors), the winner type takes it all, while if this probability is linear in the number of older neighbors of a given type, there is always coexistence.

Notes on Section 9.4.

Empirical properties of real-world network with community structure were studied in Stegehuis et al. (2016b).

Inhomogeneous random graphs with communities. The inhomogeneous random graphs with communities was introduced by Bollobás et al. (2011).

Configuration models with community structure. The hierarchical configuration model was introduced by van der Hofstad et al. (2011). The fit to real-world networks, particularly in the context of epidemics, was studied by Stegehuis et al. (2016a). The configuration model with household structure was investigated in Ball et al. (2009, 2010) in the context of epidemics on social networks. Particularly when studying epidemics on networks, clustering is highly relevant, as clustering slows down the spread of infectious diseases. Random intersection graph with prescribed degrees and groups are studied in a non-rigorous way in Newman (2003); Newman and Park (2003).

The configuration model with clustering was defined by Newman (2009), who studied it non-rigorously.

Random intersection graphs. Random intersection graphs were introduced by Singer (1996) and further studied in Fill et al. (2000); Karoński et al. (1999); Stark (2004). Theorem 9.17 is (Deijfen and Kets, 2009, Theorem 1.1), where the authors also proved that the clustering can be controlled. The model has also been investigated for more general distributions of groups per vertex by Godehardt and Jaworski (2003) and Jaworski et al. (2006). We refer to Bloznelis et al. (2015) for a survey of recent result.

Rybarczyk (2011) studies various properties of the random intersection graph when each vertex is in precisely $d$ groups that are all chosen uniformly at random from the collection of $m$ groups. In particular, Rybarczyk (2011) proves results on the giant as in Theorem 9.19, as well as on the diameter of the graph, which is $\Theta(\log n)$ when the model is sparse.

Mindaugas Bloznelis (2009, 2010a,b) studies a general random intersection model, where the sizes of groups are i.i.d. random variables, and then set of
the vertices in them are chosen uniformly at random from the vertex set. His results include distances Bloznelis (2009) and component sizes Bloznelis (2010a,b). Bloznelis (2013) studies degree and clustering structure in this setting.

Theorem 9.18 is proved by Kurauskas (2015), see also van der Hofstad et al. (2019b). Both papers investigate more general settings, Kurauskas (2015) allows also for settings with independent group memberships, while van der Hofstad et al. (2019b) also allows for more general group structures than the complete graph.

Theorem 9.19 is proved by van der Hofstad et al. (2019a).

Random intersection graphs with communities. van der Hofstad et al. (2019b) propose a model that combines the random intersection graph with more general communities than complete graphs. Hofstad et al. (2019b) identify the local limit, as well as the nature of the overlaps between different communities. van der Hofstad et al. (2019a) identifies the giant component, also when performing percolation on the model. See Vadon et al. (2019) for an informal description of the model, aimed at a broad audience.

Exponential random graphs. For a general introduction to exponential random graphs, we refer to Snijders et al. (2006); Wasserman and Pattison (1996). Frank and Strauss (1986) discuss the notion of Markov graphs, for which the edges of the graph form a Markov field. The general exponential random graph is only a Markov field when the subgraphs are restricted to edges, stars of any kind, and triangles. This is exemplified by Example 9.21, where general degrees were use and gave rise to a model with independent edges. Kass and Wasserman (1996) discuss relations to Bayesian statistics.

For a discussion on the relation between statistical mechanics and exponential models, we refer to Jaynes (1957). Let us know explain the relation between exponential random graphs and entropy maximization. Let \((p_x)_{x \in X}\) be a probability measure on a general discrete set \(X\). We define its entropy by

\[
H(p) = -\sum_{x \in X} p_x \log p_x. \tag{9.6.3}
\]

Shannon (1948) proved that the entropy is the unique the quantity that is positive, increases with increasing uncertainty, and is additive for independent sources of uncertainty, so it is a very natural quantity. Entropy measures the amount of randomness in a system. The relation to exponential random graphs is that they are the random graphs that, given the conserved expected values of the subgraph counts \(N_F(G_n)\), optimize the entropy. Indeed, recall that \(X = (X_{i,j})_{i \leq j \leq n}\) are the edge statuses of the graph, so that \(G_n\) is uniquely characterized by \(X\), and maximize \(H(p)\) over all the \(p\) such that \(\sum_x N_F(x) = \alpha_F\) for some given \(\alpha_F\) and all subgraphs \(F \in \mathcal{F}\), where \(\mathcal{F}\) is an appropriate set of subgraphs. Then, using Lagrange multipliers, the optimization problem reduces to

\[
p_{\beta}(x) = \frac{1}{Z} e^{\sum_{F \in \mathcal{F}} \beta_F N_F(x)}, \tag{9.6.4}
\]
where \( Z = Z_n(\bar{\beta}) \) is the normalization constant given in (9.4.34), and \( \bar{\beta} = (\bar{\beta}_F)_{F \in \mathcal{F}} \) is chosen as the solution to (9.4.35). This implies that, indeed, the exponential random graph model optimizes the entropy under this constraint.

See also Kass and Wasserman (1996) for a discussion of maximum entropy, and a reference to its long history as well as critique on the method.

An important question is how one can find the appropriate \( \bar{\beta} = (\beta_F)_{F \in \mathcal{F}} \) such that (9.4.35) holds. This is particularly difficult, since the computation of the normalization constant \( Z_n(\bar{\beta}) \) in (9.4.34) is quite hard. Often, Markov Chain Monte Carlo (MCMC) techniques are used to sample from \( p \) efficiently. In this case, such MCMC techniques perform a form of Glauber dynamics, for which (9.4.33) is the stationary distribution. One can then try to solve (9.4.35) by keeping track of the value of \( N_F \) in the simulation. However, these methods can be very slow, as well as daunting, since the behavior of \( N_F(X) \) under \( p = p_{\bar{\beta}} \) may undergo a phase transition, making \( \sum_x N_F(x)p_{\bar{\beta}}(x) \) very sensitive to small changes of \( \bar{\beta} \). See in particular Chatterjee and Diaconis (2013) for a discussion on this topic.

Bhamidi et al. (2011) (see also Bhamidi et al. (2008)) have studied the mixing time of the exponential random graph, when edges are changed dynamically in a Glauber way. The results are somewhat disappointing, since either edges are close to being i.i.d. or the mixing is very slow. These results, however, apply only to dense settings where the number of edges grows quadratically with the number of vertices. This problem is closely related to large deviations on dense Erdős-Rényi random graph. See also Chatterjee (2017) for background on such large deviations, and Chatterjee and Varadhan (2011) for the original paper.

For more background on sufficient statistics and their relation to symmetries, we refer to Diaconis (1992). For a discussion of the relation between information theory and exponential models, we refer to Shore and Johnson (1980).

Notes on Section 9.5.

The Newman-Watts small-world model. There are various ways of adding long-range connections (for example by rewiring the existing edges), and we shall focus on the models in Barbour and Reinert (2001, 2004, 2006), for which the strongest mathematical results have been obtained. Small-world models were first introduced and analyzed in Moore and Newman (2000); Newman et al. (2000a); Newman and Watts (1999), and a non-rigorous mean-field analysis of distances in small-world models was performed in Newman et al. (2000a). See Barbour and Reinert (2001) for a discussion of the differences between the exact and mean-field analyses.

Theorem 9.22 is (Barbour and Reinert, 2001, Theorem 3.9). The proof of Theorem 9.22 was extended by Barbour and Reinert (2006) to deal with discrete tori where the shortcuts also contribute one to the graphs distance, so that distances are the usual distances on discrete graphs.

Theorem 9.23 follows from (Barbour and Reinert, 2006, Theorem 2.1), which even allows \( \rho \) to grow with \( n \) (though not too quickly).
A related model was considered by Turova and Vallier (2010). Indeed, Turova and Vallier (2010) study a mixture between subcritical percolation on a finite cube and the Erdős-Rényi random graph. Using the methodology of Bollobás et al. (2007), it is shown that the phase transition is similar to the one described in Theorem 3.16. It would be of interest to verify whether the distance results in Bollobás et al. (2007) can also be used to prove that the distances grow like \( \log \nu n \), where \( n \) is the size of the graph, and \( \nu > 1 \) an appropriate constant.

Hyperbolic random graph. The hyperbolic random graph was introduced by Krioukov et al. (2010). The first rigorous results were proved by Gugelmann et al. (2012), who proved Theorem 9.24 and identified the exact asymptotic degree distribution (see (Gugelmann et al., 2012, Theorem 2.2)), as well as the positivity of the clustering in the graph (see (Gugelmann et al., 2012, Theorem 2.1)). The statement about the maximal degree in the hyperbolic graph is stated in (Gugelmann et al., 2012, Theorem 2.4). The sharpest results on the degree distribution and clustering were proved by Fountoulakis et al. (2020), who identify the exact clustering coefficient in a quite long and technical paper, who also proved Theorem 9.27.


Bläsius et al. (2018) study the size of the largest cliques in hyperbolic graphs.

Geometric inhomogeneous random graph. The geometric inhomogeneous random graph was introduced by Bringmann et al. (2015, 2016). The relation between the hyperbolic random graph and the general GIRG as described in Theorem 9.29 can be found in (Bringmann et al., 2017, Section 7), where limits are derived up to constants. Theorem 9.29 is proved in (Komjáthy and Lodewijks, 2020, Section 7), who also study its weighted distances focussing on the case where \( \tau \in (2, 3) \). Theorem 9.31 is proved by Komjáthy and Lodewijks (2020). In more detail, a coupling version of Theorem 9.31 is stated in (Komjáthy and Lodewijks, 2020, Claim 3.3), where a blown-up version of the GIRG is bounded from below and above by the limiting model with slightly smaller and larger intensities, respectively. Take a vertex uniformly at random in the GIRG. Then, whp it is also present in the lower and upper bounding limiting Poisson infinite GIRG with edge-probabilities given by \( h \). Similarly, whp none of the edges within a ball of intrinsic radius \( r \) will be different in the three models, which proves local weak convergence. Local convergence in probability would follow from a coupling of the neighborhoods of two uniformly chosen vertices in the GIRG to two independent limiting copies. Such independence is argued in (Komjáthy and Lodewijks, 2020, Proof of Theorem 2.12), so in particular around (Komjáthy and Lodewijks, 2020, (3.16)), this result is not stated.

Local weak limits as arising in Theorem 9.31, in turn, were studied by Hirsch

**Scale-free percolation.** Graph distances in scale-free percolation have been investigated in Deijfen et al. (2013); Heydenreich et al. (2016) by identifying the number of edges between $x$ and $y$ as a function of $|x - y|$ for $x, y$ in the infinite component. There is some follow-up work on scale-free percolation. Hirsch (2017) proposes a continuum model for scale-free percolation. Deprez et al. (2015) argue that scale-free percolation can be used to model real-life networks. Bringmann et al. (2015, 2016) study this model on a torus and in continuum space and coin the name geometric inhomogeneous random graphs. Heydenreich et al. (2016) establish recurrence and transience criteria. Deprez et al. (2015) show that when $\alpha \in (d, 2d)$, then the percolation function is continuous. For long-range percolation this was proved by Berger (2002). However, in full generality, continuity of the percolation function as $\lambda = \lambda_c$ when $\lambda_c > 0$ is unknown.

**Spatial preferential attachment models.** Jordan (2013) studies preferential attachment models where the vertices are located in a general metric space. Flaxman et al. (2006, 2007) study geometric preferential attachment models. Aiello et al. (2008) give the interpretation of spatial preferential attachment models in terms of influence regions.

For a relation between preferential attachment graphs with so-called fertility and aging, and a geometric competition-induced growth model for networks, we refer to Berger et al. (2004, 2005) and the references therein.

**Complex network models on the hypercubic lattice.** Spatial configuration models on the lattice were introduced by Deijfen and Jonasson (2006), see also Deijfen and Meester (2006). Deijfen (2009) study a related model where the vertices are a Poisson point process on $\mathbb{R}^d$.

### 9.7 Exercises for Chapter 9

**Exercise 9.1** (Topology of the strongly-connected component for digraphs) Let $G$ be a digraph. Prove that if $u$ and $v$ are such that $u$ is connected to $v$ and $v$ is connected to $u$, then the strongly connected components of $u$ and $v$ are the same.

**Exercise 9.2** (Sum of out- and in-degrees digraph agree) Let $G$ be a digraph for which $d_v^{\text{out}}$ and $d_v^{\text{in}}$ denote the out- and in-degree of $v \in V(G)$. Show that

$$\sum_{v \in V(G)} d_v^{\text{out}} = \sum_{v \in V(G)} d_v^{\text{in}}. \quad (9.7.1)$$

**Exercise 9.3** (LWC for randomly directed graphs) Let $(G_n)_{n \geq 1}$ be a random graph sequence that converges in probability in the local weak sense. Give each edge $e$ a random orientation, by orienting $e = \{u, v\}$ as $e = (u, v)$ with probability $\frac{1}{2}$ and as $e = (v, u)$ with probability $\frac{1}{2}$, independently across edges. Show that the resulting digraph converges in probability in the marked forward and backward and forward-backward local weak convergence sense.
9.7 Exercises for Chapter 9

Exercise 9.4 (LWC for randomly directed graphs (Cont.)) In the setting of Exercise 9.3, assume that the convergence of \((G_n)\) is in distribution in the local weak sense. Conclude that the resulting digraph converges in distribution in the marked forward and backward and forward-backward local weak convergence sense.

Exercise 9.5 (LWC for directed version of \(\text{PA}_n^{(m,\delta)}(d)\)) Consider the edges in \(\text{PA}_n^{(m,\delta)}(d)\) to be oriented from young to old, so that the resulting digraph has out-degree \(m\) and random in-degrees. Use Theorem 5.8 to show that this digraph converges in probability in the marked forward and backward and forward-backward local weak convergence sense.

Exercise 9.6 (Power-law lower bound for PageRank on directed version of \(\text{PA}_n^{(m,\delta)}(d)\)) Recall the directed version of \(\text{PA}_n^{(m,\delta)}(d)\) in Exercise 9.5. Use (9.2.10) to show that there exists a constant \(c = c(\alpha, \delta, m) > 0\) such that

\[
P(\mathbb{R}_\emptyset > r) \geq cr^{-\tau}, \quad \text{where} \quad \tau = 3 + \frac{\delta}{m} \tag{9.7.2}
\]

is the power-law exponent of \(\text{PA}_n^{(m,\delta)}(d)\). What does this say about the PageRank power-law hypothesis for the directed version of \(\text{PA}_n^{(m,\delta)}(d)\)?

Exercise 9.7 (Power-law lower bound for PageRank on digraphs with bounded out-degrees) Let \((G_n)_{n \geq 1}\) be a random digraph sequence that converges in probability in the marked backward local weak sense to \((D, \emptyset)\). Assume that there exist \(0 < a, b < \infty\) such that \(d^{\text{out}}_v \in [a, b]\) for all \(v \in V(G_n)\). Assume that

\[
P(D^{\text{in}}_\emptyset > r) \geq cr^{-\gamma} \quad \text{for some} \quad \gamma > 0.
\]

Use (9.2.10) to show that there exists a constant \(c' > 0\) such that

\[
P(\mathbb{R}_\emptyset > r) \geq c' r^{-\gamma}. \quad \tag{9.7.3}
\]

Assume further that the symmetry condition in (9.2.17) holds, and conclude that

\[
\frac{1}{n} \mathbb{E} \left[ \sum_{i,j \in [n]} X_{ij} \right] \rightarrow \mathbb{E}[W^{\text{in}}] = \mathbb{E}[W^{\text{out}}]. \tag{9.7.4}
\]

Exercise 9.8 (Mean number of edges in \(\text{DGRG}_n(w)\)) Consider the directed generalized random graph, as formulated in (9.2.14)–(9.2.15). Assume that the weight-regularity condition in (9.2.16) holds. Let \(X_{ij}\) be the indicator that there is a directed edge from \(i\) to \(j\) (with \(X_{ii} = 0\) for all \(i \in [n]\) by convention). Show that

\[
\frac{1}{n} \mathbb{E} \left[ \sum_{i,j \in [n]} X_{ij} \right] \rightarrow \frac{2\mathbb{E}[W^{\text{in}}]\mathbb{E}[W^{\text{out}}]}{\mathbb{E}[W^{\text{in}}] + W^{\text{out}}}.
\]

Assume further that the symmetry condition in (9.2.17) holds, and conclude that

\[
\frac{1}{n} \mathbb{E} \left[ \sum_{i,j \in [n]} X_{ij} \right] \rightarrow \mathbb{E}[W^{\text{in}}] = \mathbb{E}[W^{\text{out}}]. \tag{9.7.5}
\]

Exercise 9.9 (Number of edges in \(\text{DGRG}_n(w)\)) In the setting of Exercise 9.8, show that

\[
\frac{1}{n} \sum_{i,j \in [n]} X_{ij} \rightarrow \frac{2\mathbb{E}[W^{\text{in}}]\mathbb{E}[W^{\text{out}}]}{\mathbb{E}[W^{\text{in}}] + W^{\text{out}}}. \tag{9.7.6}
\]

Assume further that the symmetry condition in (9.2.17) holds, and conclude that

\[
\frac{1}{n} \sum_{i,j \in [n]} X_{ij} \rightarrow \mathbb{E}[W^{\text{in}}] = \mathbb{E}[W^{\text{out}}].
\]
Exercise 9.10 (Local weak limit of directed Erdős-Rényi random graph) Use Theorem 9.2 to describe the local weak limit of the directed Erdős-Rényi random graph.

Exercise 9.11 (LWC for finite-type directed inhomogeneous random graphs) Adapt the proof of Theorem 3.11 to prove Theorem 9.2 in the case of finite-type kernels. Here, we recall that a kernel $\kappa$ is called finite type when $(s,t) \mapsto \kappa(s,t)$ takes on finitely many values.

Exercise 9.12 (LWC for DGRG$_n(w)$) Consider the directed generalized random graph, as formulated in (9.2.14)–(9.2.15). Assume that the weight-regularity condition in (9.2.16) holds. Use Theorem 9.2 to determine the local weak limit in probability of DGRG$_n(w)$. Is this local weak limit a single- or a multi-type branching process?

Exercise 9.13 (Phase transition for directed Erdős-Rényi random graph) For the directed Erdős-Rényi random graph, show that $\zeta$ in Theorem 9.3 satisfies $\zeta > 0$ precisely when $\lambda > 1$.

Exercise 9.14 (Phase transition for directed generalized random graph) Consider the directed generalized random graph, as formulated in (9.2.14)–(9.2.15). Assume that the weight-regularity condition in (9.2.16) holds. What is the condition on the asymptotic weight distribution $(W^{(\text{out})}, W^{(\text{in})})$ in (9.2.16) that is equivalent to $\zeta > 0$ in Theorem 9.3?

Exercise 9.15 (Correlation out- and in-degree randomly directed graph) In an undirected graph $G$, randomly direct each edge by orienting $e = \{u,v\}$ as $(u,v)$ with probability $\frac{1}{2}$ and as $(v,u)$ with probability $\frac{1}{2}$, as in Exercise 9.3. Let $v \in V(G)$ be a vertex in $G$ of degree $d_v$. What is the correlation between its out- and in-degree in the randomly directed version of $G$?

Exercise 9.16 (LWC for DCM$_n(d)$ in Theorem 9.4) Give a proof of Theorem 9.4 by suitably adapting the proof of Theorem 4.1.

Exercise 9.17 (One-sided law of large numbers for SSC) Adapt the proof of Corollary 2.25 to show that when $G_n = ([n], E(G_n))$ converges in probability in the local weak sense to $(G,o)$ having distribution $\mu$, then, for every $\varepsilon > 0$ fixed,
\[ \mathbb{P}(|C_{\text{max}}| \leq n(\zeta + \varepsilon)) \to 1, \]
where $\zeta = \mu(|C(o)| = \infty)$ is the forward-backward survival probability of the limiting graph $(G,o)$.

Exercise 9.18 (Subcritical directed configuration model) Let DCM$_n(d)$ be a directed configuration model that satisfies the degree-regularity conditions in Let $C_{\text{max}}$ denote its largest strongly connected component. Use Exercise 9.17 to show that $|C_{\text{max}}|/n \to 0$ when $\zeta = 0$. This proves the subcritical result in Theorem 9.5(b).

Exercise 9.19 (The strongly connected component in temporal networks) Let
Exercise 9.20 (Degree structure in stochastic blockmodels) Recall the definition of the stochastic blockmodel in Section 9.3.1, and assume that (9.3.1) holds. What is the asymptotic expected degree of this model? When do all vertices have the same asymptotic expected degree?

Exercise 9.21 (The giant in stochastic blockmodels) Recall the definition of the stochastic blockmodel in Section 9.3.1, and assume that (9.3.1) holds. When is there a giant component?

Exercise 9.22 (Degree structure in stochastic blockmodels with unequal expected degrees) Let $n$ be even. Consider the stochastic blockmodel with 2 types and $n/2$ vertices of each of the types. Let $p_{ij} = a_1/n$ when $i, j$ have type 1, $p_{ij} = a_2/n$ when $i, j$ have type 2, and $p_{ij} = b/n$ when $i, j$ have different types. For $i \in \{1, 2\}$ and $k \in \mathbb{N}_0$, let $N_{i,k}(n)$ denote the number of vertices of type $i$ and degree $k$. Show that
\[
\frac{N_{i,k}(n)}{n} \xrightarrow{P} \mathbb{P}(\text{Poi}(\lambda_i) = k),
\]
(9.7.8)
where $\lambda_i = (a_i + b)/2$.

Exercise 9.23 (Community detection in stochastic blockmodels with unequal expected degrees) In the setting of Exercise 9.22, assume that $a_1 > a_2$. Consider the following greedy community detection algorithm: Let $\pi(i) = 1$ for the $n/2$ vertices of highest degree type 1, and $\pi(i) = 2$ for the remaining vertices (breaking ties randomly when necessary). Argue that this algorithm achieves the solvability condition in (9.3.2).

Exercise 9.24 (Parameter conditions for solvable stochastic blockmodels) Consider the stochastic blockmodel in the setting of Theorem 9.9, and assume that $(a - b)^2 > 2(a + b)$, so that the community detection problem is solvable. Show that $a - b > 2$ and thus also $a + b > 2$. Conclude that this model has a giant. Show that also the vertices of type 1 only (resulting in an Erdős-Rényi random graph of size $n/2$ and edge probability $a/n$) also has a giant. What are the conditions for the vertices of type 2 to have a giant?

Exercise 9.25 (Degree structure in degree-corrected stochastic blockmodels) Recall the definition of the degree-corrected stochastic blockmodel in (9.3.11) in Section 9.3.2, and assume that (9.3.1) holds. What is the asymptotic expected degree of a vertex $v$ of weight $x_v$ of this model? What are the restrictions on $(\kappa_{i,j})_{i,j \in S}$ such that the expected degree of vertex $v$ with weight $x_v$ is equal to $x_v(1 + o(1))$?

Exercise 9.26 (Equal average degrees in degree-corrected stochastic blockmodels) Recall the definition of the degree-corrected stochastic blockmodel in (9.3.11) in Section 9.3.2, and assume that (9.3.1) holds. Let $r \geq 2$ be arbitrary and assume
that \( \kappa_{i,j} = b \) for all \( i \neq j \), while \( \kappa_{i,i} = a \). Assume that \( \mu_i = 1/r \) for every \( i \in [r] \). Compute the asymptotic average degree of a vertex of type \( i \), and show that it is independent of \( i \).

**Exercise 9.27** (The giant in the degree-corrected stochastic blockmodels) Recall the definition of the degree-corrected stochastic blockmodel in Section 9.3.2, and assume that (9.3.1) holds. When is there a giant component?

**Exercise 9.28** (Degrees in configuration models with global communities) Recall the definition of the configuration models with global communities in Section 9.3.3, and assume that (9.3.18), (9.3.19) and (9.3.20) hold. What is the asymptotic expected degree of this model? When do all vertices have the same asymptotic expected degree?

**Exercise 9.29** (Local limit in configuration models with global communities) Recall the definition of the configuration models with global communities in Section 9.3.3, and assume that (9.3.18), (9.3.19) and (9.3.20) hold. What do you think that the local limit is of this model? [Note: No proof is expected, but a reasonable argument.]

**Exercise 9.30** (Giant in configuration models with global communities) Recall the definition of the configuration models with global communities in Section 9.3.3, and assume that (9.3.18), (9.3.19) and (9.3.20) hold. When do you think that there is there a giant component? [Note: No proof is expected, but a reasonable argument.]

**Exercise 9.31** (Degrees distribution in preferential attachment model with global communities) Show that \((p_k(\theta))_{k \geq m}\) in (9.3.28) is a probability distribution for all \( \theta \), i.e., show that \( \sum_{k \geq m} p_k(\theta) = 1 \).

**Exercise 9.32** (Global degrees distribution in preferential attachment model with global communities) In the preferential attachment models with global communities studied in Theorem 9.11, show that also the global degree distribution given by \( P_k = \frac{1}{n} \sum_{v \in [n]} 1_{\{D_v(n) = k\}} \) also converges almost surely.

**Exercise 9.33** (Global degrees distribution in preferential attachment model with global communities) In the preferential attachment models with global communities studied in Theorem 9.11, show that the global degree distribution has a power-law tail with exponent \( \tau = 1 + 1/\max_{s \in [r]} \theta_s^* \), provided that \( \mu_s > 0 \) for all \( s \in [r] \).

**Exercise 9.34** (Clustering in model with edges and triangles) Show that the global clustering coefficient in the model where each pair of vertices is independently connected with probability \( \lambda/n \), as for \( \text{ER}_n(\lambda/n) \), and each triple forms a triangle with probability \( \mu/n^2 \), independently for all triplets and independently of the status of the edges, converges to \( \mu/(\mu + \lambda^3) \).

**Exercise 9.35** (Local limit in inhomogeneous random graph with communities) Recall the definition of the inhomogeneous random graph with communities in
9.7 Exercises for Chapter 9

Section 9.4.1. What do you think that the local limit is of this model? [Note: No proof is expected, but a reasonable argument.]

Exercise 9.36 (Size-biased community size distribution in HCM) In the hierarchical configuration model introduced in Section 9.4.2, choose a vertex $o_n$ uniformly at random from $[n]$. Let $G_{o_n}$ be the community containing $o_n$. Show that (9.4.17) implies that $|V(G_{o_n})|$ converges in distribution, and identify its limiting distribution.

Exercise 9.37 (Local limit in hierarchical configuration model) Recall the definition of the hierarchical configuration model in Theorem 9.16. What do you think that the local limit is of this model? [Note: No proof is expected, but a reasonable argument.]

Exercise 9.38 (Law of large numbers for $|C_{\text{max}}|$ in hierarchical configuration model) Use Theorem 4.4 to prove the law of large numbers for the giant in the hierarchical configuration model in Theorem 9.16, and prove that $\zeta$ is given by (9.4.20).

Exercise 9.39 (Degree moments in scale-free percolation Deijfen et al. (2013)) Show that $E[D_x^p] < \infty$ when $p < \gamma$ and $E[D_x^p] = \infty$ when $p > \gamma$. In particular, the variance of the degrees is finite precisely when $\gamma > 2$.

Exercise 9.40 (Positive correlation between edge statuses in scale-free percolation) Show that, for scale-free percolation, and for all $x, y, z$ distinct and $\lambda > 0$,

$$P(\{x, y\} \text{ and } \{x, z\} \text{ occupied}) \geq P(\{x, y\} \text{ occupied})P(\{x, z\} \text{ occupied}),$$

(9.7.9)

the inequality being strict when $P(W_0 = 0) < 1$. In other words, the edge statuses are positively correlated.
Appendix A
Some facts about the metric space
structure of rooted graphs

Abstract
In this section, we highlight some properties and results about metric spaces, including separable metric spaces and Borel measures on them. These results are used throughout Chapters 1–3. We also present some of the missing details in the proof that the space of rooted graphs is a separable metric space, or a Polish space. Finally, we discuss what compact sets look like in this topology, and relate this to tightness criteria.

A.1 Metric spaces

In this section, we discuss metric spaces. We start by defining what a metric is:

Definition A.1 (Distance metric) Let $X$ be a space. A distance on $X$ is a function $d: X^2 \to [0, \infty)$ such that

(a) $0 \leq d(x, y) < \infty$ for all $x, y \in X$;
(b) $d(x, y) = d(y, x)$ for all $x, y \in X$;
(c) $d(x, z) \leq d(x, y) + d(y, z)$ for all $x, y, z \in X$.

Definition A.2 (Metric space) Let $X$ be a space and $d$ a metric on it. Then $(X, d)$ is called a metric space.

We next discuss some desirable properties of metric spaces:

Definition A.3 (Complete metric space) Let $(X, d)$ be a metric space. We say that $(X, d)$ is complete when every Cauchy sequence has a limit. Here, a Cauchy sequence is a sequence $(x_n)_{n \geq 1}$ with $x_n \in X$ such that for every $\varepsilon > 0$ there exists an $N = N(\varepsilon)$ such that $d(x_n, x_m) \leq \varepsilon$ for all $n, m \geq N$.

Definition A.4 (Separable metric space) Let $(X, d)$ be a metric space. We say that $(X, d)$ is separable if it contains a countable subset that is dense in $(X, d)$. This means that there exists a countable set $A \subseteq X$ such that, for every $x \in X$ there exists a sequence $(a_n)_{n \geq 1}$ with $a_n \in A$ such that $d(a_n, x) \to 0$.

Definition A.5 (Polish space) The metric space $(X, d)$ is called Polish when it is separable and complete.
A.2 Properties of the metric $d_{\mathcal{G}_*}$ on rooted graphs

For $(G,o) \in \mathcal{G}_*$, let

$$[G,o] = \{(G',o') : (G',o') \simeq (G,o)\} \quad \text{(A.2.1)}$$

denote the equivalence class in $\mathcal{G}_*$ corresponding to $(G,o)$. We further let

$$[\mathcal{G}_*] = \{(G,o) : (G,o) \in \mathcal{G}_*\} \quad \text{(A.2.2)}$$

denote the set of equivalence class in $\mathcal{G}_*$. This will be the set on which the distance $d_{\mathcal{G}_*}$ acts.

In this section, we prove that $([\mathcal{G}_*], d_{\mathcal{G}_*})$ is a Polish space:

**Theorem A.6** (Rooted graphs form a Polish space) $d_{\mathcal{G}_*}$ is a well-defined metric on $[\mathcal{G}_*]$. Further, the metric space $([\mathcal{G}_*], d_{\mathcal{G}_*})$ is a Polish space.

The proof of Theorem A.6 will be divided into several steps. We start in Proposition A.8 by showing that $d_{\mathcal{G}_*}$ is an ultrametric, which is a slightly stronger property than being a metric, and which also implies that $([\mathcal{G}_*], d_{\mathcal{G}_*})$ is a metric space. In Proposition A.10, we show that the metric space $([\mathcal{G}_*], d_{\mathcal{G}_*})$ is complete, and in Proposition A.12, we show that it is separable. After that, we complete the proof of Theorem A.6.

In the remainder of this section, we will often work with $r$-neighborhoods $B_r^{(o)}(o)$ of $o$ in $G$. We emphasize that we consider $B_r^{(o)}(o)$ to be a rooted graph, with root $o$ (recall (2.1.1)).

### A.2.1 Ultrametric property of $d_{\mathcal{G}_*}$

In this section, we prove that $d_{\mathcal{G}_*} : \mathcal{G}_* \times \mathcal{G}_* \to [0,1]$ is an ultrametric. One of the problems that we have to resolve is that the space of rooted graphs is only defined up to isomorphisms, which means that we have to make sure that $d_{\mathcal{G}_*}((G_1,o_1),(G_2,o_2))$ is independent of the exact representative we choose in the equivalence classes of $(G_1,o_1)$ and $(G_2,o_2)$. That is the content of the following proposition:

**Proposition A.7** ($d_{\mathcal{G}_*}$ is well-defined on $[\mathcal{G}_*]$) The equality $d_{\mathcal{G}_*}((\hat{G}_1,\hat{o}_1),(\hat{G}_2,\hat{o}_2)) = d_{\mathcal{G}_*}((G_1,o_1),(G_2,o_2))$ holds whenever $(\hat{G}_1,\hat{o}_1) \simeq (G_1,o_1)$ and $(\hat{G}_2,\hat{o}_2) \simeq (G_2,o_2)$. Consequently, $d_{\mathcal{G}_*} : [\mathcal{G}_*] \times [\mathcal{G}_*]$ is well-defined.

**Proposition A.8** (Ultrametricity) The map $d_{\mathcal{G}_*} : \mathcal{G}_* \times \mathcal{G}_* \to [0,1]$ is an ultrametric, meaning that

(a) $d_{\mathcal{G}_*}((G_1,o_1),(G_2,o_2)) = 0$ precisely when $(G_1,o_1) \simeq (G_2,o_2)$;
(b) $d_{\mathcal{G}_*}((G_1,o_1),(G_2,o_2)) = d_{\mathcal{G}_*}((G_2,o_2),(G_1,o_1))$ for all $(G_1,o_1),(G_2,o_2) \in \mathcal{G}_*$;
(c) $d_{\mathcal{G}_*}((G_1,o_1),(G_3,o_3)) \leq \max\{d_{\mathcal{G}_*}((G_1,o_1),(G_2,o_2)),d_{\mathcal{G}_*}((G_2,o_2),(G_3,o_3))\}$ for all $(G_1,o_1),(G_2,o_2),(G_3,o_3) \in \mathcal{G}_*$.

Before giving the proof of Propositions A.7–A.8, we state and prove an important ingredient in them:
Lemma A.9 (Local neighborhoods determine the graph) Let \((G_1, o_1)\) and \((G_2, o_2)\) be two connected locally finite rooted graphs such that \(B_r^{G_1}(o_1) \simeq B_r^{G_2}(o_2)\) for all \(r \geq 0\). Then \((G_1, o_1) \simeq (G_2, o_2)\).

Proof. We use a subsequence argument. Fix \(r \geq 0\), and consider the isomorphism \(\phi_r : B_r^{G_1}(o_1) \to B_r^{G_2}(o_2)\), which exists by assumption. Extend \(\phi_r\) to \((G_1, o_1)\) by defining

\[
\psi_r(v) = \begin{cases} 
\phi_r(v) & \text{for } v \in V(B_r^{G_1}(o_1)); \\
o_2 & \text{otherwise.}
\end{cases}
\]  

(A.2.3)

Our aim is to use \((\psi_r)_{r \geq 0}\) to construct an isomorphism between \((G_1, o_1)\) and \((G_2, o_2)\).

Abbreviate \(V_r^{G_1} = V(B_r^{G_1}(o_1))\). Let \(\psi_r|_{V_r^{G_1}}\) be the restriction of \(\psi_r\) to \(V_r^{G_1} = \{o_1\}\). Then we know that \(\psi_r(v) = o_2\) for every \(v \in V_r^{G_1}\) and \(r \geq 0\). We next let \(\psi_r|_{V_r^{G_1}}\) be the restriction of \(\psi_r\) to \(V_r^{G_1}\). Then, \(\psi_r|_{V_r^{G_1}}\) is an isomorphism between \(B_r^{G_1}(o_1)\) and \(B_r^{G_2}(o_2)\) for every \(r\). Since there are only finitely many such isomorphisms, the same isomorphism \(\phi_1'\) needs to be repeated infinitely many times in the sequence \((\psi_r|_{V_r^{G_1}})_{r \geq 1}\). Let \(N_1\) denote the values of \(r\) for which

\[
\psi_r|_{V_r^{G_1}} = \phi_1' \quad \forall r \in N_1.
\]  

(A.2.4)

Now we repeat this argument to \(k = 2\). Let \(\psi_r|_{V_r^{G_1}}\) be the restriction of \(\psi_r\) to \(V_r^{G_1}\). Again, \(\psi_r|_{V_r^{G_1}}\) is an isomorphism between \(B_r^{G_1}(o_1)\) and \(B_r^{G_2}(o_2)\) for every \(r\). Since there are again only finitely many such isomorphisms, the same isomorphism \(\phi_2'\) needs to be repeated infinitely many times in the sequence \((\psi_r|_{V_r^{G_1}})_{r \in N_1}\). Let \(N_2\) denote the values of \(r \in N_1\) for which

\[
\psi_r|_{V_r^{G_1}} = \phi_2' \quad \forall r \in N_2.
\]  

(A.2.5)

We now generalize this argument to general \(k \geq 2\). Let \(\psi_r|_{V_r^{G_1}}\) be the restriction of \(\psi_r\) to \(V_r^{G_1}\). Again, \(\psi_r|_{V_r^{G_1}}\) is an isomorphism between \(B_r^{G_1}(o_1)\) and \(B_r^{G_2}(o_2)\) for every \(r\). Since there are again only finitely many such isomorphisms, the same isomorphism \(\phi_k'\) needs to be repeated infinitely many times in the sequence \((\psi_r|_{V_r^{G_1}})_{r \in N_k}\). Let \(N_k\) denote the values of \(r \in N_{k-1}\) for which

\[
\psi_r|_{V_r^{G_1}} = \phi_k' \quad \forall r \in N_k.
\]  

(A.2.6)

Then, we see that \(N_k\) is a decreasing infinite set.

Let us define \(\psi'_1\) to be the first element of the sequence \((\psi_r)_{r \in N_k}\). Then, it follows that \(\psi'_l(v) = \phi_k'(v)\) for all \(l \geq k\) and all \(v \in V_r^{G_1}\). Denote \(U_0 = \{o_1\}\) and \(U_k = V_r^{G_1} \setminus V_{r-1}^{G_1}\). Since we assume that \(V(G_1)\) is connected, we have that \(\cup_{k \geq 0} U_k = V(G_1)\).

It follows that the functions \((\psi'_l)_{l \geq 1}\) converge pointwise to

\[
\psi(v) = \psi'_\infty(v) = \sum_{k \geq 1} \phi_k'(v) 1_{\{v \in U_k\}}.
\]  

(A.2.7)
The point is that (phisms. be connected locally finite rooted graphs that are compatible, meaning that $G$ ingredient in it: $\phi_k: U_k \rightarrow \phi_k(U_k)$ is bijective. Further, let $u, v \in V(G_1)$. Denote $k = \max\{d_{G_1}(o_1, u), d_{G_1}(o_1, v)\}$. Then $u, v \in V_k$. Because $\phi'_k(u), \phi'_k(v) \in V(B_k^{(G_1)}(o_1))$ and $\{\phi'_k(u), \phi'_k(v)\} \in E(B_k^{(G_1)}(o_1))$ precisely when $\{u, v\} \in E(B_k^{(G_1)}(o_1))$. Since $\psi = \phi_k$ on $V_k^{(G_1)}$, it then also follows that $\{\psi(u), \psi(v)\} \in E(B_k^{(G_1)}(o_1))$ precisely when $\{u, v\} \in E(B_k^{(G_1)}(o_1))$, as required. Finally, $\psi(o_1) = \phi_k(o_1)$ and $\phi_k(o_1) = o_2$ for every $k \geq 0$. This completes the proof. □

Proof of Proposition A.7. We note that if $(\hat{G}_1, \hat{o}_1) \simeq (G_1, o_1)$ and $(\hat{G}_2, \hat{o}_2) \simeq (G_2, o_2)$, then $B_v^{(\hat{G}_1)}(o_1) \simeq B_v^{(\hat{G}_2)}(o_2)$ if and only if $B_v^{(G_1)}(o_1) \simeq B_v^{(G_2)}(o_2)$. Therefore, $d_{\hat{G}_r}(G_1, o_1, (G_2, o_2))$ is independent of the exact choice of the representative in the equivalence class of $(G_1, o_1)$ and $(G_2, o_2)$, so that $d_{\hat{G}_r}(G_1, o_1, (G_2, o_2))$ is constant on such equivalence classes. This makes the metric $d_{\hat{G}_r}([G_1, o_1], [G_2, o_2])$ well-defined for $[G_1, o_1], [G_2, o_2] \in [\mathcal{G}].$

Proof of Proposition A.8. (a) Assume that $d_{\hat{G}_r}((G_1, o_1), (G_2, o_2)) = 0$. Then we have $B_v^{(G_1)}(o_1) \simeq B_v^{(G_2)}(o_2)$ for all $r \geq 0$, so that, by Lemma A.9, also $(G_1, o_1) \simeq (G_2, o_2)$, as required.

The proof of (b) is trivial and omitted.

For (c), let

$$r_{ij} = \sup\{r: B_v^{(G_i)}(o_i) \simeq B_v^{(G_j)}(o_j)\}. \quad (A.2.8)$$

Then $B_v^{(G_i)}(o_i) \simeq B_v^{(G_j)}(o_j)$ for all $r \leq \min\{r_{13}, r_{23}\}$ and $B_v^{(G_i)}(o_i) \simeq B_v^{(G_j)}(o_j)$ for all $r \leq \min\{r_{13}, r_{23}\}$. We conclude that $B_v^{(G_i)}(o_i) \simeq B_v^{(G_j)}(o_j)$ for all $r \leq \min\{r_{13}, r_{23}\}$, so that $r_{12} \geq \min\{r_{13}, r_{23}\}$. This implies that

$$1/(r_{13} + 1) \leq \max\{1/(r_{13} + 1), 1/(r_{23} + 1)\},$$

which implies the claim (recall (2.1.2)). □

A.2.2 Completeness of $([\mathcal{G}], d_{\hat{G}_r})$

In this section, we prove that $([\mathcal{G}], d_{\hat{G}_r})$ is complete:

**Proposition A.10** (Completeness) The metric space $([\mathcal{G}], d_{\hat{G}_r})$ is complete.

Before giving the proof of Proposition A.10, we state and prove an important ingredient in it:

**Lemma A.11** (Compatible rooted graph sequences have a limit) Let $\{(G_r, o_r)\}_{r \geq 0}$ be connected locally finite rooted graphs that are compatible, meaning that $B_v^{(G_r)}(o_r) \simeq B_v^{(G_r)}(o_r)$ for all $r \leq s$. Then there exists a connected locally finite rooted graph $(G, o)$ such that $(G_r, o_r) \simeq B_v^{(G)}(o)$. Moreover, $(G, o)$ is unique up to isomorphisms.

**Proof** The point is that $(G_r, o_r)$ may not live on the same vertex sets. Therefore, we first create a version $(G'_r, o'_r)$ that does live on the same vertex set. For this, we will first construct isomorphic copies of $(G_r, o_r)$ on a common node set, in
a compatible way. To do this, denote $V_r = \{v \in \mathbb{N} : r \leq N_r\}$, where $N_r = |V(B_r(o_r))|$. We define a sequence of bijections $\phi_r : V(B_r(o_r)) \to V_r$ recursively as follows. Let $\phi_0$ be the unique isomorphism from $V(B_0(o_0)) = \{o_0\}$ to $V_0 = \{1\}$.

Let $\psi_r$ be an isomorphism between $(G_{r-1}, o_{r-1})$ and $(B_{r-1}(o_r), o_r)$. Then we define $(G_r, o_r) = (\psi_r^{-1}(B_r(o_r)), o_r)$ for every $r > 0$, by construction. Further, $\phi_r$ is a bijection, so that $B_r^{(G_r)}(o_r') \simeq B_r^{(G_r)}(o_r) = (G_r, o_r)$.

Now we are ready to define $(G, o)$. We define the root as $o = 1$, the vertex set of $G$ by $V(G) = \bigcup_{r \geq 1} V(G_r)$, and the edge set $E(G) = \bigcup_{r \geq 1} E(G_r)$. Then it follows that $B_r^{(G)}(o) = B_r^{(G)}(o') \simeq B_r^{(G_r)}(o_r) = (G_r, o_r)$, as required. Further, $(G, o)$ is locally finite and connected, since $(G_r, o_r)$ is for every $r \geq 0$. Finally, to verify uniqueness, note that if $(G', o')$ also satisfies that $B_r^{(G')} (o') \simeq B_r^{(G_r)} (o_r) = (G_r, o_r)$ for every $r \geq 0$, then $B_r^{(G')} (o') \simeq B_r^{(G)} (o)$ for every $r \geq 0$, so that $(G, o) \simeq (G', o')$, as required.

Proof of Proposition A.10. To verify completeness, fix a Cauchy sequence $\{(G_n, o_n)\}_{n \geq 1}$ with representative rooted graphs $\{(G_n, o_n)\}_{n \geq 1}$. Then, for every $\varepsilon > 0$, there exists an $N = N(\varepsilon) \geq 0$ such that, for every $n, m \geq N$,

$$d_{s_*}([G_n, o_n], [G_m, o_m]) \leq \varepsilon.$$  \hfill (A.2.10)

We note that, by Proposition A.7,

$$d_{s_*}([G_n, o_n], [G_m, o_m]) = d_{s_*}((G_n, o_n), (G_m, o_m)),$$  \hfill (A.2.11)

so from now on, we can work with the representatives instead. Since

$$d_{s_*}((G_n, o_n), (G_m, o_m)) \leq \varepsilon,$$  \hfill (A.2.12)

we obtain that for all $r \leq 1/\varepsilon - 1$ and $n, m \geq N$, $B_r^{(G_n)}(o_n) \simeq B_r^{(G_m)}(o_m)$.

Equivalently, this implies that for every $r \geq 1$, there exists an $n_r$ such that, for all

$$B_r^{(G_n)}(o_n) \simeq B_r^{(G_{n_r})}(o_{n_r}).$$  \hfill (A.2.13)

Clearly, we may select $n_r$ such that $n_r$ is strictly increasing. Define $(G'_r, o'_r) = B_r^{(G_{n_r})}(o_{n_r})$. Then $\{(G'_r, o'_r)\}_{r \geq 0}$ forms a compatible sequence. By Lemma A.11, there exists a locally finite rooted graph $(G, o)$ such that $B_r^{(G)}(o) \simeq (G'_r, o'_r)$.

But then also

$$B_r^{(G_n)}(o_n) \simeq B_r^{(G_{n_r})}(o_{n_r}) = (G'_r(o'_r), o'_r).$$  \hfill (A.2.14)

This, in turn, implies that, for all $n \geq n_r$,

$$d_{s_*}((G, o), (G_n, o_n)) \leq 1/(r + 1).$$  \hfill (A.2.15)
A.2.3 Separability of \( ([\mathcal{G}], d_{\mathcal{G}}) \)

In this section, we prove that \( ([\mathcal{G}], d_{\mathcal{G}}) \) is separable:

**Proposition A.12** (Separability) The metric space \( ([\mathcal{G}], d_{\mathcal{G}}) \) is separable.

**Proof** We need to show that there exists a countable dense subset in \( ([\mathcal{G}], d_{\mathcal{G}}) \). Consider the set of all finite rooted graphs, which is certainly countable. Also, \( B_r^{(\mathcal{G})}(o) \) is a finite rooted graph for all \( r \geq 0 \). Finally, \( d_{\mathcal{G}}(B_r^{(\mathcal{G})}(o), (G, o)) \leq 1/(r + 1) \), so that \( B_r^{(\mathcal{G})}(o) \) converges to \((G, o)\) when \( r \to \infty \). Thus, the space of finite rooted graphs is dense and countable. This completes the proof that \( ([\mathcal{G}], d_{\mathcal{G}}) \) is separable. \( \square \)

A.2.4 \( ([\mathcal{G}], d_{\mathcal{G}}) \) is Polish: Proof of Theorem A.6

In this section, we use the above results to complete the proof of Theorem A.6:

**Proof of Theorem A.6.** The function \( d_{\mathcal{G}} \) is well-defined on \( [\mathcal{G}] \times [\mathcal{G}] \) by Proposition A.7. Proposition A.8 implies that \( d_{\mathcal{G}} \) is an (ultra)metric on \( [\mathcal{G}] \). Finally, Proposition A.10 proves that \( ([\mathcal{G}], d_{\mathcal{G}}) \) is complete, while Proposition A.12 proves that \( ([\mathcal{G}], d_{\mathcal{G}}) \) is separable. Thus, \( ([\mathcal{G}], d_{\mathcal{G}}) \) is a Polish space. \( \square \)

A.2.5 The laws of neighborhoods determine distributions on \( \mathcal{G} \)

In this section, we show that the laws of neighborhoods determine distributions on \( \mathcal{G} \), as was crucially used in the proof of Theorem 2.6:

**Proposition A.13** (Laws of neighborhoods determine distributions) Let \( \mu \) and \( \mu' \) be two distributions on \( \mathcal{G} \) such that \( \mu(B_r^{(\mathcal{G})}(o) \simeq H_*) = \mu'(B_r^{(\mathcal{G})}(o) \simeq H_*) \). Then \( \mu = \mu' \).

**Proof** The measures \( \mu \) and \( \mu' \) satisfy \( \mu = \mu' \) precisely when \( \mu(\mathcal{H}_*) = \mu'(\mathcal{H}_*) \) for every measurable \( \mathcal{H}_* \subseteq \mathcal{G} \). Fix \( \mathcal{H}_* \subseteq \mathcal{G} \). For \( r \geq 0 \), denote

\[
\mathcal{H}_*(r) = \{(G, o) : \exists (G', o') \in \mathcal{H}_* \text{ such that } B_r^{(\mathcal{G})}(o) \simeq B_r^{(\mathcal{G})}(o')\}. \tag{A.2.16}
\]

Thus, \( \mathcal{H}_*(r) \) contains those rooted graphs whose \( r \)-neighborhood is the same as that of a rooted graph in \( \mathcal{H}_* \). Clearly, \( \mathcal{H}_*(r) \searrow \mathcal{H}_* \) as \( r \to \infty \). Therefore, also \( \mu(\mathcal{H}_*(r)) \searrow \mu(\mathcal{H}_*) \) and \( \mu'(\mathcal{H}_*(r)) \searrow \mu'(\mathcal{H}_*) \).

Finally, note that \((G, o) \in \mathcal{H}_*(r)\) if and only if \( B_r^{(\mathcal{G})}(o) \in \mathcal{H}_*(r) \). Thus,

\[
\mu(\mathcal{H}_*(r)) = \sum_{H_* \in \mathcal{H}_*(r)} \mu(B_r^{(\mathcal{G})}(o) \simeq H_*). \tag{A.2.17}
\]

Since \( \mu(B_r^{(\mathcal{G})}(o) \simeq H_*) = \mu'(B_r^{(\mathcal{G})}(o) \simeq H_*) \), we conclude that

\[
\mu(\mathcal{H}_*(r)) = \sum_{H_* \in \mathcal{H}_*(r)} \mu(B_r^{(\mathcal{G})}(o) \simeq H_*) = \sum_{H_* \in \mathcal{H}_*(r)} \mu'(B_r^{(\mathcal{G})}(o) \simeq H_*) = \mu'(\mathcal{H}_*(r)), \tag{A.2.18}
\]
we conclude that $\mu(\mathcal{H}) = \mu'(\mathcal{H})$, as required.

A.2.6 Compact sets in $([\mathcal{G}], d)$ and tightness

In Theorem 2.7, we have given a tightness criterion for $([\mathcal{G}], d)$. In this section, we investigate tightness further, by considering compact sets in the metric space $([\mathcal{G}], d)$. First, we recall the definition of compactness:

**Definition A.14 (Compact sets on general metric spaces)** Let $X$ be a general metric space. A set $K$ is compact provided that every collection of open sets covering $K$ has a finite subset.

Next, we recall the definition of tightness:

**Definition A.15 (Tightness on general metric spaces)** A sequence of random variable $(X_n)_{n \geq 1}$ living on a general metric space $X$ is tight when, for every $\varepsilon > 0$, there exists a compact set $K = \mathcal{K}_\varepsilon$ such that

$$\limsup_{n \to \infty} P(X_n \in \mathcal{K}_\varepsilon^c) \leq \varepsilon. \quad (A.2.19)$$

Thus, the notion of tightness is intimately related to compact sets. For real-valued random variables, compact sets can be taken as $K = [-K, K]$. For $([\mathcal{G}], d)$, we first need to investigate what compact sets look like. This is the content of the next theorem:

**Theorem A.16 (Compact sets in $([\mathcal{G}], d)$)** For $(G, o) \in \mathcal{G}$ and $r \geq 1$, define

$$\Delta_r(G, o) = \max\{d_v^{(G)}: v \in V(B_r^{(G)}(o))\}, \quad (A.2.20)$$

where $d_v^{(G)}$ denotes the degree of $v \in V(G)$. Then, a closed family of rooted graphs $\mathcal{K} \subseteq \mathcal{G}$ is tight if and only if

$$\sup_{(G, o) \in \mathcal{K}} \Delta_r(G, o) < \infty \quad \text{for all } r \geq 1. \quad (A.2.21)$$

Theorem 2.7 thus states that the uniform integrability of $(d_{n_0}^{(G)})_{n \geq 1}$ implies (A.2.21), but not vice versa. It is highly interesting that the compact sets in $([\mathcal{G}], d)$ can be described so explicitly.

**Proof** Recall from (Rudin, 1991, Theorem A.4) that a closed set $\mathcal{K}$ is compact when it is totally bounded, meaning that for every $\varepsilon > 0$, the set $\mathcal{K}$ can be covered by finitely many balls of radius $\varepsilon$. As a result, for every $r \geq 1$, there must be graphs $(F_1, o), \ldots, (F_{\ell}, o)$ such that $\mathcal{K}$ is covered by the finitely many open sets

$$\{(G, o): B_r^{(G)}(o) \simeq B_r^{(F_i)}(o)\}. \quad (A.2.22)$$

Equivalently, every $(G, o) \in \mathcal{K}$ satisfies $B_r^{(G)}(o) \simeq B_r^{(F_i)}(o)$ for some $i \in [\ell]$. In turn, this is equivalent to the statement that the set

$$\mathcal{A}_r = \{B_r^{(G)}(o): (G, o) \in \mathcal{K}\} \quad (A.2.23)$$

is finite for every $r \geq 1$. 
We finally prove that $\mathcal{A}_r$ is finite for every $r \geq 1$ precisely when (A.2.21) holds. Denote $\Delta_r = \sup_{(G,o) \in \mathcal{K}} \Delta_r(G,o)$. If $\Delta_r$ is finite for every $r \geq 1$, then, because every $(G,o) \in \mathcal{K}$ is connected, the graphs $B_r^{(G)}(o)$ can have at most

$$|V(B_r^{(G)}(o))| \leq 1 + \Delta_r + \cdots + \Delta_r^r$$

many vertices, so that $\mathcal{A}_r$ is finite. On the other hand, when $\Delta_r = \infty$, then $\mathcal{K}$ contains a sequence of rooted graphs $(G_i,o_i)$ such that $\Delta_r(G_i,o_i) \to \infty$, so that also $|V(B_r^{(G_i)}(o_i))| \to \infty$. Since rooted graphs with different numbers of vertices are non-isomorphic (recall Exercise 2.1), this shows that $\mathcal{A}_r$ is infinite. \hfill \Box

A.3 Notes and Discussion

Notes on Section A.1


Notes on Section A.2

This section is based to a large extent on (Leskelä, 2019, Appendix B), some parts of the presented material is copied almost verbatim from there. I am grateful to Lasse Leskelä for making me aware of the subtleties of the proof, as well as sharing his preliminary version of these notes.


Chatterjee, S., and Durrett, R. 2009. Contact processes on random graphs with power law degree distributions have critical value 0. *Ann. Probab.*, 37(6), 2332–2356.


Csanádi, G. 2006. Dynamics of Citation Networks. 4131, 698–709.


References


References


References


Albert-Barabási model, 188
Box, George E.P., 344
“All models are wrong but some are useful”, 344
Branching process
  delayed, 272
  infinite mean, 257
  marked, 106
  multitype, 93
  two-stage, 106
Central limit theorem, 195
Chung-Lu model
  attack vulnerability, 112
  fluctuations of typical distance, 238
Citation networks, 340
  exponential growth, 340
  power-law degree sequences, 340
Clique, 216
Clustering coefficient, 52
Collaboration graph
  random intersection graph, 358
Community structure network, 339
  macroscopic communities, 339
  microscopic communities, 339
Complexity, 110
Configuration model, 121, 245
  average distance, 246
  clustering, 356, 357
  connectivity, 145
  core, 253
  diameter, 263
  diameter core, 253
  directed, 347
  distance, 246
  forward degree distribution, 123
  household structure, 356
  near-critical behavior, 151
  path counting, 248
  phase transition, 127
Connected component
  complexity, 110
Connectivity
  configuration model, 145
De Finetti’s Theorem, 160
Death process, 133
Directed network, 339
  arc, 340
Distribution
  Bernoulli, 160
  Mixed Poisson, 88
Erdős-Rényi random graph
  diameter, 264
Exchangeable random variables, 160
Helly’s Theorem, 162
Hopcount, 6
Household model, 356
Inhomogeneous random graph, 83
  diameter, 236
  fluctuations of typical distance, 238
  graphical, 85
  irreducible, 85
  small world, 202
  sparsity, 85
  typical graph distance, 201, 237
  Ultra-small world, 204
Inhomogeneous random graphs
  dual kernel, 236
  duality principle, 236
Internet, 6
  hopcount, 6
Internet Movie Data base
  random intersection graph, 358
Maclaurin inequality, 249
Pólya urn scheme
  limit theorem, 162, 163
Path
  self-avoiding, 206
Path counting
  configuration model, 248
Paths, 206
  self-avoiding, 206
percolation
  long-range, 365
Perron-Frobenius theorem, 94
Phase transition
  configuration model, 127
Preferential attachment model, 159, 279
  diameter, 289
  typical distances, 289
Random graph
  typical graph distance, 201
Random intersection graph, 358
Random intersection graphs, 358
Random variables
  exchangeable, 160
Real-world network, 3
Real-world networks
  - citation networks, 340
  - communities, 339
  - directed, 339
scale-free percolation network, 363
Scale-free trees, 280
Self-avoiding path, 206
Size-biased distribution, 106
Small world, 202
spatial preferential attachment models, 368
Stochastic domination, 106
Theorem
  - De Finetti, 160
  - Helly, 162
  - Perron-Frobenius, 94
Tree
  - height, 280
Two-regular graph
  - diameter, 277
  - longest cycle, 278
Ultra-small world, 202, 204
Uniform recursive trees, 331
Universality
  - typical distances, 291
### Todo List

| TO DO 1.1: Refer to recent discussion on power-laws in real-world networks. | 6 |
| To be added! | 139 |
| Integrate Proposition 5.20 more fluently in the text! | 192 |
| Incorporate Proposition 5.23 more fluently in the text! | 202 |
| TO DO 5.1: Add local weak convergence for BPAM | 206 |
| TO DO 5.2: Extend the discussion of the phase transition for BPAM | 208 |
| TO DO 5.3: Add further results! | 209 |
| TO DO 6.1: Add case of distances in real-world network... | 219 |
| TO DO 9.1: Add a discussion on terrorist networks | 361 |
| Explain the $1 + 2\rho$! | 405 |
| Update from here! | 413 |
| TO DO 9.14: Introduce the spatial configuration model and results of Deijfen et al. | 422 |