

RANDOM GRAPHS AND COMPLEX NETWORKS

Volume 2

Remco van der Hofstad

February 29, 2020

Aan Mad, Max en Lars
het licht in mijn leven

Ter nagedachtenis aan mijn ouders
die me altijd aangemoedigd hebben

CONTENTS

<i>List of illustrations</i>	ix
<i>List of tables</i>	xi
<i>Preface</i>	xiii
<i>Course outline</i>	xvii

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	17
1.3.4 Switching algorithms for uniform random graphs	22
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 weak convergence of random graphs	37
2.1 The metric space of rooted graphs	38
2.2 Local weak convergence of graphs	39
2.3 Local weak convergence of random graphs	43
2.3.1 Definition of local weak convergence of random graphs	43
2.3.2 Convenient criterion for local weak convergence of random graphs	45
2.3.3 Local weak convergence and completeness of the limit	46
2.3.4 Local weak convergence of Erdős-Rényi random graphs	47
2.4 Consequences of local weak convergence: local functionals	51
2.4.1 Local weak convergence and convergence of neighborhoods	51
2.4.2 Local weak convergence and clustering coefficients	52
2.4.3 Neighborhoods of edges and degree-degree dependencies	54
2.4.4 PageRank and local weak convergence	57
2.5 The giant component is almost local	59
2.5.1 The giant in Erdős-Rényi random graphs	64
2.5.2 Outline of the remainder of this book	70
2.6 Notes and discussion	71
2.7 Exercises for Chapter 2	72

Part II Connected components in random graphs 77

3	Phase transition in general inhomogeneous random graphs	81
3.1	Motivation	81
3.2	Definition of the model	83
3.2.1	Examples of inhomogeneous random graphs	86
3.3	Degree sequence of inhomogeneous random graphs	88
3.3.1	Degree sequence of finitely-types case	89
3.3.2	Finite-type approximations of bounded kernels	89
3.3.3	Degree sequences of general inhomogeneous random graphs	91
3.4	Multitype branching processes	93
3.4.1	Multitype branching processes with finitely many types	93
3.4.2	Survival vs. extinction of multitype branching processes	94
3.4.3	Poisson multitype branching processes	95
3.5	Local weak convergence for inhomogeneous random graphs	100
3.5.1	Local weak convergence: finitely many types	100
3.5.2	Local weak convergence: infinitely many types	104
3.5.3	Comparison to branching processes	105
3.5.4	Local weak convergence to unimodular trees for $\text{GRG}_n(\mathbf{w})$	108
3.6	The phase transition for inhomogeneous random graphs	109
3.7	Related results for inhomogeneous random graphs	113
3.8	Notes and discussion	116
3.9	Exercises for Chapter 3	117
4	The phase transition in the configuration model	121
4.1	Local weak convergence to unimodular trees for $\text{CM}_n(\mathbf{d})$	122
4.2	Phase transition in the configuration model	127
4.2.1	Finding the largest component	131
4.2.2	Analysis of the algorithm for $\text{CM}_n(\mathbf{d})$	133
4.2.3	Proof of Theorem 4.4	138
4.2.4	The giant component of related random graphs	142
4.3	Connectivity of $\text{CM}_n(\mathbf{d})$	144
4.4	Related results for the configuration model	150
4.5	Notes and discussion	153
4.6	Exercises for Chapter 4	154
5	Connected components in preferential attachment models	159
5.1	Exchangeable random variables and Pólya urn schemes	159
5.2	Local weak convergence of preferential attachment models	168
5.2.1	Local weak convergence of PAMs with fixed number of edges	168
5.2.2	Local weak convergence Bernoulli preferential attachment	186
5.3	Connectivity of preferential attachment models	187
5.3.1	Giant component for Bernoulli preferential attachment	188
5.4	Further results for preferential attachment models	190
5.5	Notes and discussion	190
5.6	Exercises for Chapter 5	191
	Part III Small-world properties of random graphs	195
6	Small-world phenomena in inhomogeneous random graphs	199

6.1	Small-world effect in inhomogeneous random graphs	199
6.2	Lower bounds on typical distances in IRGs	203
6.2.1	Logarithmic lower bound distances for finite-variance degrees	203
6.2.2	log log lower bound on distances for infinite-variance degrees	208
6.3	The log log upper bound	213
6.4	Path counting and log upper bound for finite-variance weights	218
6.4.1	Path-counting techniques	219
6.4.2	Logarithmic distance bounds for finite-variance weights	225
6.4.3	Distances for $\text{IRG}_n(\kappa)$: Proofs of Theorems 3.16 and 6.1(ii)	231
6.5	Related results on distances for inhomogeneous random graphs	234
6.5.1	The diameter in inhomogeneous random graphs	234
6.5.2	Distance fluctuations for $\text{GRG}_n(\mathbf{w})$ with finite-variance degrees	235
6.6	Notes and discussion	236
6.7	Exercises for Chapter 6	237
7	Small-world phenomena in the configuration model	241
7.1	The small-world phenomenon in $\text{CM}_n(\mathbf{d})$	242
7.2	Proofs of small-world results $\text{CM}_n(\mathbf{d})$	243
7.2.1	Branching process approximation	243
7.2.2	Path-counting techniques	244
7.2.3	A log log upper bound on the diameter core $\tau \in (2, 3)$	249
7.3	Branching processes with infinite mean	253
7.4	The diameter of the configuration model	259
7.4.1	The diameter of the configuration model: logarithmic case	259
7.4.2	Diameter of the configuration model for $\tau \in (2, 3)$: log log case	261
7.5	Related results for the configuration model	266
7.5.1	Distances for infinite-mean degrees	266
7.5.2	Fluctuation of distances for finite-variance degrees	268
7.5.3	Fluctuation of distances for infinite-variance degrees	269
7.6	Notes and discussion	271
7.7	Exercises for Chapter 7	273
8	Small-world phenomena in preferential attachment models	275
8.1	Logarithmic distances in preferential attachment trees	276
8.2	Small-world effect in preferential attachment models	285
8.3	Path counting in preferential attachment models	287
8.4	Small-world effect in PA models: lower bounds	294
8.4.1	Logarithmic lower bounds on distances for $\delta > 0$	294
8.4.2	Lower bounds on distances for $\delta = 0$ and $m \geq 2$	296
8.4.3	Typical distance in PA-models: log log-lower bound	297
8.5	Small-world effect in PA models: upper bounds	303
8.5.1	Logarithmic upper bounds for $\delta > 0$	303
8.5.2	The diameter of the core	304
8.6	Diameters in preferential attachment models	308
8.7	Further results on distances in preferential attachment models	309
8.8	Notes and discussion	309
8.9	Exercises for Chapter 8	309

Part IV	Related models and problems	311
9	Related Models	315
9.1	Directed random graphs	315
9.1.1	Directed inhomogeneous random graphs	316
9.1.2	The directed configuration model	318
9.1.3	Directed preferential attachment models	320
9.2	Random graphs with community structure	321
9.2.1	Stochastic block model	321
9.2.2	Inhomogeneous random graphs with communities	321
9.2.3	Configuration models with community structure	326
9.2.4	Configuration model with clustering	327
9.2.5	Random intersection graphs	328
9.2.6	Exponential random graphs and maximal entropy	329
9.3	Spatial random graphs	329
9.3.1	Small-world model	329
9.3.2	Geometric inhomogeneous random graphs	331
9.3.3	Scale-free percolation	331
9.3.4	Hyperbolic random graphs	336
9.3.5	Spatial configuration models	338
9.3.6	Spatial preferential attachment models	338
9.4	Notes and discussion	340
9.5	Exercises for Chapter 9	341
Appendix	Some facts about metric spaces	343
	<i>References</i>	349
	<i>Index</i>	357
	<i>Todo list</i>	361

ILLUSTRATIONS

1.1	Degree sequence of the Autonomous Systems graph in 2014	6
1.2	log-log plot of the degree sequence in the Internet Movie Data base	7
1.3	The probability mass functions of the in- and out-degree sequence in the WWW from various data sets	7
1.4	AS and Internet hopcount	8
1.5	Typical distances in the Internet Movie Data base	8
1.6	Forward and backward switchings	25
6.1	A 12-step self-avoiding path connecting vertices i and j .	204
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.	209
6.3	An example of a pair of paths (π, ρ) and its corresponding shape.	222
7.1	Number of AS traversed in hopcount data (blue) compared to the model (purple) with $\tau = 2.25, n = 10, 940$.	241
7.2	Empirical probability mass function of the $\text{dist}_{\text{CM}_n(\mathbf{a})}(o_1, o_2)$ for $\tau = 1.8$ and $n = 10^3, 10^4, 10^5$.	267
9.1	The WWW according to Broder et al.	316

TABLES

- 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. 81
- 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. 121

PREFACE

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, reading seminars, 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 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, Łuczak 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 Albert and Barabási (2002) has attracted

over 27000 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 Chapter 1, 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 techniques such as local-weak convergence. 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. However, 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, on April 9, 2013, to 2346 and,

on April 21, 2016, to 2986. 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.url>), 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, on April 9, 2013, to 13,987 and, on April 21, 2016, to 19,841. 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 Network Pages at <http://www.networkspages.nl>. The Network Pages 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 Leeuwen, 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 appearing, 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 Deijfen, Michel Dekking, Henri van den Esker, Lorenzo Federico, Lucas Gerin, Jesse Goodman, Rajat Hazra, Markus Heydenreich, Frank den Hollander, Yusuke Ide, Lancelot James, Martin van Jole, Willemien Kets, Júlia Komjáthy, John Lapeyre, Lasse Leskelä, Nelly Litvak, Norio Konno, Abbas Mehrabian, Mislav Mišković, Mirko Moscatelli, Jan Nagel, Sidharthan Nair, Alex Olssen, Mariana Olvera-Cravioto, Helena Peña, Nathan Ross, Karoly Simon, 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. Artëm Sapozhnikov, Maren Eckhoff and Gerard Hooghiemstra read and commented the October 2011 version. Sándor Kolumbán read large parts of the October 2015 version, and spotted many errors, typos and inconsistencies.

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 \LaTeX skills. A big thanks for that! Also my thanks for suggestions and help with figures to Marko Boon, Alessandro Garavaglia, Dimitri Krioukov, Vincent Kusters, 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.

COURSE OUTLINE

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. We then continue to discuss an important technique in this book, called *local-weak convergence*.

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 is 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 examples 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 1.1, 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 graph models, as introduced in Volume 1. 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 $V = [n] := \{1, \dots, n\}$. 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

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

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 τ ,

$$n_k \approx c_n k^{-\tau}, \quad (1.1.4)$$

then

$$\log n_k \approx \log c_n - \tau \log k, \quad (1.1.5)$$

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 τ can be estimated by the slope of the line in the log-log plot. Naturally, we must have that

$$\sum_k n_k = n < \infty, \quad (1.1.6)$$

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

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.

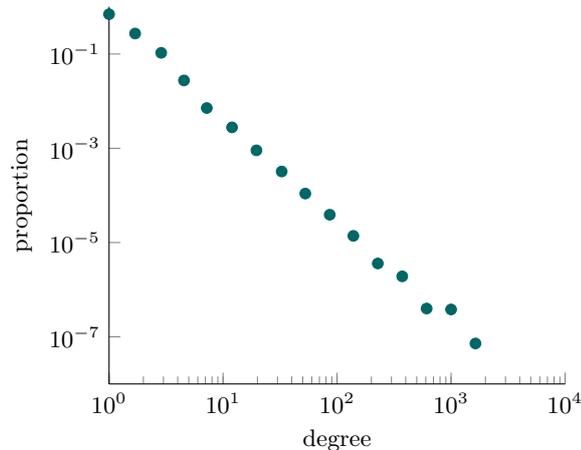


Figure 1.1 (a) Log-log plot of the probability mass function of the degree sequence of Autonomous Systems (AS) on April 2014 on a log-log scale from Krioukov et al. (2012) (data courtesy of Dmitri Krioukov).

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, 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),

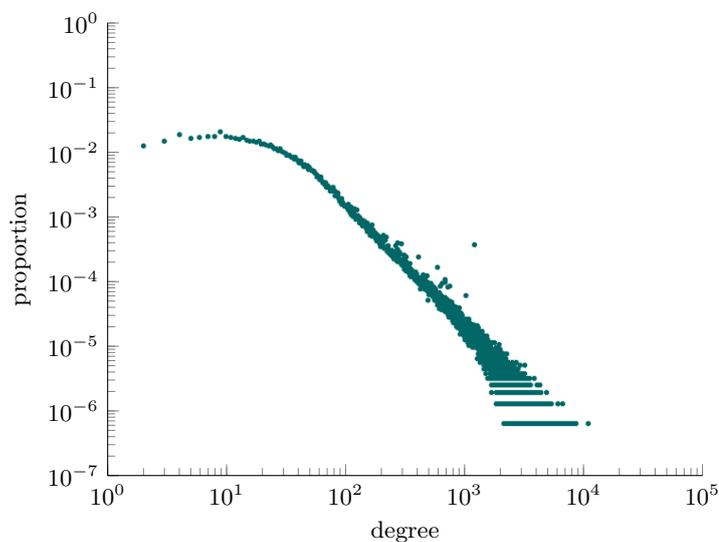


Figure 1.2 Log-log plot of the degree sequence in the Internet Movie Data base in 2007.

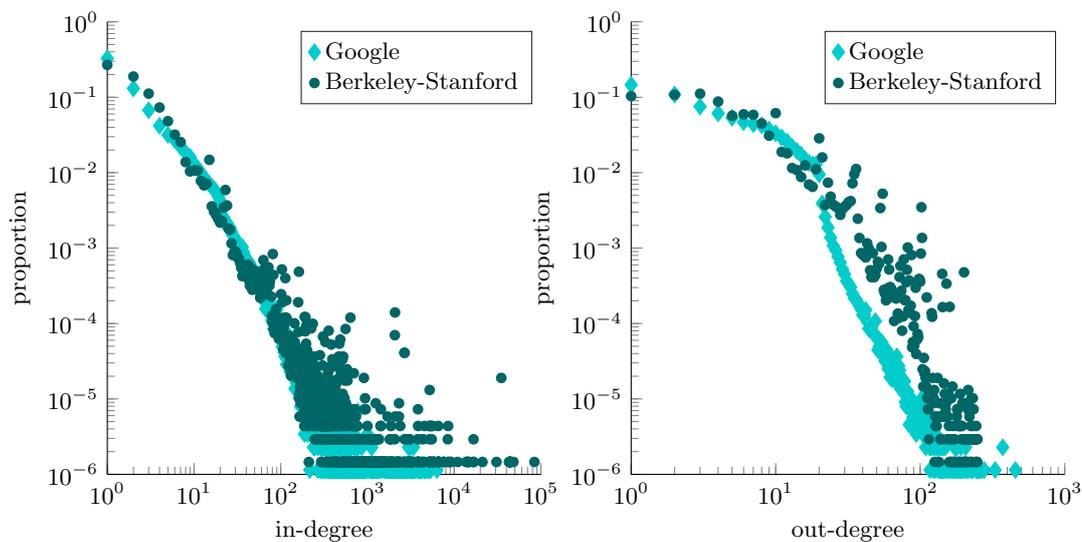


Figure 1.3 The probability mass function of the in- and out- degree sequences in the Berkeley-Stanford and Google competition graph data sets of the WWW in Leskovec et al. (2009). (a) in-degree; (b) out-degree.

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

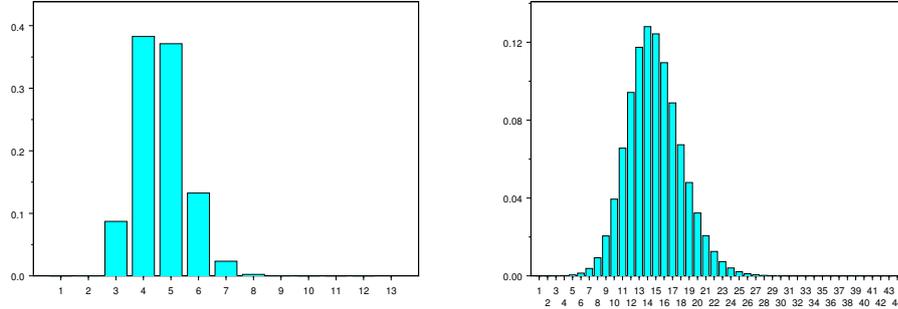


Figure 1.4 (a) Proportion of AS traversed in hopcount data. (b) Internet hopcount data. Courtesy of Hongsuda Tangmunarunkit.

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.

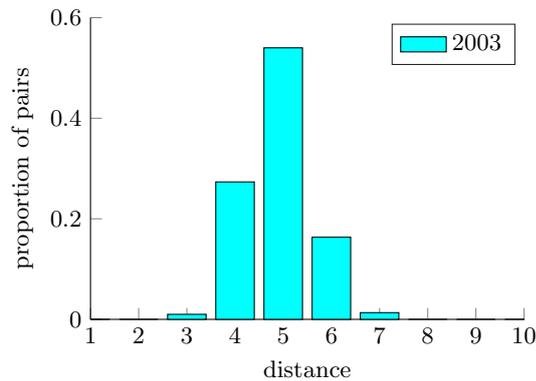


Figure 1.5 Typical distances in the Internet Movie Data base in 2003.

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

$$\text{dist}_G(o_1, o_2). \quad (1.1.7)$$

The quantity in (1.1.7) is a random variable even for *deterministic* graphs due to the occurrence of the two, uniformly at random, vertices $o_1, o_2 \in [n]$. 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

$$\text{diam}(G) = \max_{u, v \in [n]} \text{dist}_G(u, v). \quad (1.1.8)$$

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 than 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, mathematically, we model real-world networks by *graph sequences* $(G_n)_{n \geq 1}$, where G_n has size n and we take the limit $n \rightarrow \infty$. 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

$$\limsup_{n \rightarrow \infty} \mathbb{E}[D_n] = \limsup_{n \rightarrow \infty} \frac{1}{n} \sum_{i \in [n]} d_i < \infty. \quad (1.2.1)$$

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

$$D_n \xrightarrow{d} D, \quad (1.2.2)$$

where \xrightarrow{d} 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

$$\lim_{n \rightarrow \infty} \mathbb{P}(\text{dist}_G(o_1, o_2) \leq K \log n) = 1. \quad (1.2.3)$$

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 *dynamic* random graphs. 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 $[n] = \{1, \dots, n\}$, 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 $\binom{n}{2}$ 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

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

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

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

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

$$p = \frac{\lambda}{n}, \quad (1.3.3)$$

and study the graph as λ is fixed while $n \rightarrow \infty$. In this regime, we know that

$$D_n \xrightarrow{d} D, \quad (1.3.4)$$

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 $k \geq 0$,

$$P_k^{(n)} = \frac{1}{n} \sum_{i \in [n]} \mathbb{1}_{\{d_i = k\}} \xrightarrow{\mathbb{P}} p_k \equiv e^{-\lambda} \frac{\lambda^k}{k!}. \quad (1.3.5)$$

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 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 graphs 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 $\mathbf{w} = (w_i)_{i \in [n]}$ are the *vertex weights*, and ℓ_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 $\text{GRG}_n(\mathbf{w})$. 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 $\mathbf{w}^{(n)} = (w_i^{(n)})_{i \in [n]}$. 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 $\text{GRG}_n(\mathbf{w})$ is close to many other inhomogeneous random graph models, such as the *random graph with given prescribed degrees* or

Chung-Lu model, where instead

$$p_{ij} = p_{ij}^{(\text{CL})} = \min(w_i w_j / \ell_n, 1), \quad (1.3.8)$$

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

$$p_{ij} = p_{ij}^{(\text{NR})} = 1 - \exp(-w_i w_j / \ell_n). \quad (1.3.9)$$

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 $\mathbf{w} = (w_i)_{i \in [n]}$. 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

$$F_n(x) = \frac{1}{n} \sum_{i \in [n]} \mathbb{1}_{\{w_i \leq x\}}, \quad x \geq 0. \quad (1.3.10)$$

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 $n \rightarrow \infty$ the following conditions hold:*

(a) **Weak convergence of vertex weight.** As $n \rightarrow \infty$,

$$W_n \xrightarrow{d} W, \quad (1.3.11)$$

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,

$$\lim_{n \rightarrow \infty} F_n(x) = F(x). \quad (1.3.12)$$

(b) **Convergence of average vertex weight.** As $n \rightarrow \infty$,

$$\mathbb{E}[W_n] \rightarrow \mathbb{E}[W], \quad (1.3.13)$$

where W_n and W have distribution functions F_n and F , respectively. Further, we assume that $\mathbb{E}[W] > 0$.

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

$$\mathbb{E}[W_n^2] \rightarrow \mathbb{E}[W^2]. \quad (1.3.14)$$

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 $\text{GRG}_n(\mathbf{w})$ converges to the expectation of the limiting weight variable. In turn, this implies that the average degree in $\text{GRG}_n(\mathbf{w})$ 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 $\mathbf{w} = (w_i)_{i \in [n]}$ 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

$$w_i = [1 - F]^{-1}(i/n), \quad (1.3.15)$$

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)])

$$F_n(x) = \frac{1}{n}([\!|nF(x)|\!] + 1) \wedge 1. \quad (1.3.17)$$

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 $w_i w_j / (\sum_{k \in [n]} w_k)$, 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 $\text{GRG}_n(\mathbf{w})$. 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 $\text{GRG}_n(\mathbf{w})$. We denote the probability mass function of a *mixed Poisson distribution* by p_k , i.e., for $k \geq 0$,

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

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

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

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

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

$$D_n \xrightarrow{d} D, \quad (1.3.22)$$

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

Recall from Section 1.1 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,

$$\mathbb{P}(W > w) = \frac{c}{w^{\tau-1}}(1 + o(1)), \quad (1.3.23)$$

and then also, for w large,

$$\mathbb{P}(D > w) = \frac{c}{w^{\tau-1}}(1 + o(1)). \quad (1.3.24)$$

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 $\text{GRG}_n(\mathbf{w})$ 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 $i \in [n]$,

$$\mathbb{P}(X = x \mid d_i(X) = d_i \forall i \in [n]) = \frac{1}{\#\{y: d_i(y) = d_i \forall i \in [n]\}}, \quad (1.3.25)$$

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*

$$p_{ij} = \frac{w_i w_j}{\ell_n + w_i w_j}, \quad (1.3.26)$$

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 $\mathbf{d} = (d_i)_{i \in [n]}$. 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 *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

$$\ell_n = \sum_{j \in [n]} d_j \tag{1.3.27}$$

is even. We wish to construct a simple graph such that $\mathbf{d} = (d_i)_{i \in [n]}$ 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 ℓ_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 *configuration model with degree sequence \mathbf{d}* , abbreviated as $\text{CM}_n(\mathbf{d})$.

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 *exchangeable*. It can even be done in a *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 $\text{CM}_n(\mathbf{d})$ is. To do so, note that $\text{CM}_n(\mathbf{d})$ is characterized by the random vector $(X_{ij})_{1 \leq i < j \leq n}$,

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

$$d_i = X_{ii} + \sum_{j \in [n]} X_{ij} \quad (1.3.28)$$

In terms of this notation, and writing $G = (x_{ij})_{i,j \in [n]}$ to denote a multigraph on the vertices $[n]$,

$$\mathbb{P}(\text{CM}_n(\mathbf{d}) = G) = \frac{1}{(\ell_n - 1)!!} \frac{\prod_{i \in [n]} d_i!}{\prod_{i \in [n]} 2^{x_{ii}} \prod_{1 \leq i < j \leq n} x_{ij}!}. \quad (1.3.29)$$

See e.g., [Volume 1, Proposition 7.7] for this result. In particular, $\mathbb{P}(\text{CM}_n(\mathbf{d}) = G)$ 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 $\text{GRG}_n(\mathbf{w})$, we again impose *regularity conditions* on the degree sequence \mathbf{d} . In order to state these assumptions, we introduce some notation. We denote the degree of a uniformly chosen vertex o in $[n]$ by $D_n = d_o$. The random variable D_n has distribution function F_n given by

$$F_n(x) = \frac{1}{n} \sum_{j \in [n]} \mathbb{1}_{\{d_j \leq x\}}, \quad (1.3.30)$$

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 $n \rightarrow \infty$,*

$$D_n \xrightarrow{d} D, \quad (1.3.31)$$

where D_n and D have distribution functions F_n and F , respectively. Equivalently, for any x ,

$$\lim_{n \rightarrow \infty} F_n(x) = F(x). \quad (1.3.32)$$

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

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

$$\mathbb{E}[D_n] \rightarrow \mathbb{E}[D], \quad (1.3.33)$$

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

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

$$\mathbb{E}[D_n^2] \rightarrow \mathbb{E}[D^2], \quad (1.3.34)$$

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 $\text{CM}_n(\mathbf{d}) = (X_{ij})_{1 \leq i < j \leq n}$ by its erased version $\text{ECM}_n(\mathbf{d}) = (X_{ij}^{(\text{er})})_{1 \leq i < j \leq n}$, where $X_{ii}^{(\text{er})} \equiv 0$, while $X_{ij}^{(\text{er})} = 1$ 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 $\mathbf{D}^{(\text{er})} = (D_i^{(\text{er})})_{i \in [n]}$, so that

$$D_i^{(\text{er})} = d_i - 2s_i - m_i, \quad (1.3.35)$$

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

$$m_i = \sum_{j \neq i} (x_{ij} - 1) \mathbb{1}_{\{x_{ij} \geq 2\}} \quad (1.3.36)$$

is the number of multiple edges removed from i .

Denote the empirical degree sequence $(p_k^{(n)})_{k \geq 1}$ in $\text{CM}_n(\mathbf{d})$ by

$$p_k^{(n)} = \mathbb{P}(D_n = k) = \frac{1}{n} \sum_{i \in [n]} \mathbb{1}_{\{d_i = k\}}, \quad (1.3.37)$$

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

$$P_k^{(\text{er})} = \frac{1}{n} \sum_{i \in [n]} \mathbb{1}_{\{D_i^{(\text{er})} = k\}}. \quad (1.3.38)$$

From the notation it is clear that $(p_k^{(n)})_{k \geq 1}$ is a *deterministic* sequence when $\mathbf{d} = (d_i)_{i \in [n]}$ is deterministic, while $(P_k^{(\text{er})})_{k \geq 1}$ is a *random* sequence, since the erased degrees $(D_i^{(\text{er})})_{i \in [n]}$ form a random vector even when $\mathbf{d} = (d_i)_{i \in [n]}$ 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 \mathbf{d} satisfying Conditions 1.5(a)-(b), the degree sequence of the erased configuration model $(P_k^{(\text{er})})_{k \geq 1}$ converges in probability to $(p_k)_{k \geq 1}$. More precisely, for every $\varepsilon > 0$,*

$$\mathbb{P}\left(\sum_{k=1}^{\infty} |P_k^{(\text{er})} - p_k| \geq \varepsilon\right) \rightarrow 0, \quad (1.3.39)$$

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])

$$\lim_{n \rightarrow \infty} \mathbb{P}(\text{CM}_n(\mathbf{d}) \text{ is a simple graph}) = e^{-\nu/2 - \nu^2/4}, \quad (1.3.40)$$

where

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

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), $\text{CM}_n(\mathbf{d})$

conditioned on simplicity is a *uniform random graph* with the prescribed degree sequence. We denote this random graph by $UG_n(\mathbf{d})$. We will return to the difficulty of generating simple graphs with infinite-variance degrees later in this chapter.

Relation GRG and CM

Since $CM_n(\mathbf{d})$ conditioned on simplicity yields a uniform (simple) random graph with these degrees, and by (1.3.25), also $GRG_n(\mathbf{w})$ 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 $GRG_n(\mathbf{w})$ by proving them for $CM_n(\mathbf{d})$ under the appropriate degree conditions, and then proving that $GRG_n(\mathbf{w})$ 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 $GRG_n(\mathbf{w})$ through those for appropriate $CM_n(\mathbf{d})$'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 $GRG_n(\mathbf{w})$, and let $\mathbf{d} = (d_i)_{i \in [n]}$. Then, \mathbf{d} satisfies Conditions 1.5(a)-(b) in probability when \mathbf{w} satisfies Conditions 1.1(a)-(b), where*

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

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

1.3.4 SWITCHING ALGORITHMS FOR UNIFORM RANDOM GRAPHS

So far, we have focussed 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* 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 \mathbf{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 is aperiodic 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 \mathbf{d} which we denoted by $\text{UG}_n(\mathbf{d})$.

The above method works rather generally, and will, in the limit of infinitely many switches, produce a sample from $\text{UG}_n(\mathbf{d})$. As a result, this chain is the method of choice to produce a sample of $\text{UG}_n(\mathbf{d})$ 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 $\text{UG}_n(\mathbf{d})$, there can be at most one edge between two vertices. Then, the edge probabilities for $\text{UG}_n(\mathbf{d})$ 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 \mathbf{d} satisfies*

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

for some $C_F > 0$ and $\tau \in (2, 3)$. Assume further that $\ell_n/n \rightarrow 0$. 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 $\{u, v\} \notin U$,

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

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

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

$$1 - \mathbb{P}(u \sim v) = (1 + o(1)) \frac{\ell_n}{d_u d_v}. \quad (1.3.45)$$

In the erased configuration model, on the other hand,

$$1 - \mathbb{P}(u \sim v) \leq e^{-d_u d_v / \ell_n}. \quad (1.3.46)$$

Thus, the probability that two high-degree vertices are not connected is much smaller for $\text{ECM}_n(\mathbf{d})$ than for $\text{UG}_n(\mathbf{d})$. Instead, $\mathbb{P}(u \sim v) \approx \frac{d_u d_v}{\ell_n + d_u d_v}$, as it would be for $\text{GRG}_n(\mathbf{w})$ when $\mathbf{w} = \mathbf{d}$, which indicates once more that $\text{GRG}_n(\mathbf{w})$ and $\text{UG}_n(\mathbf{d})$ 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 \mathbf{d} satisfies (1.3.43) for some $C > 0$ and $\tau \in (2, 3)$. For any graph G whose degree sequence is \mathbf{d} , the number of two-paths starting from any specified vertex is $o(n)$.*

Proof Without loss of generality we may assume that $d_1 \geq d_2 \geq \dots \geq d_n$. 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

$$C_F n d_i^{1-\tau} \geq n[1 - F_n](d_i) \geq i \quad (1.3.47)$$

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 $\sum_{i=1}^{d_1} d_i$, which is at most

$$\begin{aligned} \sum_{i \geq 1}^{d_1} \left(\frac{C_F n}{i} \right)^{1/(\tau-1)} &= (C_F n)^{1/(\tau-1)} \sum_{i=1}^{d_1} i^{-1/(\tau-1)} \\ &= O\left(n^{1/(\tau-1)}\right) d_1^{\frac{\tau-2}{\tau-1}} = O\left(n^{(2\tau-3)/(\tau-1)^2}\right), \end{aligned} \quad (1.3.48)$$

since $d_1 \leq (C_F n)^{1/(\tau-1)}$. Since $\tau \in (2, 3)$, the above is $o(n)$. \square

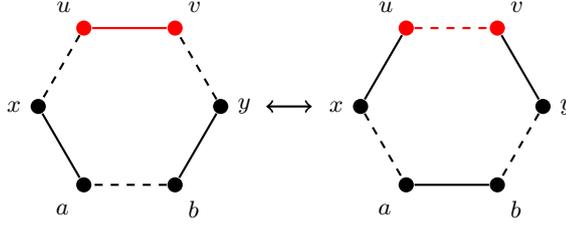


Figure 1.6 Forward and backward switchings

Proof of Theorem 1.8. To estimate $\mathbb{P}(u \sim v \mid E_U)$, 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

$$\mathbb{P}(u \sim v \mid E_U) = \frac{|S|}{|S| + |\bar{S}|} = \frac{1}{1 + |\bar{S}|/|S|}. \quad (1.3.49)$$

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 \bar{S} , denoted by $b(G')$. The number of total switchings between S and \bar{S} is equal to $|S|\mathbb{E}[f(G)] = |\bar{S}|\mathbb{E}[b(G')]$, where the expectation is over a uniformly random $G \in S$ and $G' \in \bar{S}$ respectively. Consequently,

$$\frac{|\bar{S}|}{|S|} = \frac{\mathbb{E}[f(G)]}{\mathbb{E}[b(G')]} \quad (1.3.50)$$

Given an arbitrary graph $G \in S$, the number of ways of carrying out a forward

switching is at most ℓ_n^2 , since there are at most ℓ_n ways to choose $\{x, a\}$, and at most ℓ_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 $8 \cdot \ell_n d_{\max} = o(\ell_n^2)$ choices, where d_{\max} bounds the number of ways to choose a vertex adjacent to u or v and ℓ_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 $G \in \mathcal{S}$ is $(1 + o(1))\ell_n^2$. Thus,

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

Given a graph $G' \in \bar{\mathcal{S}}$, 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 $o((d_u - |U_u|)(d_v - |U_v|)\ell_n)$. 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 $O((d_u - |U_u|)(d_v - |U_v|)) = o((d_u - |U_u|)(d_v - |U_v|)\ell_n)$.

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 $O((d_u - |U_u|)(d_v - |U_v|)d_{\max}) = o((d_u - |U_u|)(d_v - |U_v|)\ell_n)$ total choices, since $d_{\max} = O(n^{1/(\tau-1)}) = o(n) = o(L_n)$. 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

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

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

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

and thus (1.3.49) yields

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

□

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 $(\text{PA}_t^{(m,\delta)})_{t \geq 1}$ and which, for every time t , yields a graph of t vertices and mt edges for some $m = 1, 2, \dots$. We start by defining the model for $m = 1$ when the graph consists of a collection of trees. In this case, $\text{PA}_1^{(1,\delta)}$ consists of a single vertex with a single self-loop. We denote the vertices of $\text{PA}_t^{(1,\delta)}$ by $v_1^{(1)}, \dots, v_t^{(1)}$. We denote the degree of vertex $v_i^{(1)}$ in $\text{PA}_t^{(1,\delta)}$ 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 $\text{PA}_t^{(1,\delta)}$, the growth rule to obtain $\text{PA}_{t+1}^{(1,\delta)}$ 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 $v_{t+1}^{(1)}$ with probability $(1 + \delta)/(t(2 + \delta) + (1 + \delta))$, and to vertex $v_i^{(1)} \in \text{PA}_t^{(1, \delta)}$ 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,

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

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 $\text{PA}_t^{(1, \delta)}$.

The model with $m > 1$ is defined in terms of the model for $m = 1$ as follows. Fix $\delta \geq -m$. We start with $\text{PA}_{mt}^{(1, \delta/m)}$, and denote the vertices in $\text{PA}_{mt}^{(1, \delta/m)}$ by $v_1^{(1)}, \dots, v_{mt}^{(1)}$. Then we identify or collapse the m vertices $v_1^{(1)}, \dots, v_m^{(1)}$ in $\text{PA}_{mt}^{(1, \delta/m)}$ to become vertex $v_1^{(m)}$ in $\text{PA}_t^{(m, \delta)}$. In doing so, we let all the edges that are incident to any of the vertices in $v_1^{(1)}, \dots, v_m^{(1)}$ be incident to the new vertex $v_1^{(m)}$ in $\text{PA}_t^{(m, \delta)}$. Then, we collapse the m vertices $v_{m+1}^{(1)}, \dots, v_{2m}^{(1)}$ in $\text{PA}_{mt}^{(1, \delta/m)}$ to become vertex $v_2^{(m)}$ in $\text{PA}_t^{(m, \delta)}$, etc. More generally, we collapse the m vertices $v_{(j-1)m+1}^{(1)}, \dots, v_{jm}^{(1)}$ in $\text{PA}_{mt}^{(1, \delta/m)}$ to become vertex $v_j^{(m)}$ in $\text{PA}_t^{(m, \delta)}$. This defines the model for general $m \geq 1$. The resulting graph $\text{PA}_t^{(m, \delta)}$ 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 $(\text{PA}_t^{(m, \delta)})_{t \geq 1}$ 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 $(\text{PA}_t^{(m, \delta)})_{t \geq 1}$.

Degrees of fixed vertices

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

$$\Gamma(t) = \int_0^\infty x^{t-1} e^{-x} dx. \quad (1.3.56)$$

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) *Fix $m \geq 1$ and $\delta > -m$. Then, $D_i(t)/t^{1/(2+\delta/m)}$ converges almost surely to a random variable ξ_i as $t \rightarrow \infty$.*

It turns out that also $t^{-1/(2+\delta/m)} \max_{i \in [t]} D_i(t) \xrightarrow{a.s.} M$ 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 $t^{-1/(2+\delta/m)} \max_{i \in [t]} D_i(t) \xrightarrow{a.s.} M$ suggests that the degree of a random vertex satisfies a power law with power-law exponent $\tau = 3 + \delta/m$, 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

$$P_k(t) = \frac{1}{t} \sum_{i=1}^t \mathbb{1}_{\{D_i(t)=k\}} \quad (1.3.57)$$

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 $k = 0, \dots, m-1$ and, for $k \geq m$,

$$p_k = (2 + \delta/m) \frac{\Gamma(k + \delta) \Gamma(m + 2 + \delta + \delta/m)}{\Gamma(m + \delta) \Gamma(k + 3 + \delta + \delta/m)} \quad (1.3.58)$$

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 $\text{PA}_t^{(m, \delta)}$, as shown in the following theorem:

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

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

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 $k \rightarrow \infty$,

$$p_k = c_{m, \delta} k^{-\tau} (1 + O(1/k)), \quad (1.3.60)$$

where

$$\tau = 3 + \delta/m > 2, \quad \text{and} \quad c_{m, \delta} = (2 + \delta/m) \frac{\Gamma(m + 2 + \delta + \delta/m)}{\Gamma(m + \delta)}. \quad (1.3.61)$$

Therefore, by Theorem 1.12 and (1.3.60), the asymptotic degree sequence of $\text{PA}_t^{(m, \delta)}$ is close to a power law with exponent $\tau = 3 + \delta/m$. 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 $(\text{PA}_t^{(m,\delta)}(b))_{t \geq 1}$ model, in which the self-loops for $m = 1$ in (1.3.55) are not allowed, so that

$$\mathbb{P}(v_{t+1}^{(1)} \rightarrow v_i^{(1)} \mid \text{PA}_t^{(1,\delta)}(b)) = \frac{D_i(t) + \delta}{t(2 + \delta)} \quad \text{for } i \in [t]. \quad (1.3.62)$$

The model for $m \geq 2$ is again defined in terms of the model $(\text{PA}_t^{(1,\delta)}(b))_{t \geq 1}$ for $m = 1$ by collapsing blocks of m vertices. The advantage of $(\text{PA}_t^{(m,\delta)}(b))_{t \geq 1}$ compared to $(\text{PA}_t^{(m,\delta)})_{t \geq 1}$ is that $(\text{PA}_t^{(m,\delta)}(b))_{t \geq 1}$ is naturally *connected*, while $(\text{PA}_t^{(m,\delta)})_{t \geq 1}$ 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 $(\text{PA}_t^{(m,\delta)}(c))_{t \geq 1}$. In this case, the model for $m = 1$ is the same as $(\text{PA}_t^{(1,\delta)}(b))_{t \geq 1}$, while for $m \geq 2$ and $j \in \{0, \dots, m-1\}$, we attach the $(j+1)$ st edge of vertex $v_{t+1}^{(m)}$ to vertex $v_i^{(m)}$ for $i \in [t]$ with probability

$$\mathbb{P}(v_{t+1,j+1}^{(m)} \rightarrow v_i^{(m)} \mid \text{PA}_{t,j}^{(m,\delta)}(c)) = \frac{D_i(t,j) + \delta}{t(2m + \delta)} \quad \text{for } i \in [t]. \quad (1.3.63)$$

Here, $D_i(t,j)$ is the degree of vertex $v_i^{(m)}$ after the connection of the edges incident to the first t vertices, as well as the first j edges incident to vertex $v_{t+1}^{(m)}$. Many other adaptations are possible, and have been investigated in the literature, such as settings where the m edges incident to $v_{t+1}^{(m)}$ 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 $\text{BPA}_1^{(f)}$ 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

$$\frac{f(D_i(t-1))}{t-1}, \quad (1.3.64)$$

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 $t \rightarrow \infty$, i.e.,

$$P_k(t) \equiv \frac{1}{t} \sum_{i \in [t]} \mathbb{1}_{\{D_i(t)=k\}} \xrightarrow{\mathbb{P}} p_k, \quad \text{where} \quad p_k = \frac{1}{1+f(k)} \prod_{l=0}^{k-1} \frac{f(l)}{1+f(l)}. \quad (1.3.65)$$

In particular, $\log(1/p_k)/\log(k) \rightarrow 1+1/\gamma$ when $f(k)/k \rightarrow \gamma \in (0, 1)$ (see Exercise 1.12). Remarkably, when $f(k) = \gamma k + \beta$, the power-law exponent of the degree distribution does not depend on β . 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 deciding 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 $x \geq 1$,*

$$1 - F_x(x) \leq C_x x^{-(\tau-1)}. \quad (1.4.1)$$

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

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

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

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

Proof We note that for any cumulative distribution function $x \mapsto F(x)$ on the non-negative reals, we have the *partial integration identity*

$$\begin{aligned} \int_u^\infty f(x) F(dx) &= f(u)[1 - F(u)] + \int_u^\infty [f(x) - f(u)] F(dx) \\ &= f(u)[1 - F(u)] + \int_u^\infty \int_0^x f'(y) dy F(dx) \\ &= f(u)[1 - F(u)] + \int_u^\infty f'(y) \int_y^\infty F(dx) dy \\ &= f(u)[1 - F(u)] + \int_u^\infty f'(y)[1 - F(y)] dy. \end{aligned} \quad (1.4.4)$$

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 $\ell > 0$,

$$\begin{aligned} \mathbb{E}[X^a \mathbb{1}_{\{X > \ell\}}] &= x^a \mathbb{P}(X > \ell) + \int_{\ell}^{\infty} ax^{a-1} \mathbb{P}(X > x) dx \\ &\leq C_x x^{a-(\tau-1)} + aC_x \int_{\ell}^{\infty} x^{a-1} x^{-(\tau-1)} \leq C_{a,\tau} \ell^{a-(\tau-1)}. \end{aligned} \quad (1.4.5)$$

as required. \square

An important notion in many graphs is the *size-biased* version X^* of a non-negative random variable X that is given by

$$\mathbb{P}(X^* \leq x) = \frac{\mathbb{E}[X \mathbb{1}_{\{X \leq x\}}]}{\mathbb{E}[X]}. \quad (1.4.6)$$

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 $x \geq 1$,*

$$1 - F_x(x) \leq C_x x^{-(\tau-1)}. \quad (1.4.7)$$

Assume that $\tau > 2$, so that $\mathbb{E}[X] < \infty$. Further, assume that $\mathbb{E}[X] > 0$. Then, there exists a constant C such that

$$1 - F_x^*(x) \leq C_x^* x^{-(\tau-2)}. \quad (1.4.8)$$

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

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., $\mathbb{P}(X = 1) = 1 - \mathbb{P}(X = 0) = p$. 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 λ . 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/λ and variance r/λ^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 $f_x(x) = x^{a-1} (1-x)^{b-1} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}$.

We sometimes abuse notation, and write e.g., $\mathbb{P}(\text{Bin}(n, p) = k)$ to denote $\mathbb{P}(X = k)$ 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 $(\mathcal{E}_n)_{n \geq 1}$ occurs *with high probability* (whp) when $\lim_{n \rightarrow \infty} \mathbb{P}(\mathcal{E}_n) = 1$. We further write $f(n) = O(g(n))$ if $|f(n)|/|g(n)|$ is uniformly bounded from above by a positive constant as $n \rightarrow \infty$, $f(n) = \Theta(g(n))$ 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 $n \rightarrow \infty$. 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 $X_n \xrightarrow{\mathbb{P}} X$ 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 $X_n = O_{\mathbb{P}}(Y_n)$ when $|X_n|/Y_n$ is a tight sequence of random variables and $X_n = o_{\mathbb{P}}(Y_n)$ when $|X_n|/Y_n \xrightarrow{\mathbb{P}} 0$. Finally, we write that $X_n = o_{\mathbb{P}}(Y_n)$ when $X_n/Y_n \xrightarrow{\mathbb{P}} 0$.

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 \mathbf{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 $\emptyset v_1 \cdots v_k \in V(\mathbf{t})$, then also $\emptyset v_1 \cdots v_{k-1} \in V(\mathbf{t})$ and $\emptyset v_1 \cdots (v_k - 1) \in \mathbf{t}$ 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 $v \in V(\mathbf{t})$, let $|v|$ be its height. Thus $|v| = k$ when $v = \emptyset v_1 \cdots v_k$ and $|\emptyset| = 0$. Let $u, v \in V(\mathbf{t})$. Then $u < v$ when either $|u| < |v|$, or when $|u| = |v|$ and $u = \emptyset u_1 \cdots u_k$ and $v = \emptyset v_1 \cdots v_k$ are such that $(u_1, \dots, u_k) < (v_1, \dots, v_k)$ in the lexicographic sense.

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

$$s_i = s_{i-1} + x_i - 1 \quad \text{for } i \geq 1, \quad \text{with} \quad s_0 = 1, \quad (1.5.1)$$

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

1.6 NOTES AND DISCUSSION

Notes on Sections 1.1–1.3

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) *Prove (1.1.3).*

Exercise 1.2 (Uniform random graph) *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) *Show that, for every $\alpha > 0$, and with $p_k = e^{-\lambda} \frac{\lambda^k}{k!}$ the Poisson probability mass function,*

$$\lim_{k \rightarrow \infty} e^{\alpha k} p_k = 0. \quad (1.7.1)$$

Exercise 1.4 (Weight of uniformly chosen vertex) *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 $GRG_n(\mathbf{w})$) *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 $o_\varepsilon(n)$.*

Exercise 1.10 (Degrees grow to infinity a.s.) Fix $m = 1$ and $i \geq 1$. Prove that $D_i(t) \xrightarrow{\text{a.s.}} \infty$, by using that $\sum_{s=i}^t I_s \preceq D_i(t)$, where $(I_t)_{t \geq i}$ 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

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

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 $f(k) \sim \gamma k^\alpha$ 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 WEAK 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 describes 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 weak 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 weak convergence in several probabilistic senses, and show that local weak convergence is equivalent to the convergence of subgraph counts. Then we discuss several implications of local weak convergence, concerning local neighborhoods, clustering, assortativity and PageRank. We further investigate the relation between local weak 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 (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 weak convergence to random graphs, for which there are several notions of convergence alike for convergence of real-valued random variables. In Section 2.4, we discuss consequences of local weak 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 weak 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 be finite or infinite, 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.

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

$$\begin{aligned} V(B_r(o)) &= \{u : d_G(o, u) \leq r\}, \\ E(B_r(o)) &= \{\{u, v\} \in E(G) : d_G(o, u), d_G(o, v) \leq r\}. \end{aligned} \tag{2.1.1}$$

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

Definition 2.2 (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 $\phi(o_1) = \phi(o_2)$ 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 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.

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

Definition 2.3 (Metric on rooted graphs) *Let (G_1, o_1) and (G_2, o_2) be two*

rooted connected graphs, and write $B_r^{(G_i)}(o_i)$ for the neighborhood of vertex o_i in G_i . Let

$$R^* = \sup\{r: B_r^{(G_1)}(o_1) \simeq B_r^{(G_2)}(o_2)\}. \quad (2.1.2)$$

Define

$$d_{\mathcal{G}_*}((G_1, o_1), (G_2, o_2)) = 1/(R^* + 1). \quad (2.1.3)$$

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 $B_r^{(G_1)}(o_1)$ is isomorphic to $B_r^{(G_2)}(o_2)$. When $R^* = \infty$, then $B_r^{(G_1)}(o_1)$ is isomorphic to $B_r^{(G_2)}(o_2)$ 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 below where this is worked out in detail.

2.2 LOCAL WEAK CONVERGENCE OF GRAPHS

We next define local weak convergence of finite graphs:

Definition 2.4 (Local weak convergence) *Let $G_n = ([n], E(G_n))$ denote a finite (possibly disconnected) graph. Let (G_n, o_n) be the rooted graph obtained by letting $o_n \in [n]$ be chosen uniformly at random and restricting G_n to the connected component of o_n in G_n .*

We say that G_n converges in the local weak sense to (G, o) , which is a (possibly) random element of \mathcal{G}_ having law μ , when, for every bounded and continuous function $h: \mathcal{G}_* \mapsto \mathbb{R}$,*

$$\mathbb{E}_n[h(G_n, o_n)] \rightarrow \mathbb{E}[h(G, o)], \quad (2.2.1)$$

where the expectation on the right-hand side of (2.2.1) is w.r.t. (G, o) having law μ , while the expectation \mathbb{E}_n is w.r.t. the random vertex o_n only.

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 (2.2.1) is equal to

$$\mathbb{E}_n[h(G_n, o_n)] = \frac{1}{n} \sum_{u \in [n]} h(G_n, u). \quad (2.2.2)$$

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 1.6 we discuss examples of how local weak convergence may be used to obtain interesting consequences for graphs, such as its clustering and assortativity. The notion will also be crucially used in this book. We continue by discussing a convenient criterion for proving local weak convergence.

Criterion for local weak convergence

We next provide a convenient criterion for local weak convergence:

Theorem 2.5 (Criterion for local weak convergence) *The sequence of finite graphs $(G_n)_{n \geq 1}$ converges in the local weak sense to (G, o) precisely when, for every rooted graph $H_\star \in \mathcal{G}_\star$,*

$$p^{(n)}(H) = \frac{1}{n} \sum_{u \in [n]} \mathbb{1}_{\{B_r^{(G_n)}(u) \simeq H_\star\}} \rightarrow \mathbb{P}(B_r(o) \simeq H_\star), \quad (2.2.3)$$

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

Proof First of all, local weak convergence implies that (2.2.3) holds, since we can take $h(G, o) = \mathbb{1}_{\{B_r(o) \simeq H_\star\}}$, and $h: \mathcal{G}_\star \mapsto \{0, 1\}$ is continuous (see Exercise 2.6).

For the other side, we need to prove that if (2.2.3) holds, then, $\mathbb{E}_n[h(G_n, o_n)] \rightarrow \mathbb{E}[h(G, o)]$ for every bounded and continuous function $h: \mathcal{G}_\star \mapsto \mathbb{R}$. Fix ε . Take $r = r(\delta)$ so large that

$$d_{\mathcal{G}_\star}((G, o), B_o(r)) \leq \delta. \quad (2.2.4)$$

Next, by continuity of h , we can take $\delta > 0$ so small that, for every rooted graph (G, o) ,

$$|h(G, o) - h(B_o(r))| \leq \varepsilon/4. \quad (2.2.5)$$

We conclude that

$$\begin{aligned} & \left| \mathbb{E}_n[h(G_n, o_n)] - \mathbb{E}[h(G, o)] \right| \\ & \leq \left| \mathbb{E}_n[h(B_r^{(G_n)}(o_n))] - \mathbb{E}[h(B_r(o))] \right| + \varepsilon/2. \end{aligned} \quad (2.2.6)$$

Fix k large. Let $E_{r,k}(G, o)$ be the event that there exists a vertex v at distance at most r from o in G such that $d_v(G) > k$. Since the limiting graph (G, o) is locally finite, for every $\eta > 0$, we can take k so large that

$$\mathbb{P}(E_{r,k}(G, o)) \leq \eta. \quad (2.2.7)$$

Further, we can write

$$\mathbb{P}(E_{r,k}(G_n, o_n)) = 1 - \mathbb{P}(E_{r,k}(G_n, o_n)^c). \quad (2.2.8)$$

Now, there is just a *finite* number of rooted (G, o) graphs such that $E_{r,k}(G, o)^c$ holds (see Exercise 2.7). Thus, by (2.2.3),

$$\mathbb{P}(E_{r,k}(G_n, o_n)^c) \rightarrow \mathbb{P}(E_{r,k}(G, o)^c). \quad (2.2.9)$$

Take n_0 so large that, for all $n \geq n_0$,

$$|\mathbb{P}(E_{r,k}(G_n, o_n)^c) - \mathbb{P}(E_{r,k}(G, o)^c)| \leq \eta. \quad (2.2.10)$$

Then, also $\mathbb{P}(E_{r,k}(G_n, o_n)) \leq 2\eta$. We conclude that

$$\begin{aligned} & \left| \mathbb{E}_n[h(G_n, o_n)] - \mathbb{E}[h(G, o)] \right| \\ & \leq \left| \mathbb{E}_n[h(B_r^{(G_n)}(o_n)) \mathbb{1}_{E_{r,k}(G_n, o_n)^c}] - \mathbb{E}[h(B_r(o)) \mathbb{1}_{E_{r,k}(G, o)^c}] \right| + \varepsilon/2 + 2\eta \|h\|_\infty, \end{aligned} \quad (2.2.11)$$

where $\|h\|_\infty = \sup_{H_\star \in \mathcal{G}_\star} |h(H_\star)| < \infty$, since h is a bounded function. Again, since there is just a *finite* number of rooted (G, o) graphs for which $E_{r,k}(G, o)^c$ holds, by (2.2.3),

$$\mathbb{E}_n[h(B_r^{(G_n)}(o_n)) \mathbb{1}_{E_{r,k}(G_n, o_n)^c}] \rightarrow \mathbb{E}[h(B_r(o)) \mathbb{1}_{E_{r,k}(G, o)^c}]. \quad (2.2.12)$$

Then, pick $n_1 \geq 1$ so large that, for all $n \geq n_1$,

$$\left| \mathbb{E}_n[h(B_r^{(G_n)}(o_n)) \mathbb{1}_{E_{r,k}(G_n, o_n)^c}] - \mathbb{E}[h(B_r(o)) \mathbb{1}_{E_{r,k}(G, o)^c}] \right| \leq \varepsilon/4. \quad (2.2.13)$$

Taking $\eta = \varepsilon/(8\|h\|_\infty)$ then shows that, for all $n \geq n_0 \vee n_1$,

$$\left| \mathbb{E}_n[h(G_n, o_n)] - \mathbb{E}[h(G, o)] \right| \leq \varepsilon, \quad (2.2.14)$$

which proves the claim. \square

Theorem 2.5 shows that the proportion of vertices in G_n whose neighborhood looks like H_\star converges to a (possibly random) limit. 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.

We continue by giving a *tightness criterium* for local weak convergence:

Theorem 2.6 (Tightness criterion for local weak convergence) *Let $(G_n)_{n \geq 1}$ be a sequence of marked graphs, where $V(G_n) = [n]$. 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 $[n]$ 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 and let $U(G) = (G, U_G)$ be the rooted graph obtained by rooting G at U_G . We need to show that if $\{\deg_G(U_G) : G \in A\}$ is a uniformly integrable sequence of random variables, then the family A is tight. Let

$$f(d) = \sup_{G \in A} \mathbb{E}[\deg_G(U_G) \mathbb{1}_{\{\deg_G(U_G) > d\}}]. \quad (2.2.15)$$

By assumption, $\lim_{d \rightarrow \infty} f(d) = 0$. Write $m(G) = \mathbb{E}[\deg_G(U_G)]$. 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,

$$D(G)[(G, v)] = \frac{\deg_G(v)}{m(G)} \cdot U(G)[(G, v)], \quad (2.2.16)$$

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 $\{D(G) : G \in A\}$ is tight. Note that $\{\deg_G(D_G) : G \in A\}$ 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,

$$\mathbb{P}\left(F_{r+1}^M(D_G) \mid \deg_G(D_G)\right) \leq \deg_G(D_G) \mathbb{P}\left(F_r^M(X) \mid \deg_G(D_G)\right). \quad (2.2.17)$$

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

$$\mathbb{P}\left(F_r^M(X)\right) \leq \varepsilon \quad (2.2.18)$$

for all $G \in A$. This clearly implies that $\{D(G) : G \in A\}$ 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 $\mathbb{P}(\deg_G(D_G) > d) \leq \varepsilon/2$ for all $G \in A$. Also, choose M so large that $\mathbb{P}(F_r^M(D_G)) \leq \varepsilon/(2d)$ for all $G \in A$. Write F for the event such that $\deg_G(D_G) > d$. Then, by conditioning on $\deg_G(D_G)$, we see that

$$\begin{aligned} \mathbb{P}\left(F_{r+1}^M(D_G)\right) &\leq \mathbb{P}(F) + \mathbb{E}\left[\mathbb{1}_{F^c} \mathbb{P}\left(F_{r+1}^M(D_G) \mid \deg_G(D_G)\right)\right] \\ &\leq \varepsilon/2 + \mathbb{E}\left[\mathbb{1}_{F^c} \deg_G(D_G) \mathbb{P}\left(F_r^M(D_G) \mid \deg_G(D_G)\right)\right] \\ &\leq \varepsilon/2 + \mathbb{E}\left[\mathbb{1}_{F^c} d \mathbb{P}\left(F_r^M(D_G) \mid \deg_G(D_G)\right)\right] \\ &\leq \varepsilon/2 + d \mathbb{P}\left(F_r^M(D_G)\right) \\ &\leq \varepsilon/2 + d\varepsilon/(2d) = \varepsilon, \end{aligned} \quad (2.2.19)$$

for all $G \in A$, which proves the claim. \square

The needed uniform integrability in Theorem 2.6 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 $D_n = d_{o_n}^{(n)}$ for the degree of a uniform vertex in our graph. When $(d_{o_n}^{(n)})_{n \geq 1}$ forms a uniformly integrable sequence of random variables, there exists a subsequence along which D_n^* , the size-biased version of $D_n = d_{o_n}^{(n)}$, converges in distribution (see Exercise 2.11 below).

For the configuration model, Conditions 1.5(a)-(b) imply that $(d_{o_n}^{(n)})_{n \geq 1}$ 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 *neighbors* of the uniform vertex o_n , a distribution that is given by D_n^* . Of course, for local weak convergence to hold, one needs at least that the degrees of neighbors converges. Thus, at least for the configuration model, we can fully understand why the uniform integrability of $(d_{o_n}^{(n)})_{n \geq 1}$ is needed.

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.

2.3 LOCAL WEAK 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 2.3.3 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 2.3.4 by presenting the example of local weak convergence in probability of the Erdős-Rényi random graph $ER_n(\lambda/n)$.

2.3.1 DEFINITION OF LOCAL WEAK 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 2.7 (Local weak convergence of random graphs) *Let $G_n = ([n], E(G_n))$ denote a finite (possibly disconnected) random graph. Then,*

- (a) *We say that G_n converges in distribution in the local weak sense to (G, o) when*

$$\mathbb{E}[h(G_n, o_n)] \rightarrow \mathbb{E}_{\bar{\mu}}[h(G, o)], \quad (2.3.1)$$

where the expectation on the right-hand side of (2.3.1) is w.r.t. (G, o) having law $\bar{\mu}$, for every bounded and continuous function $h: \mathcal{G}_ \mapsto \mathbb{R}$, while the expectation \mathbb{E} is w.r.t. the random vertex o_n and the random graph G_n .*

- (b) *We say that G_n converges in probability in the local weak sense to (G, o) when*

$$\mathbb{E}_n[h(G_n, o_n)] \xrightarrow{\mathbb{P}} \mathbb{E}_{\mu}[h(G, o)], \quad (2.3.2)$$

where the expectation on the right-hand side of (2.3.2) is w.r.t. (G, o) having (possibly random) law μ , for every bounded and continuous function $h: \mathcal{G}_* \mapsto \mathbb{R}$, while the expectation \mathbb{E}_n is w.r.t. the random vertex o_n only.

(c) We say that G_n converges almost surely in the local weak sense to (G, o) when

$$\mathbb{E}_n[h(G_n, o_n)] \xrightarrow{a.s.} \mathbb{E}_\mu[h(G, o)], \quad (2.3.3)$$

where the expectation on the right-hand side of (2.3.2) is w.r.t. (G, o) having law μ , for every bounded and continuous function $h: \mathcal{G}_* \mapsto \mathbb{R}$, while the expectation \mathbb{E}_n is w.r.t. the random vertex o_n only.

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 μ for convergence in probability and almost surely in Corollary 2.8 below.

When we have convergence in probability in the local weak sense, then $\mathbb{E}_n[h(G_n, o_n)]$, 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 μ might be a *random* probability distribution on \mathcal{G}_* . Thus, this is sometimes called a *quenched* result. When we have convergence in distribution in the local weak sense, instead, then only expectations w.r.t. the random graph of the form $\mathbb{E}[h(G_n, o_n)]$ converge. This corresponds to an *annealed* result. Certainly when considering complicated functionals of the random graph, we would prefer the limit *not* to depend on the specific random graph, and be deterministic instead. For this, one needs the stronger notion of convergence in probability in the local weak sense.

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

Corollary 2.8 (Relation between local weak limits) *Suppose that G_n converges in distribution in the local weak sense to (\bar{G}, \bar{o}) having law $\bar{\mu}$, and in probability in the local weak sense to (G, o) having law μ . Then $\bar{\mu}(\cdot) = \mathbb{E}[\mu(\cdot)]$. In particular,*

$$\mathbb{E}_{\bar{\mu}}[h(G, o)] = \mathbb{E}[\mathbb{E}_\mu[h(G, o)]] \quad (2.3.4)$$

Proof Note that $\mathbb{E}_\mu[h(G, o)]$ is a bounded random variable, and so is $\mathbb{E}_n[h(G_n, o_n)]$. Therefore, by Dominated Convergence [Volume 1, Theorem A.1], also the expectations converge. Therefore,

$$\mathbb{E}[h(G_n, o_n)] = \mathbb{E}[\mathbb{E}_n[h(G_n, o_n)]] \rightarrow \mathbb{E}[\mathbb{E}_\mu[h(G, o)]] \quad (2.3.5)$$

which completes the proof. \square

In most of our examples, the law μ 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 $\text{Poi}(X)$, as for the local weak limit of $\text{ER}_n(X/n)$, we draw X *only once*. See Section 2.3.4 for more details on local weak convergence for $\text{Poi}(\lambda)$.

We have added the notion of local weak 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 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 weak 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.

Exercise 2.14 shows a consequence of local weak convergence in probability that can be quite useful, as it shows that the neighborhoods of two independently chosen uniform vertices are independent.

2.3.2 CONVENIENT CRITERION FOR LOCAL WEAK CONVERGENCE OF RANDOM GRAPHS

We next discuss a convenient criterion for local weak convergence, inspired by Theorem 2.5:

Theorem 2.9 (Criterion for local weak convergence of random graphs) *Let $(G_n)_{n \geq 1}$ be a sequence of rooted graphs. Then,*

- (a) G_n converges in distribution in the local weak sense to (G, o) having law $\bar{\mu}$ when, for every rooted graph $H_\star \in \mathcal{G}_\star$,

$$\mathbb{E}[p^{(n)}(H_\star)] = \frac{1}{n} \sum_{u \in [n]} \mathbb{P}(B_r^{(G_n)}(u) \simeq H_\star) \rightarrow \bar{\mu}(B_r(o) \simeq H_\star), \quad (2.3.6)$$

where $B_r(o)$ is the rooted r -neighborhood of o in (G, o) and we view $B_r^{(G_n)}(u)$ as a rooted graph with root u in G_n .

- (b) G_n converges in probability in the local weak sense to (G, o) having law μ when for every rooted graph $H_\star \in \mathcal{G}_\star$,

$$p^{(n)}(H_\star) = \frac{1}{n} \sum_{u \in [n]} \mathbb{1}_{\{B_r^{(G_n)}(u) \simeq H_\star\}} \xrightarrow{\mathbb{P}} \mu(B_r(o) \simeq H_\star), \quad (2.3.7)$$

where $B_r(o)$ is the rooted r -neighborhood of o in (G, o) and we view $B_r^{(G_n)}(u)$ as a rooted graph with root u in G_n .

- (c) G_n converges almost surely in the local weak sense to (G, o) having law μ when

for every rooted graph $H_\star \in \mathcal{G}_\star$,

$$p^{(n)}(H_\star) = \frac{1}{n} \sum_{u \in [n]} \mathbb{1}_{\{B_r^{(G_n)}(u) \simeq H_\star\}} \xrightarrow{a.s.} \mu(B_r(o) \simeq H_\star), \quad (2.3.8)$$

where $B_r(o)$ is the rooted r -neighborhood of o in (G, o) and we view $B_r^{(G_n)}(u)$ as a rooted graph with root u in G_n .

Proof This follows directly from Theorem 2.5. \square

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.

Local weak convergence of marked graphs

Benjamini, Lyons and Schramm (2015) allow for more general *marks*, a notion that will also be convenient for us, 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 Ξ , called the mark space. Here Ξ maps from $V(G)$ and $E(G)$ to Ξ . Images in Ξ 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 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 α 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 μ_n, μ on \mathcal{G}_\star , we write $\mu_n \xrightarrow{d} \mu$ as $n \rightarrow \infty$ when μ_n converges weakly to μ with respect to this metric. See Exercise 2.15 for an application of marked graphs to *directed* graphs.

2.3.3 LOCAL WEAK CONVERGENCE AND COMPLETENESS OF THE LIMIT

For many random graph models, the local weak limit is almost surely contained in a smaller set of rooted graphs. The most common example is when the random graph converges in the local weak sense to a *tree* (recall that trees are rooted, see Section 1.5), but it can apply more generally. In this case, it turns out to be enough to prove the convergence in Definition 2.7 only for elements that the limit (G, o) takes values in. Let us explain this in more detail.

Let $\mathcal{T}_\star \subset \mathcal{G}_\star$ be a subset of the space of rooted graphs. Let $\mathcal{T}_\star(r) \subseteq \mathcal{T}_\star$ be the subset of \mathcal{T}_\star of graphs for which the distance between any vertex and the root is at most r . Then, we have the following result:

Theorem 2.10 (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}_\star having law $\bar{\mu}$. Let $\mathcal{T}_\star \subset \mathcal{G}_\star$ be a subset of the space of rooted graphs. Assume that $\bar{\mu}((G, o) \in \mathcal{T}_\star) = 1$.*

Then, G_n converges in distribution in the local weak sense to (G, o) when (2.3.6) holds for all $H_\star \in \mathcal{T}_\star(k)$.

Similar extensions hold for local weak convergence in probability in (2.3.7) and almost surely in (2.3.8), with $\bar{\mu}$ replaced by μ .

We will apply Theorem 2.10 in particular when the limit is a.s. a tree. Then, Theorem 2.10 implies that we only have to investigate H_\star that are trees themselves.

Proof The set $\mathcal{T}_\star(k)$ is countable. Therefore, since $\bar{\mu}((G, o) \in \mathcal{T}_\star) = 1$, for every $\varepsilon > 0$, there exists an $m = m(\varepsilon)$ and a subset $\mathcal{T}_\star(r, m)$ of size at most m such that $\bar{\mu}(B_r(o) \in \mathcal{T}_\star(r, m)) \geq 1 - \varepsilon$. Fix this set. Then we bound

$$\begin{aligned} \mathbb{P}(B_r^{(G_n)}(o) \notin \mathcal{T}_\star(r)) &= 1 - \mathbb{P}(B_r^{(G_n)}(o) \in \mathcal{T}_\star(r)) \\ &\leq 1 - \mathbb{P}(B_r^{(G_n)}(o) \in \mathcal{T}_\star(r, m)). \end{aligned} \quad (2.3.9)$$

Therefore,

$$\begin{aligned} \limsup_{n \rightarrow \infty} \mathbb{P}(B_r^{(G_n)}(o) \notin \mathcal{T}_\star(r)) &\leq 1 - \liminf_{n \rightarrow \infty} \mathbb{P}(B_r^{(G_n)}(o) \in \mathcal{T}_\star(r, m)) \\ &= 1 - \bar{\mu}(B_r(o) \in \mathcal{T}_\star(r, m)) \leq 1 - (1 - \varepsilon) = \varepsilon. \end{aligned} \quad (2.3.10)$$

Since $\varepsilon > 0$ is arbitrary, we conclude that $\mathbb{P}(B_r^{(G_n)}(o) \notin \mathcal{T}_\star(r)) \rightarrow 0$. In particular, this means that, for any $H_\star \notin \mathcal{T}_\star(k)$,

$$\mathbb{P}(B_r^{(G_n)}(o) \simeq H_\star) \rightarrow 0 = \bar{\mu}(B_r(o) \simeq H_\star). \quad (2.3.11)$$

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

The extensions to convergence in probability and almost surely follow similarly. \square

2.3.4 LOCAL WEAK CONVERGENCE OF ERDŐS-RÉNYI RANDOM GRAPHS

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

Theorem 2.11 (Local weak convergence of Erdős-Rényi random graph) *Fix $\lambda > 0$. $\text{ER}_n(\lambda/n)$ converges in probability in the local weak sense to a Poisson branching process with mean offspring λ .*

Proof We will rely on Theorem 2.10, and prove convergence of subgraph proportions as in (2.3.8) 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 \mathbf{t} . Then, in order to prove that $\text{ER}_n(\lambda/n)$ converges locally weakly in probability to a $\text{Poi}(\lambda)$ branching process, we need to show that, for every tree \mathbf{t} ,

$$p^{(n)}(\mathbf{t}) = \frac{1}{n} \sum_{u \in [n]} \mathbb{1}_{\{B_r^{(G_n)}(u) \simeq \mathbf{t}\}} \xrightarrow{\mathbb{P}} \mu(B_r(o) \simeq \mathbf{t}), \quad (2.3.12)$$

where G_n equals $\text{ER}_n(\lambda/n)$, and the law μ of (G, o) is that of a $\text{Poi}(\lambda)$ branching process. We see that in this case, μ is *deterministic*, as it will be in most examples. In (2.3.12), we may without loss of generality assume that \mathbf{t} is a finite tree of depth at most r , since otherwise the right hand side equals zero.

Of course, for the event $\{B_r(o) \simeq \mathbf{t}\}$ to occur, the order of the tree \mathbf{t} is irrelevant. It will be convenient to also order the vertices in $\text{ER}_n(\lambda/n)$. This will be achieved by ordering the forward children of a vertex in $B_r^{(G_n)}(o_n)$ according to their vertex labels. Recall the breadth-first exploration of the tree \mathbf{t} , which is described in terms of $(x_i)_{i=0}^t$ as in (1.5.1) and the corresponding vertices $(a_i)_{i=0}^t$. In terms of this notation,

$$\mu(B_r(o) \simeq \mathbf{t}) = \prod_{i \in [t]: \text{dist}(\emptyset, a_i) < r} e^{-\lambda} \frac{\lambda^{x_i}}{x_i!}, \quad (2.3.13)$$

where $\text{dist}(\emptyset, v)$ is the tree distance between $v \in V(\mathbf{t})$ and the root. We note that $B_r(o) \simeq \mathbf{t}$ implies nothing for the vertices that are at distance exactly r away from the root \emptyset , which is why we restrict to vertices v with $\text{dist}(\emptyset, a_i) < r$ in (2.3.13). This identifies the right hand side of (2.3.12).

To prove (2.3.12), we use a second moment method. Denote

$$N_n(\mathbf{t}) = \sum_{u \in [n]} \mathbb{1}_{\{B_r^{(G_n)}(u) \simeq \mathbf{t}\}}. \quad (2.3.14)$$

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

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

$$\mathbb{E}[N_n(\mathbf{t})] = \sum_{u \in [n]} \mathbb{P}(B_r^{(G_n)}(u) \simeq \mathbf{t}) = n\mathbb{P}(B_r^{(G_n)}(1) \simeq \mathbf{t}), \quad (2.3.15)$$

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 a tree in Section 1.5. Let $v_i \in [n]$ denote the vertex label of the i th vertex that is explored in the breadth-first exploration. Let X_i denote the number of forward neighbors of v_i , except when v_i is at graph distance r from vertex 1, in which case $X_i = 0$ by convention. Further, let Y_i denote the number of edges to already found, but not yet explored, vertices. Then, $B_r^{(G_n)}(1) \simeq \mathbf{t}$ occurs precisely when $(X_i, Y_i) = (x_i, 0)$ for all $i \in [t]$. Therefore,

$$\begin{aligned} \mathbb{P}(B_r^{(G_n)}(1) \simeq \mathbf{t}) &= \mathbb{P}((X_i, Y_i) = (x_i, 0) \forall i \in [t]) \\ &= \mathbb{P}((X_{[t]}, Y_{[t]}) = (x_{[t]}, 0_{[t]}) \forall i \in [t]), \end{aligned} \quad (2.3.16)$$

where we abbreviate $x_{[t]} = (x_1, \dots, x_t)$. We condition to write this as

$$\mathbb{P}(B_r^{(G_n)}(1) \simeq \mathbf{t}) = \prod_{i=1}^t \mathbb{P}((X_i, Y_i) = (x_{[i]}, 0_{[i]}) \mid (X_{[i-1]}, Y_{[i-1]}) = (x_{[i-1]}, 0_{[i-1]})). \quad (2.3.17)$$

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,

$$X_i \sim \text{Bin}(n_i, \lambda/n), \quad (2.3.18)$$

where $n_i = n - s_i - i - 1$, and $X_i = 0$ otherwise. Further, for all $i \in [t]$,

$$Y_i \sim \text{Bin}(s_{i-1}, \lambda/n), \quad (2.3.19)$$

and Y_i are conditionally independent. Note that the distance of v_i from vertex 1 is exactly equal to the distance of the corresponding vertex $a_i \in V(\mathbf{t})$ to the root $\emptyset \in V(\mathbf{t})$. Therefore,

$$\begin{aligned} \mathbb{P}(B_r^{(G_n)}(1) \simeq \mathbf{t}) &= \prod_{i \in [t]: \text{dist}(\emptyset, a_i) < r} \mathbb{P}(\text{Bin}(n_i, \lambda/n) = x_i) \left(1 - \frac{\lambda}{n}\right)^{s_i} \\ &\rightarrow \prod_{i \in [t]: \text{dist}(\emptyset, a_i) < r} e^{-\lambda} \frac{\lambda^{x_i}}{x_i!} = \mu(B_r(o) \simeq \mathbf{t}). \end{aligned} \quad (2.3.20)$$

We conclude that

$$\mathbb{P}(B_r^{(G_n)}(1) \simeq \mathbf{t}) \rightarrow \mu(B_r(o) \simeq \mathbf{t}), \quad (2.3.21)$$

so that

$$\frac{1}{n} \mathbb{E}[N_n(\mathbf{t})] \rightarrow \mu(B_r(o) \simeq \mathbf{t}). \quad (2.3.22)$$

For the second moment of $N_n(\mathbf{t})$, we compute

$$\begin{aligned} \mathbb{E}[N_n(\mathbf{t})^2] &= \sum_{u_1, u_2 \in [n]} \mathbb{P}(B_r^{(G_n)}(u_1) \simeq \mathbf{t}, B_r^{(G_n)}(u_2) \simeq \mathbf{t}) \\ &= n \mathbb{P}(B_r^{(G_n)}(1) \simeq \mathbf{t}) + n(n-1) \mathbb{P}(B_r^{(G_n)}(1) \simeq \mathbf{t}, B_r^{(G_n)}(2) \simeq \mathbf{t}). \end{aligned} \quad (2.3.23)$$

We have already computed the asymptotics of the first term, so we are left with the second term. We claim that

$$\mathbb{P}(B_r^{(G_n)}(1) \simeq \mathbf{t}, B_r^{(G_n)}(2) \simeq \mathbf{t}) \rightarrow \mu(B_r(o) \simeq \mathbf{t})^2, \quad (2.3.24)$$

which then implies that $\text{Var}(N_n(\mathbf{t}))/\mathbb{E}[N_n(\mathbf{t})]^2 \rightarrow 0$, so that, by the Chebychev inequality,

$$\frac{N_n(\mathbf{t})}{\mathbb{E}[N_n(\mathbf{t})]} \xrightarrow{\mathbb{P}} 1, \quad (2.3.25)$$

In turn, by (2.3.22), this implies that

$$p^{(n)}(\mathbf{t}) = \frac{1}{n} N_n(\mathbf{t}) \xrightarrow{\mathbb{P}} \mu(B_r(o) \simeq \mathbf{t}), \quad (2.3.26)$$

as required.

We are left to show (2.3.24). 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

$$\begin{aligned} & \mathbb{P}(B_r^{(G_n)}(1) \simeq \mathbf{t}, B_r^{(G_n)}(2) \simeq \mathbf{t}) \\ &= \mathbb{P}(B_r^{(G_n)}(1) \simeq \mathbf{t}, B_r^{(G_n)}(2) \simeq \mathbf{t}, \text{dist}_{G_n}(1, 2) > 2r) \\ & \quad + \mathbb{P}(B_r^{(G_n)}(1) \simeq \mathbf{t}, B_r^{(G_n)}(2) \simeq \mathbf{t}, \text{dist}_{G_n}(1, 2) \leq 2r). \end{aligned} \quad (2.3.27)$$

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 (2.3.27) can be bounded by

$$\begin{aligned} & \mathbb{P}(B_r^{(G_n)}(1) \simeq \mathbf{t}, B_r^{(G_n)}(2) \simeq \mathbf{t}, \text{dist}_{G_n}(1, 2) \leq 2r) \\ &= \mathbb{E} \left[\mathbb{1}_{\{B_r^{(G_n)}(1) \simeq \mathbf{t}\}} \frac{|B_{2r}^{(G_n)}(1)| - 1}{n-1} \right] \\ &\leq \frac{1}{n} \mathbb{E}[|B_{2r}^{(G_n)}(1)|]. \end{aligned} \quad (2.3.28)$$

Write

$$\mathbb{E}[|B_{2r}^{(G_n)}(1)|] = \sum_{k=0}^{2r} \mathbb{E}[|\partial B_k^{(G_n)}(1)|]. \quad (2.3.29)$$

It is not hard to show that (see Exercise 2.16)

$$\mathbb{E}[|\partial B_k^{(G_n)}(1)|] \leq \lambda^k. \quad (2.3.30)$$

This implies that

$$\mathbb{P}(B_r^{(G_n)}(1) \simeq \mathbf{t}, B_r^{(G_n)}(2) \simeq \mathbf{t}, \text{dist}_{G_n}(1, 2) \leq 2r) = o(1). \quad (2.3.31)$$

We rewrite the first term in in (2.3.27) as

$$\mathbb{P}(B_r^{(G_n)}(1) \simeq \mathbf{t}, \text{dist}_{G_n}(1, 2) > 2r) \mathbb{P}(B_r^{(G_n)}(2) \simeq \mathbf{t} \mid B_r^{(G_n)}(1) \simeq \mathbf{t}, \text{dist}_{G_n}(1, 2) > 2r). \quad (2.3.32)$$

By (2.3.21) and (2.3.31),

$$\mathbb{P}(B_r^{(G_n)}(1) \simeq \mathbf{t}, \text{dist}_{G_n}(1, 2) > 2r) \rightarrow \mu(B_r(o) \simeq \mathbf{t}). \quad (2.3.33)$$

Further, recall that $t = |V(\mathbf{t})|$ denotes the number of vertices in \mathbf{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 $\mathbb{P}(B_r^{(G_n)}(2) \simeq \mathbf{t} \mid B_r^{(G_n)}(1) \simeq \mathbf{t}, \text{dist}_{G_n}(1, 2) > 2r)$ equals the probability that $B_r^{(G_n)}(2) \simeq \mathbf{t}$ in $\text{ER}_{n-t}(\lambda/n)$. Since t is bounded, and as in (2.3.21),

$$\mathbb{P}(B_r^{(G_n)}(2) \simeq \mathbf{t} \mid B_r^{(G_n)}(1) \simeq \mathbf{t}, \text{dist}_{G_n}(1, 2) > 2r) \rightarrow \mu(B_r(o) \simeq \mathbf{t}). \quad (2.3.34)$$

Equations (2.3.33)–(2.3.34) complete the proof of local weak convergence in probability of $\text{ER}_n(\lambda/n)$ to a $\text{Poi}(\lambda)$ branching process. \square

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 $N_n(\mathbf{t})$ with explicit computations that allow us to show the appropriate adaptation of (2.3.26).

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 2.12 (Weak convergence of neighborhood sizes) *Let $G_n = ([n], E(G_n))$ denote a finite (possibly disconnected) random graph. Assume that G_n converges in distribution in the local weak sense to (G, o) having law $\bar{\mu}$ on \mathcal{G}_* . Then, for every $m \geq 1$,*

$$\left(|\partial B_{o_n}^{(n)}(k)|\right)_{k=0}^m \xrightarrow{d} \left(|\partial B_r(o)|\right)_{k=1}^m. \tag{2.4.1}$$

Further, assume that G_n converges in probability in the local weak sense to (G, o) having law μ on \mathcal{G}_ . Then, for every $m \geq 1$, with $o_n^{(1)}, o_n^{(2)}$ two independent uniformly chosen vertices in $[n]$,*

$$\left(\left(|\partial B_k^{(G_n)}(o_n^{(1)})|, |\partial B_k^{(G_n)}(o_n^{(2)})|\right)\right)_{k=0}^m \xrightarrow{d} \left(\left(|\partial B_k^{(G)}(o^{(1)})|, |\partial B_k^{(G)}(o^{(2)})|\right)\right)_{k=1}^m, \tag{2.4.2}$$

where the two limiting neighborhood sizes are independent given μ .

Proof This follows immediately, since the function

$$h(G, o) = \mathbb{1}_{\{|\partial B_r(o)| = \ell_k \forall k \leq m\}} \tag{2.4.3}$$

is a bounded continuous function for every m and ℓ_1, \dots, ℓ_m (see Exercise 2.19). The proof of (2.4.2) is Exercise 2.20. Here, it is crucial to note that when μ is a random probability measure on \mathcal{G}_* , that the limits in (2.4.2) are corresponding to two independent (G, o) having law μ , but with the *same* μ . It is here that the randomness of μ manifests itself. Recall also the example below Corollary 2.8. \square

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

Corollary 2.13 (Large distances) *Let $G_n = ([n], E(G_n))$ denote a finite (possibly disconnected) random graph. Let (G, o) be a random variable on \mathcal{G}_* having law μ . Let o_1, o_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 (G, o) . Then,*

$$\text{dist}_{G_n}(o_1, o_2) \xrightarrow{\mathbb{P}} \infty. \tag{2.4.4}$$

Proof It suffices to prove that, for every $m \geq 1$,

$$\mathbb{P}(\text{dist}_{G_n}(o_1, o_2) \leq m) = o(1). \tag{2.4.5}$$

For this, we use that o_2 is chosen uniformly at random from $[n]$ independently of o_1 , so that

$$\mathbb{P}(\text{dist}_{G_n}(o_1, o_2) \leq m) = \mathbb{E}[|B_m^{(G_n)}(o_1)|/n]. \quad (2.4.6)$$

By Corollary 2.12, $|B_m^{(G_n)}(o_1)|$ is a tight random variable, so that $|B_m^{(G_n)}(o_1)|/n \xrightarrow{\mathbb{P}} 0$. Further, $|B_m(o_1)|/n \leq 1$. Thus, by dominated convergence ([Volume 1, Theorem A.1]), $\mathbb{E}[|B_m^{(G_n)}(o_1)|/n] = o(1)$ for every $m \geq 1$, 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 $G_n = ([n], E_n)$, we let

$$W_{G_n} = \sum_{1 \leq i, j, k \leq n} \mathbb{1}_{\{ij, jk \in E_n\}} = \sum_{v \in [n]} D_v(D_v - 1) \quad (2.4.7)$$

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 (2.4.7). We further let

$$\Delta_{G_n} = \sum_{1 \leq i, j, k \leq n} \mathbb{1}_{\{ij, jk, ik \in E_n\}} \quad (2.4.8)$$

denote six times the number of triangles in G_n . The clustering coefficient CC_{G_n} in G_n is defined as

$$\text{CC}_{G_n} = \frac{\Delta_{G_n}}{W_{G_n}}. \quad (2.4.9)$$

Informally, the 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 2.14 (Convergence of global clustering coefficient) *Let $(G_n)_{n \geq 1}$ be a sequence of rooted graphs. Let (G, o) be a random variable on \mathcal{G}_* having law μ . Assume that G_n converges in probability in the local weak sense to (G, o) . Further, assume that $D_n = d_{o_n}^{(n)}$ is such that D_n^2 is uniformly integrable, and $\mu(d_o > 1) > 0$. Then*

$$\text{CC}_{G_n} \xrightarrow{\mathbb{P}} 3 \frac{\mathbb{E}_\mu[\Delta_\mu(o)]}{\mathbb{E}_\mu[d_o(d_o - 1)]}, \quad (2.4.10)$$

where $\Delta_\mu(o) = \sum_{u, v \in \partial B_1(o)} \mathbb{1}_{\{\{u, v\} \in E(G)\}}$ denotes the number of triangles that contain o as a vertex.

Proof We write

$$\text{CC}_{G_n} = 3 \frac{\mathbb{E}_n[\Delta_{G_n}(o_n)]}{\mathbb{E}_n[d_{o_n}^{(n)}(d_{o_n}^{(n)} - 1)]}, \quad (2.4.11)$$

where the expectation is with respect to the uniform choice of $o_n \in [n]$, and D_{o_n} denotes the degree of o_n in G_n , while $\Delta_{G_n}(o_n) = \sum_{u,v \in \partial B_1^{(G_n)}(o_n)} \mathbb{1}_{\{\{u,v\} \in E(G_n)\}}$ denotes twice the number of triangles that o_n is part of.

By local weak convergence, $\Delta_{G_n}(o_n) \xrightarrow{d} \Delta_\mu(o)$ and $d_{o_n}^{(n)}(d_{o_n}^{(n)} - 1) \xrightarrow{d} d_o(d_o - 1)$. 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 $D_n = d_{o_n}^{(n)}$. We split

$$\begin{aligned} & \mathbb{E}_n[d_{o_n}^{(n)}(d_{o_n}^{(n)} - 1)] \\ &= \mathbb{E}_n[d_{o_n}^{(n)}(d_{o_n}^{(n)} - 1) \mathbb{1}_{\{(d_{o_n}^{(n)})^2 \leq K\}}] + \mathbb{E}_n[d_{o_n}^{(n)}(d_{o_n}^{(n)} - 1) \mathbb{1}_{\{(d_{o_n}^{(n)})^2 > K\}}]. \end{aligned} \quad (2.4.12)$$

By local weak convergence in probability (recall Corollary 2.12),

$$\mathbb{E}_n[d_{o_n}^{(n)}(d_{o_n}^{(n)} - 1) \mathbb{1}_{\{(d_{o_n}^{(n)})^2 \leq K\}}] \xrightarrow{\mathbb{P}} \mathbb{E}_\mu[d_o(d_o - 1) \mathbb{1}_{\{d_o^2 \leq K\}}], \quad (2.4.13)$$

since $h(G, o) = d_o(d_o - 1) \mathbb{1}_{\{d_o^2 \leq K\}}$ is a bounded continuous function. Further, by uniform integrability of $(d_{o_n}^{(n)})^2$ and with \mathbb{E} denoting the expectation with respect to o_n as well as the random graph, for every $\varepsilon > 0$, there exists an $N = N(\varepsilon)$ sufficiently large such that, uniformly in $n \geq N(\varepsilon)$.

$$\mathbb{E}[d_{o_n}^{(n)}(d_{o_n}^{(n)} - 1) \mathbb{1}_{\{(d_{o_n}^{(n)})^2 > K\}}] \leq \mathbb{E}[(d_{o_n}^{(n)})^2 \mathbb{1}_{\{(d_{o_n}^{(n)})^2 > K\}}] \leq \varepsilon^2. \quad (2.4.14)$$

Then, by the Markov inequality,

$$\mathbb{P}\left(\mathbb{E}_n[d_{o_n}^{(n)}(d_{o_n}^{(n)} - 1) \mathbb{1}_{\{(d_{o_n}^{(n)})^2 > K\}}] \geq \varepsilon\right) \leq \frac{1}{\varepsilon} \mathbb{E}[(d_{o_n}^{(n)})^2 \mathbb{1}_{\{(d_{o_n}^{(n)})^2 > K\}}] \leq \varepsilon. \quad (2.4.15)$$

It follows that $\mathbb{E}_n[d_{o_n}^{(n)}(d_{o_n}^{(n)} - 1)] \xrightarrow{\mathbb{P}} \mathbb{E}_\mu[d_o(d_o - 1)]$, as required. Since $\mu(d_o > 1) > 0$, also $\mathbb{E}_\mu[d_o(d_o - 1)] > 0$, so that also $1/\mathbb{E}_n[d_{o_n}^{(n)}(d_{o_n}^{(n)} - 1)] \xrightarrow{\mathbb{P}} 1/\mathbb{E}_\mu[d_o(d_o - 1)]$.

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

$$\mathbb{E}_n[\Delta_{G_n}(o_n)] = \mathbb{E}_n[\Delta_{G_n}(o_n) \mathbb{1}_{\{(d_{o_n}^{(n)})^2 \leq K\}}] + \mathbb{E}_n[\Delta_{G_n}(o_n) \mathbb{1}_{\{(d_{o_n}^{(n)})^2 > K\}}]. \quad (2.4.16)$$

Then, we observe that since $\Delta_{G_n}(o_n) \leq d_{o_n}^{(n)}(d_{o_n}^{(n)} - 1)$, 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 $\mathbb{E}[(d_{o_n}^{(n)})^2 \mathbb{1}_{\{(d_{o_n}^{(n)})^2 > K\}}]$, which can be treated as above. This completes the proof. \square

We see that in order to obtain convergence of the clustering coefficient, we need an additional uniform integrability condition on the degree distribution, where more precisely, we need that D_n^2 is uniformly integrable, with $D_n = d_{o_n}^{(n)}$ 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

$$\text{CC}_{G_n}(v) = 3 \frac{\Delta_{G_n}(v)}{d_v(d_v - 1)}, \quad (2.4.17)$$

where $\Delta_{G_n}(v) = \sum_{s,t \in \partial B_1^{(G_n)}(v)} \mathbb{1}_{\{\{s,t\} \in E(G_n)\}}$ is the total number of triangles that

v is part of. Then, we let

$$\overline{\text{CC}}_{G_n} = \frac{1}{n} \sum_{v \in [n]} \text{CC}_{G_n}(v) \quad (2.4.18)$$

be the local clustering coefficient. The following theorem implies its convergence:

Theorem 2.15 (Convergence of local clustering coefficient) *Let $(G_n)_{n \geq 1}$ be a sequence of rooted graphs. Let (G, o) be a random variable on \mathcal{G}_* having law μ . Assume that G_n converges in probability in the local weak sense to (G, o) . Then*

$$\overline{\text{CC}}_{G_n} \xrightarrow{\mathbb{P}} 3\mathbb{E}_\mu \left[\frac{\Delta_\mu(o)}{d_o(d_o - 1)} \right]. \quad (2.4.19)$$

Proof We now write

$$\overline{\text{CC}}_{G_n} = 3\mathbb{E}_n \left[\frac{\Delta_{G_n}(o_n)}{d_{o_n}^{(n)}(d_{o_n}^{(n)} - 1)} \right], \quad (2.4.20)$$

and note that $h(G, o) = \Delta_G(o)/[d_o(d_o - 1)]$ is a bounded continuous functional. Therefore,

$$\mathbb{E}_n \left[\frac{\Delta_{G_n}(o_n)}{d_{o_n}^{(n)}(d_{o_n}^{(n)} - 1)} \right] \xrightarrow{\mathbb{P}} \mathbb{E}_\mu \left[\frac{\Delta_\mu(o)}{d_o(d_o - 1)} \right], \quad (2.4.21)$$

as required. \square

There are more versions of the clustering coefficient. A discussion of the convergence of the 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. 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.

Let us start 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 $\underline{e} = u$ and $\bar{e} = v$ for the starting and endpoint of e . The *degree-degree distribution* is given by

$$p_{k,l}^{(n)} = \frac{1}{2|E(G_n)|} \sum_e \mathbb{1}_{\{d_{\underline{e}}=k, d_{\bar{e}}=l\}}. \quad (2.4.22)$$

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

Theorem 2.16 (Degree-degree convergence) *Let $(G_n)_{n \geq 1}$ be a sequence of rooted graphs. Let (G, o) be a random variable on \mathcal{G}_* having law μ . Assume that G_n*

converges in probability in the local weak sense to (G, o) , and that $\mu(d_o \geq 1) > 0$. Then, for every k, l with $k \geq 1$,

$$p_{k,l}^{(n)} \xrightarrow{\mathbb{P}} k\mu(d_o = k, d_V = l), \tag{2.4.23}$$

where V is a neighbor of o chosen uniformly at random.

Proof We recall (2.4.22), and note that

$$\frac{2}{n}|E(G_n)| = \frac{1}{n} \sum_{v \in [n]} d_v = \mathbb{E}_n[d_{o_n}]. \tag{2.4.24}$$

Therefore, by local weak convergence in probability

$$\frac{2}{n}|E(G_n)| \xrightarrow{\mathbb{P}} \mathbb{E}_\mu[d_o]. \tag{2.4.25}$$

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

Further, we rewrite

$$\begin{aligned} \frac{1}{n} \sum_e \mathbb{1}_{\{d_e=k, d_{\bar{e}}=l\}} &= k \frac{1}{n} \sum_{u \in [n]} \mathbb{1}_{\{d_u=k\}} \left(\frac{1}{k} \sum_{v: v \sim u} \mathbb{1}_{\{d_u=l\}} \right) \\ &= k \mathbb{E}_n \left[\mathbb{1}_{\{d_{o_n}=k, d_V=l\}} \right], \end{aligned} \tag{2.4.26}$$

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 of (d_o, d_V) conditionally on (G, o) is a deterministic function of $B_2(o_n)$,

$$\frac{1}{n} \sum_e \mathbb{1}_{\{d_e=k, d_{\bar{e}}=l\}} \xrightarrow{\mathbb{P}} k\mu(d_o = k, d_V = l). \tag{2.4.27}$$

Taking the ratio proves the claim. □

We next discuss the consequences for the assortativity coefficient (recall [Volume 1, Section 1.5]). Define the *assortativity coefficient* as

$$\rho_{G_n} = \frac{\sum_{i,j \in [n]} (\mathbb{1}_{\{ij \in E'(G_n)\}} - d_i d_j / |E'(G_n)|) d_i d_j}{\sum_{i,j \in [n]} (d_i \mathbb{1}_{\{i=j\}} - d_i d_j / |E'(G_n)|) d_i d_j}, \tag{2.4.28}$$

where $E'(G_n)$ is the collection of directed edges. We can recognize ρ_{G_n} in (2.4.28) as the *empirical correlation coefficient* of the two-dimensional sequence of variables $(d_e, d_{\bar{e}})_{e \in E'(G_n)}$. As a result, it is the correlation between the coordinates of the two-dimensional random variable of which $(p_{k,l}^{(n)})_{k,l \geq 1}$ is the joint probability mass function. We can rewrite the assortativity coefficient ρ_G more conveniently as

$$\rho_{G_n} = \frac{\sum_{ij \in E'(G_n)} d_i d_j - (\sum_{i \in [n]} d_i^2)^2 / |E'(G_n)|}{\sum_{i \in [n]} d_i^3 - (\sum_{i \in [n]} d_i^2)^2 / |E'(G_n)|}. \tag{2.4.29}$$

The following theorem gives conditions for the convergence of ρ_{G_n} when G_n converges locally in distribution:

Theorem 2.17 (Assortativity convergence) *Let $(G_n)_{n \geq 1}$ be a sequence of rooted graphs. Let (G, o) be a random variable on \mathcal{G}_* having law μ . Assume that G_n converges in probability in the local weak sense to (G, o) , and that $\mu(d_o \geq 1) > 0$. Further, assume that $D_n = d_{o_n}^{(n)}$ is such that D_n^3 is uniformly integrable. Then,*

$$\rho_{G_n} \xrightarrow{\mathbb{P}} \frac{\mathbb{E}_\mu[d_o^2 d_V] - \mathbb{E}_\mu[d_o^2]^2 / \mathbb{E}_\mu[d_o]}{\mathbb{E}_\mu[d_o^3] - \mathbb{E}_\mu[d_o^2]^2 / \mathbb{E}_\mu[d_o]}, \quad (2.4.30)$$

where V is a neighbor of o chosen uniformly at random.

Proof We start with (2.4.29), and consider the various terms. We divide all sums by n . Then, by local weak convergence,

$$\frac{1}{n} |E'(G_n)| = \mathbb{E}_n[d_{o_n}^{(n)}] \xrightarrow{\mathbb{P}} \mathbb{E}_\mu[d_o]. \quad (2.4.31)$$

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

$$\frac{1}{n} \sum_{i \in [n]} d_i^2 = \mathbb{E}_n[(d_{o_n}^{(n)})^2] \xrightarrow{\mathbb{P}} \mathbb{E}_\mu[d_o^2]. \quad (2.4.32)$$

Further, again by local weak convergence and uniform integrability of $(d_{o_n}^{(n)})^3$,

$$\frac{1}{n} \sum_{i \in [n]} d_i^3 = \mathbb{E}_n[(d_{o_n}^{(n)})^3] \xrightarrow{\mathbb{P}} \mathbb{E}_\mu[d_o^3]. \quad (2.4.33)$$

This identifies the limits of all but one of the sums appearing in (2.4.29). Details are left to the reader in Exercise 2.17.

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

$$\begin{aligned} \frac{1}{n} \sum_{ij \in E'(G_n)} d_i d_j &= \frac{1}{n} \sum_{u \in [n]} d_u^2 \left(\frac{1}{d_u} \sum_{v: v \sim u} d_v \right) \\ &= \mathbb{E}_n[d_{o_n}^2 d_V], \end{aligned} \quad (2.4.34)$$

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 that a truncation argument is needed.

We split

$$\frac{1}{n} \sum_{ij \in E'(G_n)} d_i d_j = \frac{1}{n} \sum_{ij \in E'(G_n)} d_i d_j \mathbb{1}_{\{d_i \leq K, d_j \leq K\}} + \frac{1}{n} \sum_{ij \in E'(G_n)} d_i d_j (1 - \mathbb{1}_{\{d_i \leq K, d_j \leq K\}}). \quad (2.4.35)$$

We now rewrite, similarly as above,

$$\frac{1}{n} \sum_{ij \in E'(G_n)} d_i d_j \mathbb{1}_{\{d_i \leq K, d_j \leq K\}} = \mathbb{E}_n[d_{o_n}^2 d_V \mathbb{1}_{\{d_{o_n} \leq K, d_V \leq K\}}]. \quad (2.4.36)$$

By local weak convergence (or by Theorem 2.16), since the functional is now bounded and continuous,

$$\frac{1}{n} \sum_{ij \in E'(G_n)} d_i d_j \mathbb{1}_{\{d_i \leq K, d_j \leq K\}} \xrightarrow{\mathbb{P}} \mathbb{E}_\mu [d_o^2 d_V \mathbb{1}_{\{d_{o_n} \leq K, d_V \leq K\}}]. \quad (2.4.37)$$

We are left to show that the second contribution in (2.4.35) is small. We bound this contribution as

$$\frac{1}{n} \sum_{ij \in E'(G_n)} d_i d_j (\mathbb{1}_{\{d_i > K\}} + \mathbb{1}_{\{d_j > K\}}) = \frac{2}{n} \sum_{ij \in E'(G_n)} d_i d_j \mathbb{1}_{\{d_i > K\}}. \quad (2.4.38)$$

We now use Cauchy-Schwarz to bound this as

$$\begin{aligned} \frac{1}{n} \sum_{ij \in E'(G_n)} d_i d_j (\mathbb{1}_{\{d_i > K\}} + \mathbb{1}_{\{d_j > K\}}) & \quad (2.4.39) \\ & \leq \frac{1}{n} \sqrt{\sum_{ij \in E'(G_n)} d_i^2 \mathbb{1}_{\{d_i > K\}}} \sqrt{\sum_{ij \in E'(G_n)} d_j^2} \\ & = \mathbb{E}_n [(d_{o_n}^{(n)})^3 \mathbb{1}_{\{d_{o_n}^{(n)} > K\}}]^{1/2} \mathbb{E}_n [(d_{o_n}^{(n)})^3]^{1/2}. \end{aligned}$$

By uniform integrability of $(d_{o_n}^{(n)})^3$, there exists $K = K(\varepsilon)$ and $N = N(\varepsilon)$ such that, for all $n \geq N$,

$$\mathbb{E}[(d_{o_n}^{(n)})^3 \mathbb{1}_{\{d_{o_n}^{(n)} > K\}}] \leq \varepsilon^2/2. \quad (2.4.40)$$

In turn, by the Markov inequality, this implies that

$$\mathbb{P}\left(\mathbb{E}_n [(d_{o_n}^{(n)})^3 \mathbb{1}_{\{d_{o_n}^{(n)} > K\}}] \geq \varepsilon\right) \leq \frac{1}{\varepsilon} \mathbb{E}[(d_{o_n}^{(n)})^3 \mathbb{1}_{\{d_{o_n}^{(n)} > K\}}] \leq \varepsilon/2. \quad (2.4.41)$$

As a result, whp,

$$\frac{1}{n} \sum_{ij \in E'(G_n)} d_i d_j (\mathbb{1}_{\{d_i > K\}} + \mathbb{1}_{\{d_j > K\}}) \leq \varepsilon. \quad (2.4.42)$$

This completes the proof. □

2.4.4 PAGERANK AND LOCAL WEAK CONVERGENCE

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 $d_i^{(\text{out})} \geq 1$ 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 $d_i^{(\text{out})}$. Fix $\alpha \in (0, 1)$. Then, we let the vector of PageRanks $(R_i)_{i \in [n]}$ be the unique solution to the equation

$$R_i = \alpha \sum_{j \rightarrow i} \frac{R_j}{d_j^{(\text{out})}} + 1 - \alpha, \quad (2.4.43)$$

satisfying the normalization

$$\sum_{i \in [n]} R_i = n. \quad (2.4.44)$$

The parameter $\alpha \in (0, 1)$ is called the *damping factor*, and guarantees that (2.4.43) 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 (2.4.43), namely

$$\pi_i = \alpha \sum_{j \rightarrow i} \frac{\pi_j}{d_j^{(\text{out})}} + \frac{1 - \alpha}{n}, \quad (2.4.45)$$

Therefore, $(\pi_i)_{i \in [n]}$ is the stationary distribution of a random walker that, with probability α 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 (2.4.43) becomes

$$R_i = \alpha \sum_{j \rightarrow i} \frac{R_j}{d_j^{(\text{out})}} + (1 - \alpha) \sum_{j \in \mathcal{D}} R_j + 1 - \alpha, \quad (2.4.46)$$

where $\mathcal{D} \subseteq [n]$ denotes the collection of dangling nodes (and, when $i \in \mathcal{D}$, the first term is zero by convention, and (2.4.46) reduces to (2.4.43)).

The damping factor α 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 α 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 (possibly directed) graph sequence $(G_n)_{n \geq 1}$ that converges locally weakly:

Theorem 2.18 (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 local weak sense, then there exists a limiting distribution R_\emptyset , with $\mathbb{E}[R_\emptyset] \leq 1$, such that*

$$R_{o_n}^{(G_n)} \xrightarrow{d} R_\emptyset. \quad (2.4.47)$$

- (ii) *If G_n converges in probability in the local weak sense, then there exists a limiting distribution R_\emptyset , with $\mathbb{E}[R_\emptyset] \leq 1$, such that, for every $r > 0$,*

$$\frac{1}{n} \sum_{v \in [n]} \mathbb{1}_{\{R_v^{(G_n)} > r\}} \xrightarrow{\mathbb{P}} \mathbb{P}(R_\emptyset > r). \quad (2.4.48)$$

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. Clearly, the size of the largest connected component $|\mathcal{C}_{\max}|$ is *not* continuous in the local weak convergence sense (see Exercise 2.21), as it is a *global* object. However, local weak convergence still tells us a useful story about the existence of a giant, as well as its size:

Corollary 2.19 (Upper bound on the giant) *Let $G_n = ([n], E(G_n))$ denote a finite (possibly disconnected) random graph. Let (G, o) be a random variable on \mathcal{G}_* having law μ . Assume that G_n converges in probability in the local weak sense to (G, o) having distribution μ , and write $\zeta = \mu(|\mathcal{C}(o)| = \infty)$ for the survival probability of the limiting graph (G, o) . Then, for every $\varepsilon > 0$ fixed,*

$$\mathbb{P}(|\mathcal{C}_{\max}| \leq n(\zeta + \varepsilon)) \rightarrow 1. \quad (2.5.1)$$

In particular, Corollary 2.19 implies that $|\mathcal{C}_{\max}|/n \xrightarrow{\mathbb{P}} 0$ when $\zeta = 0$ (see Exercise 2.22).

Proof Define

$$Z_{\geq k} = \sum_{v \in [n]} \mathbb{1}_{\{|\mathcal{C}(v)| \geq k\}}. \quad (2.5.2)$$

Assume that G_n converges in probability in the local weak sense to (G, o) . Then, we conclude that with $\zeta_{\geq k} = \mu(|\mathcal{C}(o)| \geq k)$ (see Exercise 2.23),

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

For every $k \geq 1$,

$$\{|\mathcal{C}_{\max}| \geq k\} = \{Z_{\geq k} \geq k\}, \quad (2.5.4)$$

and, on the event that $Z_{\geq k} \geq 1$, also $|\mathcal{C}_{\max}| \leq Z_{\geq k}$. Thus, if $\zeta_{\geq k} > 0$ for every k , we obtain that, for every $k \geq 1$ and every $\varepsilon > 0$, and all n large enough,

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

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

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

□

We conclude that while local weak convergence cannot determine the size of the largest connected component, it *can* prove an upper bound on $|\mathcal{C}_{\max}|$. In this book, we will often extend this to $|\mathcal{C}_{\max}|/n \xrightarrow{\mathbb{P}} \zeta$, but this is no longer a consequence of local weak convergence directly. In Exercise 2.24, you are asked to give an example where $|\mathcal{C}_{\max}|/n \xrightarrow{\mathbb{P}} \eta < \zeta$. Therefore, in general, more involved arguments must be used. We next prove that one, relatively simple, condition suffices:

Theorem 2.20 (The giant is almost local) *Let $G_n = ([n], E(G_n))$ denote a finite (possibly disconnected) random graph. Let (G, o) be a random variable on \mathcal{G}_* having law μ . Assume that G_n converges in probability in the local weak sense to (G, o) and write $\zeta = \mu(|\mathcal{C}(o)| = \infty)$ for the survival probability of the limiting graph (G, o) . Assume that $\zeta > 0$ and that*

$$\lim_{k \rightarrow \infty} \limsup_{n \rightarrow \infty} \frac{1}{n^2} \mathbb{E} \left[\#\{x, y \in [n] : |\mathcal{C}(x)|, |\mathcal{C}(y)| \geq k, x \not\leftrightarrow y\} \right] = 0. \quad (2.5.7)$$

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

$$\frac{|\mathcal{C}_{\max}|}{n} \xrightarrow{\mathbb{P}} \zeta, \quad \frac{|\mathcal{C}_{(2)}|}{n} \xrightarrow{\mathbb{P}} 0. \quad (2.5.8)$$

Theorem 2.20 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 2.26.

Theorem 2.20 will be useful when we can easily show that vertices with large clusters are likely to be connected. There are many almost equivalent choices for the condition in (2.5.7), at least when the local weak limit (G, o) is a (homogeneous) branching process. Indeed, we may also consider x, y for which the diameters of $\mathcal{C}(x)$ and $\mathcal{C}(y)$ exceed k , or $|\partial B_k(x)|, |\partial B_k(y)| \geq k$. See Exercise ??.

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

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

Lemma 2.21 (Convergence of sum of squares of cluster sizes) *Under the conditions of Theorem 2.20,*

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

and

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

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

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

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

$$\frac{1}{n} \sum_{i \geq 1} |\mathcal{C}_{(i)}| \mathbb{1}_{\{|\mathcal{C}_{(i)}| \geq k\}} = \zeta + o_{k,\mathbb{P}}(1). \quad (2.5.13)$$

Further, by our main assumption in (2.5.7),

$$\begin{aligned} \frac{1}{n^2} \#\{x, y \in [n]: |\mathcal{C}(x)|, |\mathcal{C}(y)| \geq k, x \not\leftrightarrow y\} \\ = \frac{1}{n^2} \sum_{\substack{i, j \geq 1 \\ i \neq j}} |\mathcal{C}_{(i)}| |\mathcal{C}_{(j)}| \mathbb{1}_{\{|\mathcal{C}_{(i)}|, |\mathcal{C}_{(j)}| \geq k\}} = o_{k,\mathbb{P}}(1), \end{aligned} \quad (2.5.14)$$

by the Markov inequality. We conclude that

$$\begin{aligned} \frac{1}{n^2} \sum_{i \geq 1} |\mathcal{C}_{(i)}|^2 \mathbb{1}_{\{|\mathcal{C}_{(i)}| \geq k\}} &= \left(\frac{1}{n} \sum_{i \geq 1} |\mathcal{C}_{(i)}| \mathbb{1}_{\{|\mathcal{C}_{(i)}| \geq k\}} \right)^2 + o_{k,\mathbb{P}}(1) \\ &= \zeta^2 + o_{k,\mathbb{P}}(1), \end{aligned} \quad (2.5.15)$$

by (2.5.12). This proves (2.5.11). Finally,

$$\frac{1}{n^2} \sum_{i \geq 1} |\mathcal{C}_{(i)}| \mathbb{1}_{\{|\mathcal{C}_{(i)}| < k\}} \leq k \frac{1}{n^2} \sum_{i \geq 1} |\mathcal{C}_{(i)}| \mathbb{1}_{\{|\mathcal{C}_{(i)}| < k\}} \leq \frac{k}{n}, \quad (2.5.16)$$

so that

$$\frac{1}{n^2} \sum_{i \geq 1} |\mathcal{C}_{(i)}|^2 = \frac{1}{n^2} \sum_{i \geq 1} |\mathcal{C}_{(i)}|^2 \mathbb{1}_{\{|\mathcal{C}_{(i)}| \geq k\}} + O\left(\frac{k}{n}\right), \quad (2.5.17)$$

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

We are now ready to complete the proof of Theorem 2.20, and we start by explaining the ideas behind this. Define the (possibly random) probability measure $(q_{i,n})_{i \geq 1}$ by

$$q_{i,n} = \frac{|\mathcal{C}_{(i)}| \mathbb{1}_{\{|\mathcal{C}_{(i)}| \geq k\}}}{\sum_{j \geq 1} |\mathcal{C}_{(j)}| \mathbb{1}_{\{|\mathcal{C}_{(j)}| \geq k\}}}. \quad (2.5.18)$$

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

$$q_{1,n} = \frac{|\mathcal{C}_{\max}| \mathbb{1}_{\{|\mathcal{C}_{\max}| \geq k\}}}{\sum_{j \geq 1} |\mathcal{C}_{(j)}| \mathbb{1}_{\{|\mathcal{C}_{(j)}| \geq k\}}} = 1 + o_{k,\mathbb{P}}(1),$$

which, together with $\sum_{j \geq 1} |\mathcal{C}_{(j)}| \mathbb{1}_{\{|\mathcal{C}_{(j)}| \geq k\}} = n\zeta(1 + o_{k,\mathbb{P}}(1))$ implies that $|\mathcal{C}_{\max}|/n \xrightarrow{\mathbb{P}} \zeta$. Further, $q_{2,n} = o_{k,\mathbb{P}}(1)$, which implies that $|\mathcal{C}_{(2)}|/n \xrightarrow{\mathbb{P}} 0$. This is what we aim to prove. We now fill in the details:

Proof of Theorem 2.20. By Lemma 2.21,

$$\frac{1}{n^2} \sum_{i \geq 1} |\mathcal{C}_{(i)}|^2 \mathbb{1}_{\{|\mathcal{C}_{(i)}| \geq k\}} = \zeta^2 + o_{k,\mathbb{F}}(1). \quad (2.5.19)$$

Denote $x_i = |\mathcal{C}_{(i)}| \mathbb{1}_{\{|\mathcal{C}_{(i)}| \geq k\}}/n$. Then, by (2.5.12) and (2.5.19), for every $\varepsilon > 0$ and whp,

$$\sum_{i \geq 1} x_i \leq \zeta + \varepsilon/4, \quad \sum_{i \geq 1} x_i^2 \geq \zeta^2 - \varepsilon^2. \quad (2.5.20)$$

Now suppose by contradiction that $x_{1,n} = |\mathcal{C}_{(1)}| \mathbb{1}_{\{|\mathcal{C}_{(1)}| \geq k\}}/n \leq \zeta - \varepsilon$ for some $\varepsilon > 0$. Then

$$\max \sum_{i \geq 1} x_i^2 \leq (\zeta - \varepsilon)^2 + (5\varepsilon/4)^2 < \zeta(1 - \varepsilon), \quad (2.5.21)$$

where the maximum is taken over all vectors $(x_i)_{i \geq 1}$ such that $x_1 \leq \zeta - \varepsilon$ and $\sum_{i \geq 1} x_i \leq \zeta + \varepsilon/4$, and where $\varepsilon > 0$ is taken sufficiently small. However, (2.5.21) contradicts (2.5.20) when $\zeta > 0$ and $\varepsilon > 0$ is sufficiently small. We conclude that, whp, $x_{1,n} = |\mathcal{C}_{\max}| \mathbb{1}_{\{|\mathcal{C}_{\max}| \geq k\}}/n = |\mathcal{C}_{\max}|/n \geq \zeta - \varepsilon$. Combining this with Corollary 2.19, we obtain that $|\mathcal{C}_{\max}|/n \xrightarrow{\mathbb{P}} \zeta$, as required.

Similarly, suppose that $x_{2,n} = |\mathcal{C}_{(2)}| \mathbb{1}_{\{|\mathcal{C}_{(2)}| \geq k\}}/n \geq \varepsilon$. Then, since $\sum_{i \geq 1} x_i \leq \zeta + \varepsilon/4$ by (2.5.20), we must have that $x_1 \leq \zeta - 3\varepsilon/4$. However, by the previous argument, the probability that $x_1 = |\mathcal{C}_{\max}|/n \leq \zeta - 3\varepsilon/4$ vanishes. \square

We next extend Theorem 2.20 somewhat, and investigate the structure of vertices in the giant. For this, we let $v_k(\mathcal{C}_{\max})$ denote the number of vertices with degree k in the giant component, and we recall that $E(\mathcal{C}_{\max})$ denotes the number of edges in the giant component:

Theorem 2.22 (Properties of the giant) *Under the assumptions of Theorem 2.20,*

$$\frac{v_\ell(\mathcal{C}_{\max})}{n} \xrightarrow{\mathbb{P}} \mu(|\mathcal{C}(o)| = \infty, d_o = \ell). \quad (2.5.22)$$

Further, assume that $D_n = d_{o_n}^{(n)}$ is uniformly integrable. Then,

$$\frac{E(\mathcal{C}_{\max})}{n} \xrightarrow{\mathbb{P}} \frac{1}{2} \mathbb{E}_\mu \left[d_o \mathbb{1}_{\{|\mathcal{C}(o)| = \infty\}} \right]. \quad (2.5.23)$$

Proof We now define, for $k \geq 1$ and $A \subseteq \mathbb{N}$,

$$Z_{A, \geq k} = \sum_{v \in [n]} \mathbb{1}_{\{|\mathcal{C}(v)| \geq k, D_v \in A\}}. \quad (2.5.24)$$

Assume that G_n converges in probability in the local weak sense to (G, o) . Then, we conclude that with $\zeta_{A, \geq k} = \mu(|\mathcal{C}(o)| \geq k, d_o \in A)$,

$$\frac{Z_{A, \geq k}}{n} \xrightarrow{\mathbb{P}} \mu(|\mathcal{C}(o)| \geq k, d_o \in A). \quad (2.5.25)$$

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

$$\frac{1}{n} \sum_{a \in A} v_a(\mathcal{C}_{\max}) \leq \frac{Z_{A, \geq k}}{n} \xrightarrow{\mathbb{P}} \mu(|\mathcal{C}(o)| \geq k, d_o \in A). \quad (2.5.26)$$

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

$$\lim_{n \rightarrow \infty} \mathbb{P}\left(\frac{1}{n} [|\mathcal{C}_{\max}| - v_\ell(\mathcal{C}_{\max})] \leq \mu(|\mathcal{C}(o)| \geq k, d_o \neq \ell) + \varepsilon/2\right) = 1. \quad (2.5.27)$$

We argue by contradiction. Suppose that, for some ℓ ,

$$\liminf_{n \rightarrow \infty} \mathbb{P}\left(\frac{v_\ell(\mathcal{C}_{\max})}{n} \leq \mu(|\mathcal{C}(o)| = \infty, d_o = \ell) - \varepsilon\right) = \kappa > 0. \quad (2.5.28)$$

Then, along the subsequence $(n_i)_{i \geq 1}$ that attains the liminf in (2.5.28), with strictly positive probability, and using (2.5.27),

$$\frac{|\mathcal{C}_{\max}|}{n} = \frac{1}{n} [|\mathcal{C}_{\max}| - v_\ell(\mathcal{C}_{\max})] + \frac{v_\ell(\mathcal{C}_{\max})}{n} \leq \mu(|\mathcal{C}(o)| \geq k) - \varepsilon/2, \quad (2.5.29)$$

which contradicts Theorem 2.20. We conclude that (2.5.28) cannot hold, so that (2.5.22) follows.

For (2.5.23), we note that

$$E(\mathcal{C}_{\max}) = \frac{1}{2} \sum_{\ell \geq 1} \ell v_\ell(\mathcal{C}_{\max}). \quad (2.5.30)$$

We divide by n and split the sum over ℓ into small and large ℓ as

$$\frac{E(\mathcal{C}_{\max})}{n} = \frac{1}{2n} \sum_{\ell \in [K]} \ell v_\ell(\mathcal{C}_{\max}) + \frac{1}{2n} \sum_{\ell > K} \ell v_\ell(\mathcal{C}_{\max}). \quad (2.5.31)$$

For the first term in (2.5.31), by (2.5.22),

$$\begin{aligned} \frac{1}{2n} \sum_{\ell \in [K]} \ell v_\ell(\mathcal{C}_{\max}) &\xrightarrow{\mathbb{P}} \frac{1}{2} \sum_{\ell \in [K]} \mu(|\mathcal{C}(o)| = \infty, d_o = \ell) \\ &= \frac{1}{2} \mathbb{E}_\mu \left[d_o \mathbb{1}_{\{|\mathcal{C}(o)| = \infty, d_o \in [K]\}} \right]. \end{aligned} \quad (2.5.32)$$

For the second term in (2.5.31), we bound, with n_ℓ the number of vertices in G_n of degree ℓ ,

$$\frac{1}{2n} \sum_{\ell > K} \ell v_\ell(\mathcal{C}_{\max}) \leq \frac{1}{2} \sum_{\ell > K} \ell \frac{n_\ell}{n} = \mathbb{E}_n [d_{o_n}^{(n)} \mathbb{1}_{\{d_{o_n}^{(n)} > K\}}]. \quad (2.5.33)$$

By uniform integrability,

$$\lim_{K \rightarrow \infty} \limsup_{n \rightarrow \infty} \mathbb{E} [d_{o_n}^{(n)} \mathbb{1}_{\{d_{o_n}^{(n)} > K\}}] = 0. \quad (2.5.34)$$

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

$$\mathbb{P}\left(\mathbb{E}_n [d_{o_n}^{(n)} \mathbb{1}_{\{d_{o_n}^{(n)} > K\}}] > \varepsilon\right) \rightarrow 0. \quad (2.5.35)$$

This completes the proof of (2.5.23). \square

2.5.1 THE GIANT IN ERDŐS-RÉNYI RANDOM GRAPHS

In this section, we use the local weak convergence in probability in Theorem 2.11, combined with the fact that the ‘giant is almost local’ in Theorem 2.20, to identify the phase transition and size of the giant in the Erdős-Rényi random graph $\text{ER}_n(\lambda/n)$:

Theorem 2.23 (Phase transition Erdős-Rényi random graph) *Fix $\lambda > 0$, and let \mathcal{C}_{\max} be the largest connected component of the Erdős-Rényi random graph $\text{ER}_n(\lambda/n)$, and $\mathcal{C}_{(2)}$ is second largest connected component (breaking ties arbitrarily). Then,*

$$\frac{|\mathcal{C}_{\max}|}{n} \xrightarrow{\mathbb{P}} \zeta_\lambda, \quad \frac{|\mathcal{C}_{(2)}|}{n} \xrightarrow{\mathbb{P}} 0, \quad (2.5.36)$$

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

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

$$\frac{v_\ell(\mathcal{C}_{\max})}{n} \xrightarrow{\mathbb{P}} e^{-\lambda} \frac{\lambda^\ell}{\ell!} [1 - \eta_\lambda^\ell], \quad (2.5.37)$$

and

$$\frac{E(\mathcal{C}_{\max})}{n} \xrightarrow{\mathbb{P}} \frac{1}{2} \lambda [1 - \eta_\lambda^2]. \quad (2.5.38)$$

The law of large numbers for the giant in Theorem 2.23 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 2.20 holds. Indeed, local weak convergence in probability follows from Theorem 2.11. Then, the claim in (2.5.36) follows directly from Theorem 2.20, while (2.5.37)–(2.5.38) follow from Theorem 2.22, and the observations that, for a Poisson branching process with mean offspring $\lambda > 1$,

$$\mu(|\mathcal{C}(o)| = \infty, d_o = \ell) = e^{-\lambda} \frac{\lambda^\ell}{\ell!} [1 - \eta_\lambda^\ell], \quad \mathbb{E}_\mu \left[d_o \mathbb{1}_{\{|\mathcal{C}(o)| = \infty\}} \right] = \lambda [1 - \eta_\lambda^2]. \quad (2.5.39)$$

Therefore, we are left to show that (2.5.7) holds. This will be performed in several steps, and we start by massaging the condition in (2.5.7) in a more appropriate form.

The condition (2.5.7) revisited.

We start by rephrasing the condition in (2.5.7) in a form that is more convenient to deal with $\text{ER}_n(\lambda/n)$:

Lemma 2.24 (Condition (2.5.7) revisited) *Under the assumptions of Theorem 2.20, the condition in (2.5.7) holds when*

$$\lim_{r \rightarrow \infty} \limsup_{n \rightarrow \infty} \frac{1}{n^2} \mathbb{E} \left[\#\{x, y \in [n] : |\partial B_r(x)|, |\partial B_r(x)| \geq r, x \not\leftrightarrow y\} \right] = 0, \quad (2.5.40)$$

when there exists $r = r_k \rightarrow \infty$ such that

$$\mu(|\mathcal{C}(o)| \geq k, |\partial B_r(o)| < r) \rightarrow 0, \quad \mu(|\mathcal{C}(o)| < k, |\partial B_r(o)| \geq r) \rightarrow 0. \quad (2.5.41)$$

Proof Denote

$$P_r^{(2)} = \#\{x, y \in [n] : |\partial B_r(x)|, |\partial B_r(x)| \geq r, x \not\leftrightarrow y\}, \quad (2.5.42)$$

$$Q_k^{(2)} = \#\{x, y \in [n] : |\mathcal{C}(x)|, |\mathcal{C}(y)| \geq k, x \not\leftrightarrow y\}. \quad (2.5.43)$$

Then,

$$|P_r^{(2)} - Q_k^{(2)}| \leq 2n[Z_{<r, \geq k} + Z_{\geq r, <k}], \quad (2.5.44)$$

where

$$Z_{<r, \geq k} = \sum_{v \in [n]} \mathbb{1}_{\{|\partial B_r(x)| < r, |\mathcal{C}(v)| \geq k\}}, \quad Z_{\geq r, <k} = \sum_{v \in [n]} \mathbb{1}_{\{|\partial B_r(x)| \geq r, |\mathcal{C}(v)| < k\}}. \quad (2.5.45)$$

Therefore, by local weak convergence in probability,

$$\begin{aligned} \frac{1}{n^2} |P_r^{(2)} - Q_k^{(2)}| &\leq \frac{2}{n} [Z_{<r, \geq k} + Z_{\geq r, <k}] \\ &\xrightarrow{\mathbb{P}} 2\mu(|\mathcal{C}(o)| \geq k, |\partial B_r(o)| < r) + 2\mu(|\mathcal{C}(o)| < k, |\partial B_r(o)| \geq r). \end{aligned} \quad (2.5.46)$$

Take $r = r_k$ as in (2.5.41). Then, the right hand side vanishes, so that, by Dominated Convergence also

$$\lim_{k \rightarrow \infty} \limsup_{n \rightarrow \infty} \frac{1}{n^2} \mathbb{E}[|P_{r_k}^{(2)} - Q_k^{(2)}|] = 0. \quad (2.5.47)$$

We arrive at

$$\lim_{k \rightarrow \infty} \limsup_{n \rightarrow \infty} \frac{1}{n^2} \mathbb{E}[Q_k^{(2)}] \leq \lim_{k \rightarrow \infty} \limsup_{n \rightarrow \infty} \frac{1}{n^2} \mathbb{E}[P_{r_k}^{(2)}] = 0, \quad (2.5.48)$$

by (2.5.40) and since $r_k \rightarrow \infty$ when $k \rightarrow \infty$. \square

The assumption in (2.5.41) is often easily verified. For example, for the Erdős-Rényi random graph $\text{ER}_n(\lambda/n)$ with $\lambda > 1$, we can take $r = k$ and use that on the event of survival (recall [Volume 1, Theorem 3.9]),

$$\lambda^{-r} |\partial B_r(x)| \xrightarrow{a.s.} M, \quad (2.5.49)$$

where $M > 0$ by [Volume 1, Theorem 3.10]. Therefore, $\mu(|\mathcal{C}(o)| \geq k, |\partial B_k(x)| < k) \rightarrow 0$. Further, $\mu(|\mathcal{C}(o)| < k, |\partial B_r(x)| \geq r) = 0$ trivially. However, there are examples where (2.5.41) fails, and then also the equivalence of (2.5.7) and (2.5.40) may be false.

Concentration of binomials.

Note that

$$\begin{aligned} & \frac{1}{n^2} \mathbb{E} \left[\#\{x, y \in [n] : |\partial B_r(x)|, |\partial B_r(y)| \geq r, x \not\leftrightarrow y\} \right] \\ &= \mathbb{P}(|\partial B_r(o_1)|, |\partial B_r(o_2)| \geq r, o_1 \not\leftrightarrow o_2). \end{aligned} \quad (2.5.50)$$

We condition on $|\partial B_r(o_i)| = b_0^{(i)} \geq r$, and $|B_r(o_i)| = s_0^{(i)}$, so that

$$\begin{aligned} & \mathbb{P}(|\partial B_r(o_1)|, |\partial B_r(o_2)| \geq r, o_1 \not\leftrightarrow o_2) \\ &= \sum_{b_0^{(1)}, b_0^{(2)}, s_0^{(1)}, s_0^{(2)}} \mathbb{P}(o_1 \not\leftrightarrow o_2 \mid |\partial B_r(o_i)| = b_0^{(i)}, |B_r(o_i)| = s_0^{(i)}, i \in \{1, 2\}) \\ & \quad \times \mathbb{P}(|\partial B_r(o_i)| = b_0^{(i)}, |B_r(o_i)| = s_0^{(i)}, i \in \{1, 2\}). \end{aligned} \quad (2.5.51)$$

Our aim is to show that, for every $\varepsilon > 0$, we can find $r = r_\varepsilon$ such that, for every $b_0^{(1)}, b_0^{(2)} \geq r$ and $s_0^{(1)}, s_0^{(2)}$ fixed,

$$\limsup_{n \rightarrow \infty} \mathbb{P}(o_1 \not\leftrightarrow o_2 \mid |\partial B_r(o_i)| = r_i, |B_r(o_i)| = b_i, i \in \{1, 2\}) \leq \varepsilon. \quad (2.5.52)$$

Conditionally on $|\partial B_r(o_1)| = b_0^{(1)}, |\partial B_r(o_2)| = b_0^{(2)}, |B_{r-1}(o_1)| = s_0^{(1)}$ and $|B_{r-1}(o_2)| = s_0^{(2)}$,

$$|\partial B_{r+1}(o_1)| \sim \text{Bin}(n_r, p_r), \quad (2.5.53)$$

where

$$n_r^{(1)} = n - b_0^{(1)} - s_0^{(1)} - s_0^{(2)}, \quad p_r^{(1)} = 1 - \left(1 - \frac{\lambda}{n}\right)^{b_0^{(1)}}. \quad (2.5.54)$$

Here, we note that the vertices in $\partial B_r(o_2)$ play a special role, 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 (2.5.54). 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 2.25 (Concentration sums of Bernoulli's) *Let $X_i \sim \text{Be}(p_i)$, $i \in [n]$, be independent Bernoulli distributed random variables, and write $X = \sum_{i=1}^n X_i$. Then, for every $\delta > 0$,*

$$\mathbb{P}(|X - \mathbb{E}[X]| \geq \delta \mathbb{E}[X]) \leq 2 \exp\left(-\frac{\delta^2 \mathbb{E}[X]}{2(1 + \delta/3)}\right). \quad (2.5.55)$$

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

Lemma 2.25 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 $a \log_\lambda n$ 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 2.25 is still quite small; and (2) we have good control of 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 ε -dependence in (2.5.55).

We will prove (2.5.52) 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.

Neighborhood growth in Erdős-Rényi random graph: first vertex.

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

$$\bar{s}_k^{(i)} = s_0^{(i)} + \sum_{l=0}^{k-1} \bar{b}_l^{(i)} \quad (2.5.56)$$

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

$$\bar{s}_k^{(i)} \leq \frac{b_0^{(i)}}{1 - \lambda(1+\varepsilon)} n^{a(1+\varepsilon)/(1-\varepsilon)} \quad (2.5.57)$$

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

With the above choices, we note that, conditionally on $|\partial B_{r+k-1}(o_1)| = b_k^{(1)} \in [\underline{b}_k^{(1)}, \bar{b}_k^{(1)}]$ and $|\partial B_{r+k-1}(o_1)| = s_k^{(1)} \leq \bar{s}_k^{(1)}$,

$$|\partial B_{r+k}(o_1)| \sim \text{Bin}(n_k^{(1)}, p_k^{(1)}), \quad (2.5.58)$$

where

$$n_k^{(1)} = n - s_k^{(1)} - s_0^{(2)}, \quad p_k^{(1)} = 1 - \left(1 - \frac{\lambda}{n}\right)^{b_k^{(1)}}. \quad (2.5.59)$$

We aim to apply Lemma 2.25 to $|\partial B_{r+k}(o_1)|$ and with $\delta = \varepsilon/2$, for which it suffices to prove bounds on the (conditional) expectation $n_k^{(1)} p_k^{(1)}$. Using that

$$\frac{b_k^{(1)} \lambda}{n} - \frac{(b_k^{(1)} \lambda)^2}{2n^2} \leq p_k^{(1)} \leq \frac{b_k^{(1)} \lambda}{n}, \quad (2.5.60)$$

so that

$$n_k^{(1)} p_k^{(1)} \leq n \frac{b_k^{(1)} \lambda}{n} = \lambda b_k^{(1)}, \quad (2.5.61)$$

which provides the upper bound on $n_k^{(1)} p_k^{(1)}$. For the lower bound, we use the lower bound in (2.5.60) to note that $p_k^{(1)} \geq (1 - \varepsilon/4) \lambda b_k^{(1)}$. Further, $n_k^{(1)} \geq (1 - \varepsilon/4)n$ uniformly in $k \leq k_\varepsilon$. We conclude that

$$n_k^{(1)} p_k^{(1)} \geq (1 - \varepsilon/4)^2 n \frac{b_k^{(1)} \lambda}{n} \geq (1 - \varepsilon/2) \lambda b_k^{(1)}. \quad (2.5.62)$$

By Lemma 2.25 with $\delta = \varepsilon/2$, therefore, with \mathcal{F}_k denoting the σ -field generated by $(|\partial B_{r+l}(o_1)|)_{l=0}^k$,

$$\begin{aligned} & \mathbb{P}\left(|\partial B_{r+k}(o_1)| - \lambda b_{k-1}^{(1)} \geq (\varepsilon/2)\lambda b_{k-1}^{(1)} \mid \mathcal{F}_{k-1}\right) \\ & \leq \mathbb{P}\left(|\partial B_{r+k}(o_1)| - \mathbb{E}|\partial B_{r+k}(o_1)| \geq (\varepsilon/2)\mathbb{E}|\partial B_{r+k}(o_1)| \mid \mathcal{F}_{k-1}\right) \\ & \leq 2 \exp\left(-\frac{\varepsilon^2(1-\varepsilon/2)\lambda b_k^{(1)}}{8(1+\varepsilon/6)}\right) = 2 \exp(-q\lambda b_{k-1}^{(1)}), \end{aligned} \quad (2.5.63)$$

where $q = \varepsilon^2(1-\varepsilon/2)/[8(1+\varepsilon/6)] > 0$.

We conclude that

$$\mathbb{P}\left(\underline{b}_k^{(1)} \leq |\partial B_{r+k}(o_1)| \leq \bar{b}_k^{(1)} \quad \forall k \leq k_\varepsilon \mid \mathcal{F}_0\right) \geq 1 - 2 \sum_{l=1}^{k_\varepsilon-1} e^{-q\lambda \underline{b}_{k-1}^{(1)}}. \quad (2.5.64)$$

Neighborhood growth in Erdős-Rényi random graph: second vertex.

We next grow the neighborhoods from vertex 2 in a very similar way, and we focus on the differences only. In the whole argument below, we condition on $|\partial B_{r+l}(o_1)| = b_l^{(1)} \in [\underline{b}_l^{(1)}, \bar{b}_l^{(1)}]$ and $|\partial B_{r+l}(o_1)| = s_l^{(1)} \leq \bar{s}_l^{(1)}$ for all $l \leq k_\varepsilon$. We now note that, conditionally on the above, as well as on $|\partial B_{r+k-1}(o_2)| = b_k^{(2)} \in [\underline{b}_k^{(2)}, \bar{b}_k^{(2)}]$ and $|\partial B_{r+k-1}(o_2)| = s_k^{(2)} \leq \bar{s}_k^{(2)}$,

$$|\partial B_{r+k}(o_2)| \sim \text{Bin}(n_k^{(2)}, p_k^{(2)}), \quad (2.5.65)$$

where now

$$n_k^{(2)} = n - s_{k_\varepsilon-1}^{(1)} - s_k^{(2)}, \quad p_k^{(2)} = 1 - \left(1 - \frac{\lambda}{n}\right)^{b_k^{(2)}}. \quad (2.5.66)$$

Following the argument in the previous paragraph, we are lead to the conclusion that

$$\begin{aligned} & \mathbb{P}\left(\underline{b}_k^{(i)} \leq |\partial B_{r+k}(o_i)| \leq \bar{b}_k^{(i)} \quad \forall k \leq k_\varepsilon \mid \mathcal{F}_0\right) \\ & \geq 1 - 2 \sum_{l=1}^{k_\varepsilon-1} [e^{-q\lambda \underline{b}_{k-1}^{(1)}} + e^{-q\lambda \underline{b}_{k-1}^{(2)}}], \end{aligned} \quad (2.5.67)$$

where again $q = \varepsilon^2(1-\varepsilon/2)/[8(1+\varepsilon/6)] > 0$.

Now we are ready to complete the proof of Theorem 2.23:

Completion of the proof of Theorem 2.23. We use Lemma 2.5.40, and conclude that we need to show (2.5.52). Recall (2.5.67), and that $\underline{b}_k^{(i)} = b_0^{(i)}[\lambda(1-\varepsilon)]^k$, where $b_0^{(i)} \geq r = r_\varepsilon$. Take $r = r_\varepsilon$ so large that

$$\sum_{l=1}^{k_\varepsilon-1} [e^{-q\lambda \underline{b}_{k-1}^{(1)}} + e^{-q\lambda \underline{b}_{k-1}^{(2)}}] \leq \varepsilon/2. \quad (2.5.68)$$

Denote the good event by

$$\mathcal{E}_{r_\varepsilon, k_\varepsilon} = \{\underline{b}_k^{(i)} \leq |\partial B_{r_\varepsilon+k}(o_i)| \leq \bar{b}_k^{(i)} \ \forall k \leq k_\varepsilon, |\partial B_{r_\varepsilon}(o_i)| \geq r_\varepsilon\}. \quad (2.5.69)$$

On $\mathcal{E}_{r_\varepsilon, k_\varepsilon}$, we have that $|\partial B_{r_\varepsilon+k_\varepsilon}(o_i)| \geq \underline{b}_{k_\varepsilon}^{(i)} = b_0^{(i)}[\lambda(1-\varepsilon)]^{k_\varepsilon} \geq r_\varepsilon n^{a(1+\varepsilon)/(1-\varepsilon)}$, where we choose $a \in (\frac{1}{2}, 1)$ so that $a(1+\varepsilon)/(1-\varepsilon) \in (\frac{1}{2}, 1)$.

There are two possibilities. The first is that $B_{r_\varepsilon+k_\varepsilon}(o_1) \cap B_{r_\varepsilon+k_\varepsilon}(o_2) \neq \emptyset$, in which case $\text{dist}_{\text{ER}_n(\lambda/n)}(o_1, o_2) \leq 2(k_\varepsilon + r_\varepsilon) < \infty$. When, however, $B_{r_\varepsilon+k_\varepsilon}(o_1) \cap B_{r_\varepsilon+k_\varepsilon}(o_2) = \emptyset$, we use that Moreover,

$$|\partial B_{r_\varepsilon+k_\varepsilon}(o_i)| \geq \bar{b}_{k_\varepsilon}^{(i)} \geq r_\varepsilon [\lambda(1-\varepsilon)]^{k_\varepsilon} = r_\varepsilon n^a. \quad (2.5.70)$$

Therefore, the number of direct edges between $\partial B_{r_\varepsilon+k_\varepsilon}(o_1)$ and $\partial B_{r_\varepsilon+k_\varepsilon}(o_2)$ is at least $(r_\varepsilon n^a)^2 \gg n$ when $a > \frac{1}{2}$. Therefore,

$$\mathbb{P}\left(\text{dist}_{\text{ER}_n(\lambda/n)}(o_1, o_2) > 2(k_\varepsilon + r_\varepsilon) + 1 \mid \mathcal{E}_{r_\varepsilon, k_\varepsilon}\right) \leq \left(1 - \frac{\lambda}{n}\right)^{(r_\varepsilon n^a)^2} = o(1). \quad (2.5.71)$$

We conclude that

$$\mathbb{P}\left(\text{dist}_{\text{ER}_n(\lambda/n)}(o_1, o_2) \leq 2(k_\varepsilon + r_\varepsilon) + 1 \mid \mathcal{E}_{r_\varepsilon, k_\varepsilon}\right) = 1 - o(1). \quad (2.5.72)$$

□

In the above proof, we have also identifies 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 2.26 (Small-world nature Erdős-Rényi random graph) *Fix $\lambda > 1$. Then, conditionally on $o_1 \longleftrightarrow o_2$,*

$$\frac{\text{dist}_{\text{ER}_n(\lambda/n)}(o_1, o_2)}{\log n} \xrightarrow{\mathbb{P}} \frac{1}{\log \lambda}. \quad (2.5.73)$$

Proof The lower bound follows directly from (2.3.30), which implies that

$$\mathbb{P}(\text{dist}_{\text{ER}_n(\lambda/n)}(o_1, o_2) \leq k) = \mathbb{E}\left[|B_k(o_1)|/n\right] \leq \sum_{l=0}^k \frac{\lambda^l}{n} = \frac{\lambda^{k+1} - 1}{n(\lambda - 1)}. \quad (2.5.74)$$

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

$$\mathbb{P}(\text{dist}_{\text{ER}_n(\lambda/n)}(o_1, o_2) \leq (1 - \varepsilon) \log_\lambda n) \rightarrow 0 \quad (2.5.75)$$

For the upper bound, we start by noting that

$$\mathbb{P}(o_1 \longleftrightarrow o_2 \mid \text{ER}_n(\lambda/n)) = \frac{1}{n^2} \sum_{i \geq 1} |\mathcal{C}_{(i)}|^2 \xrightarrow{\mathbb{P}} \zeta_\lambda^2, \quad (2.5.76)$$

by (2.5.10) in Lemma 2.21. Thus, conditioning on $o_1 \longleftrightarrow o_2$ is asymptotically the same as $o_1, o_2 \in \mathcal{C}_{\max}$. The upper bound then follows from the fact that the event $\mathcal{E}_{r_\varepsilon, k_\varepsilon}$ in (2.5.69) holds with probability at least $1 - \varepsilon$, and, on the event $\mathcal{E}_{r_\varepsilon, k_\varepsilon}$,

$$\text{dist}_{\text{ER}_n(\lambda/n)}(o_1, o_2) \leq 2(k_\varepsilon + r_\varepsilon) + 1, \quad (2.5.77)$$

whp by (2.5.72). □

2.5.2 OUTLINE OF THE REMAINDER OF THIS BOOK

The results proved for the Erdős-Rényi random graph allow us to provide a brief glimpse into the content of the remainder of this book. There, we will prove results about local weak 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 2**, we focus on local weak convergence as in Theorem 2.11, but then applied to these models. Further, we investigate the size of the giant component as in Theorems 2.22 and 2.23. The proofs are often more involved than the corresponding results for the Erdős-Rényi random graph, since these random graphs models are inhomogeneous, and sometimes also lack the independence of the edge statuses. Therefore, the proofs will require detailed knowledge of the models in question.

In **Part 3**, we focus on the small-world nature of these random graph models, as in Theorem 2.26. 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 2.26. 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 τ 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 weak convergence results in Part 2 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 4**, we investigate some adaptations of the models discussed in Parts 2 and 3. 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 4 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. [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 [2.6](#) 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.

Notes on Section 2.3

We thank Shankar Bhamidi for useful discussions about possibly random limits in Definition [2.7](#). As far as we could tell, a precise definition of local weak convergence in its various senses has not appeared explicitly in the literature.

The analysis of $\mathbb{P}(B_r^{(G_n)}(1) \simeq \mathfrak{t})$ in [\(2.3.21\)](#) has already appeared in [Volume 1, Section [4.1.2](#)], see in particular [Volume 1, [\(4.1.12\)](#)]. We have redone the analysis here to keep the exposition self-contained.

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 [2.14](#) and [2.15](#) 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 the clustering of vertices of degree k to be

$$c_{G_n}(k) = \frac{1}{n_k} \sum_{v \in [n]: d_v^{(n)}=k} \frac{\Delta_{G_n}(v)}{k(k-1)}. \quad (2.6.1)$$

It is not hard to adapt the proof of Theorem [2.15](#) to show that, under its as-

sumptions, $c_{G_n}(k) \xrightarrow{\mathbb{P}} c_G(k)$, where

$$c_G(k) = \mathbb{E}_\mu \left[\frac{\Delta_G(o)}{k(k-1)} \mid d_o = k \right]. \quad (2.6.2)$$

See Exercise 2.30.

The convergence of the assortativity coefficient in Theorem 2.17 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), Litvak and van der Hofstad (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, ρ_G in (2.4.29) 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 Litvak and van der Hofstad (2013) prove that if a limiting value of ρ_G exists for a sequence of networks, then $\liminf_{n \rightarrow \infty} \rho_G \geq 0$, so no asymptotically disassortative graph sequences exist. Also, there is an example where the limit exists in distribution and the limit is a proper random variable. 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, \dots, x_n with ranks r_1, \dots, r_n , x_i is the r_i th largest of x_1, \dots, 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*

$$p_k(G) = \frac{1}{|V(G)|} \sum_{v \in V(G)} \mathbb{1}_{\{d_v(G)=k\}}, \quad (2.7.1)$$

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 defines as in (2.1.1). Show that $d_{\mathcal{G}_*}(B_r(o), (G, o)) \leq 1/(r + 1)$. 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 (\mathcal{G}_* is separable) Use Exercise 2.4 above to show that the set of rooted graphs \mathcal{G}_* 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 $H_* \in \mathcal{G}_*$. Show that $h: \mathcal{G}_* \mapsto \{0, 1\}$ given by $h(G, o) = \mathbb{1}_{\{B_r(o) \simeq H_*\}}$ 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 at most 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_n^* , the size-biased version of $D_n = d_{o_n}^{(n)}$, converges in distribution.

Exercise 2.12 (Uniform integrability and degree regularity condition) For $\text{CM}_n(\mathbf{d})$, 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 (Local weak limit of random 2-regular graph) Show that the configuration model $\text{CM}_n(\mathbf{d})$ with $d_i = 2$ for all $i \in [n]$ converges in the local weak sense in probability to \mathbb{Z} . Conclude that the same applies to the random 2-regular graph.

Exercise 2.14 (Independent neighborhoods of different vertices) Let G_n converge in probability in the local weak sense to (G, o) . Let $(o_n^{(1)}, o_n^{(2)})$ be two indepen-

dent uniformly chosen vertices in $[n]$. Show that $(G_n, o_n^{(1)})$ and $(G_n, o_n^{(2)})$ jointly converge to two independent copies of (G, o) .

Exercise 2.15 (Directed graphs as marked graphs) *There are several ways to describe directed graphs as marked graphs. Give one.*

Exercise 2.16 (Expected boundary of balls in Erdős-Rényi random graphs) *Prove that, for $\text{ER}_n(\lambda/n)$ and every $k \geq 0$*

$$\mathbb{E}[|\partial B_k^{(G_n)}(1)|] \leq \lambda^k. \quad (2.7.2)$$

This can be done, for example, by showing that, for every $k \geq 1$

$$\mathbb{E}[|\partial B_k^{(G_n)}(1)| \mid B_{k-1}^{(G_n)}(1)] \leq \lambda |\partial B_{k-1}^{(G_n)}(1)|, \quad (2.7.3)$$

and using induction.

Exercise 2.17 (Uniform integrability and moment convergence) *Assume that $D_n = d_{o_n}^{(n)}$ is such that D_n^3 is uniformly integrable. Prove that $\mathbb{E}_n[(d_{o_n}^{(n)})^3] \xrightarrow{\mathbb{P}} \mathbb{E}_\mu[d_o^3]$. Conclude that (2.4.32) and (2.4.33) hold. [Hint: You need to be very careful, as $\mathbb{E}_\mu[d_o^3]$ may be a random variable when μ is a random measure.]*

Exercise 2.18 (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 2.19 (Continuity of neighborhood functions) *Fix $m \geq 1$ and ℓ_1, \dots, ℓ_m . Show that*

$$h(G, o) = \mathbb{1}_{\{|\partial B_r(o)| = \ell_k \forall k \leq m\}} \quad (2.7.4)$$

is a bounded continuous function.

Exercise 2.20 (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 2.14.*

Exercise 2.21 ($|\mathcal{C}_{\max}|$ is not a continuous functional) *Show that $|\mathcal{C}_{\max}|/n$ is not a continuous functional in \mathcal{G}_* .*

Exercise 2.22 (Upper bound on $|\mathcal{C}_{\max}|$ using LWC) *Let $G_n = ([n], E(G_n))$ denote a finite (possibly disconnected) random graph. Let (G, o) be a random variable on \mathcal{G}_* having law μ . Assume that G_n converges in probability in the local weak sense to (G, o) , and assume that the survival probability of the limiting graph (G, o) satisfies $\zeta = \mathbb{P}(|\mathcal{C}(o)| = \infty) = 0$. Show that $|\mathcal{C}_{\max}|/n \xrightarrow{\mathbb{P}} 0$.*

Exercise 2.23 (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 that $Z_{\geq k}/n \xrightarrow{\mathbb{P}} \zeta_{\geq k} = \mathbb{P}(|\mathcal{C}(o)| \geq k)$ for every $k \geq 1$.*

Exercise 2.24 (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 $|\mathcal{C}_{\max}|/n \xrightarrow{\mathbb{P}} \eta < \zeta = \mathbb{P}(|\mathcal{C}(o)| = \infty)$.*

Exercise 2.25 (Local weak convergence with a subcritical limit) *Let G_n converge in probability in the local weak sense to (G, o) . Assume that $\zeta = \mathbb{P}(|\mathcal{C}(o)| = \infty) = 0$. Use (2.5.6) to prove that $|\mathcal{C}_{\max}|/n \xrightarrow{\mathbb{P}} 0$.*

Exercise 2.26 (Sufficiency of (2.5.7) for almost locality giant) *Let $G_n = ([n], E(G_n))$ denote a finite (possibly disconnected) random graph. Let (G, o) be a random variable on \mathcal{G}_* having law μ . Assume that G_n converges in probability in the local weak sense to (G, o) and write $\zeta = \mathbb{P}(|\mathcal{C}(o)| = \infty)$ for the survival probability of the limiting graph (G, o) . Assume that*

$$\limsup_{k \rightarrow \infty} \frac{1}{n^2} \limsup_{n \rightarrow \infty} \mathbb{E} \left[\#\{x, y \in [n] : |\mathcal{C}(x)|, |\mathcal{C}(y)| \geq k, x \not\leftrightarrow y\} \right] > 0. \quad (2.7.5)$$

Then, prove that for some $\varepsilon > 0$,

$$\limsup_{n \rightarrow \infty} \mathbb{P}(|\mathcal{C}_{\max}| \leq n(\zeta - \varepsilon)) > 0. \quad (2.7.6)$$

Exercise 2.27 (Sufficiency of (2.5.7) for uniqueness giant) *Under the conditions in Exercise 2.26, show that*

$$\limsup_{n \rightarrow \infty} \mathbb{P}(|\mathcal{C}_{(2)}| > \varepsilon n) > 0. \quad (2.7.7)$$

Exercise 2.28 (Lower bound on graph distances in Erdős-Rényi random graphs) *Use Exercise 2.16 to show that, for every $\varepsilon > 0$,*

$$\mathbb{P} \left(\frac{\text{dist}_{\text{ER}_n(\lambda/n)}(o_1, o_2)}{\log n} \leq \frac{1 - \varepsilon}{\log \lambda} \right) = 0. \quad (2.7.8)$$

Exercise 2.29 (Lower bound on graph distances in Erdős-Rényi random graphs) *Use Exercise 2.16 to show that*

$$\lim_{K \rightarrow \infty} \limsup_{n \rightarrow \infty} \mathbb{P} \left(\text{dist}_{\text{ER}_n(\lambda/n)}(o_1, o_2) \leq \frac{\log n}{\log \lambda} - K \right) = 0, \quad (2.7.9)$$

which is a significant extension of Exercise 2.28.

Exercise 2.30 (Convergence of the clustering spectrum) *Prove that, under the conditions of Theorem 2.15, (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 $ER_n(p)$ as well as the inhomogeneous random graphs such as $GRG_n(\mathbf{w})$ 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 $GRG_n(\mathbf{w})$, 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, multitype 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

Subject	% in Giant	Total	Original Source	Data
AD-blood	0.8542	96	Goñi et al. (2008)	Goñi et al. (2008)
MS-blood	0.8780	205	Goñi et al. (2008)	Goñi et al. (2008)
actors	0.8791	2180759	Boldi and Vigna (2004)	Boldi et al. (2011)
DBLP	0.8162	986324	Boldi and Vigna (2004)	Boldi et al. (2011)
Zebras	0.8214	28	Sundaresan et al. (2007)	Sundaresan et al. (2007)

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.*

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

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 multitype branching processes, the natural generalization of branching processes for inhomogeneous random graphs. In Section 3.5, we use multitype 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 \mathcal{S} . When there are individuals of just 2 types, as in Example 3.1, then it suffices to take $\mathcal{S} = \{1, 2\}$. 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 w_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* μ_n , where, for $A \subseteq \mathcal{S}$, $\mu_n(A)$ 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* $\kappa: \mathcal{S}^2 \rightarrow [0, \infty)$. 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 κ , 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 (\mathcal{S}, μ) , where \mathcal{S} is a separable metric space and μ is a Borel probability measure on \mathcal{S} .*

(ii) *A vertex space \mathcal{V} is a triple $(\mathcal{S}, \mu, (\mathbf{x}_n)_{n \geq 1})$, where (\mathcal{S}, μ) is a ground space and, for each $n \geq 1$, \mathbf{x}_n is a random sequence (x_1, x_2, \dots, x_n) of n points of \mathcal{S} , such*

that

$$\mu_n(A) = \#\{i: x_i \in A\}/n \rightarrow \mu(A) \quad (3.2.1)$$

for every μ -continuity set $A \subseteq \mathcal{S}$. The convergence in (3.2.1) is denoted by $\mu_n \xrightarrow{p} \mu$.

(iii) A kernel κ is a symmetric non-negative (Borel) measurable function on \mathcal{S}^2 . By a kernel on a vertex space $(\mathcal{S}, \mu, (\mathbf{x}_n)_{n \geq 1})$ we mean a kernel on (\mathcal{S}, μ) .

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 $\mathbf{p} = (p_{ij})_{1 \leq i < j \leq n}$,

$$\mathbb{E}[E(\text{IRG}_n(\mathbf{p}))] = \sum_{i < j} p_{ij}, \quad (3.2.2)$$

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 κ in the following definition:

Definition 3.3 (Setting: graphical and irreducible kernels) (i) A kernel κ is graphical if the following conditions hold:

- (a) κ is continuous a.e. on \mathcal{S}^2 ;
- (b)

$$\iint_{\mathcal{S}^2} \kappa(x, y) \mu(dx) \mu(dy) < \infty; \quad (3.2.3)$$

- (c)

$$\frac{1}{n} \mathbb{E}[E(\text{IRG}_n(\kappa))] \rightarrow \frac{1}{2} \iint_{\mathcal{S}^2} \kappa(x, y) \mu(dx) \mu(dy). \quad (3.2.4)$$

Similarly, a sequence $(\kappa_n)_{n \geq 1}$ of kernels is called graphical with limit κ when

$$y_n \rightarrow y \quad \text{and} \quad z_n \rightarrow z \quad \text{imply that} \quad \kappa_n(y_n, z_n) \rightarrow \kappa(y, z), \quad (3.2.5)$$

where κ satisfies conditions (a) and (b) above, and

$$\frac{1}{n} \mathbb{E}[E(\text{IRG}_n(\kappa_n))] \rightarrow \frac{1}{2} \iint_{\mathcal{S}^2} \kappa(x, y) \mu(dx) \mu(dy). \quad (3.2.6)$$

(ii) A kernel κ is called reducible if

$$\exists A \subseteq \mathcal{S} \quad \text{with} \quad 0 < \mu(A) < 1 \quad \text{such that} \quad \kappa = 0 \quad \text{a.e. on} \quad A \times (\mathcal{S} \setminus A);$$

otherwise κ is irreducible.

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 $\iint_{\mathcal{S}^2} \kappa(x, y) \mu(dx) \mu(dy)$. 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 $o_{\mathbb{P}}(n)$, 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 $\mathcal{S} \times \mathcal{S}$ 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, κ 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 κ is reducible, then the random graph splits into two independent random graphs on the two disjoint subsets A and $\mathcal{S} \setminus A$. 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 $\mathcal{S} = [0, 1]$, $x_i = i/n$ and μ the Lebesgue-measure on $[0, 1]$. Then, clearly, (3.2.1) is satisfied. In fact, Janson (2009b) shows that we can always restrict to $\mathcal{S} = [0, 1]$ by suitably adapting the other choices of our model. However, for notational purposes, it is more convenient to work with general \mathcal{S} . For example, where $\mathcal{S} = \{1\}$ is just a single type, the model reduces to the Erdős-Rényi random graph, and in the setting where $\mathcal{S} = [0, 1]$, 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 κ , 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

$$p_{ij}(\kappa) = p_{ij} = \frac{1}{n} [\kappa(x_i, x_j) \wedge n], \quad (3.2.7)$$

and the events that different edges are present are independent. Similarly, $\text{IRG}_n(\kappa_n)$ is defined with κ_n replacing κ in (3.2.7). Exercise 3.2 shows that the lower bound in (3.2.4) always holds for $\text{IRG}_n(\kappa)$ when κ is continuous. Further, Exercise 3.3 shows that (3.2.4) holds for $\text{IRG}_n(\kappa)$ when κ is bounded and continuous.

Bollobás et al. (2007) also allow for the choices

$$p_{ij}^{(\text{NR})}(\kappa_n) = 1 - e^{-\kappa_n(x_i, x_j)/n}, \quad (3.2.8)$$

and

$$p_{ij}^{(\text{GRG})}(\kappa_n) = p_{ij} = \frac{\kappa(x_i, x_j)}{n + \kappa(x_i, x_j)}. \quad (3.2.9)$$

All results in Bollobás et al. (2007) remain valid for the choices in (3.2.8) and (3.2.9). When

$$\sum_{i,j \in [n]} \kappa_n(x_i, x_j)^3 = o(n^{3/2}), \quad (3.2.10)$$

this follows immediately from [Volume 1, Theorem 6.18], see Exercise 3.6.

For $\text{CL}_n(\mathbf{w})$ with $\mathbf{w} = (w_i)_{i \in [n]}$ as in (1.3.15), we take $\mathcal{S} = [0, 1]$, $x_i = i/n$ and, with $\psi(x) = [1 - F]^{-1}(x)$,

$$\kappa_n(x, y) = \psi(x)\psi(y)n/\ell_n. \quad (3.2.11)$$

For $\text{CL}_n(\mathbf{w})$ with $\mathbf{w} = (w_i)_{i \in [n]}$ satisfying Condition 1.1, instead, we take $\mathcal{S} = [0, 1]$, $x_i = i/n$ and

$$\kappa_n(i/n, j/n) = w_i w_j / \mathbb{E}[W_n]. \quad (3.2.12)$$

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 \mathcal{S} is general and $\kappa(x, y) = \lambda$ for every $x, y \in \mathcal{S}$, then the edge probabilities p_{ij} given by (3.2.7) are all equal to λ/n (for $n > \lambda$). Then $\text{IRG}_n(\kappa) = \text{ER}_n(\lambda/n)$. The simplest choice here is to take $\mathcal{S} = \{1\}$.

The homogeneous bipartite random graph

Let n be even, let $\mathcal{S} = \{0, 1\}$, let $x_i = 0$ for $i \in [n/2]$ and $x_i = 1$ for $i \in [n] \setminus [n/2]$. Further, let κ be defined by $\kappa(x, y) = 0$ when $x \neq y$ and $\kappa(x, y) = \lambda$ when $x = y$. Then $\text{IRG}_n(\kappa)$ is the random bipartite graph with $n/2$ vertices in each class, where each possible edge between classes is present with probability λ/n , independently of the other edges. Exercise 3.7 investigates the validity of Definitions 3.2–3.3 for homogeneous bipartite graphs.

The stochastic block model

The stochastic block model generalizes the above setting. Let n be even, let $\mathcal{S} = \{0, 1\}$, let $x_i = 0$ for $i \in [n/2]$ and $x_i = 1$ for $i \in [n] \setminus [n/2]$. Further, let κ be defined by $\kappa(x, y) = b$ when $x \neq y$ and $\kappa(x, y) = a$ 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 b/n . A major research effort has been devoted to studying when it can be statistically detected that $a \neq b$.

Homogeneous random graphs

We call an inhomogeneous random graph *homogeneous* when $\lambda(x) = \int_{\mathcal{S}} \kappa(x, y) \mu(dy) \equiv \lambda$. Thus, despite the inhomogeneity that is present, every vertex in the graph has (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 $\mathcal{S} = \{1, \dots, r\}$. 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 κ is equivalent to an $r \times r$ matrix, and the random graph $\text{IRG}_n(\kappa)$ 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 δ , 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

$$p_{ij} = \frac{c}{\max\{i, j\}}, \quad (3.2.13)$$

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

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 $\text{IRG}_n(\kappa_n)$. As we shall see, the degree of a vertex of a given type x is asymptotically Poisson with a mean

$$\lambda(x) = \int_{\mathcal{S}} \kappa(x, y) \mu(dy) \quad (3.3.1)$$

that depends on x . This leads to a mixed Poisson distribution for the degree D of a (uniformly chosen) random vertex of $\text{IRG}_n(\kappa)$. We recall that N_k denotes the number of vertices of $\text{IRG}_n(\kappa)$ with degree k . Our main result is as follows:

Theorem 3.4 (The degree sequence of $\text{IRG}_n(\kappa)$) *Let (κ_n) be a graphical sequence of kernels with limit κ as described in Definition 3.3(i). For any fixed $k \geq 0$,*

$$N_k/n \xrightarrow{\mathbb{P}} \int_{\mathcal{S}} \frac{\lambda(x)^k}{k!} e^{-\lambda(x)} \mu(dx) \quad (3.3.2)$$

where $x \mapsto \lambda(x)$ is defined by

$$\lambda(x) = \int_{\mathcal{S}} \kappa(x, y) \mu(dy). \quad (3.3.3)$$

Equivalently,

$$N_k/n \xrightarrow{\mathbb{P}} \mathbb{P}(D = k), \quad (3.3.4)$$

where D has the mixed Poisson distribution with distribution W_λ given by

$$\mathbb{P}(W_\lambda \leq x) = \int_0^x \lambda(y) \mu(dy). \quad (3.3.5)$$

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.

3.3.1 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 $\kappa_n(i, j) \rightarrow \kappa(i, j)$ for every $i, j \in [r]$, while (3.2.1) implies that the number n_i of vertices of type i satisfies $n_i/n \rightarrow \mu_i$ 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 are the number of vertices with type j and $\mu_j = \lim_{n \rightarrow \infty} n_j/n$. Hence,

$$D_v \xrightarrow{d} \text{Poi}\left(\sum_j \mu_j \kappa(i, j)\right) = \text{Poi}(\lambda(i)), \quad (3.3.6)$$

where $\lambda(i) = \int \kappa(i, j) d\mu(j) = \sum_j \kappa(i, j) \mu_j$. Consequently,

$$\mathbb{P}(D_v = k) \rightarrow \mathbb{P}(\text{Poi}(\lambda(i)) = k) = \frac{\lambda(i)^k}{k!} e^{-\lambda(i)}. \quad (3.3.7)$$

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, \dots, n_r ,

$$\frac{1}{n} \mathbb{E}[N_{k,i}] = \frac{1}{n} n_i \mathbb{P}(D_v = k) \rightarrow \mu_i \mathbb{P}(\text{Poi}(\lambda(i)) = k). \quad (3.3.8)$$

It is easily checked that $\text{Var}(N_{k,i}) = O(n)$ (see Exercise 3.12). Hence,

$$\frac{1}{n} N_{k,i} \xrightarrow{\mathbb{P}} \mathbb{P}(\text{Poi}(\lambda(i)) = k) \mu_i = \mathbb{P}(\Xi = k), \quad (3.3.9)$$

and thus, summing over i ,

$$\frac{1}{n} N_k = \sum_i \frac{1}{n} N_{k,i} \xrightarrow{\mathbb{P}} \sum_i \mathbb{P}(\text{Poi}(\lambda(i)) = k) \mu_i = \mathbb{P}(D = k). \quad (3.3.10)$$

This proves Theorem 3.4 in the finite-type case. \square

In order to prove Theorem 3.4 in the general case, we shall be approximating a sequence of graphical kernels (κ_n) by appropriate regular finite kernels.

3.3.2 FINITE-TYPE APPROXIMATIONS OF BOUNDED KERNELS

Recall that \mathcal{S} is a separable metric space, and that μ is a Borel measure on \mathcal{S} with $\mu(\mathcal{S}) = 1$. Here the metric and topological structure of \mathcal{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 (κ_n) is a graphical sequence of bounded kernels with limit κ as described in Definition 3.3(i). Thus, we assume that

$$\sup_{n \geq 1} \sup_{x, y \in \mathcal{S}} \kappa_n(x, y) < \infty. \quad (3.3.11)$$

Our aim is to find finite-type approximations of κ_n that bound κ_n from above and below. It is here that the metric structure of \mathcal{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 \mathcal{V} with limit κ , then there exists a sequence $(\kappa_m^-)_{m \geq 1}$ of finite-type kernels on \mathcal{V} with the following properties:*

- (i) *If κ is irreducible, then so is κ_m^- for all large enough m ;*
- (ii) *$\kappa_m^-(x, y) \nearrow \kappa(x, y)$ for every $x, y \in \mathcal{S}$.*

Let us now give some details. We find these finite-type approximations by giving a partition \mathcal{P}_m of \mathcal{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 \mathcal{S} . Given a sequence of finite partitions $\mathcal{P}_m = \{A_{m1}, \dots, A_{mM_m}\}$, $m \geq 1$, of \mathcal{S} and an $x \in \mathcal{S}$, we define the function $x \mapsto i_m(x)$ by requiring that

$$x \in A_{m, i_m(x)}. \quad (3.3.12)$$

Thus, $i_m(x)$ indicates the cell in \mathcal{P}_m that x is in. For $A \subset \mathcal{S}$ we write $\text{diam}(A)$ for $\sup\{|x - y| : x, y \in A\}$. By taking \mathcal{P}_m as the dyadic partition into intervals of length 2^{-m} in \mathcal{S} , we easily see the following:

Lemma 3.6 (Approximating partition) *Fix $m \geq 1$. There exists a sequence of finite partitions $\mathcal{P}_m = \{A_{m1}, \dots, A_{mM_m}\}$ of \mathcal{S} such that*

- (i) *each A_{mi} is measurable and $\mu(\partial A_{mi}) = 0$;*
- (ii) *for each m , \mathcal{P}_{m+1} refines \mathcal{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. $x \in \mathcal{S}$, $\text{diam}(A_{m, i_m(x)}) \rightarrow 0$ as $m \rightarrow \infty$, where $i_m(x)$ is defined by (3.3.12).*

Recall that a kernel κ is a symmetric measurable function on $\mathcal{S} \times \mathcal{S}$. Fixing a sequence of partitions with the properties described in Lemma 3.6, we can define sequences of lower and upper approximations to κ by

$$\kappa_m^-(x, y) = \inf\{\kappa(x', y') : x' \in A_{m, i_m(x)}, y' \in A_{m, i_m(y)}\}, \quad (3.3.13)$$

$$\kappa_m^+(x, y) = \sup\{\kappa(x', y') : x' \in A_{m, i_m(x)}, y' \in A_{m, i_m(y)}\}. \quad (3.3.14)$$

We thus replace κ by its infimum or supremum on each $A_{mi} \times A_{mj}$. As κ_m^+ might be $+\infty$, we only use it for bounded κ_n as in (3.3.11). Obviously, κ_m^- and κ_m^+ are constant on $A_{m, i} \times A_{m, j}$ for every i, j , so that κ_m^- and κ_m^+ correspond to finite-type kernels (see Exercise 3.14).

By Lemma 3.6(ii),

$$\kappa_m^- \leq \kappa_{m+1}^- \quad \text{and} \quad \kappa_m^+ \geq \kappa_{m+1}^+. \quad (3.3.15)$$

Furthermore, if κ is continuous a.e. then, by Lemma 3.6(iii),

$$\kappa_m^-(x, y) \rightarrow \kappa(x, y) \quad \text{and} \quad \kappa_m^+(x, y) \rightarrow \kappa(x, y) \quad \text{for a.e. } (x, y) \in \mathcal{S}^2. \quad (3.3.16)$$

If (κ_n) is a graphical sequence of kernels with limit κ , we define instead

$$\kappa_m^-(x, y) := \inf\{(\kappa \wedge \kappa_n)(x', y') : x' \in A_{m, i_m(x)}, y' \in A_{m, i_m(y)}, n \geq m\}, \quad (3.3.17)$$

$$\kappa_m^+(x, y) := \sup\{(\kappa \vee \kappa_n)(x', y') : x' \in A_{m, i_m(x)}, y' \in A_{m, i_m(y)}, n \geq m\}. \quad (3.3.18)$$

Again, by Lemma 3.6, we have $\kappa_m^- \leq \kappa_{m+1}^-$, and, by Lemma 3.6(iii) and 3.3(ii),

$$\kappa_m^-(x, y) \nearrow \kappa(x, y) \quad \text{as } m \rightarrow \infty, \text{ for a.e. } (x, y) \in \mathcal{S}^2. \quad (3.3.19)$$

Since $\kappa_m^- \leq \kappa$, we can obviously construct our random graph so that $\text{IRG}_n(\kappa_m^-) \subseteq \text{IRG}_n(\kappa_n)$, and in the sequel we assume this. See also Exercise 3.15. Similarly, we shall assume that $\text{IRG}_n(\kappa_m^+) \supseteq \text{IRG}_n(\kappa_n)$ when κ_n is bounded as in (3.3.11). Moreover, when $n \geq m$,

$$\kappa_n \geq \kappa_m^-, \quad (3.3.20)$$

and we may assume that $\text{IRG}_n(\kappa_m^-) \subseteq \text{IRG}_n(\kappa_n)$. 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.

3.3.3 DEGREE SEQUENCES 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 $\varepsilon > 0$ be given. From (3.2.6) and monotone convergence, there is an m such that

$$\iint_{\mathcal{S}^2} \kappa_m^-(x, y) \mu(dx) \mu(dy) > \iint_{\mathcal{S}^2} \kappa(x, y) \mu(dx) \mu(dy) - \varepsilon. \quad (3.3.21)$$

For $n \geq m$, we have $\kappa_m^- \leq \kappa_n$ by (3.3.20), so we may assume that $\text{IRG}_n(\kappa_m^-) \subseteq \text{IRG}_n(\kappa_n)$. Then, by (3.2.6) and (3.3.21),

$$\begin{aligned} & \frac{1}{n} E(\text{IRG}_n(\kappa_n) \setminus \text{IRG}_n(\kappa_m^-)) \\ &= \frac{1}{n} E(G(n, \kappa_n)) - \frac{1}{n} E(\text{IRG}_n(\kappa_m^-)) \\ & \xrightarrow{\mathbb{P}} \frac{1}{2} \iint_{\mathcal{S}^2} \kappa(x, y) \mu(dx) \mu(dy) - \frac{1}{2} \iint_{\mathcal{S}^2} \kappa_m^-(x, y) \mu(dx) \mu(dy) < \frac{\varepsilon}{2}, \end{aligned} \quad (3.3.22)$$

so that, whp $E(\text{IRG}_n(\kappa_n) \setminus \text{IRG}_n(\kappa_m^-)) < \varepsilon n$. Let us write $N_k^{(m)}$ for the number of vertices of degree k in $\text{IRG}_n(\kappa_m^-)$. It follows that, whp,

$$|N_k^{(m)} - N_k| < 2\varepsilon n. \quad (3.3.23)$$

Writing $\Xi^{(m)}$ for the equivalent of Ξ defined using κ_m^- in place of κ , by the proof for the finite-type case, $N_k^{(m)}/n \xrightarrow{\mathbb{P}} \mathbb{P}(D^{(m)} = k)$. Thus, whp,

$$|N_k^{(m)}/n - \mathbb{P}(D^{(m)} = k)| < \varepsilon. \quad (3.3.24)$$

Finally, we have $\mathbb{E}[\Xi] = \int_{\mathcal{S}} \lambda(x) \mu(dx) = \iint_{\mathcal{S}^2} \kappa(x, y) \mu(dx) \mu(dy)$. Since $\lambda^{(m)}(x) \leq \lambda(x)$, we can couple the branching processes such that $\Xi^{(m)} \leq \Xi$, and thus

$$\begin{aligned} \mathbb{P}(D \neq D^{(m)}) &= \mathbb{P}(\Xi - \Xi^{(m)} \geq 1) \leq \mathbb{E}[\Xi - \Xi^{(m)}] \\ &= \iint_{\mathcal{S}^2} \kappa(x, y) \mu(dx) \mu(dy) - \iint_{\mathcal{S}^2} \kappa_m^-(x, y) \mu(dx) \mu(dy) < \varepsilon. \end{aligned} \quad (3.3.25)$$

Combining (3.3.23), (3.3.24) and (3.3.25), we see that $|N_k/n - \mathbb{P}(\Xi = k)| < 4\varepsilon$ whp. \square

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 κ_n as in (3.3.11). As a result, we will need to study the effect of bounding κ_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 \mathcal{S} having distribution μ . 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 (κ_n) be a graphical sequence of kernels with limit κ . Suppose that $\mathbb{P}(W > t) = \mu(x: \lambda(x) > t) = at^{-(\tau-1)}(1 + o(1))$ as $t \rightarrow \infty$, for some $a > 0$ and $\tau > 2$. Then*

$$N_{\geq k}/n \xrightarrow{\mathbb{P}} \mathbb{P}(D \geq k) \sim ak^{-(\tau-1)}, \quad (3.3.26)$$

where the first limit is for k fixed and $n \rightarrow \infty$, and the second for $k \rightarrow \infty$. In particular, $\lim_{n \rightarrow \infty} \mathbb{P}(D \geq k) \sim ak^{-(\tau-1)}$ as $k \rightarrow \infty$.

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 $t \rightarrow \infty$,

$$\begin{aligned} \mathbb{P}(\text{Poi}(W) > t \mid \Lambda > (1 + \varepsilon)t) &\rightarrow 1, \\ \text{and } \mathbb{P}(\text{Poi}(W) > t \mid W < (1 - \varepsilon)t) &= o(t^{-(\tau-1)}). \end{aligned}$$

It follows that $\mathbb{P}(D > t) = \mathbb{P}(\text{Poi}(W) > t) = at^{-(\tau-1)}(1 + o(1))$ as $t \rightarrow \infty$. Exercise 3.16 asks you to fill the details on this argument. \square

This result shows that the general inhomogeneous random graph does include natural cases with power-law degree distributions. Recall that we have already

observed in [Volume 1, Theorem 6.7] that this is the case for the $\text{GRG}_n(\mathbf{w})$ when the weights sequence \mathbf{w} is chosen appropriately.

3.4 MULTITYPE BRANCHING PROCESSES

In order to study further properties of $\text{IRG}_n(\kappa_n)$, 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 $\text{IRG}_n(\kappa_n)$. 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 κ and μ , we can even have a continuum of types. Such branching processes are called *multitype branching processes*. In this section, we discuss some of the basics and we shall quickly go to the special case of multitype branching processes where every offspring has a Poisson distribution.

3.4.1 MULTITYPE BRANCHING PROCESSES WITH FINITELY MANY TYPES

Multitype 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 $\mathbf{j} = (j_1, \dots, j_r) \in \mathbb{N}_0^r$ 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 \mathbf{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 $\mathbf{Z}_n^{(i)} = (Z_{n,1}^{(i)}, \dots, Z_{n,r}^{(i)})$. We are interested in the survival or extinction of multitype branching processes, and in the growth of the generation sizes. In the multitype case, we are naturally lead to a matrix setup. We now discuss the survival versus extinction of multitype branching processes. We denote the survival probability of the multitype branching process when starting from a single individual of type i by

$$\zeta^{(i)} = \mathbb{P}(\mathbf{Z}_n^{(i)} \neq \mathbf{0} \text{ for all } n), \quad (3.4.1)$$

and we let $\boldsymbol{\zeta} = (\zeta^{(1)}, \dots, \zeta^{(r)})$. Our first aim is to investigate when $\boldsymbol{\zeta} = \mathbf{0}$.

Multitype branching processes and generating functions

We write $\mathbf{p}(\mathbf{j}) = (p_j^{(1)}, \dots, p_j^{(r)})$ and we let

$$G^{(i)}(\mathbf{s}) = \sum_{\mathbf{j}} p_j^{(i)} \prod_{a=1}^r s_a^{j_a} \quad (3.4.2)$$

be the joint moment generating function of the offspring of an individual of type i . We write $\mathbf{G}(\mathbf{s}) = (G^{(1)}(\mathbf{s}), \dots, G^{(r)}(\mathbf{s}))$ for the vector of generating functions. We now generalize Theorem 3.1 to the multitype case. Let ζ satisfy $\zeta = \mathbf{1} - \mathbf{G}(\mathbf{1} - \zeta)$. By convexity of $\mathbf{s} \mapsto \mathbf{G}(\mathbf{s})$, there is at most one non-zero solution to the equation $\mathbf{s} = \mathbf{G}(\mathbf{s})$ which is not equal to $\mathbf{0}$. Define

$$G_n^{(i)}(\mathbf{s}) = \mathbb{E}\left[\prod_{a=1}^r s_a^{Z_{n,a}^{(i)}}\right], \quad (3.4.3)$$

and $\mathbf{G}_n(\mathbf{s}) = (G_n^{(1)}(\mathbf{s}), \dots, G_n^{(r)}(\mathbf{s}))$. Then, we have that $\mathbf{G}_{n+1}(\mathbf{s}) = \mathbf{G}_n(\mathbf{G}(\mathbf{s})) = \mathbf{G}(\mathbf{G}_n(\mathbf{s}))$ and $\zeta = \mathbf{1} - \lim_{n \rightarrow \infty} \mathbf{G}_n(\mathbf{0})$. 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 $\mathbf{T}_\kappa = (\kappa_{ij}\mu(j))_{i,j=1}^r$.

We note that when $\mathbf{G}(\mathbf{s}) = \mathbf{M}\mathbf{s}$ for some matrix \mathbf{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 multitype 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 multitype branching process is non-singular.

3.4.2 SURVIVAL VS. EXTINCTION OF MULTITYPE BRANCHING PROCESSES

We continue to describe the survival versus extinction of multitype branching processes in terms of the mean offspring. Let κ_{ij} denote the expected offspring of type j of a single individual of type i , and let $\mathbf{T}_\kappa = (\kappa_{ij}\mu(j))_{i,j=1}^r$ be the matrix of expected offsprings. We shall assume that there exists an l such that the matrix \mathbf{T}_κ^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 \mathbf{T}_κ has a unique largest eigenvalue $\|\mathbf{T}_\kappa\|$ with non-negative left-eigenvector \mathbf{x}_κ , and the eigenvalue $\|\mathbf{T}_\kappa\|$ can be computed as

$$\|\mathbf{T}_\kappa\| = \sup_{\mathbf{x}: \|\mathbf{x}\|_2 \leq 1} \|\mathbf{T}_\kappa \mathbf{x}\|_2, \quad \text{where} \quad \|\mathbf{x}\|_2 = \sqrt{\sum_{i=1}^r x_i^2}. \quad (3.4.4)$$

We note that

$$\mathbb{E}[\mathbf{Z}_{n+1}^{(i)} \mid \mathbf{Z}_n^{(i)} = \mathbf{z}] = \mathbf{T}_\kappa \mathbf{z}, \quad (3.4.5)$$

so that

$$\mathbb{E}[\mathbf{Z}_{n+1}^{(i)}] = \mathbf{T}_\kappa^n \mathbf{e}^{(i)}. \quad (3.4.6)$$

where \mathbf{T}_κ^n denotes the n -fold application of the matrix \mathbf{T}_κ , and $\mathbf{e}^{(i)}$ 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 multitype branching processes, as we shall now discuss in more detail.

First, when $\|\mathbf{T}_\kappa\| < 1$,

$$\mathbb{E}[\mathbf{Z}_{n+1}^{(i)}] \leq \|\mathbf{T}_\kappa\|^n \|\mathbf{e}^{(i)}\|_2, \quad (3.4.7)$$

which converges to 0 exponentially fast. Therefore, by the Markov inequality ([Volume 1, Theorem 2.17]), the multitype branching process dies out a.s. When $\|\mathbf{T}_\kappa\| > 1$, on the other hand, the sequence

$$M_n = \mathbf{x}_\kappa \mathbf{Z}_{n+1}^{(i)} \|\mathbf{T}_\kappa\|^{-n} \quad (3.4.8)$$

is a non-negative martingale, by (3.4.5) and the fact that \mathbf{x}_κ is a left-eigenvector with eigenvalue $\|\mathbf{T}_\kappa\|$, since $\mathbf{x}_\kappa \mathbf{T}_\kappa = \|\mathbf{T}_\kappa\| \mathbf{x}_\kappa$. 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 $M_n \xrightarrow{a.s.} M_\infty$ and $\mathbb{E}[M_n] \rightarrow \mathbb{E}[M_\infty]$. More precisely, there is a multitype analog of the Kesten-Stigum Theorem ([Volume 1, Theorem 3.10]). Since $\mathbb{E}[M_n] = \mathbb{E}[M_0] = \mathbf{x}_\kappa \mathbf{e}^{(i)} > 0$, we thus have that $\mathbf{Z}_{n+1}^{(i)}$ 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 $\mathbf{Z}_{n+1}^{(i)} \xrightarrow{\mathbb{P}} \mathbf{0}$ when $\|\mathbf{T}_\kappa\| = 1$. See e.g. (Harris, 1963, Sections II.6-II.7). We conclude that, for non-singular and irreducible multitype branching processes, we have that $\zeta > 0$ precisely when $\|\mathbf{T}_\kappa\| > 1$. This is the content of the following theorem:

Theorem 3.8 (Survival vs. extinction of finite-type branching processes) *sec-MulBP* Let $(\mathbf{Z}_n^{(i)})_{n \geq 0}$ be a multitype branching process with offspring matrix \mathbf{T}_κ on the type space $[r]$. Assume that there exists an l such that the matrix \mathbf{T}_κ^l 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 ζ is the largest solution to $\zeta = \mathbf{1} - \mathbf{G}(\mathbf{1} - \zeta)$, and $\zeta > 0$ precisely when $\|\mathbf{T}_\kappa\| > 1$.

(b) Assume that $\|\mathbf{T}_\kappa\| > 1$. Let \mathbf{x}_κ be the unique positive left-eigenvector of \mathbf{T}_κ . Then, the martingale $M_n = \mathbf{x}_\kappa \mathbf{Z}_{n+1}^{(i)} \|\mathbf{T}_\kappa\|^{-n}$ converges to a.s. to a non-negative limit on the event of survival precisely when $\mathbb{E}[Z_1^{(i)} \log(Z_1^{(i)})] < \infty$ for all $i \in [r]$, where $Z_1^{(i)} = \|\mathbf{Z}_1^{(i)}\|_1$ is the total offspring of a type i individual.

3.4.3 POISSON MULTITYPE BRANCHING PROCESSES

We now specialize to Poisson multitype branching processes as these turn out to be the most relevant in the random graph setting. We call a multitype branching processes *Poisson* when all the number of children of each type are independent Poisson random variables. Thus, $\mathbf{Z}^{(i)} = (Z_{1,1}^{(i)}, \dots, Z_{1,r}^{(i)})$ is a vector of independent Poisson random variables with means $(\kappa_{1,i}, \dots, \kappa_{r,i})$. As we see later, Poisson multitype branching processes arise naturally when exploring a component of $\text{IRG}_n(\kappa)$ 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 multitype branching processes with finitely many types

For Poisson multitype branching processes with finitely many types, we obtain that

$$G^{(i)}(\mathbf{s}) = \mathbb{E}\left[\prod_{a=1}^r s_a^{Z_{1,a}^{(i)}}\right] = e^{\sum_{a=1}^r \kappa_{a,i} \mu^{(i)}(s_a - 1)} = e^{(\mathbf{T}_\kappa(\mathbf{s}-\mathbf{1}))_i}. \quad (3.4.9)$$

Thus, the vector of survival probabilities ζ satisfies

$$\zeta = \mathbf{1} - e^{-\mathbf{T}_\kappa \zeta}, \quad (3.4.10)$$

where, for a matrix \mathbf{M} , we recall that $e^{\mathbf{M}} = \sum_{n \geq 0} \mathbf{M}^n / n!$ denotes the matrix exponential. This leads us to the investigation of eigenfunctions of non-linear operators of the form $f \mapsto 1 - e^{-\mathbf{T}_\kappa f}$.

There is a beautiful property of Poisson random variables that allows us to construct a Poisson multitype 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 λ . Perform X multinomial trials, where the i^{th} outcome appears with probability p_i for some probabilities $(p_i)_{i=1}^k$. 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 $(\lambda p_i)_{i=1}^k$.*

Proof Let $(x_i)_{i=1}^k$ denote a sequence of non-negative integers, denote $x = \sum_{i=1}^k x_i$ and compute

$$\begin{aligned} \mathbb{P}((X_i)_{i=1}^k = (x_i)_{i=1}^k) &= \mathbb{P}(X = x) \mathbb{P}((X_i)_{i=1}^k = (x_i)_{i=1}^k \mid X = x) \\ &= e^{-\lambda} \frac{\lambda^x}{x!} \binom{x}{x_1, x_2, \dots, x_k} p_1^{x_1} \cdots p_k^{x_k} = \prod_{i=1}^k e^{-\lambda p_i} \frac{\lambda^{x_i}}{(x_i)!}. \end{aligned} \quad (3.4.11)$$

□

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 $\mathbf{Z}^{(i)}$ has the same distribution as $(Z_{ij})_{j \in [r]}$.

We now extend the above setting of finite-type Poisson multitype branching processes to the infinite-type case. Again, we prove results in the infinite-types case by reducing to the finite-type case.

Poisson multitype branching processes with infinitely many types

Let κ be a kernel. We define the Poisson multitype branching processes with kernel κ as follows. Each individual of type $x \in \mathcal{S}$ is replaced in the next generation by a set of individuals distributed as a Poisson process on \mathcal{S} with intensity $\kappa(x, y)\mu(dy)$. Thus, the number of children with types in a subset $A \subseteq \mathcal{S}$ 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 multitype branching process with kernel κ , starting from an ancestor of type $x \in \mathcal{S}$. Set

$$\zeta_\kappa = \int_{\mathcal{S}} \zeta_\kappa(x)\mu(dx). \quad (3.4.12)$$

Again, it can be seen in a similar way as above that $\zeta_\kappa > 0$ if and only if $\|\mathbf{T}_\kappa\| > 1$, where now the linear operator \mathbf{T}_κ is defined, for $f: \mathcal{S} \rightarrow \mathbb{R}$,

$$(\mathbf{T}_\kappa f)(x) = \int_{\mathcal{S}} \kappa(x, y)f(y)\mu(dy), \quad (3.4.13)$$

for any (measurable) function f such that this integral is defined (finite or $+\infty$) for a.e. $x \in \mathcal{S}$.

Note that $\mathbf{T}_\kappa f$ is defined for every $f \geq 0$, with $0 \leq \mathbf{T}_\kappa f \leq \infty$. If $\kappa \in L^1(\mathcal{S} \times \mathcal{S})$, as we shall assume throughout, then $\mathbf{T}_\kappa f$ is also defined for every bounded f . In this case $\mathbf{T}_\kappa f \in L^1(\mathcal{S})$ and thus $\mathbf{T}_\kappa f$ is finite a.e.

As we shall see, the analysis of multitype 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

$$\|\mathbf{T}_\kappa\| = \sup \left\{ \|\mathbf{T}_\kappa f\|_2 : f \geq 0, \|f\|_2 \leq 1 \right\} \leq \infty. \quad (3.4.14)$$

When finite, $\|\mathbf{T}_\kappa\|$ is the norm of \mathbf{T}_κ as an operator on $L^2(\mathcal{S})$; it is infinite if \mathbf{T}_κ does not define a bounded operator on L^2 . The norm $\|\mathbf{T}_\kappa\|$ is at most the Hilbert-Schmidt norm of \mathbf{T}_κ :

$$\|\mathbf{T}_\kappa\| \leq \|\mathbf{T}_\kappa\|_{\text{HS}} = \|\kappa\|_{L^2(\mathcal{S} \times \mathcal{S})} = \left(\iint_{\mathcal{S}^2} \kappa(x, y)^2 \mu(dx)\mu(dy) \right)^{1/2}. \quad (3.4.15)$$

We also define the non-linear operator Φ_κ by

$$(\Phi_\kappa f)(x) = 1 - e^{-(\mathbf{T}_\kappa f)(x)}, \quad x \in \mathcal{S}, \quad (3.4.16)$$

for $f \geq 0$. Note that for such f we have $0 \leq \mathbf{T}_\kappa f \leq \infty$, and thus $0 \leq \Phi_\kappa f \leq 1$. We shall characterize the survival probability $\zeta_\kappa(x)$, and thus ζ_κ , in terms of the non-linear operator Φ_κ , showing essentially that the function $x \mapsto \zeta_\kappa(x)$ is the maximal fixed point of the non-linear operator Φ_κ (recall (3.4.10)). Again, the survival probability satisfies that $\zeta_\kappa > 0$ precisely when $\|\mathbf{T}_\kappa\| > 1$, recall the finite-types case discussed in detail above.

We call a multitype branching process *supercritical* when $\|\mathbf{T}_\kappa\| > 1$, *critical*

when $\|\mathbf{T}_\kappa\| < 1$, and *subcritical* when $\|\mathbf{T}_\kappa\| < 1$. Then, the above discussion can be summarized by saying that a multitype 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 ψ is an eigenvector with eigenvalue

$$\|\mathbf{T}_\kappa\| = \int_{\mathcal{S}} \psi(y)^2 \mu(dy) = \|\psi\|_{L^2(\mu)}^2. \quad (3.4.17)$$

Further, the left eigenvector equals $\mu(x)\psi(x)$. Thus, the rank-1 multitype 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 \mathbf{T}_κ . It also turns out that the rank-1 multitype 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

$$\lambda(x) = \int_{\mathcal{S}} \psi(x)\psi(y)\mu(dy) = \psi(x) \int_{\mathcal{S}} \psi(y)\mu(y), \quad (3.4.18)$$

so that an offspring of an individual of type x receives mark y with probability

$$p(x, y) = \frac{T_\kappa(x, y)\mu(y)}{\lambda(x)} = \frac{\psi(x)\psi(y)\mu(y)}{\psi(x) \int_{\mathcal{S}} \psi(z)\mu(dz)} = \frac{\psi(y)\mu(y)}{\int_{\mathcal{S}} \psi(z)\mu(dz)}. \quad (3.4.19)$$

We conclude that every individual chooses its type *independently* of the type of its parent. This means that this multitype branching process reduces to a single type branching process with offspring distribution that is $\text{Poi}(W_\lambda)$, where

$$\mathbb{P}(W_\lambda \in A) = \frac{\int_A \psi(y)\mu(dy)}{\int_{\mathcal{S}} \psi(z)\mu(dz)}. \quad (3.4.20)$$

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

$$\begin{aligned} (T_\kappa\phi)(x) &= \int_{\mathcal{S}} \kappa(x, y)\phi(y)\mu(dy) = \int_{\mathcal{S}} [\alpha(x) + \alpha(y)](a\alpha(y) + b)\mu(dy) \\ &= \alpha(x)(a\|\alpha\|_{L^1(\mu)} + b) + (a\|\alpha\|_{L^2(\mu)} + b\|\alpha\|_{L^1(\mu)}) = \lambda(a\alpha(x) + b). \end{aligned} \quad (3.4.21)$$

Solving for a, b, λ leads to $a\|\alpha\|_{L^1(\mu)} + b = a\lambda$, $a\|\alpha\|_{L^2(\mu)}^2 + b\|\alpha\|_{L^1(\mu)} = \lambda b$, so that the vector $(a, b)^T$ is the eigenvector with eigenvalue λ of the matrix

$$\begin{bmatrix} \|\alpha\|_{L^1(\mu)} & 1 \\ \|\alpha\|_{L^2(\mu)}^2 & \|\alpha\|_{L^1(\mu)} \end{bmatrix} \quad (3.4.22)$$

Solving this equation leads to eigenvalues

$$\lambda = \|\alpha\|_{L^1(\mu)} \pm \|\alpha\|_{L^2(\mu)}, \quad (3.4.23)$$

and the corresponding eigenvectors $(1, \pm\|\alpha\|_{L^2(\mu)})^T$. Clearly, the maximal eigenvalue equals $\lambda = \|\alpha\|_{L^2(\mu)} + \|\alpha\|_{L^1(\mu)}$, with corresponding $L^2(\mu)$ -normalized eigenvector

$$\phi(x) = \frac{\alpha(x) + \|\alpha\|_{L^2(\mu)}}{2(\|\alpha\|_{L^2(\mu)}^2 + \|\alpha\|_{L^1(\mu)}\|\alpha\|_{L^2(\mu)})}. \quad (3.4.24)$$

All other eigenvectors can be chosen to be orthogonal to α and 1, so that this corresponds to a rank-2 setting.

Unimodular Poisson branching processes

In our results, we will be interested in multitype Poisson branching processes that start from the type distribution μ . Thus, we fix the root \emptyset in the branching process tree, and give it a random type Q satisfying that $\mathbb{P}(Q \in A) = \mu(A)$, for any measurable $A \subseteq \mathcal{X}$. 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 $\emptyset a_1 \cdots a_k$, where $a_i \in \mathbb{N}$. When applied to 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 multitype Poisson branching processes) *Let (κ_n) be a sequence of kernels such that $\kappa_n(x, y) \nearrow \kappa(x, y)$. Let $\text{BP}_{\leq k}^{(n)}$ denote the first k generations of the Poisson multitype branching process with kernel κ_n and $\text{BP}_{\leq k}$ that of the Poisson multitype branching process with kernel κ . Then $\text{BP}_{\leq k}^{(n)} \xrightarrow{d} \text{BP}_{\leq k}$. Further, let $\zeta_{\geq k}^{(n)}(x)$ denote the probability that an individual of type x has at least k descendants. Then, $\zeta_{\geq k}^{(n)}(x) \nearrow \zeta_{\geq k}(x)$.*

Proof Since $\kappa_n(x, y) \nearrow \kappa(x, y)$, we can write

$$\kappa(x, y) = \sum_{n \geq 1} \Delta \kappa_n(x, y), \quad \text{where} \quad \Delta \kappa_n(x, y) = \kappa_{n-1}(x, y) - \kappa_{n-1}(x, y). \quad (3.4.25)$$

We can represent this by a sum of independent Poisson multitype 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 $\text{BP}_{\leq k}^{(n)}$ is obtained by keeping all vertices with labels at most n , while $\text{BP}_{\leq k}$ is obtained by keeping all vertices. Consequently, $\text{BP}_{\leq k}^{(n)} \xrightarrow{d} \text{BP}_{\leq k}$ follows. Further, $1 - \zeta_{\geq k}^{(n)}(x) = \zeta_{< k}^{(n)}(x)$ is a continuous function of $\text{BP}_{\leq k}^{(n)}$ and thus also converges. \square

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 κ_n is an irreducible graphical kernel converging to some limiting kernel κ . Then $\text{IRG}_n(\kappa_n)$ converges in probability in the local weak sense to the unimodular multitype marked Galton-Watson tree, where*

▷ *the root has offspring distribution $\text{Poi}(W_\lambda)$ 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

$$\mathbb{P}(Q(x) \in A) = \frac{\int_A \kappa(x, y) \mu(dy)}{\int_S \kappa(x, y) \mu(dy)}. \quad (3.5.3)$$

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 (2.3.7) for Theorem 3.11 in the finite-type case, we introduce some notation. Fix a rooted tree (\mathbf{t}, q) of k generations, and where the vertex $v \in V(\mathbf{t})$ has type $q(v)$ for all $v \in V(\mathbf{t})$. It will again be convenient to think of \mathbf{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 $a_i \in \mathbb{N}$.

Let

$$N_n(\mathbf{t}, q) = \sum_{v \in [n]} \mathbb{1}_{\{B_k^{(n;Q)}(v) \simeq (\mathbf{t}, q)\}} \quad (3.5.4)$$

denote the number of vertices whose local neighborhood up to generation t including their types equals (\mathbf{t}, q) . Here, in $B_k^{(n;Q)}(v)$, we record of the types of the vertices in $B_k^{(G_n)}(v)$. Theorem 2.9 implies that in order to prove Theorem 4.1, we need to show that

$$\frac{N_n(\mathbf{t}, q)}{n} \xrightarrow{\mathbb{P}} \mathbb{P}(\mathbf{BP}_{\leq k} \simeq (\mathbf{t}, q)). \quad (3.5.5)$$

Indeed, since there are only a *finite* number of types, (3.5.5) also implies that $N_n(\mathbf{t})/n \xrightarrow{\mathbb{P}} \rho(\mathbf{t})$, where

$$\rho(\mathbf{t}) = \sum_q \mathbb{P}(\mathbf{BP}_{\leq k} \simeq (\mathbf{t}, q)) \quad (3.5.6)$$

is the probability that the branching process produces a certain tree (and ignoring the types). To this, we can then apply Theorem 2.9. 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 $\mathbb{E}[N_n(\mathbf{t}, q)]/n \rightarrow \mathbb{P}(\mathbf{BP}_{\leq k} \simeq (\mathbf{t}, q))$, after which we prove that $\text{Var}(N_n(\mathbf{t}, q)) = o(n^2)$. Then, by the Chebychev inequality ([Volume 1, Theorem 2.18]), (4.1.5) follows.

Local weak convergence: first moment

We start by noting that

$$\frac{1}{n} \mathbb{E}[N_n(\mathbf{t})] = \mathbb{P}(B_k^{(n;Q)}(o) \simeq (\mathbf{t}, q)), \quad (3.5.7)$$

where $o \in [n]$ is a vertex chosen uniformly at random. Our aim is to prove that $\mathbb{P}(B_k^{(n;Q)}(o) \simeq (\mathbf{t}, q)) \rightarrow \mathbb{P}(\mathbf{BP}_{\leq k} \simeq (\mathbf{t}, q))$.

Let us start with the branching process and analyze $\mathbb{P}(\mathbf{BP}_{\leq k} \simeq (\mathbf{t}, q))$. Fix a vertex $v \in \mathbf{t}$ of type $q(v)$. The probability of seeing a sequence of d_v children of (ordered) types $(q(v1), \dots, q(vd))$ equals

$$e^{-\lambda(q(v))} \frac{\lambda(q(v))^{d_v}}{d_v!} \prod_{j=1}^{d_v} \frac{\kappa(q(v), q(vj))}{\lambda(q(v))} = e^{-\lambda(q(v))} \frac{1}{d_v!} \prod_{j=1}^{d_v} \kappa(q(v), q(vj)) \mu(vj), \quad (3.5.8)$$

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 $v \in V(\mathbf{t})$ with $|v| \leq k-1$, so that

$$\mathbb{P}(\mathbf{BP}_{\leq k} \simeq (\mathbf{t}, q)) = \prod_{v \in T: |v| \leq k-1} e^{-\lambda(q(v))} \frac{1}{d_v!} \prod_{j=1}^{d_v} \kappa(q(v), q(vj)) \mu(vj). \quad (3.5.9)$$

For a comparison to the graph exploration, it will turn out to be convenient to rewrite this probability slightly. Let $\mathbf{t}_{\leq k-1} = \{v: |v| \leq k-1\}$ denote the vertices in the first $k-1$ generations of \mathbf{t} and let $|\mathbf{t}_{\leq k-1}|$ denote its size. We can order the elements of $\mathbf{t}_{\leq k-1}$ in their lexicographic ordering as $(v_i)_{i=1}^{|\mathbf{t}_{\leq k-1}|}$. The lexicographic ordering is the ordering in which the vertices are explored in the breadth-first exploration. Then we can rewrite

$$\mathbb{P}(\text{BP}_{\leq k} \simeq (\mathbf{t}, q)) = \prod_{i=1}^{|\mathbf{t}_{\leq k-1}|} e^{-\lambda(q(v_i))} \frac{1}{d_{v_i}!} \prod_{j=1}^{d_{v_i}} \kappa(q(v_i), q(v_i j)) \mu(v_i j). \quad (3.5.10)$$

Let us now turn to $\text{IRG}_n(\kappa_n)$. Fix a vertex $v \in [n]$ of type $q(v)$. The probability of seeing a sequence of d_v neighbors of (ordered) types $(q(v_1), \dots, q(vd))$ equals

$$\frac{1}{d_v!} \prod_{q \in \mathcal{S}} (1 - \kappa_n(q(v), q))^{n_q - m_q} \prod_{j=1}^{d_v} \frac{\kappa_n(q(v), q(v_j))}{n} [n_{qj} - m_{q(v_j)}(j-1)], \quad (3.5.11)$$

where $m_q = \#\{i: q(v_i) = q\}$ is the number of type q vertices in $(q(v_1), \dots, q(vd))$ and $m_{q(v_j)}(j) = \#\{i \leq j: q(v_i) = q\}$ is the number of type q vertices in $(q(v_1), \dots, q(v_j))$. 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 $n \rightarrow \infty$, and since $n_q/n \rightarrow \mu(q)$, $\kappa_n(q(v), q) \rightarrow \kappa(q(v), q)$ for every $q \in \mathcal{S}$,

$$\begin{aligned} & \frac{1}{d_v!} \prod_{q \in \mathcal{S}} \left(1 - \frac{\kappa_n(q(v), q)}{n}\right)^{n_q - m_q} \prod_{j=1}^{d_v} \kappa_n(q(v), q(v_j)) \frac{[n_{qj} - m_{q(v_j)}(j-1)]}{n} \\ & \rightarrow e^{-\lambda(q(v))} \frac{1}{d_v!} \prod_{i=1}^{d_v} \kappa(q(v), q(v_i)) \mu(q(v_i)), \end{aligned} \quad (3.5.12)$$

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 $\mathbf{t}_{\leq k-1}$ as $(v_i)_{i=1}^{|\mathbf{t}_{\leq k-1}|}$, and, for a type q , let $m_q(i) = \#\{j \in [i]: q(v_j) = q\}$ denote the number of type q individuals in (\mathbf{t}, q) encountered up to and including the i th exploration. Then,

$$\begin{aligned} \mathbb{P}(B_k^{(n; Q)}(o) \simeq (\mathbf{t}, q)) &= \prod_{i=1}^{|\mathbf{t}_{\leq k-1}|} \frac{1}{d_{v_i}!} \prod_{q \in \mathcal{S}} \left(1 - \frac{\kappa_n(q(v_i), q)}{n}\right)^{n_q - m_q(i-1)} \\ &\quad \times \prod_{j=1}^{d_{v_i}} \frac{\kappa_n(q(v_i), q(v_i j))}{n} [n_{qj} - m_{q(v_j)}(i+j-1)]. \end{aligned} \quad (3.5.13)$$

As $n \rightarrow \infty$, 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

Local weak convergence: second moment

Here, we study the second moment of $N_n(\mathbf{t}, q)$, and show that it is close to the first moment squared:

Lemma 3.12 (Concentration of the number of trees) *As $n \rightarrow \infty$,*

$$\frac{\text{Var}(N_n(\mathbf{t}, q)^2)}{n^2} \rightarrow 0. \quad (3.5.14)$$

Consequently, $N_n(\mathbf{t}, q)/n \xrightarrow{\mathbb{P}} \mathbb{P}(\text{BP}_{\leq k} \simeq (\mathbf{t}, q))$.

Proof We start by computing

$$\frac{\mathbb{E}[N_n(\mathbf{t}, q)^2]}{n^2} = \mathbb{P}(B_{o_1}^{(n, Q)}(k) = B_{o_2}^{(n, Q)}(k) = (\mathbf{t}, q)), \quad (3.5.15)$$

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, $\text{dist}_{\text{IRG}_n(\kappa_n)}(o_1, o_2) > 2k$. For this, we take the expectation with respect to o_2 to obtain that

$$\mathbb{P}(\text{dist}_{\text{IRG}_n(\kappa_n)}(o_1, o_2) \leq 2k) = \frac{1}{n} \mathbb{E}[|B_{2k}^{(n, Q)}(o_1)|]. \quad (3.5.16)$$

By (3.5.13),

$$\mathbb{P}(B_{2k}^{(n, Q)}(o_1) \simeq (\mathbf{t}, q)) \leq \prod_{i=1}^{|\mathbf{t}_{\leq k-1}|} \frac{1}{d_{v_i}!} \prod_{j=1}^{d_{v_i}} \kappa_n(q(v_i), q(v_i j)) \frac{n_{qj}}{n}. \quad (3.5.17)$$

Denote the mean-offspring matrix $\mathbf{M}^{(n)}$ by $\mathbf{M}_{ij}^{(n)} = \kappa_n(q(v_i), q(v_i j)) n_{qj}/n$, and write $\mu_i^{(n)} = n_i/n$. Then,

$$\mathbb{P}(\text{dist}_{\text{IRG}_n(\kappa_n)}(o_1, o_2) \leq 2k) \leq \frac{1}{n} \sum_{\ell \leq 2k} (\mu^{(n)})^T (\mathbf{M}^{(n)})^\ell \mathbf{1}. \quad (3.5.18)$$

Let λ_n denote the largest eigenvalue of $\mathbf{M}^{(n)}$, so that

$$\begin{aligned} \mathbb{P}(\text{dist}_{\text{IRG}_n(\kappa_n)}(o_1, o_2) \leq 2k) &\leq \frac{1}{n} \sum_{\ell \leq 2k} \|\mu^{(n)}\|_2 \lambda_n^\ell \|\mathbf{1}\|_2 \\ &= \frac{\|\mu^{(n)}\|_2 \|\mathbf{1}\|_2}{n} \sum_{\ell \leq 2k} \lambda_n^\ell \\ &= \frac{\|\mu^{(n)}\|_2 \|\mathbf{1}\|_2}{n} \frac{\lambda_n^{2k+1} - 1}{\lambda_n - 1}. \end{aligned} \quad (3.5.19)$$

As $n \rightarrow \infty$, $\mathbf{M}_{ij}^{(n)} \rightarrow (T_\kappa)_{ij}$ for every $i, j \in [m]$, so that also $\lambda_n \rightarrow \lambda = \|\mathbf{T}_\kappa\|$. We conclude that, for every $k \in \mathbb{N}$ fixed, $\mathbb{P}(\text{dist}_{\text{IRG}_n(\kappa_n)}(o_1, o_2) \leq 2k) = o(1)$. See Exercises 3.23–3.24 for extensions of this result for distances in inhomogeneous random graphs with finitely many types.

We conclude that

$$\frac{\mathbb{E}[N_n(\mathbf{t}, q)^2]}{n^2} = \mathbb{P}(B_k^{(n;Q)}(o_1), B_k^{(n;Q)}(o_2) \simeq (\mathbf{t}, q), o_2 \notin B_{2k}^{(n)}(o_1)) + o(1). \quad (3.5.20)$$

We now condition on $B_{o_1}^{(n;Q)}(k) \simeq (\mathbf{t}, q)$, and write

$$\begin{aligned} & \mathbb{P}(B_k^{(n;Q)}(o_1), B_k^{(n;Q)}(o_2) \simeq (\mathbf{t}, q), o_2 \notin B_{2k}^{(n)}(o_1)) \\ &= \mathbb{P}(B_k^{(n;Q)}(o_2) \simeq (\mathbf{t}, q) \mid B_{o_1}^{(n;Q)}(k) \simeq (\mathbf{t}, q), o_2 \notin B_{2k}^{(n)}(o_1)) \\ & \quad \times \mathbb{P}(B_k^{(n;Q)}(o_1) \simeq (\mathbf{t}, q), o_2 \notin B_{2k}^{(n)}(o_1)). \end{aligned} \quad (3.5.21)$$

We already know that $\mathbb{P}(B_k^{(n;Q)}(o_1) \simeq (\mathbf{t}, q)) \rightarrow \mathbb{P}(\mathbf{BP}_{\leq k} \simeq (\mathbf{t}, q))$, so that also

$$\mathbb{P}(B_{o_1}^{(n;Q)}(k) \simeq (\mathbf{t}, q), o_2 \notin B_{2k}^{(n)}(o_1)) \rightarrow \mathbb{P}(\mathbf{BP}_{\leq k} \simeq (\mathbf{t}, q)). \quad (3.5.22)$$

In Exercise 3.25, you prove that indeed (3.5.22) holds.

We next investigate the conditional probability, by noting that, conditionally on $B_k^{(n;Q)}(o_1) \simeq (\mathbf{t}, q)$ and $o_2 \notin B_{2k}^{(n)}(o_1)$, the probability that $B_k^{(n;Q)}(o_2) = (\mathbf{t}, q)$ is the same as the probability that $B_k^{(n;Q)}(o_2) \simeq (\mathbf{t}, q)$ in $\text{IRG}_{n'}(\kappa_{n'})$ which is obtained by removing the vertices in $B_k^{(n;Q)}(o_1)$ as well as the edges from them. We conclude that the resulting random graph has $n' = n - |\mathbf{t}|$ 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 (\mathbf{t}, q) . Further, $\kappa_{n'}(i, j) = \kappa_n(i, j)n'/n$. The whole point is that $\kappa_{n'}(i, j) \rightarrow \kappa(i, j)$ and $n'_q/n \rightarrow \mu(q)$ still hold. Therefore, also

$$\begin{aligned} & \mathbb{P}(B_k^{(n;Q)}(o_2) \simeq (\mathbf{t}, q) \mid B_k^{(n;Q)}(o_1) = (\mathbf{t}, q), o_2 \notin B_{2k}^{(n)}(o_1)) \\ & \rightarrow \mathbb{P}(\mathbf{BP}_{\leq k} \simeq (\mathbf{t}, q)). \end{aligned} \quad (3.5.23)$$

and we have proved that $\mathbb{E}[N_n(\mathbf{t}, q)^2]/n^2 \rightarrow \mathbb{P}(\mathbf{BP}_{\leq k} \simeq (\mathbf{t}, q))^2$. From this, (3.5.14) follows directly since $\mathbb{E}[N_n(\mathbf{t}, q)]/n \rightarrow \mathbb{P}(\mathbf{BP}_{\leq k} \simeq (\mathbf{t}, q))$. As a result, $\mathbb{E}[N_n(\mathbf{t}, q)]/n$ is concentrated and thus $N_n(\mathbf{t}, q)/n \xrightarrow{\mathbb{P}} \mathbb{P}(\mathbf{BP}_{\leq k} \simeq (\mathbf{t}, q))$, as required. \square

Lemma 3.12 completes the proof of Theorem 3.11 in the finite-types case. \square

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 (κ_n) . Again define κ_m^- by (3.3.17), so that $\kappa_n \geq \kappa_m^-$. Couple $\text{IRG}_n(\kappa_m^-)$ and $\text{IRG}_n(\kappa_n)$ such that $E(\text{IRG}_n(\kappa_m^-)) \subseteq E(\text{IRG}_n(\kappa_n))$. Let $\varepsilon' > 0$ be given. Recall (3.3.22), which shows that, whp, we can take m so large that the bound

$$E(\text{IRG}_n(\kappa_m^-)) \subseteq E(\text{IRG}_n(\kappa_n)) = \sum_{u \in [n]} (D_u - D_u^{(m)}) \leq \varepsilon' n \quad (3.5.24)$$

holds whp. We let K denote the maximal degree in T . Let $N_n^{(m)}(\mathbf{t}, q)$ denote $N_n(\mathbf{t}, q)$ for the kernel κ_m^- (and keep $N_n(\mathbf{t}, q)$ for (3.5.4) for the kernel κ_n). If a vertex v is such that $B_k^{(n;Q)}(v) \simeq (\mathbf{t}, q)$ 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,

$$\begin{aligned} & |N_n^{(m)}(\mathbf{t}, q) - N_n(\mathbf{t}, q)| \\ & \leq \sum_{u,v} \mathbb{1}_{\{u \in B_{k-1}^{(n;Q)}(v), B_k^{(n;Q)}(v) \simeq (\mathbf{t}, q) \text{ in } \text{IRG}_n(\kappa_m^-)\}} \mathbb{1}_{\{D_u \neq D_u^{(m)}\}} \\ & \quad + \sum_{u,v} \mathbb{1}_{\{u \in B_{k-1}^{(n;Q)}(v), B_k^{(n;Q)}(v) \simeq (\mathbf{t}, q) \text{ in } \text{IRG}_n(\kappa_n)\}} \mathbb{1}_{\{D_u^{(m)} \neq D_u\}}. \end{aligned} \quad (3.5.25)$$

Recall that the maximal degree of any vertex in T is K . Further, if $B_k^{(n;Q)}(v) \simeq (\mathbf{t}, q)$ and $u \in B_{k-1}^{(n;Q)}(v)$, then all the vertices on the path between u and v have degree at most K . Therefore,

$$\sum_v \mathbb{1}_{\{u \in B_{k-1}^{(n;Q)}(v), B_k^{(n;Q)}(v) \simeq (\mathbf{t}, q) \text{ in } \text{IRG}_n(\kappa_m^-)\}} \leq \sum_{\ell \leq k-1} K^\ell \leq \frac{K^k - 1}{K - 1}, \quad (3.5.26)$$

and in the same way,

$$\sum_v \mathbb{1}_{\{u \in B_{k-1}^{(n;Q)}(v), B_k^{(n;Q)}(v) \simeq (\mathbf{t}, q) \text{ in } \text{IRG}_n(\kappa_n)\}} \leq \frac{K^k - 1}{K - 1}. \quad (3.5.27)$$

We thus conclude that whp

$$\begin{aligned} |N_n^{(m)}(\mathbf{t}, q) - N_n(\mathbf{t}, q)| & \leq 2 \frac{K^k - 1}{K - 1} \sum_{u \in [n]} \mathbb{1}_{\{D_u^{(m)} \neq D_u\}} \\ & \leq 2 \frac{K^k - 1}{K - 1} \sum_{u \in [n]} (D_u^{(m)} - D_u) \leq 2 \frac{K^k - 1}{K - 1} \varepsilon' n. \end{aligned} \quad (3.5.28)$$

Taking $\varepsilon' = \varepsilon(K - 1)/[2(K^k - 1)]$, we thus obtain that

$$|N_n^{(m)}(\mathbf{t}, q) - N_n(\mathbf{t}, q)| \leq \varepsilon n, \quad (3.5.29)$$

as required.

For $N_n^{(m)}(\mathbf{t}, q)$, we can use Theorem 3.11 in the finite-types case to obtain that

$$\frac{1}{n} N_n^{(m)}(\mathbf{t}, q) \xrightarrow{\mathbb{P}} \mathbb{P}(\text{BP}_{\leq k}^{(m)} \simeq (\mathbf{t}, q)). \quad (3.5.30)$$

The fact that m can be taken so large that $|\mathbb{P}(\text{BP}_{\leq k}^{(m)} \simeq (\mathbf{t}, q)) - \mathbb{P}(\text{BP}_{\leq k} \simeq (\mathbf{t}, q))| \leq \varepsilon$ 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

process with a mixed Poisson offspring distribution.¹ 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

$$\mathbb{P}(M = m) = w_m/\ell_n, \quad m \in [n]. \quad (3.5.31)$$

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 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 $(X_w, M_w)_w$ a *marked mixed-Poisson branching process* (MMPBP).

Clearly, $w_o = w_{M_\emptyset}$ 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 given by

$$\begin{aligned} \mathbb{P}(w_M \leq x) &= \sum_{m=1}^n \mathbb{1}_{\{w_m \leq x\}} \mathbb{P}(M = m) = \frac{1}{\ell_n} \sum_{m=1}^n w_m \mathbb{1}_{\{w_m \leq x\}} \\ &= \mathbb{P}(W_n^* \leq x) = F_n^*(x), \end{aligned} \quad (3.5.32)$$

where W_n^* is the *size-biased distribution* of W_n and F_n^* is given by

$$F_n^*(x) = \frac{1}{\ell_n} \sum_{m=1}^n w_m \mathbb{1}_{\{w_m \leq x\}}. \quad (3.5.33)$$

When we are only interested in *numbers of individuals*, then we obtain a unimodular branching process since the random variables $(X_w)_w$ 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 $(X_w, M_w)_w$ gives us a way to retrieve the cluster of the vertex M_\emptyset , something that would not be possible on the basis of $(X_w)_w$ only.

In order to define the cluster exploration in $\text{NR}_n(\mathbf{w})$, we introduce a *thinning* that guarantees that we only inspect a vertex once. We think of M_w as being the vertex label in $\text{NR}_n(\mathbf{w})$ 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

¹ In van den Esker et al. (2008), the unimodular branching process is called a delayed branching process, and also the term two-stage branchign process has been used.

their marks correspond to vertices in $\text{NR}_n(\mathbf{w})$ that have not yet appeared. The thinning ensures this. To describe the thinning, we set \emptyset unthinned, and, for w with $w \neq \emptyset$, we thin w when either (i) one of the tree vertices on the (unique) path between the root \emptyset 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 $\mathcal{C}(o)$ 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*

$$\left(\{M_w : w \text{ unthinned}, |w| = k \} \right)_{k \geq 0}. \tag{3.5.34}$$

Proof We prove the two statements simultaneously. By construction, the distribution of o is the same as that of M_\emptyset , the mark of the root of the marked mixed Poisson branching process. We continue by proving that the direct neighbors of the root \emptyset agree in both constructions. In $\text{NR}_n(\mathbf{w})$, the direct neighbors are equal to $\{j \in [n] \setminus \{\emptyset\} : I_{lj} = 1\}$, where $(I_{lj})_{j \in [n] \setminus \{\emptyset\}}$ are independent $\text{Be}(p_{lj})$ random variables with $p_{lj} = 1 - e^{-w_l w_j / \ell_n}$.

We now prove that the same is true for the marked mixed Poisson branching process. Conditionally on $M_\emptyset = l$, 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 $(X_{\emptyset,j})_{j \in [n]}$, where $X_{\emptyset,j}$ is the number of offspring of the root that receive mark j , is a vector of independent Poisson random variables with parameters $w_l w_j / \ell_n$. Due to the thinning, a mark occurs precisely when $X_{\emptyset,j} \geq 1$. Therefore, the mark j occurs, independently for all $j \in [n]$, with probability $1 - e^{-w_l w_j / \ell_n} = p_{jk}^{(\text{NR})}$. 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 $\text{NR}_n(\mathbf{w})$.

Next, we look at the number of new elements of $\mathcal{C}(o)$ 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 $w_l w_j / \ell_n$. 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 $1 - e^{-w_l w_j / \ell_n} = p_{jk}^{(\text{NR})}$, as required. \square

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 $I_i = [n_{\leq i}] \setminus [n_{\leq i-1}]$ (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

$$\mathbb{P}(M(j) = \ell) = \frac{1}{n_i}, \quad \ell \in I_j. \quad (3.5.35)$$

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 multitype 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 $\{M_w : w \text{ unthinned}\}$, the marks of unthinned vertices encountered in the marked multitype 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

$$\left(\{M_w : w \text{ unthinned}, |w| = k \} \right)_{k \geq 0}. \quad (3.5.36)$$

You are asked to prove Proposition 3.14 in Exercise 3.28.

3.5.4 LOCAL WEAK CONVERGENCE TO UNIMODULAR TREES FOR $\text{GRG}_n(\mathbf{w})$

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 $\text{GRG}_n(\mathbf{w})$) *Assume that Condition 1.1(a)-(b) holds. Then $\text{GRG}_n(\mathbf{w})$ converges locally-weakly in probability to the*

unimodular Galton-Watson tree with offspring distribution $(p_k)_{k \geq 0}$ given by

$$p_k = \mathbb{P}(D = k) = \mathbb{E} \left[e^{-W} \frac{W^k}{k!} \right]. \quad (3.5.37)$$

This result also applies to $\text{NR}_n(\mathbf{w})$ and $\text{CL}_n(\mathbf{w})$ 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 $\text{CM}_n(\mathbf{d})$ proved in Theorem 4.1, and the relation between $\text{GRG}_n(\mathbf{w})$ and $\text{CM}_n(\mathbf{d})$ discussed in Section 1.3 and using Theorem 1.7. This approach is interesting in itself, since it allows for general proofs for $\text{GRG}_n(\mathbf{w})$ by proving the result first for $\text{CM}_n(\mathbf{d})$, and then merely extending it to $\text{GRG}_n(\mathbf{w})$. 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 $\text{IRG}_n(\kappa)$. The main result shows that there is a giant component when the associated multitype branching process is supercritical, while otherwise there is not:

Theorem 3.16 (Giant component of IRG) *Let (κ_n) be a sequence of irreducible graphical kernels with limit κ , and let \mathcal{C}_{\max} denote the largest connected component of $\text{IRG}_n(\kappa_n)$. Then,*

$$|\mathcal{C}_{\max}|/n \xrightarrow{\mathbb{P}} \zeta_\kappa. \quad (3.6.1)$$

In all cases $\zeta_\kappa < 1$, while $\zeta_\kappa > 0$ precisely when $\|\mathbf{T}_\kappa\| > 1$.

Theorem 3.16 is a generalization of the law of large numbers for the largest connected component in [Volume 1, Theorem 4.8] for $\text{ER}_n(\lambda/n)$ (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 $\mathcal{S} = \{1, 2\}$ and

$$\kappa_n(x, y) = \kappa(x, y) = \lambda \mathbb{1}_{\{x \neq y\}}/2. \quad (3.6.2)$$

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, $\|\mathbf{T}_\kappa\| = \lambda$ with corresponding eigenfunction $f(x) = 1$ for all $x \in \mathcal{S}$. 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 κ_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, $\text{IRG}_n(\kappa)$ 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 [Söderberg \(2002, 2003a,c,b\)](#), 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 κ_n is given by [\(3.2.12\)](#), i.e., $\kappa_n(i/n, j/n) = w_i w_j / \mathbb{E}[W_n]$.

We first assume that [Condition 1.1\(a\)-\(c\)](#) holds, so that in particular $\mathbb{E}[W^2] < \infty$, 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 (κ_n) converges to κ , where the limit κ is given by (recall [\(3.9.2\)](#))

$$\kappa(x, y) = \psi(x)\psi(y)/\mathbb{E}[W], \quad (3.6.3)$$

where $\psi(x) = [1 - F]^{-1}(x)$. Then, we note that for each $f \geq 0$ with $\|f\|_2 = 1$,

$$(\mathbf{T}_\kappa f)(x) = \psi(x) \frac{\int_{\mathcal{S}} \psi(x) f(x) \mu(dx)}{\int_{\mathcal{S}} \psi(x) \mu(dx)}, \quad (3.6.4)$$

so that $\|\mathbf{T}_\kappa f\|_2 = \int_{\mathcal{S}} \psi(x) f(x) \mu(dx) / \int_{\mathcal{S}} \psi(x) \mu(dx)$, which is maximal when $f(x) = \psi(x) / \|\psi\|_2$. We conclude that $\|\mathbf{T}_\kappa\| = \|\psi\|_2^2 / \int_{\mathcal{S}} \psi(x) \mu(dx) = \mathbb{E}[W^2] / \mathbb{E}[W]$. Thus,

$$\|\mathbf{T}_\kappa\| = \mathbb{E}[W^2] / \mathbb{E}[W], \quad (3.6.5)$$

and we recover the results in [Chung and Lu \(2004, 2006b\)](#) in the case where $\mathbb{E}[W^2] < \infty$. In the case where $\mathbb{E}[W^2] = \infty$, on the other hand, we see that $\|\mathbf{T}_\kappa f\|_2 = \infty$ for every f with $\|f\|_2 = 1$ such that $\int_{\mathcal{S}} \psi(x) f(x) \mu(dx) \neq 0$, so that $\|\mathbf{T}_\kappa\| = \infty$, so that $\text{CL}_n(\mathbf{w})$ is always supercritical in this regime.

While [Theorem 3.16](#) identifies the phase transition in $\text{IRG}_n(\kappa_n)$, 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 $\text{GRG}_n(\mathbf{w})$, $\text{CL}_n(\mathbf{w})$ or $\text{NR}_n(\mathbf{w})$, letting $n \rightarrow \infty$. Denote $p_k = \mathbb{P}(\text{Poi}(W) = k)$ as defined below [\(1.3.22\)](#).*

Let \mathcal{C}_{\max} and $\mathcal{C}_{(2)}$ be the largest and second largest components of $\text{GRG}_n(\mathbf{w})$, $\text{CL}_n(\mathbf{w})$ or $\text{NR}_n(\mathbf{w})$.

(a) If $\nu = \mathbb{E}[W^2]/\mathbb{E}[W] > 1$, then there exist $\xi \in (0, 1)$, $\zeta \in (0, 1)$ such that

$$\begin{aligned} |\mathcal{C}_{\max}|/n &\xrightarrow{\mathbb{P}} \zeta, \\ v_k(\mathcal{C}_{\max})/n &\xrightarrow{\mathbb{P}} p_k(1 - \xi^k), \text{ for every } k \geq 0, \\ |E(\mathcal{C}_{\max})|/n &\xrightarrow{\mathbb{P}} \frac{1}{2}\mathbb{E}[W](1 - \xi^2). \end{aligned}$$

while $|\mathcal{C}_{(2)}|/n \xrightarrow{\mathbb{P}} 0$ and $|E(\mathcal{C}_{(2)})|/n \xrightarrow{\mathbb{P}} 0$. Further, $\frac{1}{2}\mathbb{E}[W](1 - \xi^2) < \zeta$, so that the complexity of the giant is linear.

(b) If $\nu = \mathbb{E}[W^2]/\mathbb{E}[W] \leq 1$, then $|\mathcal{C}_{\max}|/n \xrightarrow{\mathbb{P}} 0$ and $|E(\mathcal{C}_{\max})|/n \xrightarrow{\mathbb{P}} 0$.

The above results apply to $\text{GRG}_n(\mathbf{w})$ and $\text{CL}_n(\mathbf{w})$ under the same conditions.

The proof of Theorem 3.17 is deferred to Section 4.2.2 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 ξ 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,

$$\xi = G_{\text{Poi}(W^*)}(\xi) = \mathbb{E}[e^{W^*(\xi-1)}], \quad (3.6.6)$$

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 $v_k(\mathcal{C}_{\max})/n \xrightarrow{\mathbb{P}} p_k(1 - \xi^k)$ and $|\mathcal{C}_{\max}|/n \xrightarrow{\mathbb{P}} \zeta$, it must be that

$$\zeta = \sum_{k \geq 0} p_k(1 - \xi^k) = 1 - G_D(\xi), \quad (3.6.7)$$

where $G_D(s) = \mathbb{E}[s^D]$ is the probability generating function of $D = \text{Poi}(W)$. We also note that $|E(\mathcal{C}_{\max})|/n \xrightarrow{\mathbb{P}} \eta$ with $\eta = \frac{1}{2}\mathbb{E}[W](1 - \xi^2)$, so that

$$\begin{aligned} \eta &= \frac{1}{2} \sum_{k \geq 0} k p_k(1 - \xi^k) = \frac{1}{2} \mathbb{E}[W] \sum_{k \geq 0} \frac{k p_k}{\mathbb{E}[W]} (1 - \xi^k) \\ &= \frac{1}{2} \mathbb{E}[W] (1 - \xi G_{\text{Poi}(W^*)}(\xi)) = \frac{1}{2} \mathbb{E}[W] (1 - \xi^2), \end{aligned} \quad (3.6.8)$$

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,

$$\sum_{k \geq 0} k p_k(1 - \xi^k) > \sum_{k \geq 0} k p_k \sum_{k \geq 0} (1 - \xi^k) p_k = \mathbb{E}[W] \zeta. \quad (3.6.9)$$

As a result,

$$\eta > \frac{1}{2} \mathbb{E}[W] \zeta. \quad (3.6.10)$$

Thus, the average degree η/ζ 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 $k \geq 1$,

$$\sum_{i=0}^{k-1} \xi^i \leq k(1 + \xi^{k-1})/2 \quad (3.6.11)$$

with strict inequality for $k \geq 3$. Multiply by $1 - \xi$ to obtain

$$1 - \xi^k \leq k(1 - \xi)(1 + \xi^{k-1})/2, \quad (3.6.12)$$

again for every $k \geq 1$, again with strict inequality for $k \geq 3$. Now multiply by p_k and sum to get

$$\sum_k p_k(1 - \xi^k) \leq (1 - \xi) \sum_k k p_k(1 + \xi^{k-1})/2. \quad (3.6.13)$$

The lhs of (3.6.13) equals ζ by (3.6.7). We next investigate the rhs of (3.6.13). Recall that

$$\sum_k k p_k = \mathbb{E}[W], \quad (3.6.14)$$

and

$$\sum_k \frac{k p_k}{\mathbb{E}[W]} \xi^{k-1} = \xi. \quad (3.6.15)$$

Hence, the rhs of (3.6.13) is

$$(1 - \xi)(\mathbb{E}[W] + \mathbb{E}[W]\xi)/2 = \mathbb{E}[W](1 - \xi^2)/2, \quad (3.6.16)$$

which is the limit in probability of $|E(\mathcal{C}_{\max})|/n$. \square

We close this section by discussing the consequences of the phase transition for the attack vulnerability of $\text{CL}_n(\mathbf{w})$:

Attack vulnerability of $\text{CL}_n(\mathbf{w})$

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 $\text{CL}_n(\mathbf{w})$, the giant component remains to exist.

We start by addressing the case of random attack for the $\text{CL}_n(\mathbf{w})$ model under Condition 1.1(a)-(c), where $\mathbb{E}[W^2] < \infty$. 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 $(1-p) \times p_{ij}$. For a branching process, this identification is exact, and we have that $\zeta_{\kappa,p} = (1-p)\zeta_{(1-p)\kappa}$, 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 $\text{CL}_n(\mathbf{w})$, this equality is only asymptotic. In the case where $\mathbb{E}[W^2] < \infty$, so that $\nu < \infty$, this means that there exists a critical value $p_c = 1 - 1/\nu$, such that if $p < p_c$, the $\text{CL}_n(\mathbf{w})$ where vertices are removed with probability p , the giant component persists, while if $p > p_c$, then the giant component is destroyed. Thus, when $\mathbb{E}[W^2] < \infty$, the $\text{CL}_n(\mathbf{w})$ is *sensitive to random attacks*. When $\mathbb{E}[W^2] = \infty$, 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 \mathbf{w} is replaced with $\mathbf{w}(p)$, which is equal to $w_i(p) = w_i \mathbb{1}_{\{i > np\}}$, 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 ν is replaced with $\nu(p)$ given by

$$\nu(p) = \mathbb{E}[\psi(U)^2 \mathbb{1}_{\{U > p\}}] / \mathbb{E}[W], \quad (3.6.18)$$

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 , $\mathbb{E}[[1 - F]^{-1}(U)^2 \mathbb{1}_{\{U > p\}}] < \infty$, so that, for p sufficiently close to 1, $\nu(p) < 1$ (see Exercise 3.38). Thus, the $\text{CL}_n(\mathbf{w})$ 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 $\text{ER}_n(\lambda/n)$, it is well known that in the subcritical case for which $\lambda < 1$, the stronger bound $|\mathcal{C}_{\max}| = \Theta(\log n)$ holds (see [Volume 1, Theorems 4.4–4.5]), and that in the supercritical case for which $\lambda > 1$, $|\mathcal{C}_{(2)}| = \Theta(\log n)$. 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 $\text{IRG}_n(\kappa_n)$, where (κ_n) is a graphical sequence of kernels with limit κ . Then,*

- (i) *if κ is subcritical and $\sup_{x,y,n} \kappa_n(x,y) < \infty$, then $|\mathcal{C}_{\max}| = O_{\mathbb{P}}(\log n)$.*
- (ii) *if κ is supercritical, κ is irreducible, and either $\inf_{x,y,n} \kappa_n(x,y) > 0$ or $\sup_{x,y,n} \kappa_n(x,y) < \infty$, then $|\mathcal{C}_{(2)}| = O_{\mathbb{P}}(\log n)$.*

When $\lim_{n \rightarrow \infty} \sup_{x,y} \kappa_n(x,y) = \infty$, 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 $\|\mathbf{T}_{\kappa}\| = \nu = \mathbb{E}[W^2]/\mathbb{E}[W]$, 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 $\mathbb{P}(W \geq w) \sim w^{-(\tau-1)}$, 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 \mathbf{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*

$$[1 - F_n](x) \leq c_2 x^{-(\tau-1)}. \quad (3.7.1)$$

Then, for $\text{NR}_n(\mathbf{w})$ with $\Delta = \max_{j \in [n]} w_j$,

$$|\mathcal{C}_{\max}| = \frac{\Delta}{1 - \nu} + o_{\mathbb{P}}(n^{1/(\tau-1)}). \quad (3.7.2)$$

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 \mathbf{w} as in (1.3.15), let F satisfy

$$[1 - F](x) = cx^{-(\tau-1)}(1 + o(1)). \quad (3.7.3)$$

Then, $\Delta = w_1 = [1 - F]^{-1}(1/n) = (cn)^{1/(\tau-1)}(1 + o(1))$. Therefore,

$$|\mathcal{C}_{\max}| = (cn)^{1/(\tau-1)}/(1 - \nu) + o(n^{1/(\tau-1)}). \quad (3.7.4)$$

Thus, the largest connected component is much larger than for $\text{ER}_n(\lambda/n)$ 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

$1/(1 - \nu)$. 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 $1/(1 - \nu)$ different vertices, leading to a cluster size of roughly $w_i/(1 - \nu)$. 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 $|\mathcal{C}_{\max}|$ to be of order $\log n$ are. Intuitively, when $\Delta \gg \log n$, then $|\mathcal{C}_{\max}| = \Delta/(1 - \nu)(1 + o_{\mathbb{P}}(1))$, whereas if $\Delta = \Theta(\log n)$, then $|\mathcal{C}_{\max}| = \Theta_{\mathbb{P}}(\log n)$ as well. In Turova (2011), it was proved that $|\mathcal{C}_{\max}|/\log n$ converges in probability to a finite constant when $\nu < 1$ and the weights are i.i.d. with distribution function F with $\mathbb{E}[e^{\alpha W}] < \infty$ for some $\alpha > 0$, 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 $\mathbb{E}[W^3] < \infty$ or not.

Theorem 3.20 (The critical behavior with finite third moments) *Fix the Norros-Reittu random graph with weights $\mathbf{w}(t) = \mathbf{w}(1 + tn^{(\tau-3)(\tau-1)})$. Assume that $\nu = 1$, that the weight sequence \mathbf{w} satisfies Condition 1.1(a)-(c), and further assume that*

$$\mathbb{E}[W_n] = \mathbb{E}[W] + o(n^{-1/3}), \quad \mathbb{E}[W_n^2] = \mathbb{E}[W^2] + o(n^{-1/3}), \quad \mathbb{E}[W_n^3] = \mathbb{E}[W^3] + o(1) \tag{3.7.5}$$

Let $(|\mathcal{C}_{(i)}(t)|)_{i \geq 1}$ denote the clusters of $\text{NR}_n(\mathbf{w}(t))$ with $\mathbf{w}(t) = (1 + tn^{-1/3})\mathbf{w}$, ordered in size. Then, as $n \rightarrow \infty$, for all $t \in \mathbb{R}$,

$$(n^{-2/3}|\mathcal{C}_{(i)}(t)|)_{i \geq 1} \xrightarrow{d} (\gamma_i^*(t))_{i \geq 1}, \tag{3.7.6}$$

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 $\text{ER}_n(\lambda/n)$ in the scaling window (see Theorem 5.7).

When $\mathbb{E}[W^{3-\varepsilon}] = \infty$ for some $\varepsilon > 0$, 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 $\tau \in (3, 4)$) *Fix the Norros-Reittu random graph with weights $\mathbf{w}(t) = \mathbf{w}(1 + tn^{(\tau-3)(\tau-1)})$ defined in (1.3.15). Assume that $\nu = 1$ and that there exists a $\tau \in (3, 4)$ and $0 < c_F < \infty$ such that*

$$\lim_{x \rightarrow \infty} x^{\tau-1}[1 - F(x)] = c_F. \tag{3.7.7}$$

Let $(|\mathcal{C}_{(i)}(t)|)_{i \geq 1}$ denote the clusters of $\text{NR}_n(\mathbf{w}(t))$, ordered in size. Then, as $n \rightarrow \infty$, for all $t \in \mathbb{R}$,

$$(n^{-(\tau-2)/(\tau-1)}|\mathcal{C}_{(i)}(t)|)_{i \geq 1} \xrightarrow{d} (\gamma_i(t))_{i \geq 1}, \tag{3.7.8}$$

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 $\mathbb{E}[W^2]$. Now we see that the critical behavior is rather different when $\mathbb{E}[W^3] < \infty$ or $\mathbb{E}[W^3] = \infty$. 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 $\mathbb{E}[W^3] < \infty$. The proof of Theorems 3.20 and 3.21 also reveals that the structure of large critical clusters is quite different. When $\mathbb{E}[W^3] < \infty$, 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 \mathbf{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 $\text{ER}_n(p)$.

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 multitype 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 $\text{NR}_n(\mathbf{w})$ and Poisson branching processes were first exploited to prove versions of Theorem 6.3.

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 εn vertices if we remove any collection of δn 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 $\mathcal{S} = [0, 1]$ and $p_{ij} = \kappa(i/n, j/n)/n$ with $\kappa: [0, 1]^2 \rightarrow [0, \infty)$ being continuous, then the model is the Erdős-Rényi random graph with edge probability λ/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 $\kappa: \mathcal{S} \times \mathcal{S}$ is continuous, then*

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \mathbb{E}[E(\text{IRG}_n(\kappa))] \geq \frac{1}{2} \iint_{\mathcal{S}^2} \kappa(x, y) \mu(dx) \mu(dy), \quad (3.9.1)$$

so that the lower bound in (3.2.4) generally holds.

Exercise 3.3 (Expected number of edges) *Show that when $\kappa: \mathcal{S} \times \mathcal{S}$ is bounded and continuous, then (3.2.4) holds.*

Exercise 3.4 (The Chung-Lu model) *Prove that when κ is given by*

$$\kappa(x, y) = [1 - F]^{-1}(x)[1 - F]^{-1}(y)/\mathbb{E}[W], \quad (3.9.2)$$

then κ is graphical precisely when $\mathbb{E}[W] < \infty$, where W has distribution function F . Further, κ is always irreducible.

Exercise 3.5 (The Chung-Lu model repeated) *Let $\tilde{w}_i = [1-F]^{-1}(i/n)\sqrt{n\mathbb{E}[W]}/\ell_n$ and $w_i = [1-F]^{-1}(i/n)$ as in [Volume 1, (6.2.14)]. Then $\text{CL}_n(\tilde{\mathbf{w}})$ and $\text{CL}_n(\mathbf{w})$ are asymptotically equivalent whenever $(\frac{\mathbb{E}[W_n]}{\ell_n} - 1)^2 = o(n)$.*

Exercise 3.6 (Asymptotic equivalence for general IRGs) *Prove that the random graphs $\text{IRG}_n(\mathbf{p})$ with p_{ij} as in (3.2.7) is asymptotically equivalent to $\text{IRG}_n(\mathbf{p})$ with $p_{ij} = p_{ij}^{(\text{NR})}(\kappa_n)$ and to $\text{IRG}_n(\mathbf{p})$ with $p_{ij} = p_{ij}^{(\text{GRG})}(\kappa_n)$ 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 block model 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*

$$\lim_{n \rightarrow \infty} n_i/n = \mu_i. \quad (3.9.3)$$

Exercise 3.12 (Variance of number of vertices of type i and degree k) *Let $\text{IRG}_n(\kappa_n)$ be a finite-type inhomogeneous random graph with graphical sequence of kernels κ_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 $\text{IRG}_n(\kappa_n)$ be an inhomogeneous random graph with graphical sequence of kernels κ_n that converge to κ . Show that the proportion of isolated vertices converges to*

$$\frac{1}{n} N_0(n) \xrightarrow{\mathbb{P}} p_0 = \int e^{-\lambda(x)} \mu(dx). \quad (3.9.4)$$

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 κ_m^- and κ_m^+ in (3.3.13) and (3.3.14) are of finite type.*

Exercise 3.15 (Inclusion of graphs for larger κ) *Let $\kappa' \leq \kappa$ hold a.e. Show that we can couple $\text{IRG}_n(\kappa')$ and $\text{IRG}_n(\kappa)$ such that $\text{IRG}_n(\kappa') \subseteq \text{IRG}_n(\kappa)$.*

Exercise 3.16 (Tails of Poisson variables) *Use stochastic domination of Poisson random variables with different variables, as well as concentration properties of Poisson variables, to complete the proof of (??).*

Exercise 3.17 (Power-laws for sum kernels) *Let $\kappa(x, y) = \alpha(x) + \alpha(y)$ for a continuous function $\alpha: [0, 1] \mapsto [0, \infty)$. 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 α ?*

Exercise 3.18 (Irreducibility of multitype 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 multitype 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 $T_\kappa^l(i, j) > 0$, where $T_\kappa(i, j) = \kappa_{ij}\mu_j$ is the mean offspring matrix.*

Exercise 3.20 (Singularity of multitype branching process) *Prove that $\mathbf{G}(\mathbf{s}) = \mathbf{M}\mathbf{s}$ for some matrix \mathbf{M} precisely when each individual in the Markov chain has exactly one offspring.*

Exercise 3.21 (Erdős-Rényi random graph) *Prove that $\text{NR}_n(\mathbf{w}) = \text{ER}_n(\lambda/n)$ when \mathbf{w} is constant with $w_i = -n \log(1 - \lambda/n)$.*

Exercise 3.22 (Homogeneous Poisson multitype branching processes) *Consider a homogeneous Poisson multitype branching process with parameter λ . Show that the function $\phi(x) = 1$ is an eigenvector of T_κ with eigenvalue λ . 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 $\|\mathbf{T}_\kappa\| > 1$) *Fix $\varepsilon > 0$. Use (3.5.19) to show that $\text{dist}_{\text{IRG}_n(\kappa_n)}(o_1, o_2) \geq (1 - \varepsilon) \log_{\|\mathbf{T}_\kappa\|}(n)$ whp when $\|\mathbf{T}_\kappa\| > 1$.*

Exercise 3.24 (No large component when $\|\mathbf{T}_\kappa\| < 1$) *Use (3.5.19) to show that $\text{dist}_{\text{IRG}_n(\kappa_n)}(o_1, o_2) = \infty$ whp when $\|\mathbf{T}_\kappa\| < 1$.*

Exercise 3.25 (Proof of no-overlap property in (3.5.22)) *Prove that $\mathbb{P}(B_k^{(n; Q)}(o_1) = (\mathbf{t}, q), o_2 \in B_{2k}^{(n)}(o_1)) \rightarrow 0$, 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 $|\mathcal{C}(o)| \preceq T^*$, 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 $\text{ER}_n(\lambda/n)$) *Use Theorem 3.15 to show that also $\text{ER}_n(\lambda/n)$ converges locally weakly in probability to the Galton-Watson tree with Poisson offspring distribution with parameter λ .*

Exercise 3.28 (Coupling to a multitype Poisson branching process) *Prove Proposition 3.14 by adapting the proof of Proposition 3.13.*

Exercise 3.29 (Phase transition for $r = 2$) *Let ζ_κ denote the survival probability of a two-types multitype branching process. Compute ζ_κ 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 (κ_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 \mathcal{C}_{\max} for $\text{ER}_n(\lambda/n)$) *Prove that, for the Erdős-Rényi random graph, Theorem 3.16 implies that $|\mathcal{C}_{\max}|/n \xrightarrow{\mathbb{P}} \zeta_\lambda$, where ζ_λ is the survival probability of a Poisson branching process with mean λ 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 ζ^2 .*

Exercise 3.33 (The size of small components for $\text{CL}_n(\mathbf{w})$) *Use Theorem 3.18 to prove that, for $\text{CL}_n(\mathbf{w})$ with weights given by (1.3.15) and with $1 < \nu < \infty$, the second largest cluster has size $|\mathcal{C}_{(2)}| = O_{\mathbb{P}}(\log n)$ when W has bounded support or is a.s. bounded below by $\varepsilon > 0$, while if $\nu < 1$, $|\mathcal{C}_{\max}| = O(\log n)$ 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 $\text{GRG}_n(\mathbf{w})$ as in Theorem 3.17. Show that the proportion of vertices of the giant component \mathcal{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 $\text{GRG}_n(\mathbf{w})$ as in Theorem 3.17. Show that when $\xi < 1$, the proportion of vertices outside the giant component \mathcal{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.

Subject	% in Giant	Size	Source	Data
California roads	0.9958	1,965,206	Leskovec et al. (2009)	Leskovec and Krevl (2014)
Facebook	0.9991	721m	Ugander et al. (Preprint (2011))	Ugander et al. (Preprint (2011))
Hyves	0.996	8,047,530	Corten (2012)	Corten (2012)
arXiv astro-ph	0.9538	18,771	Leskovec et al. (2007)	Kunegis (2017)
US power grid	1	4,941	Watts and Strogatz (1998)	Kunegis (2017)
Jazz-musicians	1	198	Gleiser and Danon (2003)	Kunegis (2017)

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 $\text{CM}_n(\mathbf{d})$

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 $\text{CM}_n(\mathbf{d})$ 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

$$p_k^* = \frac{(k+1)p_{k+1}}{\mathbb{E}[D]}. \quad (4.1.1)$$

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

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 $kp_k^{(n)}/\mathbb{E}[D_n]$, where again $p_k^{(n)} = n_k/n$ denotes the proportion of vertices with degree k , and

$$\mathbb{E}[D_n] = \frac{1}{n} \sum_{i \in [n]} d_i = \frac{1}{n} \sum_{i \in [n]} \sum_{k=0}^{\infty} k \mathbb{1}_{\{d_i=k\}} = \sum_{k=0}^{\infty} kp_k^{(n)} \quad (4.1.2)$$

is the average degree in $\text{CM}_n(\mathbf{d})$. Thus, $(kp_k^{(n)}/\mathbb{E}[D_n])_{k \geq 0}$ 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

$$p_k^{*(n)} = \frac{(k+1)p_{k+1}^{(n)}}{\mathbb{E}[D_n]}. \quad (4.1.3)$$

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 $p_k^{*(n)} \rightarrow p_k^*$, 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 $\text{CM}_n(\mathbf{d})$ 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 $\ell_n/n = \mathbb{E}[D_n]/n \rightarrow \mathbb{E}[D] > 0$. 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 (2.3.7) for Theorem 4.1, we introduce some notation. Fix a rooted tree \mathbf{t} with k generations, and let

$$N_n(\mathbf{t}) = \sum_{v \in [n]} \mathbb{1}_{\{B_k^{(n)}(v) \simeq \mathbf{t}\}} \quad (4.1.4)$$

denote the number of vertices whose local neighborhood up to generation t equals \mathbf{t} . By Theorem 2.9, in order to prove Theorem 4.1, we need to show that

$$\frac{N_n(\mathbf{t})}{n} \xrightarrow{\mathbb{P}} \mathbb{P}(\text{BP}_{\leq k} \simeq \mathbf{t}). \quad (4.1.5)$$

Here, we also rely on Theorem 2.10 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 $\mathbb{E}[N_n(\mathbf{t})]/n \rightarrow \mathbb{P}(\mathbf{t}(p, k) \simeq \mathbf{t})$, after which we prove that $\text{Var}(N_n(\mathbf{t})) = o(n^2)$. 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

$$\mathbb{P}(D_n^* = k) = \frac{k}{\mathbb{E}[D_n]} \mathbb{P}(D_n = k), \quad k \in \mathbb{N}, \quad (4.1.6)$$

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), $D_n \xrightarrow{d} D$ and $D_n^* \xrightarrow{d} D^*$, which implies that $\text{BP}_n(t) \xrightarrow{d} \text{BP}(t)$ for every t finite, where $\text{BP}(t)$ is the restriction of the unimodular branching process BP with offspring distribution $(p_k)_{k \geq 1}$ for which $p_k = \mathbb{P}(D = k)$ to its first t individuals (see Exercise 4.1). For future reference, denote the offspring distribution of the above branching process by

$$p_k^* = \mathbb{P}(D^* - 1 = k) \quad (4.1.7)$$

Note that, when \mathbf{t} is a fixed rooted tree of k generations, then $\text{BP}_{\leq k} \simeq \mathbf{t}$ precisely when $\text{BP}(t_k) \simeq \mathbf{t}$, where t_k denotes the number of vertices in \mathbf{t} .

We let $(\mathbf{G}_n(t))_{t \in \mathbb{N}_0}$ denote the graph exploration process from a uniformly chosen vertex $o \in [n]$. Here $\mathbf{G}_n(t)$ is the exploration up to t vertices, in the breadth-first manner. In particular, from $(\mathbf{G}_n(t))_{t \in \mathbb{N}_0}$ we can retrieve $(B_o(t))_{t \in \mathbb{N}_0}$ 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 $(\mathbf{G}_n(t))_{0 \leq t \leq m_n}$ is equal to $(\text{BP}_n(t))_{0 \leq t \leq m_n}$ whenever $m_n \rightarrow \infty$ arbitrarily slowly. In the statement, we write $(\widehat{\mathbf{G}}_n(t), \widehat{\text{BP}}_n(t))_{t \in \mathbb{N}_0}$ for the coupling of $(\mathbf{G}_n(t))_{0 \leq t \leq m_n}$ 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 $(\widehat{\mathbf{G}}_n(t), \widehat{\text{BP}}_n(t))_{t \in \mathbb{N}_0}$ of $(\mathbf{G}_n(t))_{0 \leq t \leq m_n}$ and $(\text{BP}_n(t))_{0 \leq t \leq m_n}$ such that*

$$\mathbb{P}\left((\widehat{\mathbf{G}}_n(t))_{0 \leq t \leq m_n} \neq (\widehat{\text{BP}}_n(t))_{0 \leq t \leq m_n}\right) = o(1), \quad (4.1.8)$$

whenever $m_n \rightarrow \infty$ arbitrarily slowly. Consequently, $\mathbb{E}[N_n(\mathbf{t})]/n \rightarrow \mathbb{P}(\text{BP}_{\leq k} \simeq \mathbf{t})$.

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.

Proof We let the offspring of the root of the branching process \widehat{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, $\widehat{D}_n = d_o$, so that also $\widehat{G}_n(1) = \widehat{\text{BP}}_n(1)$. We next explain how to *jointly* construct $(\widehat{G}_n(t), \widehat{\text{BP}}_n(t))_{0 \leq t \leq m}$ given that we have already constructed $(\widehat{G}_n(t), \widehat{\text{BP}}_n(t))_{0 \leq t \leq m-1}$.

To obtain $(\widehat{G}_n(t))_{0 \leq t \leq m}$, 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 $(\widehat{\text{BP}}_n(t))_{0 \leq t \leq m-1}$ 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 $(d_{U_t} - 1)_{t \geq 0}$ is i.i.d. When y_m is still free, i.e., has not yet been paired in $(\widehat{G}_n(t))_{0 \leq t \leq m-1}$, then we also let x_m be paired to y_m , and we have constructed $(\widehat{G}_n(t))_{0 \leq t \leq m}$. However, a problem arises when y_m has already been paired in $(\widehat{G}_n(t))_{0 \leq t \leq m-1}$, 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 $(\widehat{G}_n(t))_{t \leq m}$ and $(\widehat{\text{BP}}_n(t))_{0 \leq t \leq m}$. We now provide bounds on the probability that an error occurs before time m_n .

There are two sources of differences between $(\widehat{G}_n(t))_{t \geq 0}$ and $(\widehat{\text{BP}}_n(t))_{t \geq 0}$:

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 $(\widehat{G}_n(t))_{0 \leq t \leq m-1}$, but it is incident to a half-edge that has already been paired in $(\widehat{G}_n(t))_{0 \leq t \leq m-1}$. In particular, the vertex to which it is incident has already appeared in $(\widehat{G}_n(t))_{0 \leq t \leq m-1}$ and it is being re-used in the branching process. In this case, a *copy* of the vertex appears in $(\widehat{\text{BP}}_n(t))_{0 \leq t \leq m}$, while a *cycle* appears in $(\widehat{G}_n(t))_{0 \leq t \leq m}$.

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 $(\widehat{G}_n(t))_{t \leq m}$. The probability that the half-edge q_m equals one of these two half-edges is at most

$$\frac{2m-1}{\ell_n}. \tag{4.1.9}$$

Hence the probability that a half-edge is being re-used before time m_n is at most

$$\sum_{m=1}^{m_n} \frac{2m-1}{\ell_n} = \frac{m_n^2}{\ell_n} = o(1), \tag{4.1.10}$$

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/ℓ_n . The probability that vertex i is drawn twice before time m_n is at most

$$\frac{m_n(m_n - 1)}{2} \frac{d_i^2}{\ell_n^2}. \quad (4.1.11)$$

By the union bound, the probability that there exists a vertex that is chosen twice up to time m_n is at most

$$\frac{m_n(m_n - 1)}{2\ell_n} \sum_{i \in [n]} \frac{d_i^2}{\ell_n} \leq m_n^2 \frac{d_{\max}}{\ell_n} = o(1), \quad (4.1.12)$$

by Condition 7.8 when $m_n = o(\sqrt{n/d_{\max}})$.

Completion of the proof

In order to show that $\mathbb{E}[N_n(\mathbf{t})]/n \rightarrow \mathbb{P}(\mathbf{BP}_{\leq k} \simeq \mathbf{t})$, we let t_k denote the number of individuals in the first $k-1$ generations in \mathbf{t} , and let $(\mathbf{t}(t))_{t \in [t_k]}$ be its breadth-first exploration. Then,

$$\mathbb{E}[N_n(\mathbf{t})]/n = \mathbb{P}((\mathbf{G}_n(t))_{t \in [t_k]} = (\mathbf{t}(t))_{t \in [t_k]}), \quad (4.1.13)$$

so that

$$\begin{aligned} \mathbb{P}((\mathbf{G}_n(t))_{t \in [t_k]} = (\mathbf{t}(t))_{t \in [t_k]}) &= \mathbb{P}((\mathbf{BP}_n(t))_{t \in [t_k]} = (\mathbf{t}(t))_{t \in [t_k]}) + o(1) \\ &= \mathbb{P}((\mathbf{BP}(t))_{t \in [t_k]} = (\mathbf{t}(t))_{t \in [t_k]}) + o(1) \\ &= \mathbb{P}(\mathbf{BP}_{\leq k} \simeq \mathbf{t}) + o(1), \end{aligned} \quad (4.1.14)$$

where the first equality is (4.1.8), while the second is the statement that $\mathbf{BP}_n(t) \xrightarrow{d} \mathbf{BP}(t)$ for every t finite from Exercise 4.1. This proves the claim. \square

Local weak convergence for the configuration model: second moment

Here, we study the second moment of $N_n(\mathbf{t})$, 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,*

$$\frac{\mathbb{E}[N_n(\mathbf{t})^2]}{n^2} \rightarrow \mathbb{P}(\mathbf{BP}_{\leq k} \simeq \mathbf{t})^2. \quad (4.1.15)$$

Consequently, $N_n(\mathbf{t})/n \xrightarrow{\mathbb{P}} \mathbb{P}(\mathbf{BP}_{\leq k} \simeq \mathbf{t})$.

Proof We start by computing

$$\frac{\mathbb{E}[N_n(\mathbf{t})^2]}{n^2} = \mathbb{P}(B_k^{(n)}(o_1), B_k^{(n)}(o_2) \simeq \mathbf{t}), \quad (4.1.16)$$

where $o_1, o_2 \in [n]$ are two vertices chosen uniformly at random from $[n]$, independently. Since $|B_k^{(n)}(o_1)| \xrightarrow{d} |\mathbf{BP}_{\leq k}|$, which is a tight random variable, $o_2 \notin B_{2k}^{(n)}(o_1)$ with high probability, so that also

$$\frac{\mathbb{E}[N_n(\mathbf{t})^2]}{n^2} = \mathbb{P}(B_k^{(n)}(o_1), B_k^{(n)}(o_2) \simeq \mathbf{t}, o_2 \notin B_{2k}^{(n)}(o_1)) + o(1). \quad (4.1.17)$$

We now condition on $B_k^{(n)}(o_1) = \mathbf{t}$, and write

$$\begin{aligned} & \mathbb{P}(B_k^{(n)}(o_1), B_k^{(n)}(o_2) \simeq \mathbf{t}, o_2 \notin B_{2k}^{(n)}(o_1)) \\ &= \mathbb{P}(B_k^{(n)}(o_2) \simeq \mathbf{t} \mid B_k^{(n)}(o_1) \simeq \mathbf{t}, o_2 \notin B_{2k}^{(n)}(o_1)) \\ & \quad \times \mathbb{P}(B_k^{(n)}(o_1) \simeq \mathbf{t}, o_2 \notin B_{2k}^{(n)}(o_1)). \end{aligned} \quad (4.1.18)$$

We already know that $\mathbb{P}(B_k^{(n)}(o_1) \simeq \mathbf{t}) \rightarrow \mathbb{P}(\mathbf{BP}_{\leq k} \simeq \mathbf{t})$, so that also

$$\mathbb{P}(B_k^{(n)}(o_1) \simeq \mathbf{t}, o_2 \notin B_{2k}^{(n)}(o_1)) \rightarrow \mathbb{P}(\mathbf{BP}_{\leq k} \simeq \mathbf{t}). \quad (4.1.19)$$

In Exercise 4.2, you prove that indeed (4.1.19) holds.

We note that, conditionally on $B_k^{(n)}(o_1) \simeq \mathbf{t}, o_2 \notin B_k^{(n)}(o_1)$, the probability that $B_k^{(n)}(o_2) \simeq \mathbf{t}$ is the *same* as the probability that $B_k^{(n)}(o_2) \simeq \mathbf{t}$ in $\text{CM}_{n'}(\mathbf{d}')$ which is obtained by removing all vertices in $B_k^{(n)}(o_1)$. Thus, since $B_k^{(n)}(o_1) \simeq \mathbf{t}$, we have that $n' = n - |V(T)|$ and \mathbf{d}' is the corresponding degree sequence. The whole point is that the degree distribution \mathbf{d}' still satisfies Conditions 7.8(a)-(b). Therefore, also

$$\begin{aligned} & \mathbb{P}(B_k^{(n)}(o_2) \simeq \mathbf{t} \mid B_k^{(n)}(o_1) \simeq \mathbf{t}, o_2 \notin B_k^{(n)}(o_1)) \\ & \rightarrow \mathbb{P}(\mathbf{BP}_{\leq k} \simeq \mathbf{t}), \end{aligned} \quad (4.1.20)$$

and we have proved (4.1.15). Since $\mathbb{E}[N_n(\mathbf{t})]/n \rightarrow \mathbb{P}(\mathbf{BP}_{\leq k} \simeq \mathbf{t})$ and $\mathbb{E}[N_n(\mathbf{t})^2]/n^2 \rightarrow \mathbb{P}(\mathbf{BP}_{\leq k} \simeq \mathbf{t})^2$, it follows that $\text{Var}(N_n(\mathbf{t})/n) \rightarrow 0$, so that $N_n(\mathbf{t})/n$ is concentrated. Since also $\mathbb{E}[N_n(\mathbf{t})]/n \rightarrow \mathbb{P}(\mathbf{BP}_{\leq k} \simeq \mathbf{t})$, we obtain that $N_n(\mathbf{t})/n \xrightarrow{\mathbb{P}} \mathbb{P}(\mathbf{BP}_{\leq k} \simeq \mathbf{t})$, as required. \square

Lemma 4.3 completes the proof of Theorem 4.1. \square

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 $\text{CM}_n(\mathbf{d})$, where in most cases, the degrees $\mathbf{d} = (d_i)_{i \in [n]}$ 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,

$$\mathbb{P}(D_n = k) = n_k/n, \quad (4.2.1)$$

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 $\text{CM}_n(\mathbf{d})$ is the following:

Theorem 4.4 (Phase transition in $\text{CM}_n(\mathbf{d})$) *Suppose that Condition 1.5(a)-(b) hold and consider the random graph $\text{CM}_n(\mathbf{d})$, letting $n \rightarrow \infty$. Assume that $p_2 = \mathbb{P}(D = 2) < 1$. Let \mathcal{C}_{\max} and $\mathcal{C}_{(2)}$ be the largest and second largest components of $\text{CM}_n(\mathbf{d})$ (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*

$$\begin{aligned} |\mathcal{C}_{\max}|/n &\xrightarrow{\mathbb{P}} \zeta, \\ v_k(\mathcal{C}_{\max})/n &\xrightarrow{\mathbb{P}} p_k(1 - \xi^k) \quad \text{for every } k \geq 0, \\ |E(\mathcal{C}_{\max})|/n &\xrightarrow{\mathbb{P}} \frac{1}{2}\mathbb{E}[D](1 - \xi^2). \end{aligned}$$

while $|\mathcal{C}_{(2)}|/n \xrightarrow{\mathbb{P}} 0$ and $|E(\mathcal{C}_{(2)})|/n \xrightarrow{\mathbb{P}} 0$.

(b) *If $\nu = \mathbb{E}[D(D-1)]/\mathbb{E}[D] \leq 1$, then $|\mathcal{C}_{\max}|/n \xrightarrow{\mathbb{P}} 0$ and $|E(\mathcal{C}_{\max})|/n \xrightarrow{\mathbb{P}} 0$.*

Consequently, the same result holds for the uniform random graph with degree sequence \mathbf{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 ξ as the extinction probability of a branching process, and ζ as the survival probability of the unimodular Galton-Watson tree that appears in Theorem 4.1. Thus, ζ satisfies

$$\zeta = \sum_{k \geq 1} p_k(1 - \xi^k), \quad (4.2.2)$$

where ξ is the extinction probability of the branching process with offspring distribution $(p_k^*)_{k \geq 0}$, which satisfies

$$\xi = \sum_{k \geq 0} p_k^* \xi^k. \quad (4.2.3)$$

Clearly, $\xi = 1$ precisely when

$$\nu = \sum_{k \geq 0} k p_k^* \leq 1. \quad (4.2.4)$$

By (4.1.1), we can rewrite

$$\nu = \frac{1}{\mathbb{E}[D]} \sum_{k \geq 0} k(k+1)p_{k+1} = \mathbb{E}[D(D-1)]/\mathbb{E}[D], \quad (4.2.5)$$

which explains the condition on ν in Theorem 4.4(a). Further, to understand the asymptotics of $v_k(\mathcal{C}_{\max})$, 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 $v_k(\mathcal{C}_{\max})/n \xrightarrow{\mathbb{P}} p_k(1 - \xi^k)$. 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 $|E(\mathcal{C}_{\max})|/n \xrightarrow{\mathbb{P}} \frac{1}{2}\mathbb{E}[D](1 - \xi^2)$. Therefore, all results in Theorem 4.4 have a simple explanation in terms of the branching process approximation of the connected component for $\text{CM}_n(\mathbf{d})$ 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

$$G_D(x) = \sum_{k=0}^{\infty} p_k x^k = \mathbb{E}[x^D] \quad (4.2.6)$$

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

$$G_D^*(x) = \mathbb{E}[x^{D^*}] = \sum_{k=1}^{\infty} p_k^* x^k = G_D'(x)/G_D'(1), \quad (4.2.7)$$

$$H(x) = \mathbb{E}[D]x(x - G_D^*(x)). \quad (4.2.8)$$

Note that $G_D^*(1) = 1$, and thus $H(0) = H(1) = 0$. Note also that

$$H'(1) = 2\mathbb{E}[D] - \sum_k k^2 p_k = \mathbb{E}[2D - D^2] = -\mathbb{E}[D(D-2)] \quad (4.2.9)$$

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 ξ 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 4.2.2 below. We now remark upon the result and on the conditions arising in it.

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 $\text{CM}_n(\mathbf{d})$ is given by the Ewen's sampling formula $\text{ESF}(1/2)$, see e.g. Arratia et al. (2003). This implies that $|\mathcal{C}_{\max}|/n$ 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 $|\mathcal{C}_{(2)}|/n$ (and for $|\mathcal{C}_{(3)}|/n, \dots$), 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 $n_1 \rightarrow \infty$ be such that $n_1/n \rightarrow 0$, 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 $n_1 \rightarrow \infty$, 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 $|\mathcal{C}_{\max}| = o_{\mathbb{P}}(n)$ (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., $n_4 \rightarrow \infty$ such that $n_4/n \rightarrow 0$, 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 $\text{CM}_n(\mathbf{d})$ 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, $|\mathcal{C}_{\max}| = n - o_{\mathbb{P}}(n)$, 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 $p_2 = 1$.

Reduction to the case where $\mathbb{P}(D = 1) = p_1 > 0$

In our proof, it is convenient to assume that $p_1 = \mathbb{P}(D = 1) > 0$. 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 $d_{\min} = \min\{k: p_k > 0\}$ be the minimum of the support of the asymptotic degree distribution D . Fix $\varepsilon > 0$, and assume that $\varepsilon < p_{d_{\min}}$. Consider the configuration model with $\tilde{n} = n + 2d_{\min}\varepsilon n$, and degree sequence $\tilde{\mathbf{d}} = (\tilde{d}_i)_{i \in [n]}$ with $\tilde{n}_k = n_k$ for all $k > d_{\min}$, $\tilde{n}_{d_{\min}} = n_{d_{\min}} - \varepsilon n$, $\tilde{n}_1 = 2d_{\min}\varepsilon n$. This configuration model can be obtained from $\text{CM}_n(\mathbf{d})$ by replacing εn vertices of degree d_{\min} by d_{\min} vertices having degree 1, as if we have ‘forgotten’ that these vertices are actually equal.

Clearly, $\text{CM}_n(\mathbf{d})$ can be retrieved by identifying εn collections of d_{\min} vertices of degree 1 to a single vertex of degree d_{\min} . When \mathbf{d} satisfies Condition 1.5(a)-(b), then so does $\tilde{\mathbf{d}}$ with limiting degree distribution $\tilde{p}_1 = 2d_{\min}\varepsilon/$, $\tilde{p}_{d_{\min}} = (p_{d_{\min}} - \varepsilon)/(1 + 2d_{\min}\varepsilon)$, $\tilde{p}_k = p_k/(1 + 2d_{\min}\varepsilon)$ for all $k > d_{\min}$. The above procedure clearly makes $|\mathcal{C}_{\max}|$ smaller. Further, with ζ_ε denoting the limit of $|\mathcal{C}_{\max}|/\tilde{n}$ for $\tilde{\mathbf{d}}$, we have that $\zeta_\varepsilon \rightarrow 1$ 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 in terms of a simple continuous-time algorithm in Section 4.2.1 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 $n \rightarrow \infty$, these quantities all converge in probability to *deterministic functions* described in terms of the functions $x \mapsto H(x)$ and $x \mapsto G_D^*(x)$ 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 4.2.2, 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 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-edge. 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 $\text{CM}_n(\mathbf{d})$ 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 τ_x to the half-edge x , where τ_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 $\text{CM}_n(\mathbf{d})$ 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.
- 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.

4.2.2 ANALYSIS OF THE ALGORITHM FOR $CM_n(\mathbf{d})$

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

$$L(t) = S(t) + A(t) \quad (4.2.10)$$

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 $n \rightarrow \infty$:

Proposition 4.5 (Number of living half-edges) *As $n \rightarrow \infty$, for any $t_0 \geq 0$ fixed,*

$$\sup_{0 \leq t \leq t_0} |n^{-1}L(t) - \mathbb{E}[D_n]e^{-2t}| \xrightarrow{\mathbb{P}} 0. \quad (4.2.11)$$

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 $(N^{(x)}(t))_{t \geq 0}$ be a Markov process such that $N^{(x)}(t) = x$ a.s., and the dynamics of $t \mapsto (N^{(x)}(t))_{t \geq 0}$ is such that when it is in position y , then it jumps down by d at rate γy . In other words, the waiting time until the next event is $\text{Exp}(1/\gamma y)$ and each jump is of size d downwards. Then, for every $t_0 \geq 0$,*

$$\mathbb{E} \left[\sup_{t \leq t_0} |N^{(x)}(t) - e^{-\gamma dt} x|^2 \right] \leq 8d(e^{\gamma dt_0} - 1)x + 8d^2. \quad (4.2.12)$$

Proof 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 $x, x-1, x-2, \dots, 0$, 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 $(M_n)_{n \geq 0}$,

$$\mathbb{E} \left[\sup_{m \leq n} |M_m - \mathbb{E}[M_m]|^2 \right] \leq \text{Var}(M_n). \quad (4.2.13)$$

Application of Doob's martingale inequality yields

$$\begin{aligned} \mathbb{E} \left[\sup_{t \leq t_0} |N^{(x)}(t) - e^{-\gamma t} x|^2 \right] &\leq \mathbb{E} \left[\sup_{t \leq t_0} |e^{\gamma t} N^{(x)}(t) - x|^2 \right] \leq 4\mathbb{E} \left[(e^{\gamma t_0} N^{(x)}(t_0) - x)^2 \right] \\ &= 4e^{2\gamma t_0} \text{Var}(N^{(x)}(t_0)) \leq 4(e^{\gamma t_0} - 1)x. \end{aligned} \quad (4.2.14)$$

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 $(N^{(\lfloor x \rfloor)}(t))_{t \geq 0}$ with different initial values such that whenever the smaller one jumps by 1, so does the other. This coupling keeps

$$|N^{(x)}(t) - N^{(\lfloor x \rfloor)}(t)| < 1 \quad (4.2.15)$$

for all $t \geq 0$, and thus,

$$\sup_{t \leq t_0} |N^{(\lfloor x \rfloor)}(t) - e^{-\gamma t} \lfloor x \rfloor| \leq \sup_{t \leq t_0} |N^{(x)}(t) - e^{-\gamma t} x| + 2, \quad (4.2.16)$$

so that by (4.2.14), in turn,

$$\mathbb{E} \left[\sup_{t \leq t_0} |N^{(x)}(t) - e^{-\gamma t} x|^2 \right] \leq 8(e^{\gamma t_0} - 1)x + 8. \quad (4.2.17)$$

Finally, for a general $d > 0$, we observe that $N^{(x)}(t)/d$ is a process of the same type with the parameters (γ, d, x) replaced by $(\gamma d, 1, x/d)$, and the general result follows from (4.2.17) and (4.2.14). \square

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 $\gamma = 1$. \square

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

$$S(t) = \sum_{k=1}^{\infty} kV_k(t). \quad (4.2.18)$$

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 $\text{CM}_n(\mathbf{d})$. 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 $(\tilde{V}_k(t))_{t \geq 0}$, ignoring the effect of Step 1, 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

$$\tilde{S}(t) = \sum_k k \tilde{V}_k(t) \quad (4.2.19)$$

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 $\tilde{S}(t)$ ignores the effect of Step 1 in an identical way as $\tilde{V}_k(t)$.

Recall the functions G_D, G_D^* from (4.2.6)–(4.2.7), and define

$$h(x) = x \mathbb{E}[D] G_D^*(x). \quad (4.2.20)$$

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 $n \rightarrow \infty$, for any $t_0 \geq 0$ fixed*

$$\sup_{t \leq t_0} |n^{-1} \tilde{V}_k(t) - p_k e^{-kt}| \xrightarrow{\mathbb{P}} 0 \quad (4.2.21)$$

for every $k \geq 0$ and

$$\sup_{t \leq t_0} |n^{-1} \sum_{k=0}^{\infty} \tilde{V}_k(t) - G_D(e^{-t})| \xrightarrow{\mathbb{P}} 0, \quad (4.2.22)$$

$$\sup_{t \leq t_0} |n^{-1} \tilde{S}(t) - h(e^{-t})| \xrightarrow{\mathbb{P}} 0. \quad (4.2.23)$$

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 , $\sum_{k > K} kn_k/n = \mathbb{E}[D_n | D_n > k] < \varepsilon$. We may further assume (or deduce from Fatou's inequality) that $\sum_{k > K} kp_k < \varepsilon$, and obtain by (4.2.21) that, whp,

$$\begin{aligned} \sup_{t \leq t_0} |n^{-1} \tilde{S}(t) - h(e^{-t})| &= \sup_{t \leq t_0} \left| \sum_{k=1}^{\infty} k(n^{-1} \tilde{V}_k(t) - p_k e^{-kt}) \right| \\ &\leq \sum_{k=1}^K k \sup_{t \leq t_0} |n^{-1} \tilde{V}_k(t) - p_k e^{-kt}| + \sum_{k > K} k \left(\frac{n_k}{n} + p_k \right) \\ &\leq \varepsilon + \varepsilon + \varepsilon, \end{aligned}$$

proving (4.2.23). An almost identical argument yields (4.2.22). \square

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 $\text{CM}_n(\mathbf{d})$. Then*

$$0 \leq \tilde{S}(t) - S(t) < \sup_{0 \leq s \leq t} (\tilde{S}(s) - L(s)) + d_{\max}. \quad (4.2.24)$$

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 4.8. We then conclude that $\tilde{S}(t) - S(t) = o_{\mathbb{P}}(n)$, and so they have the same limit after rescaling by n . Let us now prove Lemma 4.8:

Proof Clearly, $V_k(t) \leq \tilde{V}_k(t)$, 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

$$\tilde{S}(t) - S(t) = \tilde{S}(t) - L(t) + A(t) < \tilde{S}(t) - L(t) + d_{\max}. \quad (4.2.25)$$

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 $n \rightarrow \infty$. Recall that $A(t) = L(t) - S(t)$ denotes the number of awakened vertices and let

$$\tilde{A}(t) = L(t) - \tilde{S}(t) = A(t) - (\tilde{S}(t) - S(t)) \quad (4.2.26)$$

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 a similar way as $\tilde{S}(t)$ is used as 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 $t_0 \geq 0$,

$$\sup_{t \leq t_0} |n^{-1} \tilde{A}(t) - H(e^{-t})| \xrightarrow{\mathbb{P}} 0. \quad (4.2.27)$$

Lemma 4.8 can be rewritten as

$$0 \leq \tilde{S}(t) - S(t) < -\inf_{s \leq t} \tilde{A}(s) + d_{\max}. \quad (4.2.28)$$

By (4.2.26) and (4.2.28),

$$\tilde{A}(t) \leq A(t) < \tilde{A}(t) - \inf_{s \leq t} \tilde{A}(s) + d_{\max}, \quad (4.2.29)$$

which, perhaps, illuminates the relation between $A(t)$ and $\tilde{A}(t)$. Recall that connected components are explored between subsequent zeros of the process $t \mapsto A(t)$. The function $t \mapsto H(e^{-t})$, 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 $\inf_{s \leq t} \tilde{A}(s)$ 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 $\tilde{A}(t)$ 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 $\phi(x) = \mathbb{E}[D](x - G_D^*(x))$ 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 ϕ 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 ξ and the result follows. \square

Now we are in the position to complete the proof of Theorem 4.4 in the following section.

4.2.3 PROOF OF THEOREM 4.4

We start with the proof of Theorem 4.4(i). Let ξ 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

$$n^{-1} \inf_{t \leq \theta} \tilde{A}(t) = \inf_{t \leq \theta} n^{-1} \tilde{A}(t) - \inf_{t \leq \theta} H(e^{-t}) \xrightarrow{\mathbb{P}} 0. \quad (4.2.30)$$

Further, by Condition 1.5(b), $d_{\max} = o(n)$, and thus $n^{-1}d_{\max} \rightarrow 0$. Consequently, (4.2.28) and (4.2.30) yield

$$\sup_{t \leq \theta} n^{-1} |A(t) - \tilde{A}(t)| = \sup_{t \leq \theta} n^{-1} |\tilde{S}(t) - S(t)| \xrightarrow{\mathbb{P}} 0. \quad (4.2.31)$$

Thus, by (4.2.27),

$$\sup_{t \leq \theta} |n^{-1} A(t) - H(e^{-t})| \xrightarrow{\mathbb{P}} 0. \quad (4.2.32)$$

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 $n^{-1}\tilde{A}(\theta + \varepsilon) \xrightarrow{\mathbb{P}} H(e^{-(\theta+\varepsilon)})$, while $A(\theta + \varepsilon) \geq 0$. Thus, with $\Delta = |H(e^{-\theta-\varepsilon})|/2 > 0$, whp

$$\tilde{S}(\theta + \varepsilon) - S(\theta + \varepsilon) = A(\theta + \varepsilon) - \tilde{A}(\theta + \varepsilon) \geq -\tilde{A}(\theta + \varepsilon) > n\Delta, \quad (4.2.33)$$

while (4.2.31) yields that $\tilde{S}(\theta) - S(\theta) < n\Delta$ whp. Consequently, whp $\tilde{S}(\theta + \varepsilon) - S(\theta + \varepsilon) > \tilde{S}(\theta) - S(\theta)$, so whp Step 1 is performed between the times θ 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 any $\varepsilon > 0$, and whp $0 \leq T_1 \leq \varepsilon$ and $\theta - \varepsilon \leq T_2 \leq \theta + \varepsilon$. In other words, $T_1 \xrightarrow{\mathbb{P}} 0$ and $T_2 \xrightarrow{\mathbb{P}} \theta$. We conclude that we have found one component that is explored between time $T_1 \xrightarrow{\mathbb{P}} 0$ and time $T_2 \xrightarrow{\mathbb{P}} \theta$. 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 $T_1^* \xrightarrow{\mathbb{P}} t_1$ and $T_2^* \xrightarrow{\mathbb{P}} t_2$ where $0 \leq t_1 \leq t_2 \leq \theta < \infty$. If \mathcal{C}^* is the union of all components explored between T_1^* and T_2^* , then*

$$v_k(\mathcal{C}^*)/n \xrightarrow{\mathbb{P}} p_k(e^{-kt_1} - e^{-kt_2}), \quad k \geq 0, \quad (4.2.34)$$

$$|\mathcal{C}^*|/n \xrightarrow{\mathbb{P}} G_D(e^{-t_1}) - G_D(e^{-t_2}), \quad (4.2.35)$$

$$|E(\mathcal{C}^*)|/n \xrightarrow{\mathbb{P}} \frac{1}{2}h(e^{-t_1}) - \frac{1}{2}h(e^{-t_2}). \quad (4.2.36)$$

In particular, if $t_1 = t_2$, then $|\mathcal{C}^*|/n \xrightarrow{\mathbb{P}} 0$ and $|E(\mathcal{C}^*)| \xrightarrow{\mathbb{P}} 0$.

Below, we apply Proposition 4.10 to $T_1 = o_p(1)$ and T_2 , where $T_2 = \theta + o_p(1)$. We can identify the values of the above constants for $t_1 = 0$ and $t_2 = \theta$ as $e^{-kt_1} = 1$, $e^{-kt_2} = \xi$, $G_D(e^{-t_1}) = 1$, $G_D(e^{-t_2}) = 1 - \zeta$, $h(e^{-t_1}) = 2\mathbb{E}[D]$, $h(e^{-t_2}) = 2\mathbb{E}[D]\xi^2$ (see Exercise 4.6).

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 $V_k(t-) = \lim_{s \nearrow t} V_k(s)$)

$$v_k(\mathcal{C}^*) = V_k(T_1^*-) - V_k(T_2^* -), \quad k \geq 1. \quad (4.2.37)$$

Since $T_2^* \xrightarrow{\mathbb{P}} t_2 \leq \theta$ and H is continuous, we obtain that $\inf_{t \leq T_2^*} H(e^{-t}) \xrightarrow{\mathbb{P}} \inf_{t \leq t_2} \tilde{H}(e^{-t}) = 0$, 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 $n^{-1} \inf_{t \leq T_2^*} \tilde{A}(t) \xrightarrow{\mathbb{P}} 0$ and thus also

$$\sup_{t \leq T_2^*} n^{-1} |\tilde{S}(t) - S(t)| \xrightarrow{\mathbb{P}} 0. \quad (4.2.38)$$

Since $\tilde{V}_j(t) \geq V_j(t)$ for every j and $t \geq 0$,

$$\tilde{V}_k(t) - V_k(t) \leq k^{-1} \sum_{j=1}^{\infty} j(\tilde{V}_j(t) - V_j(t)) = k^{-1}(\tilde{S}(t) - S(t)), \quad k \geq 1. \quad (4.2.39)$$

Hence (4.2.38) implies, for every $k \geq 1$, $\sup_{t \leq T_2^*} |\tilde{V}_k(t) - V_k(t)| = o_{\mathbb{P}}(n)$. Consequently, using Lemma 4.7, for $j = 1, 2$,

$$V_k(T_j^* -) = \tilde{V}_k(T_j^* -) + o_{\mathbb{P}}(n) = np_k e^{-kT_j^*} + o_{\mathbb{P}}(n) = np_k e^{-kt_j} + o_{\mathbb{P}}(n), \quad (4.2.40)$$

and (4.2.34) follows by (4.2.37). Similarly, using $\sum_{k=0}^{\infty} (\tilde{V}_k(t) - V_k(t)) \leq \tilde{S}(t) - S(t)$,

$$\begin{aligned} |\mathcal{C}^*| &= \sum_{k=1}^{\infty} (V_k(T_1^* -) - V_k(T_2^* -)) = \sum_{k=1}^{\infty} (\tilde{V}_k(T_1^* -) - \tilde{V}_k(T_2^* -)) + o_{\mathbb{P}}(n) \\ &= nG_D(e^{-T_1^*}) - nG_D(e^{-T_2^*}) + o_{\mathbb{P}}(n), \end{aligned} \quad (4.2.41)$$

and

$$\begin{aligned} 2|E(\mathcal{C}^*)| &= \sum_{k=1}^{\infty} k(V_k(T_1^* -) - V_k(T_2^* -)) = \sum_{k=1}^{\infty} k(\tilde{V}_k(T_1^* -) - \tilde{V}_k(T_2^* -)) + o_{\mathbb{P}}(n) \\ &= nh(e^{-T_1^*}) - nh(e^{-T_2^*}) + o_{\mathbb{P}}(n), \end{aligned} \quad (4.2.42)$$

and (4.2.35) and (4.2.36) follow from the convergence $T_i^* \xrightarrow{\mathbb{P}} t_i$ and the continuity of $t \mapsto G_D(e^{-t})$ and $t \mapsto h(e^{-t})$. \square

We are now ready to complete the proof of Theorem 4.4:

Proof of Theorem 4.4. Let \mathcal{C}'_{\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, $T_1 \xrightarrow{\mathbb{P}} 0$ and $T_2 \xrightarrow{\mathbb{P}} \theta$. The cluster \mathcal{C}'_{\max} is our candidate for the giant component \mathcal{C}_{\max} , and we next prove that indeed it is, whp, the largest connected component.

By Proposition 4.10, with $t_1 = 0$ and $t_2 = \theta$,

$$|v_k(\mathcal{C}'_{\max})|/n \xrightarrow{\mathbb{P}} p_k(1 - e^{-kt}), \quad (4.2.43)$$

$$|\mathcal{C}'_{\max}|/n \xrightarrow{\mathbb{P}} G_D(1) - G_D(e^{-\theta}) = 1 - G_D(\xi), \quad (4.2.44)$$

$$|E(\mathcal{C}'_{\max})|/n \xrightarrow{\mathbb{P}} \frac{1}{2}(h(1) - h(e^{-\theta})) = \frac{1}{2}(h(1) - h(\xi)) = \frac{\mathbb{E}[D]}{2}(1 - \xi^2), \quad (4.2.45)$$

using Exercise 4.6. We have found one large component \mathcal{C}'_{\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}'_{\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 $T_1 \xrightarrow{\mathbb{P}} 0$, the total number of vertices and edges in *all* components found before \mathcal{C}'_{\max} , i.e., before time T_1 , is $o_{\mathbb{P}}(n)$. Hence, recalling that $\ell_n = \Theta(n)$ by Condition 1.5(b),

$$\mathbb{P}(\text{a component } \mathcal{C} \text{ with } |E(\mathcal{C})| \geq \eta\ell_n \text{ is found before } \mathcal{C}'_{\max}) \rightarrow 0. \quad (4.2.46)$$

We conclude that whp no component containing at least $\eta\ell_n$ half-edges is found before \mathcal{C}'_{\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}'_{\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 $d_{\max} = o(n)$ each time Step 1 is performed, we obtain from (4.2.38) that

$$\sup_{t \leq T_3} (\tilde{S}(t) - S(t)) \leq \sup_{t \leq T_2} (\tilde{S}(t) - S(t)) + d_{\max} = o_{\mathbb{P}}(n). \quad (4.2.47)$$

Comparing this to (4.2.33), for every $\varepsilon > 0$ and whp, we have that $\theta + \varepsilon > T_3$. Since also $T_3 > T_2 \xrightarrow{\mathbb{P}} \theta$, it follows that $T_3 \xrightarrow{\mathbb{P}} \theta$. 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 $|\mathcal{C}'|/n \xrightarrow{\mathbb{P}} 0$ and $|E(\mathcal{C}')| \xrightarrow{\mathbb{P}} 0$.

On the other hand, if there would exist a component $\mathcal{C} \neq \mathcal{C}'_{\max}$ in $\text{CM}_n(\mathbf{d})$ with at least $\eta\ell_n$ edges that has not been found before \mathcal{C}'_{\max} , then with probability at least η , 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,

$$\begin{aligned} \mathbb{P}(\text{a component } \mathcal{C} \text{ with } |E(\mathcal{C})| \geq \eta \ell_n \text{ is found after } \mathcal{C}'_{\max}) & \quad (4.2.48) \\ & \leq \eta^{-1} \mathbb{P}(|E(\mathcal{C}')| \geq \eta \ell_n) \rightarrow 0, \end{aligned}$$

since $|E(\mathcal{C}')| \xrightarrow{\mathbb{P}} 0$.

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 \mathcal{C}'_{\max} that has at least $\eta \ell_n$ edges. As a result, we must have that $\mathcal{C}'_{\max} = \mathcal{C}_{\max}$, where \mathcal{C}_{\max} is the largest component. Further, again whp, $|E(\mathcal{C}_{(2)})| < \eta \ell_n$. Consequently, the results for \mathcal{C}_{\max} follow from (4.2.43)-(4.2.45). We have further shown that $|E(\mathcal{C}_{(2)})|/\ell_n \xrightarrow{\mathbb{P}} 0$, which implies that $|E(\mathcal{C}_{(2)})|/n \xrightarrow{\mathbb{P}} 0$ and $|\mathcal{C}_{(2)}|/n \xrightarrow{\mathbb{P}} 0$ because $\ell_n = \Theta(n)$ and $|\mathcal{C}_{(2)}| \leq |E(\mathcal{C}_{(2)})| + 1$. This completes the proof of Theorem 4.4(i). \square

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,

$$\sup_{t \leq T_2} |A(t) - \tilde{A}(t)| = \sup_{t \leq T_2} |\tilde{S}(t) - S(t)| \leq 2d_{\max} = o(n). \quad (4.2.49)$$

For every $\varepsilon > 0$, $n^{-1} \tilde{A}(\varepsilon) \xrightarrow{\mathbb{P}} H(e^{-\varepsilon}) < 0$ 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, $T_2 \xrightarrow{\mathbb{P}} 0$. We apply Proposition 4.10 (which holds in this case too, with $\theta = 0$) and find that if \mathcal{C} is the first component found, then $|E(\mathcal{C})|/n \xrightarrow{\mathbb{P}} 0$.

Let $\eta > 0$. If $|E(\mathcal{C}_{\max})| \geq \eta \ell_n$, then the probability that the first half-edge chosen by Step 1 belongs to \mathcal{C}_{\max} , and thus $\mathcal{C} = \mathcal{C}_{\max}$, is $2|E(\mathcal{C}_{\max})|/(2\ell_n) \geq \eta$, and hence,

$$\mathbb{P}(|E(\mathcal{C}_{\max})| \geq \eta \ell_n) \leq \eta^{-1} \mathbb{P}(|E(\mathcal{C})| \geq \eta \ell_n) \rightarrow 0. \quad (4.2.50)$$

The results follows since $\ell_n = \Theta(n)$ by Condition 1.5(b) and $|\mathcal{C}_{\max}| \leq |E(\mathcal{C}_{\max})| + 1$. This completes the proof of Theorem 4.4(ii), and thus that of Theorem 4.4. \square

4.2.4 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 $\text{UG}_n(\mathbf{d})$ denotes a uniform simple random graph with degrees \mathbf{d} (see [Volume 1, Section 7.5]). The results in Theorem 4.4 also hold for $\text{UG}_n(\mathbf{d})$ when we assume that Conditions 1.5(a)-(c) hold:

Theorem 4.11 (Phase transition in $\text{UG}_n(\mathbf{d})$) *Let \mathbf{d} satisfy Conditions 1.5(a)-(c). Then, the results in Theorem 4.4 also hold for a uniform simple graph with degree sequence \mathbf{d} .*

Proof By [Volume 1, Corollary 7.17], and since $\mathbf{d} = (d_i)_{i \in [n]}$ satisfies Condition 1.5(a)-(c), any event \mathcal{E}_n that occurs whp for $\text{CM}_n(\mathbf{d})$, also occurs whp for $\text{UG}_n(\mathbf{d})$. By Theorem 4.4, the event \mathcal{E}_n that $\{||\mathcal{C}_{\max}|/n - \zeta| \leq \varepsilon\}$ occurs whp for $\text{CM}_n(\mathbf{d})$, so it also holds whp for $\text{UG}_n(\mathbf{d})$. 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 $\text{UG}_n(\mathbf{d})$ for $\nu = \infty$) *Let \mathbf{d} satisfy Conditions 1.5(a)-(b), and assume that there exists $p > 1$ such that*

$$\mathbb{E}[D_n^p] \rightarrow \mathbb{E}[D^p]. \quad (4.2.51)$$

Then, the results in Theorem 4.4 also hold for a uniform simple graph with degree sequence \mathbf{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

$$\mathbb{P}(|\mathcal{C}_{\max}| - \zeta n| > \varepsilon n) \leq e^{-\delta n}, \quad (4.2.52)$$

and

$$\mathbb{P}(|v_k(\mathcal{C}_{\max}) - p_k(1 - \xi^k)n| > \varepsilon n) \leq e^{-\delta n}. \quad (4.2.53)$$

This *exponential concentration* is quite convenient, as it allows us to extend the result to the setting of uniform random graphs by conditioning $\text{CM}_n(\mathbf{d})$ to be simple. Indeed, since Conditions 1.5(a)-(b) imply that

$$\mathbb{P}(\text{CM}_n(\mathbf{d}) \text{ simple}) = e^{-o(n)}, \quad (4.2.54)$$

it follows that the results also hold for the uniform simple random graph $\text{UG}_n(\mathbf{d})$ when Conditions 1.5(a)-(b) hold. See Exercise 4.9 below. \square

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 $\text{GRG}_n(\mathbf{w})$) *Let \mathbf{w} satisfy Condition 1.1(a)-(c). Then, the results in Theorem 4.4 also hold for $\text{GRG}_n(\mathbf{w})$, $\text{CL}_n(\mathbf{w})$ and $\text{NR}_n(\mathbf{w})$.*

Proof Let d_i be the degree of vertex i in $\text{GRG}_n(\mathbf{w})$ 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 $\text{GRG}_n(\mathbf{w})$ conditionally on the degrees \mathbf{d} and $\text{CM}_n(\mathbf{d})$ 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 $\text{GRG}_n(\mathbf{w})$. By [Volume 1, Theorem 6.20], the same result applies to

$\text{CL}_n(\mathbf{w})$, and by [Volume 1, Exercise 6.39], also to $\text{NR}_n(\mathbf{w})$. The fact that Conditions 1.1(a)-(c) imply that Conditions 1.5(a)-(c) hold for $\text{GRG}_n(\mathbf{w})$ is stated in Theorem 1.7. \square

Unfortunately, when $\nu = \infty$, we cannot rely on the fact that by [Volume 1, Theorem 7.18], the law of $\text{GRG}_n(\mathbf{w})$ conditionally on the degrees \mathbf{d} and $\text{CM}_n(\mathbf{d})$ conditionally on being simple agree. Indeed, when $\nu = \infty$, the probability that $\text{CM}_n(\mathbf{d})$ 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 $\text{GRG}_n(\mathbf{w})$ in terms of the edge probabilities can be used rather conveniently:

Theorem 4.14 (Phase transition in $\text{GRG}_n(\mathbf{w})$) *Let \mathbf{w} satisfy Conditions 1.1(a)-(b). Then, the results in Theorem 4.4 also hold for $\text{GRG}_n(\mathbf{w})$, $\text{CL}_n(\mathbf{w})$ and $\text{NR}_n(\mathbf{w})$.*

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 $\text{NR}_n(\mathbf{w})$, the proof for $\text{GRG}_n(\mathbf{w})$ and $\text{CL}_n(\mathbf{w})$ being similar. The required upper bound $|\mathcal{C}_{\max}|/n \leq \zeta_m + o_{\mathbb{P}}(1)$ follows by the local weak convergence in Theorem 3.11 and (2.5.6).

For the lower bound, we bound $\text{NR}_n(\mathbf{w})$ from below by a random graph with edge probabilities

$$p_{ij}^{(K)} = 1 - e^{-(w_i \wedge K)(w_j \wedge K)/\ell_n}. \quad (4.2.55)$$

Therefore, also $|\mathcal{C}_{\max}| \preceq |\mathcal{C}_{\max}^{(K)}|$, 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

$$w_i^{(K)} = (w_i \wedge K) \frac{1}{\ell_n} \sum_{j \in [n]} (w_j \wedge K), \quad (4.2.56)$$

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 $|\mathcal{C}_{\max}|/n \xrightarrow{\mathbb{P}} \zeta$. \square

4.3 CONNECTIVITY OF $\text{CM}_n(\mathbf{d})$

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 $\text{CM}_n(\mathbf{d})$ is whp *connected*, i.e., $\mathcal{C}_{\max} = [n]$ and $|\mathcal{C}_{\max}| = n$.

Our main result shows that $\text{CM}_n(\mathbf{d})$ 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 $\text{CM}_n(\mathbf{d})$) *Assume that Conditions 1.5(a)-(b) hold. Further, assume that $d_i \geq 3$ for every $i \in [n]$. Then $\text{CM}_n(\mathbf{d})$ is connected whp. More precisely,*

$$\mathbb{P}(\text{CM}_n(\mathbf{d}) \text{ disconnected}) = o(1). \quad (4.3.1)$$

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 $\text{CM}_n(\mathbf{d})$ 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 $\text{ER}_n(p)$ 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 $\text{CM}_n(\mathbf{d})$ 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

$$\ell_n(\mathcal{I}) = \sum_{i \in \mathcal{I}} d_i. \quad (4.3.2)$$

Since $d_i \geq 3$, we can use Theorem 4.4 to conclude that most edges are in \mathcal{C}_{\max} , and $\mathcal{I} \neq \mathcal{C}_{\max}$. 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

$$(\ell_n(\mathcal{I}) - 1)!!(\ell_n(\mathcal{I}^c) - 1)!!. \quad (4.3.3)$$

As a result,

$$\begin{aligned} \mathbb{P}(\text{CM}_n(\mathbf{d}) \text{ disconnected}; \mathcal{E}) &\leq \sum_{\mathcal{I} \subset [n]} \frac{(\ell_n(\mathcal{I}) - 1)!! (\ell_n(\mathcal{I}^c) - 1)!!}{(\ell_n - 1)!!} \\ &= \sum_{\mathcal{I} \subset [n]} \prod_{j=1}^{\ell_n(\mathcal{I})/2} \frac{\ell_n(\mathcal{I}) - 2j + 1}{\ell_n - 2j + 1}, \end{aligned} \quad (4.3.4)$$

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

$$f(x) = \frac{\prod_{j=1}^x (2x - 2j + 1)}{\prod_{j=1}^x (\ell_n - 2j + 1)} = \frac{\prod_{i=0}^{x-1} (2i + 1)}{\prod_{k=0}^{x-1} (\ell_n - 2k + 1)} = \prod_{j=0}^{x-1} \frac{2i + 1}{\ell_n - 2i - 1}, \quad (4.3.6)$$

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

$$\mathbb{P}(\text{CM}_n(\mathbf{d}) \text{ disconnected}) \leq \sum_{\mathcal{I} \subset [n]} f(\lceil 3|\mathcal{I}|/2 \rceil) = \sum_{m=1}^{\lfloor n/2 \rfloor} \binom{n}{m} f(\lceil 3m/2 \rceil), \quad (4.3.9)$$

with $m = |\mathcal{I}|$.

Note that, for m odd,

$$\frac{f(2\lceil 3(m+1)/2 \rceil)}{f(2\lceil 3m/2 \rceil)} = \frac{f((3m+1)/2 + 1)}{f((3m+1)/2)} = \frac{3m+3}{\ell_n - 3m - 2}. \quad (4.3.10)$$

while, for m even,

$$\frac{f(\lceil 3(m+1)/2 \rceil)}{f(\lceil 3m/2 \rceil)} = \frac{f(3m/2 + 2)}{f(3m/2)} = \frac{3m+5}{\ell_n - 3m - 5} \frac{3m+3}{\ell_n - 3m - 3}, \quad (4.3.11)$$

Define

$$h_n(m) = \binom{n}{m} f(\lceil 3m/2 \rceil), \quad (4.3.12)$$

so that

$$\mathbb{P}(\text{CM}_n(\mathbf{d}) \text{ disconnected}) \leq \sum_{m=1}^{\lfloor n/2 \rfloor} h_n(m). \quad (4.3.13)$$

Then,

$$\frac{h_n(m+1)}{h_n(m)} = \frac{n-m}{m+1} \frac{f(\lceil 3(m+1)/2 \rceil)}{f(\lceil 3m/2 \rceil)}, \quad (4.3.14)$$

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 $\ell_n \geq 3n$,

$$\frac{h_n(m+1)}{h_n(m)} = \frac{3(n-m)}{\ell_n - 3m - 5} \frac{3m+5}{\ell_n - 3m - 3} \leq \frac{n-m}{n-m-1} \frac{m+2}{n-m-2}. \quad (4.3.16)$$

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 $m \geq 3$,

$$\begin{aligned} h_n(m) &= h_n(3) \prod_{j=3}^m \frac{h_n(j+1)}{h_n(j)} \leq h_n(3) \prod_{j=3}^{\lfloor n/2 \rfloor} (1 + c/n) \\ &\leq h_n(3)(1 + c/n)^{\lfloor n/2 \rfloor} \leq h_n(3)e^{c/2}, \end{aligned} \quad (4.3.18)$$

so that

$$\begin{aligned} \mathbb{P}(\text{CM}_n(\mathbf{d}) \text{ disconnected}; \mathcal{E}) &\leq \sum_{m=1}^{\varepsilon n} h_n(m) \leq h_n(1) + h_n(2) + \sum_{m=3}^{\lfloor n/2 \rfloor} h_n(m) \\ &\leq h_n(1) + h_n(2) + nh_n(3)e^{c/2}/2. \end{aligned} \quad (4.3.19)$$

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 $\lceil 3m/2 \rceil = 5$ 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

$$\mathbb{P}(\text{CM}_n(\mathbf{d}) \text{ disconnected}; \mathcal{E}) = O(1/n). \quad (4.3.21)$$

This is even stronger than required, and one would be tempted to believe that also $\mathbb{P}(\text{CM}_n(\mathbf{d}) \text{ disconnected}) = O(1/n)$, but this proof does not show this since we start with the assumption that the non-giant (collection of) component(s) \mathcal{I} satisfies $\ell_n(\mathcal{I}) \leq \ell_n/4$. \square

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 $\text{CM}_n(\mathbf{d})$ is with positive probability *disconnected* when either $n_1 \gg n^{1/2}$:

Proposition 4.16 (Disconnectivity of $\text{CM}_n(\mathbf{d})$ when $n_1 \gg n^{1/2}$) *Let Condition 1.5(a)-(b) hold, and assume that $n_1 \gg n^{1/2}$. Then,*

$$\lim_{n \rightarrow \infty} \mathbb{P}(\text{CM}_n(\mathbf{d}) \text{ connected}) = 0. \quad (4.3.22)$$

Proof We note that $\text{CM}_n(\mathbf{d})$ 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,

$$\begin{aligned} \mathbb{P}(\text{CM}_n(\mathbf{d}) \text{ contains no 2-pair}) &= \prod_{i=1}^{n_1} \frac{\ell_n - n_1 - i + 1}{\ell_n - 2i + 1} \\ &= \prod_{i=1}^{n_1} \left(1 - \frac{n_1 - i}{\ell_n - 2i + 1}\right). \end{aligned} \quad (4.3.23)$$

Since, for each $i \geq 1$,

$$1 - \frac{n_1 - i}{\ell_n - 2i + 1} \leq 1 - \frac{n_1 - i}{\ell_n} \leq e^{-(n_1 - i)/\ell_n}, \quad (4.3.24)$$

we arrive at

$$\begin{aligned} \mathbb{P}(\text{CM}_n(\mathbf{d}) \text{ contains no 2-pair}) &\leq \prod_{i=1}^{n_1} e^{-(n_1 - i)/\ell_n} \\ &= e^{-n_1(n_1 - 1)/[2\ell_n]} = o(1), \end{aligned} \quad (4.3.25)$$

since $\ell_n = \Theta(n)$ and $n_1 \gg n^{1/2}$. □

Proposition 4.17 (Disconnectivity of $\text{CM}_n(\mathbf{d})$ when $p_2 > 0$) *Let Conditions 1.5(a)-(b) hold, and assume that $p_2 > 0$. Then,*

$$\limsup_{n \rightarrow \infty} \mathbb{P}(\text{CM}_n(\mathbf{d}) \text{ connected}) < 1. \quad (4.3.26)$$

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

$$\mathbb{E}[P(2)] = \frac{2n_2(n_2 - 1)}{2(\ell_n - 1)(\ell_n - 3)}, \quad (4.3.27)$$

since there are $n_2(n_2 - 1)/2$ pairs of vertices of degree 2, and the probability

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 $n_2/n \rightarrow p_2$,

$$\mathbb{E}[P(2)] \rightarrow p_2^2/\mathbb{E}[D]^2 \equiv \lambda_2. \quad (4.3.28)$$

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

$$\mathbb{P}(\text{CM}_n(\mathbf{d}) \text{ disconnected}) \geq \mathbb{P}(P(2) > 0) \rightarrow 1 - e^{-\lambda_2} > 0, \quad (4.3.29)$$

as required. The proof that Theorem 2.6 can be applied is left as Exercise 4.11 below. \square

We close this section with a detailed result on the size of the giant component when $d_{\min} \geq 2$:

Theorem 4.18 (Connectivity of $\text{CM}_n(\mathbf{d})$ 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*

$$n - |\mathcal{C}_{\max}| \xrightarrow{d} \sum_{k \geq 2} kX_k, \quad (4.3.30)$$

where $(X_k)_{k \geq 2}$ are independent Poisson random variables with parameters $\lambda_2^k/(2k)$ with $\lambda_2 = 2p_2/\mathbb{E}[D]$. Consequently,

$$\mathbb{P}(\text{CM}_n(\mathbf{d}) \text{ connected}) \rightarrow e^{-\sum_{k \geq 2} \lambda_2^k/(2k)} \in (0, 1). \quad (4.3.31)$$

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

$$n - |\mathcal{C}_{\max}| \geq \sum_{k \geq 2} kP(k). \quad (4.3.32)$$

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 $\lambda_2^k/(2k) = \lim_n \mathbb{E}[P(k)]$. 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}_{\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. \square

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 $\text{CM}_n(\mathbf{d})$ is closely related to the largest degree:

Theorem 4.19 (Subcritical phase for $\text{CM}_n(\mathbf{d})$) *Let \mathbf{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*

$$[1 - F_n](x) \leq c_2 x^{-(\tau-1)}. \quad (4.4.1)$$

Then, for $\text{CM}_n(\mathbf{d})$ with $d_{\max} = \max_{j \in [n]} d_j$,

$$|\mathcal{C}_{\max}| = \frac{d_{\max}}{1 - \nu} + o_p(n^{1/(\tau-1)}). \quad (4.4.2)$$

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 $d_{\max} = \Theta(n^{1/(\tau-1)})$, as it would be in the case when the degrees obey a power law with exponent τ (for example, in the case where the degrees are i.i.d.). When $d_{\max} = o(n^{1/(\tau-1)})$, instead, Theorem 4.19 implies that $|\mathcal{C}_{\max}| = o_p(n^{1/(\tau-1)})$, a less precise result. Bear in mind that Theorem 4.19 only gives sharp asymptotics of $|\mathcal{C}_{\max}|$ when $d_{\max} = \Theta(n^{1/(\tau-1)})$ (see Exercise 4.19).

The intuition behind Theorem 4.19 is that from the vertex of maximal degree, there are d_{\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_{\max} trees are close to d_{\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 $|\mathcal{C}_{\max}| = d_{\max}/(1-\nu)(1+o_{\mathbb{P}}(1))$ 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 $\text{CM}_n(\mathbf{d})$, and these are further extended in van der Hofstad et al. (Preprint 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 $\text{CM}_n(\mathbf{d})$ with finite third moments) *Let \mathbf{d} satisfy Condition 1.5(a)-(c) with $\nu = \mathbb{E}[D(D-1)]/\mathbb{E}[D] = 1$.*

(a) *Assume further that $\alpha_n = \nu_n - 1 = \mathbb{E}[D_n(D_n - 2)]/\mathbb{E}[D_n] > 0$ is such that $\alpha_n \gg n^{-1/3}$, and that*

$$\mathbb{E}[D_n^{3+\varepsilon}] = O(1) \tag{4.4.3}$$

for some $\varepsilon > 0$. Let $\beta = \mathbb{E}[D(D-1)(D-2)]/\mathbb{E}[D] > 0$. Then, $\text{CM}_n(\mathbf{d})$ satisfies

$$\begin{aligned} |\mathcal{C}_{\max}| &= \frac{2}{\mathbb{E}[D]\beta} n\alpha_n + o_{\mathbb{P}}(n\alpha_n), \\ |v_k(\mathcal{C}_{\max})| &= \frac{2\mathbb{E}[D]}{\beta} k p_k n\alpha_n + o_{\mathbb{P}}(n\alpha_n), \text{ for every } k \geq 0, \\ |E(\mathcal{C}_{\max})| &= \frac{2\mathbb{E}[D]^2}{\beta} n\alpha_n + o_{\mathbb{P}}(n\alpha_n), \end{aligned}$$

while $|\mathcal{C}_{(2)}| = o_{\mathbb{P}}(n\alpha_n)$ and $|E(\mathcal{C}_{(2)})| = o_{\mathbb{P}}(n\alpha_n)$.

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 $\text{CM}_n(\mathbf{d})$ with infinite third moments) *Let \mathbf{d} satisfy Condition 1.5(a)-(c) with $\nu = \mathbb{E}[D(D-1)]/\mathbb{E}[D] = 1$. Let ζ_n^* denote the survival probability of a branching process with offspring distribution $D_n^* - 1$. Assume further that $\alpha_n = \nu_n - 1 = \mathbb{E}[D_n(D_n - 2)]/\mathbb{E}[D_n] > 0$ is such that $\alpha_n \gg n^{-1/3}(\mathbb{E}[D_n^3])^{2/3}$. Then, $\text{CM}_n(\mathbf{d})$ satisfies*

$$\begin{aligned} |\mathcal{C}_{\max}| &= \mathbb{E}[D]\zeta_n^* n(1 + o_{\mathbb{P}}(1)), \\ |v_k(\mathcal{C}_{\max})| &= k p_k \zeta_n^* n(1 + o_{\mathbb{P}}(1)), \text{ for every } k \geq 0, \\ |E(\mathcal{C}_{\max})| &= \mathbb{E}[D]\zeta_n^* n(1 + o_{\mathbb{P}}(1)), \end{aligned}$$

while $|\mathcal{C}_{(2)}| = o_{\mathbb{P}}(n\zeta_n^)$ and $|E(\mathcal{C}_{(2)})| = o_{\mathbb{P}}(n\zeta_n^*)$.*

The asymptotics of $|\mathcal{C}_{\max}|$ in Theorem 4.20 can be understood by the fact that, for a branching process with offspring distribution X having mean $\mathbb{E}[X] = 1 + \varepsilon$ where ε is small, the survival probability ζ satisfies $\zeta = 2\varepsilon/\text{Var}(X)(1 + o(1))$ (see Exercise 4.25). Therefore, the survival probability ζ^* of the branching process with offspring distribution $D^* - 1$ is close to $2\varepsilon/\beta$, where we note that $\beta = \text{Var}(D^* - 1) = \text{Var}(D^*)$. Since the limit of $|\mathcal{C}_{\max}|/n$ ζ satisfies

$$\zeta = \sum_{k=1}^{\infty} p_k (1 - (1 - \zeta^*)^k), \quad (4.4.4)$$

we further obtain that

$$\zeta = \zeta^* \mathbb{E}[D](1 + o(1)). \quad (4.4.5)$$

The results on $|v_k(\mathcal{C}_{\max})|$ and $|E(\mathcal{C}_{\max})|$ 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 $\zeta_n^* = \Theta(\alpha_n^{1/(\tau-3)})$ (recall also Exercise 4.26), and the restriction on α_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 $\text{CM}_n(\mathbf{d})$ for i.i.d. degrees:

Theorem 4.22 (Weak convergence of the ordered critical clusters: finite third moments) *Let $\mathbf{d} = (d_i)_{i \in [n]}$ be a sequence of i.i.d. random variables having the same distribution as D satisfying $\nu = \mathbb{E}[D(D-1)]/\mathbb{E}[D] = 1$. Let $(|\mathcal{C}_{(i)}|)_{i \geq 1}$ denote the clusters of $\text{CM}_n(\mathbf{d})$, ordered in size. Let $\mathbb{E}[D^3] < \infty$. Then, as $n \rightarrow \infty$,*

$$(n^{-2/3}|\mathcal{C}_{(i)}|)_{i \geq 1} \xrightarrow{d} (\gamma_i)_{i \geq 1}, \quad (4.4.6)$$

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-Rényi 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-Rényi 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 $\mathbf{d} = (d_i)_{i \in [n]}$ be a sequence of i.i.d. random variables having the same distribution as D satisfying $\nu = \mathbb{E}[D(D-1)]/\mathbb{E}[D] = 1$. Let $(|\mathcal{C}_{(i)}|)_{i \geq 1}$ denote the clusters of*

$\text{CM}_n(\mathbf{d})$, ordered in size. Let the distribution function F of D satisfy that there exists a $\tau \in (3, 4)$ and $0 < c_F < \infty$ such that

$$\lim_{x \rightarrow \infty} x^{\tau-1}[1 - F(x)] = c_F. \quad (4.4.7)$$

Then, as $n \rightarrow \infty$,

$$(n^{-(\tau-2)/(\tau-1)}|\mathcal{C}_{(i)}|)_{i \geq 1} \xrightarrow{d} (\gamma_i)_{i \geq 1}, \quad (4.4.8)$$

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^ρ , 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 (2010) 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 Bollobás 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 εn of $v_k(\mathcal{C}_{\max})$ is exponentially small for the configuration model. Since the probability of simplicity in $\text{CM}_n(\mathbf{d})$ 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 $d_{\min} \geq 2$ and Wormald (1981), who proved r -connectivity when $d_{\min} = r$.

Notes on Section 4.4

Theorem 4.19 is proved by Janson (2008). (Janson, 2008, Theorem 1.1) shows that $|\mathcal{C}_{\max}| \leq An^{1/(\tau-1)}$ 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 $d_{(j)}/(1-\nu) + o_p(n^{1/(\tau-1)})$.

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 $\mathbb{E}[D_n^3] \rightarrow \mathbb{E}[D^3]$ by van der Hofstad et al. (Preprint 2016). Interestingly, the behavior is *different* when Conditions 1.5(a)-(b) hold, but $\mathbb{E}[D_n^3] \rightarrow \infty$ sufficiently fast. Also this case is studied by van der Hofstad et al. (Preprint 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. (Preprint (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 $\mathbb{P}(B_k^{(n)}(o_1) \simeq \mathbf{t}, o_2 \in B_{2k}^{(n)}(o_1)) \rightarrow 0$, 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*

let $\mathcal{C}(1)$ denote the cluster size of vertex 1. Show that

$$|\mathcal{C}(1)|/n \xrightarrow{d} T, \quad (4.6.1)$$

where $\mathbb{P}(T \leq x) = 1 - \sqrt{1-x}$.

Exercise 4.4 (Cluster size in a 2-regular graph with some degree-1 vertices) Let $n_1 \rightarrow \infty$ with $n_1/n \rightarrow 0$, and $n_2 = n - n_1$. Let $\mathcal{C}(1)$ denote the cluster size of vertex 1. Show that

$$|\mathcal{C}(1)|/n \xrightarrow{\mathbb{P}} 0. \quad (4.6.2)$$

Exercise 4.5 (Cluster size in a 2-regular graph with some degree-4 vertices) Let $n_4 \rightarrow \infty$ with $n_4/n \rightarrow 0$, and $n_2 = n - n_4$. Let $\mathcal{C}(1)$ denote the cluster size of vertex 1. Show that

$$|\mathcal{C}(1)|/n \xrightarrow{\mathbb{P}} 1. \quad (4.6.3)$$

Exercise 4.6 (Limiting constants) Prove that for $t_1 = 0$ and $t_2 = \theta$, $e^{-kt_1} = 1$, $e^{-kt_2} = \xi$, $G_D(e^{-t_1}) = 1$, $G_D(e^{-t_2}) = 1 - \zeta$, $h(e^{-t_1}) = 2\mathbb{E}[D]$, $h(e^{-t_2}) = 2\mathbb{E}[D]\xi^2$.

Exercise 4.7 (Theorem 4.4 for $\text{GRG}_n(\mathbf{w})$ with $\nu = \infty$) Prove that all statements in Theorem 4.4 hold for the $\text{GRG}_n(\mathbf{w})$ in Theorem 4.14 when $\nu = \infty$.

Exercise 4.8 (Number of vertices with degree k) Let \mathbf{w} satisfy Conditions 1.1(a)-(b). Adapt the proof of $|\mathcal{C}_{\max}|/n \xrightarrow{\mathbb{P}} \zeta$ to show that also $v_k(\mathcal{C}_{\max})/n \xrightarrow{\mathbb{P}} p_k(1 - \xi^k)$ for $\text{NR}_n(\mathbf{w})$.

Exercise 4.9 (Phase transition in $\text{UG}_n(\mathbf{d})$ 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 $\lambda_2 = 2p_2^2/\mathbb{E}[D]^2$,

$$\mathbb{E}[(P(2))_k] \rightarrow \lambda_2^k. \quad (4.6.4)$$

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 $i \in [n]$,

$$\mathbb{P}(\text{there exists an isolated vertex}) \leq \frac{3n}{(2\ell_n - 1)(2\ell_n - 3)}. \quad (4.6.5)$$

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 $i \in [n]$,

$$\mathbb{P}(\text{there exists a cluster of size 2}) \leq \frac{15n(n-1)}{(2\ell_n - 1)(2\ell_n - 3)(2\ell_n - 5)}. \quad (4.6.6)$$

Exercise 4.15 (Cycles in $\text{CM}_n(\mathbf{d})$) Let $P(k)$ denote the number of k -cycles consisting of degree 2 vertices, for $k \geq 2$. Let $\lambda_2 = 2p_2/\mathbb{E}[D]$. 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 $\lambda_2^k/(2k) = \lim_n \mathbb{E}[P(k)]$.

Exercise 4.16 (\mathcal{C}_{\max} when $d_{\min} = 2$) Consider $\text{CM}_n(\mathbf{d})$ with $d_{\min} = 2$ and assume that $\mathbb{P}(D \geq 3) > 0$. Show that Theorem 4.18 holds if $\mathbb{P}(\exists v: d_v \geq 3 \text{ and } v \notin \mathcal{C}_{\max}) = o(1)$.

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 $\frac{k(k-1)}{2\mathbb{E}[D]^2} \sum_{\ell \geq 1} (2p_2/\mathbb{E}[D])^\ell$.

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 $|\mathcal{C}_{\max}| = d_{\max}/(1-\nu)(1+o_{\mathbb{P}}(1))$ precisely when $d_{\max} = \Theta(n^{1/(\tau-1)})$.

Exercise 4.20 (Sub-polynomial subcritical clusters) Use Theorem 4.19 to prove that $|\mathcal{C}_{\max}| = o_{\mathbb{P}}(n^\varepsilon)$ for every $\varepsilon > 0$ when (4.4.1) holds for every $\tau > 1$ (where the constant c_2 is allowed to depend on τ). 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 $|\mathcal{C}_{\max}| = d_{\max}/(1-\nu)(1+o_{\mathbb{P}}(1))$ 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 $|\mathcal{C}_{\max}| = d_{\max}/(1-\nu)(1+o_{\mathbb{P}}(1))$ to hold?

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 $\mathbb{E}[D_n^3] \rightarrow \mathbb{E}[D^3]$.*

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 ζ_n^* is the survival probability of the branching process with offspring distribution $D_n^* - 1$. Show that $\zeta_n^* = \Theta(\alpha_n^{1/(\tau-3)})$ and thus that $|\mathcal{C}_{\max}| = \Theta(\alpha_n^{1/(\tau-3)}n)$. 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 weak 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 5.3, we investigate the connectivity of $\text{PA}_t^{(m,\delta)}$. Section 5.4 highlights some further results for preferential attachment models. We close this chapter in Section 5.5 with notes and discussion, and in Section 5.6 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 $(\text{PA}_t^{(m,\delta)})_{t \geq 1}$, 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 $(\text{PA}_{1,\delta/m}(t))_{t \geq 1}$. We sometimes also discuss other variants of the model, such as $(\text{PA}_{m,\delta}^{(b)}(t))_{t \geq 1}$, 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

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).

Clearly, when a sequence of random variables is i.i.d., then it is also exchangeable (see Exercise 5.2). Remarkably, however, the distribution of an infinite sequence of random variables is always a *mixture* of independent and identically distributed 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 5.1 (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 $\mathbb{P}(U \in [0, 1]) = 1$ such that, for all $1 \leq k \leq n$*

$$\mathbb{P}(X_1 = \cdots = X_k = 1, X_{k+1} = \cdots = X_n = 0) = \mathbb{E}[U^k(1-U)^{n-k}]. \quad (5.1.1)$$

De Finetti's Theorem (Theorem 5.1) 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 the proof we see that Theorem 5.1 in fact holds more generally, for example, when X_i takes on at most a *finite* number of values. Since we only rely on Theorem 5.1 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.,

$$S_n = \sum_{k=1}^n X_k. \quad (5.1.2)$$

Then Theorem 5.1 is equivalent to the statement that

$$\mathbb{P}(S_n = k) = \mathbb{E}\left[\mathbb{P}(\text{Bin}(n, U) = k)\right]. \quad (5.1.3)$$

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

$$\lim_{n \rightarrow \infty} \mathbb{P}(S_n \in (an, bn)) = \int_a^b f(u) du, \quad (5.1.4)$$

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 5.1). Equation (5.1.4) follows by noting that $S_n/n \xrightarrow{a.s.} U$ by the strong law of large numbers applied to the conditional law given U . In Exercise 5.3, you are asked to fill in the details.

Proof of Theorem 5.1. 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

$$\begin{aligned} \mathbb{P}(X_1 = \cdots = X_k = 1, X_{k+1} = \cdots = X_n = 0) \\ = \sum_{j=k}^m \mathbb{P}(X_1 = \cdots = X_k = 1, X_{k+1} = \cdots = X_n = 0 \mid S_m = j) \mathbb{P}(S_m = j). \end{aligned} \quad (5.1.5)$$

By exchangeability and conditionally on $S_m = j$, each sequence $(X_i)_{i=1}^m$ 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

$$\mathbb{P}(X_1 = \cdots = X_k = 1, X_{k+1} = \cdots = X_n = 0 \mid S_m = j) = \frac{\binom{m-n}{j-k}}{\binom{m}{j}}. \quad (5.1.6)$$

Writing $(m)_k = m \cdot (m-1) \cdots (m-k+1)$ for the k th factorial moment of m , we therefore arrive at

$$\begin{aligned} \mathbb{P}(X_1 = \cdots = X_k = 1, X_{k+1} = \cdots = X_n = 0) \\ = \sum_{j=k}^m \frac{(j)_k (m-n)_j}{(m)_j} \mathbb{P}(S_m = j). \end{aligned} \quad (5.1.7)$$

When $m \rightarrow \infty$ and for k and n with $k \leq n$ fixed,

$$\frac{(j)_k (m-n)_j}{(m)_j} = \left(\frac{j}{m}\right)^k \left(1 - \frac{j}{m}\right)^{n-k} + o(1), \quad (5.1.8)$$

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, $[j]_k \leq (\varepsilon m)^k$.

Recall that $S_m = j$, so that

$$\begin{aligned} \mathbb{P}(X_1 = \cdots = X_k = 1, X_{k+1} = \cdots = X_n = 0) \\ = \lim_{m \rightarrow \infty} \mathbb{E}[Y_m^k (1 - Y_m)^{n-k}], \end{aligned} \quad (5.1.9)$$

where $Y_m = S_m/m$. Note that it is here that we make use of the fact that $(X_i)_{i \geq 1}$ 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 $\lim_{l \rightarrow \infty} m_l = \infty$ and a random variable U such that $Y_{m_l} \xrightarrow{d} U$. Since the random variable $Y_m^k (1 - Y_m)^{n-k}$ is uniformly bounded for each k, n , Lebeque's Dominated Convergence Theorem ([Volume 1, Theorem A.1]) gives that

$$\begin{aligned} \lim_{m \rightarrow \infty} \mathbb{E}[Y_m^k (1 - Y_m)^{n-k}] &= \lim_{l \rightarrow \infty} \mathbb{E}[Y_{m_l}^k (1 - Y_{m_l})^{n-k}] \\ &= \mathbb{E}[U^k (1 - U)^{n-k}]. \end{aligned} \quad (5.1.10)$$

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 $\lim_{l \rightarrow \infty} m'_l = \infty$ 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 $Y_m \xrightarrow{d} U$. \square

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. 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,

$$\begin{aligned} \mathbb{P}(X_1 = X_2 = 1) &= \frac{b(b-1)}{(b+r)(b+r-1)} \\ &< \left(\frac{b}{b+r}\right)^2 = \mathbb{P}(X_1 = 1)\mathbb{P}(X_2 = 1), \end{aligned} \quad (5.1.11)$$

so that X_1 and X_2 are *negatively* correlated. As a result, (5.1.1) fails.

Pólya urn schemes

An important application of De Finetti's Theorem (Theorem 5.1) 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 $W_b, W_r: \mathbb{N} \rightarrow (0, \infty)$ 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 5.2 (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 $n \rightarrow \infty$,*

$$\frac{B_n}{B_n + R_n} \xrightarrow{a.s.} U, \quad (5.1.14)$$

where U has a Beta-distribution with parameters $a = b_0 + a_b$ and $b = r_0 + a_r$, and

$$\mathbb{P}(B_n = B_0 + k) = \mathbb{E} \left[\mathbb{P}(\text{Bin}(n, U) = k) \right]. \quad (5.1.15)$$

Before proving Theorem 5.2, 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 5.2 implies that this is a mere binomial experiment.

Proof of Theorem 5.2. 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 $(x_t)_{t=1}^n$,

$$\mathbb{P}((X_t)_{t=1}^n = (x_t)_{t=1}^n) = \prod_{t=1}^n \frac{W_b(b_t)^{x_t} W_r(r_t)^{1-x_t}}{W_b(b_t) + W_r(r_t)}, \quad (5.1.17)$$

where $b_t = b_0 + \sum_{j=1}^t x_j$ and $r_t = R_0 - B_0 + t - b_t$. Denote $k = \sum_{t=1}^n x_t$. Then, by (5.1.13) and (5.1.16),

$$\prod_{t=1}^n (W_b(b_t) + W_r(r_t)) = \prod_{t=0}^{n-1} (b_0 + r_0 + a_b + a_r + t), \quad (5.1.18)$$

while

$$\prod_{t=1}^n W_b(b_t)^{x_t} = \prod_{m=0}^{k-1} (b_0 + a_b + m), \quad \prod_{t=1}^n W_r(r_t)^{1-x_t} = \prod_{j=0}^{n-k-1} (r_0 + a_r + j). \quad (5.1.19)$$

Thus, we arrive at

$$\mathbb{P}((X_t)_{t=1}^n = (x_t)_{t=1}^n) = \frac{\prod_{m=0}^{k-1} (b+m) \prod_{j=0}^{n-k-1} (r+j)}{\prod_{t=0}^{n-1} (b+r+t)}, \quad (5.1.20)$$

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 5.1), 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,

$$\begin{aligned} \mathbb{P}(S_n = k) &= \binom{n}{k} \frac{\prod_{m=0}^{k-1} (b+m) \prod_{j=0}^{n-k-1} (r+j)}{\prod_{t=0}^{n-1} (b+r+t)} \\ &= \frac{\Gamma(n+1)}{\Gamma(k+1)\Gamma(n-k+1)} \times \frac{\Gamma(k+b)}{\Gamma(b)} \times \frac{\Gamma(n-k+r)}{\Gamma(r)} \times \frac{\Gamma(b+r)}{\Gamma(n+b+r)} \\ &= \frac{\Gamma(b+r)}{\Gamma(r)\Gamma(b)} \times \frac{\Gamma(k+b)}{\Gamma(k+1)} \times \frac{\Gamma(n-k+r)}{\Gamma(n-k+1)} \times \frac{\Gamma(n+1)}{\Gamma(n+b+r)}. \end{aligned} \quad (5.1.21)$$

For k and $n-k$ large, by [Volume 1, (8.3.9)],

$$\mathbb{P}(S_n = k) = \frac{\Gamma(b+r)}{\Gamma(r)\Gamma(b)} \frac{k^{b-1} (n-k)^{r-1}}{n^{b+r-1}} (1 + o(1)). \quad (5.1.22)$$

Taking $k = \lceil un \rceil$ (recall (5.1.4))

$$\lim_{n \rightarrow \infty} n \mathbb{P}(S_n = \lceil un \rceil) = \frac{\Gamma(b+r)}{\Gamma(r)\Gamma(b)} u^{b-1} (1-u)^{r-1}, \quad (5.1.23)$$

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). \square

Multiple urn extensions

We next explain how Theorem 5.2 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

$$W_i(k) = a_i + k. \quad (5.1.24)$$

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 $a_{[j, \ell]} = \sum_{i=j}^{\ell} a_i$ and $C_{[j, \ell]}(n) =$

$\sum_{j=i}^{\ell} C_j(n)$. We first view the balls of color 1 and the other colors as a two type urn. Thus,

$$\frac{C_1(n)}{n} \xrightarrow{a.s.} U_1, \tag{5.1.25}$$

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 $2, \dots, \ell$. This implies that

$$\frac{C_2(n)}{n} \xrightarrow{a.s.} U_2(1 - U_1), \tag{5.1.26}$$

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

$$\frac{C_i(n)}{n} \xrightarrow{a.s.} U_i \prod_{j=1}^{i-1} (1 - U_j), \tag{5.1.27}$$

where U_i is independent of (U_1, \dots, 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 5.2 to urns with multiple colors, but also gives an appealing independence structure of the various limits.

Applications to scale-free trees

We close this section by discussing applications of Pólya urn schemes to scale-free trees. We start at time $t = 2$ 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 $\delta > -1$. We do not allow for self-loops, so that indeed we obtain a tree. This is a generalization of $(PA_t^{(1,\delta)}(b))_{t \geq 2}^\infty$, in which we are more flexible in choosing the initial graph. The model for $(PA_t^{(1,\delta)}(b))_{t \geq 1}$ arises when $d_1 = 3$ and $d_2 = 1$ (see Exercise 5.8).

We decompose the growing tree in two trees. For $i = 1, 2$, we let $T_i(t)$ be the tree of vertices that are closer to i than to $3 - i$. Thus, the tree $T_2(t)$ consists of those vertices for which the path in the tree from the vertex to the root passes through vertex 2, and $T_1(t)$ consists of the remainder of the scale-free tree. Let $S_i(t) = |T_i(t)|$ denote the number of vertices in $T_i(t)$. Clearly, $S_1(t) + S_2(t) = t$, which is the total number of vertices of the tree at time t . We can apply Theorem 5.2 to describe the relative sizes of $T_1(t)$ and $T_2(t)$:

Theorem 5.3 (Tree decomposition for scale-free trees) *As $t \rightarrow \infty$,*

$$\frac{S_1(t)}{S_1(t) + S_2(t)} \xrightarrow{a.s.} U, \tag{5.1.28}$$

where U has a Beta-distribution with parameters $a = (d_1 + \delta)/(2 + \delta)$ and $b = (d_2 + \delta)/(2 + \delta)$, and

$$\mathbb{P}(S_1(t) = k) = \mathbb{E} \left[\mathbb{P}(\text{Bin}(t-1, U) = k-1) \right]. \quad (5.1.29)$$

By Theorem 5.3, 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 5.3. The evolution of $(S_1(t))_{t \geq 2}$ can be viewed as a Pólya urn scheme. Indeed, when $S_1(t) = s_1(t)$, then the probability of attaching the $(t+1)$ st vertex to $T_1(t)$ is equal to

$$\frac{(2s_1(t) + d_1 - 2) + \delta s_1(t)}{(2s_1(t) + d_1 - 2) + \delta s_1(t) + 2(s_2(t) + d_2) + \delta s_2(t)}, \quad (5.1.30)$$

since the number of vertices in $T_i(t)$ equals $S_i(t)$, while the total degree of $T_i(t)$ equals $(2S_i(t) + d_i - 2)$. We can rewrite this as

$$\frac{s_1(t) + (d_1 - 2)/(2 + \delta)}{s_1(t) + s_2(t) + (d_1 + d_2 - 4)/(2 + \delta)}, \quad (5.1.31)$$

which is equal to (5.1.12) in the case (5.1.13) when $r_0 = b_0 = 1$ and $a_b = (d_1 - 2)/(2 + \delta)$, $a_r = (d_2 - 2)/(2 + \delta)$. Therefore, the proof of Theorem 5.3 follows directly from Theorem 5.2. \square

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 $t \geq k$ at which an edge is attached to one of the k initial vertices. We work with $(\text{PA}_t^{(1,\delta)})_{t \geq 1}$, so that we start at time $t = 1$ with one vertices 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 5.4 (Relative degrees in scale-free trees) *For $(\text{PA}_t^{(1,\delta)})_{t \geq 1}$, as $t \rightarrow \infty$,*

$$\frac{D_k(t)}{D_1(t) + \cdots + D_k(t)} \xrightarrow{a.s.} B_k, \quad (5.1.32)$$

where B_k has a Beta-distribution with parameters $a = 1 + \delta$ and $b = (k-1)(2 + \delta)$.

By Theorem 1.11, $D_k(t)t^{-1/(2+\delta)} \xrightarrow{a.s.} \xi_k$, where ξ_k is positive almost surely by the argument in the proof of [Volume 1, Theorem 8.14]. It thus follows from Theorem 5.4 that $B_k = \xi_k/(\xi_1 + \cdots + \xi_k)$. We conclude that Theorem 5.4 allows to identify properties of the law of the limiting degrees.

Proof of Theorem 5.4. Denote the sequence of stopping times $(\tau_k(n))_{n \geq 2k-1}$, by $\tau_k(2k-1) = k-1$, and

$$\tau_k(n) = \inf\{t: D_1(t) + \cdots + D_k(t) = n\}, \quad (5.1.33)$$

i.e., $\tau_k(n)$ is the time where the total degree of vertices $1, \dots, k$ equals n . Since $D_j(n) \xrightarrow{a.s.} \infty$ as $n \rightarrow \infty$ for every j , $\tau_k(n) < \infty$ for every n . Moreover, since $\tau_k(n) \xrightarrow{a.s.} \infty$,

$$\begin{aligned} \lim_{t \rightarrow \infty} \frac{D_k(t)}{D_1(t) + \dots + D_k(t)} &= \lim_{n \rightarrow \infty} \frac{D_k(\tau_k(n))}{D_1(\tau_k(n)) + \dots + D_k(\tau_k(n))} \\ &= \lim_{n \rightarrow \infty} \frac{D_k(\tau_k(n))}{n}. \end{aligned} \tag{5.1.34}$$

Now, the random variables $((D_k(\tau_k(n)), D_1(\tau_k(n)) + \dots + D_{k-1}(\tau_k(n))))_{n \geq 2k-1}$ form a Pólya urn scheme, with $D_k(\tau_k(2k-1)) = 1$, and $D_1(\tau_k(2k-1)) + \dots + D_{k-1}(\tau_k(2k-1)) = 2k-2$. The edge at time $\tau_k(n)$ is attached to vertex k with probability

$$\frac{D_k(\tau_k(n)) + \delta}{n + k\delta}, \tag{5.1.35}$$

which are the probabilities of a Pólya urn scheme in the linear weight case in (5.1.13) when $a_b = \delta, a_r = (k-1)\delta, b_0 = 1, r_0 = 2(k-1)$. Thus, the statement follows from Theorem 5.2. \square

Theorem 5.4 is easily extended to $(\text{PA}_t^{(1,\delta)}(b))_{t \geq 1}$:

Theorem 5.5 (Relative degrees in scale-free trees) *For $(\text{PA}_t^{(1,\delta)}(b))_{t \geq 1}$, as $t \rightarrow \infty$,*

$$\frac{D_k(t)}{D_1(t) + \dots + D_k(t)} \xrightarrow{a.s.} B_k, \tag{5.1.36}$$

where B_k has a Beta-distribution with parameters $a = 1 + \delta$ and $b = (2k-1) + (k-1)\delta$.

Of course, the dynamics for $(\text{PA}_t^{(1,\delta)}(b))_{t \geq 1}$ are slightly different. Indeed, now the random variables $((D_k(\tau_k(n)), D_1(\tau_k(n)) + \dots + D_{k-1}(\tau_k(n))))_{n \geq 2k-1}$ form a Pólya urn scheme, with $D_k(\tau_k(2k)) = 1$, and $D_1(\tau_k(2k)) + \dots + D_{k-1}(\tau_k(2k)) = 2k-1$. The edge at time $\tau_k(n)$ is attached to vertex k with probability

$$\frac{D_k(\tau_k(n)) + \delta}{n + k\delta}, \tag{5.1.37}$$

which are the probabilities of a Pólya urn scheme in the linear weight case in (5.1.13) when $a_b = \delta, a_r = (k-1)\delta, b_0 = 1, r_0 = 2k-1$. Thus, again the statement follows from Theorem 5.2. See Exercise 5.10 for the complete proof. Remarkably, even though $(\text{PA}_t^{(1,\delta)})_{t \geq 1}$ and $(\text{PA}_t^{(1,\delta)}(b))_{t \geq 1}$ have the same asymptotic degree distribution, the limiting degrees in Theorems 5.4 and 5.5 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 multitype 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 $(\text{PA}_t^{(m,\delta)}(c))_{t \geq 1}$, 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 τ_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 β . 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 β* , $D_1(\tau_k)$ 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 $\text{PA}_n^{(m,\delta)}(c)$ can again be given rather explicitly in terms of the arising random variables. Let us now give some details.

We continue by introducing the limiting graph. Let

$$u = \frac{\delta}{2m}, \tag{5.2.1}$$

and write

$$\chi = \frac{1 + 2u}{1 + u} = \frac{2(m + \delta)}{2m + \delta}, \quad \psi = \frac{1 - \chi}{\chi} = \frac{1}{1 + 2u} = \frac{m}{m + \delta}. \tag{5.2.2}$$

The exponent ψ 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 $\text{PA}_t^{(m,\delta)}(c)$ 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 $\text{PA}_n^{(m,\delta)}(c)$ 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_n 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, o) be defined by its vertices labeled by finite words, as in the Ulam-Harris labelling of trees,

$$w = w_1 w_2 \cdots w_l, \tag{5.2.3}$$

each carrying a label R or L , which are defined recursively as follows: The root \emptyset has a position $X_\emptyset = U_\emptyset^\chi$, where U_\emptyset is chosen uniformly at random in $[0, 1]$. Here the uniform random variable U_\emptyset corresponds to the uniform vertex from which we start exploring the graph. Thus, we can think of the uniform random variable U_\emptyset as corresponding to the age of the vertex. The ‘location’ of a vertex in the tree will correspond to the age to the power χ , and will be a more convenient parametrization of the ages of the vertices involved.

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 wj be the j th child of w , i.e., $wj = \emptyset w_1 w_2 \cdots w_l j$, and set

$$m_-(w) = \begin{cases} m & \text{if } w \text{ is the root or of type } L, \\ m - 1 & \text{if } w \text{ is of type } R. \end{cases} \tag{5.2.4}$$

Recall that a Gamma distribution Y with parameters r and λ has density $f_Y(x) = x^{r-1} \lambda^r e^{-\lambda x} / \Gamma(r)$ for $x \geq 0$ 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

$$\Gamma_w \sim \begin{cases} Y & \text{if } w \text{ is the root or of type } R, \\ Y^* & \text{if } w \text{ is of type } L, \end{cases} \tag{5.2.5}$$

independently of everything else.

Let $w_1, \dots, w_{m_-(w)}$ be the children of w type L , and let their locations $X_{w_1}, \dots, X_{w_{m_-(w)}}$ be sampled i.i.d. and uniformly at random from $[0, X_w]$ (where we recall that the locations correspond to the age raised to the power χ). Further, let $(X_{w(m_-(w)+j)})_{j \geq 1}$ be the (ordered) points of a Poisson point process with

intensity

$$\rho_w(x) = \Gamma_w \frac{\psi x^{\psi-1}}{x_w^\psi} \quad (5.2.6)$$

on $[X_w, 1]$, and the vertices $(w(m_-(w) + j))_{j \geq 1}$ have type R . The children of w are the vertices wj 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 $\psi = m/(m + \delta) > 0$, so the intensity ρ_w in (5.2.6) 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 *Pólya-point tree*, and the point process $(x_w)_w$ the *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 (x_w, t_w) , where $x_w \in [0, 1]$ corresponds to the (rescaled) 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 $\text{PA}_n^{(m, \delta)}(c)$:

Theorem 5.6 (Local weak convergence of preferential attachment models: fixed edges) *Fix $\delta \geq 0$ and $m \geq 1$. The preferential attachment model $\text{PA}_n^{(m, \delta)}(c)$ converges locally weakly in probability to the Pólya-point tree.*

Theorem 5.6 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 5.5 for more details, also on the history of Theorem 5.6.

Degree structure of $\text{PA}_t^{(m, \delta)}(c)$

Before turning to the proof of Theorem 5.6, we use it to describe some properties of the degrees of vertices in $\text{PA}_n^{(m, \delta)}(c)$. This is done in the following lemma:

Lemma 5.7 (Degree sequence of $\text{PA}_t^{(m, \delta)}(c)$) *Let $D_o(n)$ be the degree at time n of a vertex chosen uniformly at random from $[n]$ in $\text{PA}_n^{(m, \delta)}(c)$. Then,*

$$\mathbb{P}(D_o(n) = k) \rightarrow p_k = \frac{2m + \delta}{m} \frac{\Gamma(m + 2 + \delta + \delta/m)}{\Gamma(m + \delta)} \frac{\Gamma(k + \delta)}{\Gamma(k + 3 + \delta + \delta/m)}, \quad (5.2.7)$$

and, with $D_o(n)'$ be the degree at time n of one of the m older neighbors of o ,

$$\mathbb{P}(D_o(n)' = k) \rightarrow p'_k = \frac{2m + \delta}{m^2} \frac{\Gamma(m + 2 + \delta + \delta/m)}{\Gamma(m + \delta)} \frac{(k + 1)\Gamma(k + 1 + \delta)}{\Gamma(k + 4 + \delta + \delta/m)}. \quad (5.2.8)$$

Further, let $P_k(n)$ denote the proportion of vertices of degree k in $\text{PA}_n^{(m, \delta)}(c)$, and $P'_k(n)$ the proportion of older neighbors of vertices of degree k . Then $P_k(n) \xrightarrow{\mathbb{P}} p_k$ and $P'_k(n) \xrightarrow{\mathbb{P}} p'_k$.

5.2 Local weak convergence of preferential attachment models 171

Note that the limiting degree distribution in (5.2.7) is equal to that for $\text{PA}_n^{(m,\delta)}$ in (1.3.58), again exemplifying that the details of the model have little influence on the limiting degree sequence. It is not hard to see from Lemma 5.7 that

$$p_k = c_{m,\delta} k^{-\tau} (1 + O(1/k)), \quad p'_k = c'_{m,\delta} k^{-(\tau-1)} (1 + O(1/k)), \quad (5.2.9)$$

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 5.7. 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 position $X_\varnothing = U_\varnothing^\chi$ of the root of the Pólya graph, where U_\varnothing is standard uniform. Let D be the degree of the root. Conditionally on $X_\varnothing = x$, the degree D is m plus a Poisson variable with parameter

$$\frac{\Gamma_\varnothing}{x_\varnothing} \int_{x_\varnothing}^1 \psi x^{\psi-1} dx = \Gamma_\varnothing \frac{1 - x^\psi}{x^\psi} \equiv \Gamma_\varnothing \kappa(x), \quad (5.2.10)$$

where Γ_\varnothing is a Gamma variable with parameters $r = m + \delta$ and $\lambda = 1$. Thus, taking expectations with respect to Γ_\varnothing , we obtain

$$\begin{aligned} \mathbb{P}(D = k \mid X_\varnothing = x) &= \int_0^\infty \mathbb{P}(D = k \mid X_\varnothing = x, \Gamma_\varnothing = y) \frac{y^{m+\delta}}{\Gamma(m+\delta+1)} e^{-y} dy \\ &= \int_0^\infty e^{-y\kappa(x)} \frac{(y\kappa(x))^{k-m}}{(k-m)!} \frac{y^{m+\delta}}{\Gamma(m+\delta+1)} e^{-y} dy \\ &= \frac{\kappa(x)^{k-m}}{(1+\kappa(x))^{k-m+m+\delta}} \frac{\Gamma(k+1+\delta)}{(k-m)! \Gamma(m+\delta+1)} \\ &= (1-x^\psi)^{k-m} x^{\psi(m+\delta)} dx \frac{\Gamma(k+1+\delta)}{(k-m)! \Gamma(m+\delta+1)}, \end{aligned} \quad (5.2.11)$$

where we use that $\kappa(x)/(1+\kappa(x)) = 1 - x^\psi$. We thus conclude that, using that $X_\varnothing = U_\varnothing^\chi$ where U_\varnothing is standard uniform, has density $x^{1/\chi-1}/\chi = (\psi+1)x^\psi$ and that $\psi \cdot (m+\delta) = m$

$$\begin{aligned} \mathbb{P}(D = k) &= (\psi+1) \int_0^1 \mathbb{P}(D = k \mid X_\varnothing = x) x^\psi dx \\ &= (\psi+1) \int_0^1 (1-x^\psi)^{k-m} x^{m+\psi} \frac{\Gamma(k+1+\delta)}{(k-m)! \Gamma(m+\delta+1)} dx. \end{aligned} \quad (5.2.12)$$

We now use the integral transform $u = x^\psi$, for which $dx = (1/\psi)u^{(\psi-1)/\psi} du =$

$u^{\delta/m} du / \psi$, to arrive at

$$\begin{aligned} \mathbb{P}(D = k) &= \frac{\psi + 1}{\psi} \frac{\Gamma(k + 1 + \delta)}{(k - m)! \Gamma(m + \delta + 1)} \int_0^1 (1 - u)^{k-m} u^{m+\delta+1+\delta/m} du \\ &= \frac{\psi + 1}{\psi} \frac{\Gamma(k + 1 + \delta)}{(k - m)! \Gamma(m + \delta + 1)} \frac{\Gamma(k - m + 1) \Gamma(m + \delta + 2 + \delta/m)}{\Gamma(k + 3 + \delta + \delta/m)} \\ &= \frac{\psi + 1}{\psi} \frac{\Gamma(k + 1 + \delta) \Gamma(m + \delta + 2 + \delta/m)}{\Gamma(m + \delta + 1) \Gamma(k + 3 + \delta + \delta/m)}, \end{aligned} \quad (5.2.13)$$

as required.

We next extend this to convergence in distribution of $D'_o(n)$, 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 $D'_o(n)$. 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 position $X_\varnothing = U_\varnothing^\chi$ of the root of the Pólya graph, where U_\varnothing is standard uniform, and recall that the position X_{\varnothing_1} of one of the m older vertices to which \varnothing is connected is uniform on $[0, X_\varnothing]$. Let D be the degree of the root. Conditionally on $X_{\varnothing_1} = y$, the degree D is m plus a Poisson variable with parameter

$$\frac{\Gamma_{\varnothing_1}}{y} \int_y^1 \psi x^{\psi-1} dx = \Gamma_{\varnothing_1} \frac{1 - y^\psi}{y^\psi} \equiv \Gamma_{\varnothing_1} \kappa(x), \quad (5.2.14)$$

where Γ_{\varnothing_1} is a Gamma variable with parameters $r = m + 1 + \delta$ and $\lambda = 1$. Thus, taking expectations with respect to Γ_{\varnothing_1} and X_\varnothing , we obtain as before

$$\mathbb{P}(D = k \mid X_{\varnothing_1} = y) = \frac{\Gamma(k + 1 + \delta)}{(k - m)! \Gamma(m + \delta + 1)} (1 - y^\psi)^{k-m} y^{m(m+1+\delta)/(m+\delta)}, \quad (5.2.15)$$

where we use that $\kappa(x)/(1 + \kappa(x)) = 1 - x^\psi$. We thus conclude that

$$\begin{aligned} \mathbb{P}(D' = k) &= (\psi + 1) \int_0^1 x^\psi \frac{1}{x} \int_0^x \mathbb{P}(D = k \mid X_{\varnothing_1} = y) dy dx \\ &= (\psi + 1) \frac{\Gamma(k + 1 + \delta)}{(k - m)! \Gamma(m + \delta + 1)} \\ &\quad \times \int_0^1 x^{\psi-1} \int_0^x (1 - y^\psi)^{k-m} y^{m(m+1+\delta)/(m+\delta)} dy dx \\ &= \frac{\psi + 1}{\psi^2} \frac{\Gamma(k + 1 + \delta)}{(k - m)! \Gamma(m + \delta + 1)} \int_0^1 \int_0^u (1 - v)^{k-m} v^{m+1+\delta+\delta/m} dv du, \end{aligned} \quad (5.2.16)$$

where we now used the integral transform $u = x^\psi$ and $v = y^\psi$. Interchanging the

integrals over u and v leads to

$$\begin{aligned} \mathbb{P}(D' = k) &= \frac{\psi + 1}{\psi^2} \frac{\Gamma(k + 1 + \delta)}{(k - m)! \Gamma(m + \delta + 1)} \int_0^1 (1 - v)^{k-m+1} v^{m+1+\delta+\delta/m} dv \\ &= \frac{2m + \delta}{m^2} \frac{\Gamma(m + 2 + \delta + \delta/m)}{\Gamma(m + \delta)} \frac{(k + 1)\Gamma(k + 1 + \delta)}{\Gamma(k + 4 + \delta + \delta/m)}, \end{aligned} \tag{5.2.17}$$

as required.

The statements that $P_k(n) \xrightarrow{\mathbb{P}} p_k$ and $P'_k(n) \xrightarrow{\mathbb{P}} p'_k$ follow from the fact that Theorem 5.6 states local weak convergence in probability. \square

The proof of Theorem 5.6 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 $\text{PA}_n^{(m,\delta)}(c)$ that relies on independent random variables. We explain this now.

Finite-graph Pólya version of $\text{PA}_t^{(m,\delta)}(c)$

We now explain the finite-graph Pólya version of $\text{PA}_t^{(m,\delta)}(c)$. 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.,

$$\psi_j \sim \beta(m + \delta, (2j - 3)m + \delta(j - 1)). \tag{5.2.18}$$

Here we recall that Y has a Beta distribution with parameters (a, b) when $f_Y(y) = y^{a-1}(1 - y)^{b-1}B(a, b)$ for $y \in [0, 1]$, where $B(a, b) = \Gamma(a + b)/\Gamma(a)\Gamma(b)$ is the Beta-function. Define

$$\varphi_j^{(n)} = \psi_j \prod_{i=j+1}^n (1 - \psi_i), \quad S_k^{(n)} = \sum_{j=1}^k \varphi_j^{(n)} = \prod_{i=k+1}^n (1 - \psi_i). \tag{5.2.19}$$

Here the latter equality follows simply by induction on $k \geq 1$ (see Exercise 5.16). Finally, let $I_k = [S_{k-1}^{(n)}, S_k^{(n)})$. We now construct a graph as follows:

- ▷ Conditioned on ψ_1, \dots, ψ_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 $[0, S_{k-1}^{(n)}]$;
- ▷ Join two vertices j and k if $j < k$ and $U_{k,i} \in I_j$ for some $i \in [m]$ (with multiple edges between j and k if there are several such i);
- ▷ Call the resulting random multi-graph the *finite-size Pólya graph of size n* .

The main result for $\text{PA}_n^{(m,\delta)}(c)$ is as follows:

Theorem 5.8 (Finite-graph Pólya version of $\text{PA}_n^{(m,\delta)}(c)$) *Fix $\delta \geq 0$ and $m \geq 1$. Then, the distribution of $\text{PA}_n^{(m,\delta)}(c)$ is the same as that of the finite-size Pólya graph of size n .*

The nice thing about Theorems 5.6–5.8 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 5.8 to derive properties of the number of multiple edges in $\text{PA}_n^{(m,\delta)}(c)$ for $m = 2$.

In terms of the above Pólya-point tree, the proof will show that the Gamma variables Y that define the “strength” Γ_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 5.10 and 5.11 below).

Before moving to the proof of Theorem 5.8, 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:

$$F_k = \lim_{n \rightarrow \infty} \left(\frac{n}{k}\right)^\chi \prod_{j=k+1}^{\ell} (1 - \psi_j). \quad (5.2.20)$$

In Exercise 5.20 you are asked to prove that the almost sure limit in (5.2.20) exists. Then, the limiting distribution of fixed vertices is as follows:

Lemma 5.9 (Degree structure of $\text{PA}_t^{(m,\delta)}(c)$: fixed vertices) *Consider $\text{PA}_t^{(m,\delta)}(c)$ with $m \geq 1$ and $\delta > -m$. Then,*

$$n^{-(1-\chi)} D_k(n) \xrightarrow{d} (2m + \delta) k^\chi \psi_k F_k. \quad (5.2.21)$$

We prove Lemma 5.9 below.

Let us give some insight into the proof of Theorem 5.8, after which we give the full proof. There is a close connection between the preferential attachment model and the Pólya urn model in the following sense: every new connection 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 β* , $D_1(\tau_k)$ 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 $(D_1(\tau'_k), D_2(\tau'_k), \dots, D_n(\tau'_k))_{k \geq 1}$ that is valid for all $k, n \geq 1$ fixed. Here $\tau'_k = \int \{t: D_1(t) + \dots, D_n(t) = k\}$ is the first time where the total degree of the vertices in $[n]$ is equal to k . Note that $\tau'_{2mn} = n$. Further, the random variables $(D_1(\tau'_k), D_2(\tau'_k), \dots, D_n(\tau'_k))_{k=1}^{2mn}$ determine the law of $(\text{PA}_t^{(m,\delta)}(b))_{t \in [n]}$ 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 ψ_j in (5.2.18) follows from tracing back the graph construction obtained in this way. We next use this to complete the proof of Theorem 5.8:

Let us now make this intuition more precise. 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 $1, \dots, k-1$. 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

$$\frac{d_k + \delta}{2m(n-1) + (1 + \delta)(i-1)}, \tag{5.2.22}$$

while the probability that it is connected to a vertex in $[k-1]$ is equal to

$$\frac{d_{<k} + \delta(k-1)}{2m(n-1) + (1 + \delta)(i-1)}. \tag{5.2.23}$$

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 n to $[n-1]$ is attached to $[k-1]$ is $(d_{<k} + \delta(k-1))/Z$, where $Z = k\delta + d_{<k+1}$ 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 ψ_k and $1 - \psi_k$, where ψ_k has the $\beta(m + \delta, (2k-3)m + \delta(k-1))$ distribution. We next use this to complete the proof of Theorem 5.8:

Completion of the proof of Theorem 5.8. Using the two-urn process as an inductive input, we can now construct the Pólya graph defined in Theorem 5.8 in a similar way as it was done for Pólya urns with multiple colors in (5.1.27).

Let $X_t \in \{1, 2, \dots, \lceil t/m \rceil\}$ be the vertex receiving the t th edge in the sequential model (the other endpoint of this edge being the vertex $\lceil t/m \rceil + 1$). 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 ψ_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 ψ_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 $(1 - \psi_2)(1 - \psi_3)$, $\psi_2(1 - \psi_3)$ and ψ_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 $t = m + 1, \dots, 2m$, the distribution of $X_t \in \{1, 2\}$ is described by a two-urn model with strengths $1 - \psi_2$ and ψ_2 , where $\psi_2 \sim B_2$.
- ▷ In general, for $t = m(k-1) + 1, \dots, km$, the distribution of $X_t \in [k]$ is described

by a k -urn model with strengths

$$\varphi_j^{(k)} = \psi_j \prod_{i=j+1}^k (1 - \psi_i), \quad j = 1, \dots, k. \quad (5.2.24)$$

Here ψ_k is chosen at the beginning of the k th stage, independently of the previously chosen strengths $\psi_1, \dots, \psi_{k-1}$ (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 5.8 as follows. By (5.2.19), $S_k = \prod_{j=k+1}^n (1 - \psi_j)$. 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 5.8, and shows that the process derived above is indeed the process given in the theorem. \square

Asymptotics of $(\psi_k)_{k \in [n]}$ and $(S_k^{(n)})_{k \in [n]}$

We next continue to analyse the random variables in Theorem 5.8 to prepare us for proving local weak convergence in Theorem 5.6. We start by analyzing ψ_k for k large:

Lemma 5.10 (Gamma asymptotics of Beta variables) *As $k \rightarrow \infty$, $k\psi_k \xrightarrow{d} Y$, where Y has a Gamma distribution with $r = m + \delta$ and $\lambda = 2m + \delta$.*

More precisely, take $f_k(x)$ such that $\mathbb{P}(\psi_k \leq f_k(x)) = \mathbb{P}(\chi_k \leq x)$, where χ_k has a Gamma distribution with $r = m + \delta$ and $\lambda = 1$ (so that $Y \stackrel{d}{=} X/(2m + \delta)$). For every $\varepsilon > 0$, there exists $K = K_\varepsilon \geq 1$ sufficiently large. Then, for all $k \geq K$ and $x \leq (\log k)^2$,

$$\frac{1 - \varepsilon}{k(2m + \delta)} x \leq f_k(x) \leq \frac{1 + \varepsilon}{k(2m + \delta)} x. \quad (5.2.25)$$

Further, $\chi_k \leq (\log k)^2$ for all $k \geq K$ and with probability at least $1 - \varepsilon$.

Proof Fix $x \geq 0$. We compute that

$$\begin{aligned} \mathbb{P}(k\psi_k \leq x) &= \frac{\Gamma(m + \delta + (2k - 3)m + \delta(k - 1))}{\Gamma(m + \delta)\Gamma((2k - 3)m + \delta(k - 1))} \\ &\quad \times \int_0^{x/k} u^{m+\delta-1} (1 - u)^{(2k-3)m+\delta(k-1)-1} du \\ &= (1 + o(1)) \frac{k^{m+\delta}}{\Gamma(m + \delta)} k^{-(m+\delta)} \int_0^x u^{m+\delta-1} (1 - u/k)^{(2k-3)m+\delta(k-1)-1} du. \end{aligned} \quad (5.2.26)$$

For every $u > 0$, $(1 - u/k)^{(2k-3)m+\delta(k-1)-1} \rightarrow e^{-u(2m+\delta)}$, so that dominated convergence implies that

$$\mathbb{P}(k\psi_k \leq x) \rightarrow \int_0^x \frac{u^{m+\delta-1}}{\Gamma(m + \delta)} e^{-u(2m+\delta)} du, \quad (5.2.27)$$

as required. We refrain from proving the other statements, that are more technical versions of the above argument. \square

By Lemma 5.10, we see that indeed the Beta random variables $(\psi_k)_{k \in [n]}$ give rise to the Gamma random variables in (5.2.5). This explains the relevance of the Gamma random variables in the Pólya-point graph.

We continue by analyzing the asymptotics for the random variables $(S_k^{(n)})_{k \in [n]}$:

Lemma 5.11 (Asymptotics of $S_k^{(n)}$) *Recall that $\chi = (m + \delta)/(2m + \delta)$. For every $\varepsilon > 0$, there exists $\eta > 0$ and $K < \infty$ such that for all $n \geq K$ and with probability at least $1 - \varepsilon$,*

$$\max_{k \in [n]} \left| S_k^{(n)} - \left(\frac{k}{n}\right)^\chi \right| \leq \eta, \quad (5.2.28)$$

and

$$\max_{k \in [n] \setminus [K]} \left| S_k^{(n)} - \left(\frac{k}{n}\right)^\chi \right| \leq \varepsilon \left(\frac{k}{n}\right)^\chi. \quad (5.2.29)$$

Proof We give the intuition behind Lemma 5.11. We recall from (5.2.19) that $S_k^{(n)} = \prod_{i=k+1}^n (1 - \psi_i)$, where $(\psi_k)_{k \in [n]}$ are independent random variables. We start by investigating $S'_n = \prod_{i=1}^n (1 - \psi_i)$, so that $S_k^{(n)} = S'_k/S'_n$. We write

$$\log S'_n = \sum_{i=1}^n \log(1 - \psi_i). \quad (5.2.30)$$

Thus, using Lemma 5.10 and with $(\chi_i)_{i \geq 1}$ an i.i.d. sequence of Gamma random variables with parameters $r = m + \delta$ and using that $k\psi_k \xrightarrow{d} \chi_k/(2m + 1)$,

$$\begin{aligned} \log S'_n &\approx - \sum_{i=1}^n \psi_i \approx - \frac{1}{2m + \delta} \sum_{i=1}^n \chi_i/i \\ &\approx - \frac{1}{2m + \delta} \sum_{i=1}^n \frac{\mathbb{E}[\chi_i]}{i} = \frac{m + \delta}{2m + \delta} \log n = \chi \log n. \end{aligned} \quad (5.2.31)$$

Thus, $S_k = S'_k/S'_n \approx \left(\frac{k}{n}\right)^\chi$, as required. The proof can be completed using martingale techniques and is omitted here. \square

We use the above analysis to complete the proof of Lemma 5.9:

Proof of Lemma 5.9. We use a conditional second moment method on $D_k(n)$, conditionally on $(\psi_k)_{k \in [n]}$. Denote

$$F_k^{(n)} = \left(\frac{n}{k}\right)^\chi \prod_{j=k+1}^{\ell} (1 - \psi_j). \quad (5.2.32)$$

Then,

$$\mathbb{E}[D_k(n) \mid (\psi_k)_{k \in [n]}] = m \sum_{\ell=k+1}^n \varphi_k^{(\ell)}, \quad (5.2.33)$$

and

$$\text{Var}\left(D_k(n) \mid (\psi_k)_{k \in [n]}\right) \leq m \sum_{\ell=k+1}^n \varphi_k^{(\ell)}. \quad (5.2.34)$$

It thus suffices to show that $n^{1-\chi} m \sum_{\ell=k+1}^n \varphi_k^{(\ell)} \xrightarrow{a.s.} \frac{m}{1-\chi} k^\chi \psi_k F_k$. For this, we note that since $\varphi_k^{(\ell)} = \psi_k \prod_{j=k+1}^{\ell} (1 - \psi_j)$, we arrive at

$$m \sum_{\ell=k+1}^n \varphi_k^{(\ell)} = m \psi_k \sum_{\ell=k+1}^n \left(\frac{k}{\ell}\right)^\chi F_k^{(\ell)} = m k^\chi \psi_k \sum_{\ell=k+1}^n \ell^{-\chi} F_k^{(\ell)}. \tag{5.2.35}$$

Since $\chi \in (0, 1)$ and $F_k^{(\ell)} \xrightarrow{a.s.} F_k$ as $\ell \rightarrow \infty$, we obtain that

$$n^{1-\chi} \sum_{\ell=k+1}^n \ell^{-\chi} F_k^{(\ell)} \xrightarrow{a.s.} F_k / (1 - \chi). \tag{5.2.36}$$

This completes the proof. □

Regularity properties of the Pólya-point tree and overview

We next discuss some properties of the Pólya-point tree, showing that the tree (\mathcal{T}, \emptyset) is well defined and, within the r -neighborhood of the root, all random variables used in its description are uniformly bounded:

Lemma 5.12 (Regularity Pólya-point tree) *Fix $r \geq 1$ and $\varepsilon > 0$. Then there exist constants $\eta > 0$ and $K < \infty$ such that, with probability at least $1 - \varepsilon$, $B_r(\emptyset) \leq K$, $X_w \geq \eta$ for all $w \in B_r(\emptyset)$, and further $\Gamma_w \leq K, \rho_w(\cdot) \leq K$ for all $w \in B_r(\emptyset)$, while finally, $\min_{w, w' \in B_r(\emptyset)} |X_w - X_{w'}| \geq \eta$.*

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 δ . □

Let us describe how the proof of Theorem 5.6 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 (H, y) , using a second moment method. Let

$$N_{n,r}(H, y) = \sum_{v \in [n]} \mathbb{1}_{\{B_r^{(n)}(v) \simeq (H, y)\}}, \tag{5.2.37}$$

where $B_r^{(n)}(v)$ is the r -neighborhood of v in $\text{PA}_n^{(m, \delta)}(c)$. With $B_r(\emptyset)$ the r -neighborhood of \emptyset in the Pólya-point graph (\mathcal{T}, \emptyset) , we will show that

$$\frac{N_{n,r}(H, y)}{n} \xrightarrow{\mathbb{P}} \mathbb{P}(B_r(\emptyset) \simeq (H, y)). \tag{5.2.38}$$

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

$$\mathbb{E}[N_{n,r}(H, y)/n] = \mathbb{P}(B_r^{(n)}(o_n) \simeq (H, y)). \quad (5.2.39)$$

We do this using a *coupling approach*, by showing that we can couple $B_r^{(n)}(o_n)$, together with the types of the vertices involved, with $B_r(\emptyset)$ 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 (H, y) . Let us now give the details of the convergence of the expected number of one-neighborhoods of a uniform point.

Exploration of $\text{PA}_n^{(m,\delta)}(c)$: one-neighborhood

In this step, we discuss how we can couple the one-neighborhood of a uniformly chosen vertex to that in (\mathcal{T}, \emptyset) , so as to prove Theorem 5.6 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 $\text{PA}_n^{(m,\delta)}(c)$ and $B_1(\emptyset)$ in (\mathcal{T}, \emptyset) in such a way that all the variables involved are close:

Lemma 5.13 (Coupling one-neighborhood of uniform vertex in $\text{PA}_n^{(m,\delta)}(c)$ and Pólya-point tree) *Fix $\varepsilon > 0$. Let o_n be chosen uniformly at random from $[n]$. Let*

$$v_{o_n 1}, \dots, v_{o_n(m+q_{o_n})}$$

be the neighbors of o_n , ordered such that the edge $\{o_n, v_{o_n i}\}$ 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 $o_n = v_\emptyset$, together with the variables $(S_{v_{\emptyset i}}^{(n)})_{i \geq 1}$, and $B_1(\emptyset)$, together with its positions $(X_{\emptyset i})_{i \geq 1}$, can be coupled such that

- (i) $B_1^{(n)}(o_n) \simeq B_1(\emptyset)$ and $|B_1(\emptyset)| \leq K$;
- (ii) $|X_{\emptyset k} - S_{v_{\emptyset k}}^{(n)}| \leq \eta$ for all k for which $ok \in B_1(\emptyset)$;
- (iii) $v_{\emptyset 1}, \dots, v_{\emptyset(m+q_\emptyset)}$ are all distinct and $v_{\emptyset k} \geq \eta n$ for all $v_{\emptyset k} \in B_1^{(n)}(o_n)$;
- (iv) $\Gamma_{\emptyset k} \leq K$ for all k such that $\emptyset k \in B_1(\emptyset)$.

Consequently, $\mathbb{P}(B_1^{(n)}(o_n) \simeq (H, y)) \rightarrow \mathbb{P}(B_1(\emptyset) \simeq (H, y))$.

We start with Lemma 5.13, 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 $X_{\emptyset 1}, \dots, X_{\emptyset(m+q_\emptyset)}$ 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

$$S_{v_{\emptyset i-1}}^{(n)} \leq \frac{X_{\emptyset i}}{X_\emptyset} S_{v_\emptyset}^{(n)} \leq S_{v_{\emptyset i}}^{(n)}. \quad (5.2.40)$$

By Theorem 5.8 and the observation that $U_{v_{\emptyset i}} = X_{\emptyset i}/X_\emptyset$ is a collection of i.i.d.

uniform random variables on $[0, 1]$, it follows that whp $v_{\emptyset i} \in n[X_{\emptyset i} - \delta, X_{\emptyset i} + \delta]$, as required. In more detail, given $\varepsilon > 0$, choose $\delta > 0$ and $K > 0$ such that the regularity properties in Lemma 5.12 hold for $r = 1$. By Lemma 5.11, for n large enough,

$$|\bar{S}_{v_{\emptyset}}^{(n)} - X_{\emptyset}| \leq \eta, \quad \text{and} \quad |\bar{S}_{v_{\emptyset i}}^{(n)} - X_{\emptyset i}| \leq \eta \quad \forall i \in [m], \quad (5.2.41)$$

with probability at least $1 - 2\varepsilon$.

We continue by investigating the limiting distribution of the remaining neighbors $v_{\emptyset 1}, \dots, v_{\emptyset(m+q_{\emptyset}^{(n)})}$. Along the way, we will prove that $q_{\emptyset}^{(n)} = q_{\emptyset}$ whp. Note that by Theorem 5.8 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

$$P_{v \rightarrow v_{\emptyset}} = \frac{\varphi_{v_{\emptyset}}^{(v-1)}}{S_{v-1}^{(v-1)}} = \frac{S_{v_{\emptyset}}^{(n)}}{S_{v-1}^{(n)}} \psi_{v_{\emptyset}}, \quad (5.2.42)$$

where the latter equality follows by (5.2.19).

Denote

$$N_{v_{\emptyset}}(y) = \sum_{v=v_{\emptyset}+1}^{ny} \sum_{i=1}^m \mathbb{1}_{\{X_{v,i}=v_{\emptyset}\}}. \quad (5.2.43)$$

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 5.12, $v_{\emptyset} \geq n\eta$ with probability at least $1 - \varepsilon$. Recall that $\psi_k = f_k(\chi_k)$, where χ_k has a Gamma distribution with parameters $\lambda = 1$ and $r = m + \delta$. Thus, we may apply Lemmas 5.10 and 5.11, which yield

$$(1 - \eta)\hat{P}_{v \rightarrow v_{\emptyset}} \leq P_{v \rightarrow v_{\emptyset}} \leq (1 + \eta)\hat{P}_{v \rightarrow v_{\emptyset}}, \quad (5.2.44)$$

where

$$\hat{P}_{v \rightarrow v_{\emptyset}} = \frac{\chi_{v_{\emptyset}}}{(2m + \delta)n} \frac{n}{v_{\emptyset}} \left(\frac{v_{\emptyset}}{v}\right)^x. \quad (5.2.45)$$

Letting $\hat{N}_{v_{\emptyset}}(y) = \sum_{v=v_{\emptyset}+1}^{ny} \sum_{i=1}^m \mathbb{1}_{\{\hat{X}_{v,i}=v_{\emptyset}\}}$, where $(\hat{X}_{v,i})_{v,i}$ are conditionally independent given $(\psi_k)_{k \geq 1}$ (or, equivalently, given $(\chi_k)_{k \geq 1}$, we can couple $\hat{N}_{v_{\emptyset}}(y)$ to a Poisson random variable with parameter

$$\frac{\chi_{v_{\emptyset}}}{2(1 + \delta/m)U_{\emptyset}} \int_0^y \left(\frac{U_{\emptyset}}{s}\right)^x ds \quad (5.2.46)$$

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 $\hat{N}_{v_{\emptyset}}(x^x)$ converges to a Poisson random variable with parameter

$$\frac{\chi_{v_{\emptyset}}}{U_{\emptyset}^{\psi}} \int_0^x \psi s^{\psi-1} ds = \frac{\Gamma_{\emptyset}}{U_{\emptyset}^{\psi}} \int_0^x \psi s^{\psi-1} ds, \quad (5.2.47)$$

on $[U_\emptyset, 1]$. It is not hard to extend this result to the statement that $(\hat{N}_{v_\emptyset}(x^\chi))_{x \in [U_\emptyset, 1]}$ converges in distribution to an inhomogeneous Poisson process with intensity $\Gamma_\emptyset \psi x^{\psi-1} / U_\emptyset^\psi = \rho_\emptyset(x)$, as required. In particular, this shows that $q_\emptyset^{(n)} = q_\emptyset$ with probability at least $1 - \varepsilon$.

We conclude the proof of (i)–(ii) by investigating the differences between $X_{\emptyset i}$ and $S_{v_{\emptyset i}}^{(n)}$ using Lemma 5.11 to bound the difference between $S_{v_{\emptyset i}}^{(n)}$ and $(v_{\emptyset i}/n)^x$, which yields that for n large enough and with probability at least $1 - 3\varepsilon$, $q_\emptyset^{(n)} = q_\emptyset$ and the bounds $q_0 + m \leq N$, $\chi_{v_\emptyset} = \Gamma_{v_\emptyset} \leq K$ and

$$|X_{\emptyset k} - S_{v_{\emptyset k}}^{(n)}| \leq \eta \quad \text{for all } i \in [m + q_0] \setminus [q] \quad (5.2.48)$$

hold, as required. This completes the proof of parts (i)–(ii).

For (iii), we note that by (5.2.41) and (5.2.48), and another application of Lemma 5.11, there exists an $\eta > 0$ such that $v_{\emptyset k} \geq \eta n$ with probability at least $1 - 4\varepsilon$ and $|v_{\emptyset k} - v_{\emptyset l}| \geq \eta n$ for all $k, l \leq m + q_\emptyset$. As a result, $(v_{\emptyset k})_{k \in [m+q_\emptyset]}$ are all distinct. This completes the proof of (iii).

For (iv), we note that the claimed bounds are obviously true when the sequence $(\chi_{v_{\emptyset k}})_{k \in [m+q_0]}$ would be *independent*. However, this is not true, since the fact that $v_{\emptyset k}$ is a neighbor of v_\emptyset ‘tilts’ the distribution of the above random variables. We will provide a coupling between $(\chi_{v_{\emptyset k}})_{k \in [m+q_0]}$ and a collection of independent random variables. Let us now provide the details.

We assume that $\Gamma_\emptyset = \chi_{v_\emptyset} \leq K$, that $q_\emptyset^{(n)} = q_\emptyset$ and that $(v_{\emptyset k})_{k \in [m+q_0]}$ are all distinct with $\min_{k \in [m+q_0]} v_{\emptyset k} \geq \eta n$. Let \mathcal{E} denote the event that v_\emptyset is the uniform random vertex and that the neighbors of v_\emptyset are $(v_{\emptyset k})_{k \in [m+q_0]}$. Let $\chi^{\mathcal{E}, \gamma_\emptyset}$ denote the conditional distribution of $(\chi_k)_{k \in [n] \setminus \{v_\emptyset\}}$ conditioned on $\Gamma_\emptyset = \gamma_\emptyset$ and \mathcal{E} . Further, let $(\hat{\chi}_k)_{k \in [n]}$ denote an independent collection of random variables with distribution

$$\hat{\chi}_{v_{\emptyset k}} \sim Y_{v_{\emptyset k}}^* \quad \forall k \in [m], \quad \hat{\chi}_k \sim Y_k \quad \forall k \in [n] \setminus [m]. \quad (5.2.49)$$

Let $\rho(\cdot \mid \mathcal{E}, \chi_{v_\emptyset} = \gamma_\emptyset)$ denote the density of the (multi-dimensional) random variable $(\hat{\chi}_k)_{k \in [n]}$, and let \mathbb{P} be the joint distribution of $\text{PA}_n^{(m, \delta)}(c)$ and the random variables $(\chi_k)_{k \in [n]}$. By Bayes’ Theorem,

$$\rho(\cdot \mid \mathcal{E}, \chi_{v_\emptyset} = \gamma_\emptyset) = \frac{\mathbb{P}(\mathcal{E} \mid \cdot, \chi_{v_\emptyset} = \gamma_\emptyset)}{\mathbb{P}(\mathcal{E} \mid \chi_{v_\emptyset} = \gamma_\emptyset)} \rho_\emptyset(\cdot), \quad (5.2.50)$$

where $\rho_\emptyset(\cdot)$ is the original density of the random variables $(\chi_k)_{k \in [n] \setminus \{v_\emptyset\}}$. We denote the corresponding probability distributions and expectations as \mathbb{P}_\emptyset and \mathbb{E}_\emptyset , respectively. We conclude that we have to compare $\mathbb{P}(\mathcal{E} \mid \cdot, \chi_{v_\emptyset})$ to $\mathbb{P}(\mathcal{E} \mid \cdot, \chi_{v_\emptyset})$. Since we are interested in the distribution of $(\chi_k)_{k \in [n]}$, this amounts to comparing $\mathbb{P}(\mathcal{E} \mid (\chi_k)_{k \in [n]})$ to $\mathbb{P}(\mathcal{E} \mid \chi_{v_\emptyset})$. For this, we need to compute $\mathbb{P}(\mathcal{E} \mid (\chi_k)_{k \in [n]})$ (and note that $\mathbb{P}(\mathcal{E} \mid \chi_{v_\emptyset}) = \mathbb{E}_\emptyset[\mathbb{P}(\mathcal{E} \mid (\chi_k)_{k \in [n]})]$). This is our next step.

By Theorem 5.8, we can explicitly compute

$$\begin{aligned} \mathbb{P}(\mathcal{E} \mid (\chi_k)_{k \in [n]}) &= m! \prod_{i=1}^m P_{v_\emptyset \rightarrow v_{\emptyset i}} \prod_{j=1}^{q_\emptyset} m P_{v_{\emptyset j} \rightarrow v_\emptyset} (1 - P_{v_{\emptyset j} \rightarrow v_\emptyset})^{m-1} \\ &\times \prod_{v > v_\emptyset : v \notin \{v_{\emptyset(m+1)}, \dots, v_{\emptyset(m+q)}\}} (1 - P_{v \rightarrow v_\emptyset})^m, \end{aligned} \quad (5.2.51)$$

by the conditional independence given $(\psi_k)_{k \in [n]}$, which is equivalent to conditional independence given $(\chi_k)_{k \in [n]}$. We rewrite

$$\mathbb{P}(\mathcal{E} \mid (\chi_k)_{k \in [n]}) = m! \prod_{i=1}^m P_{v_\emptyset \rightarrow v_{\emptyset i}} \prod_{j=1}^{q_\emptyset} \frac{m P_{v_{\emptyset j} \rightarrow v_\emptyset}}{1 - P_{v_{\emptyset j} \rightarrow v_\emptyset}} \prod_{v > v_\emptyset} (1 - P_{v \rightarrow v_\emptyset})^m, \quad (5.2.52)$$

where $P_{v \rightarrow v'}$ is given, and using (5.2.19), by

$$P_{v \rightarrow v'} = \frac{\varphi_{v'}^{(v-1)}}{S_{v-1}^{(v-1)}} = \frac{S_{v'}^{(n)}}{S_{v-1}^{(n)}} \psi_{v'}. \quad (5.2.53)$$

We now perform asymptotics of this formula. We use Lemma 5.11 to show that we can for every $\eta > 0$, there exists n_0 sufficiently large, such that, with probability at least $1 - \varepsilon$,

$$\begin{aligned} (1 - \eta) \mathbb{P}(\mathcal{E} \mid (\chi_k)_{k \in [n]}) & \\ &\leq m! \prod_{i=1}^m \psi_{v_\emptyset} \left(\frac{v_\emptyset}{v_{\emptyset i}}\right)^x \prod_{j=1}^{q_\emptyset} m \psi_{v_\emptyset} \left(\frac{v_{\emptyset j}}{v_\emptyset}\right)^x e^{-m \psi_{v_\emptyset} \sum_{v > v_\emptyset} \left(\frac{v_\emptyset}{v}\right)^x} \\ &\leq (1 + \eta) \mathbb{P}(\mathcal{E} \mid (\chi_k)_{k \in [n]}). \end{aligned} \quad (5.2.54)$$

To bound $\mathbb{P}(\mathcal{E} \mid \chi_{v_\emptyset}) = \mathbb{E}_\emptyset[\mathbb{P}(\mathcal{E} \mid (\chi_k)_{k \in [n]})]$, we also need a good bound on “bad” event, for which we use the deterministic upper bound

$$\begin{aligned} \mathbb{P}(\mathcal{E} \mid (\chi_k)_{k \in [n]}) &\leq m! \prod_{i=1}^m P_{v_\emptyset \rightarrow v_{\emptyset i}} \prod_{j=1}^{q_\emptyset} m P_{v_{\emptyset j} \rightarrow v_\emptyset} \leq m! (m \psi_{v_\emptyset})^q \prod_{i=1}^m \psi_{v_{\emptyset i}} \\ &\leq C' m! \prod_{i=1}^m \psi_{v_\emptyset} \left(\frac{v_\emptyset}{v_{\emptyset i}}\right)^x \prod_{j=1}^{q_\emptyset} m \psi_{v_\emptyset} \left(\frac{v_{\emptyset j}}{v_\emptyset}\right)^x e^{-m \psi_{v_\emptyset} \sum_{v > v_\emptyset} \left(\frac{v_\emptyset}{v}\right)^x}, \end{aligned} \quad (5.2.55)$$

where $C' = \eta^{-(m+N)} \sup_{n \geq 1} e^{m n f_{\delta_n}(K)} < \infty$.

We conclude that with probability at least $1 - \varepsilon/2$ with respect to \mathbb{P}_\emptyset ,

$$\sqrt{1 - \eta} \prod_{i=1}^m \frac{\psi_{v_{\emptyset i}}}{\mathbb{E}_\emptyset[\psi_{v_{\emptyset i}}]} \leq \frac{\mathbb{P}(\mathcal{E} \mid (\chi_k)_{k \in [n]})}{\mathbb{P}(\mathcal{E} \mid \chi_\emptyset)} \leq \sqrt{1 + \eta} \prod_{i=1}^m \frac{\psi_{v_{\emptyset i}}}{\mathbb{E}_\emptyset[\psi_{v_{\emptyset i}}]}. \quad (5.2.56)$$

With the help of Lemma 5.10, with probability at least $1 - \varepsilon$,

$$(1 - \eta) \prod_{i=1}^m \frac{\chi_{v_{\emptyset i}}}{\mathbb{E}_\emptyset[\chi_{v_{\emptyset i}}]} \leq \frac{\mathbb{P}(\mathcal{E} \mid (\chi_k)_{k \in [n]})}{\mathbb{P}(\mathcal{E} \mid \chi_\emptyset)} \leq (1 + \eta) \prod_{i=1}^m \frac{\chi_{v_{\emptyset i}}}{\mathbb{E}_\emptyset[\chi_{v_{\emptyset i}}]}. \quad (5.2.57)$$

The factors $\chi_{v_{\emptyset i}}/\mathbb{E}_\emptyset[\chi_{v_{\emptyset i}}]$ correspond to the size-biasing of $\hat{\chi}_{v_{\emptyset k}}$ for $k \in [m]$ as

indicated in (5.2.49). Since this is of product structure, asymptotic independence of $(\hat{X}_{v_{\emptyset k}})_{k \in [n] \setminus \{v_{\emptyset}\}}$ remains.

We conclude that with probability at least $1 - \varepsilon$ with respect to \mathbb{P}_{\emptyset} ,

$$(1 - \eta)\hat{\rho}(\cdot) \leq \rho(\cdot \mid \mathcal{E}, \chi_{v_{\emptyset}} = \gamma_{\emptyset}) \leq (1 + \eta)\hat{\rho}(\cdot), \tag{5.2.58}$$

where $\hat{\rho}$ is the density of $(\hat{X}_{v_{\emptyset k}})_{k \in [n] \setminus \{v_{\emptyset}\}}$, whose law we denote by $\hat{\mathbb{P}}$.

To complete the proof, we use [Volume 1, (2.2.8)] to compute that

$$d_{\text{TV}}(\rho(\cdot \mid \mathcal{E}, \chi_{v_{\emptyset}}), \hat{\rho}(\cdot)) = 1 - \int \rho(\vec{\chi} \mid \mathcal{E}, \chi_{v_{\emptyset}}) \wedge \hat{\rho}(\vec{\chi}) d\vec{\chi}. \tag{5.2.59}$$

Let Ω_{\emptyset} denote the where (5.2.58) holds, so that $\mathbb{P}_{\emptyset}(\Omega \mid \mathcal{E}, \chi_{v_{\emptyset}}) \geq 1 - \eta$. Then, we bound

$$d_{\text{TV}}(\rho(\cdot \mid \mathcal{E}, \chi_{v_{\emptyset}}), \hat{\rho}(\cdot)) \leq 1 - \int_{\Omega_{\emptyset}} \rho(\vec{\chi} \mid \mathcal{E}, \chi_{v_{\emptyset}}) \wedge \hat{\rho}(\vec{\chi}) d\vec{\chi}. \tag{5.2.60}$$

On Ω_{\emptyset} , (5.2.58) holds, so that $\hat{\rho}(\vec{\chi}) \geq \rho(\vec{\chi} \mid \mathcal{E}, \chi_{v_{\emptyset}})/(1 + \eta)$. Thus,

$$\begin{aligned} d_{\text{TV}}(\rho(\cdot \mid \mathcal{E}, \chi_{v_{\emptyset}}), \hat{\rho}(\cdot)) &\leq 1 - \frac{1}{1 + \eta} \int_{\Omega_{\emptyset}} \rho(\vec{\chi} \mid \mathcal{E}, \chi_{v_{\emptyset}}) \\ &= 1 - \frac{1}{1 + \eta} \mathbb{P}_{\emptyset}(\Omega \mid \mathcal{E}, \chi_{v_{\emptyset}}) \leq 1 - \frac{1 - \eta}{1 + \eta} = \frac{2\eta}{1 + \eta} \leq 2\eta. \end{aligned} \tag{5.2.61}$$

We conclude that

$$d_{\text{TV}}(\rho(\cdot \mid A, \chi_{v_{\emptyset}}), \hat{\rho}(\cdot)) \leq 2\eta. \tag{5.2.62}$$

Choosing η so small that $2\eta \leq \varepsilon$, this shows that we can couple $(\chi_k^{\mathcal{E}, \gamma_{\emptyset}})_{k \in [n] \setminus \{v_{\emptyset}\}}$ and $(\hat{X}_k)_{k \in [n] \setminus \{v_{\emptyset}\}}$ 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 $(\hat{X}_k)_{k \in [n] \setminus \{v_{\emptyset}\}}$ as these are independent Gamma distributions. This completes the proof of (iv).

The fact that $\mathbb{P}(B_1^{(n)}(o_n) \simeq (H, y)) \rightarrow \mathbb{P}(B_1(\emptyset) \simeq (H, y))$ follows directly from the fact that we have coupled such that $q_{\emptyset}^{(n)} = q_{\emptyset}$, and this occurs with probability at least $1 - \varepsilon$. Since $\varepsilon > 0$ is arbitrary, the statement follows. \square

Exploration of $\text{PA}_n^{(m, \delta)}(c)$: general neighborhood

In this step, we discuss how we can explore the one-neighborhood of a uniformly chosen vertex, so as to prove Theorem 5.6.

Proposition 5.14 (Coupling general neighborhood of uniform vertex in $\text{PA}_n^{(m, \delta)}(c)$ and Pólya-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 $\{v_w, v_{w_i}\}$ 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 $1 - \varepsilon$, $B_r^{(n)}(o_n)$, where $o_n = v_{\emptyset}$, together with the variables $(S_{v_w}^{(n)})_{|w| \leq r}$, and $B_r(\emptyset)$, together with its positions $(X_w)_{i \geq 1}$, can be coupled such that*

- (1) $B_r^{(n)}(o_n) \simeq B_r(\emptyset)$ and $|B_r(\emptyset)| \leq K$;

- (2) $|X_w - S_{v_w}^{(n)}| \leq \eta$ for all $w \in B_r(\emptyset)$;
- (3) $(v_w)_{w \in B_r(\emptyset)}$ are all distinct and $v_w \geq \eta n$ for all $v_w \in B_r^{(n)}(o_n)$;
- (4) $\Gamma_w \leq K$ for all $w \in B_r(\emptyset)$.

Consequently, $\mathbb{P}(B_r^{(n)}(o_n) \simeq (H, y)) \rightarrow \mathbb{P}(B_r(\emptyset) \simeq (H, y))$.

Proposition 5.14 is the natural extension to $r \geq 1$ of Lemma 5.13, and we will see that its proof runs accordingly.

Proof For $r = 1$, this follows from Lemma 5.13. This initiates the induction argument.

Assume by induction that the statement holds for $r - 1$, and fix $B_{r-1}(\emptyset)$, $(X_w)_{w \in B_{r-1}(\emptyset)}$, $(\Gamma_w)_{w \in B_{r-1}(\emptyset)}$, $(\chi_w)_{w \in B_{r-1}(\emptyset)}$ and $(v_w)_{w \in B_{r-1}(\emptyset)}$ such that Proposition 5.14 holds for them (a statement that holds with probability at least $1 - \varepsilon$ according to the induction hypothesis). We now wish to extend these statements to r .

For this, fix a vertex $w \in \partial B_{r-1}(\emptyset) = B_{r-1}(\emptyset) \setminus B_{r-2}(\emptyset)$. We explore the neighborhood of v_w in much the same way as exploring the neighborhood of v_\emptyset in the proof of Lemma 5.13, but we have much more information already that we need to keep in check. To this end, we note that for all $w' \in B_{r-1}(\emptyset)$, the neighborhood of $v_{w'}$ is already determined by our conditioning on $(v_w)_{w \in B_{r-1}(\emptyset)}$. This implies in particular that none of the edges sent out of $v_{w'}$ can be attached to a vertex $v \in B_{r-1}^{(n)}(o_n)$, except when w' is of type R and $v_{w'}$ happens to be the parent of v_w , in which case the edge between $v_{w'}$ and v_w is already present. To determine the children of type L of v_w , we therefore have to condition on not hitting the set $B_{r-1}^{(n)}(o_n)$. Apart from this, the process of determining the children of v_w is exactly the same as that for v_\emptyset in the proof of Lemma 5.13. Since $|B_{r-1}^{(n)}(o_n)| \leq N$, $v_{w'} \geq \eta n$ and $\chi_{v_{w'}} \leq K$ for all $w' \in B_{r-1}^{(n)}(o_n)$, it follows that

$$\sum_{v_{w'} \in B_{r-1}^{(n)}(o_n)} \varphi_{v_{w'}}^{(v_w)} \leq \frac{C}{n} \quad (5.2.63)$$

for some $C > 0$. This implies that conditioning on $v_{w'} \notin B_{r-1}^{(n)}(o_n)$ has only a negligible effect on the distribution of the children of v_w . We thus proceed as in the proof of Lemma 5.13 to obtain a coupling between a sequence of i.i.d. values $(X_{wi})_{i \in [m]}$ of type L and the children $(v_{wi})_{i \in [m]}$. As before, we have that $|S_{v_{wi}}^{(n)} - X_{wi}| \leq \eta$, for n large enough, with probability at least $1 - \varepsilon$.

We can continue with this process for all $v_w \in \partial B_{r-1}^{(n)}(o_n) = B_{r-1}^{(n)}(o_n) \setminus B_{r-2}^{(n)}(o_n)$, thus producing the set L_r consisting of all the type L children of vertex $v_w \in B_r^{(n)}(o_n) \setminus B_{r-1}^{(n)}(o_n)$. It is not hard to see that with probability tending to 1 as $n \rightarrow \infty$, the set L_r has no intersection with $B_{r-1}^{(n)}(o_n)$, 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

in $U_{r-1} \subseteq B_{r-1}^{(n)}(o_n) \setminus B_{r-2}^{(n)}(o_n)$. 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 (5.2.53), 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 (5.2.53), (2) means that we have to use the modified probability

$$\tilde{P}_{v \rightarrow v'} = \frac{\varphi_{v'}^{(v-1)}}{\tilde{S}_{v-1}^{(v-1)}}, \quad (5.2.64)$$

where

$$\tilde{S}_{v-1}^{(v-1)} = \sum_{u > v-1 : u \notin B_{r-1}^{(n)}(o_n) \cup U_r} \varphi_u^{(v-1)}. \quad (5.2.65)$$

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 5.11 to approximate $\tilde{P}_{v \rightarrow v'}$ by $\hat{P}_{v \rightarrow v'}$, which equals

$$\hat{P}_{v \rightarrow v'} = \frac{1}{n} \frac{\chi_{v'}}{2m + \delta} \frac{n}{v'} \left(\frac{v'}{v}\right)^x. \quad (5.2.66)$$

From here onwards, the proof is a straightforward adaptation of the proof of Lemma 5.13, and we omit further details.

The fact that $\mathbb{P}(B_r^{(n)}(o_n) \simeq (H, y)) \rightarrow \mathbb{P}(B_r(\emptyset) \simeq (H, y))$ follows since we have coupled the neighborhoods such that with probability at least $1 - \varepsilon$, $q_w^{(n)} = q_w$ for all $|w| \leq r$. Since $\varepsilon > 0$ is arbitrary, the statement follows. \square

Exploration of $\text{PA}_n^{(m,\delta)}(c)$: two neighborhoods

In this part, we argue that

$$\mathbb{P}(B_r^{(n)}(o_n^{(1)}), B_r^{(n)}(o_n^{(2)}) \simeq (H, y)) \rightarrow \mathbb{P}(B_r(\emptyset) \simeq (H, y))^2, \quad (5.2.67)$$

where $o_n^{(1)}, o_n^{(2)}$ are two vertices chosen uniformly at random and independently from $[n]$.

A second moment method then immediately shows that, for every rooted tree (H, y) ,

$$\frac{1}{n} \sum_{v \in [n]} \mathbb{P}(B_1^{(n)}(v) \simeq (H, y)) \xrightarrow{\mathbb{P}} \mathbb{P}(B_r(\emptyset) \simeq (H, y)), \quad (5.2.68)$$

which completes the proof of the local weak convergence in probability in Theorem 5.6.

We note that

$$\mathbb{P}(B_r^{(n)}(o_n^{(1)}) \cap B_r^{(n)}(o_n^{(2)}) \neq \emptyset) = \mathbb{P}(o_n^{(2)} \notin B_{2r}^{(n)}(o_n^{(1)})) = 1 - o(1), \quad (5.2.69)$$

since $|B_{2r}^{(n)}(o_n^{(1)})|$ is a tight sequence of random variables. Thus, it suffices to show that

$$\mathbb{P}(B_r^{(n)}(o_n^{(1)}), B_r^{(n)}(o_n^{(2)}) \simeq (H, y), B_r^{(n)}(o_n^{(1)}) \cap B_r^{(n)}(o_n^{(2)}) = \emptyset) \rightarrow \mathbb{P}(B_r(\emptyset) \simeq (H, y))^2. \quad (5.2.70)$$

To see (5.2.67), we condition on $B_r^{(n)}(o_n^{(1)})$, which is such that $B_r(\emptyset^{(1)})$, $(X_w)_{w \in B_r(\emptyset^{(1)})}$, $(\Gamma_w)_{w \in B_r(\emptyset)}$, $(\chi_w)_{w \in B_r(\emptyset)}$ and $(v_w)_{w \in B_r(\emptyset)}$ such that Proposition 5.14 holds. This occurs whp, as Proposition 5.14. Consider

$$\mathbb{P}\left(B_r^{(n)}(o_n^{(2)}) \simeq (H, y) \mid B_r^{(n)}(o_n^{(1)}), o_n^{(2)} \notin B_{2r}^{(n)}(o_n^{(1)})\right). \quad (5.2.71)$$

We adapt the proof of Proposition 5.14. Conditioning on $B_r^{(n)}(o_n^{(1)}), o_n^{(2)} \notin B_{2r}^{(n)}(o_n^{(1)})$ has the effect that (5.2.64) is now modified to

$$\tilde{P}_{v \rightarrow v'} = \frac{\varphi_{v'}^{(v-1)}}{\tilde{S}_{v-1,2}^{(v-1)}}, \quad (5.2.72)$$

where now

$$\tilde{S}_{v-1}^{(v-1)} = \sum_{u > v-1: u \notin B_{r-1}^{(n)}(o_n) \cup U_r \cup B_{2r}^{(n)}(o_n^{(1)})} \varphi_u^{(v-1)}. \quad (5.2.73)$$

This again makes very little difference, so that indeed, for $B_r^{(n)}(o_n^{(1)})$, which is such that $B_r(\emptyset^{(1)})$, $(X_w)_{w \in B_r(\emptyset^{(1)})}$, $(\Gamma_w)_{w \in B_r(\emptyset)}$, $(\chi_w)_{w \in B_r(\emptyset)}$ and $(v_w)_{w \in B_r(\emptyset)}$ such that Proposition 5.14 holds,

$$\begin{aligned} & \mathbb{P}\left(B_r^{(n)}(o_n^{(2)}) \simeq (H, y) \mid B_r^{(n)}(o_n^{(1)}), o_n^{(2)} \notin B_{2r}^{(n)}(o_n^{(1)})\right) \\ & \xrightarrow{P} \mathbb{P}(B_r(\emptyset) \simeq (H, y)), \end{aligned} \quad (5.2.74)$$

as required. \square

5.2.2 LOCAL WEAK CONVERGENCE BERNOULLI PREFERENTIAL ATTACHMENT

Recall the Bernoulli preferential attachment model $(\text{BPA}_t^{(f)})_{t \geq 1}$ in Section 1.3.6. A special case is the $(\text{BPA}_t^{(f)})_{t \geq 1}$ with an affine attachment function f , i.e., the setting where there exist $\gamma, \beta > 0$ such that

$$f(k) = \gamma k + \beta. \quad (5.2.75)$$

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 $(\text{PA}_t^{(m,\delta)})_{t \geq 1}$) 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 5.15 (Local weak convergence of preferential attachment models: conditionally independent edges) *Fix $\delta \geq 0$. The preferential attachment model 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

5.3 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

$$N_t = I_1 + I_2 + \cdots + I_t, \quad (5.3.1)$$

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

$$\mathbb{P}(I_i = 1) = \frac{1 + \delta}{(2 + \delta)(i - 1) + 1 + \delta}. \quad (5.3.2)$$

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 5.16 (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 5.16. 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

$$\mathbb{P}(I_t = 1) = \prod_{e=1}^m \frac{2e - 1 + \delta}{(2m + \delta)t + (2e - 1 + \delta)}. \quad (5.3.3)$$

For $m \geq 2$,

$$\sum_{t=2}^{\infty} \mathbb{P}(I_t = 1) < \infty, \quad (5.3.4)$$

so that, almost surely, $I_t = 1$ only occurs *finitely* often. As a result, $\lim_{t \rightarrow \infty} N_t < \infty$ almost surely since $N_t \leq 1 + \sum_{i=2}^{\infty} I_i$. 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 $K \rightarrow \infty$, $\sum_{t=K}^{\infty} \mathbb{1}_{\{N_t > N_{t-1}\}} = 0$. We condition on $\sum_{t=K}^{\infty} \mathbb{1}_{\{N_t > N_{t-1}\}} = 0$, 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 σ -algebra generated by $(\text{PA}_t^{(m,\delta)})_{t=1}^s$. We are left to prove that for t sufficiently large, the vertices $1, \dots, K$ 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 $\mathbb{P}(N_{2t} - N_t \leq -1 \mid \mathcal{F}_K, N_t \geq 2)$ 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

$$\frac{m + \delta}{2(2m + \delta)t} \frac{(m + \delta)t/K}{2(2m + \delta)t} \geq \frac{\varepsilon}{t}, \quad (5.3.5)$$

for some $\varepsilon > 0$. Thus, the probability that this happens for at least one $s \in [2t] \setminus [t]$ is at least

$$1 - \left(1 - \frac{\varepsilon}{t}\right)^t, \quad (5.3.6)$$

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, $N_t \xrightarrow{\text{a.s.}} 1$, 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 $\text{PA}_{2t}^{(m,\delta)}$ is whp connected for all $t \geq T$, where T is large, which implies Theorem 5.16. \square

This proves that $\text{PA}_{2t}^{(m,\delta)}$ is whp connected for all $t \geq T$, where T is large, which implies Theorem 5.16. \square

5.3.1 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 5.17 (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*

$$\gamma \geq \frac{1}{2} \quad \text{or} \quad \beta > \frac{(1/2 - \gamma)^2}{1 - \gamma}. \quad (5.3.7)$$

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 γ 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 $f: \{0, 1, 2, \dots\} \mapsto (0, \infty)$ *concave* when

$$f(0) \leq 1 \quad \text{and} \quad \Delta f(k) := f(k + 1) - f(k) < 1 \text{ for all } k \geq 0. \quad (5.3.8)$$

By concavity the limit

$$\gamma := \lim_{n \rightarrow \infty} \frac{f(n)}{n} = \min_{k \geq 0} \Delta f(k) \quad (5.3.9)$$

exists. γ plays a crucial role in the analysis, as can already be observed in Theorem 5.17.

Let $(Z_t)_{t \geq 0}$ be a pure birth Markov process started with birth rate $f(k)$ when it is in state k . Let $\mathcal{S} := \{\ell\} \cup [0, \infty]$ be a type space. Given a $\alpha \in (0, 1)$, define the increasing functions M , respectively, M^τ , by

$$M(t) = \int_0^t e^{-s} \mathbb{E}[f(Z_s)] ds, \quad M(t) = \mathbb{E}[Z_t], \quad M^\tau(t) = \mathbb{E}[Z_t \mid \Delta Z_\tau = 1] - \mathbb{1}_{[\tau, \infty)}(t), \quad (5.3.10)$$

Next, define a linear operator A_α on the Banach space $C(\mathcal{S})$ of continuous, bounded functions on $\mathcal{S} := \{\ell\} \cup [0, \infty]$ with ℓ being a (non-numerical) symbol, by

$$(A_\alpha g)(\tau) = \int_0^\infty g(t) e^{\alpha t} dM(t) + \int_0^\infty g(t) e^{-\alpha t} dM^\tau(t). \quad (5.3.11)$$

The operator A_α 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 5.18 (Existence of a giant component) *No giant component exists if and only if there exists $0 < \alpha < 1$ such that A_α is a compact operator with spectral radius $\rho(A_\alpha) \leq 1$.*

It turns out that A_α 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_α 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.

TO DO 5.2: Extend the discussion of the phase transition for BPAM

5.4 FURTHER RESULTS FOR PREFERENTIAL ATTACHMENT MODELS

TO DO 5.3: Add further results!

5.5 NOTES AND DISCUSSION

Notes on Section 5.1

The proof of Theorem 5.1 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 5.6 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 5.6 is (Berger et al., 2014, Theorem 2.2). Theorem 5.6 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 5.8 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 5.8 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 5.8 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 5.4 for more details.

Theorem 5.15 is proved by Dereich and Mörters (2009, 2011, 2013) for the Bernoulli preferential attachment model.

Notes on Section 5.3

The result on the giant component for preferential attachment models with conditionally independent edges in Theorem 5.18 is proved by Dereich and Mörters (2009, 2011, 2013).

Notes on Section 5.4

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).

5.6 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 5.1)) *Use De Finetti's Theorem (Theorem 5.1) to prove that $S_n/n \xrightarrow{\text{a.s.}} U$. 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*

$$\mathbb{P}(X_k = X_n = 1) \geq \mathbb{P}(X_k = 1)\mathbb{P}(X_n = 1). \quad (5.6.1)$$

Prove that equality holds if and only if there exists a p such that $\mathbb{P}(U = p) = 1$.

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 5.3 is such that

$$\frac{S_1(n)}{S_1(n) + S_2(n)} \xrightarrow{\text{a.s.}} U, \quad (5.6.2)$$

where U is uniform on $[0, 1]$. Use this to prove that $\mathbb{P}(S_1(n) = k) = 1/n$ for each $k \in [n]$.

Exercise 5.8 (Scale-free trees) Recall the model studied in Theorem 5.3, 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 $(\text{PA}_t^{(1,\delta)}(b))_{t \geq 1}$ arises when $d_1 = 3$ and $d_2 = 1$. What does Theorem 5.3 imply for $(\text{PA}_t^{(1,\delta)}(b))_{t \geq 1}$?

Exercise 5.9 (Relative degrees of vertices 1 and 2) Use Theorem 5.4 to compute $\lim_{t \rightarrow \infty} \mathbb{P}(D_2(t) \geq xD_1(t))$ for $(\text{PA}_t^{(1,\delta)})_{t \geq 1}$.

Exercise 5.10 (Proof of Theorem 5.5) Complete the proof of Theorem 5.5.

Exercise 5.11 (Size-biased version of Gamma) Let X have a Gamma distribution with parameter r . Show that its size-biased version X^* has a Gamma distribution with parameter $r + 1$.

Exercise 5.12 (Power-law exponents in $\text{PA}_t^{(m,\delta)}(c)$) Prove the power-law relations in (5.2.9) and identify $c_{m,\delta}$ and $c'_{m,\delta}$.

Exercise 5.13 (Joint law (D, D') for $\text{PA}_t^{(m,\delta)}(c)$ (Berger et al., 2014, Lemma 5.3)) Adapt the proof of Lemma 5.7 to show that, for $j \geq m$ and $k \geq m + 1$,

$$\begin{aligned} \mathbb{P}(D = j, D' = k) & \\ &= \frac{2m + \delta}{m^2} \frac{\Gamma(k + 1 + \delta)}{k! \Gamma(m + 1 + \delta)} \frac{\Gamma(j + \delta)}{j! \Gamma(m + \delta)} \\ &\quad \times \int_0^1 \int_v^1 (1 - v)^k v^{m+1+\delta+\delta/m} (1 - u)^j u^{m+\delta} du dv. \end{aligned} \quad (5.6.3)$$

Exercise 5.14 (Joint law (D, D') for $\text{PA}_t^{(m,\delta)}(c)$ (Berger et al., 2014, Lemma 5.3)) Use Lemma 5.7 and Exercise 5.13 to show that, for fixed $j \geq m$ and as $k \rightarrow \infty$,

$$\mathbb{P}(D' = k \mid D = j) = C_j k^{-(2+\delta/m)} (1 + O(1/k)). \quad (5.6.4)$$

Exercise 5.15 (Joint law (D, D') for $\text{PA}_t^{(m,\delta)}(c)$ (Berger et al., 2014, Lemma

5.3)) Use Lemma 5.7 and Exercise 5.13 to show that, for fixed $k \geq m + 1$ and as $j \rightarrow \infty$,

$$\mathbb{P}(D = j \mid D' = k) = \tilde{C}_j j^{-(m+4+\delta+\delta/m)}(1 + O(1/j)). \quad (5.6.5)$$

Exercise 5.16 (Simple form of S_k in (5.2.19)) Prove, using induction on k , that $S_k^{(n)} = \prod_{i=k+1}^n (1 - \psi_i)$ in (5.2.19) holds.

Exercise 5.17 (Multiple edges and Theorem 5.8) Fix $m = 2$. Let M_t denote the number of multiple edges in $\text{PA}_t^{(m,\delta)}(c)$. Use Theorem 5.8 to show that

$$\mathbb{E}[M_{t+1} - M_t] = \sum_{k=1}^t \mathbb{E}\left[\left(\frac{\varphi_k}{S_{t-1}^{(n)}}\right)^2\right]. \quad (5.6.6)$$

Exercise 5.18 (Multiple edges and Theorem 5.8 (cont.)) Fix $m = 2$. Let M_t denote the number of multiple edges in $\text{PA}_t^{(m,\delta)}(c)$. Compute $\mathbb{E}\left[\left(\frac{\varphi_k}{S_{t-1}^{(n)}}\right)^2\right]$ and use this to show that $\mathbb{E}[M_t]/\log t \rightarrow c$, and identify $c > 0$.

Exercise 5.19 (Multiple edges and Theorem 5.8 (cont.)) Fix $m = 2$. Let M_t denote the number of multiple edges in $\text{PA}_t^{(m,\delta)}(c)$. Use Theorem 5.8 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

$$\mathbb{P}(M_{t+1} - M_t = 1 \mid (\psi_k)_{k \geq 1}) = \sum_{k=1}^{t-1} \left(\frac{\varphi_k}{S_{t-1}^{(n)}}\right)^2. \quad (5.6.7)$$

Exercise 5.20 (Almost sure limit of F_k in (5.2.20)) Fix $k \geq 1$. Prove that $(M_n(k))_{n \geq k+1}$, where

$$M_n(k) = \prod_{j=k+1}^n \frac{1 - \psi_j}{\mathbb{E}[1 - \psi_j]}, \quad (5.6.8)$$

is a multiplicative positive martingale and using the Martingale Convergence Theorem ([Volume 1, Theorem 2.24]) and thus converges. Conclude that (5.2.20) holds by showing that $\prod_{j=k+1}^n \mathbb{E}[1 - \psi_j] = (k/n)^x(1 + o(1))$.

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 μ such that $\mathbb{E}[|E(\text{BPA}_t^{(f)})|]/t \rightarrow \mu$.

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 $|E(\text{BPA}_t^{(f)})|/t \xrightarrow{\mathbb{P}} \mu$. Show that $D_t(t) \xrightarrow{d} \text{Poi}(\mu)$.

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

$$\mathbb{E}[N_t] = \frac{1 + \delta}{2 + \delta} \log t(1 + o(1)), \quad \text{Var}(N_t) = \frac{1 + \delta}{2 + \delta} \log t(1 + o(1)). \quad (5.6.9)$$

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 $N_t/\log t \xrightarrow{\mathbb{P}} (1 + \delta)/(2 + \delta)$.

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 $S_t/\log t \xrightarrow{\mathbb{P}} m(m + 1 + 2\delta)/[2(2m + \delta)]$.

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 $S_t/\log t \xrightarrow{\mathbb{P}} (m - 1)(m + 2\delta)/[2(2m + \delta)]$.

Exercise 5.27 (All-time connectivity for $(\text{PA}_t^{(m,\delta)})_{t \geq 1}$) Fix $m \geq 2$. Compute the probability that $(\text{PA}_t^{(m,\delta)})_{t \geq 1}$ 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 macroscopic 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 random graph model with power-law degrees having power-law exponent τ , 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 general 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 $\text{IRG}_n(\kappa_n)$ where, as usual, (κ_n) is a graphical sequence of kernels with limit κ .

Recall that we write $\text{dist}_G(i, j)$ 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 $\text{dist}_G(i, j) = \infty$ when i, j are in different connected components. We define the *typical graph distance* to be $\text{dist}_G(o_1, o_2)$, 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 $\text{dist}_{\text{IRG}_n(\kappa_n)}(o_1, o_2) = \infty$. By Theorem 3.16, $\mathbb{P}(\text{dist}_{\text{IRG}_n(\kappa_n)}(o_1, o_2) = \infty) \rightarrow 1 - \zeta^2 > 0$, since $\zeta < 1$ (see Exercise 3.32). In particular, when $\zeta = 0$, which is equivalent to $\nu \leq 1$, $\mathbb{P}(\text{dist}_{\text{IRG}_n(\kappa_n)}(o_1, o_2) = \infty) \rightarrow 1$. 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 $\nu = \|\mathbf{T}_\kappa\| \in (1, \infty)$. When $\|\mathbf{T}_\kappa\| = \infty$, instead, then our results also prove that $\text{dist}_{\text{IRG}_n(\kappa_n)}(o_1, o_2) = o_{\mathbb{P}}(\log n)$, 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 (κ_n) be graphical sequence of kernels with limit κ , and with $\nu = \|\mathbf{T}_\kappa\| \in (1, \infty)$. Let $\varepsilon > 0$ be fixed. Then, for $\text{IRG}_n(\kappa_n)$,*

(i) *If $\sup_{x,y,n} \kappa_n(x, y) < \infty$, then*

$$\mathbb{P}(\text{dist}_{\text{IRG}_n(\kappa_n)}(o_1, o_2) \leq (1 - \varepsilon) \log_\nu n) = o(1). \quad (6.1.1)$$

(ii) *If κ is irreducible, then*

$$\mathbb{P}(\text{dist}_{\text{IRG}_n(\kappa_n)}(o_1, o_2) \leq (1 + \varepsilon) \log_\nu n) = \zeta_\kappa^2 + o(1). \quad (6.1.2)$$

In the terminology of [Volume 1, Section 1.4], Theorem 6.1(ii) implies that $\text{IRG}_n(\kappa)$ is a *small world* when κ is irreducible and $\nu = \|\mathbf{T}_\kappa\| < \infty$. 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 (3.4.8), a Poisson multitype branching process with kernel κ has neighborhoods that grow exponentially, i.e., the number of vertices at distance k grows like $\|\mathbf{T}_\kappa\|^k$. 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 of order n , so that we need that $\|\mathbf{T}_\kappa\|^k = \nu^k \sim n$, 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 τ satisfies $\tau \in (2, 3)$.

Distances in rank-1 IRGs with finite-variance weights

We continue by investigating the behavior of $\text{dist}_{\text{NR}_n(\mathbf{w})}(o_1, o_2)$ for $\text{NR}_n(\mathbf{w})$ in the case where the weights have finite variance:

Theorem 6.2 (Typical distances in $\text{NR}_n(\mathbf{w})$ for finite-variance weights) *In the Norros-Reittu model $\text{NR}_n(\mathbf{w})$, where the weights $\mathbf{w} = (w_i)_{i \in [n]}$ satisfy Condition 1.1(a)-(c) and where $\nu > 1$, conditionally on $\text{dist}_{\text{NR}_n(\mathbf{w})}(o_1, o_2) < \infty$,*

$$\frac{\text{dist}_{\text{NR}_n(\mathbf{w})}(o_1, o_2)}{\log n} \xrightarrow{\mathbb{P}} 1/\log \nu. \quad (6.1.3)$$

The same result applies, under the same conditions, to $\text{GRG}_n(\mathbf{w})$ and $\text{CL}_n(\mathbf{w})$.

Theorem 6.2 can be seen as a special case of Theorem 6.1. However, in Theorem 6.1(i), we require that κ_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 $\text{NR}_n(\mathbf{w})$ 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 $\nu = \sum_{k \geq 0} k p_k^* < \infty$, the number of vertices at distance k is close to $M \nu^k$, where M is the martingale limit of Z_k / ν^k . To know what $\text{dist}_{\text{NR}_n(\mathbf{w})}(o_1, o_2)$ 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 $\text{dist}_{\text{NR}_n(\mathbf{w})}(o_1, o_2)$. 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 ν^k / ℓ_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 $(1 + \varepsilon) \log_\nu n$ 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 $\text{dist}_{\text{NR}_n(\mathbf{w})}(o_1, o_2) = o_\mathbb{P}(\log n)$, 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 $\text{NR}_n(\mathbf{w})$, in the case where the degrees obey a power-law with degree exponent τ satisfying

that $\tau \in (2, 3)$. In this case, $\nu = \infty$, so that $\text{dist}_{\text{NR}_n(\mathbf{w})}(o_1, o_2) = o_{\mathbb{P}}(\log n)$ (recall Exercise 6.2). It turns out that the precise scaling of $\text{dist}_{\text{NR}_n(\mathbf{w})}(o_1, o_2)$ depends sensitively on the precise way how $\nu_n \rightarrow \infty$. Below, we assume that they obey a power-law with exponent τ 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 $\text{GRG}_n(\mathbf{w})$ and the Chung-Lu model $\text{CL}_n(\mathbf{w})$. In this section, we discuss the setting where the weights \mathbf{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 ,

$$c_1 x^{-(\tau-1+\delta)} \leq [1 - F_n](x) \leq c_2 x^{-(\tau-1-\delta)}, \quad (6.1.4)$$

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 $\text{NR}_n(\mathbf{w})$ for $\tau \in (2, 3)$) *Fix the Norros-Reittu model $\text{NR}_n(\mathbf{w})$, where the weights $\mathbf{w} = (w_i)_{i \in [n]}$ satisfy Conditions 1.1(a)-(b) and (6.1.4). Then, conditionally on $\text{dist}_{\text{NR}_n(\mathbf{w})}(o_1, o_2) < \infty$,*

$$\frac{\text{dist}_{\text{NR}_n(\mathbf{w})}(o_1, o_2)}{\log \log n} \xrightarrow{\mathbb{P}} \frac{2}{|\log(\tau - 2)|}. \quad (6.1.5)$$

The same results apply, under the same conditions, to $\text{CL}_n(\mathbf{w})$ and $\text{GRG}_n(\mathbf{w})$.

Theorem 6.3 implies that $\text{NR}_n(\mathbf{w})$ with \mathbf{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 $\text{NR}_n(\mathbf{w})$ is a comparison to branching processes, which is particularly pretty for $\text{NR}_n(\mathbf{w})$. 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 $k \rightarrow \infty$. 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.

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 $\text{NR}_n(\mathbf{w})$ is

$$e^{-\sum_{i: w_i > y} w w_i / \ell_n} = e^{-w[1 - F_n^*](y)}, \tag{6.1.6}$$

where $F_n^*(y) = \sum_{i: w_i \leq y} w_i / \ell_n$ is the distribution function of W_n^* introduced in (6.2.33). When (6.1.4) holds, it follows that $[1 - F_n^*](y)$ 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 $\text{NR}_n(\mathbf{w})$ is approximately $e^{-w y^{-(\tau-2)}}$. 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 $\text{NR}_n(\mathbf{w})$, 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 $\text{NR}_n(\mathbf{w})$, 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 $\text{NR}_n(\mathbf{w})$. The main result is as follows:

Theorem 6.4 (Logarithmic lower bound graph distances $\text{NR}_n(\mathbf{w})$) *Assume that*

$$\limsup_{n \rightarrow \infty} \nu_n = \nu, \tag{6.2.1}$$

where $\nu \in (1, \infty)$ and

$$\nu_n = \mathbb{E}[W_n^2] / \mathbb{E}[W_n] = \sum_{i \in [n]} w_i^2 / \sum_{i \in [n]} w_i. \tag{6.2.2}$$

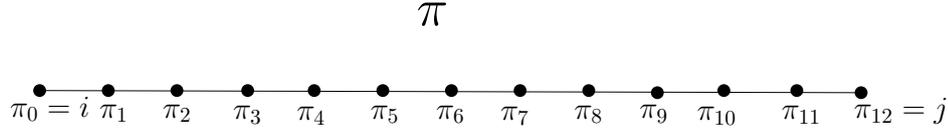


Figure 6.1 A 12-step self-avoiding path connecting vertices i and j .

Then, for any $\varepsilon > 0$,

$$\mathbb{P}(\text{dist}_{\text{NR}_n(\mathbf{w})}(o_1, o_2) \leq (1 - \varepsilon) \log_\nu n) = o(1). \quad (6.2.3)$$

The same results hold for $\text{CL}_n(\mathbf{w})$ and $\text{GRG}_n(\mathbf{w})$ 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

$$\begin{aligned} \mathbb{P}(\text{dist}_{\text{NR}_n(\mathbf{w})}(o_1, o_2) \leq k_n) &= \frac{1}{n^2} \sum_{i, j \in [n]} \mathbb{P}(\text{dist}_{\text{NR}_n(\mathbf{w})}(i, j) \leq k_n) \\ &= \frac{1}{n^2} \sum_{i, j \in [n]} \sum_{k=0}^{k_n} \mathbb{P}(\text{dist}_{\text{NR}_n(\mathbf{w})}(i, j) = k). \end{aligned} \quad (6.2.4)$$

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 $\pi = (\pi_0, \dots, \pi_k)$ 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 π *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 $\text{dist}_{\text{NR}_n(\mathbf{w})}(i, j) = k$, there must be path of length k such that all edges (π_l, π_{l+1}) are occupied in $\text{NR}_n(\mathbf{w})$, for $l = 0, \dots, k-1$. The probability in $\text{NR}_n(\mathbf{w})$ that the edge (π_l, π_{l+1}) is occupied is equal to

$$1 - e^{-w_{\pi_l} w_{\pi_{l+1}} / \ell_n} \leq w_{\pi_l} w_{\pi_{l+1}} / \ell_n. \quad (6.2.5)$$

For $\text{CL}_n(\mathbf{w})$ and $\text{GRG}_n(\mathbf{w})$, an identical upper bound holds, which explains why the proof of Theorem 6.4 for $\text{NR}_n(\mathbf{w})$ applies verbatim to those models.

We say that π is *occupied* when all edges in π are occupied in $\text{NR}_n(\mathbf{w})$. Then, by the union bound or Boole's inequality,

$$\begin{aligned} \mathbb{P}(\text{dist}_{\text{NR}_n(\mathbf{w})}(i, j) = k) &\leq \mathbb{P}(\exists \pi \in \mathcal{P}_k(i, j) : \pi \text{ occupied}) \\ &\leq \sum_{\pi \in \mathcal{P}_k(i, j)} \mathbb{P}(\pi \text{ occupied}). \end{aligned} \tag{6.2.6}$$

For any path $\pi \in \mathcal{P}_k(i, j)$,

$$\begin{aligned} \mathbb{P}(\pi \text{ occupied}) &= \prod_{s=0}^{k-1} \mathbb{P}((\pi_s, \pi_{s+1}) \text{ occupied}) \leq \prod_{l=0}^{k-1} w_{\pi_l} w_{\pi_{l+1}} / \ell_n \\ &= \frac{w_{\pi_0} w_{\pi_k}}{\ell_n} \prod_{l=1}^k w_{\pi_l}^2 / \ell_n = \frac{w_i w_j}{\ell_n} \prod_{l=1}^k w_{\pi_l}^2 / \ell_n. \end{aligned} \tag{6.2.7}$$

Therefore,

$$\begin{aligned} \mathbb{P}(\text{dist}_{\text{NR}_n(\mathbf{w})}(i, j) = k) &\leq \frac{w_i w_j}{\ell_n} \sum_{\pi \in \mathcal{P}_k(i, j)} \prod_{l=1}^k \frac{w_{\pi_l}^2}{\ell_n} \\ &\leq \frac{w_i w_j}{\ell_n} \prod_{l=1}^k \left(\sum_{\pi_l \in [n]} \frac{w_{\pi_l}^2}{\ell_n} \right) = \frac{w_i w_j}{\ell_n} \nu_n^k, \end{aligned} \tag{6.2.8}$$

where ν_n is defined in (6.2.2). We conclude that

$$\begin{aligned} \mathbb{P}(\text{dist}_{\text{NR}_n(\mathbf{w})}(i, j) \leq k_n) &\leq \frac{1}{n^2} \sum_{i, j \in [n]} \sum_{k=0}^{k_n} \frac{w_i w_j}{\ell_n} \nu_n^k = \frac{\ell_n}{n^2} \sum_{k=0}^{k_n} \nu_n^k \\ &= \frac{\ell_n}{n^2} \frac{\nu_n^{k_n+1} - 1}{\nu_n - 1}. \end{aligned} \tag{6.2.9}$$

By (6.2.1), $\limsup_{n \rightarrow \infty} \nu_n = \nu \in (1, \infty)$, so that, for n large enough, $\nu_n \geq (\nu - \delta) > 1$, while $\ell_n/n = \mathbb{E}[W_n] \rightarrow \mathbb{E}[W] < \infty$. Thus, since $\nu \mapsto (\nu^{k+1} - 1)/(\nu - 1)$ is increasing for every integer $k \geq 0$,

$$\mathbb{P}(\text{dist}_{\text{NR}_n(\mathbf{w})}(o_1, o_2) \leq k_n) \leq O((\nu - \delta)^{k_n} / n) = o(1), \tag{6.2.10}$$

when $\delta = \delta(\varepsilon) > 0$ is chosen such that $(1 - \varepsilon)/\log(\nu - \delta) < 1$, and since $k_n = \lceil (1 - \varepsilon) \log_\nu n \rceil$. This completes the proof of Theorem 6.4. \square

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 $(\text{dist}_{\text{NR}_n(\mathbf{w})}(o_1, o_2) \leq \frac{\log n}{\log \nu_n})_-$ is tight.

We close this section by extending the above result to settings where ν_n is not necessarily bounded, the most interesting case being $\tau = 3$:

Corollary 6.5 (Lower bound graph distances $\text{NR}_n(\mathbf{w})$ for $\tau = 3$) *Let ν_n be given in (6.2.2). Then, for any $\varepsilon > 0$,*

$$\mathbb{P}\left(\text{dist}_{\text{NR}_n(\mathbf{w})}(o_1, o_2) \leq (1 - \varepsilon) \log_{\nu_n} n\right) = o(1). \quad (6.2.11)$$

The same results hold for $\text{CL}_n(\mathbf{w})$ and $\text{GRG}_n(\mathbf{w})$ 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

$$\mathbb{P}\left(\text{dist}_{\text{NR}_n(\mathbf{w})}(o_1, o_2) \leq (1 - \varepsilon) \frac{\log n}{\log \log n}\right) = o(1). \quad (6.2.12)$$

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

$$\mathbb{P}(\text{dist}_{\text{IRG}_n(\kappa_n)}(i, j) = k) \leq \sum_{i_1, \dots, i_{k-1} \in [n]} \prod_{l=0}^{k-1} \frac{\kappa_n(x_{i_l}, x_{i_{l+1}})}{n}, \quad (6.2.13)$$

where $i_0 = i, i_k = j$ and we can restrict the vertices to be *distinct*, so that

$$\mathbb{P}(\text{dist}_{\text{IRG}_n(\kappa_n)}(i, j) = k) \leq \frac{1}{n^k} \sum_{i_0, i_1, \dots, i_{k-1}, i_k \in [n]} \prod_{l=0}^{k-1} \kappa_n(x_{i_l}, x_{i_{l+1}}). \quad (6.2.14)$$

If the above $(k + 1)$ -dimensional discrete integrals could be replaced by the continuous integral, then we would arrive at

$$\frac{1}{n} \int_{\mathcal{S}} \cdots \int_{\mathcal{S}} \prod_{l=0}^k \kappa(x_l, x_{l+1}) \prod_{i=0}^k \mu(dx_i) = \frac{1}{n} \|\mathbf{T}_{\kappa}^{k+1} \mathbf{1}\|_1, \quad (6.2.15)$$

which is bounded from above by $\frac{1}{n} \|\mathbf{T}_{\kappa}\|^{k+1}$. Repeating the bound in (6.2.10) would then prove that, when $\nu = \|\mathbf{T}_{\kappa}\| > 1$,

$$\mathbb{P}(\text{dist}_{\text{IRG}_n(\kappa_n)}(i, j) \leq (1 - \varepsilon) \log_{\nu} n) = o(1). \quad (6.2.16)$$

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

$$\begin{aligned} \mathbb{P}(\text{dist}_{\text{IRG}_n(\kappa)}(i, j) = k) &\leq \frac{1}{n^2} \sum_{v_1, \dots, v_{k-1} \in [n]} \prod_{l=0}^{k-1} \frac{\kappa_n(x_{v_l}, x_{v_{l+1}})}{n} \\ &\leq \frac{1}{n^2} \sum_{i_1, \dots, i_{k-1} \in [n]} \prod_{l=0}^{k-1} \kappa^{(n)}(i_l, i_{l+1}) \frac{n_{i_{l+1}}}{n}, \end{aligned} \quad (6.2.17)$$

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 $n_i/n \rightarrow p_i$ and that $\kappa^{(n)}(i, j) \rightarrow \kappa_{ij}$. This also implies that $\|\mathbf{T}_{\kappa_n}\| \rightarrow \nu$, where ν is largest eigenvalue of the matrix $(m_{ij})_{i, j \in [r]}$ with $m_{ij} = \kappa_{ij} p_j$. Denoting $m_{ij}^{(n)} = \kappa^{(n)}(i, j) n_j/n \rightarrow m_{ij}$, we obtain

$$\begin{aligned} \mathbb{P}(\text{dist}_{\text{IRG}_n(\kappa)}(i, j) = k) &\leq \frac{1}{n^2} \sum_{v_1, \dots, v_{k-1} \in [n]} \prod_{l=0}^{k-1} \frac{\kappa_n(x_{v_l}, x_{v_{l+1}})}{n} \\ &\leq \frac{1}{n} \langle \boldsymbol{\mu}^T, [M^{(n)}]^k \mathbf{1} \rangle, \end{aligned} \quad (6.2.18)$$

where $\mathbf{1}$ is the all-one vector, $\boldsymbol{\mu}_i = n_i/n \rightarrow p_i$, and $M_{ij}^{(n)} = m_{ij}^{(n)}$. Obviously, where there are $N < \infty$ types,

$$\langle \boldsymbol{\mu}^T, [M^{(n)}]^k \mathbf{1} \rangle \leq \|M^{(n)}\|^k \|\boldsymbol{\mu}\| \|\mathbf{1}\| \leq \|M^{(n)}\|^k \sqrt{N}. \quad (6.2.19)$$

Thus,

$$\mathbb{P}(\text{dist}_{\text{IRG}_n(\kappa)}(i, j) = k) \leq \frac{\sqrt{N}}{n} \|M^{(n)}\|^k. \quad (6.2.20)$$

We conclude that

$$\mathbb{P}(d_{\text{IRG}_n(\kappa_n)}(o_1, o_2) \leq (1 - \varepsilon) \log_{\nu_n} n) = o(1), \quad (6.2.21)$$

where $\nu_n = \|M^{(n)}\| \rightarrow \nu$. 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 κ_n from above by κ_m^+ , which is of finite-type. Then, use the fact that $\|\mathbf{T}_{\kappa_m^+}\| \downarrow \|\mathbf{T}_{\kappa}\| = \nu > 1$ to conclude that $\mathbb{P}(d_{\text{IRG}_n(\kappa_n)}(o_1, o_2) \leq (1 - \varepsilon) \log_{\nu} n) = o(1)$ 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 $\|\mathbf{T}_{\kappa}\| = \infty$, which, for example for $\text{CL}_n(\boldsymbol{w})$, is the case when F has infinite second moment. (Bollobás et al., 2007, Theorem 3.14(iv)) states that when $\|\mathbf{T}_{\kappa}\| = \infty$, the typical graph distance is smaller than $\log n$. More precisely, (Bollobás et al., 2007, Theorem 3.14(iv)) states

that if κ is irreducible and $\|\mathbf{T}_\kappa\| = \infty$, then there is a function $f(n) = o(\log n)$ such that

$$\mathbb{P}(\text{dist}_{\text{IRG}_n(\kappa)}(i, j) \leq f(n)) = \zeta_\kappa^2 + o(1). \tag{6.2.22}$$

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 $\text{NR}_n(\mathbf{w})$ for $\tau \in (2, 3)$. The main result we prove is the following theorem:

Theorem 6.6 (Loglog lower bound on typical distances in $\text{NR}_n(\mathbf{w})$) *Suppose that the weights $\mathbf{w} = (w_i)_{i \in [n]}$ satisfy Condition 1.1(a) and that there exists a $\tau \in (2, 3)$ and c_2 such that, for all $x \geq 1$,*

$$[1 - F_n](x) \leq c_2 x^{-(\tau-1)}, \tag{6.2.23}$$

Then, for every $\varepsilon > 0$,

$$\mathbb{P}\left(\text{dist}_{\text{NR}_n(\mathbf{w})}(i, j) \leq (1 - \varepsilon) \frac{2 \log \log n}{|\log(\tau - 2)|}\right) = o(1). \tag{6.2.24}$$

The same results hold for $\text{CL}_n(\mathbf{w})$ and $\text{GRG}_n(\mathbf{w})$ 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), ν_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 ν_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

$$\sum_{j: w_j \geq y} \frac{w w_j}{\ell_n} = w[1 - F_n^*](y), \tag{6.2.25}$$

which is small when y is too large. On the other hand, the main contribution to ν_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 $\nu_n = \sum_{s \in [n]} w_s^2 / \ell_n$, 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.

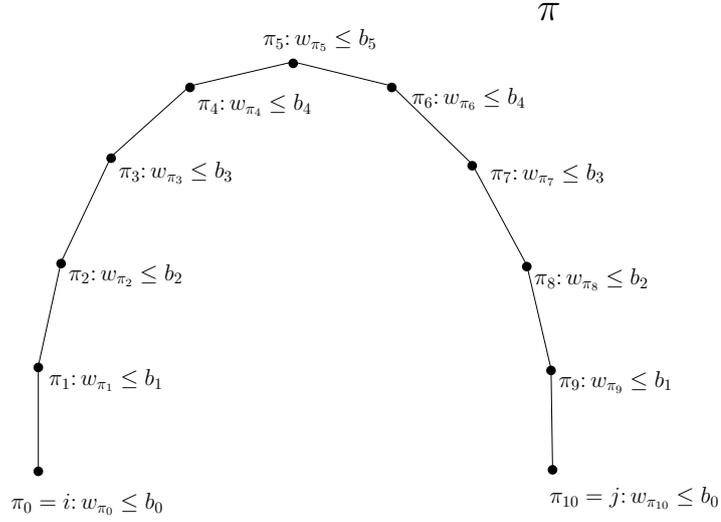


Figure 6.2 A 10-step good path connecting i and j and the upper bounds on the weight of its vertices. The height of a vertex is high for vertices with large weights.

We now present the details for this argument. We again start from

$$\mathbb{P}(\text{dist}_{\text{NR}_n(w)}(o_1, o_2) \leq k_n) = \frac{1}{n^2} \sum_{i, j \in [n]} \mathbb{P}(\text{dist}_{\text{NR}_n(w)}(i, j) \leq k_n). \quad (6.2.26)$$

When $\text{dist}_{\text{NR}_n(w)}(i, j) \leq k_n$, 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 $l = 0, \dots, k$, and bad otherwise. The condition $w_{\pi_l} \leq b_l \wedge b_{k-l}$ for every $l = 0, \dots, k$ 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 $(k-l)^{\text{th}}$ 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

$$\mathcal{E}_k(i, j) = \{\exists \pi \in \mathcal{GP}_k(i, j) : \pi \text{ occupied}\} \quad (6.2.27)$$

denote the event that there exists a good path of length k .

When $\text{dist}_{\text{NR}_n(w)}(i, j) \leq k_n$, 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

$$\mathcal{BP}_k(i) = \{\pi \in \mathcal{P}_k(i) : w_{\pi_l} > b_l, w_{\pi_s} \leq b_s \forall s < l\} \quad (6.2.28)$$

denote the set of *bad paths* of length k , i.e., those $\pi \in \mathcal{P}_k(i)$ that are not in $\mathcal{GP}_k(i, \pi_k)$. Let $\mathcal{F}_l(i)$ be the event that there exists a bad path of length l starting from i , i.e.,

$$\mathcal{F}_l(i) = \{\exists \pi \in \mathcal{BP}_l(i) : \pi \text{ occupied}\}. \quad (6.2.29)$$

Then, since $\text{dist}_{\text{NR}_n(w)}(i, j) \leq k_n$ implies that there either is a good path or a bad path,

$$\{\text{dist}_{\text{NR}_n(w)}(i, j) \leq k_n\} \subseteq \bigcup_{k \leq k_n} (\mathcal{F}_k(i) \cup \mathcal{F}_k(j) \cup \mathcal{E}_k(i, j)), \quad (6.2.30)$$

so that, by Boole's inequality,

$$\mathbb{P}(\text{dist}_{\text{NR}_n(w)}(i, j) \leq k_n) \leq \sum_{k=0}^{k_n} [\mathbb{P}(\mathcal{F}_k(i)) + \mathbb{P}(\mathcal{F}_k(j)) + \mathbb{P}(\mathcal{E}_k(i, j))]. \quad (6.2.31)$$

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

$$\nu_n(b) = \frac{1}{\ell_n} \sum_{i \in [n]} w_i^2 \mathbb{1}_{\{w_i \leq b\}}, \quad (6.2.32)$$

be the restriction of ν_n to vertices with weights at most b , and let

$$F_n^*(x) = \frac{1}{\ell_n} \sum_{i \in [n]} w_i \mathbb{1}_{\{w_i \leq x\}} \quad (6.2.33)$$

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,*

$$\mathbb{P}(\mathcal{F}_k(i)) \leq w_i [1 - F_n^*](b_k) \prod_{l=1}^{k-1} \nu_n(b_l), \quad (6.2.34)$$

and

$$\mathbb{P}(\mathcal{E}_k(i, j)) \leq \frac{w_i w_j}{\ell_n} \prod_{l=1}^{k-1} \nu_n(b_l \wedge b_{k-l}), \quad (6.2.35)$$

where

$$\nu_n(b) \leq c_\nu b^{3-\tau}. \quad (6.2.36)$$

When $b_l = \infty$ for each l , the bound in (6.2.35) equals that obtained in (6.2.8).

Proof We start by proving (6.2.34). By Boole's inequality,

$$\mathbb{P}(\mathcal{F}_k(i)) = \mathbb{P}(\exists \pi \in \mathcal{BP}_l(i) : \pi \text{ occupied}) \leq \sum_{\pi \in \mathcal{BP}_l(i)} \mathbb{P}(\pi \text{ occupied}). \quad (6.2.37)$$

By (6.2.7), (6.2.32) and (6.2.33),

$$\begin{aligned} \mathbb{P}(\mathcal{F}_k(i)) &\leq \sum_{\pi \in \mathcal{BP}_l(i)} \frac{w_i w_{\pi_k}}{\ell_n} \prod_{l=1}^{k-1} w_{\pi_l}^2 / \ell_n \\ &\leq w_i \sum_{\pi_k : w_{\pi_k} \geq b_k} \frac{w_{\pi_k}}{\ell_n} \times \prod_{l=1}^{k-1} \sum_{\pi_l : w_{\pi_l} \leq b_l} w_{\pi_l}^2 / \ell_n \\ &= w_i [1 - F_n^*](b_k) \prod_{l=1}^k \nu_n(b_l). \end{aligned} \quad (6.2.38)$$

The same bound applies to $\text{CL}_n(\mathbf{w})$ and $\text{GRG}_n(\mathbf{w})$.

The proof of (6.2.35) is similar. Indeed, by (6.2.7),

$$\mathbb{P}(\mathcal{E}_k(i, j)) \leq \sum_{\pi \in \mathcal{GP}_k(i, j)} \frac{w_i w_j}{\ell_n} \prod_{l=1}^{k-1} w_{\pi_l}^2 / \ell_n \leq \frac{w_i w_j}{\ell_n} \prod_{l=1}^{k-1} \nu(b_l \wedge b_{k-l}). \quad (6.2.39)$$

Now follow the steps in the proof of (6.2.34). Finally, (6.2.42) follows from (1.4.2) 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 $\text{CL}_n(\mathbf{w})$ and $\text{GRG}_n(\mathbf{w})$. \square

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 $\mathbf{w} = (w_i)_{i \in [n]}$ satisfy Conditions 1.1(a)–(b) and that there exist $\tau \in (2, 3)$ and c_2 such that, for all $x \geq 1$,*

$$[1 - F_n](x) \leq c_2 x^{-(\tau-1)}. \quad (6.2.40)$$

Then, there exists a constant $c_2^ > 0$ such that, for all $x \geq 1$,*

$$[1 - F_n^*](x) \leq c_2^* x^{-(\tau-2)}, \quad (6.2.41)$$

and there exists a $c_\nu > 0$ such that for all $b \geq 1$,

$$\nu_n(b) \leq c_\nu b^{3-\tau}. \quad (6.2.42)$$

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)). \square

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),

$$\mathbb{P}(\text{dist}_{\text{NR}_n(w)}(i, j) \leq k_n) \leq \frac{1}{n} + \sum_{k=1}^{k_n} \left[\frac{2}{n} \sum_{i \in [n]} \mathbb{P}(\mathcal{F}_k(i)) + \frac{1}{n^2} \sum_{i, j \in [n]: i \neq j} \mathbb{P}(\mathcal{E}_k(i, j)) \right],$$

where the contribution $1/n$ is due to $i = j$ for which $\text{dist}_{\text{NR}_n(w)}(i, i) = 0$. We use Lemmas 6.7 and 6.8 to provide bounds on $\mathbb{P}(\mathcal{F}_k(i))$ and $\mathbb{P}(\mathcal{E}_k(i, j))$. 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 $\mathbb{P}(\mathcal{F}_k(i))$ is small, and, for this choice of $(b_l)_{l=0}^{\infty}$, we show that $\mathbb{P}(\mathcal{E}_k(i, j))$ 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 $\mathbb{P}(\mathcal{F}_k(i)) \leq [1 - F_n^*](b_k) \prod_{l=0}^{k-1} \nu_n(b_l)$ is small. Below (6.1.6), it is argued that $b_k \approx b_{k-1}^{1/(\tau-2)}$. 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

$$a = 1/(\tau - 2 - \delta) > 1. \quad (6.2.43)$$

Take $b_0 = e^A$ for some constant $A \geq 0$ sufficiently large and define $(b_l)_{l \geq 0}$ recursively by

$$b_l = b_{l-1}^a, \quad \text{so that} \quad b_l = b_0^{a^l} = e^{A(\tau-2-\delta)^{-l}}. \quad (6.2.44)$$

We start from (6.2.31). By Lemma 6.7, we obtain an upper bound on $\mathbb{P}(\mathcal{F}_k(i))$ 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

$$\begin{aligned} \prod_{l=1}^{k-1} \nu_n(b_l) &\leq \prod_{l=1}^{k-1} c_\nu b_l^{\tau-3} = c_\nu^k e^{K(3-\tau) \sum_{l=1}^{k-1} a^l} \\ &\leq c_\nu^{k-1} e^{K(3-\tau)a^k/(a-1)} = c_2^* w_i c_\nu^{k-1} b_k^{(3-\tau)/(a-1)}. \end{aligned} \quad (6.2.45)$$

Combining (6.2.45) with the bound on $[1 - F_n^*](b_k)$ in Lemma 6.8 yields

$$\mathbb{P}(\mathcal{F}_k(i)) \leq c_2^* w_i c_\nu^k b_k^{-(\tau-2)+(3-\tau)/(a-1)}. \quad (6.2.46)$$

Since $3 - \tau + \delta < 1$ when $\tau \in (2, 3)$ and $\delta \in (0, \tau - 2)$,

$$\begin{aligned} (\tau - 2) - (3 - \tau)/(a - 1) &= (\tau - 2) - (3 - \tau)(\tau - 2 - \delta)/(3 - \tau + \delta) \\ &= \delta/(3 - \tau + \delta) > \delta, \end{aligned} \quad (6.2.47)$$

so that

$$\mathbb{P}(\mathcal{F}_k(i)) \leq c_2^* w_i c_\nu^k b_k^{-\delta}. \quad (6.2.48)$$

As a result, for each $\delta > 0$

$$\begin{aligned} \frac{1}{n} \sum_{i \in [n]} \sum_{k=0}^{k_n} \mathbb{P}(\mathcal{F}_k(i)) &\leq c_2^* \frac{1}{n} \sum_{i \in [n]} w_i \mathbb{1}_{\{w_i > K\}} + \frac{1}{n} \sum_{i \in [n]} c_2^* w_i \sum_{k \geq 1} c_\nu^k b_k^{-\delta} \\ &= O(c_\nu^k b_k^{-\delta}) \leq \varepsilon, \end{aligned} \quad (6.2.49)$$

by (6.2.44) and when we take $A = A(\delta, \varepsilon)$ sufficiently large.

Similarly, since $b_l \geq 1$, by (6.2.45),

$$\mathbb{P}(\mathcal{E}_k(i, j)) \leq \frac{w_i w_j}{\ell_n} \prod_{l=1}^{k-1} \nu_n(b_l \wedge b_{k-l}) \leq \frac{w_i w_j}{\ell_n} c_\nu^{k-1} b_{\lfloor k/2 \rfloor}^{2(3-\tau)/(a-1)}, \quad (6.2.50)$$

so that, using further that $l \mapsto b_l$ is increasing,

$$\begin{aligned} \sum_{k=1}^{k_n} \frac{1}{n^2} \sum_{i, j \in [n]} \mathbb{P}(\mathcal{E}_k(i, j)) &\leq \frac{1}{n^2} \sum_{k=1}^{k_n} \sum_{i, j \in [n]} \frac{w_i w_j}{\ell_n} c_\nu^{k-1} b_{\lfloor k/2 \rfloor}^{2(3-\tau)/(a-1)} \\ &\leq \frac{\ell_n}{n^2} k_n c_\nu^{k_n-1} b_{\lfloor k_n/2 \rfloor}^{2(3-\tau)/(a-1)}, \end{aligned} \quad (6.2.51)$$

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 $(\tau - 2 - \delta)^{-(k_n+1)/2} \leq (\log n)^{1-\varepsilon/4}$. Then, by (6.2.44),

$$b_{\lfloor k_n/2 \rfloor} \leq e^{A(\tau-2-\delta)^{-(k_n+1)/2}} \leq e^{A(\log n)^{1-\varepsilon/4}}, \quad (6.2.52)$$

and we conclude that

$$\sum_{k=1}^{k_n} \frac{1}{n^2} \sum_{i, j \in [n]} \mathbb{P}(\mathcal{E}_k(i, j)) \leq \frac{\ell_n}{n^2} k_n c_\nu^{k_n} \exp(2A(3-\tau)(\log n)^{1-\varepsilon/4}) = o(1), \quad (6.2.53)$$

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 $\left(\text{dist}_{\text{NR}_n(\mathbf{w})}(o_1, o_2) \leq \frac{2 \log \log n}{|\log(\tau-2)|} \right)_-$ 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 $\text{NR}_n(\mathbf{w})$ 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 $\mathbf{w} = (w_i)_{i \in [n]}$ satisfies Conditions 1.1(a)-(b) and (6.3.1). Then, for every $\varepsilon > 0$, as $n \rightarrow \infty$,*

$$\mathbb{P}\left(\text{dist}_{\text{NR}_n(\mathbf{w})}(o_1, o_2) \leq \frac{2(1+\varepsilon) \log \log n}{|\log(\tau-2)|} \mid \text{dist}_{\text{NR}_n(\mathbf{w})}(o_1, o_2) < \infty\right) \rightarrow 1. \quad (6.3.2)$$

The same results hold for $\text{CL}_n(\mathbf{w})$ and $\text{GRG}_n(\mathbf{w})$ 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^α , 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 $\frac{(1+\varepsilon) \log \log n}{|\log(\tau-2)|}$. The latter is only true when the vertex is in the giant connected component, a fact we need to carefully take 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 $a_1 a_2$ are occupied for all $a_1, a_2 \in A$. We continue by proving that, whp, Giant_n forms a clique:

Lemma 6.10 (High-weight vertices form clique) *Under the conditions of Theorem 6.9,*

$$\mathbb{P}(\text{Giant}_n \text{ does not form clique}) \leq n^2 e^{-n^{2\alpha}/\ell_n} \quad (6.3.4)$$

The same results hold for $\text{CL}_n(\mathbf{w})$ under the same conditions, for $\text{GRG}_n(\mathbf{w})$ the diameter of 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 Giant_n , so that

$$\mathbb{P}(\text{Giant}_n \text{ does not form clique}) \leq n^2 \max_{a_1, a_2 \in \text{Giant}_n} \mathbb{P}(a_1 a_2 \text{ vacant}). \quad (6.3.5)$$

The edge $a_1 a_2$ is vacant with probability

$$\mathbb{P}(a_1 a_2 \text{ vacant}) = e^{-w_{a_1} w_{a_2} / \ell_n} \leq e^{-n^{2\alpha} / \ell_n}, \quad (6.3.6)$$

since $w_a \geq n^\alpha$ for every $a \in \text{Giant}_n$. Multiplying out gives the result. For $\text{CL}_n(\mathbf{w})$, $\mathbb{P}(a_1 a_2 \text{ vacant}) = 0$, so the same proof applies.

For $\text{GRG}_n(\mathbf{w})$, we need to strengthen this analysis slightly. Indeed, for $\text{GRG}_n(\mathbf{w})$, for all $a_1, a_2 \in \text{Giant}_n$,

$$\mathbb{P}(a_1 a_2 \text{ present}) \geq \frac{n^{2\alpha}}{\ell_n + n^{2\alpha}} = 1 - \Theta(n^{1-2\alpha}) \geq \frac{1}{2}. \quad (6.3.7)$$

Thus, the diameter of 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

$$\mathbb{P}(\text{diam}(\text{ER}_n(\frac{1}{2})) > 2) \leq n^2 \mathbb{P}(\text{dist}_{\text{ER}_n(\frac{1}{2})}(1, 2) > 2). \quad (6.3.8)$$

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

$$\mathbb{P}(\text{dist}_{\text{ER}_n(\frac{1}{2})}(1, 2) > 2) = (1 - \frac{1}{4})^{n-2}, \quad (6.3.9)$$

which, combined with (6.3.8), completes the proof. \square

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 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*

$$\mathbb{P}\left(\text{dist}_{\text{NR}_n(\mathbf{w})}(i, \text{Giant}_n) \geq (1 + \varepsilon) \frac{\log \log n}{|\log(\tau - 2)|}\right) \leq ce^{-c_1^* w_i^\eta}. \quad (6.3.10)$$

Consequently, with $\mathcal{W}_m(i) = \sum_{k \in \partial B_m(i)} w_k$ denoting the weight of vertices at graph distance m from i ,

$$\mathbb{P}\left(\text{dist}_{\text{NR}_n(\mathbf{w})}(\mathcal{N}_m(i), \text{Giant}_n) \geq (1 + \varepsilon) \frac{\log \log n}{|\log(\tau - 2)|} \mid B_m(i)\right) \leq ce^{-c_1^* \mathcal{W}_m(i)^\eta}. \quad (6.3.11)$$

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,

$$x_\ell = \max\{j \in [n]: x_{\ell-1}j \text{ occupied}\}. \quad (6.3.12)$$

Thus, x_ℓ 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),

$$\mathbb{P}(w_{x_{\ell+1}} < w_{x_\ell}^a \mid (x_s)_{s \leq \ell}) = e^{-w_{x_\ell} \sum_{t: w_t \geq w_{x_\ell}^a} w_t / \ell_n} = e^{-w_{x_\ell} [1 - F_n^*](w_{x_\ell}^a)}. \quad (6.3.13)$$

We split the argument depending on whether $w_{x_\ell}^a \leq n^\alpha$ or not. Firstly, when $w_{x_\ell}^a \leq n^\alpha$, by (6.3.1) and uniformly for $x \leq n^\alpha$,

$$[1 - F_n^*](x) \geq \frac{xn}{\ell_n}[1 - F_n](x) \geq c_1^* x^{-(\tau-2)}, \tag{6.3.14}$$

where, for n large enough, we can take $c_1^* = c_1/(2\mathbb{E}[W])$. Therefore,

$$\mathbb{P}(w_{x_{\ell+1}} < w_{x_\ell}^a \mid (x_s)_{s \leq \ell}) \leq e^{-c_1^* w_{x_\ell}^{1-(\tau-2)a}} \leq e^{-c_1^* w_{x_\ell}^\delta}, \tag{6.3.15}$$

since $a = 1/(\tau - 2 + \delta) > 1$ so that $1 - (\tau - 2)a = a\delta > \delta$.

Secondly, when $w_{x_\ell}^a > n^\alpha$, but $w_{x_\ell} < n^\alpha$, we can use (6.3.14) for $x = n^\alpha$ to obtain

$$\mathbb{P}(w_{x_{\ell+1}} < n^\alpha \mid (x_s)_{s \leq \ell}) \leq e^{-c_1^* w_{x_\ell} n^{\alpha(\tau-2)}} \leq e^{-c_1^* n^{\alpha[1-(\tau-2)]/a}} \leq e^{-c_1^* n^{\alpha\delta/a}}. \tag{6.3.16}$$

Therefore, in both cases, and with $\eta = \alpha\delta/a$,

$$\mathbb{P}(w_{x_{\ell+1}} < (n^\alpha \wedge w_{x_\ell}^a) \mid (x_s)_{s \leq \ell}) \leq e^{-c_1^* w_{x_\ell}^\eta}. \tag{6.3.17}$$

As a result, when x_ℓ is such that w_{x_ℓ} is quite large, whp, $w_{x_{\ell+1}} \geq w_{x_\ell}$. This produces, whp, a short path to 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 $w_{x_{\ell+1}} > w_{x_\ell}^a$ 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) $w_{x_\ell} \geq w_{x_0}^{\alpha^\ell} = w_i^{\alpha^\ell}$ for every $\ell \leq k - 1$,
- (2) $\text{dist}_{\text{NR}_n(\mathbf{w})}(i, \text{Giant}_n) \leq k$.

By (1), $w_{x_{k-1}} \geq w_i^{\alpha^{k-1}}$, and $w_{x_{k-1}} \in [n^{\alpha/a}, n^\alpha]$. Therefore, $w_i^{\alpha^{k-1}} \leq n^\alpha$, which, in turn, implies that

$$a^{k-1} \leq \alpha \log n, \quad \text{or} \quad k - 1 \leq (\log \log n + \log \alpha)(\log a). \tag{6.3.18}$$

Let $k_n = (1 + \varepsilon) \frac{\log \log n}{|\log(\tau-2)|}$. By (1) and (2), when $\text{dist}_{\text{NR}_n(\mathbf{w})}(i, \text{Giant}_n) > k_n$ occurs, then there must exist an $\ell \leq k_n$ such that $w_{x_{\ell+1}} \leq n^\alpha \wedge w_{x_\ell}^a$. We conclude that

$$\begin{aligned} \mathbb{P}\left(\text{dist}_{\text{NR}_n(\mathbf{w})}(i, \text{Giant}_n) \geq k_n\right) &\leq \sum_{\ell=0}^{k_n} \mathbb{P}(w_{x_{\ell+1}} \leq w_{x_\ell}^a) \\ &\leq \sum_{\ell=0}^{k_n} \mathbb{E}[\mathbb{P}(w_{x_{\ell+1}} \leq w_{x_\ell}^a \mid (x_s)_{s \leq \ell})] \\ &\leq \sum_{\ell=0}^{k_n} \mathbb{E}[e^{-c_1^* w_{x_\ell}^\eta}] \leq \sum_{\ell=0}^{k_n} e^{-c_1^* w_i^{\delta a^\ell}} \leq ce^{-c_1^* w_i^\eta}. \end{aligned} \tag{6.3.19}$$

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 $\mathcal{W}_m(i) = \sum_{k \in B_m(i)} w_k$. \square

Completion of the proof of Theorem 6.9

To prove the upper bound in Theorem 6.9, for $\varepsilon \in (0, 1)$, we take

$$k_n = (1 + \varepsilon) \frac{\log \log n}{|\log(\tau - 2)|}, \quad (6.3.20)$$

so that it suffices to show, for every $\varepsilon > 0$,

$$\lim_{n \rightarrow \infty} \mathbb{P}(\text{dist}_{\text{NR}_n(\mathbf{w})}(o_1, o_2) \leq 2k_n \mid \text{dist}_{\text{NR}_n(\mathbf{w})}(o_1, o_2) < \infty) = 1. \quad (6.3.21)$$

Since

$$\begin{aligned} & \mathbb{P}(\text{dist}_{\text{NR}_n(\mathbf{w})}(o_1, o_2) \leq 2k_n \mid \text{dist}_{\text{NR}_n(\mathbf{w})}(o_1, o_2) < \infty) \\ &= \frac{\mathbb{P}(\text{dist}_{\text{NR}_n(\mathbf{w})}(o_1, o_2) \leq 2k_n)}{\mathbb{P}(\text{dist}_{\text{NR}_n(\mathbf{w})}(o_1, o_2) < \infty)}, \end{aligned} \quad (6.3.22)$$

this follows from the two bounds

$$\liminf_{n \rightarrow \infty} \mathbb{P}(\text{dist}_{\text{NR}_n(\mathbf{w})}(o_1, o_2) < \infty) \leq \zeta^2, \quad (6.3.23)$$

$$\limsup_{n \rightarrow \infty} \mathbb{P}(\text{dist}_{\text{NR}_n(\mathbf{w})}(o_1, o_2) \leq 2k_n) \geq \zeta^2, \quad (6.3.24)$$

with $\zeta > 0$ the survival probability of the underlying branching process approximation to the neighborhoods of $\text{NR}_n(\mathbf{w})$ identified in Theorem 3.15. For (6.3.23), we split, for some $m \geq 1$,

$$\begin{aligned} & \mathbb{P}(\text{dist}_{\text{NR}_n(\mathbf{w})}(o_1, o_2) < \infty) \\ & \leq \mathbb{P}(\text{dist}_{\text{NR}_n(\mathbf{w})}(o_1, o_2) \leq 2m) \\ & \quad + \mathbb{P}(|\partial B_m(o_1)| > 0, |\partial B_m(o_2)| > 0, \text{dist}_{\text{NR}_n(\mathbf{w})}(o_1, o_2) > 2m). \end{aligned} \quad (6.3.25)$$

By Corollary 2.13, $\mathbb{P}(\text{dist}_{\text{NR}_n(\mathbf{w})}(o_1, o_2) \leq 2m) = o(1)$, and, by (2.4.2) in Corollary 2.12,

$$\begin{aligned} & \mathbb{P}(|\partial B_m(o_1)| > 0, |\partial B_m(o_2)| > 0, \text{dist}_{\text{NR}_n(\mathbf{w})}(o_1, o_2) > 2m) \\ & = \mathbb{P}(|\partial B_m(o)| > 0)^2 + o(1), \end{aligned} \quad (6.3.26)$$

where $\partial B_m(o)$ is the set of vertices at distance m in the local weak limit of $\text{NR}_n(\mathbf{w})$ identified in Theorem 3.15. The right-hand side of (6.3.26) converges to ζ^2 when $m \rightarrow \infty$. This proves (6.3.23).

To prove (6.3.24), we fix $m \geq 1$ and write

$$\begin{aligned} & \mathbb{P}(2m < \text{dist}_{\text{NR}_n(\mathbf{w})}(o_1, o_2) \leq 2k_n) \\ & \geq \mathbb{P}(\text{dist}_{\text{NR}_n(\mathbf{w})}(o_i, \text{Giant}_n) \leq k_n, i = 1, 2, \text{dist}_{\text{NR}_n(\mathbf{w})}(o_1, o_2) > 2m) \\ & \geq \mathbb{P}(|\partial B_m(o_1)| > 0, |\partial B_m(o_2)| > 0, \text{dist}_{\text{NR}_n(\mathbf{w})}(o_1, o_2) > 2m) \\ & \quad - 2\mathbb{P}(\text{dist}_{\text{NR}_n(\mathbf{w})}(o_1, \text{Giant}_n) < k_n, |\partial B_m(o_1)| > 0). \end{aligned} \quad (6.3.27)$$

By (6.3.26), the first term converges to ζ_m^2 , which in turn converges to ζ^2 when $m \rightarrow \infty$.

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

$$\begin{aligned} & \mathbb{P}(\text{dist}_{\text{NR}_n(\mathbf{w})}(o_1, \text{Giant}_n) < k_n, |B_m(o_1)| > 0) \\ &= \mathbb{E} \left[\mathbb{P}(\text{dist}_{\text{NR}_n(\mathbf{w})}(o_1, \text{Giant}_n) > k_n \mid B_m(o_1)) \mathbb{1}_{\{|\partial B_m(o_1)| > 0\}} \right]. \end{aligned} \tag{6.3.28}$$

By Proposition 6.11,

$$\mathbb{P}(\text{dist}_{\text{NR}_n(\mathbf{w})}(o_1, \text{Giant}_n) > k_n \mid B_m(o_1)) \leq ce^{-c_1^* \mathcal{W}_m(o_1)^\eta}. \tag{6.3.29}$$

By Theorem 3.15 and Conditions 1.1(a)-(b), we obtain that

$$\mathcal{W}_m(o_1)^\eta \xrightarrow{d} \sum_{i=1}^{|\partial B_m(o)|} W_i^*, \tag{6.3.30}$$

where $(W_i^*)_{i \geq 1}$ are i.i.d. copies of random variables with distribution function F^* . Therefore,

$$\mathcal{W}_m(o_1) \xrightarrow{\mathbb{P}} \infty \tag{6.3.31}$$

when first $n \rightarrow \infty$ followed by $m \rightarrow \infty$, and we use that $|\partial B_m(o)| \xrightarrow{\mathbb{P}} \infty$ since $|\partial B_m(o)| > 0$. As a result,

$$\mathbb{P}(\text{dist}_{\text{NR}_n(\mathbf{w})}(o_1, \text{Giant}_n) > k_n \mid B_m(o_1)) \mathbb{1}_{\{|\partial B_m(o_1)| > 0\}} \xrightarrow{\mathbb{P}} 0, \tag{6.3.32}$$

which by Lebesgues Dominated Convergence Theorem [Volume 1, Theorem A.1] implies that

$$\mathbb{E} \left[e^{-c_1^* \mathcal{W}_m(V_1)^\eta} \mathbb{1}_{\{|\partial B_m(o_1)| > 0\}} \right] \rightarrow 0, \tag{6.3.33}$$

when first $n \rightarrow \infty$ followed by $m \rightarrow \infty$. This proves (6.3.24), and thus completes the proof of the upper bound in Theorem 6.3. \square

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 $(1 + \varepsilon) \log_\nu n$ 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 219

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 \mathcal{I} with edge probabilities $p_{ij} = u_i u_j$, for some weights $(u_i)_{i \in \mathcal{I}}$. We obtain $\text{CL}_n(\mathbf{w})$ by taking $u_i = w_i / \sqrt{\ell_n}$ and $\mathcal{I} = [n]$. Since the $\text{NR}_n(\mathbf{w})$ random graph is closely related to $\text{CL}_n(\mathbf{w})$, this suffices for our purposes.

For $a, b \in \mathcal{I}$ and $k \geq 1$, let

$$N_k(a, b) = \#\{\pi \in \mathcal{P}_k(a, b) : \pi \text{ occupied}\} \quad (6.4.1)$$

denote the number of paths of length k between the vertices a and b . Let

$$n_k(a, b) = \mathbb{E}[N_k(a, b)] \quad (6.4.2)$$

denote the expected number of occupied paths of length k connecting a and b . Define

$$\bar{n}_k(a, b) = u_a u_b \left(\sum_{i \in \mathcal{I} \setminus \{a, b\}} u_i^2 \right)^{k-1}, \quad \underline{n}_k(a, b) = u_a u_b \left(\sum_{i \in \mathcal{I}_{a, b, k}} u_i^2 \right)^{k-1}, \quad (6.4.3)$$

where $\mathcal{I}_{a, b, k}$ is the subset of \mathcal{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)

$$n_k(a, b) \leq \bar{n}_k(a, b). \quad (6.4.4)$$

In this section, we will prove that $\underline{n}_k(a, b)$ 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

$$\nu_{\mathcal{I}} = \sum_{i \in \mathcal{I}} u_i^2, \quad \gamma_{\mathcal{I}} = \sum_{i \in \mathcal{I}} u_i^3 \quad (6.4.5)$$

denote the sums of squares and third powers of $(u_i)_{i \in \mathcal{I}}$, respectively. Our aim is to show that whp 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 $k \geq 1$, $a, b \in \mathcal{I}$ and $(u_i)_{i \in \mathcal{I}}$,*

$$\mathbb{E}[N_k(a, b)] \geq \underline{n}_k(a, b), \quad (6.4.6)$$

while, assuming that $\nu_{\mathcal{I}} > 1$,

$$\text{Var}(N_k(a, b)) \leq n_k(a, b) \quad (6.4.7)$$

$$+ \bar{n}_k(a, b)^2 \left(\frac{\gamma_{\mathcal{I}} \nu_{\mathcal{I}}^2}{\nu_{\mathcal{I}} - 1} \left(\frac{1}{u_a} + \frac{1}{u_b} \right) + \frac{\gamma_{\mathcal{I}}^2 \nu_{\mathcal{I}}}{u_a u_b (\nu_{\mathcal{I}} - 1)^2} + e_k \right),$$

where

$$e_k = \left(1 + \frac{\gamma_{\mathcal{I}}}{u_a \nu_{\mathcal{I}}}\right) \left(1 + \frac{\gamma_{\mathcal{I}}}{u_b \nu_{\mathcal{I}}}\right) \frac{\nu_{\mathcal{I}}}{\nu_{\mathcal{I}} - 1} \left(e^{2k^3 \gamma_{\mathcal{I}}^2 / \nu_{\mathcal{I}}^3} - 1\right).$$

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 $\mathbb{E}[N_k(a, b)] = n_k(a, b) \rightarrow \infty$, by taking \mathcal{I} is a large subset of $[n]$ and $u_i = w_i / \sqrt{\ell_n}$. In this case, $\nu_{\mathcal{I}} \approx \nu_n \approx \nu > 1$. In our applications of Proposition 6.12, the ratio $\bar{n}_k(a, b) / \underline{n}_k(a, b)$ will be bounded, and $k^3 \gamma_{\mathcal{I}}^2 / \nu_{\mathcal{I}}^3 = o(1)$, 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, $\gamma_{\mathcal{I}} / u_a$ and $\gamma_{\mathcal{I}} / u_b$ are typically small, so that

$$\frac{\text{Var}(N_k(a, b))}{\mathbb{E}[N_k(a, b)]^2} \approx \frac{\gamma_{\mathcal{I}} \nu_{\mathcal{I}}^2}{\nu_{\mathcal{I}} - 1} \left(\frac{1}{u_a} + \frac{1}{u_b}\right) + \frac{\gamma_{\mathcal{I}}^2 \nu_{\mathcal{I}}}{u_a u_b (\nu_{\mathcal{I}} - 1)^2} \quad (6.4.8)$$

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

$$N_k(a, b) = \sum_{\pi \in \mathcal{P}_k(a, b)} \mathbb{1}_{\{\pi \text{ occupied}\}}. \quad (6.4.9)$$

As a result,

$$\begin{aligned} \mathbb{E}[N_k(a, b)] &= \sum_{\pi \in \mathcal{P}_k(a, b)} \mathbb{P}(\pi \text{ occupied}) = \sum_{\pi \in \mathcal{P}_k(a, b)} \prod_{l=0}^{k-1} u_{\pi_l} u_{\pi_{l+1}} \\ &= u_{\pi_0} u_{\pi_k} \sum_{\pi \in \mathcal{P}_k(a, b)} \prod_{l=1}^{k-1} u_{\pi_l}^2. \end{aligned} \quad (6.4.10)$$

For $\pi \in \mathcal{P}_k(a, b)$, $\pi_0 = a, \pi_k = b$. Further,

$$\sum_{\pi \in \mathcal{P}_k(a, b)} \prod_{l=1}^{k-1} u_{\pi_l}^2 = \sum_{i_1, \dots, i_{k-1} \in \mathcal{I} \setminus \{a, b\}}^* \prod_{l=1}^{k-1} u_{\pi_l}^2, \quad (6.4.11)$$

where we recall that $\sum_{i_1, \dots, i_r \in \mathcal{I}}^*$ denotes a sum over *distinct* indices. Each sum over i_j yields a factor that is at least $\sum_{i \in \mathcal{I}_{a, b, k}} u_i^2$, which proves (6.4.6).

To compute $\text{Var}(N_k(a, b))$, we again start from (6.4.9), which yields

$$\begin{aligned} \text{Var}(N_k(a, b)) &= \sum_{\pi, \rho \in \mathcal{P}_k(a, b)} \left[\mathbb{P}(\pi, \rho \text{ occupied}) \right. \\ &\quad \left. - \mathbb{P}(\pi \text{ occupied}) \mathbb{P}(\rho \text{ occupied}) \right]. \end{aligned} \quad (6.4.12)$$

6.4 Path counting and log upper bound for finite-variance weights 221

For π, ρ , we denote by $\pi \cap \rho$ the edges the paths π and ρ have in common. The occupation statuses of π and ρ are independent precisely when $\pi \cap \rho = \emptyset$, so that

$$\text{Var}(N_k(a, b)) \leq \sum_{\substack{\pi, \rho \in \mathcal{P}_k(a, b) \\ \pi \cap \rho \neq \emptyset}} \mathbb{P}(\pi, \rho \text{ occupied}). \quad (6.4.13)$$

Define $\rho \setminus \pi$ to be the edges in ρ that are *not* part of π , so that

$$\begin{aligned} \mathbb{P}(\pi, \rho \text{ occupied}) &= \mathbb{P}(\pi \text{ occupied})\mathbb{P}(\rho \text{ occupied} \mid \pi \text{ occupied}) \\ &= \prod_{l=0}^k u_{\pi_l} u_{\pi_{l+1}} \prod_{e \in \rho \setminus \pi} u_{\bar{e}} u_{\underline{e}}, \end{aligned} \quad (6.4.14)$$

where, for an edge $e = \{x, y\}$, we write $\bar{e} = x, \underline{e} = y$. When $\pi = \rho$, then

$$\mathbb{P}(\pi, \rho \text{ occupied}) = \mathbb{P}(\pi \text{ occupied}), \quad (6.4.15)$$

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 (π, ρ) 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 π and ρ 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 ρ that are not part of π .

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 π . We can view these subpaths as excursions of the path ρ from the walk π . By construction, between two excursions, there is at least one edge that π and ρ have in common.

Fix m . We define $\text{Shape}(\pi, \rho)$, the *shape* of the pair (π, ρ) , by

$$\text{Shape}(\pi, \rho) = (\vec{x}_{m+1}, \vec{s}_m, \vec{t}_m, \vec{o}_{m+1}, \vec{r}_{m+1}), \quad (6.4.16)$$

where

- (1) $\vec{x}_{m+1} \in \mathbb{N}_0^{m+1}$ and $x_j \geq 0$ is the length of the subpath in $\rho \cap \pi$ in between the $(j - 1)^{\text{st}}$ 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 $j \in \{2, \dots, m\}$, $x_j \geq 1$;
- (2) $\vec{s}_m \in \mathbb{N}^m$ and $s_j \geq 1$ is the number of edges in the j^{th} subpath of $\pi \setminus \rho$;
- (3) $\vec{t}_m \in \mathbb{N}^m$ and $t_j \geq 1$ is the number of edges in the j^{th} subpath of $\rho \setminus \pi$;
- (4) $\vec{o}_{m+1} \in [m + 1]^{m+1}$ and o_j is the order of the j^{th} common subpath in $\rho \cap \pi$ of the path π in ρ , i.e., $o_2 = 5$ means that the second subpath that π has in common with ρ is the 5th subpath that ρ has in common with π . Note that $o_1 = 1$ and $o_{m+1} = m + 1$, since π and ρ start and end in a and b , respectively;

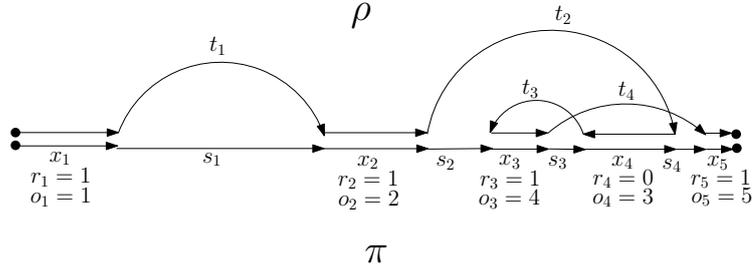


Figure 6.3 An example of a pair of paths (π, ρ) and its corresponding shape.

- (5) $\vec{r}_{m+1} \in \{0, 1\}^{m+1}$ is such that r_j describes the direction in which the j^{th} common subpath in $\rho \cap \pi$ of the path π is traversed by ρ , with $r_j = 1$ when the direction is the same for π and ρ 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 π and ρ . See Figure 6.3 for an example of a pair of paths (π, ρ) 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 (π, ρ) with m excursions and l common edges, so that (6.4.17) hold. Then,

$$\text{Var}(N_k(a, b)) \leq n_k(a, b) + \sum_{l=1}^{k-2} \sum_{m=1}^{k-l} \sum_{\sigma \in \text{Shape}_{m,l}} \sum_{\substack{\pi, \rho \in \mathcal{P}_k(a, b) \\ \text{Shape}(\pi, \rho) = \sigma}} \mathbb{P}(\pi, \rho \text{ occupied}). \tag{6.4.18}$$

When $\text{Shape}(\pi, \rho) = \sigma$ for some $\sigma \in \text{Shape}_{m,l}$, and since π and ρ 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 π , 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 ρ that are not part of π are part of precisely 2 edges. By construction, the $k + 1$ vertices of both π and ρ are disjoint, but the remaining $k - l - m$ vertices in ρ may intersect those of π . Therefore, denoting $a_1 = \delta_{x_1,0}$, $a_{m+1} = \delta_{x_{m+1},0}$,

$$\mathbb{P}(\pi, \rho \text{ occupied}) = u_a^{1+a_1} u_b^{1+a_{m+1}} \prod_{s=1}^{2m-a_1-a_{m+1}} u_{v_s}^3 \prod_{t=2m-a_1-a_{m+1}+1}^{2(k-1)-l-m} u_{v_t}^2, \tag{6.4.19}$$

where $(v_1, \dots, v_{k+1+l-m}) \in \mathcal{I}^{k-1+l-m}$.

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 223

$\text{Shape}(\pi, \rho) = \sigma$ from above by summing (6.4.19) over *all* $(v_1, \dots, v_{k-1+l-m}) \in \mathcal{I}^{k-1+l-m}$, to obtain for any $\sigma \in \text{Shape}_{m,l}$,

$$\begin{aligned} & \sum_{\substack{\pi, \rho \in \mathcal{P}_k(a, b) \\ \text{Shape}(\pi, \rho) = \sigma}} \mathbb{P}(\pi, \rho \text{ occupied}) \\ & \leq u_a u_b \gamma_{\mathcal{I}}^{2m} \nu_{\mathcal{I}}^{2k-1-3m-l} \left(\frac{u_a \nu_{\mathcal{I}}}{\gamma_{\mathcal{I}}} \right)^{\delta_{x_1,0}} \left(\frac{u_b \nu_{\mathcal{I}}}{\gamma_{\mathcal{I}}} \right)^{\delta_{x_{m+1},0}} \\ & = \bar{n}_k(a, b)^2 \gamma_{\mathcal{I}}^{2(m+1)} \nu_{\mathcal{I}}^{-3(m-1)-l} \left(\frac{\gamma_{\mathcal{I}}}{u_a \nu_{\mathcal{I}}} \right)^{1-\delta_{x_1,0}} \left(\frac{\gamma_{\mathcal{I}}}{u_b \nu_{\mathcal{I}}} \right)^{1-\delta_{x_{m+1},0}}. \end{aligned} \quad (6.4.20)$$

Therefore, we arrive at

$$\begin{aligned} \text{Var}(N_k(a, b)) & \leq n_k(a, b) + \bar{n}_k(a, b)^2 \sum_{l=1}^{k-2} \sum_{m=1}^{k-l} \gamma_{\mathcal{I}}^{2(m-1)} \nu_{\mathcal{I}}^{-3(m-1)-l} \\ & \quad \times \sum_{\sigma \in \text{Shape}_{m,l}} \left(\frac{\gamma_{\mathcal{I}}}{u_a \nu_{\mathcal{I}}} \right)^{1-\delta_{x_1,0}} \left(\frac{\gamma_{\mathcal{I}}}{u_b \nu_{\mathcal{I}}} \right)^{1-\delta_{x_{m+1},0}}. \end{aligned} \quad (6.4.21)$$

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*

$$2^{m-1}(m-1)! \binom{k-l-1}{m-1}^2 \binom{l}{m-a_1-a_{m+1}}. \quad (6.4.22)$$

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 $\vec{\sigma}_{m+1}$ of the common parts when we have fixed the directions the paths can be traversed.

In counting the number of $\vec{x}_{m+1}, \vec{s}_m, \vec{t}_m$, we repeatedly use the fact that there are $\binom{a-1}{b-1}$ possible sequences $(y_1, \dots, y_b) \in \mathbb{N}_0^b$ such that $\sum_{j=1}^b y_j = a$. This can be seen by representing a as a sequence of a ones, separated by $a-1$ zeros. We draw b zeros, which we can do in $\binom{a-1}{b-1}$ possible ways. Then, we note that a sequence $(y_1, \dots, y_b) \in \mathbb{N}_0^b$ such that $\sum_{j=1}^b y_j = a$ can be obtained uniquely by letting y_i be the number of ones in between the $(i-1)^{\text{st}}$ and i^{th} chosen zero. Similarly, there are $\binom{a+b-1}{b-1}$ possible sequences $(y_1, \dots, y_b) \in \mathbb{N}^b$ such that $\sum_{j=1}^b y_j = a$, since we can apply the previous equality to $(y_1+1, \dots, y_b+1) \in \mathbb{N}^b$.

Using the above, we continue to count the number of shapes. The number of $(s_1, \dots, s_m) \in \mathbb{N}^m$ such that $s_j \geq 1$ and $\sum_{j=1}^m s_j = k-l$ equals

$$\binom{k-l-1}{m-1}. \quad (6.4.23)$$

The same applies to $(t_1, \dots, t_m) \in \mathbb{N}^m$ such that $t_j \geq 1$ and $\sum_{j=1}^m t_j = l$. 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 $x_1 = x_2 = 0$, 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

$$\binom{l}{m - a_1 - a_{m+1}} \quad (6.4.24)$$

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} . \square

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

$$\begin{aligned} & 2^{m-1}(m-1)! \binom{k-l-1}{m-1}^2 \binom{l}{m-a_1-a_{m+1}} \\ & \times \left(\frac{2\gamma_x^2}{\nu_x^3}\right)^{m-1} \nu_x^{-l} \left(\frac{\gamma_x}{u_a \nu_x}\right)^{1-a_1} \left(\frac{\gamma_x}{u_b \nu_x}\right)^{1-a_{m+1}} \end{aligned} \quad (6.4.25)$$

over $l \in [k-2]$, $m \in [k-l]$ and $a_1, a_{m+1} \in \{0, 1\}$, where, by convention, $\binom{l}{-1} = 0$.

We start with $m = 1$, for which we obtain that the sum of (6.4.25) over the other variables equals

$$\begin{aligned} & \gamma_x \left(\frac{1}{u_a} + \frac{1}{u_b}\right) \sum_{l=1}^{\infty} \nu_x^{-(l-1)} + \frac{\gamma_x^2}{u_a u_b \nu_x} \sum_{l=1}^{\infty} l \nu_x^{-(l-1)} \\ & = \frac{\gamma_x \nu_x^2}{\nu_x - 1} \left(\frac{1}{u_a} + \frac{1}{u_b}\right) + \frac{\gamma_x^2 \nu_x}{u_a u_b (\nu_x - 1)^2}, \end{aligned} \quad (6.4.26)$$

where we use that, for $a \in [0, 1)$,

$$\sum_{l=0}^{\infty} a^{-l} = a/(1-a), \quad \sum_{l=0}^{\infty} l a^{-(l-1)} = a^2/(1-a)^2. \quad (6.4.27)$$

The terms in (6.4.26) are the first two terms appearing on the right-hand side of (6.4.7).

This leaves us to bound the contribution when $m \geq 2$. We continue by bounding

$$\binom{k-l-1}{m-1} (m-1)! = \frac{1}{(m-1)!} \left(\frac{(k-l-1)!}{(k-l-m)!}\right)^2 \leq \frac{k^{2(m-1)}}{(m-1)!}, \quad (6.4.28)$$

and, using that $\binom{a}{b} \leq a^b/b!$ and $l \leq k$,

$$\binom{l}{m-a_1-a_{m+1}} \leq \frac{l^{m-a_1-a_{m+1}}}{(m-a_1-a_{m+1})!} \leq k^m. \quad (6.4.29)$$

6.4 Path counting and log upper bound for finite-variance weights 225

Therefore, the number of shapes in $\text{Shape}_{m,l}$ is, for each $l \geq 1$ and $m \geq 2$, bounded by

$$2^{m-1} \frac{k^{2(m-1)}}{(m-1)!} k^m = k \frac{(2k^3)^{m-1}}{(m-1)!}. \quad (6.4.30)$$

Since the above is independent of l , we can start by summing (6.4.25) over $l \geq 1$, and over $a_1, a_{m+1} \in \{0, 1\}$ to obtain a bound of the form

$$\begin{aligned} & k \left(1 + \frac{\gamma_{\mathcal{I}}}{u_a \nu_{\mathcal{I}}}\right) \left(1 + \frac{\gamma_{\mathcal{I}}}{u_b \nu_{\mathcal{I}}}\right) \frac{\nu_{\mathcal{I}}}{\nu_{\mathcal{I}} - 1} \sum_{m \geq 2} \frac{(2k^3)^{m-1}}{(m-1)!} \left(\frac{\gamma_{\mathcal{I}}^2}{\nu_{\mathcal{I}}^3}\right)^{m-1} \\ & = k \left(1 + \frac{\gamma_{\mathcal{I}}}{u_a \nu_{\mathcal{I}}}\right) \left(1 + \frac{\gamma_{\mathcal{I}}}{u_b \nu_{\mathcal{I}}}\right) \frac{\nu_{\mathcal{I}}}{\nu_{\mathcal{I}} - 1} \left(e^{2k^3 \gamma_{\mathcal{I}}^2 / \nu_{\mathcal{I}}^3} - 1\right). \end{aligned} \quad (6.4.31)$$

The term in (6.4.31) is the last term appearing on the right-hand side of (6.4.7). Summing the bounds in (6.4.26) and (6.4.31) proves (6.4.7). \square

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 $\text{NR}_n(\mathbf{w})$) *Assume that Conditions 1.1(a)–(c) hold, where $\nu = \mathbb{E}[W^2]/\mathbb{E}[W] \in (1, \infty)$. Then, for any $\varepsilon > 0$,*

$$\mathbb{P}(\text{dist}_{\text{NR}_n(\mathbf{w})}(o_1, o_2) \leq (1 + \varepsilon) \log_{\nu} n \mid \text{dist}_{\text{NR}_n(\mathbf{w})}(o_1, o_2) < \infty) = 1 + o(1). \quad (6.4.32)$$

The same results hold for $\text{CL}_n(\mathbf{w})$ and $\text{GRG}_n(\mathbf{w})$ 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 ζ_m^2 , where $\zeta_m = \mathbb{P}(\partial B_m(o) \neq \emptyset)$ is the probability that the branching process that is the local weak limit of $\text{NR}_n(\mathbf{w})$ survives to generation m . See also Corollary 2.12 that proves the asymptotic independence of the neighborhoods of o_1 and o_2 . Then, $\zeta_m \rightarrow \zeta$ when $m \rightarrow \infty$, and, conditionally on $|\partial B_m(o)| > 0$, $|\partial B_m(o)| \geq M$ whp, for any M and as $m \rightarrow \infty$. This explains the branching-process approximation.

We now state the precise branching-process approximation result that we will rely upon. We take $u_i = w_i/\sqrt{\ell_n}$,

$$a = \partial B_m(o_1), \quad b = \partial B_m(o_2), \quad (6.4.33)$$

so that

$$\begin{aligned} u_a &= \frac{1}{\sqrt{\ell_n}} \sum_{i \in \partial B_m(o_1)} w_i = \mathcal{W}_m(o_1)/\sqrt{\ell_n}, \\ u_b &= \frac{1}{\sqrt{\ell_n}} \sum_{i \in \partial B_m(o_2)} w_i = \mathcal{W}_m(o_2)/\sqrt{\ell_n}. \end{aligned} \quad (6.4.34)$$

We formalize the above ideas in the following lemma:

Lemma 6.15 (Branching process approximation) *As $n \rightarrow \infty$,*

$$(\mathcal{W}_m(o_1), \mathcal{W}_m(o_2)) \xrightarrow{d} \left(\sum_{j=1}^{Z_m^{(1)}} W^{*(1)}(j), \sum_{j=1}^{Z_m^{(2)}} W^{*(2)}(j) \right), \quad (6.4.35)$$

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 2.12, $|\partial B_m(o_1)|$ 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 $\text{NR}_n(\mathbf{w})$ 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, $\mathcal{W}_m(o_i) = \sum_{j=1}^{|\partial B_m(o_i)|} W_n^*(j)$, where $(W_n^*(j))_{j \geq 1}$ are i.i.d. copies of W_n^* . By Condition 1.1(a), $W_n^* \xrightarrow{d} W^*$, so that $\mathcal{W}_m(o_i) \xrightarrow{d} \sum_{j=1}^{|\partial B_m(o_i)|} W^{*(i)}(j)$. The joint convergence follows in a similar fashion, now using local weak convergence in probability. \square

Second moment method and path counting

Fix $k = k_n = (1 + \varepsilon) \log_\nu n - 2m$. 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 $\partial B_m(o_2) \neq \emptyset$, there exist a path of length $k_n - 2m$ connecting $\partial B_m(o_1)$ 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 $\text{dist}_{\text{NR}_n(\mathbf{w})}(o_1, o_2) \leq k_n - 2m$ occurs whp. For this, we take $u_i = w_i/\sqrt{\ell_n}$. 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

$$\mathcal{I}_{a,b} = \{i \in [n] : w_i \leq K\} \setminus (B_m(o_1) \cup B_m(o_2)). \quad (6.4.36)$$

In order to apply Proposition 6.12, we start by investigating the constants appearing in it:

6.4 Path counting and log upper bound for finite-variance weights 227

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 $k = (1 + \varepsilon) \log_\nu n$,*

$$n_k(a, b) \xrightarrow{\mathbb{P}} \infty, \quad \bar{n}_k(a, b) = (1 + o_{\mathbb{P}}(1)) \underline{n}_k(a, b), \quad (6.4.37)$$

and, as $n \rightarrow \infty$,

$$\frac{\text{Var}(N_k(a, b))}{\mathbb{E}[N_k(a, b)]^2} \leq \frac{K\nu^2}{\nu - 1} \left(\frac{1}{\sqrt{\ell_n u_a}} + \frac{1}{\sqrt{\ell_n u_b}} \right) + \frac{K^2\nu^2}{(\nu - 1)\ell_n u_a u_b} + o_{\mathbb{P}}(1). \quad (6.4.38)$$

Proof By (6.4.3),

$$\underline{n}_k(a, b) = u_a u_b \nu_{\mathcal{I}_{a,b}}^{k-1}, \quad (6.4.39)$$

and

$$\frac{\bar{n}_k(a, b)}{\underline{n}_k(a, b)} = (\nu_{\mathcal{I}_{a,b}} / \nu_{\mathcal{I}_{a,b,k}})^{k-1}. \quad (6.4.40)$$

We start by investigating $\nu_{\mathcal{I}}$. Denote

$$\nu(K) = \frac{\mathbb{E}[W^2 \mathbb{1}_{\{W \leq K\}}]}{\mathbb{E}[W]}. \quad (6.4.41)$$

Then, by (6.4.36) and the fact that $B_m(o_1)$ and $B_m(o_2)$ contain a finite number of vertices,

$$\lim_{n \rightarrow \infty} \nu_{\mathcal{I}_{a,b}} = \nu(K). \quad (6.4.42)$$

The same applies to $\nu_{\mathcal{I}_{a,b,k}}$. Then, with $K > 0$ chosen so large that $\nu(K) \geq \nu - \varepsilon/2$ and with $k = (1 + \varepsilon) \log_\nu n$,

$$\underline{n}_k(a, b) = u_a u_b \nu_{\mathcal{I}_{a,b}}^{k-1} = \frac{\mathcal{W}_m(o_1) \mathcal{W}_m(o_2)}{\ell_n} n^{(1+\varepsilon) \log \nu_{\mathcal{I}_{a,b}} / \log \nu - 1} \xrightarrow{\mathbb{P}} \infty, \quad (6.4.43)$$

where K and n are so large that $(1 + \varepsilon)\nu(K)/\nu > 1$. This proves the first property in (6.4.37).

To prove the second property in (6.4.37), we note that the set $\mathcal{I}_{a,b,k}$ is obtained from $\mathcal{I}_{a,b}$ by removing the k vertices with highest weight. Since $w_i \leq K$ for all $i \in \mathcal{I}$ (recall (6.4.36)), $\nu_{\mathcal{I}_{a,b}} \leq \nu_{\mathcal{I}_{a,b,k}} + kK/\ell_n$. Since $k \leq A \log n$, we therefore arrive at

$$\frac{\bar{n}_k(a, b)}{\underline{n}_k(a, b)} \leq (1 + kK/(\ell_n \nu_{\mathcal{I}_{a,b,k}}))^{k-1} = e^{k^2 K / (\ell_n \nu_{\mathcal{I}_{a,b,k}})} \xrightarrow{\mathbb{P}} 1, \quad (6.4.44)$$

as required.

To prove (6.4.38), we rely on Proposition 6.12. We have already shown that $n_k(a, b) = \mathbb{E}[N_k(a, b)] \rightarrow \infty$, so that the first term on the right-hand side of (6.4.7) is $o_{\mathbb{P}}(\mathbb{E}[N_k(a, b)]^2)$. Further, by (6.4.36),

$$\gamma_{\mathcal{I}} \leq \nu_{\mathcal{I}} (\max_{i \in \mathcal{I}} u_i) \leq \frac{\nu_{\mathcal{I}} K}{\sqrt{\ell_n}}, \quad (6.4.45)$$

so that, for $k \leq A \log n$ with $A > 1$ fixed,

$$\left(1 + \frac{\gamma_{\mathcal{I}}}{u_a \nu_{\mathcal{I}}}\right) \left(1 + \frac{\gamma_{\mathcal{I}}}{u_b \nu_{\mathcal{I}}}\right) k \left(e^{2k^3 \gamma_{\mathcal{I}}^2 / \nu_{\mathcal{I}}^3} - 1\right) = o_{\mathbb{P}}(1). \tag{6.4.46}$$

Substituting these bounds into (6.4.38) and using (6.4.37) 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

$$\mathbb{P}(k_n < \text{dist}_{\text{NR}_n(w)}(o_1, o_2) < \infty) = o(1). \tag{6.4.47}$$

Indeed, then $\mathbb{P}(\text{dist}_{\text{NR}_n(w)}(o_1, o_2) > k_n \mid \text{dist}_{\text{NR}_n(w)}(o_1, o_2) < \infty) = o(1)$ since, $\mathbb{P}(\text{dist}_{\text{NR}_n(w)}(o_1, o_2) < \infty) \rightarrow \zeta^2 > 0$ by Theorem 3.17. We rewrite

$$\begin{aligned} & \mathbb{P}(k_n < \text{dist}_{\text{NR}_n(w)}(o_1, o_2) < \infty) && (6.4.48) \\ &= \mathbb{P}(k_n < \text{dist}_{\text{NR}_n(w)}(o_1, o_2) < \infty, \partial B_m(o_1) \neq \emptyset, \partial B_m(o_2) \neq \emptyset) \\ &\leq \mathbb{P}(N_{k_n-2m}(\partial B_m(o_1), \partial B_m(o_2)) = 0, \partial B_m(o_1) \neq \emptyset, \partial B_m(o_2) \neq \emptyset). \end{aligned}$$

Recall that $k = k_n = (1 + \varepsilon) \log_{\nu} n$. By the Chebychev inequality [Volume 1, Theorem 2.18], the conditional probability of $\{\text{dist}_{\text{NR}_n(w)}(o_1, o_2) > k_n\}$ given $B_m(o_1), B_m(o_2)$ is at most

$$\frac{\text{Var}(N_{k-2m}(a, b))}{\mathbb{E}[N_{k-2m}(a, b)]^2} \leq \frac{K \nu^2}{\nu - 1} \left(\frac{1}{\sqrt{\ell_n u_a}} + \frac{1}{\sqrt{\ell_n u_b}} \right) + \frac{K^2 \nu^2}{(\nu - 1) \ell_n u_a u_b} + o_{\mathbb{P}}(1). \tag{6.4.49}$$

When $\partial B_m(o_1) \neq \emptyset$ and $\partial B_m(o_2) \neq \emptyset$, by (6.4.34),

$$\frac{1}{\sqrt{\ell_n u_a}} + \frac{1}{\sqrt{\ell_n u_b}} \xrightarrow{\mathbb{P}} \left(\sum_{j=1}^{Z_m^{(1)}} W^{*(1)}(j) \right)^{-1} + \left(\sum_{j=1}^{Z_m^{(2)}} W^{*(2)}(j) \right)^{-1} \xrightarrow{\mathbb{P}} 0, \tag{6.4.50}$$

when $m \rightarrow \infty$. Therefore,

$$\mathbb{P}\left(N_{k-2m}(a, b) = 0 \mid \partial B_m(o_1) \neq \emptyset, \partial B_m(o_2) \neq \emptyset\right) \xrightarrow{\mathbb{P}} 0, \tag{6.4.51}$$

and, by Lebesgues' Dominated Convergence Theorem [Volume 1, Theorem A.1],

$$\mathbb{P}(\text{dist}_{\text{NR}_n(w)}(o_1, o_2) > k_n, \partial B_m(o_1) \neq \emptyset, \partial B_m(o_2) \neq \emptyset) \rightarrow 0, \tag{6.4.52}$$

when first $n \rightarrow \infty$ followed by $m \rightarrow \infty$, which completes the proof. \square

We close this section by describing what happens when $\tau = 3$, and there are no slowly-varying functions.

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:

6.4 Path counting and log upper bound for finite-variance weights 229

Theorem 6.17 (Graph distances $\text{NR}_n(\mathbf{w})$ 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 $x \leq n^\alpha$,*

$$[1 - F_n](x) \geq c_1/x^2, \quad (6.4.53)$$

and for all $x \geq 0$,

$$[1 - F_n](x) \leq c_2/x^2. \quad (6.4.54)$$

Then, conditionally on $\text{dist}_{\text{NR}_n(\mathbf{w})}(o_1, o_2) < \infty$,

$$\frac{\text{dist}_{\text{NR}_n(\mathbf{w})}(o_1, o_2) \log \log n}{\log n} \xrightarrow{\mathbb{P}} 1. \quad (6.4.55)$$

The same results hold for $\text{CL}_n(\mathbf{w})$ and $\text{GRG}_n(\mathbf{w})$ 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

$$\alpha_n = e^{\nu_n^{1-\eta}}. \quad (6.4.56)$$

Define the *core* of $\text{NR}_n(\mathbf{w})$ to be

$$\text{Core}_n = \{i: w_i \geq \alpha_n\}. \quad (6.4.57)$$

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.,*

$$\mathbb{P}(V'_i = j) = \frac{w_j}{\sum_{v \in \text{Core}_n} w_v}, \quad (6.4.58)$$

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

$$\mathbb{P}(H'_n \leq \frac{(1 + \varepsilon) \log n}{\log \log n}) \rightarrow 1. \quad (6.4.59)$$

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 $\eta > 0$,*

$$\mathbb{P}(d_{\text{NR}_n(\mathbf{w})}(o_1, \text{Core}_n) \leq \nu_n^{1-\eta}, d_{\text{NR}_n(\mathbf{w})}(o_2, \text{Core}_n) \leq \nu_n^{1-\eta}) \rightarrow \zeta^2. \quad (6.4.60)$$

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

$$\begin{aligned} d_{\text{NR}_n(\mathbf{w})}(o_1, o_2) &\leq d_{\text{NR}_n(\mathbf{w})}(o_1, \text{Core}_n) + d_{\text{NR}_n(\mathbf{w})}(o_2, \text{Core}_n) \\ &\quad + d_{\text{NR}_n(\mathbf{w})}(V'_1, V'_2), \end{aligned} \quad (6.4.61)$$

where $V'_1, V'_2 \in \text{Core}_n$ 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, $V'_1, V'_2 \in \text{Core}_n$ 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 $\nu_n^{1-\eta} \leq \varepsilon k_n / 4$,

$$\begin{aligned} \mathbb{P}(d_{\text{NR}_n(\mathbf{w})}(o_1, o_2) \leq (1 + \varepsilon)k_n) & \quad (6.4.62) \\ & \geq \mathbb{P}(d_{\text{NR}_n(\mathbf{w})}(o_i, \text{Core}_n) \leq \nu_n^{1-\eta}, i = 1, 2) \\ & \quad \times \mathbb{P}(d_{\text{NR}_n(\mathbf{w})}(V'_1, V'_2) \leq (1 + \varepsilon/2)k_n \mid d_{\text{NR}_n(\mathbf{w})}(o_i, \text{Core}_n) \leq \nu_n^{1-\eta}, i = 1, 2). \end{aligned}$$

By Proposition 6.19, the first probability converges to ζ^2 , and by Proposition 6.18, the second probability converges to 1. We conclude that

$$\mathbb{P}(d_{\text{NR}_n(\mathbf{w})}(o_1, o_2) \leq (1 + \varepsilon) \frac{\log n}{\log \log n}) \rightarrow \zeta^2.$$

Since also $\mathbb{P}(d_{\text{NR}_n(\mathbf{w})}(o_1, o_2) < \infty) \rightarrow \zeta^2$, this completes the proof. \square

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

$$a = V'_1, \quad b = V'_2, \quad \mathcal{I} = \{w_i : w_i \in [K, \sqrt{\alpha_n}]\}. \quad (6.4.63)$$

The whole point is that there exists a constant $c > 0$ such that

$$\nu_x \geq c \log \alpha_n = c \nu_n^{1-\eta}, \quad (6.4.64)$$

while $u_a \geq \alpha_n / \sqrt{\ell_n}$, $u_b \geq \alpha_n / \sqrt{\ell_n}$, so that

$$\mathbb{E}[N_k(a, b)] \approx \alpha_n^2 c^k \nu_n^{k(1-\eta)} / \ell_n \rightarrow \infty \quad (6.4.65)$$

for $k = \log n / ((1 - \eta) \log \nu_n) \leq (1 + \varepsilon/2) \log n / \log \nu_n$ when η is such that $1/(1 - \eta) \leq 1 + \varepsilon/2$. Further,

$$\gamma_x \leq \sqrt{\alpha_n} / \sqrt{\ell_n}, \quad (6.4.66)$$

so that $\text{Var}(N_k(a, b)) / \mathbb{E}[N_k(a, b)]^2 \rightarrow 0$ by Proposition 6.12. Since $\nu_n = \Theta(\log n)$ when $\tau = 3$ (see (6.4.53)–(6.4.54) in Theorem 6.17), this completes the proof. \square

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 ζ^2 when first $n \rightarrow \infty$ followed by $m \rightarrow \infty$. Then, we perform a second moment method on the number of paths between $\partial B_m(o_i)$ and Core_n . For this, we take $k = \nu_n^{1-\eta}$ and

$$a = \partial B_m(o_1), \quad b = \text{Core}_n, \quad \mathcal{I} = \{i : w_i \leq K\} \setminus (B_m(o_1) \cup B_m(o_2)). \quad (6.4.67)$$

Then we follow the proof in (6.4.49)–(6.4.52) to show that

$$\mathbb{P}(d_{\text{NR}_n(\mathbf{w})}(o_1, \text{Core}_n) > \nu_n^{1-\eta}, \partial B_m(o_1) \neq \emptyset, \partial B_m(o_2) \neq \emptyset) \rightarrow 0, \quad (6.4.68)$$

6.4 Path counting and log upper bound for finite-variance weights 231

as required. Note, for this, that, conditionally on $B_m(o_1), B_m(o_2)$

$$\mathbb{E}[N_k(a, b)] \approx \nu(K)^k \mathcal{W}_m(o_1) \frac{1}{\ell_n} \sum_{i \in \text{Core}_n} w_i, \quad (6.4.69)$$

where $\nu(K) \rightarrow \infty$ as $K \rightarrow \infty$, and where, by (6.4.53),

$$\frac{1}{n} \sum_{i \in \text{Core}_n} w_i \geq \alpha_n [1 - F_n](\alpha_n) \geq c/\alpha_n. \quad (6.4.70)$$

Therefore, $\mathbb{E}[N_k(a, b)] \rightarrow \infty$ as soon as $k \geq 2 \log \alpha_n / \log \nu(K)$, which is satisfied for K sufficiently large and $k = \nu_n^{1-\eta}$. Exercise 6.22 asks you to complete the above proof. \square

6.4.3 DISTANCES FOR $\text{IRG}_n(\kappa)$: 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 (κ_n) is graphical sequence of kernels with limit irreducible κ , and with $\nu = \|\mathbf{T}_\kappa\| \in (1, \infty)$. 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 κ is bounded, i.e., $\sup_{x,y} \kappa(x, y) < \infty$. Indeed, we can always stochastically dominate v with an unbounded κ_n from below by $\text{IRG}_n(\kappa_n)$ with a bounded kernel that approximates it arbitrarily well. Since graph distances increase by decreasing κ_n , if we prove Theorem 6.1(ii) in the bounded case, the unbounded case will follow immediately. Further, we can approximate a bounded κ_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 $i, j \in [k]$,

$$\lim_{n \rightarrow \infty} \kappa_{n; i, j} = \kappa_{i, j}, \quad (6.4.71)$$

and

$$\lim_{n \rightarrow \infty} \mu_n(i) = \frac{1}{n} \#\{v \in [n] : i_v = i\} = \mu(i). \quad (6.4.72)$$

In this case, $\|\mathbf{T}_{\kappa_n}\|$ is the largest eigenvalue of the matrix $\mathbf{M}_{i,j}^{(n)} = \kappa_{n; i, j} \mu_n(j)$, which converges to the largest eigenvalue of the matrix $\mathbf{M}_{i,j} = \kappa_{n; i, j} \mu(j)$, which equals $\nu = \|\mathbf{T}_\kappa\| \in (1, \infty)$, 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

$$\mathbb{P}(d_{\text{IRG}_n(\kappa_n)}(o_1, o_2) \leq (1 + \varepsilon) \log_\nu n \mid B_m(o_1), B_m(o_2)) = 1 + o_\mathbb{P}(1). \quad (6.4.73)$$

We follow the proof of Theorem 6.14, and rely on path-counting techniques. We again take

$$a = \partial B_m(o_1), \quad b = \partial B_m(o_2), \quad (6.4.74)$$

and

$$\mathcal{I}_{a,b} = [n] \setminus (B_m(o_1) \cup B_m(o_2)). \quad (6.4.75)$$

Recall that

$$N_k(a, b) = \#\{\pi \in \mathcal{P}_k(a, b) : \pi \text{ occupied}\}. \quad (6.4.76)$$

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

$$\begin{aligned} \mathbb{E}[N_k(a, b)] &= \sum_{\pi \in \mathcal{P}_k(a, b)} \mathbb{P}(\pi \text{ occupied}) = \sum_{\pi \in \mathcal{P}_k(a, b)} \prod_{l=0}^k \frac{\kappa_n(\pi_l, \pi_{l+1})}{n} \\ &\leq \frac{1}{n} \langle \mathbf{x}, T_{\kappa_n}^k \mathbf{y} \rangle, \end{aligned} \quad (6.4.77)$$

where $\mathbf{x} = (x_i)_{i=1}^k$, $\mathbf{y} = (y_i)_{i=1}^k$ 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 $(\underline{\mu}_n - k)/\underline{\mu}_n$, where $\underline{\mu}_n = \min_{j \in [k]} \mu_n(j) \rightarrow \min_{j \in [k]} \mu(j) > 0$, 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 $(Z_m^{(1, j_1)}(i))_{i \in [k]}$, which, by Theorem 3.8(b), is close to $M_\infty \mathbf{x}_{\kappa}(i)$ for some strictly positive random variable M_∞ . We conclude that

$$\mathbf{x} \xrightarrow{d} \mathbf{Z}_m^{(1, j_1)}(i), \quad \mathbf{x} \xrightarrow{d} \mathbf{Z}_m^{(2, j_2)}(i), \quad (6.4.78)$$

where the limiting branching processes are independent. Equation (6.4.78) 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 $\partial B_m(o_1), \partial B_m(o_2) \neq \emptyset$,

$$\mathbb{E}[N_k(a, b)] \xrightarrow{\mathbb{P}} \infty. \quad (6.4.79)$$

This completes the analysis of the first moment.

We next analyse $\text{Var}(N_k(a, b))$ and prove that $\text{Var}(N_k(a, b))/\mathbb{E}[N_k(a, b)]^2 \xrightarrow{\mathbb{P}} \infty$. Here our proof will be slightly more sketchy.

Recall (6.4.18). In this case, we compute, to replace (6.4.19),

$$\mathbb{P}(\pi, \rho \text{ occupied}) = \mathbb{P}(\pi \text{ occupied})\mathbb{P}(\rho \text{ occupied} \mid \pi \text{ occupied}). \quad (6.4.80)$$

Recall the definition of a *shape* in (6.4.16). Fix $\sigma \in \text{Shape}_{m, l}$ and $\rho \in \mathcal{P}_k(a, b)$ with $\text{Shape}(\pi, \rho) = \sigma$. The factor $\mathbb{P}(\rho \text{ occupied} \mid \pi \text{ occupied})$, summed out over the free vertices of ρ (i.e., the ones that are not also vertices in π) gives rise to

6.4 Path counting and log upper bound for finite-variance weights 233

m factors of the form $T_{\kappa_n}^{t_i}(i_{\pi_{u_i}}, i_{\pi_{v_i}})/n$ for $i \in [m+1]$ and some vertices π_{u_i} and π_{v_i} in the path $(\pi_i)_{i=0}^k$. We use that, uniformly in $t \geq 1$,

$$\frac{k^2}{n} \max_{i,j \in [k]} T_{\kappa_n}^t(i, j) \leq \frac{Ck^2}{n} \|\mathbf{T}_{\kappa_n}\|^t, \quad (6.4.81)$$

where k^2 is an upper bound on the number of choices for π_{u_i} and π_{v_i} . Thus, for every of the m subpaths of length t_i we obtain a factor $\frac{Ck^2}{n} \|\mathbf{T}_{\kappa_n}\|^{t_i}$. We arrive at

$$\begin{aligned} & \sum_{\substack{\pi, \rho \in \mathcal{P}_k(a, b) \\ \text{Shape}(\pi, \rho) = \sigma}} \mathbb{P}(\pi, \rho \text{ occupied}) \\ & \leq \mathbb{E}[N_k(a, b)] \prod_{i=1}^m \frac{Ck^2}{n} \|\mathbf{T}_{\kappa_n}\|^{t_i} \leq \mathbb{E}[N_k(a, b)] \|\mathbf{T}_{\kappa_n}\|^{k-l} \left(\frac{Ck^2}{n}\right)^m. \end{aligned} \quad (6.4.82)$$

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 (6.4.47) in the proof of Theorem 6.14. We omit further details. \square

Concentration of the giant in Theorem 3.16

By Theorem 3.11, we know that $\text{IRG}_n(\kappa_n)$ converges in probability in the local weak sense. Therefore, we aim to apply Theorem 2.20, for which we need to verify that (2.5.7) holds.

We rewrite the expectation appearing in (2.5.7) as

$$\begin{aligned} & \frac{1}{n^2} \mathbb{E} \left[\#\{x, y \in [n] : |\mathcal{C}(x)|, |\mathcal{C}(y)| \geq k, x \not\leftrightarrow y\} \right] \\ & = \mathbb{P}(|\mathcal{C}(o_1)|, |\mathcal{C}(o_2)| \geq k, o_1 \not\leftrightarrow o_2). \end{aligned} \quad (6.4.83)$$

We condition on $B_k(o_1)$ and $B_k(o_2)$, and note that the events that $\{|\mathcal{C}(o_1)| \geq k\}$ and $\{|\mathcal{C}(o_2)| \geq k\}$ are measurable with respect to $B_k(o_1)$ and $B_k(o_2)$ to obtain

$$\begin{aligned} & \mathbb{P}(|\mathcal{C}(o_1)|, |\mathcal{C}(o_2)| \geq k, o_1 \not\leftrightarrow o_2) \\ & = \mathbb{E} \left[\mathbb{1}_{\{|\mathcal{C}(o_1)|, |\mathcal{C}(o_2)| \geq k\}} \mathbb{P}(o_1 \not\leftrightarrow o_2 \mid B_k(o_1), B_k(o_2)) \right]. \end{aligned} \quad (6.4.84)$$

By Exercise 6.23, together with Theorem 3.11,

$$\lim_{k \rightarrow \infty} \lim_{n \rightarrow \infty} \mathbb{P}(\partial B_k(o_1) = \emptyset, |\mathcal{C}(o_1)| \geq k) = 0. \quad (6.4.85)$$

Therefore,

$$\begin{aligned} & \mathbb{P}(|\mathcal{C}(o_1)|, |\mathcal{C}(o_2)| \geq k, o_1 \not\leftrightarrow o_2) \\ & = \mathbb{E} \left[\mathbb{1}_{\{|\mathcal{C}(o_1)|, |\mathcal{C}(o_2)| \geq k, \partial B_k(o_1), \partial B_k(o_2) \neq \emptyset\}} \mathbb{P}(o_1 \not\leftrightarrow o_2 \mid B_k(o_1), B_k(o_2)) \right]. \end{aligned} \quad (6.4.86)$$

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,

$$\begin{aligned} & \mathbb{P}(|\mathcal{C}(o_1)|, |\mathcal{C}(o_2)| \geq k, o_1 \not\leftrightarrow o_2) \\ &= \mathbb{E} \left[\mathbb{1}_{\{\partial B_k(o_1), \partial B_k(o_2) \neq \emptyset\}} \mathbb{P}(o_1 \not\leftrightarrow o_2 \mid B_k(o_1), B_k(o_2)) \right] + o_k(1), \end{aligned} \quad (6.4.87)$$

where $o_k(1)$ denotes a quantity that tends to zero when first $n \rightarrow \infty$ followed by $k \rightarrow \infty$.

In the proof of Theorem 6.1(ii), we have shown that, on $\{\partial B_k(o_1), \partial B_k(o_2) \neq \emptyset\}$,

$$\mathbb{P}(d_{\text{IRG}_n(\kappa_n)}(o_1, o_2) \leq (1 + \varepsilon) \log_\nu n \mid B_k(o_1), B_k(o_2)) = 1 + o_\varepsilon(1). \quad (6.4.88)$$

In particular, on $\{\partial B_k(o_1), \partial B_k(o_2) \neq \emptyset\}$,

$$\mathbb{P}(o_1 \not\leftrightarrow o_2 \mid B_k(o_1), B_k(o_2)) = o_\varepsilon(1). \quad (6.4.89)$$

Since also $\mathbb{P}(o_1 \not\leftrightarrow o_2 \mid B_k(o_1), B_k(o_2)) \leq 1$, the Dominated Convergence Theorem [Volume 1, Theorem A.1] completes the proof of (2.5.7) for $\text{IRG}_n(\kappa_n)$, as required. \square

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

$$\text{diam}(G) = \max_{u, v: \text{dist}_G(u, v) < \infty} \text{dist}_G(u, v). \quad (6.5.1)$$

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 κ by a duality principle for $\text{IRG}_n(\kappa)$. Before we state the results, we introduce the notion of the *dual kernel*:

Definition 6.20 (Dual kernel for $\text{IRG}_n(\kappa)$) *Let (κ_n) be a sequence of supercritical kernels with limit κ . The dual kernel is the kernel $\widehat{\kappa}$ defined by $\widehat{\kappa}(x, y) = \kappa(x, y)$, with reference measure $d\widehat{\mu}(x) = (1 - \zeta_\kappa(x))\mu(dx)$.*

The dual kernel shall describe the graph after the removal of the giant component. Here, the reference measure $\widehat{\mu}$ measures the size of the graph. In this case, a vertex x is in the giant component with probability $1 - \zeta_\kappa(x)$, in which case it must be removed. Thus, $\widehat{\mu}$ describes the proportion of vertices of the various

6.5 Related results on distances for inhomogeneous random graphs 235

types which are outside the giant component. As before, we define the operator $\mathbf{T}_{\widehat{\kappa}}$ by the equality

$$(\mathbf{T}_{\widehat{\kappa}}f)(x) = \int_S \widehat{\kappa}(x, y)f(y)d\widehat{\mu}(y) = \int_S \kappa(x, y)f(y)[1 - \zeta_{\kappa}(x)]\mu(dy), \quad (6.5.2)$$

and we write $\|\mathbf{T}_{\widehat{\kappa}}\|$ for

$$\|\mathbf{T}_{\widehat{\kappa}}\| = \sup \{ \|\mathbf{T}_{\widehat{\kappa}}f\|_2 : f \geq 0, \|f\|_{\widehat{\mu}, 2} = 1 \}, \quad (6.5.3)$$

where

$$\|f\|_{\widehat{\mu}, 2}^2 = \int_S f^2(x)\widehat{\mu}(dx). \quad (6.5.4)$$

Theorem 6.21 (The diameter of $\text{IRG}_n(\kappa)$ in the finite-types case) *Let (κ_n) be a sequence of kernels with limit κ , which has finitely many types. If $0 < \|\mathbf{T}_{\kappa}\| < 1$, then*

$$\frac{\text{diam}(\text{IRG}_n(\kappa_n))}{\log n} \xrightarrow{\mathbb{P}} \frac{1}{\log \|\mathbf{T}_{\kappa}\|^{-1}} \quad (6.5.5)$$

as $n \rightarrow \infty$. If $\|\mathbf{T}_{\kappa}\| > 1$ and κ irreducible, then

$$\frac{\text{diam}(\text{IRG}_n(\kappa_n))}{\log n} \xrightarrow{\mathbb{P}} \frac{2}{\log \|\mathbf{T}_{\widehat{\kappa}}\|^{-1}} + \frac{1}{\log \|\mathbf{T}_{\kappa}\|}, \quad (6.5.6)$$

where $\widehat{\kappa}$ is the dual kernel to κ .

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 $d_{\text{IRG}_n(\kappa)(o_1, o_2)}/\log n$ conditioned on $d_{\text{IRG}_n(\kappa)(o_1, o_2)} < \infty$. This effect is particularly noticeable in the case when $\tau \in (2, 3)$, where $d_{\text{IRG}_n(\kappa)(o_1, o_2)}/\log \log n$, conditionally on $d_{\text{IRG}_n(\kappa)(o_1, o_2)} < \infty$, converges in probability 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 $\text{IRG}_n(\kappa)$ arises due to very thin lines of length of order $\log n$. Since these thin lines involve only very few vertices, they will not contribute to $d_{\text{IRG}_n(\kappa)(o_1, o_2)}$, 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 $\text{ER}_n(\lambda/n)$.

We will not prove Theorem 6.21 here. For $\text{GRG}_n(\mathbf{w})$, it will follow from Theorem 7.16, which proves a related result for the configuration model.

6.5.2 DISTANCE FLUCTUATIONS FOR $\text{GRG}_n(\mathbf{w})$ 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

$$1 - F(x) \leq cx^{-(\tau-1)}. \quad (6.5.7)$$

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 $\text{CL}_n(\mathbf{w})$) *Assume that (6.5.7) is satisfied, and let $\nu = \mathbb{E}[W^2]/\mathbb{E}[W] > 1$. For $k \geq 1$, define $a_k = \lfloor \log_\nu k \rfloor - \log_\nu k \in (-1, 0]$. Then, for $\text{CL}_n(\mathbf{w})$ with \mathbf{w} as in (1.3.15), there exist random variables $(R_a)_{a \in (-1, 0]}$ with $\limsup_{K \rightarrow \infty} \sup_{a \in (-1, 0]} \mathbb{P}(|R_a| < K) = 1$ such that, as $n \rightarrow \infty$ and for all $k \in \mathbb{Z}$,*

$$\begin{aligned} \mathbb{P}(d_{\text{NR}_n(\mathbf{w})}(o_1, o_2) - \lfloor \log_\nu n \rfloor = k \mid d_{\text{NR}_n(\mathbf{w})}(o_1, o_2) < \infty) \\ = \mathbb{P}(R_{a_n} = k) + o(1). \end{aligned} \quad (6.5.8)$$

While Theorem 6.1 implies that, conditionally on o_1 and o_2 being connected, $d_{\text{NR}_n(\mathbf{w})}(o_1, o_2)/\log n \xrightarrow{\mathbb{P}} 1/\log \nu$, Theorem 6.22 implies that the fluctuations of $d_{\text{NR}_n(\mathbf{w})}(o_1, o_2)$ 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 $\text{CL}_n(\mathbf{w})$, and depend sensitively on a , which implies that although $\left(d_{\text{NR}_n(\mathbf{w})}(o_1, o_2) - \lfloor \log_\nu n \rfloor\right)_{n=2}^\infty$ 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.

6.6 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 $\text{dist}_{\text{NR}_n(\mathbf{w})}(o_1, o_2)$ when $\nu < \infty$ as in Theorem 6.14 often rely on branching process comparisons up to a generation $m = m_n \rightarrow \infty$.

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 extends a previous result of Chung and Lu (2001) that proves logarithmic asymptotics of $\text{diam}(\text{ER}_n(\lambda/n))$, 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 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

$$d_{\text{ER}_n(\lambda/n)}(o_1, o_2) / \log n \xrightarrow{\mathbb{P}} 1 / \log \lambda$$

conditionally on $d_{\text{ER}_n(\lambda/n)}(o_1, o_2) < \infty$.

Exercise 6.2 (Typical distances when $\nu = \infty$) Prove that

$$d_{\text{NR}_n(\mathbf{w})}(o_1, o_2) / \log n \xrightarrow{\mathbb{P}} 0$$

when $\nu = \infty$, by using an appropriate truncation argument and monotonicity.

Exercise 6.3 (Power-law tails in key example of deterministic weights) *Let \mathbf{w} be defined as in (1.3.15), and assume that F satisfies*

$$1 - F(x) = x^{-(\tau-1)}L(x), \quad (6.7.1)$$

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 $\mathbf{w} = (w_i)_{i \in [n]}$ 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 ε , we can find a constant $K = K(\varepsilon) > 0$ such that*

$$\mathbb{P}(d_{\text{NR}_n(\mathbf{w})}(o_1, o_2) \leq \log_{\nu_n}(n) - K) \leq \varepsilon. \quad (6.7.2)$$

Conclude that if $\log \nu_n = \log \nu + o(1/\log n)$, then the same statement holds with $\log_{\nu} n$ replacing $\log_{\nu_n} n$.

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 $\nu_n/\log n \rightarrow c$. Use Corollary 6.5 to obtain that for any $\varepsilon > 0$,*

$$\mathbb{P}\left(d_{\text{NR}_n(\mathbf{w})}(o_1, o_2) \leq (1 - \varepsilon) \frac{\log n}{\log \log n}\right) = o(1). \quad (6.7.3)$$

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 $d_{\text{NR}_n(\mathbf{w})}(o_1, o_2) \geq (\tau - 1)/(\tau - 3)$ in this case. How useful is this bound?*

Exercise 6.11 (Convergence in probability of typical distance in $\text{IRG}_n(\kappa_n)$) *Suppose that the graphical sequence of kernels (κ_n) satisfies $\sup_{x, y, n} \kappa_n(x, y) < \infty$, where the limit κ is irreducible and $\nu = \|\mathbf{T}_\kappa\| > 1$. Prove that Theorem 3.16 together with Theorem 6.1(i–ii) imply that, conditionally on $d_{\text{IRG}_n(\kappa_n)}(o_1, o_2) < \infty$,*

$$d_{\text{IRG}_n(\kappa_n)}(o_1, o_2)/\log n \xrightarrow{\mathbb{P}} 1/\log \nu. \quad (6.7.4)$$

Exercise 6.12 (Convergence in probability of typical distance in $\text{IRG}_n(\kappa_n)$)

Suppose that the graphical sequence of kernels (κ_n) converges to κ , where κ is irreducible and $\|\mathbf{T}_\kappa\| = \infty$. Prove that Theorem 3.16 together with Theorem 6.1(iii) imply, conditionally on $d_{\text{IRG}_n(\kappa_n)}(o_1, o_2) < \infty$,

$$d_{\text{IRG}_n(\kappa_n)}(o_1, o_2) / \log n \xrightarrow{\mathbb{P}} 0. \tag{6.7.5}$$

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 $a \neq b$,

$$\begin{aligned} \mathbb{P}(\text{dist}_{\text{NR}_n(\mathbf{w})}(a, b) \leq k_n) &\leq \frac{w_a w_b}{\ell_n} \sum_{k=1}^{k_n} \prod_{l=1}^{k-1} \nu_n(b_l \wedge b_{k-l}) \\ &+ (w_a + w_b) \sum_{k=1}^{k^*} [1 - F_n^*](b_k) \prod_{l=1}^k \nu_n(b_l). \end{aligned} \tag{6.7.6}$$

Exercise 6.14 (Lower bound on fluctuations*) Adapt the proof of Theorem 6.6 to show that for every ε , we can find a constant $K = K(\varepsilon) > 0$ such that

$$\mathbb{P}(d_{\text{NR}_n(\mathbf{w})}(o_1, o_2) \leq \frac{2 \log \log n}{|\log(\tau - 2)|} - K) \leq \varepsilon. \tag{6.7.7}$$

Hint: choose $b_k = Lb_{k-1}^{1/(\tau-2)}$, where the constant $L > 0$ 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 $\text{Var}(N_k(a, b)) \leq \mathbb{E}[N_k(a, b)]$ for $k = 2$.

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 $\text{ER}_n(\lambda/n)$) 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 $k \leq K \log n$,

$$\mathbb{E}[N_k(A, B)] = \lambda^k |A| |B| \left(1 - \frac{|A| + |B|}{n}\right)^k (1 + o(1)). \tag{6.7.8}$$

Use Proposition 6.12 to bound the variance of $N_k(A, B)$, and prove that

$$N_k(A, B) / \mathbb{E}[N_k(A, B)] \xrightarrow{\mathbb{P}} 1 \tag{6.7.9}$$

when $|A|, |B| \rightarrow \infty$ with $|A| + |B| = o(n/k)$.

Exercise 6.19 (ν_n bound for $\tau = 3$) Prove that (6.4.53) and (6.4.54) imply that $\nu_x \geq c \log \alpha_n$ by using

$$\frac{1}{n} \sum_{i \in \mathcal{I}} w_i^2 = \mathbb{E}[W_n^2 \mathbb{1}_{\{W_n \in [K, \sqrt{\alpha_n}]\}}] = 2 \int_K^{\sqrt{\alpha_n}} x [F_n(\sqrt{\alpha_n}) - F_n(x)] dx. \tag{6.7.10}$$

Exercise 6.20 (Expected number of paths within Core_n diverges) *Prove that*

$$\mathbb{E}[N_k(a, b)] \rightarrow \infty$$

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*

$$\text{Var}(N_k(a, b)) / \mathbb{E}[N_k(a, b)]^2 \rightarrow 0$$

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 (6.4.49)–(6.4.52).*

Exercise 6.23 (Size versus boundary of large clusters) *Let the graph $G = (V, E)$ converge locally weakly in probability. Prove that*

$$\lim_{k \rightarrow \infty} \lim_{n \rightarrow \infty} \mathbb{P}(\partial B_k(o_1) = \emptyset, |\mathcal{C}(o_1)| \geq k) = 0. \quad (6.7.11)$$

Exercise 6.24 (Diameter of $\text{ER}_n(\lambda/n)$) *Recall Theorem 6.21. For $\text{ER}_n(\lambda/n)$, show that $\|\mathbf{T}_\kappa\| = \lambda$ and $\|\mathbf{T}_{\hat{\kappa}}\| = \mu_\lambda$, where μ_λ is the dual parameter in [Volume 1, (3.6.6)], so that Theorem 6.21 becomes*

$$\frac{\text{diam}(\text{ER}_n(\lambda/n))}{\log n} \xrightarrow{\mathbb{P}} \frac{2}{\log \mu_\lambda^{-1}} + \frac{1}{\log \lambda}. \quad (6.7.12)$$

Exercise 6.25 (Finite variance degrees when (6.5.7) holds) *Prove that (6.5.7) implies that $\mathbb{E}[W^2] < \infty$. 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 $\text{CL}_n(\mathbf{w})$) *Prove that, under the conditions of Theorem 6.22, and conditionally on $d_{\text{NR}_n(\mathbf{w})}(o_1, o_2) < \infty$, the sequence $\left(d_{\text{NR}_n(\mathbf{w})}(o_1, o_2) - \lfloor \log_\nu n \rfloor\right)_{n=2}^\infty$ is tight.*

Exercise 6.27 (Non-convergence of centered typical graph distances in $\text{CL}_n(\mathbf{w})$) *Prove that, under the conditions of Theorem 6.22, and conditionally on $d_{\text{NR}_n(\mathbf{w})}(o_1, o_2) < \infty$, the sequence $d_{\text{NR}_n(\mathbf{w})}(o_1, o_2) - \lfloor \log_\nu n \rfloor$ 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 $\text{GRG}_n(\mathbf{w})$ and $\text{NR}_n(\mathbf{w})$) *Use [Volume 1, Theorem 6.18] to prove that Theorem 6.22 holds verbatim for $\text{GRG}_n(\mathbf{w})$ and $\text{NR}_n(\mathbf{w})$ 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 $CM_n(\mathbf{d})$ 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 $CM_n(\mathbf{d})$ and the AS-counts are quite close.

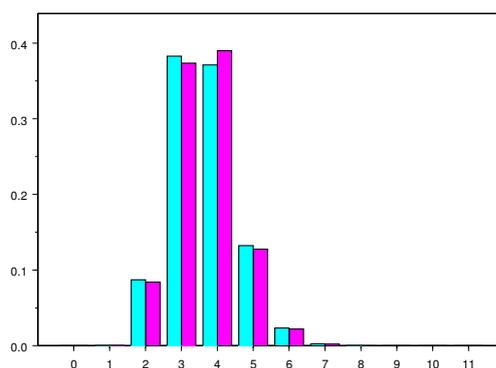


Figure 7.1 Number of AS traversed in hopcount data (blue) compared to the model (purple) with $\tau = 2.25, n = 10,940$.

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 $\text{CM}_n(\mathbf{d})$

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 $\text{CM}_n(\mathbf{d})$ when Conditions 1.5(a)-(c) hold:

Theorem 7.1 (Typical distances in $\text{CM}_n(\mathbf{d})$ for finite-variance degrees) *In the configuration model $\text{CM}_n(\mathbf{d})$, where the degrees $\mathbf{d} = (d_i)_{i \in [n]}$ satisfy Conditions 1.5(a)-(c) and where $\nu > 1$, conditionally on $\text{dist}_{\text{CM}_n(\mathbf{d})}(o_1, o_2) < \infty$,*

$$\text{dist}_{\text{CM}_n(\mathbf{d})}(o_1, o_2) / \log n \xrightarrow{\mathbb{P}} 1 / \log \nu. \quad (7.1.1)$$

Theorem 7.1 shows that the typical distances in $\text{CM}_n(\mathbf{d})$ 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 $\text{NR}_n(\mathbf{w})$ 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 ,

$$c_1 x^{-(\tau-1+\delta)} \leq [1 - F_n](x) \leq c_2 x^{-(\tau-1-\delta)}, \quad (7.1.2)$$

where the upper bound holds 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 typical distances of $\text{CM}_n(\mathbf{d})$ under the infinite-variance condition in (7.1.2) are identified in the following theorem:

Theorem 7.2 (Typical distances in $\text{CM}_n(\mathbf{d})$ for $\tau \in (2, 3)$) *Let the degrees $\mathbf{d} = (d_i)_{i \in [n]}$ in the configuration model $\text{CM}_n(\mathbf{d})$ satisfy Conditions 1.5(a)-(b) and (7.1.2). Then, conditionally on $\text{dist}_{\text{CM}_n(\mathbf{d})}(o_1, o_2) < \infty$,*

$$\frac{\text{dist}_{\text{CM}_n(\mathbf{d})}(o_1, o_2)}{\log \log n} \xrightarrow{\mathbb{P}} \frac{2}{|\log(\tau - 2)|}. \quad (7.1.3)$$

Theorem 7.2 is similar in spirit to Theorem 6.3 for $\text{NR}_n(\mathbf{w})$.

7.2 PROOFS OF SMALL-WORLD RESULTS $\text{CM}_n(\mathbf{d})$

In this section, we give the proofs of Theorems 7.1 and 7.2 describing the small-world properties in $\text{CM}_n(\mathbf{d})$. 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 $\text{CM}_n(\mathbf{d})$. 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 $\text{CM}_n(\mathbf{d})$ 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 $Z_m^{(n;i)} = Z_m(V_i)$, so that $Z_m^{(n;i)}$ 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 $(Z_l^{(1)}, Z_l^{(2)})_{k=0}^m$ 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 $(Z_l^{(1)}, Z_l^{(2)})_{l \geq 0}$ 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 $m \geq 1$,*

$$\mathbb{P}(\text{dist}_{\text{CM}_n(\mathbf{d})}(o_1, o_2) \leq 2m) = o(1), \quad (7.2.1)$$

and

$$(Z_l^{(n;1)}, Z_l^{(n;2)})_{l=0}^m \xrightarrow{d} (Z_l^{(1)}, Z_l^{(2)})_{l=0}^m. \quad (7.2.2)$$

Corollary 7.3 follows from the local-weak convergence in probability in Theorem 4.1, combined with Corollaries 2.12–2.13.

7.2.2 PATH-COUNTING TECHNIQUES

In this section, we present path-counting techniques similar to those in Section 6.4.1. Since $\text{CM}_n(\mathbf{d})$ is a multigraph, and not a simple graph as $\text{NR}_n(\mathbf{w})$, we need to be precise what a path in $\text{CM}_n(\mathbf{d})$ is. We start by introducing some notation.

A *path* π of length k in $\text{CM}_n(\mathbf{d})$ means a sequence

$$\pi = \{(\pi_0, s_0), (\pi_1, s_1, t_1), \dots, (\pi_{k-1}, s_{k-1}, t_{k-1}), (\pi_k, t_k)\}, \quad (7.2.3)$$

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 π_i and $t_{i+1} \in [d_{\pi_{i+1}}]$ denotes the label of the half-edge incident to π_{i+1} . In particular, multiple edges between π_i and π_{i+1} give rise to *distinct* paths through the same vertices. For a path π , we write $\pi \subseteq \text{CM}_n(\mathbf{d})$ when the path π in (7.2.3) is present in $\text{CM}_n(\mathbf{d})$, so that the half-edge corresponding to s_i is paired with the half-edge corresponding to t_{i+1} for $i = 0, \dots, k-1$. Without loss of generality, we assume throughout that the path π is *simple*, i.e., π_0, \dots, π_k are distinct vertices.

In this section, we perform first and second moment computations on the number of paths present in $\text{CM}_n(\mathbf{d})$. We start by proving upper bounds on the expected number of paths.

Upper bounds on the expected number of paths in $\text{CM}_n(\mathbf{d})$

For $a, b \in [n]$, $\mathcal{I} \subseteq [n]$ and $k \geq 1$, we let $\mathcal{P}_k(a, b) = \mathcal{P}_k(a, b; \mathcal{I})$ denote the set of k -paths that only use vertices in \mathcal{I} , and we let

$$N_k(a, b) = N_k(a, b; \mathcal{I}) = \#\{\pi \in \mathcal{P}_k(a, b) : \pi \subseteq \text{CM}_n(\mathbf{d})\} \quad (7.2.4)$$

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 $(d_i)_{i \in [n]}$,*

$$\mathbb{E}[N_k(a, b)] \leq \frac{d_a d_b \ell_n}{(\ell_n - 2k + 1)(\ell_n - 2k)} \nu_x^{k-1}, \quad (7.2.5)$$

where

$$\nu_{\mathcal{I}} = \sum_{i \in \mathcal{I} \setminus \{a,b\}} \frac{d_i(d_i - 1)}{\ell_n}. \tag{7.2.6}$$

Proof The probability that the path π in (7.2.3) is present in $\text{CM}_n(\mathbf{d})$ is equal to

$$\prod_{i=1}^k \frac{1}{\ell_n - 2i + 1}, \tag{7.2.7}$$

and the number of paths with fixed vertices π_0, \dots, π_k is equal to

$$d_{\pi_0} \left(\prod_{i=1}^{k-1} d_{\pi_i}(d_{\pi_i} - 1) \right) d_{\pi_k}. \tag{7.2.8}$$

Substituting $\pi_0 = a, \pi_k = b$, we arrive at

$$\mathbb{E}[N_k(a, b)] = \frac{d_a d_b}{\ell_n - 2k + 1} \sum_{\pi_1, \dots, \pi_{k-1}}^* \prod_{i=1}^{k-1} \frac{d_{\pi_i}(d_{\pi_i} - 1)}{\ell_n - 2i + 1}, \tag{7.2.9}$$

where the sum is over distinct elements of $\mathcal{I} \setminus \{a, b\}$. Let R denote the subset of vertices of $\mathcal{I} \setminus \{a, b\}$ for which $d_i \geq 2$. Then,

$$\mathbb{E}[N_k(a, b)] = \frac{d_a d_b}{\ell_n - 2k + 1} \sum_{\pi_1, \dots, \pi_{k-1} \in R}^* \prod_{i=1}^{k-1} \frac{d_{\pi_i}(d_{\pi_i} - 1)}{\ell_n - 2i + 1}, \tag{7.2.10}$$

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 $a_i \geq 0$,

$$\frac{(r - k + 1)!}{r!} \sum_{\pi_1, \dots, \pi_{k-1} \in R}^* \prod_{i=1}^{k-1} a_i \leq \left(\frac{1}{r} \sum_{i \in R} a_i \right)^{k-1}. \tag{7.2.11}$$

Let $a_i = d_i(d_i - 1)$, so that

$$\sum_{i \in R} a_i = \ell_n \nu_{\mathcal{I}}. \tag{7.2.12}$$

We arrive at

$$\begin{aligned} \mathbb{E}[N_k(a, b)] &= \frac{d_a d_b}{\ell_n - 2k + 1} (\ell_n \nu_{\mathcal{I}} / r)^{k-1} \prod_{i=1}^{k-1} \frac{(r - i + 1)}{(\ell_n - 2i + 1)} \\ &\leq \frac{d_a d_b}{\ell_n - 2k + 1} \frac{\ell_n}{\ell_n - 2k} \nu_{\mathcal{I}}^{k-1} \prod_{i=0}^{k-2} \frac{(1 - \frac{i}{r})}{(1 - \frac{2i}{\ell_n})}. \end{aligned} \tag{7.2.13}$$

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 $\text{CM}_n(\mathbf{d})$

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 $\text{CM}_n(\mathbf{d})$) *Assume that*

$$\limsup_{n \rightarrow \infty} \nu_n > 1, \tag{7.2.14}$$

where

$$\nu_n = \mathbb{E}[D_n(D_n - 1)]/\mathbb{E}[D_n]. \tag{7.2.15}$$

Then, for any $\varepsilon > 0$,

$$\mathbb{P}(\text{dist}_{\text{CM}_n(\mathbf{d})}(o_1, o_2) \leq (1 - \varepsilon) \log_{\nu_n} n) = o(1). \tag{7.2.16}$$

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 $\text{NR}_n(\mathbf{w})$ 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 $\text{CM}_n(\mathbf{d})$) *Suppose that the weights $\mathbf{d} = (d_i)_{i \in [n]}$ satisfy Condition 1.5(a) and that there exists a $\tau \in (2, 3)$ and c_2 such that, for all $x \geq 1$,*

$$[1 - F_n](x) \leq c_2 x^{-(\tau-1)}, \tag{7.2.17}$$

Then, for every $\varepsilon > 0$,

$$\mathbb{P}\left(\text{dist}_{\text{CM}_n(\mathbf{d})}(o_1, o_2) \leq (1 - \varepsilon) \frac{2 \log \log n}{|\log(\tau - 2)|}\right) = o(1). \tag{7.2.18}$$

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 (π_0, \dots, π_k) , (7.2.7)-(7.2.8) yield that the probability that there exists edges between π_{i-1} and π_i for all $i = 1, \dots, k$ in $\text{CM}_n(\mathbf{d})$ is bounded above by

$$\frac{d_{\pi_0} d_{\pi_k}}{\ell_n - 2k + 1} \prod_{i=1}^{k-1} \frac{d_{\pi_i} (d_{\pi_i} - 1)}{\ell_n - 2i + 1}. \tag{7.2.19}$$

Equation (7.2.19) replaces the similar identity (6.2.7) for $\text{NR}_n(\mathbf{w})$. We see that w_{π_0} and w_{π_k} in (6.2.7) are replaced with d_{π_0} and d_{π_k} in (7.2.19), and, for $i = 1, \dots, k - 1$, 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 ℓ_n in (6.2.7) is replaced with $(\ell_n - 2i + 1)$ in (7.2.19).

Define, as in (6.2.32),

$$\nu_n(b) = \frac{1}{\ell_n} \sum_{i \in [n]} d_i(d_i - 1) \mathbb{1}_{\{d_i \leq b\}}. \quad (7.2.20)$$

Then, we can adapt the arguments in Section 6.2.2 to obtain that (see in particular Exercise 6.13),

$$\begin{aligned} \mathbb{P}(\text{dist}_{\text{CM}_n(\mathbf{d})}(a, b) \leq k_n) &\leq \frac{d_a d_b}{\ell_n} \sum_{k=1}^{k_n} \frac{\ell_n^k (\ell_n - 2k - 1)!!}{(\ell_n - 1)!!} \prod_{l=1}^{k-1} \nu_n(b_l \wedge b_{k-l}) \\ &\quad + (d_a + d_b) \sum_{k=1}^{k^*} \frac{\ell_n^k (\ell_n - 2k - 1)!!}{(\ell_n - 1)!!} [1 - F_n^*(b_k)] \prod_{l=1}^k \nu_n(b_l), \end{aligned} \quad (7.2.21)$$

i.e., the bound in (6.7.6) is changed by factors $\frac{\ell_n^k (\ell_n - 2k - 1)!!}{(\ell_n - 1)!!}$ in the sum. For $k = O(\log \log n)$ and when Conditions 1.5(a)-(b) hold,

$$\frac{\ell_n^k (\ell_n - 2k - 1)!!}{(\ell_n - 1)!!} = \prod_{i=1}^k \frac{\ell_n}{\ell_n - 2i + 1} = 1 + O(k^2/\ell_n) = 1 + o(1), \quad (7.2.22)$$

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. \square

Second moment method for the number of paths in $\text{CM}_n(\mathbf{d})$

We next extend the above first moment bounds on the number of paths in $\text{CM}_n(\mathbf{d})$ to second moment methods. Define

$$\bar{n}_k(a, b) = \frac{\ell_n^k (\ell_n - 2k - 1)!!}{(\ell_n - 1)!!} \frac{d_a d_b}{\ell_n} \left(\sum_{i \in \mathcal{I} \setminus \{a, b\}} \frac{d_i(d_i - 1)}{\ell_n} \right)^{k-1}, \quad (7.2.23)$$

$$\underline{n}_k(a, b) = \frac{d_a d_b}{\ell_n} \left(\sum_{i \in \mathcal{I}_{a, b, k}} \frac{d_i(d_i - 1)}{\ell_n} \right)^{k-1}, \quad (7.2.24)$$

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

$$\nu_{\mathcal{I}} = \frac{1}{\ell_n} \sum_{i \in \mathcal{I}} d_i(d_i - 1), \quad \gamma_{\mathcal{I}} = \frac{1}{\ell_n^{3/2}} \sum_{i \in \mathcal{I}} d_i(d_i - 1)(d_i - 2). \quad (7.2.25)$$

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 $(u_i)_{i \in \mathcal{I}}$,*

$$\mathbb{E}[N_k(a, b)] \geq \underline{n}_k(a, b), \quad (7.2.26)$$

while, assuming that $\nu_x > 1$,

$$\text{Var}(N_k(a, b)) \leq n_k(a, b) + \bar{n}_k(a, b)^2 \left(\frac{\gamma_x \nu_x^2}{\nu_x - 1} \left(\frac{1}{d_a} + \frac{1}{d_b} \right) + \frac{\gamma_x^2 \nu_x}{d_a d_b (\nu_x - 1)^2} + e'_k \right), \quad (7.2.27)$$

where

$$e'_k = \left(\prod_{i=1}^k \frac{\ell_n - 2i + 1}{\ell_n - 2i - 2k + 1} - 1 \right) + k \frac{\ell_n^{2k} (\ell_n - 4k - 1)!!}{(\ell_n - 1)!!} \left(1 + \frac{\gamma_x}{d_a \nu_x} \right) \left(1 + \frac{\gamma_x}{d_b \nu_x} \right) \frac{\nu_x}{\nu_x - 1} (e^{2k^3 \gamma_x^2 / \nu_x^3} - 1). \quad (7.2.28)$$

Proof The proof of (7.2.26) follows immediately from (7.2.9), together with the fact that $1/(\ell_n - 2i + 1) \geq 1/\ell_n$.

For the proof of (7.2.27), we follow the proof of (6.4.7), and discuss the differences only. We recall that

$$N_k(a, b) = \sum_{\pi \in \mathcal{P}_k(a, b)} \mathbb{1}_{\{\pi \subseteq \text{CM}_n(\mathbf{d})\}} \quad (7.2.29)$$

is the number of paths π 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

$$\begin{aligned} \text{Var}(N_k(a, b)) &= \sum_{\pi, \rho \in \mathcal{P}_k(a, b)} [\mathbb{P}(\pi, \rho \subseteq \text{CM}_n(\mathbf{d})) - \mathbb{P}(\pi \subseteq \text{CM}_n(\mathbf{d}))\mathbb{P}(\rho \subseteq \text{CM}_n(\mathbf{d}))]. \end{aligned} \quad (7.2.30)$$

Equation (7.2.30) replaces (6.4.12) for $\text{NR}_n(\mathbf{w})$.

We say that two paths π and ρ are *disjoint* when they use distinct sets of half-edges. Thus, it is possible that the vertex sets $\{\pi_1, \dots, \pi_{k-1}\}$ and $\{\rho_1, \dots, \rho_{k-1}\}$ have a non-empty intersection, but then the half-edges leading in and out of the joint vertices for π and ρ must be distinct. For $\text{NR}_n(\mathbf{w})$, pairs of paths using different edges are *independent*, so that these pairs do not contribute to $\text{Var}(N_k(a, b))$. For $\text{CM}_n(\mathbf{d})$, instead, for disjoint pairs π and ρ ,

$$\mathbb{P}(\pi, \rho \subseteq \text{CM}_n(\mathbf{d})) = \prod_{i=1}^k \frac{\ell_n - 2i + 1}{\ell_n - 2i - 2k + 1} \mathbb{P}(\pi \subseteq \text{CM}_n(\mathbf{d}))\mathbb{P}(\rho \subseteq \text{CM}_n(\mathbf{d})), \quad (7.2.31)$$

which explains the first contribution to e'_k . For the other contributions, we follow the proof of (6.4.12) for $\text{NR}_n(\mathbf{w})$, and omit further details. \square

With Proposition 7.7 at hand, we can adapt the proof of Theorem 6.14 to $\text{CM}_n(\mathbf{d})$ to prove the following theorem:

Theorem 7.8 (Logarithmic upper bound graph distances $\text{CM}_n(\mathbf{d})$) *Assume*

that Conditions 1.5(a)-(c) hold, where $\nu = \mathbb{E}[D(D - 1)]/\mathbb{E}[D] \in (1, \infty)$. Then, for any $\varepsilon > 0$,

$$\mathbb{P}(\text{dist}_{\text{CM}_n(\mathbf{d})}(o_1, o_2) \leq (1 + \varepsilon) \log_\nu n \mid \text{dist}_{\text{CM}_n(\mathbf{d})}(o_1, o_2) < \infty) = 1 + o(1). \quad (7.2.32)$$

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 $\text{CM}_n(\mathbf{d})$ 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 $\text{CM}_n(\mathbf{d})$ in Theorem 7.2 is organized as follows:

- (a) We first prove an upper bound on the diameter of the *core* of $\text{CM}_n(\mathbf{d})$, 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 $\text{CM}_n(\mathbf{d})$ in Theorem 7.2. The bound on the diameter of the core is also useful in studying the diameter of $\text{CM}_n(\mathbf{d})$ 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* 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 Core_n is with high probability bounded above by*

$$\frac{2 \log \log n}{|\log(\tau - 2)|} + 1. \quad (7.2.34)$$

Proof We note that (7.1.2) implies that, for some $\alpha \in (1/2, 1/(\tau - 1))$,

$$\max_{i \in [n]} d_i \geq u_1, \quad \text{where } u_1 = n^\alpha. \quad (7.2.35)$$

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

$$u_k = C \log n (u_{k-1})^{\tau-2}. \quad (7.2.37)$$

Then, we define

$$\Gamma_k = \{i: d_i \geq u_k\}. \quad (7.2.38)$$

We identify u_k in the following lemma:

Lemma 7.10 (Identification $(u_k)_{k \geq 1}$) *For every $k \geq 1$,*

$$u_k = C^{a_k} (\log n)^{b_k} n^{c_k}, \quad (7.2.39)$$

where

$$c_k = \alpha(\tau - 2)^{k-1}, \quad a_k = b_k = \frac{1}{3 - \tau} [1 - (\tau - 2)^{k-1}]. \quad (7.2.40)$$

Proof We note that c_k, b_k, a_k satisfy the recursions, for $k \geq 2$,

$$c_k = (\tau - 2)c_{k-1}, \quad b_k = 1 + (\tau - 2)b_{k-1}, \quad a_k = 1 + (\tau - 2)a_{k-1}, \quad (7.2.41)$$

with initial conditions $c_1 = \alpha, a_1 = b_1 = 0$. Solving the recursions yields our claim. \square

In order to study connectivity of sets in $\text{CM}_n(\mathbf{d})$, we rely on the following lemma, which is of independent interest:

Lemma 7.11 (Connectivity sets in $\text{CM}_n(\mathbf{d})$) *For any two sets of vertices $A, B \subseteq [n]$,*

$$\mathbb{P}(A \text{ not directly connected to } B) \leq e^{-d_A d_B / (2\ell_n)}, \quad (7.2.42)$$

where, for any $A \subseteq [n]$,

$$d_A = \sum_{i \in A} d_i \quad (7.2.43)$$

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

$$1 - \frac{d_B}{\ell_n - 2k + 1} \leq 1 - \frac{d_B}{\ell_n}. \quad (7.2.44)$$

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

$$\mathbb{P}(A \text{ not directly connected to } B) \leq \left(1 - \frac{d_B}{\ell_n}\right)^{d_A/2} \leq e^{-d_A d_B / (2\ell_n)}, \quad (7.2.45)$$

where we use that $1 - x \leq e^{-x}$. \square

The key step in the proof of Theorem 7.9 is the following proposition showing that whp every vertex in Γ_k is connected to a vertex in Γ_{k-1} :

Proposition 7.12 (Connectivity between Γ_{k-1} and Γ_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 $C > 2\mathbb{E}[D]/c$. Then, the probability that there exists an $i \in \Gamma_k$ that is not directly connected to Γ_{k-1} is $o(n^{-\delta})$, for some $\delta > 0$ that is independent of k .*

Proof We note that, by definition,

$$\sum_{i \in \Gamma_{k-1}} d_i \geq u_{k-1} |\Gamma_{k-1}| = u_{k-1} n [1 - F_n](u_{k-1}). \quad (7.2.46)$$

By (7.1.2), and since $k \mapsto u_k$ is decreasing with $u_1 = n^\alpha$,

$$[1 - F_n](u_{k-1}) \geq c(u_{k-1})^{1-\tau}. \quad (7.2.47)$$

As a result, we obtain that for every $k \geq 2$,

$$\sum_{i \in \Gamma_{k-1}} d_i \geq cn(u_{k-1})^{2-\tau}. \quad (7.2.48)$$

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 Γ_{k-1} is bounded by

$$ne^{-\frac{u_k n u_{k-1} [1 - F(u_{k-1})]}{2\ell_n}} \leq ne^{-\frac{cu_k (u_{k-1})^{2-\tau}}{2\mathbb{E}[D_n]}} = n^{1 - \frac{cC}{2\mathbb{E}[D_n]}}. \quad (7.2.49)$$

By Conditions 1.5(a)-(b), $\mathbb{E}[D_n] \rightarrow \mathbb{E}[D]$, so that, as $n \rightarrow \infty$ and taking $C > 2\mathbb{E}[D]/c$, we obtain the claim for any $\delta < \frac{cC}{2\mathbb{E}[D]} - 1$. \square

We now complete the proof of Theorem 7.9:

Proof of Theorem 7.9. Fix

$$k^* = \frac{\log \log n}{|\log(\tau - 2)|}. \quad (7.2.50)$$

As a result of Proposition 7.12, whp, the diameter of Γ_{k^*} is at most $2k^* + 1$, because the distance between any vertex in Γ_{k^*} and Γ_1 is at most k^* , while, by Exercise 7.4, Γ_1 forms a complete graph. Therefore, it suffices to prove that

$$\text{Core}_n \subseteq \Gamma_{k^*}. \quad (7.2.51)$$

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,

$$u_{k^*} = C^{a_{k^*}} (\log n)^{b_{k^*}} n^{c_{k^*}}. \quad (7.2.52)$$

We note that $n^{c_{k^*}} = e^{\log n (\tau - 2)^{k^*}}$. Since, for $2 < \tau < 3$,

$$x(\tau - 2)^{\frac{\log x}{|\log(\tau - 2)|}} = x \cdot x^{-1} = 1, \quad (7.2.53)$$

we find with $x = \log n$ that $n^{c_{k^*}} = e$. Further, $b_k \rightarrow 1/(\tau - 3)$ as $k \rightarrow \infty$, so that $(\log n)^{b_{k^*}} = (\log n)^{1/(3-\tau)+o(1)}$, and $a_k = b_k$, so that also $C^{a_{k^*}} = C^{1/(\tau-3)+o(1)}$. We conclude that

$$u_{k^*} = (\log n)^{1/(3-\tau)+o(1)}, \quad (7.2.54)$$

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 $\text{dist}_{\text{CM}_n(\mathbf{d})}(o_1, o_2)$ 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 $x \leq n^\alpha$,

$$[1 - F_n](x) \geq c_1 x^{-(\tau-1)}. \quad (7.2.55)$$

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 $\mathbf{d} = (d_i)_{i \in [n]}$ satisfies Conditions 1.5(a)-(b) and (7.2.55). Then, for every $\varepsilon > 0$,*

$$\lim_{n \rightarrow \infty} \mathbb{P} \left(\text{dist}_{\text{CM}_n(\mathbf{d})}(o_1, o_2) \leq \frac{2(1 + \varepsilon) \log \log n}{|\log(\tau - 2)|} \mid \text{dist}_{\text{CM}_n(\mathbf{d})}(o_1, o_2) < \infty \right) = 1. \quad (7.2.56)$$

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 $i \in \{1, 2\}$, $Z_m^{(n;i)}$ denotes the number of unpaired or free half edges incident to vertices in $B_m(o_i)$. By Corollary 7.3, $(Z_l^{(1)}, Z_l^{(2)})_{l=0}^m$ can be whp perfectly coupled to $(Z_l^{(n;1)}, Z_l^{(n;2)})_{l=0}^m$, 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 $B_m(o_1), B_m(o_2)$ which are such that $Z_m^{(n;1)} \geq 1, Z_m^{(n;2)} \geq 1$. We couple these to $(Z_m^{(1)}, Z_m^{(2)})$. We note that, conditionally on $Z_m^{(1)} \geq 1, Z_m^{(2)} \geq 1$, $Z_m^{(1)} \xrightarrow{\mathbb{P}} \infty, Z_m^{(2)} \xrightarrow{\mathbb{P}} \infty$ when $m \rightarrow \infty$.

We will condition on $(B_m(o_1), B_m(o_2))$, and denote the conditional distribution by $\tilde{\mathbb{P}}_m$, and the expectation and variance under the measure $\tilde{\mathbb{P}}_m$ by $\tilde{\mathbb{E}}_m$ and $\tilde{\text{Var}}_m$, 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 $d_{a_1} = Z_m^{(n;1)}, d_{a_2} = Z_m^{(n;2)}$ 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 $\mathcal{I} = \{i \in R : d_i \leq K\}$. Then, Proposition 7.7 gives that, conditionally on $(B_m(o_1), B_m(o_2))$ such that $Z_m^{(n;1)} \geq 1, Z_m^{(n;2)} \geq 1$,

$$\begin{aligned} \tilde{\mathbb{P}}_m(N_k(a_i, b) = 0) &\leq \widetilde{\text{Var}}_m(N_k(a_i, b)) / \tilde{\mathbb{E}}_m[N_k(a_i, b)]^2 \\ &\leq O(K)(1/Z_m^{(n;1)} + 1/Z_m^{(n;2)}) \xrightarrow{\mathbb{P}} 0, \end{aligned} \quad (7.2.57)$$

in the iterated limit where first $n \rightarrow \infty$, followed by $m \rightarrow \infty$. As a result, conditionally on $(B_m(o_1), B_m(o_2))$ such that $Z_m^{(n;1)} \geq 1, Z_m^{(n;2)} \geq 1$, with probability at least $1 - o(1)$, $N_k(a_i, b) \geq 1$, so that, on this event,

$$\text{dist}_{\text{CM}_n(\mathbf{d})}(o_1, o_2) \leq \text{diam}_{\text{CM}_n(\mathbf{d})}(\text{Core}_n) + 2k \leq \frac{2(1 + \varepsilon) \log \log n}{|\log(\tau - 2)|}. \quad (7.2.58)$$

We further use that, by Theorem 4.4,

$$\mathbb{P}(\text{dist}_{\text{CM}_n(\mathbf{d})}(o_1, o_2) < \infty) \rightarrow \zeta^2, \tag{7.2.59}$$

while

$$\mathbb{P}(Z_m^{(n;1)} \geq 1, Z_m^{(n;2)} \geq 1) \rightarrow \zeta^2 \tag{7.2.60}$$

in the iterated limit where first $n \rightarrow \infty$, followed by $m \rightarrow \infty$. As a result, with $k_n = \frac{2(1+\varepsilon)\log \log n}{|\log(\tau-2)|}$,

$$\begin{aligned} &\mathbb{P}(\text{dist}_{\text{CM}_n(\mathbf{d})}(o_1, o_2) \leq k_n \mid \text{dist}_{\text{CM}_n(\mathbf{d})}(o_1, o_2) < \infty) && (7.2.61) \\ &\geq \frac{\mathbb{P}(\text{dist}_{\text{CM}_n(\mathbf{d})}(o_1, o_2) \leq k_n, Z_m^{(n;1)} \geq 1, Z_m^{(n;2)} \geq 1)}{\mathbb{P}(\text{dist}_{\text{CM}_n(\mathbf{d})}(o_1, o_2) < \infty)} \\ &= \frac{\mathbb{P}(Z_m^{(n;1)} \geq 1, Z_m^{(n;2)} \geq 1) - o(1)}{\mathbb{P}(\text{dist}_{\text{CM}_n(\mathbf{d})}(o_1, o_2) < \infty)} = 1 - o(1), \end{aligned}$$

as required. This completes the proof of Theorem 7.13. □

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 $\text{CM}_n(\mathbf{d})$, 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*

$$x^{-\alpha-\gamma(x)} \leq 1 - F_x(x) \leq x^{-\alpha+\gamma(x)}, \quad \text{for large } x, \tag{7.3.1}$$

where $x \mapsto \gamma(x)$ satisfies

- (i) $x \mapsto x^{\gamma(x)}$ is non-decreasing,
- (ii) $\int_0^\infty \gamma(e^{e^x}) dx < \infty$, 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 α 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 $\mathbb{E}[X^s] = \infty$ 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

$$M_n = \alpha^n \log(Z_n \vee 1). \quad (7.3.2)$$

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

$$Y_i = \alpha^i \log \left(\frac{(Z_i \vee 1)}{(Z_{i-1} \vee 1)^{1/\alpha}} \right). \quad (7.3.3)$$

We can write

$$M_n = Y_1 + Y_2 + \cdots + Y_n. \quad (7.3.4)$$

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

$$\sum_{i=1}^{\infty} \mathbb{E}[|Y_i|] < \infty. \quad (7.3.5)$$

This is what we prove in the following three steps.

Inserting normalization sequences

We next investigate $\mathbb{E}[|Y_i|]$. We prove by induction on i that there exist constants $\kappa < 1$ and $C > 0$ such that

$$\mathbb{E}[|Y_i|] \leq K \kappa^i. \quad (7.3.6)$$

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

$$u_n = \inf\{x: 1 - F_x(x) \leq 1/n\}. \quad (7.3.7)$$

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

$$U_i = \alpha^i \log \left(\frac{u_{Z_{i-1} \vee 1}}{(Z_{i-1} \vee 1)^{1/\alpha}} \right), \quad V_i = \alpha^i \log \left(\frac{Z_i \vee 1}{u_{Z_{i-1} \vee 1}} \right). \quad (7.3.8)$$

Then, $Y_i = U_i + V_i$, so that

$$\mathbb{E}[|Y_i|] \leq \mathbb{E}[|U_i|] + \mathbb{E}[|V_i|]. \tag{7.3.9}$$

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 $\lim_{x \rightarrow \infty} \gamma(x) = 0$, there exists a constant $C_\varepsilon \geq 1$ such that, for all $n \geq 1$,

$$u_n \leq C_\varepsilon n^{1/\alpha + \varepsilon}, \tag{7.3.10}$$

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

$$1 + o(1) = n[1 - F_x(u_n)] \geq n[u_n^{-(\tau-1)-\gamma(u_n)}] \geq n[u_n^{-(\tau-1)} e^{\log(C_\varepsilon n^{1/\alpha + \varepsilon})^\gamma}], \tag{7.3.11}$$

which, in turn, implies that there exists a constant $c > 0$ such that

$$u_n \leq n^{1/\alpha} e^{c(\log n)^\gamma}. \tag{7.3.12}$$

In a similar way, we can show the matching lower bound $u_n \geq n^{1/\alpha} e^{-c(\log n)^\gamma}$. As a result,

$$\mathbb{E}[|U_i|] \leq c\alpha^i \mathbb{E}[(\log(Z_{i-1} \vee 1))^\gamma]. \tag{7.3.13}$$

Using the concavity of $x \mapsto x^\gamma$ for $\gamma \in [0, 1)$, as well as Jensen's Inequality, we arrive at

$$\mathbb{E}[|U_i|] \leq c\alpha^i \left(\mathbb{E}[(\log(Z_{i-1} \vee 1))] \right)^\gamma = \alpha^{i(1-\gamma)} \mathbb{E}[M_{i-1}]^\gamma. \tag{7.3.14}$$

By (7.3.4) and (7.3.6), which implies that $\mathbb{E}[M_{i-1}] \leq K\kappa/(1-\kappa)$, we arrive at

$$\mathbb{E}[|U_i|] \leq \alpha^{i(1-\gamma)} c \left(\frac{K\kappa}{1-\kappa} \right)^\gamma. \tag{7.3.15}$$

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 $(Z_i \vee 1)/(u_{Z_i \vee 1})$ should be close to a stable random variable. We make use of this fact by bounding

$$\mathbb{E}[|V_i|] \leq \alpha^i \sup_{m \geq 1} \mathbb{E}[|\log(S_m/u_m)|], \tag{7.3.16}$$

where $S_m = X_1 + \dots + X_m$, and $(X_i)_{i=1}^m$ are i.i.d. copies of the offspring distribution X . We shall prove that there exists a constant $C > 0$ such that, for all $m \geq 1$,

$$\mathbb{E}[|\log(S_m/u_m)|] \leq C. \tag{7.3.17}$$

In order to prove (7.3.17), we note that it suffices to bound

$$\mathbb{E}[(\log(S_m/u_m))_+] \leq C_+, \quad \mathbb{E}[(\log(S_m/u_m))_-] \leq C_-, \tag{7.3.18}$$

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 $\mathbb{E}[(\log(S_m/u_m))_-]$. We note that $(\log x)_- = \log(x^{-1} \vee 1)$, so that

$$\mathbb{E}[(\log(S_m/u_m))_-] = \mathbb{E}[\log(u_m/(S_m \wedge u_m))], \quad (7.3.19)$$

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

$$\mathbb{E}[\log(u_m/(S_m \wedge u_m))] \leq \mathbb{E}[\log(u_m/(X_{(m)} \wedge u_m))]. \quad (7.3.20)$$

We next use that, for $x \geq 1$, $x \mapsto \log(x)$ is concave, so that, for every s ,

$$\begin{aligned} \mathbb{E}[\log(u_m/(X_{(m)} \wedge u_m))] &= \frac{1}{s} \mathbb{E}[\log((u_m/(X_{(m)} \wedge u_m))^s)] \\ &\leq \frac{1}{s} \log(\mathbb{E}[(u_m/(X_{(m)} \wedge u_m))^s]) \\ &\leq \frac{1}{s} + \frac{1}{s} \log(u_m^s \mathbb{E}[X_{(m)}^{-s}]), \end{aligned} \quad (7.3.21)$$

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 $X_{(m)}^{-s} = (-Y_{(m)})^s$, where $Y_j = -X_j^{-1}$ and $Y_{(m)} = \max_{1 \leq j \leq m} Y_j$. Clearly, $Y_j \in [-1, 0]$ since $X_i \geq 1$, so that $\mathbb{E}[(-Y_1)^s] < \infty$. Also, $u_m Y_{(m)} = -u_m/X_{(m)}$ converges in distribution to $-E^{-1/\alpha}$, where E is exponential with mean 1, so it follows from (Pickands III, 1968, Theorem 2.1) that, as $m \rightarrow \infty$,

$$\mathbb{E}[(u_m Y_{(m)})^p] \rightarrow \mathbb{E}[E^{-1/\alpha}] < \infty. \quad (7.3.22)$$

We proceed with $\mathbb{E}[(\log(S_m/u_m))_+]$, 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

$$\mathbb{E}[\log(S_m \vee u_m/u_m)] = \frac{1}{s} \mathbb{E}[\log((S_m \vee u_m/u_m)^s)] \leq \frac{1}{s} + \log(\mathbb{E}[(S_m/u_m)^s]). \quad (7.3.23)$$

The discussion on (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 $\lambda_s(m)$ depending on s , m and F_x . Using the discussion on (Hall, 1981, Page 564), we have that $\lambda_s(m) \leq C_s m^{s/\alpha} l(m^{1/\alpha})^s$, where $l(\cdot)$ is a slowly-varying function. With some more effort, it can be shown that we can replace $l(m^{1/\alpha})$ by $\ell(m)$, which gives

$$\mathbb{E}[\log(S_m \vee u_m/u_m)] \leq \frac{1}{s} + \log \mathbb{E}\left[\left(\frac{S_m}{u_m}\right)^s\right] \leq \frac{1}{s} + \frac{C_s}{s} m^{s/\alpha} \ell(m)^s u_m^{-s} = \frac{1}{s} + 2^{s/2} \frac{C_s}{s}, \quad (7.3.24)$$

and which proves the first bound in (7.3.18) with $C_+ = \frac{1}{s} + 2^{s/2} \frac{C_s}{s}$.

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

$$\mathbb{E}[|Y_i|] \leq \alpha^{i(1-\gamma)} c \left(\frac{K\kappa}{1-\kappa}\right)^\gamma + C\alpha^i \leq K\kappa^i, \quad (7.3.25)$$

when we take $\kappa = \alpha^{1-\gamma}$ and we take K to be sufficiently large, for example $K \geq 2C$ and $K \geq 2c \left(\frac{K\kappa}{1-\kappa} \right)^\gamma$. 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 $\mathbb{E}[X] = \infty$, the survival probability $\zeta = \mathbb{P}(Z_n \geq 1 \forall n \geq 0)$ 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,

$$\frac{Z_n^{(\infty)}}{Z_n} \xrightarrow{a.s.} \xi > 0, \quad (7.3.26)$$

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, $\mathbb{P}(Z_1^{(\infty)} \geq 1) = 1$, 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,

$$\alpha^n \log(Z_n^{(\infty)} \vee 1) \xrightarrow{a.s.} Y^{(\infty)}, \quad (7.3.27)$$

and combining (7.3.26) and (7.3.27), it immediately follows that, conditionally on survival,

$$\alpha^n \log(Z_n \vee 1) \xrightarrow{a.s.} Y^{(\infty)}. \quad (7.3.28)$$

We conclude that Theorem 7.14 holds, where $Y = 0$ with probability $\eta = 1 - \zeta$ and $Y = Y^{(\infty)}$ with probability ζ . \square

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,*

$$\lim_{x \rightarrow \infty} \frac{\log \mathbb{P}(Y > x)}{x} = -1, \quad (7.3.29)$$

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),

$$Y = \sum_{n=1}^{\infty} Y_n, \quad (7.3.30)$$

where

$$Y_1 = \alpha \log(Z_1 \vee 1). \quad (7.3.31)$$

By (7.3.1),

$$\mathbb{P}(Y_1 > x) = \mathbb{P}(Z_1 > e^{x^{1/\alpha}}) = e^{-x(1+o(1))}, \quad (7.3.32)$$

which shows 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).

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

$$\mathbb{P}(\text{dist}_{\text{CM}_n(\mathbf{d})}(o_1, o_2) = k) = \frac{1}{n} \mathbb{E}[|B_k(o_1)|]. \quad (7.3.33)$$

This suggests that $\text{dist}_{\text{CM}_n(\mathbf{d})}(o_1, o_2)$ should be closely related to the value of k (if any exists) such that $\mathbb{E}[|B_k(o_1)|] = \Theta(n)$. Using the branching process approximation that

$$|B_k(o_1)| = Z_k^{(n;1)} \approx Z_k^{(1)} \approx e^{(\tau-2)^{-k} Y_1(1+o_{\mathbb{P}}(1))} \quad (7.3.34)$$

(the first approximation being true for smallish k , but not necessarily for large k), this suggest that $\text{dist}_{\text{CM}_n(\mathbf{d})}(o_1, o_2) \approx k_n$, where, by Theorem 7.14,

$$\Theta(n) = Z_{k_n}^{(1)} = e^{(\tau-2)^{-k_n} Y_1(1+o_{\mathbb{P}}(1))}, \quad (7.3.35)$$

which suggests that $\text{dist}_{\text{CM}_n(\mathbf{d})}(o_1, o_2) \approx \log \log n / |\log(\tau - 2)|$. Of course, for such values, the branching process approximation may fail miserably, and in fact it does. This is exemplified the fact that $e^{(\tau-2)^{-k_n} Y_1(1+o_{\mathbb{P}}(1))}$ can become much larger than n , which is clearly impossible for $|B_{k_n}(o_1)|$.

More intriguingly, comparing with we see that the proposed typical distances are a factor 2 too small compared to Theorem 7.2. The reason is that the double-exponential growth can clearly no longer be valid when $Z_k^{(n;1)}$ becomes too large, and thus, $Z_k^{(n;1)}$ must be far away from $Z_k^{(1)}$ in this regime. The whole problem is that we are using the branching-process approximation well beyond its expiration date.

So, let us try this again, but rather than using it for *one* neighborhood, let us use the branching process approximation from two sides. Now we rely on the statement that

$$\mathbb{P}(\text{dist}_{\text{CM}_n(\mathbf{d})}(o_1, o_2) \leq 2k) = \mathbb{P}(B_k(o_1) \cap B_k(o_2) \neq \emptyset). \quad (7.3.36)$$

Again using (7.3.35), we see that $|B_k(o_1)| \approx e^{(\tau-2)^{-k} Y_1(1+o_{\mathbb{P}}(1))}$ and $|B_k(o_2)| \approx e^{(\tau-2)^{-k} Y_2(1+o_{\mathbb{P}}(1))}$, where Y_1 and Y_2 are independent. These grow roughly at the same pace, and in particular, $|B_k(o_1)| = n^{\Theta(1)}$ roughly at the same time, namely, when $k \approx \log \log n / |\log(\tau - 2)|$. Thus, we conclude that $\text{dist}_{\text{CM}_n(\mathbf{d})}(o_1, o_2) \approx 2 \log \log n / |\log(\tau - 2)|$, as rigorously proved in Theorem 7.2. We will see that

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 $G_D^*(x)$ 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 ξ is the extinction probability of the branching process with offspring distribution p defined in (4.2.3) and further define

$$\mu = G_D^*(\xi) = \sum_{k=1}^{\infty} k \xi^{k-1} p_k. \tag{7.4.1}$$

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,*

$$\frac{\text{diam}(\text{CM}_n(\mathbf{d}))}{\log n} \xrightarrow{\mathbb{P}} \frac{\mathbb{1}_{\{\nu > 1\}}}{\log \nu} + 2 \frac{\mathbb{1}_{\{p_1 > 0\}}}{|\log \mu|} + \frac{\mathbb{1}_{\{p_1=0, p_2 > 0\}}}{|\log p_2^*|}. \tag{7.4.2}$$

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 $o_r(\log n)$. 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 $\text{CM}_n(\mathbf{d})$, 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

$$\frac{\text{diam}(\text{CM}_n(\mathbf{d}))}{\log n} \xrightarrow{\mathbb{P}} \frac{1}{\log(r-1)}. \quad (7.4.3)$$

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 $\text{diam}(\text{CM}_n(\mathbf{d})) \gg \log n$. 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 $\text{GRG}_n(\mathbf{w})$, where (recall [Volume 1, Exercise 6.1])

$$w_i = \frac{n\lambda}{n-\lambda}. \quad (7.4.4)$$

Clearly, $\mathbf{w} = (n\lambda/(n-\lambda))_{i \in [n]}$ 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

$$\frac{\text{diam}(\text{ER}_n(\lambda/n))}{\log n} \xrightarrow{\mathbb{P}} \frac{1}{\log \lambda} + \frac{2}{|\log \mu_\lambda|}. \quad (7.4.5)$$

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 $\log_\nu n$ 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 μ 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

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 $\log_\nu n$. 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 $d_{\min} \geq 3$. Then, $\text{CM}_n(\mathbf{d})$ 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 $\text{CM}_n(\mathbf{d})$ 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 $d_{\min} \geq 3$, we know by Theorem 4.15 that $\text{CM}_n(\mathbf{d})$ is whp connected. The main result is as follows:

Theorem 7.17 (Diameter of $\text{CM}_n(\mathbf{d})$ for $\tau \in (2, 3)$) *Suppose that the empirical distribution function F_n of the degrees $\mathbf{d} = (d_i)_{i \in [n]}$ satisfies Condition 1.5(a)-(b) and that (7.1.2) holds. Assume further that $d_{\min} = \min_{i \in [n]} d_i \geq 3$ and $p_{d_{\min}} =$*

$\mathbb{P}(D = d_{\min}) > 0$. Then,

$$\frac{\text{diam}(\text{CM}_n(\mathbf{d}))}{\log \log n} \xrightarrow{\mathbb{P}} \frac{2}{|\log(\tau - 2)|} + \frac{2}{\log(d_{\min} - 1)}. \quad (7.4.6)$$

When comparing Theorem 7.17 to Theorem 7.2, we see that for $d_{\min} \geq 3$, 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 $d_{\min} \geq 3$, 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_{\min} . As a result, we will see that the expected number of vertices whose $(1 - \varepsilon) \log \log n / \log(d_{\min} - 1)$ neighborhood only contains vertices with degree d_{\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 $2 \log \log n / \log(d_{\min} - 1)$.

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 $d_i = d_{\min}$, 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 $M_k \xrightarrow{\mathbb{P}} \infty$ for $k = (1 - \varepsilon) \log \log n / \log(d_{\min} - 1)$. 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 $\text{CM}_n(\mathbf{d})$ satisfy that $d_{\min} \geq 3$, $n_{d_{\min}} > d_{\min}(d_{\min} - 1)^{k-1}$. Then, for all $k \geq 1$,*

$$\mathbb{E}[M_k] = n_{d_{\min}} \prod_{i=1}^{d_{\min}(d_{\min}-1)^{k-1}} \frac{d_{\min}(n_{d_{\min}} - (i-1))}{\ell_n - 2i + 1}, \quad (7.4.7)$$

and, for k such that $d_{\min}(d_{\min} - 1)^{k-1} \leq \ell_n/8$,

$$\mathbb{E}[M_k^2] \leq \mathbb{E}[M_k]^2 + \mathbb{E}[M_k] \left[\frac{d_{\min}}{d_{\min} - 2} (d_{\min} - 1)^k + \frac{2n_{d_{\min}} d_{\min}^2 (d_{\min} - 1)^{2k}}{(d_{\min} - 2)\ell_n} \right]. \quad (7.4.8)$$

Proof We start by proving (7.4.7). We note that each vertex of degree d_{\min} has

the same probability of being minimally- k connected, and that there are precisely $n_{d_{\min}}$ vertices of degree d_{\min} , so that

$$\mathbb{E}[M_k] = n_{d_{\min}} \mathbb{P}(v \text{ is minimally-}k\text{-connected}). \quad (7.4.9)$$

Vertex v with $d_v = d_{\min}$ 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_{\min} , and no cycles occur in $B_k(v)$. When $i - 1$ half-edges are paired to distinct vertices of degree d_{\min} , then the probability that the i th half-edge is again paired to a distinct vertex of degree d_{\min} equals

$$\frac{d_{\min}(n_{d_{\min}} - (i - 1))}{\ell_n - 2i + 1}. \quad (7.4.10)$$

Since for v to be minimally- k -connected, there are $d_{\min}(d_{\min} - 1)^{k-1}$ half-edges that need to be paired to distinct vertices of degree d_{\min} , this proves (7.4.7).

To prove (7.4.8), we note that

$$\mathbb{E}[M_k^2] = \sum_{v_1, v_2 \in [n]} \mathbb{P}(v_1, v_2 \text{ are minimally-}k\text{-connected}). \quad (7.4.11)$$

We split the above probability depending on whether $B_k(o_1) \cap B_k(o_2) = \emptyset$ or not. The contribution to $\mathbb{E}[M_k^2]$ due to $B_k(o_1) \cap B_k(o_2) = \emptyset$ is, similarly to the proof of (7.4.7), equal to

$$n_{d_{\min}}(n_{d_{\min}} - i_{k-1}) \prod_{i=1}^{2i_k} \frac{d_{\min}(n_{d_{\min}} - (i - 1))}{\ell_n - 2i + 1}, \quad (7.4.12)$$

where we abbreviate $i_k = d_{\min}(d_{\min} - 1)^{k-1}$ and note that $n_{d_{\min}} - i_{k-1} > 0$, since, by assumption, $n_{d_{\min}} > d_{\min}(d_{\min} - 1)^{k-1}$.

We use that $i \mapsto d_{\min}(n_{d_{\min}} - (i - 1))/(\ell_n - 2i + 1)$ is decreasing since

$$\frac{(n_{d_{\min}} - (i - 1))}{\ell_n - 2i + 1} \geq \frac{n_{d_{\min}} - i}{\ell_n - 2i - 1} \quad (7.4.13)$$

precisely when $\ell_n \geq 2n_{d_{\min}} + 1$, and which is true since $\ell_n \geq d_{\min}n_{d_{\min}} \geq 3n_{d_{\min}}$. Therefore, the contribution to $\mathbb{E}[M_k^2]$ from v_1 and v_2 satisfying $B_k(v_1) \cap B_k(v_2) = \emptyset$ is at most

$$n_{d_{\min}}^2 \left(\prod_{i=1}^{i_k} \frac{d_{\min}(n_{d_{\min}} - (i - 1))}{\ell_n - 2i + 1} \right)^2 = \mathbb{E}[M_k^2], \quad (7.4.14)$$

which is the first contribution to the r.h.s. of (7.4.8).

We are left to deal with the contribution to $\mathbb{E}[M_k^2]$ from v_1 and v_2 such that $B_k(v_1) \cap B_k(v_2) \neq \emptyset$. When v_1 is k -minimally connected,

$$\begin{aligned} |B_k(v_1)| &= 1 + \sum_{l=1}^k d_{\min}(d_{\min} - 1)^{l-1} \\ &= 1 + d_{\min} \frac{(d_{\min} - 1)^k - 1}{d_{\min} - 2} \leq d_{\min} \frac{(d_{\min} - 1)^k}{d_{\min} - 2}. \end{aligned} \quad (7.4.15)$$

Therefore, the contribution due to $v_2 \in B_k(v_1)$ is bounded by

$$\mathbb{E}[M_k] \frac{d_{\min}}{d_{\min} - 2} (d_{\min} - 1)^k, \quad (7.4.16)$$

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 $d_{\min}(d_{\min} - 1)^k$ half-edges in $B_k(v_1)$ needs to be connected to one of the $d_{\min}(d_{\min} - 1)^{l-k}$ half-edges in $B_{l-k}(v_2)$, where $l = \text{dist}_{\text{CM}_n(d)}(v_1, v_2) \in [2k] \setminus [k]$. Conditionally on v_1 being k -minimally connected and v_2 being $l-k$ -minimally connected, the probability that this occurs is at most

$$\frac{d_{\min}(d_{\min} - 1)^k d_{\min}(d_{\min} - 1)^{l-k}}{\ell_n - 2i_k - 2i_{l-k} + 1} \leq \frac{2d_{\min}(d_{\min} - 1)^k d_{\min}(d_{\min} - 1)^{l-k-1}}{\ell_n}, \quad (7.4.17)$$

where, in the last inequality, we used that $d_{\min}(d_{\min} - 1)^{k-1} \leq \ell_n/8$. Therefore, this contribution is bounded by

$$\mathbb{E}[M_k] \sum_{l=k+1}^{2k} \mathbb{E}[M_{l-k}] \frac{2d_{\min}(d_{\min} - 1)^k d_{\min}(d_{\min} - 1)^{l-k-1}}{\ell_n}. \quad (7.4.18)$$

We bound $\mathbb{E}[M_{l-k}] \leq n_{d_{\min}}$ and sum

$$\sum_{l=k+1}^{2k} (d_{\min} - 1)^{l-k-1} \leq \frac{(d_{\min} - 1)^{2k}}{d_{\min} - 2}, \quad (7.4.19)$$

to arrive at the third and final contribution to the r.h.s. of (7.4.8). \square

To complete the proof of the lower bound on the diameter, we fix $\varepsilon > 0$ sufficiently small, and take $k^* = (1 - \varepsilon) \frac{\log \log n}{\log(d_{\min} - 1)}$. Clearly,

$$d_{\min}(d_{\min} - 1)^{k^*-1} \leq (\log n)^{1-\varepsilon} \leq \ell_n/8, \quad (7.4.20)$$

so that, in particular, we may use Lemma 7.18.

We note that $n_{d_{\min}}/n \rightarrow p_{d_{\min}}$ by Condition 1.5(a) and $p_{d_{\min}} > 0$ by assumption. Therefore, by Conditions 1.5(a)-(b), $d_{\min}n_{d_{\min}}/\ell_n \rightarrow \lambda_{d_{\min}}$, where we define $\lambda_{d_{\min}} = d_{\min}p_{d_{\min}}/\mathbb{E}[D]$. By (7.4.7) in Lemma 7.18,

$$\mathbb{E}[M_k] \geq n(p_{d_{\min}} - \delta)(\lambda_{d_{\min}} - \delta)^{d_{\min}(d_{\min} - 1)^{k-1}}. \quad (7.4.21)$$

As a result, $\mathbb{E}[M_{k^*}] \geq n(p_{d_{\min}} - \delta)(\lambda_{d_{\min}} - \delta)^{(\log n)^{1-\varepsilon}}$. Further, by (7.4.8) in Lemma 7.18,

$$\text{Var}(M_{k^*}) = o(\mathbb{E}[M_{k^*}]^2), \quad (7.4.22)$$

so that

$$M_{k^*}/\mathbb{E}[M_{k^*}] \xrightarrow{\mathbb{P}} 1. \quad (7.4.23)$$

We conclude that, whp, $M_{k^*} \geq n^{1-o(1)}$. Since each minimally- k^* -connected vertex

uses up at most

$$1 + \sum_{l=1}^{k^*} d_{\min}(d_{\min} - 1)^{l-1} = n^{o(1)} \tag{7.4.24}$$

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 \mathcal{A}_{12} denote the event that v_1, v_2 are minimally- k^* -connected with their k^* -neighborhoods being disjoint.

Conditionally on \mathcal{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 $\partial B_{k^*}(v_2)$ 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 $\tilde{\mathbf{d}}$ 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 $\varepsilon > 0$,

$$\mathbb{P}\left(\text{dist}_{\text{CM}_n(\mathbf{d})}(\partial B_{k^*}(v_1), \partial B_{k^*}(v_2)) \leq (1 - \varepsilon) \frac{2 \log \log n}{|\log(\tau - 2)|} \mid \mathcal{A}_{12}\right) = o(1). \tag{7.4.25}$$

Therefore, whp,

$$\begin{aligned} \text{diam}(\text{CM}_n(\mathbf{d})) &\geq (1 - \varepsilon) \frac{2 \log \log n}{|\log(\tau - 2)|} + 2k^* \\ &= (1 - \varepsilon) \log \log n \left[\frac{2}{|\log(\tau - 2)|} + \frac{2}{\log(d_{\min} - 1)} \right]. \end{aligned} \tag{7.4.26}$$

Since $\varepsilon > 0$ is arbitrary, this proves the lower bound on $\text{diam}(\text{CM}_n(\mathbf{d}))$ in Theorem 7.17. \square

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 $k_n = \log \log n [2/|\log(\tau - 2)| + 2/\log(d_{\min} - 1)]$,

$$\mathbb{P}(\exists v_1, v_2 \in [n]: \text{dist}_{\text{CM}_n(\mathbf{d})}(v_1, v_2) \geq (1 + \varepsilon)k_n) = o(1). \tag{7.4.27}$$

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 (7.4.27) follows when we can show that

$$\mathbb{P}(\exists v: \text{dist}_{\text{CM}_n(\mathbf{d})}(v, \text{Core}_n) \geq (1 + \varepsilon) \log \log n / \log(d_{\min} - 1)) = o(1), \tag{7.4.28}$$

and we will argue that, uniformly in $v \in [n]$,

$$\mathbb{P}(\text{dist}_{\text{CM}_n(\mathbf{d})}(v, \text{Core}_n) \geq (1 + \varepsilon) \log \log n / \log(d_{\min} - 1)) = o(1/n), \tag{7.4.29}$$

which would prove (7.4.27).

To prove (7.4.29), it is convenient to explore the neighborhood of the vertex v by only pairing up the first d_{\min} half-edges incident to v and the $d_{\min} - 1$ 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 $k^* = (1 + \varepsilon/2) \log \log n / \log (d_{\min} - 1)$ is close to

$$(d_{\min} - 1)^k \approx (\log n)^{1+\varepsilon/2}. \tag{7.4.30}$$

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 Core_n quickly, meaning in $o(\log \log n)$ steps. More precisely, the main ingredient in the upper bound is the statement that, for

$$k^* = (1 + \varepsilon/2) \log \log n / \log (d_{\min} - 1)$$

and whp, the k -exploration tree connects to Core_n whp in less than k^* . We omit further details. \square

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

$$\lim_{x \rightarrow \infty} x^{\tau-1} [1 - F](x) = c. \tag{7.5.1}$$

We study the configuration model $\text{CM}_n(\mathbf{d})$ where the degrees $\mathbf{d} = (d_i)_{i \in [n]}$ 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

$$P_i = Z_i / Z, \tag{7.5.2}$$

where $Z_i = \Gamma_i^{-1/(\tau-1)}$ and $\Gamma_i = \sum_{j=1}^i E_j$ 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, $M_{P,k} = (B_1, B_2, \dots)$, 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 $M_{P,D_1}^{(1)}$ and $M_{P,D_2}^{(2)}$ be two random variables which are conditionally independent given $P = (P_i)_{i \geq 1}$. In terms of this notation, the main result on distances in $\text{CM}_n(\mathbf{d})$ when the degrees have infinite mean is the following:

Theorem 7.19 (Distances in $\text{CM}_n(\mathbf{d})$ with i.i.d. infinite mean degrees) *Fix*

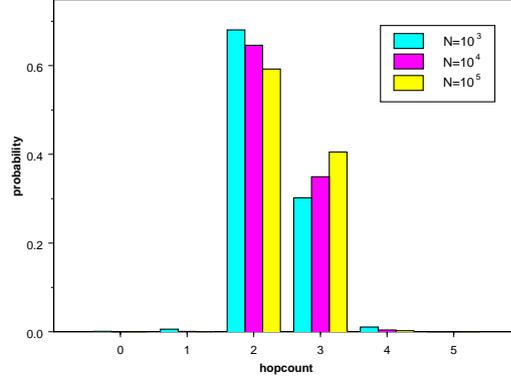


Figure 7.2 Empirical probability mass function of the $\text{dist}_{\text{CM}_n(\mathbf{d})}(o_1, o_2)$ for $\tau = 1.8$ and $n = 10^3, 10^4, 10^5$.

$\tau \in (1, 2)$ in (7.5.1) and let $(d_i)_{i \in [n]}$ be a sequence of i.i.d. copies of D . Then, $\text{CM}_n(\mathbf{d})$ satisfies

$$\lim_{n \rightarrow \infty} \mathbb{P}(\text{dist}_{\text{CM}_n(\mathbf{d})}(o_1, o_2) = 2) = 1 - \lim_{n \rightarrow \infty} \mathbb{P}(\text{dist}_{\text{CM}_n(\mathbf{d})}(o_1, o_2) = 3) = p_F \in (0, 1). \tag{7.5.3}$$

The probability p_F can be identified as the probability that $M_{P, D_1}^{(1)}$ and $M_{P, D_2}^{(2)}$ have an identical outcome, i.e., there is an outcome that occurs both in $M_{P, D_1}^{(1)}$ and $M_{P, D_2}^{(2)}$, where D_1 and D_2 are two i.i.d. copies of D .

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 $\text{dist}_{\text{CM}_n(\mathbf{d})}(o_1, o_2) = 1$ 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, $\ell_n \geq n^{1/(\tau-1)-\varepsilon}$. Therefore, on the event that $\ell_n \geq n^{1/(\tau-1)-\varepsilon}$ and $d_1 \leq \log n$ and $d_2 \leq \log n$, the probability that $\text{dist}_{\text{CM}_n(\mathbf{d})}(o_1, o_2) = 1$ is bounded above by

$$\frac{(\log n)^2}{n^{1/(\tau-1)-\varepsilon}} = o(1). \tag{7.5.4}$$

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, $M_{P, d_1}^{(1)} = (B_1^{(1)}, B_2^{(1)}, \dots)$, where $B_i^{(1)}$ is the number of edges between vertex i and the vertex with degree $d_{(n+1-i)}$. 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 $\xi_k k^{1/(\tau-1)} \xrightarrow{\mathbb{P}} 1$ as $k \rightarrow \infty$. 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 \mathbb{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,

$$\mathbb{P}_n(v_i \text{ not directly connected to } v_j) \leq e^{-\frac{d_{(n+1-i)}d_{(n+1-j)}}{2\ell_n}}. \tag{7.5.5}$$

Moreover, $d_{(n+1-i)}, d_{(n+1-j)} \geq n^{1/(\tau-1)-\varepsilon}$ with high probability for n sufficiently large and any $\varepsilon > 0$, while whp $\ell_n \leq n^{1/(\tau-1)+\varepsilon}$. As a result, whp,

$$\mathbb{P}_n(v_i \text{ not directly connected to } v_j) \leq e^{-n^{1/(\tau-1)-3\varepsilon}}. \tag{7.5.6}$$

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

$$\mathbb{P}_n(\text{dist}_{\text{CM}_n(\mathbf{d})}(o_1, o_2) \leq 3) = 1 - o(1). \tag{7.5.7}$$

This completes the proof. □

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 $(R_a)_{a \in (-1, 0]}$.

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 $(Z_k/\mathbb{E}[D]\nu^{k-1})_{k \geq 1}$ 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]):

$$\lim_{n \rightarrow \infty} \frac{Z_n}{\mathbb{E}[D]\nu^{n-1}} = \mathcal{W} \quad a.s. \tag{7.5.8}$$

In the theorem below we need two independent copies \mathcal{W}_1 and \mathcal{W}_2 of \mathcal{W} .

Theorem 7.20 (Limit law for typical distance in $\text{CM}_n(\mathbf{d})$) *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 $a_k = \lfloor \log_\nu k \rfloor - \log_\nu k \in (-1, 0]$. Then, $\text{CM}_n(\mathbf{d})$ satisfies that there exist random variables $(R_a)_{a \in (-1, 0]}$ such that as $n \rightarrow \infty$ and for all $k \in \mathbb{Z}$,

$$\begin{aligned} \mathbb{P}(\text{dist}_{\text{CM}_n(\mathbf{d})}(o_1, o_2) - \lfloor \log_\nu n \rfloor = k \mid \text{dist}_{\text{CM}_n(\mathbf{d})}(o_1, o_2) < \infty) \\ = \mathbb{P}(R_{a_n} = k) + o(1). \end{aligned} \tag{7.5.10}$$

The random variables $(R_a)_{a \in (-1, 0]}$ can be identified as

$$\mathbb{P}(R_a > k) = \mathbb{E}[\exp\{-\kappa \nu^{a+k} \mathcal{W}_1 \mathcal{W}_2\} | \mathcal{W}_1 \mathcal{W}_2 > 0,], \quad (7.5.11)$$

where \mathcal{W}_1 and \mathcal{W}_2 are independent limit copies of \mathcal{W} in (7.5.8) and where $\kappa = \mathbb{E}[D](\nu - 1)^{-1}$.

In words, Theorem 7.20 states that for $\tau > 3$, the graph distance $\text{dist}_{\text{CM}_n(\mathbf{d})}(o_1, o_2)$ 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 \mathcal{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 \mathcal{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, $\mathbb{E}[D] = r$, $\nu = r - 1$, and $\mathcal{W} = 1$ a.s. In particular, $\mathbb{P}(\text{dist}_{\text{CM}_n(\mathbf{d})}(o_1, o_2) < \infty) = 1 + o(1)$. Therefore,

$$\mathbb{P}(R_a > k) = \exp\left\{-\frac{r}{r-2}(r-1)^{a+k}\right\}, \quad (7.5.12)$$

and $\text{dist}_{\text{CM}_n(\mathbf{d})}(o_1, o_2)$ 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

$$p_j^* = p(1-p)^{j-1}, \quad \text{so that} \quad p_j = \frac{1}{j c_p} p(1-p)^{j-2}, \quad \forall j \geq 1, \quad (7.5.13)$$

and c_p is a normalization constant. For $p > \frac{1}{2}$, the law of \mathcal{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 $\text{CM}_n(\mathbf{d})$ 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

$$x^{-\tau+1-C(\log x)^{\gamma-1}} \leq 1 - F(x) \leq x^{-\tau+1+C(\log x)^{\gamma-1}}, \quad \text{for large } x. \quad (7.5.14)$$

The condition in (7.5.14) is such that the results in Theorem 7.14 apply. Then, we can identify the fluctuations of the typical graph distance in $\text{CM}_n(\mathbf{d})$ as follows:

Theorem 7.21 (Fluctuations graph distance $\text{CM}_n(\mathbf{d})$ 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, $\text{CM}_n(\mathbf{d})$ satisfies that there exist random variables $(R_a)_{a \in (-1, 0]}$ such that, as $n \rightarrow \infty$ and for all $l \in \mathbb{Z}$,

$$\begin{aligned} \mathbb{P}\left(\text{dist}_{\text{CM}_n(\mathbf{d})}(o_1, o_2) = 2 \left\lfloor \frac{\log \log n}{|\log(\tau - 2)|} \right\rfloor + l \mid \text{dist}_{\text{CM}_n(\mathbf{d})}(o_1, o_2) < \infty\right) \\ = \mathbb{P}(R_{a_n} = l) + o(1), \end{aligned} \tag{7.5.15}$$

where $a_n = \lfloor \frac{\log \log n}{|\log(\tau - 2)|} \rfloor - \frac{\log \log n}{|\log(\tau - 2)|} \in (-1, 0]$. Here, the random variables $(R_a)_{a \in (-1, 0]}$ are given by

$$\mathbb{P}(R_a > l) = \mathbb{P}\left(\min_{s \in \mathbb{Z}} [(\tau - 2)^{-s} Y_1 + (\tau - 2)^{s - c_l} Y_2] \leq (\tau - 2)^{\lfloor l/2 \rfloor + a} \mid Y_1 Y_2 > 0\right),$$

where $c_l = 1$ 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 $\text{dist}_{\text{CM}_n(\mathbf{d})}(o_1, o_2)$ 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 $\beta_n \in (1, 1/(\tau - 1)]$ such that $F_n(x) = 1$ for $x \geq n^{\beta_n(1+\varepsilon)}$ for all $\varepsilon > 0$, while for all $x \leq n^{\beta_n(1-\varepsilon)}$,*

$$1 - F_n(x) = \frac{L_n(x)}{x^{\tau-1}}, \tag{7.5.16}$$

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

$$e^{-C_1(\log x)^\eta} \leq L_n(x) \leq e^{C_1(\log x)^\eta}. \tag{7.5.17}$$

Theorem 7.23 (Fluctuations distances $\text{CM}_n(\mathbf{d})$ 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*

$$\max\{d_{\text{TV}}(F_n, F), d_{\text{TV}}(F_n^*, F^*)\} \leq n^{-\kappa\beta_n}. \tag{7.5.18}$$

When $\beta_n \rightarrow 1/(\tau - 1)$, we further require that the limit random variable Y in Theorem 7.14 has no pointmass on $(0, \infty)$. Then,

$$\text{dist}_{\text{CM}_n(\mathbf{d})}(o_1, o_2) - 2 \frac{\log \log(n^{\beta_n})}{|\log(\tau - 2)|} - \frac{1}{\beta_n(\tau - 3)} \tag{7.5.19}$$

is a tight sequence of random variables.

Which of the two terms in (7.5.19) dominates depends sensitively on the choice of β_n . When $\beta_n \rightarrow \beta \in (0, 1/(\tau - 1)]$, 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 $\beta_n = \Theta(1/\log \log n)$. Graph distances are always much smaller than $\log n$, the boundary point being $\beta_n = \Theta(1/\log n)$, 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

$$Y_i^{(n)} = (\tau - 2)^{t_n} \log(Z_{t_n}^{(n;i)}) \tag{7.5.20}$$

for an appropriate t_n , which is the first t for which $\max\{Z_{t_n}^{(n;1)}, Z_{t_n}^{(n;2)}\} \geq n^{\rho_n}$ for some appropriate $\rho_n < \beta_n$. As it turns out, $(Y_1^{(n)}, Y_2^{(n)}) \xrightarrow{d} (Y_1, Y_2)$. Further, let $b_n^{(i)}$ be the fractional part of

$$\frac{\log \log(n^{\beta_n}) - \log Y_i^{(n)}}{|\log(\tau - 2)|}.$$

Then the most precise result available for graph distances in $\text{CM}_n(\mathbf{d})$ for infinite-variance degrees is the following theorem:

Theorem 7.24 (Fluctuations distances $\text{CM}_n(\mathbf{d})$ for truncated infinite-variance degrees) *Assume that the assumptions in Theorem 7.23 hold. Then,*

$$\begin{aligned} \text{dist}_{\text{CM}_n(\mathbf{d})}(o_1, o_2) &- 2 \frac{\log \log(n^{\beta_n})}{|\log(\tau - 2)|} - b_n^{(1)} - b_n^{(2)} - \left\lceil \frac{\frac{1}{\beta_n} - (\tau - 2)^{b_n^{(1)}} - (\tau - 2)^{b_n^{(2)}}}{3 - \tau} \right\rceil \\ &\xrightarrow{d} -1 - \frac{Y_1 Y_2}{|\log(\tau - 2)|}. \end{aligned} \tag{7.5.21}$$

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 $\text{CM}_n(\mathbf{d})$ where, instead of truncating the degrees, we only allow for paths through vertices of maximal degree n^{β_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 ?. 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 $\mathbb{P}(\text{dist}_{\text{CM}_n(d)}(o_1, o_2) = 2)$ 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^α is considered, where α 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^{β_n} as in van der Hofstad and Komjáthy (2017) for the $\tau \in (2, 3)$ case.

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 $\mathbb{P}(0 < Y \leq \varepsilon)$ 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 $d_{\text{TV}}(p^{*(n)}, p^*) = 0$.*

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 (Γ_1 is a complete graph) *Use Lemma 7.11 and $\alpha > 1/2$ to show that, whp, Γ_1 in (7.2.36) forms a complete graph, i.e., whp, every $i, j \in \Gamma_1$ are direct neighbors in $\text{CM}_n(\mathbf{d})$.*

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 $\mathbb{E}[X] = \infty$ when the conditions in Theorem 7.14 are satisfied. Extend this to show that $\mathbb{E}[X^s] = \infty$ for every $s > \alpha \in (0, 1)$.*

Exercise 7.8 (Conditions in Theorem 7.14 for individuals with infinite line of descent) *Prove that $p^{(\infty)}$ in [Volume 1, (3.4.2)] satisfies the conditions in Theorem 7.14 with the function $x \mapsto \gamma^*(x)$, given by $\gamma^*(x) = \gamma(x) + c/\log x$.*

Exercise 7.9 (Convergence for $Z_n + 1$) *Show that, under the conditions of Theorem 7.14, also $\alpha^n \log(Z_n + 1)$ 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 size of the longest cycle of the 2-regular graph?*

Exercise 7.12 (Parameters for $\text{ER}_n(\lambda/n)$) *Fix $\lambda > 1$. Prove that $\nu = \lambda$ and $\mu = \mu_\lambda$, where $\mu_\lambda \in [0, 1)$ is the dual parameter, i.e., the unique μ satisfying that*

$$\mu e^{-\mu} = \lambda e^{-\lambda}. \quad (7.7.1)$$

Exercise 7.13 (Typical distance is at least 2 whp) *Complete the argument that $\mathbb{P}(\text{dist}_{\text{CM}_n(\mathbf{d})}(o_1, o_2) = 1) = o(1)$ in the proof of Theorem 7.19.*

Exercise 7.14 (Typical distance equals 2 whp for $\tau = 1$) *Let the $(d_i)_{i \in [n]}$ be a sequence of i.i.d. copies of D with distribution function F satisfying that $x \mapsto [1 - F](x)$ is slowly varying at ∞ . Prove that $\text{CM}_n(\mathbf{d})$ satisfies that $\text{dist}_{\text{CM}_n(\mathbf{d})}(o_1, o_2) \xrightarrow{\mathbb{P}} 2$.*

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 $\text{dist}_{\text{CM}_n(\mathbf{d})}(o_1, o_2) < \infty$, along the subsequence $n_k = \lfloor n_1 \nu^{k-1} \rfloor$, the sequence of random variables $\text{dist}_{\text{CM}_n(\mathbf{d})}(o_1, o_2) - \lfloor \log_\nu n_k \rfloor$ converges in distribution to $R_{a_{n_1}}$ as $k \rightarrow \infty$.*

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 $\text{dist}_{\text{CM}_n(\mathbf{d})}(o_1, o_2) < \infty$, the random variable $\text{dist}_{\text{CM}_n(\mathbf{d})}(o_1, o_2)$ is in between $(1 \pm \varepsilon) \log_\nu n$ for any $\varepsilon > 0$;*
- (ii) *conditionally on $\text{dist}_{\text{CM}_n(\mathbf{d})}(o_1, o_2) < \infty$, the random variables $\text{dist}_{\text{CM}_n(\mathbf{d})}(o_1, o_2) - \log_\nu n$ form a tight sequence, i.e.,*

$$\lim_{K \rightarrow \infty} \limsup_{n \rightarrow \infty} \mathbb{P}(|\text{dist}_{\text{CM}_n(\mathbf{d})}(o_1, o_2) - \log_\nu n| \leq K \mid \text{dist}_{\text{CM}_n(\mathbf{d})}(o_1, o_2) < \infty) = 1. \quad (7.7.2)$$

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 $A_n = \Theta(n^{\beta_n})$. Show that Condition 7.22 holds when*

$$1 - F_n(x) = \frac{e^{-x/A_n}}{x^{\tau-1}}, \quad (7.7.3)$$

with $\tau \in (2, 3)$.

Exercise 7.18 (Paths through vertices with degree constraints) *Suppose that Condition 7.22 holds for $\beta_n = 1/(\tau - 1)$. 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 $\alpha_n \gg 1/\log n$, 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. Remarkably, the setting where $\tau \in (2, 3)$ (corresponding to $\delta \in (-m, 0)$) is better understood than the setting where $\tau > 3$.

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 than they are in the corresponding generalized random graphs and 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 less complete.

Throughout this chapter, we work with the preferential attachment model defined in [Volume 1, Section 8.2], which we denote by $(\text{PA}_n^{(m,\delta)})_{n \geq 1}$, 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, (??)] for $m = 1$, and as described on page [Volume 1, 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 $(\text{PA}_n^{(m,\delta)}(b)(n))_{n \geq 1}$, in which the $m = 1$ model does not have any self-loops.

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 lower bounds on distances. In Section 8.5, we prove the matching upper bounds on graph distances. In Section 8.6, we discuss the diameter of preferential attachment models for $m \geq 2$. In Section 8.7, we discuss further results about distances in preferential attachment models. We close this chapter with notes and discussion in Section 8.8, and with exercises in Section 8.9.

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 *height* of the tree $\text{PA}_n^{(1,\delta)}$, where the height of a tree T is defined as

$$\text{height}(T) = \max_{v \in [n]} \text{dist}_T(1, v), \quad (8.1.1)$$

followed by its *diameter*:

Theorem 8.1 (Typical distance in scale-free trees) *Fix $m = 1$ and $\delta > -1$. Then, as $n \rightarrow \infty$,*

$$\frac{\text{dist}_{\text{PA}_n^{(1,\delta)}}(1, o)}{\log n} \xrightarrow{\mathbb{P}} \frac{1 + \delta}{2 + \delta}, \quad \frac{\text{height}(\text{PA}_n^{(1,\delta)})}{\log n} \xrightarrow{\mathbb{P}} \frac{2(1 + \delta)}{(2 + \delta)}. \quad (8.1.2)$$

Theorem 8.2 (Diameter of scale-free trees) *Fix $m = 1$ and $\delta > -1$. Let γ be the non-negative solution of*

$$\gamma + (1 + \delta)(1 + \log \gamma) = 0. \quad (8.1.3)$$

Then

$$\frac{\text{diam}_{\text{PA}_n^{(1,\delta)}}(o_1, o_2)}{\log n} \xrightarrow{\mathbb{P}} \frac{2(1 + \delta)}{(2 + \delta)\gamma}, \quad \frac{\text{diam}(\text{PA}_n^{(1,\delta)})}{\log n} \xrightarrow{\mathbb{P}} \frac{2(1 + \delta)}{(2 + \delta)\gamma}. \quad (8.1.4)$$

The proof of Theorems 8.1–8.2 rely on the fact that $\text{PA}_n^{(1,\delta)}$ consists of a collection of trees with precisely one self-loop. There is a close analogy between these trees and so-called *uniform recursive trees*. In uniform recursive trees, we grow a

tree such that at time 1, we have a unique 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 *random plane-oriented recursive tree*. For a uniform recursive tree of size t , it is proved by [Pittel \(1994\)](#) that the maximal distance between the root and any other vertex is with high probability equal to $\frac{1}{2\gamma} \log n(1 + o(1))$, where γ satisfies (8.1.3) 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 $\frac{1}{\gamma} \log n(1 + o(1))$. We shall make use of similar arguments below to prove [Theorem 8.2](#). Exercises 8.1–8.3 explore properties of the constant γ .

In the proof of [Theorem 8.2](#) it will be useful to work with $\text{PA}_n^{(1,\delta)}(b)$ instead of $\text{PA}_n^{(1,\delta)}$, for which a similar result holds:

Theorem 8.3 (Distances in scale-free trees $\text{PA}_n^{(1,\delta)}(b)$) *Fix $m = 1$ and $\delta > -1$, and let γ be the solution of (8.1.3). Then*

$$\frac{\text{dist}_{\text{PA}_n^{(1,\delta)}(b)}(1, o)}{\log n} \xrightarrow{\mathbb{P}} \frac{1 + \delta}{2 + \delta}, \quad \frac{\text{height}(\text{PA}_n^{(1,\delta)}(b))}{\log n} \xrightarrow{a.s.} \frac{2(1 + \delta)}{(2 + \delta)}. \quad (8.1.5)$$

and

$$\frac{\text{dist}_{\text{PA}_n^{(1,\delta)}(b)}(o_1, o_2)}{\log n} \xrightarrow{\mathbb{P}} \frac{2(1 + \delta)}{(2 + \delta)\gamma}, \quad \frac{\text{diam}(\text{PA}_n^{(1,\delta)}(b))}{\log n} \xrightarrow{\mathbb{P}} \frac{2(1 + \delta)}{(2 + \delta)\gamma}. \quad (8.1.6)$$

We start by proving the upper bound in [Theorem 8.3](#). We remark that the almost sure limit of the height in [Theorem 8.3](#) does not depend on the precise starting configuration of the graph $\text{PA}_2^{(1,\delta)}(b)$.

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 t and the root 1. Before stating the result, we need some more notation. We write $t \rightarrow s$ when in $(\text{PA}_n^{(1,\delta)}(b))_{n \geq 1}$ one of the edges of vertex t is connected to vertex s . Note that for this to happen, we need that $t > s$. For $s_1 = t > s_2 > \dots > s_k = 1$, and denoting $\vec{s}_k = (s_1, s_2, \dots, s_k)$, we write

$$E_{\vec{s}_k} = \bigcap_{i=1}^{k-1} \{s_i \rightarrow s_{i+1}\}. \quad (8.1.7)$$

For a configuration of $\text{PA}_n^{(1,\delta)}(b)$, we let $\text{dist}(t, s)$ denote the unique value of k such that $t = s_1 \rightarrow s_2 \rightarrow \dots \rightarrow s_{k-1} \rightarrow s_k = s$. Then the probability mass function of $\text{dist}(t, s)$ can be explicitly identified as follows:

Proposition 8.4 (Distribution of $\text{dist}(t, s)$ in $\text{PA}_n^{(1,\delta)}(b)$) *Fix $m = 1$ and $\delta > -1$.*

Then, for all $t > s$,

$$\mathbb{P}(\text{dist}(t, s) = k) = \left(\frac{1 + \delta}{2 + \delta}\right)^k \frac{\Gamma(t + \frac{1}{2+\delta})\Gamma(s)}{\Gamma(s + \frac{1}{2+\delta})\Gamma(t + 1)} \sum_{\vec{s}_k} \prod_{i=1}^{k-1} \frac{1}{s_i}, \tag{8.1.8}$$

where the sum is over ordered vectors $\vec{s}_k = (s_0, \dots, s_k)$ of length $k + 1$ with $s_0 = t$ and $s_k = s$. Further,

$$\mathbb{P}(t \longrightarrow s_1 \longrightarrow \dots \longrightarrow s_{k-1} \longrightarrow s) = \left(\frac{1 + \delta}{2 + \delta}\right)^k \frac{\Gamma(t + \frac{1}{2+\delta})\Gamma(s)}{\Gamma(s + \frac{1}{2+\delta})\Gamma(t + 1)} \prod_{i=1}^{k-1} \frac{1}{s_i}, \tag{8.1.9}$$

Proof of Proposition 8.4. Since the path between vertex t and s is *unique*, we obtain

$$\mathbb{P}(\text{dist}(t, s) = k) = \sum_{\vec{s}_k} \mathbb{P}\left(\bigcap_{i=0}^{k-1} \{s_i \longrightarrow s_{i+1}\}\right), \tag{8.1.10}$$

where again the sum is over all ordered vectors $\vec{s}_k = (s_0, \dots, s_k)$ of length $k + 1$ with $s_0 = t$ and $s_k = s$. Therefore, (8.1.8) follows immediately from (8.1.9).

We claim that the events $\{s_i \longrightarrow s_{i+1}\}$ are *independent*, i.e., for every sequence $\vec{s}_k = (s_0, \dots, s_k)$

$$\mathbb{P}\left(\bigcap_{i=0}^{k-1} \{s_i \longrightarrow s_{i+1}\}\right) = \prod_{i=0}^{k-1} \mathbb{P}(s_i \longrightarrow s_{i+1}). \tag{8.1.11}$$

We prove the independence in (8.1.11) by induction on k . For $k = 0$, there is nothing to prove, and this initializes the induction hypothesis. To advance the induction hypothesis in (8.1.11), we condition on $\text{PA}_{s_0-1}^{(1,\delta)}(b)$ to obtain

$$\begin{aligned} \mathbb{P}\left(\bigcap_{i=0}^{k-1} \{s_i \longrightarrow s_{i+1}\}\right) &= \mathbb{E}\left[\mathbb{P}\left(\bigcap_{i=0}^{k-1} \{s_i \longrightarrow s_{i+1}\} \mid \text{PA}_{s_1-1}^{(1,\delta)}(b)\right)\right] \\ &= \mathbb{E}\left[\mathbb{1}_{\bigcap_{i=1}^{k-1} \{s_i \longrightarrow s_{i+1}\}} \mathbb{P}(s_0 \longrightarrow s_1 \mid \text{PA}_{s_0-1}^{(1,\delta)}(b))\right], \end{aligned} \tag{8.1.12}$$

since the event $\bigcap_{i=1}^{k-1} \{s_i \longrightarrow s_{i+1}\}$ is measurable with respect to $\text{PA}_{s_0-1}^{(1,\delta)}(b)$. Furthermore, from [Volume 1, (8.2.2)],

$$\mathbb{P}(s_0 \longrightarrow s_1 \mid \text{PA}_{s_0-1}^{(1,\delta)}(b)) = \frac{D_{s_1}(s_0 - 1) + \delta}{(2 + \delta)(s_0 - 1)}. \tag{8.1.13}$$

In particular,

$$\mathbb{P}(s_0 \longrightarrow s_1) = \mathbb{E}\left[\frac{D_{s_1}(s_0 - 1) + \delta}{(2 + \delta)(s_0 - 1)}\right]. \tag{8.1.14}$$

Therefore,

$$\begin{aligned} \mathbb{P}\left(\bigcap_{i=0}^{k-1} \{s_i \rightarrow s_{i+1}\}\right) &= \mathbb{E}\left[\mathbb{1}_{\bigcap_{i=1}^{k-1} \{s_i \rightarrow s_{i+1}\}} \frac{D_{s_2}(s_1 - 1) + \delta}{(2 + \delta)(s_1 - 1)}\right] \\ &= \mathbb{P}\left(\bigcap_{i=1}^{k-1} \{s_i \rightarrow s_{i+1}\}\right) \mathbb{E}\left[\frac{D_{s_1}(s_0 - 1) + \delta}{(2 + \delta)(s_0 - 1)}\right], \end{aligned} \tag{8.1.15}$$

since the random variable $D_{s_1}(s_0 - 1)$ only depends on how many edges are connected to s_1 after time s_1 , and is thus independent of the event $\bigcap_{i=1}^{k-1} \{s_i \rightarrow s_{i+1}\}$, which only depends on the attachment of the edges *up to and including* time s_1 . We conclude that

$$\mathbb{P}\left(\bigcap_{i=0}^{k-1} \{s_i \rightarrow s_{i+1}\}\right) = \mathbb{P}(s_1 \rightarrow s_2) \mathbb{P}\left(\bigcap_{i=0}^{k-1} \{s_i \rightarrow s_{i+1}\}\right). \tag{8.1.16}$$

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

$$\mathbb{P}(t \rightarrow s_1 \rightarrow \dots \rightarrow s_{k-1} \rightarrow s) = \prod_{i=0}^{k-1} \mathbb{E}\left[\frac{D_{s_{i+1}}(s_i - 1) + \delta}{(2 + \delta)(s_i - 1)}\right]. \tag{8.1.17}$$

By [Volume 1, (8.11.3)],

$$\begin{aligned} \mathbb{E}\left[\frac{D_i(t) + \delta}{(2 + \delta)t}\right] &= (1 + \delta) \frac{\Gamma(t + \frac{1}{2+\delta})\Gamma(i)}{(2 + \delta)t\Gamma(t)\Gamma(i + \frac{1}{2+\delta})} \\ &= \frac{1 + \delta}{2 + \delta} \frac{\Gamma(t + \frac{1}{2+\delta})\Gamma(i)}{\Gamma(t + 1)\Gamma(i + \frac{1}{2+\delta})}, \end{aligned} \tag{8.1.18}$$

so that

$$\begin{aligned} \mathbb{P}(t \rightarrow s_1 \rightarrow \dots \rightarrow s_{k-1} \rightarrow s) &= \left(\frac{1 + \delta}{2 + \delta}\right)^k \prod_{i=0}^{k-1} \frac{\Gamma(s_i + \frac{1}{2+\delta})\Gamma(s_{i+1})}{\Gamma(s_i + 1)\Gamma(s_{i+1} + \frac{1}{2+\delta})} \\ &= \left(\frac{1 + \delta}{2 + \delta}\right)^k \frac{\Gamma(s_0 + \frac{1}{2+\delta})\Gamma(s_k + 1)}{\Gamma(s_k + \frac{1}{2+\delta})\Gamma(s_0)} \prod_{i=1}^{k-1} \frac{1}{s_i} \\ &= \left(\frac{1 + \delta}{2 + \delta}\right)^k \frac{\Gamma(t + \frac{1}{2+\delta})\Gamma(s)}{\Gamma(1 + \frac{1}{2+\delta})\Gamma(t + 1)} \prod_{i=1}^{k-1} \frac{1}{s_i}. \end{aligned} \tag{8.1.20}$$

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 o chosen uniformly at random from $[n]$,

$$\frac{\text{dist}(1, o)}{\log n} \leq (1 + o_{\mathbb{P}}(1)) \frac{1 + \delta}{2 + \delta}. \tag{8.1.21}$$

and, almost surely for large enough n ,

$$\frac{\text{dist}(1, n)}{\log n} \leq (1 + \varepsilon) \frac{(1 + \delta)}{(2 + \delta)\gamma}. \tag{8.1.22}$$

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 ,

$$\text{height}(\text{PA}_n^{(1, \delta)}(b)) \leq (1 + \varepsilon) \frac{(1 + \delta)}{(2 + \delta)\gamma}. \tag{8.1.23}$$

By the triangle inequality,

$$\text{dist}_{\text{PA}_n^{(1, \delta)}(b)}(o_1, o_2) \leq \text{dist}(1, o_1) + \text{dist}(1, o_2), \tag{8.1.24}$$

$$\text{diam}(\text{PA}_n^{(1, \delta)}(b)) \leq 2\text{height}(\text{PA}_n^{(1, \delta)}(b)), \tag{8.1.25}$$

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

$$\mathbb{P}(\text{dist}(1, n) = k) = \left(\frac{1 + \delta}{2 + \delta}\right)^k \frac{\Gamma(n + \frac{1}{2 + \delta})}{\Gamma(1 + \frac{1}{2 + \delta})\Gamma(n + 1)} \sum_{\vec{t}_{k-1}}^* \frac{1}{(k - 1)!} \prod_{i=1}^{k-1} \frac{1}{t_i}, \tag{8.1.26}$$

where the sum now is over all vectors $\vec{t}_{k-1} = (t_1, \dots, t_{k-1})$ 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

$$\mathbb{P}(\text{dist}(1, n) = k) \leq \left(\frac{1 + \delta}{2 + \delta}\right)^k \frac{\Gamma(n + \frac{1}{2 + \delta})}{\Gamma(1 + \frac{1}{2 + \delta})\Gamma(n + 1)} \frac{1}{(k - 1)!} \left(\sum_{s=2}^{n-1} \frac{1}{s}\right)^{k-1}. \tag{8.1.27}$$

Since $x \mapsto 1/x$ is monotonically decreasing

$$\sum_{s=2}^{n-1} \frac{1}{s} \leq \int_1^n \frac{1}{x} dx = \log n. \tag{8.1.28}$$

Also, we use [Volume 1, (8.3.9)] to bound, for some constant $C_\delta > 0$,

$$\begin{aligned} \mathbb{P}(\text{dist}(1, n) = k) &\leq C_\delta n^{-\frac{1 + \delta}{2 + \delta}} \frac{\left(\frac{1 + \delta}{2 + \delta} \log n\right)^{k-1}}{(k - 1)!} \\ &= C\mathbb{P}\left(\text{Poi}\left(\frac{1 + \delta}{2 + \delta} \log n\right) = k - 1\right). \end{aligned} \tag{8.1.29}$$

Now we are ready to prove (8.1.21). We note that o is chosen uniformly in $[n]$,

so that

$$\begin{aligned}
\mathbb{P}(\text{dist}(1, o) = k) &= \frac{1}{n} \sum_{s=1}^n \mathbb{P}(\text{dist}(v_s, v_1) = k) \leq \frac{1}{n} \sum_{s=1}^n C_\delta s^{-\frac{1+\delta}{2+\delta}} \frac{\left(\frac{1+\delta}{2+\delta} \log s\right)^{k-1}}{(k-1)!} \\
&\leq \frac{\left(\frac{1+\delta}{2+\delta} \log n\right)^{k-1}}{t(k-1)!} \sum_{s=1}^n C_\delta s^{-\frac{1+\delta}{2+\delta}} \leq C \frac{\left(\frac{1+\delta}{2+\delta} \log n\right)^{k-1}}{(k-1)!} n^{-\frac{1+\delta}{2+\delta}} \\
&= C \mathbb{P}\left(\text{Poi}\left(\frac{1+\delta}{2+\delta} \log n\right) = k-1\right). \tag{8.1.30}
\end{aligned}$$

Therefore,

$$\mathbb{P}(\text{dist}(1, o) > k) \leq C \mathbb{P}\left(\text{Poi}\left(\frac{1+\delta}{2+\delta} \log n\right) \geq k\right). \tag{8.1.31}$$

Now we fix $\varepsilon > 0$ and take $k = k_n = \frac{(1+\varepsilon)(1+\delta)}{(2+\delta)} \log n$, to arrive at

$$\mathbb{P}(\text{dist}(1, o) > k_n) \leq C \mathbb{P}\left(\text{Poi}\left(\frac{1+\delta}{2+\delta} \log n\right) \geq \frac{(1+\varepsilon)(1+\delta)}{(2+\delta)} \log n\right) = o(1), \tag{8.1.32}$$

by the law of large numbers and for any $\varepsilon > 0$, as required.

We continue to prove (8.1.22). By (8.1.29),

$$\mathbb{P}(\text{dist}(1, n) > k) \leq C_\delta \mathbb{P}\left(\text{Poi}\left(\frac{1+\delta}{2+\delta} \log n\right) \geq k\right). \tag{8.1.33}$$

Now take $k_n = a \log n$ with $a > (1+\delta)/(2+\delta)$. We use the Borel-Cantelli lemmas to see that $\text{dist}(1, n) > k_n$ will occur only finitely often when (8.1.33) is summable. We then use the large deviation bounds for Poisson random variables in [Volume 1, Exercise 2.20] with $\lambda = (1+\delta)/(2+\delta)$ to obtain that

$$\mathbb{P}(\text{dist}(1, n) > a \log n) \leq C_\delta n^{-[a(\log(a(2+\delta)/(1+\delta)) - 1) + \frac{1+\delta}{2+\delta}]}. \tag{8.1.34}$$

Let x be the solution of

$$x(\log(x(2+\delta)/(1+\delta)) - 1) + \frac{1+\delta}{2+\delta} = 1, \tag{8.1.35}$$

so that $x = \frac{(1+\delta)}{(2+\delta)\gamma}$. Then, for every $a > x$,

$$\mathbb{P}(\text{dist}(1, n) > a \log n) = O(n^{-p}), \tag{8.1.36}$$

where

$$p = [a(\log(a(2+\delta)/(1+\delta)) - 1) + (1+\delta)(2+\delta)] > 1. \tag{8.1.37}$$

As a result, by the Borel-Cantelli Lemma, the event $\{\text{dist}(1, n) > k_n\}$ occurs only finitely often, and we conclude that (8.1.22) holds. \square

Proof of the lower bound on $\text{dist}_{\text{PA}_n^{(1,\delta)}(b)}(1, o)$ in Theorem 8.3. We use (8.1.30) to obtain that

$$\mathbb{P}(\text{dist}_{\text{PA}_n^{(1,\delta)}(b)}(1, o) \leq k) \leq C\mathbb{P}\left(\text{Poi}\left(\frac{1+\delta}{2+\delta}\log n\right) \leq k\right). \tag{8.1.38}$$

Fix $k_n = \frac{(1+\delta)(1-\varepsilon)}{(2+\delta)}\log n$, and note that $\mathbb{P}(\text{dist}_{\text{PA}_n^{(1,\delta)}(b)}(1, o) \leq k_n) = o(1)$ by the law of large numbers. □

To complete the proof of $\text{height}(\text{PA}_n^{(1,\delta)}(b))/\log n \xrightarrow{a.s.} \frac{(1+\delta)(1-\varepsilon)}{(2+\delta)\gamma}$ in Theorem 8.3, we use the second moment method to prove that $\text{height}(\text{PA}_n^{(1,\delta)}(b)) \leq \frac{(1+\delta)(1-\varepsilon)}{(2+\delta)\gamma}\log n$ has vanishing probability. Together with (8.1.23), this certainly proves that

$$\text{height}(\text{PA}_n^{(1,\delta)}(b))/\log n \xrightarrow{\mathbb{P}} \frac{(1+\delta)(1-\varepsilon)}{(2+\delta)\gamma}.$$

However, since $\text{height}(\text{PA}_n^{(1,\delta)}(b))$ 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 $\text{height}(\text{PA}_n^{(1,\delta)}(b)) \leq \frac{(1+\delta)(1-\varepsilon)}{(2+\delta)\gamma}\log n$ in the following proposition:

Proposition 8.5 (Height of $\text{PA}_n^{(1,\delta)}(b)$ converges in probability) *For every $\varepsilon > 0$ there exists a $\eta = \eta(\varepsilon) > 0$ such that*

$$\mathbb{P}\left(\text{height}(\text{PA}_n^{(1,\delta)}(b)) \leq \frac{(1+\delta)(1-\varepsilon)}{(2+\delta)\gamma}\log n\right) \leq O(n^{-\eta}). \tag{8.1.39}$$

Proof of lower bound on $\text{height}(\text{PA}_n^{(1,\delta)}(b))$ 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, $\text{height}(\text{PA}_{t_k}^{(1,\delta)}(b)) \geq \frac{(1+\delta)(1-\varepsilon)}{(2+\delta)\gamma}\log t_k$. This proves the almost sure lower bound on $\text{height}(\text{PA}_n^{(1,\delta)}(b))$ along the subsequence $(t_k)_{k \geq 0}$. To extend this to an almost sure lower bound when $n \rightarrow \infty$, we use that $t \mapsto \text{height}(\text{PA}_n^{(1,\delta)}(b))$ is non-decreasing, so that, for every $t \in [t_{k-1}, t_k]$,

$$\begin{aligned} \text{height}(\text{PA}_n^{(1,\delta)}(b)) &\geq \text{height}(\text{PA}_{t_{k-1}}^{(1,\delta)}(b)) && (8.1.40) \\ &\geq \frac{(1+\delta)(1-\varepsilon)}{(2+\delta)\gamma}\log t_{k-1} \\ &\geq (1-\varepsilon)(1-\alpha)\frac{(1+\delta)}{(2+\delta)\gamma}\log n, \end{aligned}$$

where the third inequality follows from the almost sure lower bound on $\text{height}(\text{PA}_{t_{k-1}}^{(1,\delta)}(b))$. The above bound holds for all $\varepsilon, \alpha > 0$, so that letting $\varepsilon, \alpha \downarrow 0$ 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 $\pi = (\pi_0, \dots, \pi_k)$ 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

tree $\text{PA}_n^{(1,\delta)}(b)$. We write $\pi \in \text{PA}_n^{(1,\delta)}(b)$ for the event that the edge from π_i is connected to π_{i+1} for all $i = 0, \dots, k-1$. We let

$$N_k(n) = \#\{\pi \subseteq \text{PA}_n^{(1,\delta)}(b) : \pi_k \in [n]\} \quad (8.1.41)$$

denote the number of k -step paths in $\text{PA}_n^{(1,\delta)}(b)$ with an endpoint in $[n]$. By Proposition 8.4,

$$\begin{aligned} \mathbb{E}[N_k(n)] &= \left(\frac{1+\delta}{2+\delta}\right)^k \sum_{s_k=1}^n \sum_{s_0=s_k}^n \frac{\Gamma(s_0 + \frac{1}{2+\delta})\Gamma(s_k)}{\Gamma(s_k + \frac{1}{2+\delta})\Gamma(s_0+1)} \sum_{\vec{s}_k} \prod_{i=1}^{k-1} \frac{1}{s_i} \\ &\geq \left(\frac{1+\delta}{2+\delta}\right)^k \sum_{s_k=1}^n \frac{\Gamma(s_k)}{\Gamma(s_k + \frac{1}{2+\delta})} \sum_{s_0=s_k}^n \frac{\Gamma(s_0 + \frac{1}{2+\delta})}{\Gamma(s_0+1)} \sum_{\vec{s}_k} \prod_{i=1}^{k-1} \frac{1}{s_i}, \end{aligned} \quad (8.1.42)$$

where again the sum is over all ordered $\vec{s}_k = (s_0, \dots, s_k)$ with $s_k \in [n]$ and $s_0 \in [n]$. We can bound this from below by

$$\mathbb{E}[N_k(n)] \geq \left(\frac{1+\delta}{2+\delta}\right)^k \sum_{s_k=1}^n \frac{\Gamma(s_k)}{\Gamma(s_k + \frac{1}{2+\delta})} \sum_{s_0=s_k}^n \frac{\Gamma(s_0 + \frac{1}{2+\delta})}{\Gamma(s_0+1)} \frac{1}{(k-1)!} \sum_{\vec{t}_k} \prod_{i=1}^{k-1} \frac{1}{t_i}, \quad (8.1.43)$$

where now the sum is over all vectors $\vec{t}_k = (t_1, \dots, t_{k-1})$ with distinct coordinates with $t_i \in [s_k+1, s_0-1]$. For fixed s_0, s_k ,

$$\sum_{\vec{t}_k} \prod_{i=1}^{k-1} \frac{1}{t_i} \geq \left(\sum_{s=s_k+k}^{s_0-1} \frac{1}{s}\right)^{k-1}. \quad (8.1.44)$$

We can lower bound

$$\sum_{s=s_k+k}^{s_0-1} \frac{1}{s} \geq \int_{s_k+k}^{s_0} \frac{1}{x} dx = \log(s_0/(s_k+k)) \geq (1-\varepsilon) \log n \quad (8.1.45)$$

when $s_0 \geq n/2$, $s_k \leq T$ and $\log[2(T+k)] \leq \varepsilon \log n$. Thus, we conclude that

$$\mathbb{E}[N_k(n)] \geq \left(\frac{1+\delta}{2+\delta}\right)^k \sum_{s_k=1}^n \frac{\Gamma(s_k)}{\Gamma(s_k + \frac{1}{2+\delta})} \sum_{s_0=t/2}^n \frac{\Gamma(s_0 + \frac{1}{2+\delta})}{\Gamma(s_0+1)} \frac{1}{(k-1)!} [(1-\varepsilon) \log n]^{k-1}. \quad (8.1.46)$$

Using [Volume 1, (8.3.9)], we therefore arrive at

$$\begin{aligned} \mathbb{E}[N_k(n)] &\geq (1+o(1)) \left(\frac{1+\delta}{2+\delta}\right)^k \sum_{s_k=1}^n s_k^{-1/(2+\delta)} \sum_{s_0=t/2}^n s_0^{-(1+\delta)/(2+\delta)} \frac{1}{(k-1)!} [(1-\varepsilon) \log n]^k \\ &\geq c \left(\frac{1+\delta}{2+\delta}\right)^k n^{(1+\delta)/(2+\delta)} n^{1/(2+\delta)} \frac{1}{(k-1)!} [(1-\varepsilon) \log n]^{k-1} \\ &= cn^{(1+\delta)/(2+\delta)} n^{[1+(1-\varepsilon)(1+\delta)]/(2+\delta)} \mathbb{P}\left(\text{Poi}\left(\frac{1+\delta}{2+\delta}(1-\varepsilon) \log n\right) = k-1\right). \end{aligned} \quad (8.1.47)$$

When we take $k = k_n = \frac{(1+\delta)(1-\varepsilon)}{(2+\delta)\gamma} \log n$, then

$$\mathbb{P}\left(\text{Poi}\left(\frac{1+\delta}{2+\delta}(1-\varepsilon)\log n\right) = k-1\right) \geq cn^{-(1-\varepsilon)}/\sqrt{\log n}, \tag{8.1.48}$$

so that

$$\mathbb{E}[N_k(n)] \geq n^{(1+\delta)/(2+\delta)}n^{\varepsilon/(2+\delta)+o(1)}. \tag{8.1.49}$$

We next take $T = n^\varepsilon$ to arrive at

$$\mathbb{E}[N_k(n)] \geq n^{\varepsilon+o(1)}. \tag{8.1.50}$$

This provides the required lower bound on $\mathbb{E}[N_k(n)]$. 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 $\text{Var}(N_k(n)) \leq C\mathbb{E}[N_k(n)]^2n^{-\eta}$ (see Lemma 8.12 below). As a result, by the Chebychev inequality ([Volume 1, Theorem 2.18])

$$\mathbb{P}(N_k(n) = 0) \leq \frac{\text{Var}(N_k(n))}{\mathbb{E}[N_k(n)]^2} \leq Cn^{-\eta}. \tag{8.1.51}$$

Since $\text{height}(\text{PA}_n^{(1,\delta)}(b)) \geq k$ 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 $\text{PA}_n^{(1,\delta)}(b)$ in Theorem 8.3, for which we use that

$$\text{diam}(\text{PA}_n^{(1,\delta)}(b)) \leq 2 \cdot \text{height}(\text{PA}_n^{(1,\delta)}(b)). \tag{8.1.52}$$

Equation (8.1.52) together with the upper bound in Theorem 8.3 imply that

$$\limsup_{n \rightarrow \infty} \frac{\text{diam}(\text{PA}_n^{(1,\delta)}(b))}{\log n} \leq \frac{2(1+\delta)}{\gamma(2+\delta)}. \tag{8.1.53}$$

For the lower bound, we use the lower bound on $\text{diam}(\text{PA}_n^{(1,\delta)}(b))$ in Theorem 8.3 and the decomposition of scale-free trees in Theorem 5.3. Theorem 5.3 states that the scale-free tree $\text{PA}_n^{(1,\delta)}(b)$ can be decomposed into two scale-free trees, having a similar distribution as copies $\text{PA}_{S_1(t)}^{(b1)}(1, \delta)$ and $\text{PA}_{t-S_1(t)}^{(1,\delta)}(b2)$, where $(\text{PA}_n^{(1,\delta)}(b1))_{n \geq 1}$ and $(\text{PA}_n^{(1,\delta)}(b2))_{n \geq 1}$ are *independent* scale-free tree processes, and the law of $S_1(t)$ is described in (5.1.29). By this tree decomposition

$$\text{diam}(\text{PA}_n^{(1,\delta)}(b)) \geq \text{height}(\text{PA}_{S_1(n)}^{(1,\delta)}(b1)) + \text{height}(\text{PA}_{n-S_1(n)}^{(1,\delta)}(b2)). \tag{8.1.54}$$

The two trees $(\text{PA}_n^{(1,\delta)}(b1))_{n \geq 1}$ and $(\text{PA}_n^{(1,\delta)}(b2))_{n \geq 1}$ are not *exactly* equal in distribution to $(\text{PA}_n^{(1,\delta)}(b))_{n \geq 1}$, because the initial degree of the starting vertices at time $n = 2$ is different. However, the precise almost sure scaling in Theorem 5.3 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 $\text{PA}_2^{(1,\delta)}(b1)$ and $\text{PA}_2^{(1,\delta)}(b1)$ (see the remark below Theorem 8.3).

Exercise??

Since $S_1(n)/n \xrightarrow{a.s.} U$, with U having a Beta-distribution with parameters $a = \frac{3+\delta}{2+\delta}$ and $b = \frac{1+\delta}{2+\delta}$, we obtain that $\text{height}(\text{PA}_{S_1(n)}^{(1,\delta)}(b1))/\log n \xrightarrow{a.s.} \frac{(1+\delta)}{(2+\delta)\gamma}$ and $\text{height}(\text{PA}_{n-S_1(n)}^{(1,\delta)}(b2))/\log n \xrightarrow{a.s.} \frac{(1+\delta)}{(2+\delta)\gamma}$. Thus, we conclude that, almost surely for large enough n ,

$$\frac{\text{diam}(\text{PA}_n^{(1,\delta)}(b))}{\log n} \geq (1 - \varepsilon) \frac{2(1 + \delta)}{(2 + \delta)\gamma}. \tag{8.1.55}$$

Combining (8.1.53) and (8.1.55) proves the convergence of $\text{diam}(\text{PA}_n^{(1,\delta)}(b))/\log n$ in Theorem 8.3.

We proceed with the proof of the convergence of $\text{dist}_{\text{PA}_n^{(1,\delta)}(b)}(o_1, o_2)/\log n$ in Theorem 8.3. We write

$$\begin{aligned} \text{dist}_{\text{PA}_n^{(1,\delta)}(b)}(o_1, o_2) &= \text{dist}_{\text{PA}_n^{(1,\delta)}(b)}(1, o_1) + \text{dist}_{\text{PA}_n^{(1,\delta)}(b)}(1, o_2) \\ &\quad - \text{dist}_{\text{PA}_n^{(1,\delta)}(b)}(1, V), \end{aligned} \tag{8.1.56}$$

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 $\text{dist}_{\text{PA}_n^{(1,\delta)}(b)}(1, V) = o_{\mathbb{P}}(\log n)$.

Update from here!

To prove Theorem 8.2, we note that the connected components of $\text{PA}_n^{(1,\delta)}$ are similar in distribution to single scale-free tree $\text{PA}_{t_1}^{(1,\delta)}(b1), \dots, \text{PA}_{t_{N_n}}^{(1,\delta)}(bN_n)$, 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_t 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 $\varepsilon n/\log n$. Since

$$\log(\varepsilon n/\log n) = \log n(1 + o(1)), \tag{8.1.57}$$

the result follows along the same lines as in the proof of Theorem 8.3. □

8.2 SMALL-WORLD EFFECT 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. We investigate both the diameter as well as typical distances. By Theorem 5.16, $\text{PA}_n^{(m,\delta)}$ is whp connected when $m \geq 2$. Recall that in a connected graph, the typical distance or hopcount $\text{dist}_{\text{PA}_n^{(m,\delta)}}(o_1, o_2)$ is the graph distance between two vertices chosen uniformly at random from $[n]$. Recall further that the power-law degree exponent for $\text{PA}_n^{(m,\delta)}$ 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 distaces are logarithmic in the size of the graph when $\tau > 3$, and doubly logarithmic when $\tau \in (2, 3)$. We will see that similar behavior is true for $\text{PA}_n^{(m,\delta)}$. We start by analyzing the case where $\delta > 0$.

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 τ satisfies $\tau > 3$. In this case, both the diameter as well as typical distances are logarithmic in the size of the graph:

Theorem 8.6 (A $\log n$ bound for typical distances in PAMs) *Fix $m \geq 2$ and $\delta > 0$. There exist $0 < a_1 < a_2 < \infty$ such that, as $n \rightarrow \infty$,*

$$\mathbb{P}(a_1 \log n \leq \text{dist}_{\text{PA}_n^{(m,\delta)}}(o_1, o_2) \leq a_2 \log n) = 1 - o(1). \quad (8.2.1)$$

While we believe that there are constants a and b with $a < b$ such that

$$\text{dist}_{\text{PA}_n^{(m,\delta)}}(o_1, o_2) / \log n \xrightarrow{\mathbb{P}} a, \quad \text{diam}(\text{PA}_n^{(m,\delta)}) / \log n \xrightarrow{\mathbb{P}} b, \quad (8.2.2)$$

currently, a proof for this fact is still missing.

Distances in PA models with $m \geq 2$ and $\delta = 0$.

For $\delta = 0$, $\tau = 3$. For $\text{NR}_n(\mathbf{w})$, distances grow as $\log n / \log \log n$ in this case (recall Theorem 6.17). The same turns out to be true for $\text{PA}_n^{(m,\delta)}$:

Theorem 8.7 (Typical distances of $\text{PA}_n^{(m,\delta)}$ for $\delta = 0$) *Fix $m \geq 2$ and $\delta = 0$. As $n \rightarrow \infty$,*

$$\text{dist}_{\text{PA}_n^{(m,\delta)}}(o_1, o_2) \frac{\log \log n}{\log n} \xrightarrow{\mathbb{P}} 1. \quad (8.2.3)$$

Theorem 8.7 shows that distances for $\tau = 3$ are similar in $\text{PA}_n^{(m,\delta)}$ as in $\text{NR}_n(\mathbf{w})$. Interestingly, for $\text{PA}_n^{(m,\delta)}$ with $\delta = 0$, the diameter and the typical distances are close to being equal. For $\text{NR}_n(\mathbf{w})$ and $\text{CM}_n(\mathbf{d})$ with power-law exponent $\tau = 3$, this fact is not known.

 $\log \log$ 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 ($\log \log n$ asymptotics for the diameter for $\delta < 0$) *Fix $m \geq 2$ and assume that $\delta \in (-m, 0)$. As $n \rightarrow \infty$,*

$$\frac{\text{dist}_{\text{PA}_n^{(m,\delta)}}(o_1, o_2)}{\log \log n} \xrightarrow{\mathbb{P}} \frac{4}{|\log(\tau - 2)|}. \quad (8.2.4)$$

Interestingly, the term $4/|\log(\tau - 2)|$ appearing in Theorems 8.8 replaces the term $2/|\log(\tau - 2)|$ in Theorems 7.2 and 7.17 for the configuration model $\text{CM}_n(\mathbf{d})$ with power-law exponent $\tau \in (2, 3)$. Thus, typical distances are twice as big for $\text{PA}_n^{(m,\delta)}$ compared to $\text{CM}_n(\mathbf{d})$ with the same power-law exponent. This can be intuitively explained as follows. For the configuration model $\text{CM}_n(\mathbf{d})$, vertices with high degrees are likely to be directly connected (see e.g. Lemma 7.11). For $\text{PA}_n^{(m,\delta)}$, this is *not* the case. However, vertices with high degrees are likely to be

at distance two. This makes distances in $\text{PA}_n^{(m,\delta)}$ about twice as big as those for $\text{CM}_n(\mathbf{d})$ with the same degree sequence. This effect is special for $\delta < 0$ and is studied in more detail in Exercises 8.4–8.5.

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 suggests a strong form of *universality*, which is interesting in its 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 $\pi = (s_0, s_1, \dots, s_l)$ *self-avoiding* when $s_i \neq s_j$ for all $1 \leq i < j \leq l$. The following proposition studies the probability that a path is present in $\text{PA}_n^{(m,\delta)}$:

Proposition 8.9 (Path counting in $\text{PA}_n^{(m,\delta)}$) *Denote $\gamma = \frac{m}{2m+\delta}$. Let $\pi = (\pi_0, \pi_1, \dots, \pi_l)$ be a self-avoiding path of length l consisting of the $l + 1$ unordered vertices $\pi_0, \pi_1, \dots, \pi_l$. Then*

$$\mathbb{P}(\pi \subseteq \text{PA}_n^{(m,\delta)}) \leq (Cm^2)^l \prod_{i=0}^{l-1} \frac{1}{(\pi_i \wedge \pi_{i+1})^\gamma (\pi_i \vee \pi_{i+1})^{1-\gamma}}. \tag{8.3.1}$$

Paths are formed by repeatedly forming edges. When $m = 1$, paths go from later vertices to older vertices. When $m \geq 2$, this monotonicity property of paths is lost, which makes the proof harder. We start by investigating intersections of events that specify which edges are present in $\text{PA}_n^{(m,\delta)}$. We start by introducing some notation. Denote by

$$\{g(t, j) = s\}, \quad 1 \leq j \leq m, \tag{8.3.2}$$

the event that the j th edge of vertex t is attached to the earlier vertex s . For $\text{PA}_n^{(m,\delta)}$, this event means that in $\{\text{PA}_{mt}(1, \delta)\}$ the edge from vertex $m(t - 1) + j$ is attached to one of the vertices $m(s - 1) + 1, \dots, ms$.

It is a direct consequence of the definition of PA-models that the event (8.3.2) increases the preference for vertex s , and hence decreases (in a relative way) the preference for the vertices u , $1 \leq u \leq t$, $u \neq s$. It should be intuitively clear that another way of expressing this effect is to say that, for different $s_1 \neq s_2$, the events $\{g(t_1, j_1) = s_1\}$ and $\{g(t_2, j_2) = s_2\}$ are negatively correlated. We now formalize this result. For integer $n_s \geq 1$, we denote by

$$E_s = \bigcap_{i=1}^{n_s} \{g(t_i^{(s)}, j_i^{(s)}) = s\}, \tag{8.3.3}$$

the event that the j_i th edge of vertex t_i is attached to the earlier vertex s , for

$i = 1, \dots, n_s$. We start by proving that the events E_s , for different s , are *negatively correlated* for each choice of $k \geq 1$ and all possible choices of $t_i^{(s)}, j_i^{(s)}$.

Lemma 8.10 (Negative correlation for connection of edges) *For distinct s_1, s_2, \dots, s_k , both for $\text{PA}_n^{(m,\delta)}$ and for $\text{PA}_t^{(m,\delta)}(b)$,*

$$\mathbb{P}\left(\bigcap_{i=1}^k E_{s_i}\right) \leq \prod_{i=1}^k \mathbb{P}(E_{s_i}). \quad (8.3.4)$$

Proof We only prove the statement for $\text{PA}_n^{(m,\delta)}$, the proof for $\text{PA}_t^{(m,\delta)}(b)$ is identical. We use induction on the largest edge number present in the events E_{s_1}, \dots, E_{s_k} . Here, we define the edge number of the event $\{g(t, j) = s\}$ to be $m(t-1) + j$, which is the order of the edge when we consider the edges as being attached in sequence in $\text{PA}_{mt}(1, \delta/m)$.

The induction hypothesis is that (8.3.4) holds for all k and all choices of $t_i^{(s)}, j_i^{(s)}$ such that $\max_{i,s} m(t_i^{(s)} - 1) + j_i^{(s)} \leq e$, where induction is performed with respect to e .

To initialize the induction, we note that for $e = 1$, the induction hypothesis holds trivially, since $\bigcap_{i=1}^k E_{s_i}$ 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 and all choices of $t_i^{(s)}, j_i^{(s)}$ such that $\max_{i,s} m(t_i^{(s)} - 1) + j_i^{(s)} \leq e - 1$, and we extend it to all k and all choices of $t_i^{(s)}, j_i^{(s)}$ such that $\max_{i,s} m(t_i^{(s)} - 1) + j_i^{(s)} \leq e$. Clearly, for k and $t_i^{(s)}, j_i^{(s)}$ such that $\max_{i,s} m(t_i^{(s)} - 1) + j_i^{(s)} \leq e - 1$, the bound follows from the induction hypothesis, so we may restrict attention to the case that $\max_{i,s} m(t_i^{(s)} - 1) + j_i^{(s)} = e$. We note that there is a unique choice of t, j such that $m(t-1) + j = e$. There are two possibilities: (1) Either there is exactly one choice of s and $t_i^{(s)}, j_i^{(s)}$ such that $t_i^{(s)} = t, j_i^{(s)} = j$, or (2) there are at least *two* of such choices. In the latter case, $\bigcap_{s=1}^k E_s = \emptyset$, 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 $t_i^{(s)}, j_i^{(s)}$ such that $t_i^{(s)} = t, j_i^{(s)} = j$. Denote by

$$E'_s = \bigcap_{i=1: (t_i^{(s)}, j_i^{(s)}) \neq (t, j)}^{n_s} \{g(t_i^{(s)}, j_i^{(s)}) = s\} \quad (8.3.5)$$

the restriction of E_s to all *other* edges. Then we can write

$$\bigcap_{i=1}^k E_{s_i} = \{g(t, j) = s\} \cap E'_s \cap \bigcap_{i=1: s_i \neq s}^k E_{s_i}. \quad (8.3.6)$$

By construction, all the edge numbers of the events in $E'_s \cap \bigcap_{i=1: s_i \neq s}^k E_{s_i}$ are at most $e - 1$. Thus we obtain

$$\mathbb{P}\left(\bigcap_{i=1}^k E_{s_i}\right) \leq \mathbb{E}\left[\mathbb{1}_{E'_s \cap \bigcap_{i=1: s_i \neq s}^k E_{s_i}} \mathbb{P}_{e-1}(g(t, j) = s)\right], \quad (8.3.7)$$

where \mathbb{P}_{e-1} denotes the conditional probability given the edge attachments up to the $(e - 1)$ st edge connection, or, equivalently, given $\text{PA}_{e-1}(1, \delta/m)$, and we have used that the event $E'_s \cap \bigcap_{i=1}^k_{s_i \neq s} E_{s_i}$ is measurable with respect to $\text{PA}_{e-1}(1, \delta/m)$.

We compute

$$\mathbb{P}_{e-1}(g(t, j) = s) = \frac{D_s(t - 1, j - 1) + \delta}{(2m + \delta)(t - 1) + (j - 1)(2 + \delta/m) + 1 + \delta}, \tag{8.3.8}$$

where we recall that $D_s(t - 1, j - 1)$ is the degree of vertex s after $j - 1$ edges of vertex t have been attached. We wish to use the induction hypothesis. For this, we note that

$$D_s(t - 1, j - 1) = m + \sum_{(t', j') : mt' + j' \leq e-1} \mathbb{1}_{\{g(t', j') = s\}}, \tag{8.3.9}$$

where we recall that $e - 1 = m(t - 1) + j - 1$. Each of the events $\{g(t', j') = s\}$ in (8.3.9) 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 $m + \delta \geq 0$, that,

$$\begin{aligned} \mathbb{P}\left(\bigcap_{i=1}^k E_{s_i}\right) &\leq \frac{m + \delta}{(2m + \delta)(t - 1) + (j - 1)(2 + \delta/m) + 1 + \delta} \mathbb{P}(E'_s) \prod_{i=1: s_i \neq s}^k \mathbb{P}(E_{s_i}) \\ &\quad + \sum_{(t', j') : mt' + j' \leq e-1} \frac{\mathbb{P}(E'_s \cap \{g(t', j') = s\})}{(2m + \delta)(t - 1) + (j - 1)(2 + \delta/m) + 1 + \delta} \prod_{i=1: s_i \neq s}^k \mathbb{P}(E_{s_i}). \end{aligned} \tag{8.3.10}$$

We use (8.3.9) to recombine the above as

$$\mathbb{P}\left(\bigcap_{i=1}^k E_{s_i}\right) \leq \mathbb{E}\left[\mathbb{1}_{E'_s} \frac{D_s(t - 1, j - 1) + \delta}{(2m + \delta)(t - 1) + (j - 1)(2 + \delta/m) + 1 + \delta}\right] \prod_{i=1: s_i \neq s}^k \mathbb{P}(E_{s_i}), \tag{8.3.11}$$

and the advancement is completed when we note that

$$\mathbb{E}\left[\mathbb{1}_{E'_s} \frac{D_s(t - 1, j - 1) + \delta}{(2m + \delta)(t - 1) + (j - 1)(2 + \delta/m) + 1 + \delta}\right] = \mathbb{P}(E_s). \tag{8.3.12}$$

The claim in Lemma 8.10 follows by induction. □

We next study the probabilities of the events E_s when $n_s \leq 2$:

Lemma 8.11 (Edge connection events for at most two edges) *Denote $\gamma = \frac{m}{2m + \delta}$. There exist absolute constants $M_1 = M_1(\delta, m)$, $M_2 = M_2(\delta, m)$, such that*

(i) for $m = 1$ and any $n > s$,

$$\mathbb{P}(g(n, 1) = s) = (1 + \delta) \frac{\Gamma(t)\Gamma(s + \frac{1+\delta}{2+\delta})}{\Gamma(t + \frac{1+\delta}{2+\delta})\Gamma(s + 1)} \leq \frac{M_1}{n^{1-\gamma} s^\gamma}. \tag{8.3.13}$$

Consequently, for each $1 \leq j \leq m$ and $n > s$,

$$\mathbb{P}(g(n, j) = s) \leq \frac{M_1}{n^{1-\gamma} s^\gamma}. \quad (8.3.14)$$

(ii) for $m = 1$ and any $t_2 > t_1 > s$,

$$\mathbb{P}(g(t_1, 1) = s, g(t_2, 1) = s) = \frac{\Gamma(t_2)\Gamma(t_1 + \frac{1}{2+\delta})\Gamma(s + \frac{1+\delta}{2+\delta})}{\Gamma(t_2 + \frac{1+\delta}{2+\delta})\Gamma(t_1 + 1)\Gamma(s + \frac{3+\delta}{2+\delta})} (1 + \delta). \quad (8.3.15)$$

Consequently, for any $1 \leq j_1, j_2 \leq m$ and $t_2 > t_1 > s$,

$$\mathbb{P}(g(t_1, j_1) = s, g(t_2, j_2) = s) \leq \frac{M_2}{(t_1 t_2)^{1-\gamma} s^{2\gamma}}. \quad (8.3.16)$$

Proof We only prove (8.3.13) and (8.3.15), (8.3.14) and (8.3.16) follow immediately from [Volume 1, (8.3.9)].

Throughout this proof, we abbreviate $g(t) = g(t, 1)$. By the definition of $(\text{PA}_n^{(m,\delta)})_{n \geq 1}$ in terms of $(\text{PA}_n(1, \delta/m))_{n \geq 1}$, this implies the result for general $m \geq 1$, where the factors of m follow from the fact that vertex s in $\text{PA}_n^{(m,\delta)}$ corresponds to vertices $ms, \dots, m(s+1) - 1$ in $\text{PA}_{mt}(1, \delta/m)$, which are all at least ms . Note, in particular, that $g(t, j) = s$ for $m \geq 2$ in $\text{PA}_n^{(m,\delta)}$ is equivalent to $g(m(t-1) + j) \in \{m(s-1) + 1, \dots, ms\}$ in $\text{PA}_{mt}(1, \delta/m)$.

For (8.3.13), we use [Volume 1, Theorem 8.2] to compute

$$\begin{aligned} \mathbb{P}(g(t) = s) &= \mathbb{E}[\mathbb{E}[\mathbb{1}_{\{g(t)=s\}} | \text{PA}_{t-1}(1, \delta)]]] = \mathbb{E}\left[\frac{D_s(t-1) + \delta}{(2+\delta)(t-1) + 1 + \delta}\right] \\ &= (1 + \delta) \frac{\Gamma(t)\Gamma(s + \frac{1+\delta}{2+\delta})}{\Gamma(t + \frac{1+\delta}{2+\delta})\Gamma(s+1)}. \end{aligned} \quad (8.3.17)$$

Take $t_2 > t_1$. We proceed with the proof of (8.3.15) by computing

$$\begin{aligned} \mathbb{P}(g(t_1) = s, g(t_2) = s) & \quad (8.3.18) \\ &= \mathbb{E}\left[\mathbb{P}(g(t_1) = s, g(t_2) = s | \text{PA}_{t_2-1}(m, \delta))\right] \\ &= \mathbb{E}\left[\mathbb{1}_{\{g(t_1)=s\}} \left(\frac{D_s(t_2-1) + \delta}{(t_2-1)(2+\delta) + 1 + \delta}\right)\right] \\ &= \frac{1}{(t_2-1)(2+\delta) + 1 + \delta} \frac{\Gamma(t_2)\Gamma(t_1 + \frac{1+\delta}{2+\delta})}{\Gamma(t_2-1 + \frac{1+\delta}{2+\delta})\Gamma(t_1+1)} \mathbb{E}\left[\mathbb{1}_{\{g(t_1)=s\}} (D_s(t_1) + \delta)\right] \\ &= \frac{\Gamma(t_2)\Gamma(t_1 + \frac{1+\delta}{2+\delta})}{\Gamma(t_2 + \frac{1+\delta}{2+\delta})\Gamma(t_1+1)} \mathbb{E}\left[\mathbb{1}_{\{g(t_1)=s\}} (D_s(t_1) + \delta)\right] \end{aligned} \quad (8.3.19)$$

where we use the iteration, for $t_1 < u \leq t_2 - 1$,

$$\begin{aligned} \mathbb{E}\left[\mathbb{1}_{\{g(t_1)=s\}} (D_s(u) + \delta)\right] & \quad (8.3.20) \\ &= \left(1 + \frac{1}{(2+\delta)(u-1) + 1 + \delta}\right) \mathbb{E}\left[\mathbb{1}_{\{g(t_1)=s\}} (D_s(u-1) + \delta)\right]. \end{aligned}$$

We are lead to compute $\mathbb{E} [\mathbb{1}_{\{g(t_1)=s\}} (D_s(t_1) + \delta)]$. We use recursion to obtain

$$\begin{aligned} & \mathbb{E} [\mathbb{1}_{\{g(t_1)=s\}} (D_s(t_1) + \delta) | \text{PA}_{t_1-1}(m, \delta)] \\ &= \mathbb{E} [\mathbb{1}_{\{g(t_1)=s\}} (D_s(t_1) - D_s(t_1 - 1)) | \text{PA}_{t_1-1}(m, \delta)] \\ & \quad + \mathbb{E} [\mathbb{1}_{\{g(t_1)=s\}} (D_s(t_1 - 1) + \delta) | \text{PA}_{t_1-1}(m, \delta)] \\ &= \frac{(D_s(t_1 - 1) + \delta)(D_s(t_1 - 1) + 1 + \delta)}{(t_1 - 1)(2 + \delta) + 1 + \delta}. \end{aligned} \quad (8.3.21)$$

By [Volume 1, Proposition 8.15],

$$\mathbb{E}[(D_s(t) + \delta)(D_s(t) + 1 + \delta)] = \frac{2}{c_2(t)} \mathbb{E}[Z_{s,2}(t)] = \frac{c_2(s)}{c_2(t)} (2 + \delta)(1 + \delta). \quad (8.3.22)$$

Recalling that $c_k(j) = \Gamma(j + \frac{1+\delta}{2+\delta})/\Gamma(j + \frac{k+1+\delta}{2+\delta})$, this brings us to

$$\mathbb{E}[(D_s(t) + \delta)(D_s(t) + 1 + \delta)] = \frac{\Gamma(t + \frac{3+\delta}{2+\delta})\Gamma(s + \frac{1+\delta}{2+\delta})}{\Gamma(t + \frac{1+\delta}{2+\delta})\Gamma(s + \frac{3+\delta}{2+\delta})} (2 + \delta)(1 + \delta). \quad (8.3.23)$$

where M_4 is a uniform constant. Consequently,

$$\begin{aligned} & \mathbb{E} [\mathbb{1}_{\{g(t_1)=s\}} (D_s(t_1) + \delta)] \\ &= \frac{\Gamma(t_1 + \frac{1}{2+\delta})\Gamma(s + \frac{1+\delta}{2+\delta})}{[(t_1 - 1)(2 + \delta) + 1 + \delta]\Gamma(t_1 - \frac{1}{2+\delta})\Gamma(s + \frac{3+\delta}{2+\delta})} (2 + \delta)(1 + \delta) \\ &= \frac{\Gamma(t_1 + \frac{1}{2+\delta})\Gamma(s + \frac{1+\delta}{2+\delta})}{\Gamma(t_1 + \frac{1+\delta}{2+\delta})\Gamma(s + \frac{3+\delta}{2+\delta})} (1 + \delta). \end{aligned} \quad (8.3.24)$$

Combining (8.3.19), (8.3.21) and (8.3.22), we arrive at

$$\begin{aligned} & \mathbb{P}(g(t_1) = s, g(t_2) = s) \\ &= \frac{\Gamma(t_2)\Gamma(t_1 + \frac{1+\delta}{2+\delta})}{\Gamma(t_2 + \frac{1+\delta}{2+\delta})\Gamma(t_1 + 1)} \times \frac{\Gamma(t_1 + \frac{1}{2+\delta})\Gamma(s + \frac{1+\delta}{2+\delta})}{\Gamma(t_1 + \frac{1+\delta}{2+\delta})\Gamma(s + \frac{3+\delta}{2+\delta})} (1 + \delta) \\ &= \frac{\Gamma(t_2)\Gamma(t_1 + \frac{1}{2+\delta})\Gamma(s + \frac{1+\delta}{2+\delta})}{\Gamma(t_2 + \frac{1+\delta}{2+\delta})\Gamma(t_1 + 1)\Gamma(s + \frac{3+\delta}{2+\delta})} (1 + \delta), \end{aligned} \quad (8.3.25)$$

as required. \square

With Lemmas 8.10 and 8.11 at hand, we are ready to prove Proposition 8.9:

Proof of Proposition 8.9. Since π is self-avoiding, we can write $\{\pi \subseteq \text{PA}_n^{(m,\delta)}\} = \bigcap_{i=1}^k E_{s_i}$, where either

$$E_s = \{g(t, j) = s\} \quad (8.3.26)$$

for some $t > s$ and some $1 \leq j \leq m$, or

$$E_s = \{g(t_1, j_1) = g(t_2, j_2) = s\}, \quad (8.3.27)$$

for some $t_1, t_2 > s$ and some $1 \leq j_1, j_2 \leq m$. In the first case, by (8.3.13),

$$\mathbb{P}(E_s) = \mathbb{P}(g(n, j) = s) \leq \frac{M_1}{n^{1-\gamma} s^\gamma}, \tag{8.3.28}$$

whereas in the second case, according to (8.3.15),

$$\mathbb{P}(E_s) = \mathbb{P}(g(t_1, j_1) = s, g(t_2, j_2) = s) \leq \frac{M_2}{(t_1 t_2)^{1-\gamma} s^{2\gamma}} = \frac{M_2}{t_1^{1-\gamma} s^\gamma t_2^{1-\gamma} s^\gamma}. \tag{8.3.29}$$

In both cases $M_i, i = 1, 2$, is an absolute constant. Lemma 8.10 then yields (8.3.1), where the factor m^{2l} originates from the number of possible choices of $j_i \in [m]$ for $i = 1, \dots, k$ and the possible s_i that are collapsed to the same vertex. \square

Lemma 8.12 (A variance estimate on $N_k(n)$) *Recall the definition of $N_k(n)$ in (8.1.41) and let $T = n^\varepsilon$ for some $\varepsilon > 0$. Then there exists constants $C > 0$ and $\eta = \eta(\varepsilon) > 0$ such that*

$$\text{Var}(N_k(n)) \leq C \mathbb{E}[N_k(n)]^2 n^{-\eta}. \tag{8.3.30}$$

Proof By Exercise 8.6, when the path (π_0, \dots, π_k) is completely disjoint from (ρ_0, \dots, ρ_k) ,

$$\begin{aligned} & \mathbb{P}\left(\bigcap_{i=0}^{k-1} \{\pi_i \longrightarrow \pi_{i+1}\} \cap \bigcap_{i=0}^{k-1} \{\rho_i \longrightarrow \rho_{i+1}\}\right) \\ & \leq \mathbb{P}\left(\bigcap_{i=0}^{k-1} \{\pi_i \longrightarrow \pi_{i+1}\}\right) \mathbb{P}\left(\bigcap_{i=0}^{k-1} \{\rho_i \longrightarrow \rho_{i+1}\}\right). \end{aligned} \tag{8.3.31}$$

Therefore, the indicators of disjoint paths are negatively correlated. As a result, we can bound

$$\text{Var}(N_k(n)) \leq \sum_{\pi, \rho: \pi \cap \rho \neq \emptyset} \mathbb{P}(\pi, \rho \subseteq \text{PA}_n^{(1, \delta)}(b)). \tag{8.3.32}$$

Since $m = 1$, the paths π, ρ of length k must merge at some point, before moving off to their common end point in $[n]$. When $\rho = \pi$, then we obtain a contribution $\mathbb{E}[N_k(n)]$, so that from now on we assume that $\pi \neq \rho$.

Write $\pi = (\pi_0, \dots, \pi_k)$ and $\rho = (\rho_0, \dots, \rho_k)$ and $\pi \neq \rho$. Then there must be an $l \in [k - 1]$ such that $\pi_j = \rho_j$ for all $j = l, \dots, k$. For two fixed paths π and ρ for which $\pi_j = \rho_j$ for all $j = l, \dots, k$, while $\pi_{l-1} \neq \rho_{l-1}$. By Lemma 8.10, for such paths π, ρ ,

$$\begin{aligned} \mathbb{P}(\pi, \rho \subseteq \text{PA}_n^{(1, \delta)}(b)) & \leq \left(\prod_{i=1}^{l-1} \mathbb{P}(\pi_{i-1} \longrightarrow \pi_i) \mathbb{P}(\rho_{i-1} \longrightarrow \rho_i)\right) \\ & \quad \times \mathbb{P}(\pi_{l-1}, \rho_{l-1} \longrightarrow \pi_l) \left(\prod_{j=l+1}^k \mathbb{P}(\pi_{j-1} \longrightarrow \pi_j)\right). \end{aligned} \tag{8.3.33}$$

By (8.1.9) in Proposition 8.4,

$$\prod_{i=1}^{l-1} \mathbb{P}(\pi_{i-1} \longrightarrow \pi_i) = \left(\frac{1+\delta}{2+\delta}\right)^{l-1} \frac{\Gamma(\pi_0 + \frac{1}{2+\delta})\Gamma(\pi_{l-1})}{\Gamma(\pi_{l-1} + \frac{1}{2+\delta})\Gamma(\pi_0 + 1)} \prod_{i=1}^{l-2} \frac{1}{\pi_i} \quad (8.3.34)$$

By symmetry, we may assume without loss of generality that $\pi_{l-1} > \rho_{l-1}$. Then, by (8.3.15),

$$\mathbb{P}(\pi_{l-1}, \rho_{l-1} \longrightarrow \pi_l) = (1+\delta) \frac{\Gamma(\rho_{l-1} - \frac{\delta}{2+\delta})\Gamma(\pi_{l-1} - \frac{1+\delta}{2+\delta})\Gamma(\pi_l)}{\Gamma(\rho_{l-1} + \frac{1}{2+\delta})\Gamma(\pi_{l-1})\Gamma(\pi_l + \frac{2}{2+\delta})}. \quad (8.3.35)$$

As a result,

$$\begin{aligned} \mathbb{P}(\pi, \rho \subseteq \text{PA}_n^{(1,\delta)}(b)) &\leq \left(\frac{1+\delta}{2+\delta}\right)^{l+k-3} \frac{\Gamma(\pi_0 + \frac{1}{2+\delta})\Gamma(\pi_{l-1})}{\Gamma(\pi_{l-1} + \frac{1}{2+\delta})\Gamma(\pi_0 + 1)} \frac{\Gamma(\rho_0 + \frac{1}{2+\delta})\Gamma(\rho_{l-1})}{\Gamma(\rho_{l-1} + \frac{1}{2+\delta})\Gamma(\rho_0 + 1)} \\ &\quad \times (1+\delta) \frac{\Gamma(\rho_{l-1} - \frac{\delta}{2+\delta})\Gamma(\pi_{l-1} - \frac{1+\delta}{2+\delta})\Gamma(\pi_l)}{\Gamma(\rho_{l-1} + \frac{1}{2+\delta})\Gamma(\pi_{l-1})\Gamma(\pi_l + \frac{2}{2+\delta})} \prod_{i=1}^{l-2} \frac{1}{\pi_i \rho_i} \\ &\quad \times \frac{\Gamma(\pi_l + \frac{1}{2+\delta})\Gamma(\pi_k)}{\Gamma(\pi_k + \frac{1}{2+\delta})\Gamma(\pi_l + 1)} \prod_{i=l+1}^{k-1} \frac{1}{\pi_i} \\ &= (1+\delta) \left(\frac{1+\delta}{2+\delta}\right)^{l+k-3} \frac{\Gamma(\pi_0 + \frac{1}{2+\delta})}{\Gamma(\pi_0 + 1)} \frac{\Gamma(\rho_0 + \frac{1}{2+\delta})}{\Gamma(\rho_0 + 1)} \frac{\Gamma(\rho_{l-1})\Gamma(\rho_{l-1} - \frac{\delta}{2+\delta})}{\Gamma(\rho_{l-1} + \frac{1}{2+\delta})\Gamma(\rho_{l-1} + \frac{1}{2+\delta})} \\ &\quad \times \frac{\Gamma(\pi_l + \frac{1}{2+\delta})}{\Gamma(\pi_l + \frac{2}{2+\delta})} \frac{\Gamma(\pi_k)}{\Gamma(\pi_k + \frac{1}{2+\delta})} \frac{1}{(\pi_{l-1} + \frac{1}{2+\delta})} \prod_{i=1}^{l-2} \frac{1}{\pi_i \rho_i} \prod_{i=l}^{k-1} \frac{1}{\pi_i}. \end{aligned} \quad (8.3.36)$$

By [Volume 1, (8.3.9)], this can be bounded by

$$C \left(\frac{1+\delta}{2+\delta}\right)^{l+k} (\pi_0 \rho_0)^{-(1+\delta)/(2+\delta)} (\pi_l \pi_k)^{-1/(2+\delta)} \prod_{i=1}^{l-1} \frac{1}{\rho_i} \prod_{i=1}^{k-1} \frac{1}{\pi_i}, \quad (8.3.37)$$

where C is a uniform constant. We need to sum the above over $l < k$, all decreasing $\pi = (\pi_0, \dots, \pi_k)$ with $\pi_k \in [n]$ and all decreasing $(\rho_0, \dots, \rho_{l-1})$ with $\rho_{l-1} > \pi_l$. The sum can be bounded from above by summing over *all* decreasing $(\rho_0, \dots, \rho_{l-1})$ and bounding $\pi_l^{-1/(2+\delta)} \leq 1$ to obtain an upper bound as in (8.1.43)

$$\mathbb{E}[N_k(n)] \left(\frac{1+\delta}{2+\delta}\right)^l \frac{(\log n)^l}{l!} \sum_{s \in [n]} s^{-(1+\delta)/(2+\delta)}. \quad (8.3.38)$$

Therefore,

$$\text{Var}(N_k(n)) \leq \mathbb{E}[N_k(n)] + \mathbb{E}[N_k(n)] \sum_{l=1}^{k-1} \left(\frac{1+\delta}{2+\delta}\right)^l \frac{(\log n)^l}{l!} \sum_{s \in [n]} s^{-(1+\delta)/(2+\delta)}. \quad (8.3.39)$$

When $T = n^\varepsilon$, 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 $\eta = \eta(\varepsilon) > 0$.

□

Add Discussion of $\text{dist}_{\text{PA}_n^{(1,\delta)_b}}(1, V) = o_{\mathbb{P}}(\log n)$ here! Or exercise?

8.4 SMALL-WORLD EFFECT IN PA MODELS: LOWER BOUNDS

In this section we prove lower bounds on distances in $\text{PA}_n^{(m,\delta)}$ with $m \geq 2$. In Section 8.3 we start by proving an upper bound on the probability that a path exists in $\text{PA}_n^{(m,\delta)}$, which is our main tool in this section. After this we prove the lower bounds on distances for $\delta > 0$ in Section 8.4.1, for $\delta = 0$ in Section 8.4.2, and for $\delta < 0$ in Section 8.4.3.

8.4.1 LOGARITHMIC LOWER BOUNDS ON DISTANCES FOR $\delta > 0$

In this section we investigate lower bounds on the distances when $\delta > 0$, in which case $\gamma = m/(2m + \delta) < 1/2$. By Proposition 8.9,

$$\mathbb{P}\left(\text{dist}_{\text{PA}_n^{(m,\delta)}}(1, t) = k\right) \leq c^k \sum_{\pi} \prod_{j=0}^{k-1} \frac{1}{(\pi_j \wedge \pi_{j+1})^\gamma (\pi_j \vee \pi_{j+1})^{1-\gamma}}, \quad (8.4.1)$$

where $c = m^2 C$, and where the sum is over all self-avoiding paths $\pi = (\pi_0, \dots, \pi_k)$ with $\pi_k = t, \pi_0 = 1$. Define

$$f_k(i, t) = \sum_{\bar{s}} \prod_{j=0}^{k-1} \frac{1}{(\pi_j \wedge \pi_{j+1})^\gamma (\pi_j \vee \pi_{j+1})^{1-\gamma}}, \quad (8.4.2)$$

where now the sum is over all self-avoiding $\pi = (\pi_0, \dots, \pi_k)$ with $\pi_k = t, \pi_0 = i$, so that

$$\mathbb{P}\left(\text{dist}_{\text{PA}_n^{(m,\delta)}}(i, t) = k\right) \leq c^k f_k(i, t). \quad (8.4.3)$$

We study the function $f_k(i, t)$ in the following lemma:

Lemma 8.13 (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 $1 \leq i < t$ and all $k \geq 1$,*

$$f_k(i, t) \leq \frac{C_{\gamma,b}^k}{i^b n^{1-b}}. \quad (8.4.4)$$

Proof We prove the lemma using induction on $k \geq 1$. To initialize the induction hypothesis, we note that, for $1 \leq i < t$ and every $b \geq a$,

$$f_1(i, t) = \frac{1}{(i \wedge t)^\gamma (i \vee t)^{1-\gamma}} = \frac{1}{i^\gamma n^{1-\gamma}} = \frac{1}{n} \left(\frac{t}{i}\right)^\gamma \leq \frac{1}{n} \left(\frac{t}{i}\right)^b = \frac{1}{i^b n^{1-b}}. \quad (8.4.5)$$

This initializes the induction hypothesis when $C_{\gamma,b} \geq 1$. To advance the induction hypothesis, note that

$$f_k(i, t) \leq \sum_{s=1}^{i-1} \frac{1}{s^\gamma i^{1-\gamma}} f_{k-1}(s, t) + \sum_{s=i+1}^{\infty} \frac{1}{i^\gamma s^{1-\gamma}} f_{k-1}(s, t). \quad (8.4.6)$$

We now bound each of these two contributions, making use of the induction hypothesis. We bound the first sum by

$$\begin{aligned} \sum_{s=1}^{i-1} \frac{1}{s^\gamma i^{1-\gamma}} f_{k-1}(s, t) &\leq C_{\gamma,b}^{k-1} \sum_{s=1}^{i-1} \frac{1}{s^\gamma i^{1-\gamma}} \frac{1}{s^b n^{1-b}} = \frac{C_{\gamma,b}^{k-1}}{i^{1-\gamma} n^{1-b}} \sum_{s=1}^{i-1} \frac{1}{s^{\gamma+b}} \\ &\leq \frac{1}{1-\gamma-b} \frac{C_{\gamma,b}^{k-1}}{i^b n^{1-b}}, \end{aligned} \quad (8.4.7)$$

since $\gamma + b < 1$. We bound the second sum by

$$\begin{aligned} \sum_{s=i+1}^{\infty} \frac{1}{i^\gamma s^{1-\gamma}} f_{k-1}(s, t) &\leq C_{\gamma,b}^{k-1} \sum_{s=i+1}^{t-1} \frac{1}{i^\gamma s^{1-\gamma}} \frac{1}{s^b n^{1-b}} + C_{\gamma,b}^{k-1} \sum_{s=t+1}^{\infty} \frac{1}{i^\gamma s^{1-\gamma}} \frac{1}{t^b s^{1-b}} \\ &= \frac{C_{\gamma,b}^{k-1}}{i^\gamma n^{1-b}} \sum_{s=i+1}^{t-1} \frac{1}{s^{1-\gamma+b}} + \frac{C_{\gamma,b}^{k-1}}{i^\gamma t^b} \sum_{s=t+1}^{\infty} \frac{1}{s^{2-\gamma-b}} \\ &\leq \frac{1}{b-\gamma} \frac{C_{\gamma,b}^{k-1}}{i^b n^{1-b}} + \frac{1}{1-\gamma-b} \frac{C_{\gamma,b}^{k-1}}{i^b n^{1-b}}, \end{aligned} \quad (8.4.8)$$

since $1 + b - a > 1$, $2 - \gamma - b > 1$, $b > \gamma$ and $(t/i)^\gamma \leq (t/i)^b$. We conclude that

$$f_k(i, t) \leq \frac{C_{\gamma,b}^{k-1}}{i^b n^{1-b}} \left(\frac{1}{b-\gamma} + \frac{2}{1-\gamma-b} \right) \leq \frac{C_{\gamma,b}^k}{i^b n^{1-b}}, \quad (8.4.9)$$

when

$$C_{\gamma,b} = \frac{1}{b-\gamma} + \frac{2}{1-\gamma-b} \geq 1. \quad (8.4.10)$$

This advances the induction hypothesis, and completes the proof of Lemma 8.13. \square

We next prove the upper bound on the diameter of $\text{PA}_n^{(m,\delta)}$. By Lemma 8.13 and (8.4.3),

$$\mathbb{P}\left(\text{dist}_{\text{PA}_n^{(m,\delta)}}(1, t) = k\right) \leq \frac{(cC_{\gamma,b})^k}{n^{1-b}}. \quad (8.4.11)$$

As a result,

$$\mathbb{P}(\text{diam}(\text{PA}_n^{(m,\delta)}) \leq k) \leq \mathbb{P}\left(\text{dist}_{\text{PA}_n^{(m,\delta)}}(1, t) \leq k\right) \leq \frac{(cC_{\gamma,b})^{k+1}}{n^{1-b}(cC_{\gamma,b} - 1)} = o(1), \quad (8.4.12)$$

whenever $k \leq \frac{1-b}{\log(cC_{\gamma,b})} \log n$. We conclude that there exists $c_2 = c_2(m, \delta)$ such that $\text{diam}(\text{PA}_n^{(m,\delta)}) \geq c_2 \log n$ whp. \square

We next extend the above discussion to *typical* distances:

Lemma 8.14 (Typical distances for $\delta > 0$) *Fix $m \geq 1$ and $\delta > 0$. Let $\text{dist}_{\text{PA}_n^{(m,\delta)}}(o_1, o_2) = \text{dist}_{\text{PA}_n^{(m,\delta)}}(V_1, V_2)$ be the distance between two uniformly chosen vertices in $[n]$. Then, whp, for $c_2 = c_2(m, \delta) > 0$ sufficiently small, $\text{dist}_{\text{PA}_n^{(m,\delta)}}(o_1, o_2) \geq c_2 \log n$.*

Proof By Lemma 8.13, with $K = \log(cC_{a,b} \vee 2)$ and $\gamma < b < 1 - \gamma$, and for all $1 \leq i < j \leq t$,

$$\mathbb{P}\left(\text{dist}_{\text{PA}_n^{(m,\delta)}}(i, j) = k\right) \leq c^k f_k(i, j) \leq \frac{e^{Kk}}{i^b j^{1-b}}. \tag{8.4.13}$$

As a result,

$$\mathbb{P}\left(\text{dist}_{\text{PA}_n^{(m,\delta)}}(i, j) \leq c_2 \log n\right) \leq \frac{n^{Kc_2}}{i^b j^{1-b}} \frac{e^K}{e^K - 1}, \tag{8.4.14}$$

and thus, using also $\sum_{i=1}^{j-1} i^{-b} \leq j^{1-b}/(1-b)$,

$$\begin{aligned} \mathbb{P}(\text{dist}_{\text{PA}_n^{(m,\delta)}}(o_1, o_2) \leq c_2 \log n) &= \frac{1}{t^2} \mathbb{P}\left(\text{dist}_{\text{PA}_n^{(m,\delta)}}(i, j) \leq c_2 \log n\right) \tag{8.4.15} \\ &\leq 2 \sum_{1 \leq i < j \leq n} \frac{n^{Kc_2}}{i^b j^{1-b}} = O(n^{Kc_2-1}) = o(1), \end{aligned}$$

for every $c_2 > 0$ such that $Kc_2 + 1 < 2$. □

8.4.2 LOWER BOUNDS ON DISTANCES FOR $\delta = 0$ AND $m \geq 2$

In this section we investigate lower bounds on the distances in $\text{PA}_n^{(m,\delta)}$ when $m \geq 2$ and $\delta = 0$ and prove the lower bound in Theorem 8.24.

We again start from Proposition 8.9, which as we will show implies that for $\delta = 0$,

$$k = \frac{\log(t-1)}{\log(3Cm^2 \log n)} \tag{8.4.16}$$

is a lower bound for the diameter of $\text{PA}_n^{(m,\delta)}$. Consider a path π of length l consisting of the vertices $\pi_0, \pi_1, \dots, \pi_l$, then (8.3.1) in Proposition 8.9 implies that

$$\mathbb{P}(\pi \subseteq \text{PA}_n^{(m,\delta)}) \leq (Cm^2)^l \prod_{j=0}^{l-1} \frac{1}{\sqrt{\pi_j \pi_{j+1}}} = \frac{(Cm^2)^l}{\sqrt{\pi_0 \pi_l}} \prod_{j=1}^{l-1} \frac{1}{\pi_j}. \tag{8.4.17}$$

Thus, the expected number of paths π of length l between $\pi_0 = n$ and $\pi_l = n - 1$ is bounded by

$$\begin{aligned} &\frac{(Cm^2)^l}{\sqrt{n(n-1)}} \sum_{1 \leq \pi_1, \dots, \pi_{l-1} \leq t-2} \prod_{j=1}^{l-1} \frac{1}{\pi_j} = \frac{(Cm^2)^l}{\sqrt{n(n-1)}} \left(\sum_{s=1}^{n-2} \frac{1}{s}\right)^{l-1} \\ &\leq \frac{(Cm^2)^l}{n-1} (\log n)^{l-1} \leq (1/2)^l (\log n)^{-1} \rightarrow 0 \end{aligned}$$

precisely when $(2Cm^2 \log n)^l \leq n - 1$, or, equivalently,

$$l \leq \frac{\log(n-1)}{\log((2Cm^2) \log n)}. \tag{8.4.18}$$

Equality in (8.4.18) holds for k in (8.4.16). This implies that the diameter is at least L in (8.4.16), and completes the proof of Theorem 8.24. □

8.4.3 TYPICAL DISTANCE IN PA-MODELS: log log-LOWER BOUND

In this section we prove the lower bound in Theorem 8.8. We do so in a more general setting assuming an upper bound on the existence of paths in the model:

Assumption 8.15 *There exist κ and γ such that, for all l and pairwise distinct vertices $\pi_0, \dots, \pi_l \in [n]$,*

$$\mathbb{P}(\pi_0 \leftrightarrow \pi_1 \leftrightarrow \dots \leftrightarrow \pi_l) \leq \prod_{i=1}^l \kappa(\pi_{i-1} \wedge \pi_i)^{-\gamma} (\pi_i \vee \pi_{i-1})^{\gamma-1}. \tag{8.4.19}$$

By Proposition 8.9, Assumption 8.15 is satisfied for $\text{PA}_n^{(m,\delta)}$ with $\gamma = m/(2m + \delta)$. We expect log log-distances in such networks if and only if $\delta \in (-m, 0)$, so that $\frac{1}{2} < \gamma < 1$. Theorem 8.16, which is the main result in this section, gives a lower bound on the typical distance in this case:

Theorem 8.16 (Doubly logarithmic lower bound on distances PAMs) *Let $(\text{PA}_n)_{t \in \mathbb{N}}$ be a random graph model that satisfies Assumption 8.15 for some γ satisfying $\frac{1}{2} < \gamma < 1$. Then, for random vertices V_1 and V_2 chosen independently and uniformly from $[n]$ and whp as K grows large,*

$$\text{dist}_{\text{PA}_n}(V_1, V_2) \geq \frac{4 \log \log n}{\log(\gamma/(1-\gamma))} - K. \tag{8.4.20}$$

For $\text{PA}_n^{(m,\delta)}$, $\gamma = m/(2m + \delta)$, so that

$$\frac{\gamma}{1-\gamma} = \frac{m}{m+\delta} = \frac{1}{\tau-2}, \tag{8.4.21}$$

where we recall that $\tau = 3 + \delta/m$. Therefore, Theorem 8.16 proves the lower bound in Theorem 8.8.

TO DO 8.2: Adapt this argument as in ?!

The proof of Theorem 8.16 is based on a *constrained* or *truncated first order method*, similar to the ones used for $\text{NR}_n(\mathbf{w})$ in Theorem 6.6 and for $\text{CM}_n(\mathbf{d})$ in Theorem 7.6. Due to the fact that the probability for existence of paths satisfies a rather different bound compared to the bounds for $\text{NR}_n(\mathbf{w})$ and $\text{CM}_n(\mathbf{d})$, this truncated first order method looks rather different compared to the ones 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 . Then for $k_n \in \mathbb{N}$

$$\begin{aligned} \mathbb{P}(\text{dist}_{\text{PA}_n^{(m,\delta)}}(v, w) \leq 2k_n) &= \mathbb{P}\left(\bigcup_{k=1}^{2k_n} \bigcup_{\pi} \{v \leftrightarrow \pi_1 \leftrightarrow \pi_2 \leftrightarrow \dots \leftrightarrow \pi_{k-1} \leftrightarrow w\}\right) \\ &\leq \sum_{k=1}^{2k_n} \sum_{\pi} \prod_{j=1}^k p(\pi_{j-1}, \pi_j), \end{aligned} \tag{8.4.22}$$

where $\pi = (\pi_0, \dots, \pi_k)$ is any collection of pairwise distinct vertices in PA_n with $\pi_0 = v$ and $\pi_k = w$ and, for $n, m \in \mathbb{N}$, we define

$$p(n, m) = \kappa(n \wedge m)^{-\gamma} (n \vee m)^{\gamma-1}. \tag{8.4.23}$$

We assign to each path $\pi = (\pi_0, \dots, \pi_k)$ the weight

$$p(\pi) = \prod_{j=1}^k p(\pi_{j-1}, \pi_j), \tag{8.4.24}$$

and the upper bound is just the sum over the weights of all paths from v to w of length no more than $2k_n$. The shortcoming of this bound is that the paths that contribute most to the total weight are those that connect v , resp. w , quickly to vertices with extremely small indices and thus extremely high degree. Since such paths are quite unlikely, they have to be removed in order to get a reasonable estimate.

To this end we define a decreasing sequence $\ell = (\ell_k)_{k=0, \dots, k}$ of positive integers and consider a tuple of vertices $\pi = (\pi_0, \dots, \pi_k)$ as *good* if $v_l \wedge v_{k-l} \geq \ell_l$ for all $l \in \{0, \dots, k\}$. We denote the probability that there exists a good path of length k between v and w by $\mathcal{E}_k(v, w)$. We further denote by $\mathcal{F}_k(v)$ the event that there exists a bad path of length k in the network starting at v , i.e., a path $v = \pi_0 \leftrightarrow \dots \leftrightarrow \pi_l$ such that $\pi_0 \geq \ell_0, \dots, \pi_{l-1} \geq \ell_{l-1}$, but $\pi_l < \ell_l$, i.e., a path that traverses the threshold after exactly l steps. For fixed vertices $v, w \geq \ell_0$, the truncated first moment estimate is the estimate that

$$\mathbb{P}(\text{dist}_{\text{PA}_t}(v, w) \leq 2k) \leq \sum_{l=1}^k \mathbb{P}(\mathcal{F}_l(v)) + \sum_{l=1}^k \mathbb{P}(\mathcal{F}_l(w)) + \sum_{k=1}^{2k} \mathbb{P}(\mathcal{E}_k(v, w)) \tag{8.4.25}$$

where the good paths in the last sum start with $v_0 = v$ and end with $v_k = w$. Equation (8.4.25) 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 PA_n are the most likely to have large degrees. This explains why good vertices have high indices for PA_n , while good vertices have high weights for $\text{NR}_n(\mathbf{w})$.

By assumption,

$$\mathbb{P}(\pi_0 \leftrightarrow \dots \leftrightarrow \pi_k) \leq p(\pi_0, \dots, \pi_k) \tag{8.4.26}$$

so that for $v \geq \ell_0$ and $l = 1, \dots, k$, and with $\pi = (\pi_0, \dots, \pi_l)$ with $\pi_0 = v$,

$$\mathbb{P}(\mathcal{F}_l(v)) \leq \sum_{\pi_1=\ell_1}^n \dots \sum_{\pi_{l-1}=\ell_{l-1}}^n \sum_{\pi_l=1}^{\ell_{l-1}} p(\pi). \tag{8.4.27}$$

Given $\varepsilon > 0$ we choose $\ell_0 = \lceil \varepsilon t \rceil$ and $(\ell_j)_{j=0, \dots, k}$ decreasing fast enough so that the first two summands on the right hand side of (8.4.25) together are no larger than 2ε . For $l \in [k]$, we set

$$f_{l,t}(v, u) := \mathbb{1}_{\{v \geq \ell_0\}} \sum_{\pi_1=\ell_1}^n \dots \sum_{\pi_{l-1}=\ell_{l-1}}^n p(v, \pi_1, \dots, \pi_{l-1}, u), \tag{8.4.28}$$

and set $f_{0,t}(v, u) = \mathbb{1}_{\{v=u\}} \mathbb{1}_{\{u \leq n\}}$. To rephrase the truncated moment estimate in terms of f , note that p is symmetric so that for all $l \leq 2k$

$$\begin{aligned} \mathbb{P}(\mathcal{E}_l(v, w)) &\leq \sum_{\pi_1=\ell_1}^n \cdots \sum_{\pi_{\lfloor l/2 \rfloor}=\ell_{\lfloor l/2 \rfloor}}^n \cdots \sum_{\pi_{l-1}=\ell_1}^n p(v, \pi_1, \dots, \pi_{\lfloor l/2 \rfloor}) p(\pi_{\lfloor l/2 \rfloor}, \dots, \pi_{l-1}, w) \\ &= \sum_{\pi_{\lfloor l/2 \rfloor}=\ell_{\lfloor l/2 \rfloor}}^n f_{\lfloor l/2 \rfloor, t}(v, \pi_{\lfloor l/2 \rfloor}) f_{\lfloor l/2 \rfloor, t}(w, \pi_{\lfloor l/2 \rfloor}). \end{aligned} \tag{8.4.29}$$

Using the recursive representation

$$f_{k+1,t}(v, n) = \sum_{m=\ell_k}^n f_{k,t}(v, m) p(m, n), \tag{8.4.30}$$

we establish upper bounds for $f_{k,t}(v, u)$ and use these to show that the rightmost term in (8.4.25) remains small if $k = k_n$ is chosen sufficiently small. This leads to the lower bounds for the typical distance in Theorem 8.16. Let us now make these ideas precise:

Proof of Theorem 8.16. We assume that Assumption 8.15 holds for a $\gamma \in (\frac{1}{2}, 1)$ with a fixed constant κ . Recall the definition of $f_{k,t}$ and the key estimates (8.4.25), (8.4.27) and (8.4.29), which combined give

$$\begin{aligned} \mathbb{P}(\text{dist}_{\text{PA}_n}(v, w) \leq 2k_n) &\leq \sum_{k=1}^{k_n} \sum_{l=1}^{\ell_k-1} f_{k,t}(v, l) + \sum_{k=1}^{k_n} \sum_{l=1}^{\ell_k-1} f_{k,t}(w, l) \\ &\quad + \sum_{k=1}^{2k_n} \sum_{v_{\lfloor k/2 \rfloor}=\ell_{\lfloor n/2 \rfloor}}^n f_{\lfloor n/2 \rfloor, t}(v, v_{\lfloor n/2 \rfloor}) f_{\lfloor n/2 \rfloor, t}(w, v_{\lfloor n/2 \rfloor}). \end{aligned} \tag{8.4.31}$$

The remaining task of the proof is to choose $k_n \in \mathbb{N}$ and $2 \leq \ell_{k_n} \leq \dots \leq \ell_0 \leq t$ which allow the required estimates for the right-hand side. Denote the truncated version of $f_{k,t}(v, m)$ by $\bar{f}_{k,t}(v, m) = \mathbb{1}_{\{m \geq \ell_k\}} f_{k,t}(v, m)$. Our aim is to provide a majorant of the form

$$f_{k,t}(v, m) \leq \alpha_k m^{-\gamma} + \mathbb{1}_{\{m > \ell_{k-1}\}} \beta_k m^{\gamma-1} \tag{8.4.32}$$

for suitably chosen parameters $\alpha_k, \beta_k \geq 0$. Key to this choice is the following lemma:

Lemma 8.17 (A recursive bound on $f_{k,t}(v, m)$ for $\gamma \in (1/2, 1)$) *Let $\gamma \in (1/2, 1)$ and suppose that $2 \leq \ell \leq t$, $\alpha, \beta \geq 0$ and $q: [n] \rightarrow [0, \infty)$ satisfies*

$$q(m) \leq \mathbb{1}_{\{m \geq \ell\}} (\alpha m^{-\gamma} + \beta m^{\gamma-1}) \quad \text{for all } m \in [n]. \tag{8.4.33}$$

Then there exists a constant $c > 1$ depending only on γ and κ such that

$$\sum_{k=1}^n q(k) p(k, m) \leq c (\alpha \log(t/\ell) + \beta n^{2\gamma-1}) m^{-\gamma} + c \mathbb{1}_{\{m > \ell\}} (\alpha \ell^{1-2\gamma} + \beta \log(t/\ell)) m^{\gamma-1} \tag{8.4.34}$$

for all $m \in [n]$.

Proof We use (8.4.23) to rewrite

$$\begin{aligned} \sum_{k=1}^n q(k)p(k, m) &= \sum_{k=m \vee \ell}^n q(k)p(k, m) + \mathbb{1}_{\{m > \ell\}} \sum_{k=\ell}^{m-1} q(k)p(k, m) \\ &= \sum_{k=m \vee \ell}^n \kappa(\alpha k^{-\gamma} + \beta k^{\gamma-1})k^{\gamma-1}m^{-\gamma} + \mathbb{1}_{\{m > \ell\}} \sum_{k=\ell}^{m-1} \kappa(\alpha k^{-\gamma} + \beta k^{\gamma-1})k^{-\gamma}m^{\gamma-1}. \end{aligned}$$

Simplifying the sums leads to

$$\begin{aligned} \sum_{k=1}^n q(k)p(k, m) &\leq \kappa \left(\alpha \sum_{k=m \vee \ell}^n k^{-1} + \beta \sum_{k=m \vee \ell}^n k^{2\gamma-2} \right) m^{-\gamma} \\ &\quad + \kappa \mathbb{1}_{\{m > \ell\}} \left(\alpha \sum_{k=\ell}^{m-1} k^{-2\gamma} + \beta \sum_{k=\ell}^{m-1} k^{-1} \right) m^{\gamma-1} \\ &\leq \kappa \left(\alpha \log \left(\frac{m}{\ell-1} \right) + \frac{\beta}{2\gamma-1} n^{2\gamma-1} \right) m^{-\gamma} \\ &\quad + \kappa \mathbb{1}_{\{m > \ell\}} \left(\frac{\alpha}{1-2\gamma} (\ell-1)^{1-2\gamma} + \beta \log \left(\frac{m}{\ell-1} \right) \right) m^{\gamma-1}. \end{aligned} \tag{8.4.35}$$

This immediately implies the assertion since $\ell \geq 2$ by assumption. \square

We apply Lemma 8.17 iteratively. We use induction to prove that there exist $(\ell_k)_{k \geq 0}$, $(\alpha_k)_{k \geq 1}$ and $(\beta_k)_{k \geq 1}$ such that

$$f_{k,t}(v, m) \leq \alpha_k m^{-\gamma} + \beta_k m^{\gamma-1} \quad \text{for all } m \in [n]. \tag{8.4.36}$$

The sequences $(\ell_k)_{k \geq 0}$, $(\alpha_k)_{k \geq 1}$ and $(\beta_k)_{k \geq 1}$ are chosen as follows. We let $\ell_0, \alpha_1, \beta_1$ be determined by

$$\ell_0 = \lceil \varepsilon t \rceil, \quad \alpha_1 = \kappa(\varepsilon t)^{\gamma-1} \quad \text{and} \quad \beta_1 = \kappa(\varepsilon t)^{-\gamma}, \tag{8.4.37}$$

where $\varepsilon > 0$ is small.

For higher values of k , $\alpha_{k+1}, \beta_{k+1}$ and ℓ_k satisfy the recursions

$$\frac{6\varepsilon}{\pi^2 k^2} \geq \frac{1}{1-\gamma} \alpha_k \ell_k^{1-\gamma}, \tag{8.4.38}$$

where we let ℓ_k be the largest integer satisfying the above inequality under the assumption that $\ell_k \geq 2$, and

$$\alpha_{k+1} = c(\alpha_k \log(t/\ell_k) + \beta_k n^{2\gamma-1}), \quad \beta_{k+1} = c(\alpha_k \ell_k^{1-2\gamma} + \beta_k \log(t/\ell_k)), \tag{8.4.39}$$

TO DO 8.3: Check whether equality is okay??

where c is the constant in Lemma 8.17.

We start by initializing the induction. Fix $v \geq \ell_0$. Then, for all $m \in [n]$,

$$\begin{aligned} f_{1,t}(v, m) &= p(v, m) \leq k \ell_0^{\gamma-1} m^{-\gamma} + \mathbb{1}_{\{m > \ell_0\}} \kappa \ell_0^{-\gamma} m^{\gamma-1} \\ &\leq \alpha_1 m^{-\gamma} + \mathbb{1}_{\{m > \ell_0\}} \beta_1 m^{\gamma-1}. \end{aligned} \tag{8.4.40}$$

Now suppose, for some $k \in \mathbb{N}$, that we have chosen α_k, β_k and an integer ℓ_{k-1} such that

$$f_{k,t}(v, m) \leq \alpha_k m^{-\gamma} + \beta_k m^{\gamma-1} \quad \text{for all } m \in [n]. \quad (8.4.41)$$

By the induction hypothesis we can apply Lemma 8.17 with $\ell = \ell_k$ and $q(m) = \bar{f}_{k,t}(v, m) = \mathbb{1}_{\{m \geq \ell_k\}} f_{k,t}(v, m)$. Then Lemma 8.17 yields that

$$f_{k+1,t}(v, m) \leq \alpha_{k+1} m^{-\gamma} + \mathbb{1}_{\{m > \ell_k\}} \beta_{k+1} m^{\gamma-1} \quad \text{for all } m \in [n], \quad (8.4.42)$$

showing that the induction can be advanced up to the point where $\ell_k < 2$. Having advanced the induction hypothesis, we obtain that (8.4.36) holds with the given choices of $(\ell_k)_{k \geq 0}$, $(\alpha_k)_{k \geq 1}$ and $(\beta_k)_{k \geq 1}$.

We next use (8.4.36) to prove Theorem 8.16. Summing over (8.4.42) and using (8.4.36) and (8.4.38) we obtain

$$\sum_{l=1}^{\ell_k-1} f_{k,t}(v, l) \leq \frac{1}{1-\gamma} \alpha_k \ell_k^{1-\gamma} \leq \frac{6\varepsilon}{\pi^2 k^2}, \quad (8.4.43)$$

which, when summed over all $k \geq 1$ is bounded by ε . Hence the first two summands on the right-hand side in (8.4.31) together are smaller than 2ε . It remains to choose $k_n = k_n(t)$ as large as possible while ensuring that $\ell_{k_n} \geq 2$ and

$$\lim_{t \rightarrow \infty} \sum_{k=1}^{2k_n} \sum_{\pi_{\lfloor k/2 \rfloor} = \ell_{\lfloor k/2 \rfloor}}^n f_{\lfloor k/2 \rfloor, t}(v, \pi_{\lfloor k/2 \rfloor}) f_{\lceil k/2 \rceil, t}(w, \pi_{\lceil k/2 \rceil}) = 0. \quad (8.4.44)$$

To this end recall that ℓ_k is the largest integer satisfying (8.4.38) and the parameters α_k, β_k are defined via equalities in (8.4.39). To establish lower bounds for the decay of ℓ_k we investigate the growth of $\eta_k = t/\ell_k > 0$. Going backwards through the definitions yields, for $k \geq 1$, that there exists a constant $C > 0$ such that

$$\eta_{k+2}^{1-\gamma} \leq C \left[\frac{(k+2)^2}{k^2} \eta_k^\gamma + \frac{(k+2)^2}{(k+1)^2} \eta_{k+1}^{1-\gamma} \log \eta_{k+1} \right], \quad (8.4.45)$$

with $\eta_1, \eta_2 \leq C_0$ for some constant $C_0 > 0$ (which, as all constants in this paragraph, may depend on ε). Indeed, writing C for a constant that may change from line to line, using first the relation for ℓ_{k+2} in (8.4.38), followed by the equality for α_{k+2} in (8.4.39),

$$\eta_{k+2}^{1-\gamma} = n^{1-\gamma} \ell_{k+2}^{\gamma-1} = C n^{1-\gamma} \alpha_{k+2} (k+2)^2 = C n^{1-\gamma} (k+2)^2 (\alpha_{k+1} \log(\eta_{k+1}) + \beta_{k+1} n^{2\gamma-1}). \quad (8.4.46)$$

The first term on the right-hand side of (8.4.46) equals $C \frac{(k+2)^2}{(k+1)^2} \eta_{k+1}^{1-\gamma} \log(\eta_{k+1})$, which is the second term on the right-hand side of (8.4.45). For the second term on the right-hand side of (8.4.46), we use the equality for β_{k+1} in (8.4.39) to write it as

$$C n^\gamma (k+2)^2 \beta_{k+1} = C n^\gamma (k+2)^2 (\alpha_k \ell_k^{1-2\gamma} + \beta_k \log(\eta_k)). \quad (8.4.47)$$

Now the first term is equal to $C \frac{(k+2)^2}{k^2} \eta_{k+1}^\gamma$, which is the first term in (8.4.45). We

are left with $Cn^\gamma(k+2)^2\beta_k \log(\eta_k)$. Since $c\beta_k n^{2\gamma-1} \leq \alpha_{k+1}$,

$$\begin{aligned} Cn^\gamma(k+2)^2\beta_k \log(\eta_k) &\leq Cn^{1-\gamma}(k+2)^2\alpha_{k+1} \log(\eta_k) \leq Cn^{1-\gamma} \frac{(k+2)^2}{(k+1)^2} \ell_{k+1}^{-(1-\gamma)} \\ &= C\eta_{k+1}^{1-\gamma} \log(\eta_k) \leq C\eta_{k+1}^{1-\gamma} \log(\eta_{k+1}), \end{aligned} \quad (8.4.48)$$

since $k \mapsto \ell_k$ is decreasing, so that $k \mapsto \eta_k$ is increasing. This completes the proof of (8.4.45).

Lemma 8.18 (Inductive bounds on η_k) *Let $(\eta_k)_{k \geq 1}$ satisfy $\eta_0 = 1/\varepsilon$ and $\eta_1 =$, and (8.4.45) holds for $k \geq 2$. Then there exist constants $b, B > 0$ such that*

$$\eta_k \leq b \exp\left(B\left(\frac{\gamma}{1-\gamma}\right)^{k/2}\right). \quad (8.4.49)$$

Proof By assumption, (8.4.49) follows for $k = 0$ by the fact that $\eta_0 = t/\ell_0 \leq 1/\varepsilon$. By (8.4.45), we can obtain similar bounds for η_k with $k \leq 2$. Suppose that we know that (8.4.49) holds for all $l \leq k-1$ where $k \geq 3$. Then (8.4.45) together with the fact that $k/(k-1)$ and $k/(k-2)$ are bounded yields that there exists a constant $C > 0$ such that

$$\eta_k^{1-\gamma} \leq C\eta_{k-2}^\gamma + C\eta_{k-1}^{1-\gamma} \log \eta_{k-1}, \quad (8.4.50)$$

and using that $(x+y)^{1/(1-\gamma)} \leq 2^{1/(1-\gamma)}(x^{1/(1-\gamma)} + y^{1/(1-\gamma)})$ leads us to

$$\eta_k \leq C\eta_{k-2}^{\gamma/(1-\gamma)} + C\eta_{k-1}(\log \eta_{k-1})^{1/(1-\gamma)}. \quad (8.4.51)$$

Iterating the above inequality once more yields

$$\eta_k \leq C(2C)^{\gamma/(1-\gamma)}\eta_{k-4}^{\gamma/(1-\gamma)^2} + C(2C)^{\gamma/(1-\gamma)}\eta_{k-3}^{\gamma/(1-\gamma)}(\log \eta_{k-3})^{\gamma/(1-\gamma)^2} + C\eta_{k-1}(\log \eta_{k-1})^{1/(1-\gamma)}. \quad (8.4.52)$$

Iterating indefinitely yields

$$\eta_k \leq C(2C)^{\sum_{i=1}^{k/2} (\gamma/(1-\gamma))^i} \eta_0^{(\gamma/(1-\gamma))^{k/2}} + C \sum_{i=1}^{k/2} (2C)^{\sum_{j=1}^{i-1} (\gamma/(1-\gamma))^j} \eta_{k-2i+1}^{(\gamma/(1-\gamma))^{i-1}} (\log \eta_{k-2i+1})^{\gamma^{i-1}/(1-\gamma)^i}. \quad (8.4.53)$$

We now prove by induction that there exists constants b, B such that

$$\eta_k \leq b \exp\left(B\left(\frac{\gamma}{1-\gamma}\right)^{k/2}\right). \quad (8.4.54)$$

Using the induction hypothesis, we can bound η_k by

TO DO 8.4: Continue from here!

$$\eta_k^{1-\gamma} \leq CA^{\gamma(k-1)} \exp\left(B(1-\gamma)\left(\frac{\gamma}{1-\gamma}\right)^{k/2}\right) + Cb^{1-\gamma} \exp\left(B\sqrt{(1-\gamma)^3/\gamma}\left(\frac{\gamma}{1-\gamma}\right)^{k/2}\right) B\left(\frac{\gamma}{1-\gamma}\right)^{(k-1)/2}. \quad (8.4.55)$$

Since $\gamma > 1/2$, we have that $(1-\gamma)/\gamma < 1$, so that for large k the second term is negligible compared to the first term for every $k \geq 2$ when B is sufficiently

large. For $\gamma > 1/2$... Taking b large enough, we can bound the right-hand side of (8.4.55) by $b^{1-\gamma} \exp\left(B(1-\gamma)\left(\frac{\gamma}{1-\gamma}\right)^{k/2}\right)$ \square

We now use (8.4.42) and (8.4.49) to estimate

$$\begin{aligned} \sum_{n=1}^{2k_n} \sum_{u=\ell_k}^n \mu_{n^*}^{(v)}(u) \mu_{n-n^*}^{(w)}(u) &\leq 2 \sum_{k=1}^{k_n} \sum_{u=\ell_k}^n (\alpha_k u^{-\gamma} + \beta_k u^{\gamma-1})^2 \\ &\leq \frac{4}{2\gamma-1} \sum_{k=1}^{k_n} (\alpha_k^2 \ell_k^{1-2\gamma} + \beta_k^2 n^{2\gamma-1}) \\ &\leq \frac{4}{2\gamma-1} k_n (\alpha_{k_n}^2 \ell_{k_n}^{1-2\gamma} + \beta_{k_n}^2 n^{2\gamma-1}). \end{aligned} \tag{8.4.56}$$

Using (8.4.38) and (8.4.49) the first summand in the bracket can be estimated by

$$\alpha_{k_n}^2 \ell_{k_n}^{1-2\gamma} \leq \left(k_n^{-2} \frac{6\varepsilon}{\pi^2} (1-\gamma)\right)^2 \ell_{k_n}^{-1} \leq \left(\frac{6\varepsilon}{b\pi^2} (1-\gamma)\right)^2 \frac{1}{tk_n^4} \exp\left(B\left(\frac{\gamma}{1-\gamma}\right)^{k_n/2}\right). \tag{8.4.57}$$

Using equality in (8.4.39) we get $\beta_{k_n} \leq c(\alpha_{k_n} \ell_{k_n}^{1-2\gamma} + \alpha_{k_n} n^{1-2\gamma} \log(t/\ell_{k_n}))$. Noting that the second summand on the right-hand side is bounded by a multiple of the first, we find a constant $C_1 > 0$ such that $\beta_{k_n}^2 n^{2\gamma-1} \leq C_1 \alpha_{k_n}^2 \ell_{k_n}^{1-2\gamma}$, and thus, for a suitable constant $C_2 > 0$,

$$\sum_{n=1}^{2k_n} \sum_{u=\ell_k}^n \mu_{n^*}^{(v)}(u) \mu_{n-n^*}^{(w)}(u) \leq C_2 \frac{1}{tk_n^3} \exp\left(B\left(\frac{\gamma}{1-\gamma}\right)^{k_n/2}\right). \tag{8.4.58}$$

Hence, for a suitable constant $C > 0$, choosing

$$k_n \leq \frac{\log \log n}{\log \sqrt{\frac{\gamma}{1-\gamma}}} - K, \tag{8.4.59}$$

we obtain that the term we consider goes to zero of order $O((\log \log n)^{-3})$. Note from (8.4.49) that this choice also ensures that $\ell_{k_n} \geq 2$. We have thus shown that

$$\mathbb{P}(\text{dist}_{\text{PA}_n^{(m,\delta)}}(v, w) \leq 2k_n) \leq 2\varepsilon + O((\log \log n)^{-3}), \tag{8.4.60}$$

whenever $v, w \geq \ell_0 = \lceil \varepsilon t \rceil$, which implies the statement of Theorem 8.16. \square

8.5 SMALL-WORLD EFFECT IN PA MODELS: UPPER BOUNDS

8.5.1 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 that the logarithmic bounds on the diameter for $(\text{PA}_n^{(m,\delta)})_{n \geq 1}$:

Proof of Theorem 8.6. We start by proving the claim for $(\text{PA}_{m,\delta}^{(b)}(t))_{n \geq 1}$. Since $(\text{PA}_{m,\delta}^{(b)}(t))_{n \geq 1}$ is obtained from $(\text{PA}_{1,\delta/m}^{(b)}(mt))_{n \geq 1}$ by collapsing m successive vertices, $\text{diam}(\text{PA}_{m,\delta}^{(b)}(t)) \preceq \text{diam}(\text{PA}_{1,\delta/m}^{(b)}(mt))$, and the result follows from Theorem 8.3.

□

8.5.2 THE DIAMETER OF THE CORE

In this section we investigate $PA_n^{(m,\delta)}$ with $m \geq 2$ and $\delta \in (-m, 0)$ and prove the upper bounds in Theorem 8.8.

The proof of Theorem 8.25 is divided into two key steps. In the first, in Theorem 8.19, 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 the average distance for the configuration model, but substantial adaptations are necessary to deal with preferential attachment. After this, in Theorem ??, 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 8.25 for $2t$ rather than for t . Clearly, this does not make any difference for the results.

We adapt the proof of Theorem 7.9 to $PA_n^{(m,\delta)}$. We recall that

$$\tau = 3 + \frac{\delta}{m}, \tag{8.5.1}$$

so that $-m < \delta < 0$ corresponds to $\tau \in (2, 3)$. Throughout this section, we fix $m \geq 2$.

We take $\sigma > \frac{1}{3-\tau} = -\frac{m}{\delta} > 1$ and define the *core* $Core_t$ of the PA-model $PA_{m,\delta}(2t)$ to be

$$Core_t = \{i \in [n] : D_i(t) \geq (\log n)^\sigma\}, \tag{8.5.2}$$

i.e., all the vertices which at time n have degree at least $(\log n)^\sigma$.

For a graph G with vertex set $[n]$ and a given edge set, we write $dist_G(i, j)$ for the shortest-path distance between i and j in the graph G . Also, for $A \subseteq [n]$, we write

$$diam_t(A) = \max_{i,j \in A} dist_{PA_n^{(m,\delta)}}(i, j). \tag{8.5.3}$$

Then, the diameter of the core in the graph $PA_{m,\delta}(2t)$, which we denote by $diam_{2t}(Core_t)$, is bounded in the following theorem:

Theorem 8.19 (The diameter of the core) *Fix $m \geq 2$. For every $\sigma > \frac{1}{3-\tau}$, whp,*

$$diam_{2t}(Core_t) \leq (1 + o(1)) \frac{4 \log \log n}{|\log(\tau - 2)|}. \tag{8.5.4}$$

The proof of Theorem 8.19 is divided into several smaller steps. We start by proving that the diameter of the *inner core* $Inner_t$, which is defined by

$$Inner_t = \{i \in \{1, 2, \dots, t\} : D_i(t) \geq n^{\frac{1}{2(\tau-1)}} (\log n)^{-\frac{1}{2}}\}, \tag{8.5.5}$$

is, whpbounded by $C_\delta < \infty$. After this, we will show that the distance from the *outer core*, which is defined to be equal to $Outer_t = Core_n \setminus Inner_t$, 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.

Proposition 8.20 (The diameter of the inner core) *Fix $m \geq 2$ and $\delta \in (-m, 0)$. Then, whp,*

$$\text{diam}_{2t}(\text{Inner}_t) < C_\delta. \quad (8.5.6)$$

Proof We first introduce the important notion of a t -connector between a vertex $i \in \{1, 2, \dots, t\}$ and a set of vertices $A \subseteq \{1, 2, \dots, t\}$, which plays a crucial role throughout the proof. Fix a set of vertices A and a vertex i . We say that the vertex $j \in \{t+1, t+2, \dots, 2t\}$ is a t -connector between i and A if one of the edges incident to j connects to i and another edge incident to j connects to a vertex in A . Thus, when there exists a t -connector between i and A , the distance between i and A in $\text{PA}_{m,\delta}(2t)$ is at most 2.

We note that for a set of vertices A and a vertex i with degree at time n equal to $D_i(t)$, we have that, conditionally on $\text{PA}_n^{(m,\delta)}$, the probability that $j \in \{t+1, t+2, \dots, 2t\}$ is a t -connector for i and A is at least

$$\frac{(D_A(t) + \delta|A|)(D_i(t) + \delta)}{[2t(2m + \delta)]^2}, \quad (8.5.7)$$

independently of the fact whether the other vertices are t -connectors or not, and where, for any $A \subseteq \{1, 2, \dots, t\}$, we write

$$D_A(t) = \sum_{i \in A} D_i(t). \quad (8.5.8)$$

Since $d_i(t) + \delta \geq m + \delta > 0$ for every $i \leq t$, and $\delta < 0$, we have that

$$D_i(t) + \delta = D_i(t) \left(1 + \frac{\delta}{D_i(t)}\right) \geq D_i(t) \left(1 + \frac{\delta}{m}\right) = D_i(t) \frac{m + \delta}{m}, \quad (8.5.9)$$

and, thus, also $D_A(t) + \delta|A| \geq D_A(t) \frac{m + \delta}{m}$. As a result, for $\eta = (m + \delta)^2 / (2m(2m + \delta))^2 > 0$, the probability that $j \in \{t+1, t+2, \dots, 2t\}$ is a t -connector for i and A is at least $\frac{\eta D_A(t) D_i(t)}{t^2}$, independently of the fact whether the other vertices are t -connectors or not. Therefore, the probability that there is no t -connector for i and A is, conditionally on $\text{PA}_n^{(m,\delta)}$, bounded above by

$$\left(1 - \frac{\eta D_A(t) D_i(t)}{t^2}\right)^t \leq \exp \left\{ -\frac{\eta D_A(t) D_i(t)}{t} \right\}. \quad (8.5.10)$$

We shall make use of (8.5.10) in several places throughout the proof.

From [Volume 1, Theorem 8.3] whp, Inner_t contains at least \sqrt{t} vertices and denote the first \sqrt{t} vertices of Inner_t by I . Observe that for $\tau > 2$ we have $n^{(\tau-1)^{-1}-1} \downarrow 0$ so that, for any $i, j \in I$, the probability that there exists a t -connector for i and j is bounded below by

$$1 - \exp\{-\eta n^{\frac{1}{\tau-1}-1} (\log n)^{-1}\} \geq p_t \equiv n^{\frac{1}{\tau-1}-1} (\log n)^{-2}, \quad (8.5.11)$$

for t sufficiently large.

We wish to couple Inner_t with an Erdős-Rényi random graph with $n_t = \sqrt{t}$ vertices and edge probability p_t , which we denote by $\text{ER}(n_t, p_t)$. For this, for $i, j \in \{1, 2, \dots, n_t\}$, we say that an edge between i and j is present when there exists a t -connector connecting the i th and j th vertex in I . We now prove that this graph is bounded below by $\text{ER}(n_t, p_t)$. Note that (8.5.11) 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 $n_t(n_t - 1)/2$ 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_t , for every l . This would prove the claimed stochastic domination by $\text{ER}(n_t, p_t)$.

Indeed, the l th edge is present precisely when there exists a t -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 t -connect the previous edges. This removes at most $n_t(n_t - 1)/2 \leq t/2$ possible t -connectors, after which at least another $t/2$ remain. The probability that one of them is a t -connector for the i th and j th vertex in I is bounded below by, for t sufficiently large,

$$1 - \exp\{-\eta n^{\frac{1}{\tau-1}-2}(\log n)^{-1}t/2\} = 1 - \exp\{-\eta n^{\frac{1}{\tau-1}-1}(\log n)^{-1}/2\} \geq p_t \equiv n^{\frac{1}{\tau-1}-1}(\log n)^{-2}, \quad (8.5.12)$$

using $1 - e^{-x} \geq x/2$ for $x \in [0, 1]$ and $\eta/2 \geq \log n^{-1}$ for t sufficiently large.

This proves the claimed stochastic domination of the random graph on the vertices I and $\text{ER}(n_t, p_t)$. Next, we show that $\text{diam}(\text{ER}(n_t, p_t))$ 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 $p^d n^{d-1} - 2 \log n \rightarrow \infty$, while $p^{d-1} n^{d-2} - 2 \log n \rightarrow -\infty$, then $\text{diam}(\text{ER}(n, p)) = d$, whp. In our case, $n = n_t = n^{1/2}$ and $p = p_t = n^{\frac{1}{\tau-1}-1}(\log n)^{-2}$, 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 $\text{PA}_{m,\delta}(2t)$ is whp bounded by $2(\frac{\tau-1}{3-\tau} + 1)$ in this case.

We finally show that for any $i \in \text{Inner}_t \setminus I$, the probability that there does not exist a t -connector connecting i and I is small. Indeed, this probability is, since $D_I(t) \geq \sqrt{t} n^{\frac{1}{\tau-1}}(\log n)^{-1/2}$, and $D_i(t) \geq n^{\frac{1}{2(\tau-1)}}(\log n)^{-1/2}$, the probability of there not existing a t -connector is bounded above by $e^{-\eta m^{1/(\tau-1)-1/2}(\log n)^{-1}}$, which is tiny since $\tau < 3$. This proves that whp the distance between any vertex $i \in \text{Inner}_t \setminus I$ and I is bounded by 2, and, together with the fact that $\text{diam}_{2t}(I) \leq 2(\frac{\tau-1}{3-\tau} + 1)$ thus implies that $\text{diam}_{2t}(\text{Inner}_t) \leq 2(\frac{\tau-1}{3-\tau} + 2)$. \square

Proposition 8.21 (Distance between outer and inner core) *Fix $m \geq 2$. With high probability, the inner core Inner_t can be reached from any vertex in the outer core Outer_t using no more than $\frac{2 \log \log n}{|\log(\tau-2)|}$ edges in $\text{PA}_{m,\delta}(2t)$. More precisely, whp*

$$\max_{i \in \text{Outer}_t} \min_{j \in \text{Inner}_t} \text{dist}_{\text{PA}_{m,\delta}(2t)}(i, j) \leq \frac{2 \log \log n}{|\log(\tau-2)|}. \quad (8.5.13)$$

Proof Recall that

$$\text{Outer}_t = \text{Core}_t \setminus \text{Inner}_t. \quad (8.5.14)$$

and define

$$\mathcal{N}_1 = \text{Inner}_t = \{i : D_i(t) \geq \mathcal{W}_1\}, \quad (8.5.15)$$

where

$$\mathcal{W}_1 = l_t = n^{\frac{1}{2(\tau-1)}} (\log n)^{-\frac{1}{2}}. \quad (8.5.16)$$

We now recursively define a sequence u_k , for $k \geq 2$, so that for any vertex $i \in \{1, 2, \dots, t\}$ with degree at least u_k , the probability that there is no t -connector for the vertex i and the set

$$\mathcal{N}_{k-1} = \{j : D_j(t) \geq \mathcal{W}_{k-1}\}, \quad (8.5.17)$$

conditionally on $\text{PA}_n^{(m, \delta)}$ is tiny. According to (8.5.10) and [Volume 1, Exercise 8.20], this probability is at most

$$\exp \left\{ - \frac{\eta B t [u_{k-1}]^{2-\tau} u_k}{t} \right\} = o(n^{-1}), \quad (8.5.18)$$

where we define

$$\mathcal{W}_k = D \log n (\mathcal{W}_{k-1})^{\tau-2}, \quad (8.5.19)$$

with D exceeding $(\eta B)^{-1}$. By Lemma 7.10 we have

$$\mathcal{W}_k = D^{a_k} (\log n)^{b_k} n^{c_k}, \quad (8.5.20)$$

where

$$c_k = \frac{(\tau-2)^{k-1}}{2(\tau-1)}, \quad b_k = \frac{1 - (\tau-2)^{k-1}}{3-\tau} - \frac{1}{2}(\tau-2)^{k-1}, \quad a_k = \frac{1 - (\tau-2)^{k-1}}{3-\tau}. \quad (8.5.21)$$

Then, the key step in the proof of Proposition 8.21 is the following lemma:

Lemma 8.22 (Connectivity between \mathcal{N}_{k-1} and \mathcal{N}_k) *Fix $m, k \geq 2$. Then the probability that there exists an $i \in \mathcal{N}_k$ that is not at distance two from \mathcal{N}_{k-1} in $\text{PA}_{m, \delta}(2t)$ is $o(n^{-1})$.*

Proof We note that, by [Volume 1, Exercise 8.20], with probability exceeding $1 - o(n^{-1})$, for all k ,

$$\sum_{i \in \mathcal{N}_{k-1}} D_i(t) \geq B t [\mathcal{W}_{k-1}]^{2-\tau}. \quad (8.5.22)$$

On the event that the bounds in (8.5.22) hold, we obtain by (8.5.10) that the conditional probability, given $\text{PA}_{m, \delta}(t)$, that there exists an $i \in \mathcal{N}_k$ such that there is no t -connector between i and \mathcal{N}_{k-1} is bounded, using Boole's inequality, by

$$t e^{-\eta B [\mathcal{W}_{k-1}]^{2-\tau} \mathcal{W}_k} = t e^{-\eta B D \log n} = o(n^{-1}), \quad (8.5.23)$$

where we have used (8.5.19) and we have taken $D > 2(\eta B)^{-1}$. \square

We now complete the proof of Proposition 8.21. Fix

$$k^* = \left\lfloor \frac{\log \log n}{|\log(\tau - 2)|} \right\rfloor. \tag{8.5.24}$$

As a result of Lemma 8.22, we have that the distance between \mathcal{N}_{k^*} and Inner_t is at most $2k^*$. Therefore, we are done when we can show that

$$\text{Outer}_t \subseteq \{i : D_i(t) \geq (\log n)^\sigma\} \subseteq \mathcal{N}_{k^*} = \{i : D_i(t) \geq \mathcal{W}_{k^*}\}, \tag{8.5.25}$$

so that it suffices to prove that $(\log n)^\sigma \geq \mathcal{W}_{k^*}$, for any $\sigma > \frac{1}{3-\tau}$. For this, we note that, by Lemma 7.10, we have that

$$\mathcal{W}_{k^*} = D^{a_{k^*}} (\log n)^{b_{k^*}} n^{c_{k^*}}. \tag{8.5.26}$$

We have that $n^{c_{k^*}} = O(1) = (\log n)^{o(1)}$, $(\log n)^{b_{k^*}} = (\log n)^{\frac{1}{3-\tau} + o(1)}$, and $D^{a_{k^*}} = (\log n)^{o(1)}$. Thus,

$$\mathcal{W}_{k^*} = (\log n)^{\frac{1}{3-\tau} + o(1)}, \tag{8.5.27}$$

so that, by picking t sufficiently large, we can make $\sigma \geq \frac{1}{3-\tau} + o(1)$. This completes the proof of Proposition 8.21. \square

Proof of Theorem 8.19. We note that whp $\text{diam}_{2t}(\text{Core}_t) \leq C_\delta + 2k^*$, where k^* is the upper bound on $\max_{i \in \text{Outer}_t} \min_{j \in \text{Inner}_t} d_{\text{PA}_{m,\delta}(2t)}(i, j)$ in Proposition 8.21, and we have made use of Proposition 8.20. This proves Theorem 8.19. \square

Proof of the upper bound in Theorem 8.8.

TO DO 8.5: Add proof! \square

8.6 DIAMETERS IN PREFERENTIAL ATTACHMENT MODELS

In this section, we investigate the diameter in preferential attachment models.

Theorem 8.23 (A $\log n$ bound for the diameter in PAMs) *Fix $m \geq 1$ and $\delta > 0$. For $\text{PA}_n^{(m,\delta)}$ there exist $0 < b_1 < b_2 < \infty$ such that, as $n \rightarrow \infty$,*

$$\mathbb{P}(b_1 \log n \leq \text{diam}(\text{PA}_n^{(m,\delta)}) \leq b_2 \log n) = 1 - o(1). \tag{8.6.1}$$

Theorem 8.24 (Diameter of $\text{PA}_n^{(m,\delta)}$ for $\delta = 0$) *Fix $m \geq 2$ and $\delta = 0$. For $\text{PA}_n^{(m,\delta)}$, as $n \rightarrow \infty$,*

$$\text{diam}(\text{PA}_n^{(m,\delta)}) \frac{\log \log n}{\log n} \xrightarrow{\mathbb{P}} 1. \tag{8.6.2}$$

Theorem 8.25 (Diameter of $\text{PA}_n^{(m,\delta)}$ for $\delta < 0$) *Fix $m \geq 2$ and $\delta \in (-m, 0)$. For $\text{PA}_n^{(m,\delta)}$, as $n \rightarrow \infty$,*

$$\frac{\text{diam}(\text{PA}_n^{(m,\delta)})}{\log \log n} \xrightarrow{\mathbb{P}} \frac{4}{|\log(\tau - 2)|} + \frac{2}{\log m}. \tag{8.6.3}$$

8.7 Further results on distances in preferential attachment models 309

TO DO 8.6: Extend the discussion in this section.

TO DO 8.7: Explain that Theorem 8.23 is already proved

8.7 FURTHER RESULTS ON DISTANCES IN PREFERENTIAL ATTACHMENT MODELS

TO DO 8.9: Add further result, such as the ones in the BPAM

8.8 NOTES AND DISCUSSION

TO DO 8.10: Extend 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 one of the main results in Pittel (1994).

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.

Notes on Section 8.2.

A weaker version of Theorem 8.8 is proved in Dommers et al. (2010). The current theorem is inspired by Dereich et al. (2012).

Notes on Section 8.3.

Notes on Section 8.4.

Notes on Section 8.5.

Notes on Section 8.6.

Notes on Section 8.7.

8.9 EXERCISES FOR CHAPTER 8

TO DO 8.11: Design more exercises

Exercise 8.1 (Bound on γ) *Prove that the solution γ of (8.1.3) satisfies $\gamma < 1$. What does this imply for the diameter and typical distances in scale-free trees?*

Exercise 8.2 (Bound on γ) *Prove that the solution γ of (8.1.3) satisfies $\gamma \in (0, e^{-1})$.*

Exercise 8.3 (Typical distances in scale-free trees) *Use rove that*

$$\frac{\text{dist}_{\text{PA}_n^{(1,\delta)(b)}}(o_1, o_2)}{\log n} \xrightarrow{\mathbb{P}} \frac{2(1+\delta)}{(2+\delta)\gamma}.$$

Exercise 8.4 (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*

$$\lim_{t \rightarrow \infty} \mathbb{P}(\text{dist}_{\text{PA}_n^{(m, \delta)}}(v_i, v_j) \leq 2) = 1. \quad (8.9.1)$$

Exercise 8.5 (Early vertices are not at distance 2 when $\delta > 0$) *Let $\delta > 0$ and $m \geq 2$. Show that for i, j fixed*

$$\lim_{t \rightarrow \infty} \mathbb{P}(\text{dist}_{\text{PA}_n^{(m, \delta)}}(v_i, v_j) = 2) = 0. \quad (8.9.2)$$

Exercise 8.6 (Negative correlations for $m = 1$) *Show that when $m = 1$, Lemma 8.10 implies that when (π_0, \dots, π_k) contains different coordinates as (ρ_0, \dots, ρ_k) , then*

$$\mathbb{P}\left(\bigcap_{i=0}^{k-1} \{\pi_i \rightarrow \pi_{i+1}\} \cap \bigcap_{i=0}^{k-1} \{\rho_i \rightarrow \rho_{i+1}\}\right) \leq \mathbb{P}\left(\bigcap_{i=0}^{k-1} \{\pi_i \rightarrow \pi_{i+1}\}\right) \mathbb{P}\left(\bigcap_{i=0}^{k-1} \{\rho_i \rightarrow \rho_{i+1}\}\right). \quad (8.9.3)$$

Exercise 8.7 (Extension of (8.3.15) to $\text{PA}_n^{(1, \delta)}(b)$) *Prove that for $\text{PA}_n^{(1, \delta)}(b)$, (8.3.15) is replaced with*

$$\mathbb{P}(g(t_1) = s, g(t_2) = s) = (1 + \delta) \frac{\Gamma(t_1 - \delta/(2 + \delta))\Gamma(t_2 - (1 + \delta)/(2 + \delta))\Gamma(s)}{\Gamma(t_1 + 1/(2 + \delta))\Gamma(t_2)\Gamma(s + 2/(2 + \delta))}. \quad (8.9.4)$$

Part IV

Related models and problems

Summary of Part III.

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 τ , 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 $\tau > 3$.*

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 really small distances in them, 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 197) and B (see above) remain valid, and if not, to which extent they need to be adapted.

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 discussing directed versions of the random graphs studied in this book. In Section 9.2, 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. In Section 9.3, 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 9.4, and with exercises in Section 9.5.

9.1 DIRECTED RANDOM GRAPHS

TO DO 9.1: Explain directed networks from a practical setting (WWW?)

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 links to a paper, or that paper links to mine.

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 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 backwards path. See Exercise 9.1 for a description of the topology of di-graphs 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

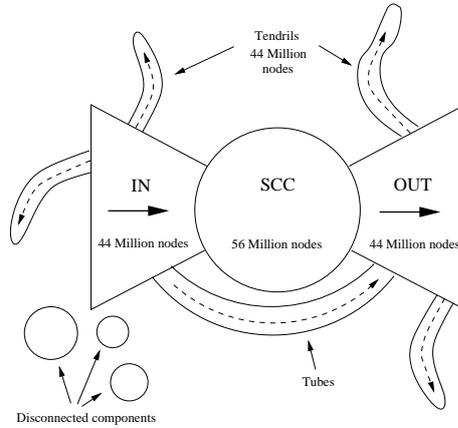


Figure 9.1 The WWW according to Broder et al. (2000).

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 9.1 (which is [Volume 1, Figure 1.19] repeated) for a description of these parts for the WWW, as well as their relative sizes.

In this section, we discuss some models of directed random graphs or random di-graphs. We start in Section 9.1.1 by discussing inhomogeneous random di-graphs, in Section 9.1.2 we continue with the directed configuration model, and in Section 9.1.3, we close with directed preferential attachment models.

9.1.1 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 μ as $n \rightarrow \infty$. That is, we assume that, for each μ -continuous Borel set $\mathcal{A} \subseteq \mathcal{S}$, as $n \rightarrow \infty$,

$$\frac{1}{n} |\{i \in [n] : x_i \in \mathcal{A}\}| \rightarrow \mu(\mathcal{A}). \quad (9.1.1)$$

Here we say that a Borel set \mathcal{A} is μ -continuous whenever its boundary $\partial\mathcal{A}$ has zero probability, i.e., $\mu(\partial\mathcal{A}) = 0$.

Given n , let D_n be the random digraph on the vertex set $(x_i)_{i \in [n]}$ with independent arcs having probabilities

$$\mathbb{P}((x_i, x_j) \in E(D_n)) = 1 \wedge (\kappa(x_i, x_j)/n), \quad i, j \in [n]. \quad (9.1.2)$$

Combining \mathcal{S} , μ , and κ , 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. Note that in the case of random graphs it is necessary to assume, in addition, that the kernel κ is symmetric. In the definition of digraphs D_n for $n \geq 2$, we do not require the symmetry of the kernel.

For large n , the phase transition in the digraph D_n can be described in terms of the survival probabilities of the related multitype Galton-Watson branching processes with type space \mathcal{S} . Let us introduce the necessary mixed-Poisson branching processes now. Given $s \in \mathcal{S}$, let $\mathcal{X}(s)$ and $\mathcal{Y}(s)$ denote the Galton-Watson processes starting at a particle of type $s \in \mathcal{S}$ such that the number of children of types in a subset $\mathcal{A} \subseteq \mathcal{S}$ of a particle of type $t \in \mathcal{S}$ has a Poisson distribution with means

$$\int_{\mathcal{A}} \kappa(t, u) \mu(du), \quad \text{and} \quad \int_{\mathcal{A}} \kappa(u, t) \mu(du), \quad (9.1.3)$$

respectively. These numbers are independent for disjoint subsets \mathcal{A} and for different particles. The critical point of the emergence of the giant SCC is determined by the averaged joint survival probability

$$\zeta = \int_{\mathcal{S}} \zeta_{\mathcal{X}}(s) \zeta_{\mathcal{Y}}(s) \mu(ds) \quad (9.1.4)$$

being positive. Here $\zeta_{\mathcal{X}}(s)$ and $\zeta_{\mathcal{Y}}(s)$ denote the non-extinction probabilities of $\mathcal{X}(s)$ and $\mathcal{Y}(s)$, respectively.

In order to establish Theorem 9.1, we need to impose further conditions on the kernel κ , like those in Chapter 3. Namely, we need to assume that the kernel κ 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 $(\mu \times \mu)(\{(s, t) \in \mathcal{A} \times (\mathcal{S} \setminus \mathcal{A}) : \kappa(s, t) = 0\}) = 0$ implies that either $\mu(\mathcal{A}) = 0$ or $\mu(\mathcal{S} \setminus \mathcal{A}) = 0$. In addition, we assume that κ 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 $n \rightarrow \infty$,

$$\frac{1}{n} |E(\text{DIRG}_n(\kappa))| \xrightarrow{\mathbb{P}} \int_{\mathcal{S} \times \mathcal{S}} \kappa(s, t) \mu(ds) \mu(dt) < \infty. \quad (9.1.5)$$

Note that here we implicitly assume that κ is integrable.

This gives rise to the following theorem:

Theorem 9.1 (Phase transition in $\text{DIRG}_n(\kappa)$) *Suppose that*

(a) *When $\nu > 1$, ζ in (9.1.4) satisfies $\zeta \in (0, 1]$ and*

$$|\mathcal{C}_{\max}|/n \xrightarrow{\mathbb{P}} \zeta, \quad (9.1.6)$$

while $|\mathcal{C}_{(2)}|/n \xrightarrow{\mathbb{P}} 0$ and $|E(\mathcal{C}_{(2)})|/n \xrightarrow{\mathbb{P}} 0$.

(b) *When $\nu < 1$, ζ in (9.1.4) satisfies $\zeta = 0$ and $|\mathcal{C}_{\max}|/n \xrightarrow{\mathbb{P}} 0$ and $|E(\mathcal{C}_{\max})|/n \xrightarrow{\mathbb{P}} 0$.*

TO DO 9.3: Add discussion directed inhomogeneous random graphs

9.1.2 THE DIRECTED CONFIGURATION MODEL

One way to obtain a directed version of $\text{CM}_n(\mathbf{d})$ is to give each edge a direction, chosen with probability $1/2$, 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. In real-world applications, correlations between in- and out-degrees can be positive or negative, depending on the precise application. Therefore, we formulate a general model of directed graphs, where we can prescribe both the in- and out-degrees of vertices.

Fix $\mathbf{d}^{(\text{in})} = (d_i^{(\text{in})})_{i \in [n]}$ to be a sequence of in-degrees, where $d_i^{(\text{in})}$ denotes the in-degree of vertex i . Similarly, we let $\mathbf{d}^{(\text{out})} = (d_i^{(\text{out})})_{i \in [n]}$ be a sequence of out-degrees. Naturally, we need that

$$\sum_{i \in [n]} d_i^{(\text{in})} = \sum_{i \in [n]} d_i^{(\text{out})} \quad (9.1.7)$$

in order for a graph with in- and out-degree sequence $\mathbf{d} = (\mathbf{d}^{(\text{in})}, \mathbf{d}^{(\text{out})})$ to exist. We think of $d_i^{(\text{in})}$ as the number of in-half-edges incident to vertex i and $d_i^{(\text{out})}$ as the number of out-half-edges incident to vertex i . The directed configuration model $\text{DCM}_n(\mathbf{d})$ 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 $d_i^{(\text{in})}$ and out-degree $d_i^{(\text{out})}$. Similarly to $\text{CM}_n(\mathbf{d})$, $\text{DCM}_n(\mathbf{d})$ 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 $(D_n^{(\text{in})}, D_n^{(\text{out})})$ denote the in- and out-degree of a vertex chosen uniformly at random from $[n]$. The following exercise investigates the limiting distribution of the number of self-loops in $\text{DCM}_n(\mathbf{d})$:

We continue to investigate the strongly connected component of $\text{DCM}_n(\mathbf{d})$. Assume, similarly to Condition 1.5(a)-(b), that $(D_n^{(\text{in})}, D_n^{(\text{out})}) \xrightarrow{d} (D^{(\text{in})}, D^{(\text{out})})$, and that $\mathbb{E}[D_n^{(\text{in})}] \rightarrow \mathbb{E}[D^{(\text{in})}]$ and $\mathbb{E}[D_n^{(\text{out})}] \rightarrow \mathbb{E}[D^{(\text{in})}]$. Naturally, by (9.1.7), this implies that $\mathbb{E}[D^{(\text{out})}] = \mathbb{E}[D^{(\text{in})}]$:

Let

$$p_{k,l} = \mathbb{P}(D^{(\text{in})} = k, D^{(\text{out})} = l) \quad (9.1.8)$$

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 $\text{DCM}_n(\mathbf{d})$. The distribution $(p_{k,l})_{k,l \geq 0}$ plays a similar role for $\text{DCM}_n(\mathbf{d})$ as $(p_k)_{k \geq 0}$ does for $\text{CM}_n(\mathbf{d})$. We further define

$$p_k^{*(\text{in})} = \sum_l l p_{k,l} / \mathbb{E}[D^{(\text{out})}], \quad p_l^{*(\text{out})} = \sum_k k p_{k,l} / \mathbb{E}[D^{(\text{in})}]. \quad (9.1.9)$$

The distributions $(g_k^{(\text{in})})_{k \geq 0}$ and $(g_k^{(\text{out})})_{k \geq 0}$ correspond to the asymptotic forward in- and out-degree of a uniformly chosen edge in $\text{CM}_n(\mathbf{d})$. Let $\theta^{(\text{in})}$ and $\theta^{(\text{out})}$ be the survival probabilities of the branching processes with offspring distributions $(p_k^{*(\text{in})})_{k \geq 0}$ and $(p_k^{*(\text{out})})_{k \geq 0}$, respectively, and define

$$\zeta^{(\text{in})} = 1 - \sum_{k,l} p_{k,l} (1 - \theta^{(\text{in})})^l, \quad \zeta^{(\text{out})} = 1 - \sum_{k,l} p_{k,l} (1 - \theta^{(\text{out})})^k. \quad (9.1.10)$$

Then, $\zeta^{(\text{out})}$ has the interpretation of the asymptotic probability that a uniform 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

$$\psi = \sum_{k,l} p_{k,l} (1 - \theta^{(\text{in})})^l (1 - \theta^{(\text{out})})^k \tag{9.1.11}$$

so that ψ 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

$$\zeta = \zeta^{(\text{out})} + \zeta^{(\text{in})} - (1 - \psi) \tag{9.1.12}$$

has the interpretation of the asymptotic probability that a uniform vertex has both a large forward and backward cluster. 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 \mathcal{C}_{\max} denote the size of the largest strongly connected component in $\text{DCM}_n(\mathbf{d})$. Finally, we let

$$\nu = \sum_{k=0}^{\infty} k p_k^{*(\text{in})} = \sum_{k,l} k l p_{k,l} / \mathbb{E}[D^{(\text{out})}], \tag{9.1.13}$$

Alternatively, $\nu = \sum_{k=0}^{\infty} k p_k^{*(\text{out})}$. Then, the main results is as follows:

Theorem 9.2 (Phase transition in $\text{DCM}_n(\mathbf{d})$) *Suppose that \mathbf{d} satisfies that*

$$(D_n^{(\text{in})}, D_n^{(\text{out})}) \xrightarrow{d} (D^{(\text{in})}, D^{(\text{out})}), \quad \mathbb{E}[D_n^{(\text{in})}] \rightarrow \mathbb{E}[D^{(\text{in})}], \quad \mathbb{E}[D_n^{(\text{out})}] \rightarrow \mathbb{E}[D^{(\text{out})}], \tag{9.1.14}$$

and

$$\mathbb{E}[D_n^{(\text{in})} D_n^{(\text{out})}] \rightarrow \mathbb{E}[D^{(\text{in})} D^{(\text{out})}]. \tag{9.1.15}$$

Further, assume that \mathbf{d} is proper, as explained below.

(a) When $\nu > 1$, ζ in (9.1.12) satisfies $\zeta \in (0, 1]$ and

$$|\mathcal{C}_{\max}|/n \xrightarrow{\mathbb{P}} \zeta, \tag{9.1.16}$$

while $|\mathcal{C}_{(2)}|/n \xrightarrow{\mathbb{P}} 0$ and $|E(\mathcal{C}_{(2)})|/n \xrightarrow{\mathbb{P}} 0$.

(b) When $\nu < 1$, ζ in (9.1.12) satisfies $\zeta = 0$ and $|\mathcal{C}_{\max}|/n \xrightarrow{\mathbb{P}} 0$ and $|E(\mathcal{C}_{\max})|/n \xrightarrow{\mathbb{P}} 0$.

Theorem 9.2 is the adaptation to $\text{DCM}_n(\mathbf{d})$ of Theorem 4.4 for $\text{CM}_n(\mathbf{d})$. In the statement of Theorem 9.2, we have assumed that \mathbf{d} is *proper*, which is a technical requirement on the degree sequence stating that (a) $\mathbb{E}[(D_n^{(\text{in})})^2] = O(1)$, $\mathbb{E}[(D_n^{(\text{out})})^2] = O(1)$; (b) $\mathbb{E}[D_n^{(\text{in})}(D_n^{(\text{out})})^2] = o(n^{1/12} \log n)$. In view of the fact that

such conditions do not appear in Theorem 4.4, these conditions are likely to be suboptimal for Theorem 9.2 to hold.

TO DO 9.3: Add discussion distances in directed random graphs

9.1.3 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 Theorem 8.3. We first describe the model. Let G_0 be any fixed initial directed graph with t_0 edges, where t_0 is some arbitrary positive integer.

We next define $G(t)$. Fix some non-negative parameters $\alpha, \beta, \gamma, \delta_{\text{in}}$ and δ_{out} , where $\alpha + \beta + \gamma = 1$. We say that we choose a vertex *according to* $f_i(t)$ when we choose vertex i with probability

$$\frac{f_i(t)}{\sum_j f_j(t)}. \quad (9.1.17)$$

Thus, the probability that we choose a vertex i is proportional to the value of the function $f_i(t)$. Also, we denote the in-degree of vertex i in $G(t)$ by $D_i^{(\text{in})}(t)$, and the out-degree of vertex i in $G(t)$ by $D_i^{(\text{out})}(t)$.

We let $G(t_0) = G_0$, where t_0 is chosen appropriately, as we will indicate below. For $t \geq t_0$, we form $G(t+1)$ from $G(t)$ according to the following growth rules:

- (A) With probability α , we add a new vertex v together with an edge from v to an existing vertex which is chosen according to $D_i^{(\text{in})}(t) + \delta_{\text{in}}$.
- (B) With probability β , we add an edge between the existing vertices v and w , where v and w are chosen independently, v according to $D_i^{(\text{in})}(t) + \delta_{\text{in}}$ and w according to $D_i^{(\text{out})}(t) + \delta_{\text{out}}$.
- (C) With probability γ , we add a vertex w and an edge from an existing vertex v to w according to $D_i^{(\text{out})}(t) + \delta_{\text{out}}$.

The above growth rule produces a graph process $(G(t))_{t \geq t_0}$ where $G(t)$ has precisely t edges. The number of vertices in $G(t)$ is denoted by $T(t)$, where $T(t) \sim \text{Bin}(t, \alpha + \gamma)$.

It is not hard to see that if $\alpha\delta_{\text{in}} + \gamma = 0$, then all vertices outside of G_0 have in-degree zero, while if $\gamma = 1$ all vertices outside of G_0 have in-degree one. Similar trivial graph processes arise when $\gamma\delta_{\text{out}} + \alpha = 0$ or $\alpha = 1$.

We exclude the above cases. Then, Bollobás et al. (2003) show that both the in-degree and the out-degree of the graph converge, in the sense that we will explain now. Denote by $(X_k(t))_{k \geq 0}$ the in-degree sequence of $G(t)$, so that

$$X_k(t) = \sum_{v \in G(t)} \mathbb{1}_{\{D_v^{(\text{in})} = k\}}, \quad (9.1.18)$$

and, similarly, let $(Y_k(t))_{k \geq 0}$ be the out-degree sequence of $G(t)$, so that

$$Y_k(t) = \sum_{v \in G(t)} \mathbb{1}_{\{D_v^{(\text{out})} = k\}}. \quad (9.1.19)$$

Denote

$$\tau_{\text{in}} = 1 + \frac{1 + \delta_{\text{in}}(\alpha + \beta)}{\alpha + \beta}, \quad \tau_{\text{out}} = 1 + \frac{1 + \delta_{\text{out}}(\gamma + \beta)}{\gamma + \beta}. \quad (9.1.20)$$

Then (Bollobás et al., 2003, Theorem 3.1) shows that there exist probability distributions $p = (p_k)_{k \geq 0}$ and $q = (q_k)_{k \geq 0}$ such that

$$X_k(t) - p_k t = o_{\mathbb{P}}(t), \quad Y_k(t) - q_k t = o_{\mathbb{P}}(t), \quad (9.1.21)$$

where, for $k \rightarrow \infty$,

$$p_k = C_{\text{in}} k^{-\tau_{\text{in}}} (1 + o(1)), \quad q_k = C_{\text{out}} k^{-\tau_{\text{out}}} (1 + o(1)). \quad (9.1.22)$$

In fact, the probability distributions p and q are determined explicitly, as in (8.4.2) above, and p and q have a similar shape as p in (8.4.2). Also, since $\delta_{\text{in}}, \delta_{\text{out}} \geq 0$, and $\alpha + \beta, \gamma + \beta \leq 1$, we again have that $\tau_{\text{in}}, \tau_{\text{out}} \in (2, \infty)$.

9.2 RANDOM GRAPHS WITH COMMUNITY STRUCTURE

TO DO 9.4: Add practical example of community structure: clustering table?

TO DO 9.5: Add discussion community structure: partition, SBM, larger overlap

9.2.1 STOCHASTIC BLOCK MODEL

TO DO 9.6: Add discussion SBM

9.2.2 INHOMOGENEOUS RANDOM GRAPHS WITH COMMUNITIES

TO DO 9.7: Make this introduction section shorter

General inhomogeneous random graphs have rather low clustering. Indeed, assuming that $\kappa_n(x, y \leq n$, we can compute that the expected number of triangles in an $\text{IRG}_n(\kappa)$ is equal to

$$\mathbb{E}[\# \text{ triangles in } \text{IRG}_n(\kappa)] = \frac{1}{n^3} \sum_{i,j,k \in [n]} \kappa_n(x_i, x_j) \kappa_n(x_j, x_k) \kappa_n(x_k, x_i). \quad (9.2.1)$$

Under relatively weak conditions on the kernel κ_n , it follows that

$$\mathbb{E}[\# \text{ triangles in } \text{IRG}_n(\kappa)] \rightarrow \int_{S^3} \kappa(x_1, x_2) \kappa(x_2, x_3) \kappa(x_3, x_1) \mu(dx_1) \mu(dx_2) \mu(dx_3). \quad (9.2.2)$$

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 a model similar to the inhomogeneous random graph $\text{IRG}_n(\kappa)$ 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 λ/n , as for $\text{ER}_n(\lambda/n)$, but also each collection of triples forms a triangle with probability μ/n^2 , independently for all triplets and independently of the status of the edges. Here the exponent n^{-2} is chosen to as to make the expected number of triangles containing a vertex bounded.

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. In order to formulate this general version of the model, we start by introducing some notation.

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 $V(F) = [r] = \{1, 2, \dots, r\}$. 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 κ_F be a measurable function from \mathcal{S}^r to $[0, \infty)$. The function κ_F is called the *kernel corresponding to F* . A sequence $\tilde{\kappa} = (\kappa_F)_{F \in \mathcal{F}}$ is a *kernel family*.

Let $\tilde{\kappa}$ be a kernel family and n an integer. We define a random graph $\text{IRG}_n(\tilde{\kappa})$ with vertex set $[n] = \{1, \dots, n\}$. First let $x_1, x_2, \dots, x_n \in \mathcal{S}$ be i.i.d. (independent and identically distributed) with the distribution μ . Given $\mathbf{x} = (x_1, \dots, x_n)$, construct $\text{IRG}_n(\tilde{\kappa})$ as follows, starting with the empty graph. For each r and each $F \in \mathcal{F}$ with $|F| = r$, and for every r -tuple of distinct vertices $(v_1, \dots, v_r) \in [n]^r$, add a copy of F on the vertices v_1, \dots, v_r (with vertex i of F mapped to v_i) with probability

$$p = p(v_1, \dots, v_r; F) = \frac{\kappa_F(x_{v_1}, \dots, x_{v_r})}{n^{r-1}}, \quad (9.2.3)$$

all these choices being independent. Here, if $p > 1$, by convention we simply add a copy with probability 1. We shall often call the added copies of the various F that together form $\text{IRG}_n(\tilde{\kappa})$ *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 (9.2.3) is that we wish to consider sparse graphs; indeed, our main interest is the case when $\text{IRG}_n(\tilde{\kappa})$ has $O(n)$ edges. As it turns out, we can be slightly more general; however, when κ_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 μ^r on \mathcal{S}^r .

There are several plausible choices for the normalization in (9.2.3). 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 (9.2.3) 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 $ER_n(\lambda/n)$; the normalization we use here differs from this by a factor of 2.

In the special case where all κ_F are zero apart from κ_{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 κ_2 for κ_{K_2} . Note that in the edge-only case, given \mathbf{x} , two vertices i and j are joined with probability

$$\frac{\kappa_2(x_i, x_j) + \kappa_2(x_j, x_i)}{n} + O\left(\frac{(\kappa_2(x_i, x_j) + \kappa_2(x_j, x_i))^2}{n^2}\right). \tag{9.2.4}$$

The correction term will never matter, so we may as well replace κ_2 by its symmetrized version. In fact, we shall always assume that κ_F is invariant under permutations of the vertices of the graph F .

For any kernel family $\tilde{\kappa}$, let κ_e be the corresponding *edge kernel*, defined by

$$\kappa_e(x, y) = \sum_F \sum_{ij \in E(F)} \int_{S^{V(F) \setminus \{i, j\}}} \kappa_F(x_1, \dots, x_{i-1}, x, x_{i+1}, \dots, x_{j-1}, y, x_{j+1}, \dots, x_{|F|}), \tag{9.2.5}$$

where the second sum runs over all $2E(F)$ ordered pairs (i, j) with $ij \in Edges(F)$, 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 $G(n, \tilde{\kappa})$ is at most $\kappa_e(x_i, x_j)/n + O(1/n^2)$. In other words, κ_e captures the edge probabilities in $G(n, \tilde{\kappa})$, but not the correlations.

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

$$\xi(\tilde{\kappa}) = \sum_{F \in \mathcal{F}} E(F) \int_{S^{|F|}} \kappa_F = \frac{1}{2} \int_{S^2} \kappa_e \leq \infty \tag{9.2.6}$$

be the *asymptotic edge density* of $\tilde{\kappa}$. Since every copy of F contributes $E(F)$ edges, the following theorem is almost obvious, provided we can ignore overlapping edges.

Theorem 9.3 (The edge density in $IRG_n(\tilde{\kappa})$) *As $n \rightarrow \infty$, $\mathbb{E}[E(IRG_n(\tilde{\kappa}))/n] \rightarrow \xi(\tilde{\kappa}) \leq \infty$. Moreover, $E(IRG_n(\tilde{\kappa}))/n$ converges in probability to the asymptotic edge density $\xi(\tilde{\kappa})$. In other words, if $\xi(\tilde{\kappa}) < \infty$ then $E(IRG_n(\tilde{\kappa})) = \xi(\tilde{\kappa})n + o_p(n)$, and if $\xi(\tilde{\kappa}) = \infty$ then $E(IRG_n(\tilde{\kappa})) > Cn$ for every constant C , whp.*

The main focus is the emergence of the giant component. By the *component structure* of a graph G , we mean the set of vertex sets of its components, i.e., the structure encoding only which vertices are in the same component, not the internal structure of the components themselves. When studying the component structure of $IRG_n(\tilde{\kappa})$, 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 $\tilde{\kappa}$ is a *clique kernel family* if the only non-zero kernels are those corresponding to complete graphs; the corresponding random graph model $\text{IRG}_n(\tilde{\kappa})$ is a *clique model*. For questions corresponding component structure, it suffices to study clique models. For clique kernels we write κ_r for κ_{K_r} ; as above, we always assume that κ_r is symmetric, here meaning invariant under all permutations of the coordinates of \mathcal{S}^r . Given a general kernel family $\tilde{\kappa}$, the corresponding (symmetrized) clique kernel family is given by $\tilde{\kappa} = (\kappa_r)_{r \geq 2}$ with

$$\kappa_r(x_1, \dots, x_r) = \sum_{F \in \mathcal{F}: |F|=r} \frac{1}{r!} \sum_{\pi \in \mathfrak{S}_r} \kappa_F(x_{\pi(1)}, \dots, x_{\pi(r)}), \tag{9.2.7}$$

where \mathfrak{S}_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 $\text{IRG}_n(\tilde{\kappa})$, we may always replace $\tilde{\kappa}$ by the corresponding clique kernel family.

It is often convenient to think of a clique model as a random *hypergraph*, with the cliques as the hyperedges; for this reason we call a clique kernel family a *hyperkernel*. Note that each unordered set of r vertices corresponds to $r!$ r -tuples, so the probability that we add a K_r on a given set of r vertices is $r! \kappa_r(x_{v_1}, \dots, x_{v_r}) / n^{r-1}$. (More precisely, this is the expected number of K_r s added with this vertex set.)

In our analysis we also consider the linear operator T_{κ_e} defined by

$$T_{\kappa_e}(f)(x) = \int_{\mathcal{S}} \kappa_e(x, y) f(y) d\mu(y), \tag{9.2.8}$$

where κ_e is defined by (9.2.5). We need to impose some sort of integrability condition on our kernel family:

Definition 9.4 (i) A kernel family $\tilde{\kappa} = (\kappa_F)_{F \in \mathcal{F}}$ is integrable if

$$\int \tilde{\kappa} = \sum_{F \in \mathcal{F}} |F| \int_{\mathcal{S}^{|F|}} \kappa_F < \infty. \tag{9.2.9}$$

This means that the expected number of atoms containing a given vertex is bounded.

(ii) A kernel family $\tilde{\kappa} = (\kappa_F)_{F \in \mathcal{F}}$ is edge integrable if

$$\sum_{F \in \mathcal{F}} E(F) \int_{\mathcal{S}^{|F|}} \kappa_F < \infty; \tag{9.2.10}$$

equivalently, $\xi(\kappa) < \infty$ or $\int_{\mathcal{S}^2} \kappa_e < \infty$. This means that the expected number of edges in $G(n, \tilde{\kappa})$ is $O(n)$, see Theorem 9.3, and thus the expected degree of a given vertex is bounded.

Note that a hyperkernel (κ_r) is integrable if and only if $\sum_{r \geq 2} r \int_{\mathcal{S}^r} \kappa_r < \infty$, and edge integrable

The main results concerning the phase transition on $\text{IRG}_n(\tilde{\kappa})$ is that if $\tilde{\kappa}$ is an

integrable kernel family satisfying a certain extra assumption, then the normalized size of the giant component in $\text{IRG}_n(\tilde{\kappa})$ is simply $\zeta(\tilde{\kappa}) + o_p(1)$. The extra assumption is an irreducibility assumption similar to Definition 3.3(ii) that essentially guarantees that the graph does not split into two pieces: we say that a symmetric kernel $\kappa_e : \mathcal{S}^2 \rightarrow [0, \infty)$ is *reducible* if

$$\exists A \subset \mathcal{S} \text{ with } 0 < \mu(A) < 1 \text{ such that } \kappa_e = 0 \text{ a.e. on } A \times (\mathcal{S} \setminus A);$$

otherwise κ_e is irreducible. Thus, κ_e is irreducible if

$$A \subseteq \mathcal{S} \text{ and } \kappa_e = 0 \text{ a.e. on } A \times (\mathcal{S} \setminus A) \text{ implies } \mu(A) = 0 \text{ or } \mu(\mathcal{S} \setminus A) = 0.$$

We are now ready to formulate the main result in this section involving the phase transition in $\text{IRG}_n(\tilde{\kappa})$. Recall that $|\mathcal{C}_{\max}|$ denotes the number of vertices in the largest connected component of the graph under consideration, and $|\mathcal{C}_{(2)}|$ the number of vertices in its second largest component.

Theorem 9.5 (The phase transition on 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 hyperkernel, given by (9.2.7). Then, there exists a $\zeta(\tilde{\kappa}) \in [0, 1)$ such that*

$$|\mathcal{C}_{\max}| = \zeta(\tilde{\kappa})n + o_p(n), \tag{9.2.11}$$

and $|\mathcal{C}_{(2)}| = o_p(n)$.

Theorem 9.5 is proved by showing that (in the hyperkernel case) the branching process that captures the ‘local structure’ of $\text{IRG}_n(\tilde{\kappa})$. For Theorem 9.5 to be useful we would like to know something about $\zeta(\tilde{\kappa})$, which can be calculated from $\zeta_{\tilde{\kappa}}$, which is in turn the largest solution to the functional equation:

$$f = 1 - e^{-S_{\tilde{\kappa}}(f)(x)}. \tag{9.2.12}$$

The question when $\zeta(\tilde{\kappa}) > 0$ is settled in the following theorem:

Theorem 9.6 (Condition for existence giant component) *Let $\tilde{\kappa}$ be an integrable hyperkernel. Then, $\zeta(\tilde{\kappa}) > 0$ if and only if $\|T_{\kappa_e}\| > 1$. Furthermore, if $\tilde{\kappa}$ is irreducible and $\|T_{\kappa_e}\| > 1$, then $\zeta_{\tilde{\kappa}}(x)$ is the unique non-zero solution to the functional equation (9.2.12), and $\zeta_{\tilde{\kappa}}(x) > 0$ holds for a.e. x .*

In general, $\|T_{\kappa_e}\|$ may be rather hard to calculate. When we suppose that each κ_r is constant, however, this can be done. Indeed, say that $\kappa_r = c_r$. Then $\kappa_e(x, y) = \sum_r r(r-1)c_r = 2\xi(\kappa)$ for all x and y , so

$$\|T_{\kappa_e}\| = 2\xi(\kappa). \tag{9.2.13}$$

This is perhaps surprising: it tells us that for such uniform hyperkernels, 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.

9.2.3 CONFIGURATION MODELS WITH COMMUNITY STRUCTURE

Hierarchical configuration model

TO DO 9.8: Present results HCM: connectivity and graph distances

The configuration model has low clustering, which often makes it inappropriate in applied contexts. Indeed, many real-world networks, in particular social networks, have a high amount of clustering instead. 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 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 $G_i = (V_i, E_i)$ and the inter-community degrees $\mathbf{d}^{(b)} = (d_u^{(b)})_{u \in V_i}$ satisfying that $\sum_{u \in V_i} d_u^{(b)} = d_i$. Naturally, the size of the graph becomes $N = \sum_{i \in [n]} |V_i|$.

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 7.8(a)-(b), while the empirical distribution of the graphs

$$\mu_n(H, \vec{\mathbf{d}}) = \frac{1}{n} \sum_{i \in [n]} \mathbb{1}_{\{G_i = H, (d_u^{(b)})_{u \in V_i} = \vec{\mathbf{d}}\}} \quad (9.2.14)$$

converges as $n \rightarrow \infty$ to some probability distribution on graphs with integer marks associated to the vertices. See Stegehuis et al. (2016a,b) for power-law relations in hierarchical configuration models and epidemic spread in them, respectively, and Hofstad et al. (Preprint 2015) for its topological properties, such as its connectivity, its clustering, etc.

The configuration model with household structure

TO DO 9.9: Adapt this discussion: special case of HCM

The configuration model has low clustering, which often makes it inappropriate in applied contexts. Indeed, in Chapter 1, we have seen that many real-world networks have a high amount of clustering instead. For example, in modeling a social network, one can expect a large amount of clustering. 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 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 uniform vertex in 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 , whereas its global structure is described by the configuration model $\text{CM}_n(\mathbf{d})$.

The number of vertices in the configuration model with household structure is given by

$$N = \sum_{i=1}^n n_i, \tag{9.2.15}$$

where n_i denotes the number of vertices in G_i . Further, assume that the empirical distribution of the graph sizes converges, i.e., there exists a distribution function F_H such that, for every $x \in \mathbb{R}$,

$$F_{H,n}(x) = \frac{1}{n} \sum_{i \in [n]} \mathbb{1}_{\{n_i \leq x\}} \rightarrow F_H(x), \tag{9.2.16}$$

and $N/n \rightarrow \mathbb{E}[H]$, where H has distribution function F_H . Assume further that the local graphs G_i are all *connected*. Then, assuming (9.2.16) and $N/n \rightarrow \mathbb{E}[H]$, the size of the largest connected component $|\mathcal{C}_{\max}|$ of the configuration model with household structure satisfies

$$|\mathcal{C}_{\max}|/N \xrightarrow{\mathbb{P}} \zeta, \tag{9.2.17}$$

where ζ is the asymptotic proportion of vertices in the giant component of $\text{CM}_n(\mathbf{d})$.

It would be of interest to investigate typical distances within the configuration model with household structure. It seems reasonable to predict that $H_N/\log N$ converges to a limit when [Volume 1, Condition 7.8(a)-(c)] holds, where the limit equals $\alpha/\log \nu$, where α denotes the typical distance in a household drawn uniformly at random. Similar results should hold for infinite variance degrees, where it can be expected that $H_N/\log \log N \xrightarrow{\mathbb{P}} 2\alpha/|\log(\tau - 2)|$.

Due to the fact that a vertex in the configuration model is a household, which in itself is a small subgraph, it seems plausible that the fluctuations of the typical distances *within* the different households give rise to larger fluctuations of the typical distances for the configuration model with household structure than for $\text{CM}_n(\mathbf{d})$ as presented in Theorem 7.20–7.21. In particular, one might expect that $(H_N - \alpha \log_\nu N)/\sqrt{\log N}$ converges to a normal random variable when [Volume 1, Condition 7.8(a)-(c)] holds, while $(H_N - 2\alpha \log \log N/|\log(\tau - 2)|)/\sqrt{\log \log N}$ converges to a normal random variable when a condition as in (7.1.2) holds with $\delta = 0$.

9.2.4 CONFIGURATION MODEL WITH CLUSTERING

The low clustering of $\text{CM}_n(\mathbf{d})$ can be resolved by introducing households as described above. Alternatively, and in the spirit of clustered inhomogeneous random graphs as described in Section 9.2.2, we can also introduce clustering directly. In the configuration model with clustering, we assign two numbers to a vertex $i \in [n]$. We let $d_i^{(\text{si})}$ denote the number of simple half-edges incident to vertex i , and we let $d_i^{(\text{tr})}$ denote the number of triangles that vertex i is part of. We say that there are $d_i^{(\text{si})}$ half-edges incident to vertex i , and $d_i^{(\text{tr})}$ 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 $\text{CM}_n(\mathbf{d})$); 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 $(D_n^{(\text{si})}, D_n^{(\text{tr})})$ denote the number of simple edges and triangles incident to a uniform vertex in $[n]$, and assume that $(D_n^{(\text{si})}, D_n^{(\text{tr})}) \xrightarrow{d} (D^{(\text{si})}, D^{(\text{tr})})$ for some limiting distribution $(D^{(\text{si})}, D^{(\text{tr})})$. Newman (2009) performs a generating function analysis of when a giant component is expected to exist. The criterion Newman finds is that

$$\left(\frac{\mathbb{E}[(D^{(\text{si})})^2]}{\mathbb{E}[D^{(\text{si})}]} - 2 \right) \left(\frac{2\mathbb{E}[(D^{(\text{tr})})^2]}{\mathbb{E}[D^{(\text{tr})}]} - 3 \right) < \frac{2\mathbb{E}[D^{(\text{si})}D^{(\text{tr})}]}{\mathbb{E}[D^{(\text{si})}]\mathbb{E}[D^{(\text{tr})}]}.$$
 (9.2.18)

When $D^{(\text{tr})} = 0$ a.s., so that there are no triangles, this reduces to

$$\frac{\mathbb{E}[(D^{(\text{si})})^2]}{\mathbb{E}[D^{(\text{si})}]} - 2 > 0,$$
 (9.2.19)

which is equivalent to $\nu = \mathbb{E}[D^{(\text{si})}(D^{(\text{si})} - 1)]/\mathbb{E}[D^{(\text{si})}] > 1$. It would be of interest to analyze this model mathematically.

9.2.5 RANDOM INTERSECTION GRAPHS

TO DO 9.10: Extend discussion of intersection graphs

Random intersection graphs with independent group-memberships

The most studied model is when there are n vertices, $m = m(n)$ groups and each vertex is independently connected to each group with probability $p(n)$ (often $m = n^\alpha$ and $p = cn^{-(1+\alpha)/2}$ which gives different behavior depending on the value of α). This model was introduced by Singer (1996) and further studied in Fill et al. (2000); Karoński et al. (1999); Stark (2004).

Deijfen and Kets (2009) studied a version where the vertices have random weights and the edge probabilities are determined by these weights (which gives a graph where both the degree distribution and the clustering can be controlled). In the above model the number of groups that a vertex belongs to is $\text{Bin}(m, p)$ -distributed. The model has also been investigated for more general distributions of groups per vertex by Godehardt and Jaworski (2003) and Jaworski et al. (2006).

Mindaugas Bloznelis has studied this more general model in quite a few papers, including distances and component sizes Bloznelis (n.d., 2010a,b).

Random intersection graphs with prescribed degrees and groups

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 $d_v^{(\text{ve})}$ groups, while group $g \in [m]$ has size $d_g^{(\text{gr})}$. 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

$$\sum_{v \in [n]} d_v^{(\text{ve})} = \sum_{g \in [m]} d_g^{(\text{gr})}. \quad (9.2.20)$$

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. This model might be appropriate as a simple model for collaboration graphs such as the IMDb and the collaboration graph among mathematicians. In the above setting, this model has not received much attention in the mathematical community.

Random intersection graphs with communities

9.2.6 EXPONENTIAL RANDOM GRAPHS AND MAXIMAL ENTROPY

TO DO 9.11: Add discussion of exponential random graphs, and show that GRG is special case

9.3 SPATIAL RANDOM GRAPHS

TO DO 9.12: Add discussion why geometry may be useful in random graphs

9.3.1 SMALL-WORLD MODEL

The models described so far do not incorporate geometry at all. An alternative approach of explaining the small-world phenomenon is to start with a finite torus, and to add random long range connections to them, independently for each pair of vertices. This gives rise to a graph which is a small perturbation of the original lattice, but has occasional long-range connections that are crucial in order to shrink 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 acquaintances in the population living far apart. The main idea is that, even though the shortcuts only form a tiny part of the connections in the graph, they are crucial in order to make it a small world.

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.

The simplest version of the model studied in [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* in [Barbour and Reinert \(2001\)](#). 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 puncturing of the circle, and the distance is then the usual distance along the punctured graph. Denote by D_n the distance between two uniformly chosen points along the punctured circle. Then, ([Barbour and Reinert, 2001](#), Theorem 3.9) states that as $n \rightarrow \infty$, $D_n(2\rho)/\log \rho n$ converges in probability to 1 when $n\rho \rightarrow \infty$, and that $\rho(D_n - \log \rho n/2)$ converges in distribution to a random variable T satisfying

$$\mathbb{P}(T > t) = \int_0^\infty \frac{e^{-y} dy}{1 + e^{2t}}. \quad (9.3.1)$$

The random variable T can also be described by

$$\mathbb{P}(T > t) = \mathbb{E}[e^{-e^{2t}W^{(1)}W^{(2)}}], \quad (9.3.2)$$

where $W^{(1)}, W^{(2)}$ are two independent exponential random variables with parameter 1. Alternatively, it can be seen 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 ([Barbour and Reinert, 2001](#), Theorem 3.9) 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. In ([Barbour and Reinert, 2001](#), Theorem 4.2), also an extension to higher dimensions is given.

The proof 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. For this, it was necessary that the average number of shortcuts per vertex $\rho \downarrow 0$, a restriction that does not appear in [Barbour and Reinert \(2001\)](#). It would be of interest to extend the results to the case of *fixed* ρ as well in the discrete setting.

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 ???. 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.

9.3.2 GEOMETRIC INHOMOGENEOUS RANDOM GRAPHS

TO DO 9.13: Introduce the GIRG

9.3.3 SCALE-FREE PERCOLATION

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 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

$$p_{xy} = 1 - e^{-\lambda W_x W_y / |x-y|^\alpha}, \quad (9.3.3)$$

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 λ . 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 9.5).

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

$$1 - F_w(w) = \mathbb{P}(W > w) = w^{-(\tau-1)}L(w), \quad (9.3.4)$$

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 $x > 0$,

$$\lim_{t \rightarrow \infty} \frac{L(tx)}{L(x)} = 1. \quad (9.3.5)$$

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 α, λ (and the dimension d). Of course, there may be many vertex-weight distributions having the asymptotics in (9.3.4) with the same τ , but the role of τ 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 9.7 (Power-law degrees for power-law weights [Deijfen et al. \(2013\)](#))
Fix $d \geq 1$.

- (a) Assume that the weight distribution satisfies (??) with $\alpha \leq d$ or $\gamma = \alpha(\tau - 1)/d \leq 1$. Then $\mathbb{P}(D_0 = \infty \mid W_0 > 0) = 1$.
- (b) Assume that the weight distribution satisfies (??) 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

$$\mathbb{P}(D_0 > s) = s^{-\gamma}\ell(s). \quad (9.3.6)$$

The fact that the degrees have a power-law distribution is why this model is called **scale-free percolation**. The parameter γ 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 9.4). We continue by studying the percolative properties of scale-free percolation. As before, we denote by λ_c the infimum of all $\lambda \geq 0$ with the property $\mathbb{P}(|\mathcal{C}(0)| = \infty) > 0$. 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 9.8 (Positivity of the critical value [Deijfen et al. \(2013\)](#)) *Assume that the weight distribution satisfies (??) 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. 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. Again we see that graph distances are rather small if $\gamma \in (1, 2)$, whereas graph distances are much larger for $\gamma > 2$.

There is some follow-up work on scale-free percolation. [Hirsch \(Preprint \(2014\)\)](#) 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. \(\(2015\),6\)](#) 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.

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

$$p_k = \mathbb{P}(D_o = k), \quad (9.3.7)$$

where D_x 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 $k \geq 0$,

$$P_k^{(n)} = \frac{1}{n} \sum_{x \in B_r} \mathbb{1}_{\{D_x = k\}}, \quad (9.3.8)$$

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

$$|x - y| \leq \delta \min\{U_x^{-q}, U_y^{-q}\}. \quad (9.3.9)$$

We can think of the ball of radius δU_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 (9.3.9). The parameter δ 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 \mathbb{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

$$\lim_{k \rightarrow \infty} k^{\tau-1} \mathbb{P}(D_o \geq k) \quad (9.3.10)$$

exists, so that the model is scale free with degree power-law exponent τ (recall (1.4.3)). The intuitive explanation of (9.3.10) 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

$$D_o = \sum_{x \in \mathbb{Z}^d} \mathbb{1}_{\{|x| \leq \min\{U_o^{-q}, U_x^{-q}\}\}} = \sum_{x: |x| \leq u^{-q}} \mathbb{1}_{\{|x| \leq U_x^{-q}\}}. \quad (9.3.11)$$

Note that the random variables $\{\mathbb{1}_{\{|x| \leq U_x^{-q}\}}\}_{x \in \mathbb{Z}^d}$ are independent Bernoulli random variables with probability of success equal to

$$\mathbb{P}(\mathbb{1}_{\{|x| \leq U_x^{-q}\}} = 1) = \mathbb{P}(U \leq |x|^{-1/q}) = |x|^{-1/q}. \tag{9.3.12}$$

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

$$\mathbb{E}[D_o \mid U_o = u] = \sum_{x: |x| \leq u^{-q}} |x|^{-1/q} = cu^{-(qd-1)}(1 + o(1)), \tag{9.3.13}$$

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

$$\mathbb{E}[D_o \mid U_o = u_k] = cu_k^{-(qd-1)}(1 + o(1)) = k, \tag{9.3.14}$$

so that $u_k = (k/c)^{-1/(qd-1)}$. This suggests that

$$\mathbb{P}(D_o \geq k) = \mathbb{P}(U \leq u_k)(1 + o(1)) = (ck)^{-1/(qd-1)}(1 + o(1)), \tag{9.3.15}$$

which explains (9.3.10).

We next turn to distances in this scale-free percolation model. For $x, y \in \mathbb{Z}^d$, we denote by $d_{\mathbb{G}_q}(x, y)$ the graph distance (or chemical distance) between the vertices x and y , i.e., the minimal number of edges in \mathbb{G}_q connecting x and y . The main result in Yukich (2006) is the following theorem:

Theorem 9.9 (Ultra-small distances for scale-free percolation) *For all $d \geq 1$ and all $q \in (1/d, \infty)$, whp as $|x| \rightarrow \infty$,*

$$d_{\mathbb{G}_q}(o, x) \leq 8 + 4 \log \log |x|. \tag{9.3.16}$$

The result in Theorem 9.9 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 $d_{\mathbb{G}_q}(o, x)/\log \log |x|$ exists, and, if so, what this limit is.

While Theorem 9.9 resembles the results in Theorem 7.21, there are a few essential differences. First of all, \mathbb{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 9.9 to the setting on finite tori, where the Euclidean norm $|x - y|$ in (9.3.9) 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 9.9. 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 9.9. 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 9.9 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.

9.3.4 HYPERBOLIC RANDOM GRAPHS

Here we consider the hyperbolic random graph where nodes 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. Let us now describe the random graph model in more detail. 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

$$\rho(r) = \beta \frac{\sinh(\beta r)}{\cosh(\beta R) - 1}, \quad (9.3.17)$$

with $\beta = (\tau - 1)/2$. Here ν parametrizes the average degree of the generated networks and $-\beta$ the negative curvature of the space. The angle 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 between two points at polar coordinates (r, ϕ) and (r', ϕ') is given by the hyperbolic law of cosines

$$\cosh(\beta x) = \cosh(\beta r) \cosh(\beta r') - \sinh(\beta r) \sinh(\beta r') \cos(\Delta(\phi, \phi')), \quad (9.3.18)$$

where $\Delta(\phi, \phi') = \pi - |\pi - |\phi - \phi'|||$ is the angle between the two angles.

For a point i with radial coordinate r_i , we define its type t_i as

$$t_i = e^{(R-r_i)/2}. \quad (9.3.19)$$

Then, the degree of vertex i can be approximated by a Poisson random variable with mean t_i [Bode et al. \(2015\)](#). Thus, with high probability

$$t_i = \Theta(D_i), \quad (9.3.20)$$

where D_i denotes the degree of vertex i . Furthermore, t_i are distributed as a power-law with exponent τ [Bode et al. \(2015\)](#), so that the degrees have a power-law distribution as well. The t_i 's can be interpreted as the weights in a rank-1 inhomogeneous random graph [Bode et al. \(2015\)](#). Let us now explain the degree structure in more detail.

We now study $c(k)$ for the hyperbolic random graph. We need to compute the probability that a triangle is formed between a vertex of degree k , a vertex i with $t_i \propto n^{\alpha_1}$ and a vertex j with $t_j \propto n^{\alpha_2}$ with $\alpha_1 \leq \alpha_2$. We can write this probability

as

$$\mathbb{P}(\Delta \text{ on types } k, n^{\alpha_1}, n^{\alpha_2}) = \mathbb{P}(k \leftrightarrow n^{\alpha_1})\mathbb{P}(k \leftrightarrow n^{\alpha_2})\mathbb{P}(n^{\alpha_1} \text{ and } n^{\alpha_2} \text{ neighbor connect}). \quad (9.3.21)$$

Two vertices with types t_i and t_j connect with probability [Bode et al. \(2015\)](#)

$$\mathbb{P}(i \leftrightarrow j \mid t_i, t_j) = \min\left(\frac{2\nu t_i t_j}{\pi n}, 1\right) (1 + o_{\mathbb{P}}(1)). \quad (9.3.22)$$

Therefore, the probability that a vertex of degree k connects with a randomly chosen vertex of type n^{α_1} can be approximated by

$$\mathbb{P}(k \leftrightarrow n^{\alpha_1}) \propto \min(kn^{\alpha_1-1}, 1) \quad (9.3.23)$$

with high probability. The third term in (9.3.21) equals the probability that the two neighbors of a vertex of degree k connect to one another, which is more involved. Two neighbors of a vertex are likely to be close to one another, which makes the probability that they connect larger. We now compute the order of magnitude of the third term in (9.3.21). Let i and j be neighbors of a vertex with degree k , with $t_i \propto n^{\alpha_1}$ and $t_j \propto n^{\alpha_2}$. Two vertices with types t_i and t_j and angular coordinates ϕ_i and ϕ_j connect to one another if the relative angle between ϕ_i and ϕ_j , $\Delta\theta$, satisfies [Bode et al. \(2015\)](#)

$$\Delta\theta \leq \frac{2\nu t_i t_j}{n}. \quad (9.3.24)$$

W.l.o.g., let the angular coordinate of the vertex with degree k be 0. Then, by (9.3.24) ϕ_i and ϕ_j satisfy

$$\begin{aligned} -\Theta(\min(kn^{\alpha_1-1}, 1)) &\leq \phi_i \leq \Theta(\min(kn^{\alpha_1-1}, 1)), \\ -\Theta(\min(kn^{\alpha_2-1}, 1)) &\leq \phi_j \leq \Theta(\min(kn^{\alpha_2-1}, 1)). \end{aligned} \quad (9.3.25)$$

Because the angular coordinates in the hyperbolic random graph are uniformly distributed, ϕ_i and ϕ_j are uniformly distributed in the above ranges. By (9.3.24), vertices i and j are connected if their relative angle is at most

$$2\nu n^{\alpha_1+\alpha_2-1}. \quad (9.3.26)$$

Thus, the probability that i and j connect, is the probability that two randomly chosen points in the intervals (9.3.25) differ in their angles by at most (9.3.26). Assume $\alpha_2 \geq \alpha_1$. Then, the probability that i and j are connected is proportional to

$$\mathbb{P}(n^{\alpha_1} \text{ and } n^{\alpha_2} \text{ neighbor connect}) \propto \min\left(\frac{n^{\alpha_1+\alpha_2-1}}{\min(n^{\alpha_2-1}, 1)}, 1\right) = \min(n^{\alpha_1} \max(n^{\alpha_2-1}, k^{-1}), 1). \quad (9.3.27)$$

Thus, (??) reduces to

$$\max_{\alpha_1, \alpha_2} n^{(\alpha_1+\alpha_2)(1-\tau)} \min(kn^{\alpha_1-1}, 1) \min(kn^{\alpha_2-1}, 1) \min(n^{\alpha_1} \max(n^{\alpha_2-1}, k^{-1}), 1). \quad (9.3.28)$$

Because of the $\min(kn^{\alpha_2-1}, 1)$ term, it is never optimal to let the max term be attained by n^{α_2-1} . Thus, the equation reduces further to

$$\max_{\alpha_1, \alpha_2} n^{(\alpha_1 + \alpha_2)(1-\tau)} \min(kn^{\alpha_1-1}, 1) \min(kn^{\alpha_2-1}, 1) \min(n^{\alpha_1} k^{-1}, 1). \tag{9.3.29}$$

The maximizers over $\alpha_1 \leq \alpha_2$ are given by

$$(n^{\alpha_1}, n^{\alpha_2}) \propto \begin{cases} (n^0, n^0), & \tau > \frac{5}{2} \\ (k, k) & \tau < \frac{5}{2}, k \ll \sqrt{n} \\ (n/k, n/k) & \tau < \frac{5}{2}, k \gg \sqrt{n}. \end{cases} \tag{9.3.30}$$

Combining this with (??) shows that

$$c(k) \propto \begin{cases} k^{-1} & \tau > \frac{5}{2} \\ k^{4-2\tau} & \tau < \frac{5}{2}, k \ll \sqrt{n} \\ k^{2\tau-6} n^{5-2\tau} & \tau < \frac{5}{2}, k \gg \sqrt{n}. \end{cases} \tag{9.3.31}$$

Taken together, this gives the following result for the clustering spectrum of the hyperbolic random graph:

Theorem 9.10 *The local clustering coefficient in the hyperbolic random graph satisfies with high probability*

$$c(k) \propto \begin{cases} k^{-1} & \tau > \frac{5}{2}, k \gg 1 \\ k^{4-2\tau} & \tau < \frac{5}{2}, 1 \ll k \ll \sqrt{n} \\ k^{2\tau-6} n^{5-2\tau} & \tau < \frac{5}{2}, k \gg \sqrt{n}. \end{cases} \tag{9.3.32}$$

In the hyperbolic random graph, typical triangles that contribute most to $c(k)$ are given by vertices of degrees as in (9.3.30). 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 .

9.3.5 SPATIAL CONFIGURATION MODELS

TO DO 9.14: Introduce the spatial configuration model and results of Deijfen et al.

9.3.6 SPATIAL PREFERENTIAL ATTACHMENT MODELS

In the past years, several spatial preferential attachment models have been considered. We shall now discuss three of such models.

Flaxman et al. (2006, 2007) study a class of geometric preferential attachment models that combines aspects of random geometric graphs and preferential attachment graphs is introduced and studied. 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

$$D_t(x_{t+1}) = \sum_{v \in V_t(x_t)} D_v^{(t)}, \tag{9.3.33}$$

where $D_v^{(t)}$ denotes the degree of vertex $v \in V_t$ in G_t . The m edges are connected to vertices (y_1, \dots, y_m) conditionally independently given (G_t, x_{t+1}) , so that, for all $v \in V_t(x_{t+1})$,

$$\mathbb{P}(y_i = v \mid G_t) = \frac{D_v^{(t)}}{\max(D_t(x_{t+1}), \alpha m A_r t)}, \tag{9.3.34}$$

while

$$\mathbb{P}(y_i = x_{t+1} \mid G_t) = 1 - \frac{D_t(x_{t+1})}{\max(D_t(x_{t+1}), \alpha m A_r t)}, \tag{9.3.35}$$

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 $\{p_k\}_{k=m}^\infty$ such that, **whp**,

$$P_k^{(n)} = p_k(1 + o(1)), \tag{9.3.36}$$

where $(p_k)_{k \geq m}^\infty$ satisfies (??) 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.

[Aiello et al. \(2008\)](#) 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 the model of [Aiello et al. \(2008\)](#) 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, [Aiello et al. \(2008\)](#) 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

$$d(x, y) = \min\{\|x - y + u\|_\infty : u \in \{0, 1, -1\}^m\}. \tag{9.3.37}$$

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

$$R(i, t) = \frac{A_1 D_i^{(t)} + A_2}{t + A_3}, \quad (9.3.38)$$

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 $pA_1 \leq 1$. One of the main results of the paper is that this model is a scale-free graph process. Indeed, denote

$$p_k = \frac{p^k}{1 + kpA_1 + pA_2} \prod_{j=0}^{k-1} \frac{jA_1 + A_2}{1 + A_2 + pA_2}, \quad (9.3.39)$$

then (Aiello et al., 2008, Theorem 1.1) shows that whp, for $k \leq (n^{1/8} / \log n)^{4pA_1/(2pA_1+1)}$, the degree sequence of the graph of size n satisfies (recall (??))

$$P_k^{(n)} = p_k(1 + o(1)), \quad (9.3.40)$$

and $(p_k)_{k \geq 0}$ satisfies (1.4.4) with $\tau = 1 + 1/(pA_1) \in [2, \infty)$. Further results involve the study of maximal in-degrees and the total number of edges.

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.

9.4 NOTES AND DISCUSSION

TO DO 9.16: Extend the notes and discussion section.

Notes on Section 9.1.

Bloznelis et al. (2012) study the general directed inhomogeneous random graph studied in this section.

Cao and Olvera-Cravioto (2017) investigate general rank-1 inhomogeneous digraphs. Lee and Olvera-Cravioto (2017) use this to prove that the limiting PageRank of such graphs exists, and that the solution obeys the same recurrence relation as on a Galton-Watson tree.

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 9.2, 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.

Chen and Olvera-Cravioto (2013) study a way to obtain nearly i.i.d. in- and out-degrees.

Notes on Section 9.2.

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 \(2003\)](#).

Notes on Section 9.3.

9.5 EXERCISES FOR CHAPTER 9

TO DO 9.17: Design more exercises for Chapter 9.

Exercise 9.1 (Topology of the strongly-connected component for di-graphs) *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 (Clustering in model with edges and triangles) *Show that the clustering coefficient in the model where each pair of vertices is independently connected with probability λ/n , as for $\text{ER}_n(\lambda/n)$ and each triple forms a triangle with probability μ/n^2 , independently for all triplets and independently of the status of the edges, converges to $\mu/(\mu + \lambda^2)$.*

Exercise 9.3 (Law of large number for $|\mathcal{C}_{\max}|$ in configuration model with households) *Use Theorem 4.4 to prove (9.2.17).*

Exercise 9.4 (Degree moments in scale-free percolation [Deijfen et al. \(2013\)](#)) *Show that $\mathbb{E}[D_0^p] < \infty$ when $p < \gamma$ and $\mathbb{E}[D_0^p] = \infty$ when $p > \gamma$. In particular, the variance of the degrees is finite precisely when $\gamma > 2$.*

Exercise 9.5 (Positive correlation between edge statuses in scale-free percolation) *Show that, for scale-free percolation, and for all x, y, z distinct and $\lambda > 0$,*

$$\mathbb{P}(\{x, y\} \text{ and } \{x, z\} \text{ occupied}) \geq \mathbb{P}(\{x, y\} \text{ occupied}) \mathbb{P}(\{x, z\} \text{ occupied}), \quad (9.5.1)$$

the inequality being strict when $\mathbb{P}(W_0 = 0) < 1$. In other words, the edge statuses are positively correlated.

APPENDIX A
SOME FACTS ABOUT
METRIC SPACES

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.

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 \mathcal{X} be a space. A distance on \mathcal{X} is a function $d: \mathcal{X}^2 \mapsto [0, \infty)$ such that*

- (a) $0 \leq d(x, y) < \infty$ for all $x, y \in \mathcal{X}$;
- (b) $d(x, y) = d(y, x)$ for all $x, y \in \mathcal{X}$;
- (c) $d(x, z) \leq d(x, y) + d(y, z)$ for all $x, y, z \in \mathcal{X}$.

Definition A.2 (Metric space) *Let \mathcal{X} be a space and d a metric on it. Then (\mathcal{X}, d) is called a metric space.*

We next discuss some desirable properties of metric spaces:

Definition A.3 (Complete metric space) *Let (\mathcal{X}, d) be a metric space. We say that (\mathcal{X}, d) is complete when every Cauchy sequence has a limit. Here, a Cauchy sequence is a sequence $(x_n)_{n \geq 1}$ with $x_n \in \mathcal{X}$ 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 (\mathcal{X}, d) be a metric space. We say that (\mathcal{X}, d) is separable if it contains a countable subset that is dense in (\mathcal{X}, d) . This means that there exists a countable set $\mathcal{A} \subseteq \mathcal{X}$ such that, for every $x \in \mathcal{X}$ there exists a sequence $(a_n)_{n \geq 1}$ with $a_n \in \mathcal{A}$ such that $d(a_n, x) \rightarrow 0$.*

Definition A.5 (Polish space) *The metric space (\mathcal{X}, d) is called Polish when it is separable and complete.*

A.2 PROPERTIES OF THE METRIC $d_{\mathcal{G}_\star}$ ON ROOTED GRAPHS

For $(G, o) \in \mathcal{G}_\star$, let

$$[G, o] = \{(G', o') : (G', o') \simeq (G, o)\} \quad (\text{A.2.1})$$

denote the equivalence class in \mathcal{G}_\star corresponding to (G, o) . We further let

$$[\mathcal{G}_\star] = \{(G, o) : (G, o) \in \mathcal{G}_\star\} \quad (\text{A.2.2})$$

denote the set of equivalence class in \mathcal{G}_\star . This will be the set on which the distance $d_{\mathcal{G}_\star}$ acts

In this section, we prove that $([\mathcal{G}_\star], d_{\mathcal{G}_\star})$ is a Polish space:

Theorem A.6 (Rooted graphs form a Polish space) *$d_{\mathcal{G}_\star}$ is a well-defined metric on $[\mathcal{G}_\star]$. Further, the metric space $([\mathcal{G}_\star], d_{\mathcal{G}_\star})$ 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 $d_{\mathcal{G}_\star}$ is an ultrametric, which is a slightly stronger property than being a metric, and which also implies that $(\mathcal{G}_\star, d_{\mathcal{G}_\star})$ is a metric space. In Proposition A.10, we show that the metric space $(\mathcal{G}_\star, d_{\mathcal{G}_\star})$ is complete, and in Proposition A.12, we show that it is separable. After that, we complete the proof of Theorem A.6.

A.2.1 ULTRAMETRIC PROPERTY OF $d_{\mathcal{G}_\star}$

In this section, we prove that $d_{\mathcal{G}_\star} : \mathcal{G}_\star \times \mathcal{G}_\star \rightarrow [0, 1]$ 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 $d_{\mathcal{G}_\star}((G_1, o_1), (G_2, o_2))$ 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 ($d_{\mathcal{G}_\star}$ is well-defined on $[\mathcal{G}_\star]$) *The equality $d_{\mathcal{G}_\star}((\hat{G}_1, \hat{o}_1), (\hat{G}_2, \hat{o}_2)) = d_{\mathcal{G}_\star}((G_1, o_1), (G_2, o_2))$ 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, $d_{\mathcal{G}_\star} : [\mathcal{G}_\star] \times [\mathcal{G}_\star]$ is well-defined.*

Proposition A.8 (Ultrametricity) *The map $d_{\mathcal{G}_\star} : \mathcal{G}_\star \times \mathcal{G}_\star \rightarrow [0, 1]$ is an ultrametric, meaning that*

- (a) $d_{\mathcal{G}_\star}((G_1, o_1), (G_2, o_2)) = 0$ precisely when $(G_1, o_1) \simeq (G_2, o_2)$;
- (b) $d_{\mathcal{G}_\star}((G_1, o_1), (G_2, o_2)) = d_{\mathcal{G}_\star}((G_2, o_2), (G_1, o_1))$ for all $(G_1, o_1), (G_2, o_2) \in \mathcal{G}_\star$;
- (c) $d_{\mathcal{G}_\star}((G_1, o_1), (G_3, o_3)) \leq \max\{d_{\mathcal{G}_\star}((G_1, o_1), (G_2, o_2)), d_{\mathcal{G}_\star}((G_2, o_2), (G_3, o_3))\}$ for all $(G_1, o_1), (G_2, o_2), (G_3, o_3) \in \mathcal{G}_\star$.

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 $(B_r^{(G_1)}(o_1), o_1) \simeq (B_r^{(G_2)}(o_2), o_2)$ 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 $\phi_r: (B_r^{(G_1)}(o_1), o_1) \rightarrow (B_r^{(G_2)}(o_2), o_2)$, which exists by assumption. Extend ϕ_r to (G_1, o_1) by defining

$$\psi_r(v) = \begin{cases} \phi_r(v) & \text{for } v \in V(B_r^{(G_1)}(o_1)); \\ o_2 & \text{otherwise.} \end{cases} \quad (\text{A.2.3})$$

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 $\psi_r|_{V_0^{(G_1)}}$ be the restriction of ψ_r to $V_0^{(G_1)} = \{o_1\}$. Then we know that $\psi_r(v) = o_2$ for every $v \in V_0^{(G_1)}$ and $r \geq 0$. We next let $\psi_r|_{V_1^{(G_1)}}$ be the restriction of ψ_r to $V_1^{(G_1)}$. Then, $\psi_r|_{V_1^{(G_1)}}$ is an isomorphism between $(B_1^{(G_1)}(o_1), o_1)$ and $(B_1^{(G_2)}(o_2), o_2)$ for every r . Since there are only *finitely* many such isomorphisms, the same isomorphism ϕ'_1 needs to be repeated infinitely many times in the sequence $(\psi_r|_{V_1^{(G_1)}})_{r \geq 1}$. Let \mathbb{N}_1 denote the values of r for which

$$\psi_r|_{V_1^{(G_1)}} = \phi'_1 \quad \forall r \in \mathbb{N}_1. \quad (\text{A.2.4})$$

Now we repeat this argument to $k = 2$. Let $\psi_r|_{V_2^{(G_1)}}$ be the restriction of ψ_r to $V_2^{(G_1)}$. Again, $\psi_r|_{V_2^{(G_1)}}$ is an isomorphism between $(B_2^{(G_1)}(o_1), o_1)$ and $(B_2^{(G_2)}(o_2), o_2)$ for every r . Since there are again only finitely many such isomorphisms, the same isomorphism ϕ'_2 needs to be repeated infinitely many times in the sequence $(\psi_r|_{V_2^{(G_1)}})_{r \in \mathbb{N}_1}$. Let \mathbb{N}_2 denote the values of $r \in \mathbb{N}_1$ for which

$$\psi_r|_{V_2^{(G_1)}} = \phi'_2 \quad \forall r \in \mathbb{N}_2. \quad (\text{A.2.5})$$

We now generalize this argument to general $k \geq 2$. Let $\psi_r|_{V_k^{(G_1)}}$ be the restriction of ψ_r to $V_k^{(G_1)}$. Again, $\psi_r|_{V_k^{(G_1)}}$ is an isomorphism between $(B_k^{(G_1)}(o_1), o_1)$ and $(B_k^{(G_2)}(o_2), o_2)$ for every r . Since there are again only finitely many such isomorphisms, the same isomorphism ϕ'_k needs to be repeated infinitely many times in the sequence $(\psi_r|_{V_k^{(G_1)}})_{r \in \mathbb{N}_{k-1}}$. Let \mathbb{N}_k denote the values of $r \in \mathbb{N}_{k-1}$ for which

$$\psi_r|_{V_k^{(G_1)}} = \phi'_k \quad \forall r \in \mathbb{N}_k. \quad (\text{A.2.6})$$

Then, we see that \mathbb{N}_k is a decreasing infinite set.

Let us define ψ'_ℓ to be the first element of the sequence $(\psi_r)_{r \in \mathbb{N}_\ell}$. Then, it follows that $\psi'_\ell(v) = \phi'_k(v)$ for all $\ell \geq k$ and all $v \in V_k^{(G_1)}$. Denote $U_0 = \{o_1\}$ and $U_k = V_k^{(G_1)} \setminus V_{k-1}^{(G_1)}$. 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 $(\psi'_\ell)_{\ell \geq 1}$ converge pointwise to

$$\psi(v) = \psi'_\infty(v) = \sum_{k \geq 1} \phi'_k(v) \mathbb{1}_{\{v \in U_k\}}. \quad (\text{A.2.7})$$

We claim that ψ is the desired isomorphism between (G_1, o_1) and (G_2, o_2) . The map ψ is clearly bijective, since $\phi'_k: U_k \rightarrow \phi'_k(U_k)$ 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 ϕ'_k is an isomorphism between $(B_k^{(G_1)}(o_1), o_1)$ and $(B_k^{(G_2)}(o_2), o_2)$, it follows that $\phi'_k(u), \phi'_k(v) \in V(B_k^{(G_2)}(o_2))$ and $\{\phi'_k(u), \phi'_k(v)\} \in E(B_k^{(G_2)}(o_2))$ 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 $\{\psi(u), \psi(v)\} \in E(B_k^{(G_2)}(o_2))$ 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. \square

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 $(B_r^{(G_1)}(o_1), o_1) \simeq (B_r^{(G_2)}(o_2), o_2)$ if and only if $(B_r^{(\hat{G}_1)}(\hat{o}_1), \hat{o}_1) \simeq (B_r^{(\hat{G}_2)}(\hat{o}_2), \hat{o}_2)$. Therefore, $d_{\mathcal{G}_*}((G_1, o_1), (G_2, o_2))$ 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 $d_{\mathcal{G}_*}((G_1, o_1), (G_2, o_2))$ is constant on such equivalence classes. This makes the metric $d_{\mathcal{G}_*}([G_1, o_1], [G_2, o_2])$ well-defined for $[G_1, o_1], [G_2, o_2] \in [\mathcal{G}_*]$. \square

Proof of Proposition A.8. (a) Assume that $d_{\mathcal{G}_*}((G_1, o_1), (G_2, o_2)) = 0$. Then we have $(B_r^{(G_1)}(o_1), o_1) \simeq (B_r^{(G_2)}(o_2), o_2)$ 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 $r_{ij} = \sup\{r : (B_r^{(G_i)}(o_i), o_i) \simeq (B_r^{(G_j)}(o_j), o_j)\}$. Then $(B_r^{(G_1)}(o_1), o_1) \simeq (B_r^{(G_3)}(o_3), o_3)$ for all $r \leq \min\{r_{13}, r_{23}\}$ and $(B_r^{(G_1)}(o_2), o_2) \simeq (B_r^{(G_3)}(o_3), o_3)$ for all $r \leq \min\{r_{13}, r_{23}\}$. We conclude that $(B_r^{(G_1)}(o_1), o_1) \simeq (B_r^{(G_2)}(o_2), o_2)$ 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)). \square

A.2.2 COMPLETENESS OF $([\mathcal{G}_*], d_{\mathcal{G}_*})$

In this section, we prove that $([\mathcal{G}_*], d_{\mathcal{G}_*})$ is complete:

Proposition A.10 (Completeness) *The metric space $([\mathcal{G}_*], d_{\mathcal{G}_*})$ 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 $(B_r^{(G_s)}(o_s), o_s) \simeq (B_r^{(G_r)}(o_r), o_r)$ for all $r \leq s$. Then there exists a connected locally finite rooted graph (G, o) such that $(G_r(o_r), o_r) \simeq (B_r^{(G)}(o), o)$. 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 $\phi_r : V(B_r(o_r)) \rightarrow V_r$ recursively as follows. Let ϕ_0 be the unique isomorphism from $V(B_0(o_0)) = \{o_0\}$ to $V_0 = \{1\}$.

Let ψ_r be an isomorphism between (G_{r-1}, o_{r-1}) and $(B_{r-1}^{(G_r)}(o_r), o_r)$, and η_r an

arbitrary bijection between $V(G_{r-1}) \setminus V(B_{r-1}^{(G_r)})$ to $V_r \setminus V_{r-1}$. Define

$$\phi_r(v) = \begin{cases} \phi_{r-1}(\psi_r^{-1}(v)) & \text{for } v \in V(B_{r-1}^{(G_r)}), \\ \eta_r(v) & \text{for } v \in V(G_{r-1}) \setminus V(B_{r-1}^{(G_r)}). \end{cases} \quad (\text{A.2.8})$$

Then we define $(G'_r, o'_r) = (\phi_r(G_r), \phi_r(o_r))$, where $\phi_r(G_r)$ is the graph consisting of the vertex set $\{\phi_r(v) : v \in V(G_r)\}$ and edge set $\{\{\phi_r(u), \phi_r(v)\} : \{u, v\} \in E(G_r)\}$. Let us derive some properties of (G'_r, o'_r) . First of all, we see that $o'_r = 1$ for every $r \geq 0$, by construction. Further, ϕ_r is a bijection, so that $(B_r^{(G'_r)}(o'_r), o'_r) \simeq (B_r^{(G_r)}(o_r), o_r) = (G_r, o_r)$.

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 $(B_r^{(G)}(o), o) = (B_r^{(G'_r)}(o'_r), o'_r) \simeq (B_r^{(G_r)}(o_r), o_r) = (G_r, o_r)$, 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 $(B_r^{(G')}(o'), o') \simeq (B_r^{(G_r)}(o_r), o_r) = (G_r, o_r)$ for every $r \geq 0$, then $(B_r^{(G')}(o'), o') \simeq (B_r^{(G)}(o), o)$ for every $r \geq 0$, so that $(G, o) \simeq (G', o')$, as required. \square

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 $n, m \geq N$,

$$d_{\mathcal{G}_*}([G_n, o_n], [G_m, o_m]) \leq \varepsilon. \quad (\text{A.2.9})$$

We note that, by Proposition A.7,

$$d_{\mathcal{G}_*}([G_n, o_n], [G_m, o_m]) = d_{\mathcal{G}_*}((G_n, o_n), (G_m, o_m)), \quad (\text{A.2.10})$$

so from now on, we can work with the representatives instead. Since

$$d_{\mathcal{G}_*}((G_n, o_n), (G_m, o_m)) \leq \varepsilon, \quad (\text{A.2.11})$$

we obtain that for all $r \leq 1/\varepsilon - 1$ and $n, m \geq N$, $(B_r^{(G_n)}(o_n), o_n) \simeq (B_r^{(G_m)}(o_m), o_m)$. Equivalently, this implies that for every $r \geq 1$, there exists an n_r such that, for all $n \geq n_r$,

$$(B_r^{(G_n)}(o_n), o_n) \simeq (B_r^{(G_{n_r})}(o_{n_r}), o_{n_r}). \quad (\text{A.2.12})$$

Clearly, we may select n_r such that n_r is strictly increasing. Define $(G'_r, o'_r) = (B_r^{(G_{n_r})}(o_{n_r}), o_{n_r})$. 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 $(B_r^{(G)}(o), o) \simeq (G'_r(o'_r), o'_r)$. But then also

$$(B_r^{(G_n)}(o_n), o_n) \simeq (B_r^{(G_{n_r})}(o_{n_r}), o_{n_r}) = (G'_r(o'_r), o'_r). \quad (\text{A.2.13})$$

This, in turn, implies that, for all $n \geq n_r$,

$$d_{\mathcal{G}_*}([G, o], [G_n, o_n]) = d_{\mathcal{G}_*}((G, o), (G_n, o_n)) \leq 1/(r + 1). \quad (\text{A.2.14})$$

Since $r \geq 1$ is arbitrary, we conclude that $[G_n, o_n]$ converges to $[G, o]$, which is in $[\mathcal{G}_*]$, so that $([\mathcal{G}_*], d_{\mathcal{G}_*})$ is complete. \square

A.2.3 SEPARABILITY OF $([\mathcal{G}_\star], d_{\mathcal{G}_\star})$

In this section, we prove that $([\mathcal{G}_\star], d_{\mathcal{G}_\star})$ is separable:

Proposition A.12 (Separability) *The metric space $(\mathcal{G}_\star, d_{\mathcal{G}_\star})$ is separable.*

Proof We need to show that there exists a countable dense subset in $(\mathcal{G}_\star, d_{\mathcal{G}_\star})$. Consider the set of all *finite* rooted graphs, which is certainly countable. Also, $(B_r^{(G)}(o), o)$ is a finite rooted graph for all $r \geq 0$. Finally, $d_{\mathcal{G}_\star}((B_r^{(G)}(o), o), (G, o)) \leq 1/(r+1)$, so that $(B_r^{(G)}(o), o)$ converges to (G, o) when $r \rightarrow \infty$. Thus, the space of finite rooted graphs is dense and countable. This completes the proof that $(\mathcal{G}_\star, d_{\mathcal{G}_\star})$ is separable. \square

A.2.4 $([\mathcal{G}_\star], d_{\mathcal{G}_\star})$ 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 $d_{\mathcal{G}_\star}$ is well-defined on $[\mathcal{G}_\star] \times [\mathcal{G}_\star]$ by Proposition A.7. Proposition A.8 implies that $d_{\mathcal{G}_\star}$ is an (ultra)metric on $[\mathcal{G}_\star]$. Finally, Proposition A.10 proves that $([\mathcal{G}_\star], d_{\mathcal{G}_\star})$ is complete, while Proposition A.12 proves that $([\mathcal{G}_\star], d_{\mathcal{G}_\star})$ is separable. Thus, $([\mathcal{G}_\star], d_{\mathcal{G}_\star})$ is a Polish space. \square

A.3 NOTES AND DISCUSSION

Notes on Section A.1

We draw inspiration from [Howes \(1995\)](#) and [Rudin \(1987, 1991\)](#).

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.

REFERENCES

- Aiello, W., Bonato, A., Cooper, C., Janssen, J., and Pralat, P. 2008. A spatial web graph model with local influence regions. *Internet Math.*, **5**(1-2), 175–196.
- Albert, R., and Barabási, A.-L. 2002. Statistical mechanics of complex networks. *Rev. Modern Phys.*, **74**(1), 47–97.
- Aldous, D. 1985. Exchangeability and related topics. Pages 1–198 of: *École d’été de probabilités de Saint-Flour, XIII—1983*. Lecture Notes in Math., vol. **1117**. Berlin: Springer.
- Aldous, D., and Lyons, R. 2007. Processes on unimodular random networks. *Electron. J. Probab.*, **12**, no. 54, 1454–1508.
- Aldous, D., and Steele, J. M. 2004. The objective method: probabilistic combinatorial optimization and local weak convergence. **110**, 1–72.
- Arratia, R., Barbour, AD, and Tavaré, S. 2003. *Logarithmic combinatorial structures: a probabilistic approach*. European Mathematical Society.
- Athreya, K., and Ney, P. 1972. *Branching processes*. New York: Springer-Verlag. Die Grundlehren der mathematischen Wissenschaften, Band 196.
- Athreya, K. B., Ghosh, A. P., and Sethuraman, S. 2008. Growth of preferential attachment random graphs via continuous-time branching processes. *Proc. Indian Acad. Sci. Math. Sci.*, **118**(3), 473–494.
- Athreya, K.B. 2007. Preferential attachment random graphs with general weight function. *Internet Math.*, **4**(4), 401–418.
- Ball, F., Sirl, D., and Trapman, P. 2009. Threshold behaviour and final outcome of an epidemic on a random network with household structure. *Adv. in Appl. Probab.*, **41**(3), 765–796.
- Ball, F., Sirl, D., and Trapman, P. 2010. Analysis of a stochastic SIR epidemic on a random network incorporating household structure. *Math. Biosci.*, **224**(2), 53–73.
- Barabási, A.-L. 2002. *Linked: The New Science of Networks*. Cambridge, Massachusetts: Perseus Publishing.
- Barabási, A.-L., and Albert, R. 1999. Emergence of scaling in random networks. *Science*, **286**(5439), 509–512.
- Barbour, A. D., and Reinert, G. 2001. Small worlds. *Random Structures Algorithms*, **19**(1), 54–74.
- Barbour, A. D., and Reinert, G. 2004. Correction: “Small worlds” [Random Structures Algorithms **19** (2001), no. 1, 54–74; MR1848027]. *Random Structures Algorithms*, **25**(1), 115.
- Barbour, A. D., and Reinert, G. 2006. Discrete small world networks. *Electron. J. Probab.*, **11**, no. 47, 1234–1283 (electronic).
- Bender, E.A., and Canfield, E.R. 1978. The asymptotic number of labelled graphs with given degree sequences. *Journal of Combinatorial Theory (A)*, **24**, 296–307.
- Benjamini, I., and Schramm, O. 2001. Recurrence of distributional limits of finite planar graphs. *Electron. J. Probab.*, **6**, no. 23, 13 pp. (electronic).
- Benjamini, I., Kesten, H., Peres, Y., and Schramm, O. 2004. Geometry of the uniform spanning forest: transitions in dimensions 4, 8, 12, *Ann. of Math. (2)*, **160**(2), 465–491.
- Benjamini, I., Lyons, R., and Schramm, O. 2015. Unimodular random trees. *Ergodic Theory Dynam. Systems*, **35**(2), 359–373.
- Berger, N. 2002. Transience, recurrence and critical behavior for long-range percolation. *Comm. Math. Phys.*, **226**(3), 531–558.
- Berger, N., Borgs, C., Chayes, J. T., D’Souza, R. M., and Kleinberg, R. D. 2004. Competition-induced preferential attachment. Pages 208–221 of: *Automata, languages and programming*. Lecture Notes in Comput. Sci., vol. **3142**. Berlin: Springer.
- Berger, N., Borgs, C., Chayes, J. T., D’Souza, R. M., and Kleinberg, R. D. 2005. Degree distribution of competition-induced preferential attachment graphs. *Combin. Probab. Comput.*, **14**(5-6), 697–721.

- Berger, N., Borgs, C., Chayes, J., and Saberi, A. 2014. Asymptotic behavior and distributional limits of preferential attachment graphs. *Ann. Probab.*, **42**(1), 1–40.
- Bhamidi, S., Hofstad, R. van der, and Hooghiemstra, G. 2010a. First passage percolation on random graphs with finite mean degrees. *Ann. Appl. Probab.*, **20**(5), 1907–1965.
- Bhamidi, S., Hofstad, R. van der, and Leeuwaarden, J. van. 2010b. Scaling limits for critical inhomogeneous random graphs with finite third moments. *Electronic Journal of Probability*, **15**, 1682–1702.
- Bhamidi, S., Hofstad, R. van der, and Leeuwaarden, J. van. 2012. Novel scaling limits for critical inhomogeneous random graphs. *Ann. Probab.*, **40**(6), 2299–2361.
- Bingham, N. H., Goldie, C. M., and Teugels, J. L. 1989. *Regular variation*. Encyclopedia of Mathematics and its Applications, vol. **27**. Cambridge: Cambridge University Press.
- Biskup, M. 2004. On the scaling of the chemical distance in long-range percolation models. *Ann. Probab.*, **32**(4), 2938–2977.
- Bloznelis, M. *Loglog distances in a power law random intersection graph*. Preprint (2008).
- Bloznelis, M. 2010a. Component evolution in general random intersection graphs. *SIAM J. Discrete Math.*, **24**(2), 639–654.
- Bloznelis, M. 2010b. The largest component in an inhomogeneous random intersection graph with clustering. *Electron. J. Combin.*, **17**(1), Research Papr 110, 17.
- Bloznelis, M., Götze, F., and Jaworski, J. 2012. Birth of a strongly connected giant in an inhomogeneous random digraph. *J. Appl. Probab.*, **49**(3), 601–611.
- Bode, M., Fountoulakis, N., and Müller, T. 2015. On the largest component of a hyperbolic model of complex networks. *Electron. J. Combin.*, **22**(3), Paper 3.24, 46.
- Boldi, P., and Vigna, S. 2004. The WebGraph Framework I: Compression Techniques. Pages 595–601 of: *Proc. of the Thirteenth International World Wide Web Conference (WWW 2004)*. Manhattan, USA: ACM Press.
- Boldi, P., Rosa, M., Santini, M., and Vigna, S. 2011. Layered Label Propagation: A MultiResolution Coordinate-Free Ordering for Compressing Social Networks. Pages 587–596 of: Srinivasan, Sadagopan, Ramamritham, Krithi, Kumar, Arun, Ravindra, M. P., Bertino, Elisa, and Kumar, Ravi (eds), *Proceedings of the 20th international conference on World Wide Web*. ACM Press.
- Bollobás, B. 1980. A probabilistic proof of an asymptotic formula for the number of labelled regular graphs. *European J. Combin.*, **1**(4), 311–316.
- Bollobás, B. 2001. *Random graphs*. Second edn. Cambridge Studies in Advanced Mathematics, vol. **73**. Cambridge: Cambridge University Press.
- Bollobás, B., and Fernandez de la Vega, W. 1982. The diameter of random regular graphs. *Combinatorica*, **2**(2), 125–134.
- Bollobás, B., and Riordan, O. 2004a. The diameter of a scale-free random graph. *Combinatorica*, **24**(1), 5–34.
- Bollobás, B., and Riordan, O. 2004b. Shortest paths and load scaling in scale-free trees. *Phys. Rev. E.*, **69**, 036114.
- Bollobás, B., and Riordan, O. 2015. An old approach to the giant component problem. *J. Combin. Theory Ser. B*, **113**, 236–260.
- Bollobás, B., Riordan, O., Spencer, J., and Tusnády, G. 2001. The degree sequence of a scale-free random graph process. *Random Structures Algorithms*, **18**(3), 279–290.
- Bollobás, B., Borgs, C., Chayes, J., and Riordan, O. 2003. Directed scale-free graphs. Pages 132–139 of: *Proceedings of the Fourteenth Annual ACM-SIAM Symposium on Discrete Algorithms (Baltimore, MD, 2003)*. New York: ACM.
- Bollobás, B., Janson, S., and Riordan, O. 2005. The phase transition in the uniformly grown random graph has infinite order. *Random Structures Algorithms*, **26**(1-2), 1–36.
- Bollobás, B., Janson, S., and Riordan, O. 2007. The phase transition in inhomogeneous random graphs. *Random Structures Algorithms*, **31**(1), 3–122.
- Bringmann, K., Keusch, R., and Lengler, J. (2015). *Geometric Inhomogeneous Random Graphs*. Preprint arXiv:1511.00576 [cs.SI].

- Bringmann, K., Keusch, R., and Lengler, J. (2016). *Average Distance in a General Class of Scale-Free Networks with Underlying Geometry*. Preprint arXiv:1602.05712 [cs.DM].
- Britton, T., Deijfen, M., and Martin-Löf, A. 2006. Generating simple random graphs with prescribed degree distribution. *J. Stat. Phys.*, **124**(6), 1377–1397.
- Broder, A., Kumar, R., Maghoul, F., Raghavan, P., Rajagopalan, S., Stata, R., Tomkins, A., and Wiener, J. 2000. Graph Structure in the Web. *Computer Networks*, **33**, 309–320.
- Callaway, D.S., Hopcroft, J.E., Kleinberg, J. M., Newman, M. E. J., and Strogatz, S.H. 2001. Are randomly grown graphs really random? *Physical Review E*, **64**, 041902.
- Cao, J., and Olvera-Cravioto, M. 2017. *On a general class of inhomogeneous random digraphs*. Preprint arXiv:1712.03319 [math.PR].
- Caravenna, F., Garavaglia, A., and van der Hofstad, R. 2019. Diameter in ultra-small scale-free random graphs. *Random Structures Algorithms*, **54**(3), 444–498.
- 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.
- Chen, N., and Olvera-Cravioto, M. 2013. Directed random graphs with given degree distributions. *Stoch. Syst.*, **3**(1), 147–186.
- Chung, F., and Lu, L. 2001. The diameter of sparse random graphs. *Adv. in Appl. Math.*, **26**(4), 257–279.
- Chung, F., and Lu, L. 2002a. The average distances in random graphs with given expected degrees. *Proc. Natl. Acad. Sci. USA*, **99**(25), 15879–15882 (electronic).
- Chung, F., and Lu, L. 2002b. Connected components in random graphs with given expected degree sequences. *Ann. Comb.*, **6**(2), 125–145.
- Chung, F., and Lu, L. 2003. The average distance in a random graph with given expected degrees. *Internet Math.*, **1**(1), 91–113.
- Chung, F., and Lu, L. 2004. Coupling online and offline analyses for random power law graphs. *Internet Math.*, **1**(4), 409–461.
- Chung, F., and Lu, L. 2006a. *Complex graphs and networks*. CBMS Regional Conference Series in Mathematics, vol. **107**. Published for the Conference Board of the Mathematical Sciences, Washington, DC.
- Chung, F., and Lu, L. 2006b. The volume of the giant component of a random graph with given expected degrees. *SIAM J. Discrete Math.*, **20**, 395–411.
- Cohen, R., and Havlin, S. 2003. Scale-free networks are ultrasmall. *Phys. Rev. Lett.*, **90**, 058701, 1–4.
- Cooper, C., and Frieze, A. 2004. The size of the largest strongly connected component of a random digraph with a given degree sequence. *Combin. Probab. Comput.*, **13**(3), 319–337.
- Cortez, R. 2012. Composition and Structure of a Large Online Social Network in the Netherlands. *PLOS ONE*, **7**(4), 1–8.
- Darling, D. A. 1970. The Galton-Watson process with infinite mean. *J. Appl. Probability*, **7**, 455–456.
- Davies, P. L. 1978. The simple branching process: a note on convergence when the mean is infinite. *J. Appl. Probab.*, **15**(3), 466–480.
- Deijfen, M., and Kets, W. 2009. Random intersection graphs with tunable degree distribution and clustering. *Probab. Engrg. Inform. Sci.*, **23**(4), 661–674.
- Deijfen, M., Hofstad, R. van der, and Hooghiemstra, G. 2013. Scale-free percolation. *Annales de l'institut Henri Poincaré (B) Probability and Statistics*, **49**(3), 817–838. cited By (since 1996)0.
- Dembo, A., and Montanari, A. 2010. Ising models on locally tree-like graphs. *The Annals of Applied Probability*, **20**(2), 565–592.
- Deprez, P, Hazra, R., and Wüthrich, M. 2015. *Inhomogeneous Long-Range Percolation for Real-Life Network Modeling*.
- Dereich, S., and Mörters, P. 2009. Random networks with sublinear preferential attachment: Degree evolutions. *Electronic Journal of Probability*, **14**, 122–1267.

- Dereich, S., and Mörters, P. 2011. Random networks with concave preferential attachment rule. *Jahresber. Dtsch. Math.-Ver.*, **113**(1), 21–40.
- Dereich, S., and Mörters, P. 2013. Random networks with sublinear preferential attachment: the giant component. *Ann. Probab.*, **41**(1), 329–384.
- Dereich, S., Mönch, C., and Mörters, P. 2012. Typical distances in ultrasmall random networks. *Adv. in Appl. Probab.*, **44**(2), 583–601.
- Dhara, S., Hofstad, R. van der, van Leeuwaarden, J., and Sen, S. 2017. Critical window for the configuration model: finite third moment degrees. *Electron. J. Probab.*, **22**, Paper No. 16, 33.
- Dhara, S., Hofstad, R. van der, Leeuwaarden, J.S.H., and Sen, S. Preprint (2016). *Heavy-tailed configuration models at criticality*. Available at arXiv:1612.00650 [math.PR].
- Ding, J., Kim, J. H., Lubetzky, E., and Peres, Y. 2010. Diameters in supercritical random graphs via first passage percolation. *Combin. Probab. Comput.*, **19**(5-6), 729–751.
- Ding, J., Kim, J. H., Lubetzky, E., and Peres, Y. 2011. Anatomy of a young giant component in the random graph. *Random Structures Algorithms*, **39**(2), 139–178.
- Dommers, S., Hofstad, R. van der, and Hooghiemstra, G. 2010. Diameters in preferential attachment graphs. *Journ. Stat. Phys.*, **139**, 72–107.
- Drmotič, M. 2009. *Random trees*. SpringerWienNewYork, Vienna. An interplay between combinatorics and probability.
- Durrett, R. 2003. Rigorous result for the CHKNS random graph model. Pages 95–104 of: *Discrete random walks (Paris, 2003)*. Discrete Math. Theor. Comput. Sci. Proc., AC. Assoc. Discrete Math. Theor. Comput. Sci., Nancy.
- Durrett, R. 2007. *Random graph dynamics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge: Cambridge University Press.
- Eckhoff, M., Goodman, J., Hofstad, R. van der, and Nardi, F.R. 2013. Short paths for first passage percolation on the complete graph. *J. Stat. Phys.*, **151**(6), 1056–1088.
- Erdős, P., and Rényi, A. 1959. On random graphs. I. *Publ. Math. Debrecen*, **6**, 290–297.
- Erdős, P., and Rényi, A. 1960. On the evolution of random graphs. *Magyar Tud. Akad. Mat. Kutató Int. Közl.*, **5**, 17–61.
- Erdős, P., and Rényi, A. 1961a. On the evolution of random graphs. *Bull. Inst. Internat. Statist.*, **38**, 343–347.
- Erdős, P., and Rényi, A. 1961b. On the strength of connectedness of a random graph. *Acta Math. Acad. Sci. Hungar.*, **12**, 261–267.
- Esker, H. van den, Hofstad, R. van der, Hooghiemstra, G., and Znamenski, D. 2006. Distances in random graphs with infinite mean degrees. *Extremes*, **8**, 111–140.
- Esker, H. van den, Hofstad, R. van der, and Hooghiemstra, G. 2008. Universality for the distance in finite variance random graphs. *J. Stat. Phys.*, **133**(1), 169–202.
- Faloutsos, C., Faloutsos, P., and Faloutsos, M. 1999. On power-law relationships of the internet topology. *Computer Communications Rev.*, **29**, 251–262.
- Federico, L., and van der Hofstad, R. 2017. Critical window for connectivity in the configuration model. *Combin. Probab. Comput.*, **26**(5), 660–680.
- Fernholz, D., and Ramachandran, V. 2007. The diameter of sparse random graphs. *Random Structures Algorithms*, **31**(4), 482–516.
- Fill, J., Scheinerman, E., and Singer-Cohen, K. 2000. Random intersection graphs when $m = \omega(n)$: an equivalence theorem relating the evolution of the $G(n, m, p)$ and $G(n, p)$ models. *Random Structures Algorithms*, **16**(2), 156–176.
- Flaxman, A., Frieze, A., and Vera, J. 2006. A geometric preferential attachment model of networks. *Internet Math.*, **3**(2), 187–205.
- Flaxman, A., Frieze, A., and Vera, J. 2007. A geometric preferential attachment model of networks II. In: *Proceedings of WAW 2007*.
- Gao, P., and Wormald, N. 2016. Enumeration of graphs with a heavy-tailed degree sequence. *Adv. Math.*, **287**, 412–450.
- Gao, P., Hofstad, R. van der, Soutwell, A., and Stegehuis, C. 2018. *Counting triangles in power-law uniform random graphs*. Available at arXiv:1812.04289 [math.PR].

- Gilbert, E. N. 1959. Random graphs. *Ann. Math. Statist.*, **30**, 1141–1144.
- Gleiser, P., and Danon, L. 2003. Community Structure in Jazz. *Advances in Complex Systems*, **06**(04), 565–573.
- Godehardt, E., and Jaworski, J. 2003. Two models of random intersection graphs for classification. Pages 67–81 of: *Exploratory data analysis in empirical research*. Stud. Classification Data Anal. Knowledge Organ. Berlin: Springer.
- Goñi, J., Esteban, F., de Mendizábal, N., Sepulcre, J., Ardanza-Trevijano, S., Agirrezabal, I., and Villoslada, P. 2008. A computational analysis of protein-protein interaction networks in neurodegenerative diseases. *BMC Systems Biology*, **2**(1), 52.
- Grimmett, G. 1999. *Percolation*. 2nd edn. Berlin: Springer.
- Hall, P. 1981. Order of magnitude of moments of sums of random variables. *J. London Math Soc.*, **24**(2), 562–568.
- Halmos, P. 1950. *Measure Theory*. D. Van Nostrand Company, Inc., New York, N. Y.
- Hardy, G. H., Littlewood, J. E., and Pólya, G. 1988. *Inequalities*. Cambridge Mathematical Library. Cambridge: Cambridge University Press. Reprint of the 1952 edition.
- Harris, T. 1963. *The theory of branching processes*. Die Grundlehren der Mathematischen Wissenschaften, Bd. 119. Berlin: Springer-Verlag.
- Hatami, H., and Molloy, M. 2010. The scaling window for a random graph with a given degree sequence. Pages 1403–1411 of: *Proceedings of the Twenty-First Annual ACM-SIAM Symposium on Discrete Algorithms*. Society for Industrial and Applied Mathematics.
- Heydenreich, Markus, Hulshof, Tim, and Jorritsma, Joost. (2016). *Structures in supercritical scale-free percolation*. Preprint arXiv:1604.08180 [math.PR], to appear in *Ann. Appl. Probab.*
- Hirsch, C. Preprint (2014). *From heavy-tailed Boolean models to scale-free Gilbert graphs*.
- Hofstad, R. van der. 2017. *Random graphs and complex networks. Volume 1*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge.
- Hofstad, R. van der. 2018+. *Random graphs and complex networks. Vol. 2*. In preparation, see <http://www.win.tue.nl/~rhofstad/NotesRGCNII.pdf>.
- Hofstad, R. van der, and Komjáthy, J. 2017. When is a scale-free graph ultra-small? *J. Stat. Phys.*, **169**(2), 223–264.
- Hofstad, R. van der, and Litvak, N. 2014. Degree-degree dependencies in random graphs with heavy-tailed degrees. *Internet Math.*, **10**(3-4), 287–334.
- Hofstad, R. van der, Hooghiemstra, G., and Van Mieghem, P. 2005. Distances in random graphs with finite variance degrees. *Random Structures Algorithms*, **27**(1), 76–123.
- Hofstad, R. van der, Hooghiemstra, G., and Znamenski, D. 2007a. Distances in random graphs with finite mean and infinite variance degrees. *Electron. J. Probab.*, **12**(25), 703–766 (electronic).
- Hofstad, R. van der, Hooghiemstra, G., and Znamenski, D. 2007b. A phase transition for the diameter of the configuration model. *Internet Math.*, **4**(1), 113–128.
- Hofstad, R. van der, Leeuwaarden, J.H.S. van, and Stegehuis, C. Preprint 2015. *Hierarchical configuration model*.
- Hofstad, R. van der, Janson, S., and Luczak, M. Preprint 2016. *Component structure of the configuration model: barely supercritical case*. Available at <http://arxiv.org/pdf/1611.05728.pdf>.
- Howes, N. 1995. *Modern analysis and topology*. Universitext. Springer-Verlag, New York.
- Janson, S. 2008. The largest component in a subcritical random graph with a power law degree distribution. *Ann. Appl. Probab.*, **18**(4), 1651–1668.
- Janson, S. 2009a. On percolation in random graphs with given vertex degrees. *Electron. J. Probab.*, **14**, no. 5, 87–118.
- Janson, S. 2009b. Standard representation of multivariate functions on a general probability space. *Electron. Commun. Probab.*, **14**, 343–346.
- Janson, S. 2010a. Asymptotic equivalence and contiguity of some random graphs. *Random Structures Algorithms*, **36**(1), 26–45.

- Janson, S. 2010b. Susceptibility of random graphs with given vertex degrees. *J. Comb.*, **1**(3-4), 357–387.
- Janson, S., and Luczak, M. 2007. A simple solution to the k -core problem. *Random Structures Algorithms*, **30**(1-2), 50–62.
- Janson, S., and Luczak, M. 2009. A new approach to the giant component problem. *Random Structures Algorithms*, **34**(2), 197–216.
- Janson, S., Luczak, T., and Rucinski, A. 2000. *Random graphs*. Wiley-Interscience Series in Discrete Mathematics and Optimization. Wiley-Interscience, New York.
- Jaworski, J., Karoński, M., and Stark, D. 2006. The degree of a typical vertex in generalized random intersection graph models. *Discrete Math.*, **306**(18), 2152–2165.
- Joseph, Adrien. 2014. The component sizes of a critical random graph with given degree sequence. *Ann. Appl. Probab.*, **24**(6), 2560–2594.
- Kallenberg, O. 2002. *Foundations of modern probability*. Second edn. Probability and its Applications (New York). New York: Springer-Verlag.
- Karoński, M., Scheinerman, E., and Singer-Cohen, K. 1999. On random intersection graphs: the subgraph problem. *Combin. Probab. Comput.*, **8**(1-2), 131–159. Recent trends in combinatorics (Mátraháza, 1995).
- Kingman, J. F. C. 1975. The first birth problem for an age-dependent branching process. *Ann. Probability*, **3**(5), 790–801.
- Krioukov, D., Papadopoulos, F., Kitsak, M., Vahdat, A., and Boguñá, M. 2010. Hyperbolic geometry of complex networks. *Phys. Rev. E* (3), **82**(3), 036106, 18.
- Krioukov, D., Kitsak, M., Sinkovits, R., Rideout, D., Meyer, D., and Boguñá, M. 2012. Network cosmology. *Scientific reports*, **2**.
- Kunegis, J. 2017. *The Koblenz Network Collection*.
- Lee, J., and Olvera-Cravioto, M. 2017. *PageRank on inhomogeneous random digraphs*. Preprint arXiv:1707.02492 [math.PR].
- Leskelä, L. 2019. *Random graphs and network statistics*. Available at <http://math.aalto.fi/~lleskela/LectureNotes004.html>.
- Leskovec, J., and Krevl, A. 2014 (Jun). *SNAP Datasets: Stanford Large Network Dataset Collection*. <http://snap.stanford.edu/data>.
- Leskovec, J., Kleinberg, J., and Faloutsos, C. 2007. Graph evolution: Densification and shrinking diameters. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, **1**(1), 2.
- Leskovec, J., Lang, K., Dasgupta, A., and Mahoney, M. 2009. Community structure in large networks: Natural cluster sizes and the absence of large well-defined clusters. *Internet Mathematics*, **6**(1), 29–123.
- Litvak, N., and Hofstad, R. van der. 2013. Uncovering disassortativity in large scale-free networks. *Physical Review E*, **87**(2), 022801.
- Luczak, T. 1992. Sparse random graphs with a given degree sequence. Pages 165–182 of: *Random graphs, Vol. 2 (Poznań, 1989)*. Wiley-Intersci. Publ. Wiley, New York.
- McKay, B.D. 1981. Subgraphs of random graphs with specified degrees. *Congressus Numerantium*, **33**, 213–223.
- McKay, B.D. 2011. Subgraphs of Random Graphs with Specified Degrees. In: *Proceedings of the International Congress of Mathematicians 2010 (ICM 2010)*. Published by Hindustan Book Agency (HBA), India. WSPC Distribute for All Markets Except in India.
- McKay, Brendan D, and Wormald, Nicholas C. 1990. Asymptotic enumeration by degree sequence of graphs of high degree. *European Journal of Combinatorics*, **11**(6), 565–580.
- Molloy, M., and Reed, B. 1995. A critical point for random graphs with a given degree sequence. *Random Structures Algorithms*, **6**(2-3), 161–179.
- Molloy, M., and Reed, B. 1998. The size of the giant component of a random graph with a given degree sequence. *Combin. Probab. Comput.*, **7**(3), 295–305.
- Moore, C., and Newman, M. E. J. 2000. Epidemics and percolation in small-world networks. *Phys. Rev. E*, **61**, 5678–5682.

- Nachmias, A., and Peres, Y. 2010. Critical percolation on random regular graphs. *Random Structures Algorithms*, **36**(2), 111–148.
- Newman, M. E. J. 2003. Properties of highly clustered networks. *Physical Review E*, **68**(2), 026121.
- Newman, M. E. J. 2009. Random Graphs with Clustering. *Phys. Rev. Lett.*, **103**(Jul), 058701.
- Newman, M. E. J. 2010. *Networks: an introduction*. Oxford University Press.
- Newman, M. E. J., and Park, J. 2003. Why social networks are different from other types of networks. *Physical Review E*, **68**(3), 036122.
- Newman, M. E. J., and Watts, D.J. 1999. Scaling and percolation in the small-world network model. *Phys. Rev. E*, **60**, 7332–7344.
- Newman, M. E. J., Moore, C., and Watts, D. J. 2000a. Mean-field solution of the small-world network model. *Phys. Rev. Lett.*, **84**, 3201–3204.
- Newman, M. E. J., Strogatz, S., and Watts, D. 2000b. Random graphs with arbitrary degree distribution and their application. *Phys. Rev. E*, **64**, 026118, 1–17.
- Newman, M. E. J., Strogatz, S., and Watts, D. 2002. Random graph models of social networks. *Proc. Nat. Acad. Sci.*, **99**, 2566–2572.
- Newman, M. E. J., Watts, D. J., and Barabási, A.-L. 2006. *The Structure and Dynamics of Networks*. Princeton Studies in Complexity. Princeton University Press.
- Norros, I., and Reittu, H. 2006. On a conditionally Poissonian graph process. *Adv. in Appl. Probab.*, **38**(1), 59–75.
- Pemantle, R. 2007. A survey of random processes with reinforcement. *Probab. Surv.*, **4**, 1–79 (electronic).
- Pickands III, J. 1968. Moment convergence of sample extremes. *Ann. Math. Statistics*, **39**, 881–889.
- Pittel, B. 1994. Note on the heights of random recursive trees and random m -ary search trees. *Random Structures Algorithms*, **5**(2), 337–347.
- Reittu, H., and Norros, I. 2004. On the power law random graph model of massive data networks. *Performance Evaluation*, **55**(1-2), 3–23.
- Riordan, O. 2012. The phase transition in the configuration model. *Combinatorics, Probability and Computing*, **21**, 1–35.
- Riordan, O., and Wormald, N. 2010. The diameter of sparse random graphs. *Combinatorics, Probability & Computing*, **19**(5-6), 835–926.
- Ross, S. M. 1996. *Stochastic processes*. Second edn. Wiley Series in Probability and Statistics: Probability and Statistics. New York: John Wiley & Sons Inc.
- Rudas, A., Tóth, B., and Valkó, B. 2007. Random trees and general branching processes. *Random Structures Algorithms*, **31**(2), 186–202.
- Rudin, W. 1987. *Real and complex analysis*. New York: McGraw-Hill Book Co.
- Rudin, W. 1991. *Functional analysis*. International Series in Pure and Applied Mathematics. New York: McGraw-Hill Inc.
- Schuh, H.-J., and Barbour, A. D. 1977. On the asymptotic behaviour of branching processes with infinite mean. *Advances in Appl. Probability*, **9**(4), 681–723.
- Seneta, E. 1973. The simple branching process with infinite mean. I. *J. Appl. Probability*, **10**, 206–212.
- Singer, K. 1996. *Random intersection graphs*. ProQuest LLC, Ann Arbor, MI. Thesis (Ph.D.)–The Johns Hopkins University.
- Smythe, R., and Mahmoud, H. 1994. A survey of recursive trees. *Teor. ĭmovĭr. Mat. Stat.*, 1–29.
- Söderberg, B. 2002. General formalism for inhomogeneous random graphs. *Phys. Rev. E* (3), **66**(6), 066121, 6.
- Söderberg, B. 2003a. Properties of random graphs with hidden color. *Phys. Rev. E* (3), **68**(2), 026107, 12.
- Söderberg, B. 2003b. Random graph models with hidden color. *Acta Physica Polonica B*, **34**, 5085–5102.
- Söderberg, B. 2003c. Random graphs with hidden color. *Phys. Rev. E* (3), **68**(1), 015102, 4.

- Solomonoff, R., and Rapoport, A. 1951. Connectivity of random nets. *Bull. Math. Biophys.*, **13**, 107–117.
- Stark, D. 2004. The vertex degree distribution of random intersection graphs. *Random Structures Algorithms*, **24**(3), 249–258.
- Stegehuis, C., Hofstad, R. van der, and Leeuwaarden, J.S.H van. 2016a. Epidemic spreading on complex networks with community structures. *Scientific Reports*, **6**, 29748.
- Stegehuis, C., Hofstad, R. van der, and Leeuwaarden, J.S.H van. 2016b. Power-law relations in random networks with communities. *Physical Review E*, **94**, 012302.
- Sundaresan, S., Fischhoff, I., Dushoff, J., and Rubenstein, D. 2007. Network metrics reveal differences in social organization between two fission–fusion species, Grevy’s zebra and onager. *Oecologia*, **151**(1), 140–149.
- Turova, T. 2013. Diffusion approximation for the components in critical inhomogeneous random graphs of rank 1. *Random Structures Algorithms*, **43**(4), 486–539.
- Turova, T.S. 2011. The largest component in subcritical inhomogeneous random graphs. *Combinatorics, Probability and Computing*, **20**(01), 131–154.
- Turova, T.S., and Vallier, T. 2010. Merging percolation on \mathbf{Z}^d and classical random graphs: Phase transition. *Random Structures Algorithms*, **36**(2), 185–217.
- Ugander, J., Karrer, B., Backstrom, L., and Marlow, C. Preprint (2011). The anatomy of the Facebook social graph. Available from <http://arxiv.org/pdf/1111.4503.pdf>.
- Watts, D. J. 1999. *Small worlds. The dynamics of networks between order and randomness*. Princeton Studies in Complexity. Princeton, NJ: Princeton University Press.
- Watts, D. J. 2003. *Six degrees. The science of a connected age*. New York: W. W. Norton & Co. Inc.
- Watts, D. J., and Strogatz, S. H. 1998. Collective dynamics of ‘small-world’ networks. *Nature*, **393**, 440–442.
- Wormald, N. C. 1981. The asymptotic connectivity of labelled regular graphs. *J. Combin. Theory Ser. B*, **31**(2), 156–167.
- Yukich, J. E. 2006. Ultra-small scale-free geometric networks. *J. Appl. Probab.*, **43**(3), 665–677.

INDEX

- Albert-Barábasi model, 161
- Branching process
 - delayed, 229
 - infinite mean, 220
 - marked, 81
 - multitype, 69
 - two-stage, 81
- Central limit theorem, 168
- Chung-Lu model
 - attack vulnerability, 88
 - fluctuations of typical distance, 203
- Clique, 187
- Collaboration graph
 - random intersection graph, 289
- Complexity, 86
- Configuration model, 97, 209
 - average distance, 209
 - clustering, 286, 287
 - connectivity, 120
 - core, 216
 - diameter, 225
 - diameter core, 216
 - directed, 278
 - distance, 209
 - forward degree distribution, 98
 - household structure, 286
 - near-critical behavior, 126
 - path counting, 211
 - phase transition, 103
- Connected component
 - complexity, 86
- Connectivity
 - configuration model, 120
- De Finetti's Theorem, 134
- Death process, 108
- Distribution
 - Bernoulli, 134
 - Mixed Poisson, 64
- Erdős-Rényi random graph
 - diameter, 226
- Exchangeable random variables, 134
- Helly's Theorem, 136
- Hopcount, 8
- Household model, 286
- Inhomogeneous random graph, 59
 - diameter, 202
 - fluctuations of typical distance, 203
 - graphical, 60
 - irreducible, 60
 - small world, 174
 - sparsity, 60
 - typical graph distance, 174, 203
 - Ultra-small world, 176
- Inhomogeneous random graphs
 - dual kernel, 202
 - duality principle, 202
- Internet, 8
 - hopcount, 8
- Internet Movie Data base
 - random intersection graph, 289
- Maclaurin inequality, 213
- Mixed marked Poisson branching process, 82
- Mixed Poisson distribution, 64
- Multitype branching process, 69
 - irreducible, 70
 - Markov chain, 70
 - singular, 70
- Pólya urn scheme
 - limit theorem, 136, 137
- Path
 - self-avoiding, 178
- Path counting
 - configuration model, 211
- Paths, 178
 - self-avoiding, 178
- percolation
 - long-range, 296
- Perron-Frobenius theorem, 70
- Phase transition
 - configuration model, 103
- Preferential attachment model, 133, 235
 - diameter, 244
 - typical distances, 244
- Random graph
 - typical graph distance, 174
- Random intersection graph, 288
- Random intersection graphs, 288
- Random variables
 - exchangeable, 134
- Real-world network, 3
- scale-free percolation network, 294
- Scale-free trees, 235
- Self-avoiding path, 178
- Size-biased distribution, 82
- Small world, 174
- spatial preferential attachment models, 298
- Stochastic domination, 82
- Theorem
 - De Finetti, 134
 - Helly, 136
 - Perron-Frobenius, 70
- Tree
 - height, 236

- Two-regular graph
 - diameter, [233](#)
 - longest cycle, [233](#)
- Ultra-small world, [174](#), [176](#)
- Uniform recursive trees, [236](#)
- Universality, [245](#)

TODO LIST

TO DO 1.1: Refer to recent discussion on power-laws in real-world networks.	6
TO DO 5.1: Add local weak convergence for BPAM	187
TO DO 5.2: Extend the discussion of the phase transition for BPAM	190
TO DO 5.3: Add further results!	190
TO DO 6.1: Add case of distances in real-world network...	199
Exercise??	284
Update from here!	285
Add Discussion of $\text{dist}_{\text{PA}_n^{(1,\delta)}(b)}(1, V) = o_{\mathbb{P}}(\log n)$ here! Or exercise?	294
TO DO 8.2: Adapt this argument as in ?!	297
TO DO 8.3: Check whether equality is okay??	300
TO DO 8.4: Continue from here!	302
TO DO 8.5: Add proof!	308
TO DO 8.6: Extend the discussion in this section.	309
TO DO 8.7: Explain that Theorem 8.23 is already proved	309
TO DO 8.9: Add further result, such as the ones in the BPAM	309
TO DO 8.10: Extend notes and discussion	309
TO DO 8.11: Design more exercises	309
TO DO 9.1: Explain directed networks from a practical setting (WWW?)	315
TO DO 9.3: Add discussion directed inhomogeneous random graphs	317
TO DO 9.3: Add discussion distances in directed random graphs	320
TO DO 9.4: Add practical example of community structure: clustering table?	321
TO DO 9.5: Add discussion community structure: partition, SBM, larger overlap	321
TO DO 9.6: Add discussion SBM	321
TO DO 9.7: Make this introduction section shorter	321
TO DO 9.8: Present results HCM: connectivity and graph distances	326
TO DO 9.9: Adapt this discussion: special case of HCM	326
TO DO 9.10: Extend discussion of intersection graphs	328
TO DO 9.11: Add discussion of exponential random graphs, and show that GRG is special case	329
TO DO 9.12: Add discussion why geometry may be useful in random graphs	329
TO DO 9.13: Introduce the GIRG	331
TO DO 9.14: Introduce the spatial configuration model and results of Deijfen et al.	338
TO DO 9.16: Extend the notes and discussion section.	340
TO DO 9.17: Design more exercises for Chapter 9.	341