

Stochastic processes on random graphs

Remco van der Hofstad

July 12, 2017

2010 Mathematical Subject Classification: 60K35, 60K37, 82B43

Keywords: Real-world networks, random graphs, local weak convergence, scale-free behavior, small worlds, percolation, first-passage percolation, competition processes, Ising model, random walks, phase transitions, critical behavior, coexistence, mixing times, cover times.

DEPARTMENT OF MATHEMATICS AND COMPUTER SCIENCE, EINDHOVEN UNIVERSITY OF TECHNOLOGY, P.O. Box 513, 5600 MB EINDHOVEN, THE NETHERLANDS.
Email: r.w.v.d.hofstad@tue.nl

Preface

In this book, we discuss *stochastic processes on random graphs*. The understanding of such processes is interesting from an applied perspective, since random graphs serve as models for real-world networks, while stochastic processes on them aim to model network functionality. The mathematics of stochastic processes on random graphs is also highly interesting, for one due to the *double randomness* that is present. Indeed, the stochastic process lives on a random structure, namely, the random graph itself. This gives rise to an intricate interplay between the random geometry of the graph and the behavior of the stochastic process on it.

This book is to a large extent based on joint work with others. Chapter 3 on first-passage percolation or routing on random graphs is based on joint work with Enrico Baroni, Shankar Bhamidi, Júlia Komjáthy and Gerard Hooghiemstra. Chapter 4 on percolation on random graphs is based on joint work with Shankar Bhamidi, Souvik Dhara, Johan van Leeuwen, Sanchayan Sen and Clara Stegehuis. Chapter 5 on Ising models on random graphs is based on joint work with Sander Dommers, Cristian Giardinà, Claudi Giberti and Maria-Luiso Prioriello. Section 6.1 in Chapter 6 on competition on random graphs is based on joint work with Enrico Baroni, Mia Deijfen and Júlia Komjáthy. Research is a team effort, and it has been a great pleasure to work with such an extended group of highly talented researchers, as well as wonderful people!

This work would not have been possible without the generous support of various institutions. The work of RvdH is supported by the Netherlands Organisation for Scientific Research (NWO) through VICI grant 639.033.806 and the Gravitation NETWORKS grant 024.002.003. This work has been written as lecture notes for the Saint Flour Summer School in Probability 2017, held at Saint Flour. RvdH warmly thanks the organisers Christophe Bahadoran, Arnaud Guillin and Laurent Serlet.

Finally, my family deserves an enormous thanks. I have often been physically, and sometimes also mentally, away while writing this text. Mad, Max, and Lars, you kept me grounded, and aware that there is something more that is worth spending time on. Mad, thanks for your support. I love you.

This is a preliminary version of this book, and I am grateful for any comments that we receive before finalizing it.

Contents

Preface	v
Chapter 1. Introduction and motivation	1
1.1. Motivation: Real-world networks	1
1.2. Random graphs and real-world networks	6
1.3. Random graph models	7
1.4. Related geometries and universality	24
1.5. Stochastic processes on random graphs and network functionality	25
1.6. Organization and notation	26
Chapter 2. Preliminaries	29
2.1. Branching processes	29
2.2. Local weak convergence of random graphs	35
2.3. The giant component in random graphs	51
2.4. Connectivity of random graphs	57
2.5. Small-world random graphs	59
2.6. Further technical ingredients	69
2.7. Notes and discussion for Chapter 2	74
Chapter 3. First passage percolation on random graphs	77
3.1. First passage percolation on general graphs and motivation	77
3.2. Fixing ideas: first-passage percolation on the complete graph	80
3.3. Markovian flow on the configuration model	94
3.4. The trick of the trade: Continuous-time branching processes	103
3.5. Universality for first-passage percolation on configuration models	116
3.6. First-passage percolation on scale-free configuration models	128
3.7. Further results for first-passage percolation	136
3.8. Notes and discussion of first-passage percolation on random graphs	145
Chapter 4. Percolation on random graphs	149
4.1. Percolation on general graphs and motivation	149
4.2. Fixing ideas: complete graph and Erdős-Rényi	155
4.3. Percolation on the configuration model: Janson's construction	165
4.4. Critical percolation on CMs with finite third-moment degrees	173
4.5. Percolation on configuration models with infinite third-moment degrees	197
4.6. Metric space convergence of critical percolation on configuration models	207
4.7. Percolation on the PAM with conditionally independent edges	211
4.8. Further results on percolation on random graphs	213
4.9. Notes and discussion on percolation on random graphs	224
Chapter 5. Ising models on random graphs	227

5.1.	The Ising model on general graphs and motivation	227
5.2.	Fixing ideas: complete graph and Curie-Weiss model	233
5.3.	Thermodynamic limits on the configuration model	243
5.4.	Critical temperature on the Ising model	261
5.5.	Ising critical exponents on locally-tree like random graphs	264
5.6.	Central limit theorem for the non-critical total spin	279
5.7.	Non-classical limit theorems for the critical total spin	282
5.8.	Further results on Ising models on random graphs	304
5.9.	Notes and discussion of Ising models on random graphs	307
Chapter 6.	Related stochastic processes on random graphs	309
6.1.	Competition processes	309
6.2.	Contact process	320
6.3.	Bootstrap percolation	320
	Bibliography	321

CHAPTER 1

Introduction and motivation

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 [160], both for the motivation as well as for the introduction of the random graph models involved.

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 webpages 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 [25] and Watts [249] for expository accounts of the discovery of network properties by Barabási, Watts and co-authors. In [223], you can find 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 [222] lists many of the empirical properties of, and scientific methods for, networks. See also [160, 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.,

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

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

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

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

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

EXERCISE 1.1 (Probability mass function typical degree). *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, 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 ,

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

then

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

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

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

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 [130] (see Figure 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 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 3 displays the degree-sequence for both the in- as well as the out-degrees in various World-Wide Web data bases.

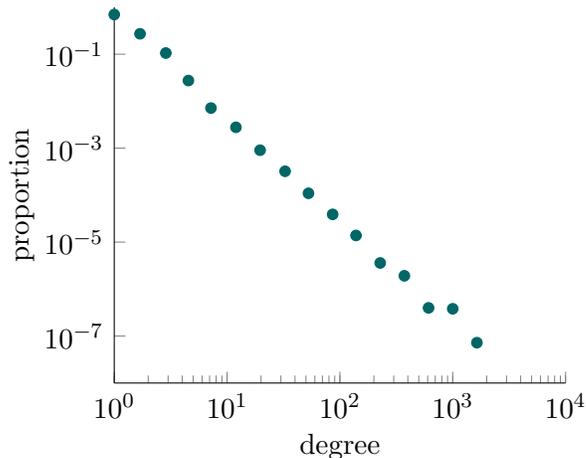


FIGURE 1. (a) Log-log plot of the probability mass function of the degree sequence of Autonomous Systems (AS) on April 2014 on a log-log scale from [202] (data courtesy of Dmitri Krioukov).

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 [248]). 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 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 4(b), the proportion of routers traversed by an e-mail message between two uniformly chosen routers, referred to as the *hopcount*, is shown. It shows that the number of routers traversed is at most 27, while the distribution resembles a Poisson probability mass function. Figure 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

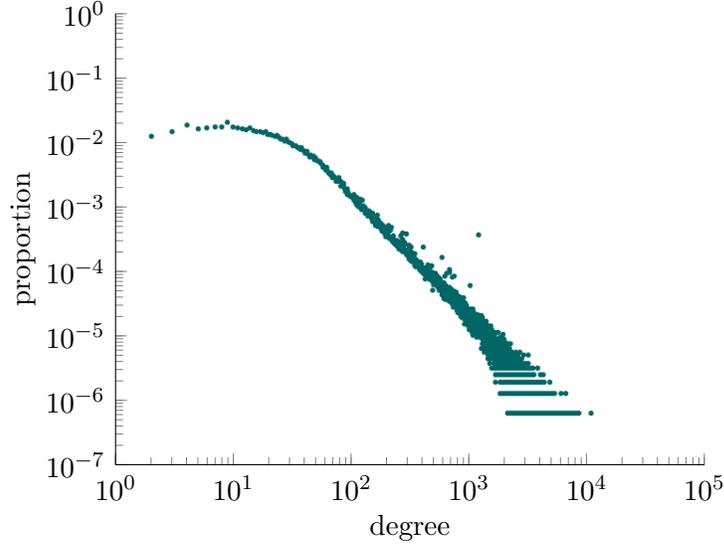


FIGURE 2. Log-log plot of the degree sequence in the Internet Movie Data base in 2007.

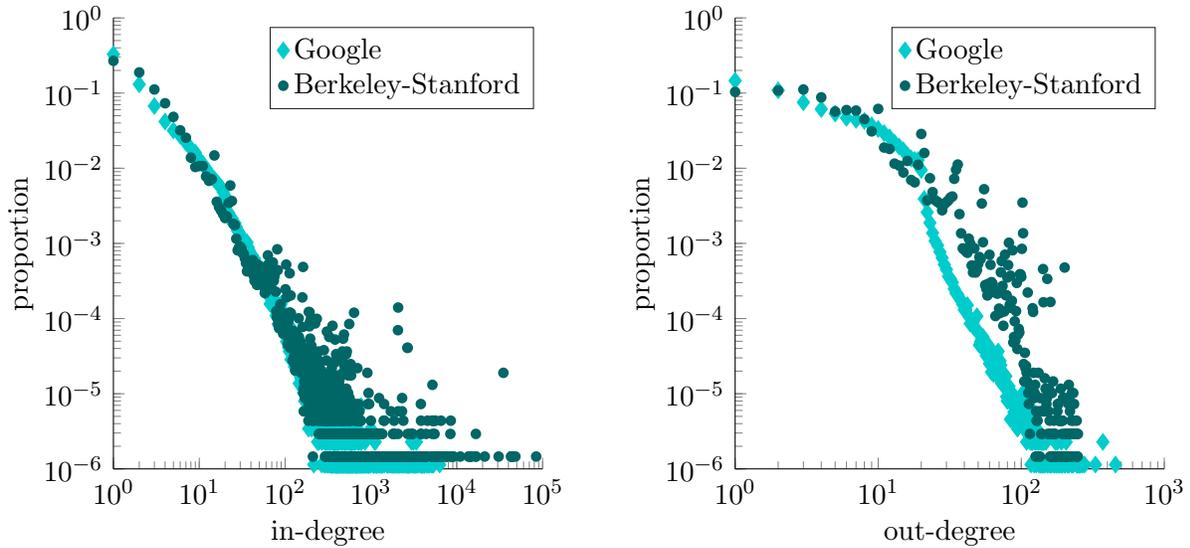


FIGURE 3. The probability mass function of the in- and out-degree sequences in the Berkeley-Stanford and Google competition graph data sets of the WWW in [204]. (a) in-degree; (b) out-degree.

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.

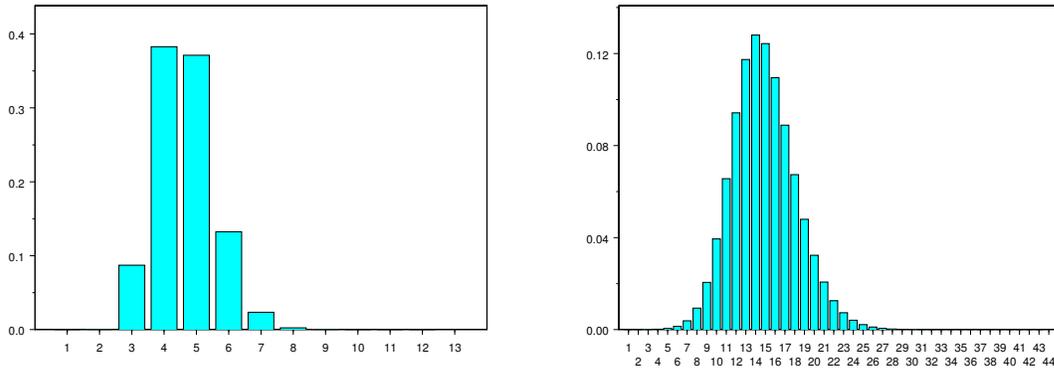


FIGURE 4. (a) Proportion of AS traversed in hopcount data. (b) Internet hopcount data. Courtesy of Hongsuda Tangmunarunkit.

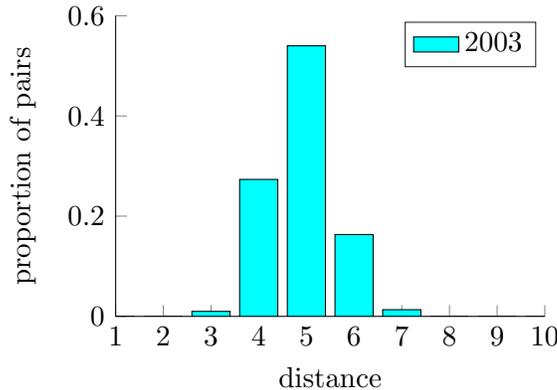


FIGURE 5. Typical distances in the Internet Movie Data base in 2007.

Let us continue by introducing graph distances, as displayed in Figures 4–5, 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 U_1 and U_2 uniformly at random from $[n]$, and we investigate the random variable

$$(1.1.7) \quad \text{dist}_G(U_1, U_2).$$

The quantity in (1.1.7) is a random variable even for *deterministic* graphs due to the occurrence of the two, uniformly at randomly, vertices $U_1, U_2 \in [n]$. Figures 4–5 display the probability mass function of this random variable for some real-world networks. Often, we will consider $\text{dist}_G(U_1, U_2)$ conditionally on $\text{dist}_G(U_1, U_2) < \infty$. This means that we consider the typical number of edges between a uniformly chosen pair of *connected* vertices. As a result, $\text{dist}_G(U_1, U_2)$ is sometimes referred to as the *typical distance*.

The nice property of $\text{dist}_G(U_1, U_2)$ 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

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

However, the diameter has several disadvantages. First, in many instances, the diameter is algorithmically more difficult to compute than the typical distances (since one has to measure the distances between all pairs of vertices and maximize over them). Second, it is a number instead of the distribution of a random variable, and therefore contains far less information than the distribution of $\text{dist}_G(U_1, U_2)$. 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

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

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

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

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

$$(1.2.3) \quad \lim_{n \rightarrow \infty} \mathbb{P}(\text{dist}_G(U_1, U_2) \leq K \log n) = 1.$$

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., [160, 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 [222] 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

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 $\text{ER}_n(p)$.

The model is named after Erdős and Rényi, since they have made profound contributions in the study of this model. See, in particular, [122, 123, 124, 125], 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 [139], and was already investigated heuristically by Solomon and Rapoport in [242]. 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 (Uniform random graph). *Consider $\text{ER}_n(p)$ 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.*

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

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

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

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

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

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

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

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

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 [160, Section 5.4]), i.e., for every $k \geq 0$,

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

It is well known that the Poisson distribution has very thin tails, even thinner than any exponential:

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,*

$$(1.3.6) \quad \lim_{k \rightarrow \infty} e^{\alpha k} p_k = 0.$$

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. This perspective will be convenient when studying percolation on random graphs in Chapter 4. Secondly, the model described here as the Erdős-Rényi random graph was actually *not* invented by Erdős and Rényi, but rather by Gilbert [139]. Erdős and Rényi [122, 123, 125], 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, Janson and Riordan [56]. We will discuss such general inhomogeneous random graphs in more detail below. We start with one key example, that has attracted the most attention in the literature so far.

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, Deijfen and Martin-Löf [68].

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

$$(1.3.7) \quad p_{ij} = p_{ij}^{(\text{GRG})} = \frac{w_i w_j}{\ell_n + w_i w_j},$$

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

$$(1.3.8) \quad \ell_n = \sum_{i \in [n]} w_i.$$

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

$$(1.3.9) \quad p_{ij} = p_{ij}^{(\text{CL})} = \min(w_i w_j / \ell_n, 1),$$

and which has been studied intensively by Chung and Lu (see [79, 80, 81, 82, 83]). A further adaptation is the so-called *Poissonian random graph* or Norros-Reittu model [226], for which

$$(1.3.10) \quad p_{ij} = p_{ij}^{(\text{NR})} = 1 - \exp(-w_i w_j / \ell_n).$$

See Janson [182] or [160, 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

$$(1.3.11) \quad F_n(x) = \frac{1}{n} \sum_{i \in [n]} \mathbb{1}_{\{w_i \leq x\}}, \quad x \geq 0.$$

We can interpret F_n as the distribution of the weight of a uniformly chosen vertex in $[n]$ (see Exercise 1.4). We denote the weight of a uniformly chosen vertex U in $[n]$ by $W_n = w_U$, so that, by Exercise 1.4, W_n has distribution function F_n .

EXERCISE 1.4 (Weight of uniformly chosen vertex). *Let U be a vertex chosen uniformly at random from $[n]$. Show that the weight w_U of U 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 $n \rightarrow \infty$,*

$$(1.3.12) \quad W_n \xrightarrow{d} W,$$

where W_n and W have distribution functions F_n and F , respectively. Equivalently, for any x for which $x \mapsto F(x)$ is continuous,

$$(1.3.13) \quad \lim_{n \rightarrow \infty} F_n(x) = F(x).$$

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

$$(1.3.14) \quad \mathbb{E}[W_n] \rightarrow \mathbb{E}[W],$$

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 $n \rightarrow \infty$,

$$(1.3.15) \quad \mathbb{E}[W_n^2] \rightarrow \mathbb{E}[W^2].$$

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

$$(1.3.16) \quad w_i = [1 - F]^{-1}(i/n),$$

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

$$(1.3.17) \quad [1 - F]^{-1}(u) = \inf\{x : [1 - F](x) \leq u\}.$$

For the choice in (1.3.16), we can explicitly compute F_n as

$$(1.3.18) \quad \begin{aligned} F_n(x) &= \frac{1}{n} \sum_{i \in [n]} \mathbb{1}_{\{w_i \leq x\}} = \frac{1}{n} \sum_{i \in [n]} \mathbb{1}_{\{[1-F]^{-1}(i/n) \leq x\}} = \frac{1}{n} \sum_{j=0}^{n-1} \mathbb{1}_{\{[1-F]^{-1}(1-\frac{j}{n}) \leq x\}} \\ &= \frac{1}{n} \sum_{j=0}^{n-1} \mathbb{1}_{\{F^{-1}(\frac{j}{n}) \leq x\}} = \frac{1}{n} \sum_{j=0}^{n-1} \mathbb{1}_{\{\frac{j}{n} \leq F(x)\}} = \frac{1}{n} (\lfloor nF(x) \rfloor + 1) \wedge 1, \end{aligned}$$

where we write $j = n - i$ in the third equality. It is not hard to see that Condition 1.1(a) holds for $(w_i)_{i \in [n]}$ as in (1.3.16), 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 the following exercise:

EXERCISE 1.5 (Domination weights). *Let W_n have distribution function F_n from (1.3.18). Show that W_n is stochastically dominated by the random variable W having distribution function F .*

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 (1.3.16), 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

$$(1.3.19) \quad d_i = \sum_{j \in [n]} X_{ij},$$

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 (1.3.7).

For $k \geq 0$, we let

$$(1.3.20) \quad P_k^{(n)} = \frac{1}{n} \sum_{i \in [n]} \mathbb{1}_{\{D_i = k\}}$$

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 $k \geq 0$,

$$(1.3.21) \quad p_k = \mathbb{E} \left[e^{-W} \frac{W^k}{k!} \right],$$

where W is a random variable having distribution function F from Condition 1.1. The main result concerning the vertex degrees, which is [160, 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 $\varepsilon > 0$,*

$$(1.3.22) \quad \mathbb{P} \left(\sum_{k=0}^{\infty} |P_k^{(n)} - p_k| \geq \varepsilon \right) \rightarrow 0,$$

where $(p_k)_{k \geq 0}$ is given by (1.3.21).

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

$$(1.3.23) \quad D_n \xrightarrow{d} D,$$

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

EXERCISE 1.6 (Degree of uniformly chosen vertex in $\text{GRG}_n(\mathbf{w})$). *Prove that, under the conditions of Theorem 1.3, (1.3.23) holds.*

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,

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

and then also, for w large,

$$(1.3.25) \quad \mathbb{P}(D > w) = \frac{c}{w^{\tau-1}}(1 + o(1)).$$

EXERCISE 1.7 (Power-law degrees in generalized random graphs). *Prove that, under the conditions of Theorem 1.3, (1.3.25) follows from (1.3.24). Does the converse also hold?*

Generalized random graph conditioned on its degrees. The generalized random graph with its edge probabilities as in (1.3.7) 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 (1.3.7). 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 $i \in [n]$,

$$(1.3.26) \quad \mathbb{P}(X = x \mid d_i(X) = d_i \forall i \in [n]) = \frac{1}{\#\{y: d_i(y) = d_i \forall i \in [n]\}},$$

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

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

$$(1.3.27) \quad p_{ij} = \frac{w_i w_j}{\ell_n + w_i w_j},$$

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

PROOF. This is [160, Theorem 6.15]. □

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 1.6(a)-(c):

THEOREM 1.5 (Regularity conditions weights and degrees). *Let d_i be the degree of vertex i in $\text{GRG}_n(\mathbf{w})$, and let $\mathbf{d} = (d_i)_{i \in [n]}$. Then, \mathbf{d} satisfies Conditions 1.6(a)-(b) in probability when \mathbf{w} satisfies Conditions 1.1(a)-(b), where*

$$(1.3.28) \quad \mathbb{P}(D = k) = \mathbb{E} \left[\frac{W^k}{k!} e^{-W} \right]$$

denotes the mixed-Poisson distribution with mixing distribution W having distribution function F in Condition 1.1(a). Further, \mathbf{d} satisfies Conditions 1.6(a)-(c) in probability when \mathbf{w} satisfies Conditions 1.1(a)-(c).

PROOF. This is [160, Theorem 7.19]. The weak convergence in Condition 1.6(a) is Theorem 1.3. \square

Theorem 1.5 allows us to prove many results for the generalized random graph by first proving them for the *configuration model*, as discussed in more detail in Section 1.3.3, and then extending them to the generalized random graph. See [160, Sections 6.6 and 7.5] for more details. This will prove to be a convenient proof strategy to deduce results for the GRG from those for the configuration model.

Related inhomogeneous random graph models. Bollobás, Janson and Riordan [56] study inhomogeneous random graphs in great detail, focussing on their degree structure, small-world nature, the existence of a giant component and various other graph properties. See also [161, Chapters 2 and 3]. The general set-up is as follows. The edge probabilities are replaced by

$$(1.3.29) \quad p_{ij} = \frac{\kappa_{i,j}}{n},$$

the factor $1/n$ making sure that the model is sparse. Often $\kappa_{i,j}$ is chosen as $\kappa_{i,j} = \kappa(x_i, x_j)$, where $(x_i)_{i \in [n]}$ are i.i.d. uniform random variables and $(x, y) \mapsto \kappa(x, y)$ is a sufficiently nice function mapping $[0, 1]^2$ to $[0, \infty)$. Of course, when the kernel $(\kappa_{i,j})_{1 \leq i < j \leq n}$ is highly variable, $\kappa_{i,j}/n$ might be larger than 1. In this case, again there are several ways to resolve this issue, for example by taking instead $p_{ij} = \min(\kappa_{i,j}/n, 1)$, or rather $p_{ij} = 1 - e^{-\kappa_{i,j}/n}$, etc. Often, this makes little difference.

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

$$(1.3.30) \quad \ell_n = \sum_{j \in [n]} d_j$$

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:

EXERCISE 1.8 (Simple graphs). *Find a simple example of a sequence $(d_i)_{i \in [n]}$ for which $\ell_n = \sum_{j \in [n]} d_j$ is even, but there is no simple graph having $(d_i)_{i \in [n]}$ as a degree sequence.*

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., [160, Definition 7.5 and Lemma 7.6] for more details on this exchangeability. Interestingly, one can compute 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 $(X_{ij})_{1 \leq i \leq j \leq n}$, where X_{ij} is the number of edges between vertex i and j . Here X_{ii} is the number of self-loops incident to vertex i , and

$$(1.3.31) \quad d_i = X_{ii} + \sum_{j \in [n]} X_{ij}$$

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

$$(1.3.32) \quad \mathbb{P}(\text{CM}_n(\mathbf{d}) = G) = \frac{1}{(\ell_n - 1)!!} \frac{\prod_{i \in [n]} d_i!}{\prod_{i \in [n]} 2^{x_{ii}} \prod_{1 \leq i < j \leq n} x_{ij}!}.$$

See e.g., [160, 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 CM 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 in [48] to study uniform random regular graphs (see also [52, Section 2.4]). The introduction was inspired by, and generalized the results in, the work of Bender and Canfield [31]. 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 in [215, 216]. 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 U in $[n]$ by $D_n = d_U$. The random variable D_n has distribution function F_n given by

$$(1.3.33) \quad F_n(x) = \frac{1}{n} \sum_{j \in [n]} \mathbb{1}_{\{d_j \leq x\}},$$

which is the *empirical distribution of the degrees*. We assume that the vertex degrees satisfy the following *regularity conditions*:

CONDITION 1.6 (Regularity conditions for vertex degrees).

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

$$(1.3.34) \quad D_n \xrightarrow{d} D,$$

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

$$(1.3.35) \quad \lim_{n \rightarrow \infty} F_n(x) = F(x).$$

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

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

$$(1.3.36) \quad \mathbb{E}[D_n] \rightarrow \mathbb{E}[D],$$

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

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

$$(1.3.37) \quad \mathbb{E}[D_n^2] \rightarrow \mathbb{E}[D^2],$$

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

The possibility to obtain a non-simple graph is a major disadvantage of the CM. 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 $\text{CM}_n(\mathbf{d}) = (X_{ij})_{1 \leq i \leq j \leq n}$ by its erased version $\text{ECM}_n(\mathbf{d}) = (X_{ij}^{(\text{er})})_{1 \leq i \leq j \leq n}$, 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, [160, 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 [160, 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

$$(1.3.38) \quad D_i^{(\text{er})} = d_i - 2s_i - m_i,$$

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

$$(1.3.39) \quad m_i = \sum_{j \neq i} (x_{ij} - 1) \mathbb{1}_{\{x_{ij} \geq 2\}}$$

is the number of multiple edges removed from i .

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

$$(1.3.40) \quad p_k^{(n)} = \mathbb{P}(D_n = k) = \frac{1}{n} \sum_{i \in [n]} \mathbb{1}_{\{d_i = k\}},$$

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

$$(1.3.41) \quad P_k^{(\text{er})} = \frac{1}{n} \sum_{i \in [n]} \mathbb{1}_{\{D_i^{(\text{er})} = k\}}.$$

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 1.7 (Degree sequence of erased configuration model with fixed degrees). *For fixed degrees \mathbf{d} satisfying Conditions 1.6(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 $\varepsilon > 0$,*

$$(1.3.42) \quad \mathbb{P}\left(\sum_{k=1}^{\infty} |P_k^{(\text{er})} - p_k| \geq \varepsilon\right) \rightarrow 0,$$

where $p_k = \mathbb{P}(D = k)$ as in Condition 1.6(a).

Theorem 1.7 indeed shows that most of the edges are kept in the erasure procedure:

EXERCISE 1.9 (Number of erased edges). *Assume that Conditions 1.6(a)-(b) hold. Show that Theorem 1.7 implies that the number of erased edges is $o_{\mathbb{P}}(n)$.*

(b) Configuration model conditioned on simplicity. The second solution to the multigraph problem of the CM 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 [68]). It turns out that, when Conditions 1.6(a)-(c) hold, then (see [160, Theorem 7.12])

$$(1.3.43) \quad \lim_{n \rightarrow \infty} \mathbb{P}(\text{CM}_n(\mathbf{d}) \text{ is a simple graph}) = e^{-\nu/2 - \nu^2/4},$$

where

$$(1.3.44) \quad \nu = \frac{\mathbb{E}[D(D-1)]}{\mathbb{E}[D]}$$

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 (1.3.32), $\text{CM}_n(\mathbf{d})$ conditioned on simplicity is a *uniform random graph* with the prescribed degree sequence. We will return to the difficulty of generating simple graphs with infinite-variance degrees in the next chapter.

Relation GRG and CM. Since $\text{CM}_n(\mathbf{d})$ conditioned on simplicity yields a uniform (simple) random graph with these degrees, and by (1.3.26), also $\text{GRG}_n(\mathbf{w})$ 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 $\text{GRG}_n(\mathbf{w})$ by proving them for $\text{CM}_n(\mathbf{d})$ under the appropriate degree conditions, and then proving that $\text{GRG}_n(\mathbf{w})$ satisfies these conditions in probability. See [160, Section 7.5], where this is worked out in great detail. We summarize the results in Theorem 2.15 below, as it will be frequently convenient to derive results for $\text{GRG}_n(\mathbf{w})$ through those for appropriate $\text{CM}_n(\mathbf{d})$'s.

1.3.4. Related configuration models. In this section, we discuss some random graph models that are adaptations, and/or extensions, of the configuration so as to deal with directed graphs or graphs with community structure.

The directed configuration model. Many real-world networks are *directed*, in the sense that edges are oriented from their starting vertex to their end vertex. For example, in the World-Wide Web, the vertices are web pages, and the edges are the hyperlinks between them, which are clearly oriented. One could naturally forget about these directions, but that would discard a wealth of information. For example, in citation networks, it makes a substantial difference whether my paper cites your paper, or your paper cites mine.

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

$$(1.3.45) \quad \sum_{i \in [n]} d_i^{(\text{in})} = \sum_{i \in [n]} d_i^{(\text{out})}$$

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 $[n]$.

Assume, similarly to Conditions 1.6(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 (1.3.45), this implies that $\mathbb{E}[D^{(\text{out})}] = \mathbb{E}[D^{(\text{in})}]$.

Let

$$(1.3.46) \quad p_{k,l} = \mathbb{P}(D^{(\text{in})} = k, D^{(\text{out})} = l)$$

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

The hierarchical configuration model. 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 1.6(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 1.6(a)-(b), while the empirical distribution of the graphs

$$(1.3.47) \quad \mu_n(H, \vec{d}) = \frac{1}{n} \sum_{i \in [n]} \mathbb{1}_{\{G_i = H, (d_u^{(b)})_{u \in V_i} = \vec{d}\}}$$

converges as $n \rightarrow \infty$ to some probability distribution on graphs with integer marks associated to the vertices. See [246, 247] for power-law relations in hierarchical configuration models and epidemic spread in them, respectively, and [171] for its topological properties, such as its connectivity, its clustering, etc.

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 [26], whose model we will generalize. Bollobás, Riordan, Spencer and Tusnády [59] studied the model by Barabási and Albert [26], and later many other papers followed on this, and related, models. See [160, 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,

$$(1.3.48) \quad \mathbb{P}(v_{t+1}^{(1)} \rightarrow v_i^{(1)} \mid \text{PA}_t^{(1,\delta)}) = \begin{cases} \frac{1 + \delta}{t(2 + \delta) + (1 + \delta)} & \text{for } i = t + 1, \\ \frac{D_i(t) + \delta}{t(2 + \delta) + (1 + \delta)} & \text{for } i \in [t]. \end{cases}$$

The above preferential attachment mechanism is called *affine*, since the attachment probabilities in (1.3.48) 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 [26] 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 $D_i(t) \xrightarrow{a.s.} \infty$:

EXERCISE 1.10 (Degrees grow to infinity a.s.). *Fix $m = 1$ and $i \geq 1$. Prove that $D_i(t) \xrightarrow{a.s.} \infty$, by using that $\sum_{s=i}^t I_s \preceq D_i(t)$, 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 $m > 1$?*

As a result, one could also call the PAM the *old-get-richer model*, which would be more appropriate, as we discuss in ore detail now. 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

$$(1.3.49) \quad \Gamma(t) = \int_0^\infty x^{t-1} e^{-x} dx.$$

The following theorem describes the evolution of the degree of fixed vertices (see [160, Theorem 8.2 and (8.3.11)]):

THEOREM 1.8 (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 $t \rightarrow \infty$.*

EXERCISE 1.11 (Degrees of fixed vertices). *Prove Theorem 1.8 for $m = 1$ and $\delta > -1$ using the martingale convergence theorem and the fact that*

$$(1.3.50) \quad M_i(t) = \frac{D_i(t) + \delta}{1 + \delta} \prod_{s=i-1}^{t-1} \frac{(2 + \delta)s + 1 + \delta}{(2 + \delta)(s + 1)}$$

is a martigale.

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 [160, 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 $\tau = 3 + \tau/m$, and that is our next item on the agenda.

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

$$(1.3.51) \quad P_k(t) = \frac{1}{t} \sum_{i=1}^t \mathbb{1}_{\{D_i(t)=k\}}$$

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 $k \geq m$,

$$(1.3.52) \quad p_k = (2 + \delta/m) \frac{\Gamma(k + \delta)\Gamma(m + 2 + \delta + \delta/m)}{\Gamma(m + \delta)\Gamma(k + 3 + \delta + \delta/m)}$$

It turns out that $(p_k)_{k \geq 0}$ is a probability mass function (see [160, 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 1.9 (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 $t \rightarrow \infty$,*

$$(1.3.53) \quad \mathbb{P}\left(\max_k |P_k(t) - p_k| \geq C\sqrt{\frac{\log t}{t}}\right) = o(1).$$

We next investigate the scale-free properties of $(p_k)_{k \geq 0}$ by investigating the asymptotics of p_k for k large. By (1.3.52) and Stirling's formula, as $k \rightarrow \infty$,

$$(1.3.54) \quad p_k = c_{m,\delta} k^{-\tau} (1 + O(1/k)),$$

where

$$(1.3.55) \quad \tau = 3 + \delta/m > 2, \quad \text{and} \quad c_{m,\delta} = (2 + \delta/m) \frac{\Gamma(m + 2 + \delta + \delta/m)}{\Gamma(m + \delta)}.$$

Therefore, by Theorem 1.9 and (1.3.54), 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.

EXERCISE 1.12 (Power-law degree sequence). *Prove (1.3.55) by using Stirling's formula. is a martigale.*

1.3.6. Related preferential attachment models. There are numerous adaptations of the simple preferential attachment model as defined here. See [160, Section 8.9] for more details. Here we discuss the directed preferential attachment model, as well as a preferential attachment model with conditionally independent edges.

A directed preferential attachment model. Bollobás, Borgs, Chayes and Riordan [53] investigate a directed preferential attachment model and prove that the degrees obey a power law similar to the one in Theorem 1.9. 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

$$(1.3.56) \quad \frac{f_i(t)}{\sum_j f_j(t)}.$$

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, Borgs, Chayes and Riordan [53] 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

$$(1.3.57) \quad X_k(t) = \sum_{v \in G(t)} \mathbb{1}_{\{D_v^{(\text{in})} = k\}},$$

and, similarly, let $(Y_k(t))_{k \geq 0}$ be the out-degree sequence of $G(t)$, so that

$$(1.3.58) \quad Y_k(t) = \sum_{v \in G(t)} \mathbb{1}_{\{D_v^{(\text{out})} = k\}}.$$

Denote

$$(1.3.59) \quad \tau_{\text{in}} = 1 + \frac{1 + \delta_{\text{in}}(\alpha + \beta)}{\alpha + \beta}, \quad \tau_{\text{out}} = 1 + \frac{1 + \delta_{\text{out}}(\gamma + \beta)}{\gamma + \beta}.$$

Then [53, Theorem 3.1] shows that there exist probability distributions $p = (p_k)_{k \geq 0}$ and $q = (q_k)_{k \geq 0}$ such that

$$(1.3.60) \quad X_k(t) - p_k t = O_{\mathbb{P}}(t), \quad Y_k(t) - q_k t = O_{\mathbb{P}}(t),$$

where, for $k \rightarrow \infty$,

$$(1.3.61) \quad p_k = C_{\text{in}} k^{-\tau_{\text{in}}}(1 + o(1)), \quad q_k = C_{\text{out}} k^{-\tau_{\text{out}}}(1 + o(1)).$$

In fact, the probability distributions p and q are determined explicitly, as in (1.3.52) above, and p and q have a similar shape as p in (1.3.52). Also, since $\delta_{\text{in}}, \delta_{\text{out}} \geq 0$, and $\alpha + \beta, \gamma + \beta \leq 1$, we again have that $\tau_{\text{in}}, \tau_{\text{out}} \in (2, \infty)$.

Preferential attachment models with conditionally independent edges. A preferential attachment models with conditionally independent edges is investigated by Dereich and Mörters in a sequence of papers [98, 99, 100]. Fix a preferential attachment function $f: \mathbb{N}_0 \mapsto (0, \infty)$. Then, the graph evolves as follows. Start with $G(1)$ 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

$$(1.3.62) \quad \frac{f(D_i(t-1))}{t-1},$$

where $D_i(t-1)$ is the in-degree of vertex i at time $t-1$. This creates the random graph $G(t)$. Note that the number of edges in the random graph process $(G(t))_{t \geq 1}$ is *not* fixed, and equal a random variable. In particular, it makes a difference whether we use the in-degree in (1.3.62).

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.,

$$(1.3.63) \quad P_k(t) \equiv \frac{1}{t} \sum_{i \in [t]} \mathbb{1}_{\{D_i(t)=k\}} \xrightarrow{\mathbb{P}} p_k, \quad \text{where} \quad p_k = \frac{1}{1 + f(k)} \prod_{l=0}^{k-1} \frac{f(l)}{1 + f(l)}.$$

In particular, $\log(1/p_k)/\log(k) \rightarrow 1 + 1/\gamma$ when $f(k)/k \rightarrow \gamma \in (0, 1)$, while $\log(1/p_k) \sim k^{1-\alpha}/(\gamma(1-\alpha))$ when $f(k) \sim \gamma k^\alpha$ for some $\alpha \in (0, 1)$. 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.

1.4. Related geometries and universality

There are tons of other graph topologies where one can expect similar results as in the random graphs discussed above. In general, the random graphs that we investigate display mean-field behavior, since most pairs of vertices can interact with one another. This is quite different from *geometric* settings, such as arising from squares or tori. For such geometries, graph distances are typically substantial, which is different from real-world networks.

Further, the random graph models investigated here are also *inhomogeneous*, and one can expect that the results depend sensitively on the amount of inhomogeneity. 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.

Random intersection graphs. A model that has received some attention recently is obtained by giving a group structure to the graph. Indeed, fix a number m of groups. Let each of the vertices in $[n]$ connect to a (possibly random) number of groups. This creates a bipartite graph of group memberships. Then we connect all the vertices that are members of the same group to one another. The resulting graph is called the *random*

intersection graph, and is quite flexible in terms of its degree distribution, clustering and connectivity structure.

The complete graph. Arguably the simplest graph model one can consider is the *complete graph*, often denoted by K_n . The vertex set of K_n is $[n]$, while the edge set equals $E_n = \{ij: 1 \leq i < j \leq n\}$. The complete graph K_n has no geometry, and is transitive in the sense that every vertex plays precisely the same role. We can see the Erdős-Rényi random graph as performing percolation on the complete graph. For processes on graphs, the complete graph is the ultimate mean-field model. As a result, we will start each chapter by studying the stochastic process in question on the complete graph, which is both enlightening as well as displays the important characteristics of the stochastic process at hand.

The hypercube. A graph that is transitive as well, but does have substantial geometry is the hypercube Q_d of degree d . For Q_d , the vertex set is $\{0, 1\}^d$, so that $n = 2^d$. Further, the edge set consists of $\{\{x, y\}: x, y \in \{0, 1\}^d, x \sim y\}$, where $x \sim y$ when x and y differ in precisely one coordinate. The hypercube is a geometric graph for which the dimension tends to infinity, which possibly explains its popularity.

Universality continued. The complete graph and, to a lesser extent, the hypercube are geometric graphs that are high-dimensional, yet completely egalitarian. Thus, one universality class for random graphs could consist of those settings where the behavior on the random graph is similar to that on the complete graph or hypercube. We will see that this is often true when the inhomogeneity is not too pronounced, while if it is, the behavior of random graphs is quite different. This brings us to the main topic of this lecture notes, stochastic processes on random graphs, as discussed in the next section.

1.5. Stochastic processes on random graphs and network functionality

We are typically not only interested in networks, but also in their functionality. Indeed, we wish to know how quickly a disease or a rumor spreads, how products compete for the market in social networks, how damage restricts the network's functionality, how people reach consensus when talking to one another, etc. Where we model networks using random graphs, we model their functionality by stochastic processes on them. In recent decades, a tremendous body of work was developed discussing stochastic processes on random graphs. In this text, we will focus on the following key aspects:

Main aim textk:

Describe stochastic processes on random graphs, such as percolation, first-passage percolation, competition, Ising models and random walks. Highlight the relation between the *behavior of stochastic processes on random graphs* and the *topology of the random graphs*. Here topology refers to graph properties such as their degrees, their graph distances, their clustering and so on.

The random graph models that we shall focus mainly on are

- ▷ Inhomogeneous random graphs, in which edges are present *independently*, possibly with inhomogeneous edge probabilities;
- ▷ The configuration model, or the random graph with *prescribed degrees*, in which we can flexibly choose the degrees structure of the random graph;
- ▷ Preferential attachment models, which are *dynamical models* for network evolution such that vertices with high degrees are more likely to attract edges of newly added vertices.

We further discuss extensions to various other random graph models that have attracted the attention in the literature.

The stochastic processes that we focus mainly on are the following:

- ▷ First passage percolation, which is a model for optimal routing or rumor spread in networks;
- ▷ Percolation, which is a model for random failure in networks;
- ▷ Ising models, which is a model for cooperative phenomena in networks;

Many more stochastic processes have been studied in the literature, and we also discuss several other examples.

The *central question in this text* is how the behavior of such processes is related to the inherent *inhomogeneity* present in the graphs, as represented in the highly variable degrees of their vertices.

In physics terms, stochastic processes on random graphs often display ‘mean-field behavior’, which means that they behave in a similar way as their counterparts on the complete graph, possibly apart from some trivial rescaling. Therefore, we start each chapter by fixing the ideas and discussing the behavior of the process on the complete graph first. We also highlight one ‘trick of the trade’, a central piece of methodology that plays a crucial role in that chapter.

While describing the results and proofs, we state many open problems, as well as exercises, along the way.

1.6. Organization and notation

Organisation. This document is organised as follows. Chapter 2 discusses preliminaries, mainly focussing on the topology of the random graphs that we make use of in the remainder of this text. In Chapter 3, we treat first-passage percolation or routing on random graphs, which is a model for the spread of a rumor, for optimal routes in random graphs or the spread of a very simple disease. Chapter 4 discusses percolation on random graphs. Chapter 5 studies Ising models on random graphs as a model for cooperative phenomena in networks. Finally, in Chapter 6, we study related stochastic processes on random graphs, such as competition on random graphs, bootstrap percolation or contact processes.

Notation. Let us introduce some standard notation. We say that a sequence of events E_n occurs *with high probability* (whp) when $\lim_{n \rightarrow \infty} \mathbb{P}(E_n) = 1$. 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 $X_n/Y_n \xrightarrow{\mathbb{P}} 0$.

We say that a sequence of events $(\mathcal{E}_n)_{n \geq 1}$ occurs *with high probability* when $\lim_{n \rightarrow \infty} \mathbb{P}(\mathcal{E}_n) = 1$. We often abbreviate this as “whp”. 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 $i \geq 1$.

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 sometimes abuse notation, and write e.g., $\mathbb{P}(\text{Bin}(n, p) = k)$ to denote $\mathbb{P}(X = k)$ when $X \sim \text{Bin}(n, p)$.

CHAPTER 2

Preliminaries

Abstract

In this chapter, we discuss some of the preliminaries of random graphs that later chapters rely upon. We discuss branching processes, and show how they can be used to describe local neighborhoods in random graphs. Then we continue to discuss properties of random graphs, such as their connectivity structure and the giant component, their small-world nature and their local weak convergence. We further discuss some related network characteristics that have received substantial attention in the literature.

Organization of this chapter. We start in Section 2.1 by discussing branching processes, and show in Section 2.2 how configuration models and generalized random graphs locally-weakly converge to them. In Section 2.3, we investigate when random graph models have a *giant component* containing a positive proportion of the vertices, and in Section 2.4 we give conditions that guarantee that these graphs are fully connected. In Section 2.5, we discuss the small-world nature of the random graphs that we have introduced in the previous chapter. We close this chapter in Section 2.6 by describing some technical results that will prove to be essential throughout this book.

2.1. Branching processes

Branching processes play an important role in random graph theory, because many random graphs locally look like a branching process. For example, in the Erdős-Rényi random graph $ER_n(\lambda/n)$, each vertex has a degree that is close to Poisson distributed with parameter λ . When exploring the neighborhood of a vertex v , say in a breadth-first manner, we inspect a vertex and see what its neighbors are outside of the vertices that have been found already. As long as we perform few of such vertex exploration steps, the number of new vertices found to be in the neighborhood of v is well described by a Poisson branching process with parameter λ . This is a highly useful heuristic, and can be made precise in many cases. As a result, many proofs in random graph theory depend sensitively on branching process results. In this section, we describe a few of those. For more details, the reader is advised to read [160, Chapter 3], or one of the classical texts on branching processes by Athreya and Ney [23], Harris [153] or Jagers [177]. We start with the phase transition in branching processes.

2.1.1. Phase transition branching processes. A branching process is the simplest possible model for a population evolving in time. Suppose that each individual

independently gives birth to a number of children with the same distribution, independently across the different individuals. For reasons that will become clear later on, we denote the offspring distribution of the branching process by $(p_i^*)_{i \geq 0}$. We denote the number of individuals in the k th generation by Z_k , where, by convention, we let $Z_0 = 1$. Then Z_k satisfies the recursion relation

$$(2.1.1) \quad Z_k = \sum_{i=1}^{Z_{k-1}} X_{k,i},$$

where $(X_{k,i})_{k,i \geq 1}$ is a doubly infinite array of i.i.d. random variables. We will often write X for a generic random variable having the law of the offspring distribution, so that $(X_{k,i})_{k,i \geq 1}$ is a doubly infinite array of i.i.d. random variables with $X_{k,i} \sim X$ for all $n, i \geq 0$. In this case, the law $(p_i^*)_{i \geq 0}$ of X is called the *offspring distribution* of the branching process.

One of the major results of branching processes is that when $\mathbb{E}[X] \leq 1$, the population dies out with probability one (unless $X_{1,1} = 1$ with probability one), while when $\mathbb{E}[X] > 1$, there is a non-zero probability that the population will survive forever. In order to state the result, we denote the *extinction probability* by

$$(2.1.2) \quad \eta^* = \mathbb{P}(\exists k: Z_k = 0).$$

Then the main result on the phase transition for branching processes is the following theorem:

THEOREM 2.1 (Survival vs. extinction for branching processes). *For a branching process with i.i.d. offspring X , $\eta = 1$ when $\mathbb{E}[X] < 1$, while $\eta < 1$ when $\mathbb{E}[X] > 1$. Further, $\eta = 1$ when $\mathbb{E}[X] = 1$ and $\mathbb{P}(X = 1) < 1$. The extinction probability η^* is the smallest solution in $[0, 1]$ of*

$$(2.1.3) \quad \eta = G_X(\eta^*),$$

with $s \mapsto G_X(s)$ the probability generating function of the offspring distribution X , i.e.,

$$(2.1.4) \quad G_X(s) = \mathbb{E}[s^X].$$

In this book, we will mainly concentrate on *critical* and *supercritical* branching processes, for which $\mathbb{E}[X] = 1$ and $\mathbb{E}[X] > 1$, respectively.

Unimodular Galton-Watson trees. As described above, we will use branching processes to approximate random graph neighborhoods. Such random graph neighborhoods can be obtained by starting from a uniform vertex in the vertex set $[n]$, and exploring its neighborhood. It turns out that starting from a uniform vertex is closely related to the notion of *local weak convergence*, as discussed in more detail in Section 2.2, and gives rise to slightly altered branching processes. Such branching processes are so-called *unimodular graphs*, which is a rooted graph that often arises as the local weak limit of a sequence of graphs (in our case, the random graphs $(G_n)_{n \geq 1}$). See Aldous and Lyons [11] and Benjamini, Lyons and Schramm [32] for more background on unimodular graphs and trees. Let us continue by defining the so-called *unimodular Galton-Watson tree* as identified by Lyons, Pemantle and Peres in [213].

To define the unimodular Galton-Watson tree, let D be an integer-valued random variable, and $(p_k)_{k \geq 0}$ its probability mass function, so that $p_i = \mathbb{P}(D = i)$. We assume

that $p_0 = 0, p_1 < 1$ and $\sum_{i \geq 0} ip_i = \mathbb{E}[D] < \infty$. Our unimodular Galton-Watson tree will be such that the root has offspring distribution $(p_i)_{i \geq 0}$, while every other individual in the branching process has offspring distribution $(p_k^*)_{k \geq 0}$ given by

$$(2.1.5) \quad p_k^* = \frac{(k+1)p_{k+1}}{\sum_{i \geq 0} ip_i} = \frac{k+1}{\mathbb{E}[D]} \mathbb{P}(D = k+1).$$

Recall that the *size-biased* version X^* of an integer-valued random variable X with $\mathbb{P}(X = 0) < 1, \mathbb{E}[X] < \infty$ is

$$(2.1.6) \quad \mathbb{P}(X^* = k) = \frac{k}{\mathbb{E}[X]} \mathbb{P}(X = k).$$

Then we can see that the probability mass function $(p_k^*)_{k \geq 0}$ satisfies $p_k^* = \mathbb{P}(D^* - 1 = k)$, i.e., it is the distribution of the size-biased version of D minus one.

Let η^* denote the extinction probability of the branching process with offspring distribution $(p_k^*)_{k \geq 0}$. Then, the extinction probability η of the unimodular Galton-Watson tree satisfies

$$(2.1.7) \quad \eta = G_D(\eta^*),$$

where $s \mapsto G_D(s)$ is the probability generating function of the random variable D :

EXERCISE 2.1 (Extinction probability unimodular Galton-Watson tree). *Prove that the extinction probability of the unimodular Galton-Watson tree satisfies $\eta = G_D(\eta^*)$.*

Poisson branching processes are very special examples of unimodular Galton-Watson trees, since they are in fact the only examples where the unimodular Galton-Watson tree is also a normal Galton-Watson tree:

EXERCISE 2.2 (Poisson Galton-Watson trees). *Show that $(p_k)_{k \geq 0} = (p_k^*)_{k \geq 0}$ if and only if $(p_k)_{k \geq 0}$ is the Poisson distribution, i.e., there exists a $\lambda > 0$ such that*

$$(2.1.8) \quad p_k = e^{-\lambda} \frac{\lambda^k}{k!}.$$

Conclude that a unimodular Galton-Watson tree is a regular Galton-Watson tree precisely when the offspring has a Poisson distribution.

2.1.2. Generation growth for branching processes with finite mean. As mentioned before, we will be mainly interested in settings where the random graph on which our processes live are supercritical. This corresponds to settings where the graphs that these branching processes describe locally have a giant component containing a positive proportion of the vertices. Since random graphs can locally be approximated by branching processes, this leads us to study supercritical branching processes. The next theorem describes the growth of their generation sizes in the case where the branching processes have finite mean:

THEOREM 2.2 (Exponential growth for supercritical branching processes). *For a branching process with i.i.d. offspring X having mean $\nu = \mathbb{E}[X] > 1$, $\nu^{-k} Z_k^* \xrightarrow{a.s.} W^*$ for some random variable W which is finite with probability 1.*

EXERCISE 2.3 (Exponential growth through martingales). *Use the martingale convergence theorem to prove Theorem 2.2.*

Let η denote the extinction probability of the branching process with i.i.d. offspring X . The Kesten-Stigum Theorem [196] states that the limiting random variable W in Theorem 2.2 satisfies $\mathbb{P}(W^* = 0) = \eta$ precisely when $\mathbb{E}[X \log(X)] < \infty$ (where, by convention, $0 \log 0 = 0$). When $\mathbb{E}[X \log(X)] = \infty$, instead, $\mathbb{P}(W^* = 0) = 1$, so that Theorem 2.2 is not very useful. In most of what we do, it will be the case that either $\mathbb{E}[X] = \infty$, corresponding to *infinite-mean* branching processes, or $\mathbb{E}[X \log(X)] < \infty$.

For unimodular Galton-Watson trees, we have that the offspring equals $X = D^* - 1$ in every vertex except for the root, while the root has offspring D . In this case, with Z_k denoting the number of individuals in the k th generation of the unimodular Galton-Watson tree, the above convergence result can be adapted to

$$(2.1.9) \quad \nu^{-k} Z_k \xrightarrow{a.s.} W,$$

where

$$(2.1.10) \quad W \stackrel{d}{=} \nu^{-1} \sum_{i=1}^D W_i^*,$$

and $(W_i^*)_{i \geq 1}$ are i.i.d. copies of the limiting random variable W^* in Theorem 2.2. Throughout this text, we will use Z_k for the generation size of the unimodular Galton-Watson tree, while Z_k^* denotes the generation size of a regular Galton-Watson tree (with the convention that its offspring distribution satisfies $X = D^* - 1$).

EXERCISE 2.4 (Exponential growth generations unimodular Galton-Watson trees). *Use Theorem 2.2 to prove (2.1.9)–(2.1.10).*

2.1.3. Exploration of Galton-Watson trees and random walks. In this section, we study the relation between branching process trees and random walks that arise through their *exploration*. Explorations of random graphs are one of the major techniques in random graph theory, and these rely on the fact that the explorations are *close to* random walks.

In branching processes, it is common to study the number of descendants of each individual in a given generation. For random graph purposes, it is often convenient to use a different construction of a branching process by sequentially investigating the number of children of each member of the population. This picture leads to a *random walk formulation of branching processes*. For more background on random walks, we refer the reader to the classic book by Spitzer [244] or the book by Grimmett and Stirzaker [146, Section 5.3]. We follow [160, Sections 3.3 and 3.5] closely.

We now give the random walk representation of a branching process. Let X'_1, X'_2, \dots be independent and identically distributed random variables with the same distribution as Z_1 . Define S'_0, S'_1, \dots by the recursion

$$(2.1.11) \quad \begin{aligned} S'_0 &= 1, \\ S'_i &= S'_{i-1} + X'_i - 1 = X'_1 + \dots + X'_i - (i - 1). \end{aligned}$$

Let T' be the smallest t for which $S'_t = 0$, i.e.,

$$(2.1.12) \quad T' = \inf\{t: S'_t = 0\} = \inf\{t: X'_1 + \dots + X'_t = t - 1\}.$$

In particular, if such a t does not exist, then we define $T' = +\infty$.

The above description turns out to be *equivalent* to the normal definition of a branching process, but records the branching process tree in a different manner. See [160, Lemma 3.6], where this is proved formally. For example, in the random walk picture, it is slightly more difficult to extract the distribution of the generation sizes, and the answer depends on how we explore the branching process tree. Usually, this exploration is performed in either a breadth-first or a depth-first order. For the distribution of $(S'_t)_{t \geq 0}$ this makes no distinction. [160, Lemma 3.6] implies that T' is equal in distribution to the *total progeny* of the branching process, and it is equal to the total number of individuals in the family tree of the initial individual. We will also see that $(X'_k)_{k \geq 1}$ in (2.1.11) and $(X_{i,n})_{i,n \geq 1}$ in the more classical description that

$$(2.1.13) \quad Z_k^* = \sum_{i=1}^{Z_{k-1}^*} X_{i,k}$$

are related through $X'_k = X_{i_k, n_k}$ for a unique i_k, n_k that depends on the order in which the tree is explored. Note that, in terms of $(Z_k^*)_{k \geq 0}$, we have

$$(2.1.14) \quad T = \sum_{k \geq 0} Z_k^*.$$

The branching process belonging to the recursion in (2.1.11) is the following. The population starts with one active individual. At time i , we select one of the active individuals in the population, and give it X'_i children. The children (if any) are set to active, and the individual itself becomes inactive. This exploration process is continued as long as there are active individuals in the population. Then, the process S'_i describes the number of active individuals after the first i individuals have been explored. The process stops when $S'_i = 0$, but the recursion can be defined for all t since this leaves the value of T' unaffected. Note that, for a branching process, (2.1.11) only makes sense as long as $i \leq T'$, since only then $S'_i \geq 0$ for all $i \leq T'$. However, (2.1.11) itself can be defined for *all* $i \geq 0$, also when $S'_i < 0$. This fact will be useful in the sequel.

From now on and with [160, Lemma 3.6] in mind, we will slightly abuse notation and write $(S_i)_{i \geq 0}$ and $(X_i)_{i \geq 1}$ instead of $(S'_i)_{i \geq 0}$ and $(X'_i)_{i \geq 1}$. Let us give one example of the use of the random-walk description for branching processes, which is an explicit formula for the distribution of the *total progeny* (i.e., the total number of individuals ever alive) of the branching process:

THEOREM 2.3 (Law of total progeny). *For a branching process with i.i.d. offspring distribution $Z_1^* = X$ and all $k \geq 0$,*

$$(2.1.15) \quad \mathbb{P}(T = k) = \frac{1}{k} \mathbb{P}(X_1 + \cdots + X_k = k - 1),$$

where $(X_i)_{i=1}^k$ are i.i.d. copies of X .

See [160, Theorem 3.14], to which we also refer for an extensive discussion, and the original proof with Keane in [168].

2.1.4. Generation growth for branching processes with infinite mean. A special case for unimodular Galton-Watson trees when the distribution of the root has finite mean, but infinite variance. In this case, the offspring distribution of all vertices except for the root have *infinite mean*. The following theorem gives a condition for *double-exponential growth* of the generation sizes of such branching processes, and is due to Davies [88]:

THEOREM 2.4 (Branching processes with infinite mean). *Let $(Z_k^*)_{k \geq 0}$ be the generation sizes of 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*

$$(2.1.16) \quad x^{-\alpha-\gamma(x)} \leq 1 - F_X(x) \leq x^{-\alpha+\gamma(x)}, \quad \text{for large } x,$$

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 $\alpha^k \log(Z_k^* \vee 1) \xrightarrow{a.s.} Y^*$, with $\mathbb{P}(Y^* = 0)$ equal to the extinction probability of $(Z_k^*)_{k \geq 0}$, whereas Y^* is non-degenerate on $(0, \infty)$.

In the analysis for the configuration model, $\alpha = \tau - 2$, as α corresponds to the tail exponent of the size-biased random variable D^* . Theorem 2.4 covers the case where the branching process has an offspring which has very thick tails. Indeed, it is not hard to show that Theorem 2.4 implies that $\mathbb{E}[X^s] = \infty$ for every $s > \alpha \in (0, 1)$ (see Exercise 2.6 below).

For unimodular Galton-Watson trees, the above convergence result can be adapted to

$$(2.1.17) \quad \alpha^k \log(Z_k \vee 1) \xrightarrow{a.s.} Y,$$

where

$$(2.1.18) \quad Y \stackrel{d}{=} \alpha \min_{i=1}^D Y_i^*,$$

and $(Y_i^*)_{i \geq 1}$ are i.i.d. copies of the limiting random variable in Theorem 2.4 for the offspring distribution $(p_k^*)_{k \geq 0}$.

We do not prove Theorem 2.4 here, and rather sketch its proof in a simpler, yet still quite general case, in which $\gamma(x) = (\log x)^{\gamma-1}$ for some $\gamma \in [0, 1)$.

EXERCISE 2.5 (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 2.4.*

Idea behind the proof of Theorem 2.4 for $\gamma(x) = (\log x)^{\gamma-1}$. Define

$$(2.1.19) \quad Y_k^* = \alpha^k \log(Z_k^* \vee 1).$$

We assume for simplicity that $\mathbb{P}(Z_1^* \geq 1) = 1$, so that $\eta = 1$. We start by splitting Y_k^* in a suitable way. For $i \geq 1$, we define

$$(2.1.20) \quad U_i = \alpha^i \log \left(\frac{(Z_i^* \vee 1)}{(Z_{i-1}^* \vee 1)^{1/\alpha}} \right),$$

so that

$$(2.1.21) \quad Y_k^* = U_1 + U_2 + \cdots + U_k.$$

From this split, it is clear that almost sure convergence of Y_k^* follows when the sum $\sum_{i=0}^{\infty} U_i$ converges, which, in turn, is the case when

$$(2.1.22) \quad \sum_{i=1}^{\infty} \mathbb{E}[|U_i|] < \infty.$$

Now, when $k \rightarrow \infty$, we have that $(Z_i^* \vee 1)/(Z_{i-1}^* \vee 1)^{1/\alpha} \xrightarrow{d} \mathcal{S}_\alpha$, where \mathcal{S}_α denotes an α -stable distribution. In particular, $\mathbb{E}[\log \mathcal{S}_\alpha] < \infty$, which suggests that indeed (2.1.22) holds. Proving (2.1.22) is more technical, and is omitted here. \square

EXERCISE 2.6 (Infinite mean under conditions Theorem 2.4). *Prove that $\mathbb{E}[X] = \infty$ when the conditions in Theorem 2.4 are satisfied. Extend this to show that $\mathbb{E}[X^s] = \infty$ for every $s > \alpha \in (0, 1)$.*

EXERCISE 2.7 (Convergence for $Z_k^* + 1$). *Show that, under the conditions of Theorem 2.4, also $\alpha^k \log(Z_k^* + 1)$ converges to Y almost surely.*

We finally state some properties of the a.s. limit Y of $(\alpha^k \log(Z_k^* \vee 1))_{k \geq 0}$, of which we omit a proof:

THEOREM 2.5 (Limiting variable for infinite-mean branching processes). *Under the conditions of Theorem 2.4,*

$$(2.1.23) \quad \lim_{x \rightarrow \infty} \frac{\log \mathbb{P}(Y^* > x)}{x} = -1,$$

where Y^* is the a.s. limit of $\alpha^k \log(Z_k^* \vee 1)$.

Theorem 2.5 can be understood from the fact that by (2.1.19)–(2.1.20),

$$(2.1.24) \quad Y^* = \sum_{i=1}^{\infty} U_i,$$

where

$$(2.1.25) \quad U_1 = \alpha \log(Z_1^* \vee 1).$$

By (2.1.16),

$$(2.1.26) \quad \mathbb{P}(U_1 > x) = \mathbb{P}(Z_1^* > e^{x^{1/\alpha}}) = e^{-x(1+o(1))},$$

which shows that Y_1 satisfies (2.1.23). The equality in (2.1.24) together with (2.1.20) suggests that the tails of Y are equal to those of Y_1 , which heuristically explains (2.1.23).

2.2. Local weak convergence of random graphs

In this section, we establish local weak convergence of $\text{CM}_n(\mathbf{d})$ to a unimodular Galton-Watson tree assuming that Conditions 1.6(a)-(b) hold. We show that the same result follows for $\text{GRG}_n(\mathbf{w})$ when Conditions 1.1(a)-(b) hold. We first recall what local weak convergence, as introduced by Benjamini and Schramm [33], means in this context. Local weak convergence to unimodular Galton-Watson trees is often called *locally tree-like* behavior. We start by discussing local weak convergence in general.

2.2.1. Local weak convergence. Let us start by discussing the notion of local weak convergence in more detail. We follow Aldous and Steele [12], using results from Benjamini, Lyons and Schramm [32] as well. A graph is *locally finite* when each of its vertices has finite degree (though not necessarily uniformly bounded). 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. Here, graphs can be finite or infinite, but we will always have graphs that are locally finite in mind. We let \mathcal{G}_* denote the space of rooted graphs, as introduced by Aldous and Steel [12].

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

$$(2.2.1) \quad V(B_o(r)) = \{u: d_G(o, u) \leq r\}, \quad E(B_o(r)) = \{\{u, v\} \in E(G): u, v \in B_o(r)\}.$$

Recall that two graphs $G_1 = (V(G_1), E(G_1))$ and $G_2 = (V(G_2), E(G_2))$ are *isomorphic*, which we write as $G_1 \simeq G_2$, when there exists a map $\phi: V(G_1) \mapsto V(G_2)$ such that $\{u, v\} \in E(G_1)$ precisely when $\{\phi(u), \phi(v)\} \in E(G_2)$. These notions allow us to turn the space of connected rooted graphs into a metric space:

DEFINITION 2.6 (Metric on rooted graphs). *Let (G_1, o_1) and (G_2, o_2) be two rooted connected graphs, and write $B_{i, v_i}(r)$ for the neighborhood of vertex v_i in G_i . Let*

$$(2.2.2) \quad R^* = \sup\{r: B_{1, o_1}(r) \simeq B_{2, o_2}(r)\}.$$

Define

$$(2.2.3) \quad d_{\mathcal{G}_*}((G_1, o_1), (G_2, o_2)) = 1/(R^* + 1).$$

The space \mathcal{G}_* of rooted graphs is a nice metric space under the metric $d_{\mathcal{G}_*}$ in (2.2.3), in that it is separable and thus Polish. The value R^* is the largest value of r for which $B_{1, o_1}(r)$ is isomorphic to $B_{2, o_2}(r)$. When $R^* = \infty$, then $B_{1, o_1}(r)$ is isomorphic to $B_{2, o_2}(r)$ for *every* $r \geq 1$, and then the rooted graphs G_1 and G_2 are the same. Various generalizations are possible. For example Aldous and Steel [12] 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 [32] 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. We will not use such more general set ups, but these will be highly relevant in many applications. Aldous and Lyons [11] also study the implications for stochastic processes, such as percolation and random walks, on unimodular graphs.

We next define local weak convergence of finite graphs:

DEFINITION 2.7 (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 convergence sense to (G, o) , which is a (possibly) random element of \mathcal{G}_* having law μ , when*

$$(2.2.4) \quad \mathbb{E}_n [h(G_n, o_n)] \rightarrow \mathbb{E}[h(G, o)],$$

where the expectation on the right-hand side of (2.2.4) 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.

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 (2.2.4) is equal to

$$(2.2.5) \quad \mathbb{E}_n [h(G_n, o_n)] = \frac{1}{n} \sum_{u \in [n]} h(G_n, u).$$

The notion of local weak convergence is hard to grasp, and it also may appear to be rather weak. In the sequel, we discuss examples of graphs that converge locally weakly. Further, see Section 2.7 for some examples of how local weak convergence may be used. The notion will also be crucially used in these notes. We continue by discussing a convenient criterion for proving local weak convergence.

Criterion for local weak convergence. While the notion of local weak convergence may seem somewhat weak, it actually turns out to be quite convenient, as we will see in the remainder of this text. Before going into various proofs of local weak convergence for the random graph models as defined in Chapter 1, we first give a convenient criterium for convergence:

THEOREM 2.8 (Criterion for local weak convergence). *Let $(G_n)_{n \geq 1}$ be a sequence of rooted graphs. Let $d_{o_n}^{(n)}$ denote the degree of o_n in G_n . Then $(G_n, o_n)_{n \geq 1}$ is tight precisely when $(d_{o_n}^{(n)})_{n \geq 1}$ forms a uniformly integrable sequence of random variables. Further, the only possible limit of (G_n, o_n) is (G, o) when, for every graph H ,*

$$(2.2.6) \quad p^{(n)}(H) = \frac{1}{n} \sum_{u \in [n]} \mathbb{1}_{\{B_u^{(n)}(k) \simeq H\}} \rightarrow \mathbb{P}(B_o(k) \simeq H),$$

where $B_u^{(n)}(k)$ is the k -neighborhood of u in G_n , and $B_o(k)$ is the k -neighborhood of o in the limiting graph (G, o) . Consequently, (G_n, o_n) converges to (G, o) when $(d_{o_n}^{(n)})_{n \geq 1}$ forms a uniformly integrable sequence of random variables and (2.2.6) holds. As a result, then also G_n locally weakly converges to (G, o) .

PROOF. The tightness statement is [32, Theorem 3.1]. Further, when (2.2.6) holds, then the bounded and continuous functions

$$(2.2.7) \quad h_H(G, o) = \mathbb{1}_{\{B_o(k) \simeq H\}}$$

satisfy (2.2.4). Since these functions uniquely identify the limit, this proves the claim. \square

We first give some examples in the form of exercises:

EXERCISE 2.8 (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 2.9 (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.*

Local weak convergence for random graphs. We next discuss settings 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 two related notions of convergence that we may consider. We start with *convergence in distribution*:

DEFINITION 2.9 (Local weak convergence). *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 . Then,*

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

$$(2.2.8) \quad \mathbb{E}[h(G_n, o_n)] \rightarrow \mathbb{E}[h(G, o)],$$

where the expectation on the right-hand side of (2.2.8) 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} 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 convergence sense to (G, o) when*

$$(2.2.9) \quad \mathbb{E}_n[h(G_n, o_n)] \xrightarrow{\mathbb{P}} \mathbb{E}[h(G, o)],$$

where the expectation on the right-hand side of (2.2.9) 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 two closely related definitions lies in what can be concluded from it. Indeed, when we have convergence in probability in the local weak convergence 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. Thus, this is sometimes called a *quenched* result. When we have convergence in distribution in the local weak convergence sense, instead, then only expectation 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 convergence sense. In Chapter 5, we see an example for the pressure in the Ising model, where convergence in probability in the local weak convergence sense implies that the pressure on a random graph converges to a *deterministic* limit.

The following exercise shows a consequence of local weak convergence in probability that can be quite useful:

EXERCISE 2.10 (Independent neighborhoods of different vertices). *Let G_n converge in probability in the local weak convergence sense to (G, o) . Let $(o_n^{(1)}, o_n^{(2)})$ be two independent 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) .*

We next discuss a convenient criterion for local weak convergence, inspired by Theorem 2.8:

THEOREM 2.10 (Criterion for local weak convergence of random graphs). *Let $(G_n)_{n \geq 1}$ be a sequence of rooted graphs. Let $d_{o_n}^{(n)}$ denote the degree of o_n in G_n . Let (G, o) be a random variable on \mathcal{G}_* having law μ . Then,*

(a) G_n converges in distribution in the local weak convergence sense to (G, o) when

$$(2.2.10) \quad \limsup_{K \rightarrow \infty} \limsup_{n \rightarrow \infty} \mathbb{P} \left(\sum_{v \in [n]} d_v^{(n)} \mathbb{1}_{\{d_v^{(n)} \geq K\}} \geq \varepsilon \right) = 0,$$

and when, for every rooted graph $(H, y) \in \mathcal{G}_*$,

$$(2.2.11) \quad \mathbb{E}[p^{(n)}(H, y)] = \frac{1}{n} \sum_{u \in [n]} \mathbb{P}(B_u^{(n)}(k) \simeq (H, y)) \rightarrow \mathbb{P}((B_o(k), o) \simeq (H, y)),$$

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

(b) G_n converges in probability in the local weak convergence sense to (G, o) when (2.2.10) holds, and when, for every rooted graph $(H, y) \in \mathcal{G}_*$,

$$(2.2.12) \quad p^{(n)}(H, y) = \frac{1}{n} \sum_{u \in [n]} \mathbb{1}_{\{B_u^{(n)}(k) \simeq (H, y)\}} \xrightarrow{\mathbb{P}} \mathbb{P}((B_o(k), o) \simeq (H, y)),$$

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

PROOF. This follows directly from Theorem 2.8. In particular, (2.2.10) shows that $\sum_{v \in [n]} \mathbb{1}_{\{d_v^{(n)} \geq K\}}$ converges to zero as first n and then K tends to infinity. Since convergence in distribution and in probability are the same when the limit is a constant, (2.2.10) is the same for local weak convergence in probability and in distribution. \square

EXERCISE 2.11 (Example of weak convergence where convergence in probability fails). *Construct an example where G_n converges in distribution in the local weak convergence sense to (G, o) , but not in probability.*

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

2.2.2. Local weak convergence to unimodular trees for $\text{CM}_n(\mathbf{d})$. We next investigate the locally tree-like nature of the configuration model. Our main result is as follows:

THEOREM 2.11 (Locally tree-like nature $\text{CM}_n(\mathbf{d})$). *Assume that Conditions 1.6(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 $p_k = \mathbb{P}(D = k)$.*

We first note that the uniform integrability of d_{o_n} is, when Condition 1.6(a) holds, equivalent to Condition 1.6(b). This explains the necessity of Conditions 1.6(a)-(b). In order to get started for the proof of (2.2.12) for Theorem 2.11, we introduce some notation. Fix a rooted tree (T, y) with k generations, and let

$$(2.2.13) \quad N_n(T, y) = \sum_{v \in [n]} \mathbb{1}_{\{B_v^{(n)}(k) \simeq (T, y)\}}$$

denote the number of vertices whose local neighborhood up to generation t equals (T, y) . By Theorem 2.10, in order to prove Theorem 2.11, we need to show that

$$(2.2.14) \quad \frac{N_n(T, y)}{n} \xrightarrow{\mathbb{P}} \mathbb{P}(\mathcal{T}(p, k) \simeq (T, y)).$$

For this, we will use a *second moment method*. We first prove that the first moment $\mathbb{E}[N_n(T, y)]/n \rightarrow \mathbb{P}(\mathcal{T}(p, k) \simeq (T, y))$, after which we prove that $\text{Var}(N_n(T, y)) = o(n^2)$. Then, by the Chebychev inequality, (2.2.14) follows.

Local weak convergence to homogeneous trees for $\text{CM}_n(\mathbf{d})$: 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

$$(2.2.15) \quad \mathbb{P}(D_n^* = k) = \frac{k}{\mathbb{E}[D_n]} \mathbb{P}(D_n = k), \quad k \in \mathbb{N},$$

is the size-biased distribution of D_n . Denote this branching process by $(\text{BP}_n(t))_{t \in \mathbb{N}_0}$. Here, $\text{BP}_n(t)$ denotes the branching process when it contains precisely t vertices, and we explore it in the breadth-first order. Clearly, by Conditions 1.6(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 $\mathcal{T}(p) \equiv \mathcal{T}(p, \infty)$ to its first t individuals:

EXERCISE 2.12 (Convergence of n -dependent branching process). *Assume that Conditions 1.6(a)-(b) hold. Prove that $D_n \xrightarrow{d} D$ and $D_n^* \xrightarrow{d} D^*$, and conclude that $\text{BP}_n(t) \xrightarrow{d} \text{BP}(t)$ for every t finite.*

For future reference, denote the offspring distribution of the above branching process by

$$(2.2.16) \quad p_k^* = \mathbb{P}(D^* - 1 = k)$$

Note that, when (T, y) is a fixed rooted tree of k generations, then $\mathcal{T}(p, k) \simeq (T, y)$ precisely when $\text{BP}(t_k) \simeq (T, y)$, where t_k denotes the number of vertices in (T, y) .

We let $(\mathbf{G}_n(t))_{t \in \mathbb{N}_0}$ denote the graph exploration process from a uniformly chosen vertex $U \in [n]$. 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 $(B_U(t))_{t \in \mathbb{N}_0}$, where $D_n = d_i$. 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 2.12 (Coupling graph exploration and branching process). *Subject to Conditions 1.6(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*

$$(2.2.17) \quad \mathbb{P}\left(\widehat{(\mathbf{G}}_n(t))_{0 \leq t \leq m_n} \neq \widehat{(\text{BP}}_n(t))_{0 \leq t \leq m_n}\right) = o(1),$$

whenever $m_n \rightarrow \infty$ arbitrarily slowly. Consequently, $\mathbb{E}[N_n(\mathcal{T})]/n \rightarrow \mathbb{P}(\mathcal{T}(p, k) \simeq \mathcal{T})$.

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 1.6(a)-(b) hold.

PROOF. We let the offspring of the root of the branching process \widehat{D}_n be equal to d_U , which is the number of neighbors of the vertex $U \in [n]$ that is chosen uniformly at random. By construction, $\widehat{D}_n = d_U$, so that also $\widehat{G}_n(1) = \widehat{BP}_n(1)$. We next explain how to *jointly* construct $(\widehat{G}_n(t), \widehat{BP}_n(t))_{0 \leq t \leq m}$ given that we have already constructed $(\widehat{G}_n(t), \widehat{BP}_n(t))_{0 \leq t \leq m-1}$.

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 $(\widehat{BP}_n(t))_{0 \leq t \leq m-1}$ 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 $(\widehat{BP}_n(t))_{0 \leq t \leq m}$. 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 $(\widehat{BP}_n(t))_{t \geq 0}$:

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 $(\widehat{BP}_n(t))_{0 \leq t \leq m}$, 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

$$(2.2.18) \quad \frac{2m-1}{\ell_n}.$$

Hence the probability that a half-edge is being re-used before time m_n is at most

$$(2.2.19) \quad \sum_{m=1}^{m_n} \frac{2m-1}{\ell_n} = \frac{m_n^2}{\ell_n} = o(1),$$

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

$$(2.2.20) \quad \frac{m_n(m_n - 1)}{2} \frac{d_i^2}{\ell_n^2}.$$

By the union bound, the probability that there exists a vertex that is chosen twice up to time m_n is at most

$$(2.2.21) \quad \frac{m_n(m_n - 1)}{2\ell_n} \sum_{i \in [n]} \frac{d_i^2}{\ell_n} \leq m_n^2 \frac{d_{\max}}{\ell_n} = o(1),$$

by Condition 1.6 when $m_n = o(\sqrt{n/d_{\max}})$.

Completion of the proof. In order to show that $\mathbb{E}[N_n(T, y)]/n \rightarrow \mathbb{P}(\mathcal{T}(p, k) \simeq (T, y))$, we let t_k denote the number of individuals in the first $k - 1$ generations in (T, y) , and let $(\mathcal{T}(t))_{0 \leq t \leq t_k}$ be its breadth-first exploration. Then,

$$(2.2.22) \quad \mathbb{E}[N_n(T, y)]/n = \mathbb{P}((\mathbf{G}_n(t))_{0 \leq t \leq t_k} = (\mathcal{T}(t))_{0 \leq t \leq t_k}),$$

so that

$$(2.2.23) \quad \begin{aligned} \mathbb{P}((\mathbf{G}_n(t))_{0 \leq t \leq t_k} = (\mathcal{T}(t))_{0 \leq t \leq t_k}) &= \mathbb{P}((\mathbf{BP}_n(t))_{0 \leq t \leq t_k} = (\mathcal{T}(t))_{0 \leq t \leq t_k}) + o(1) \\ &= \mathbb{P}((\mathbf{BP}(t))_{0 \leq t \leq t_k} = (\mathcal{T}(t))_{0 \leq t \leq t_k}) + o(1) \\ &= \mathbb{P}(\mathcal{T}(p, k) \simeq (T, y)) + o(1), \end{aligned}$$

where the first equality is (2.2.17), while the second is the statement that $\mathbf{BP}_n(t) \xrightarrow{d} \mathbf{BP}(t)$ for every t finite from Exercise 2.12. This proves the claim. \square

Local weak convergence to homogeneous trees for $\text{CM}_n(\mathbf{d})$: second moment. Here, we study the second moment of $N_n(T, y)$, and show that it is close to the first moment squared:

LEMMA 2.13 (Concentration of the number of trees). *Assume that Conditions 1.6(a)-(b) hold. Then,*

$$(2.2.24) \quad \frac{\mathbb{E}[N_n(T, y)^2]}{n^2} \rightarrow \mathbb{P}(\mathcal{T}(p, k) \simeq (T, y))^2.$$

Consequently, $N_n(T, y)/n \xrightarrow{\mathbb{P}} \mathbb{P}(\mathcal{T}(p, k) \simeq (T, y))$.

PROOF. We start by computing

$$(2.2.25) \quad \frac{\mathbb{E}[N_n(T, y)^2]}{n^2} = \mathbb{P}(B_{U_1}^{(n)}(k) = B_{U_2}^{(n)}(k) = (T, y)),$$

where $U_1, U_2 \in [n]$ are two vertices chosen uniformly at random from $[n]$, independently. With high probability, $U_2 \notin B_{U_1}^{(n)}(k)$, so that

$$(2.2.26) \quad \frac{\mathbb{E}[N_n(T, y)^2]}{n^2} = \mathbb{P}(B_{U_1}^{(n)}(k) = B_{U_2}^{(n)}(k) = (T, y), U_2 \notin B_{U_1}^{(n)}(2k)) + o(1).$$

We now condition on $B_{U_1}^{(n)}(k) = (T, y)$, and write

(2.2.27)

$$\begin{aligned} \mathbb{P}(B_{U_1}^{(n)}(k) = B_{U_2}^{(n)}(k) = (T, y), U_2 \notin B_{U_1}^{(n)}(2k)) \\ = \mathbb{P}(B_{U_2}^{(n)}(k) = (T, y), U_2 \notin B_{U_1}^{(n)}(2k) \mid B_{U_1}^{(n)}(k) = (T, y)) \mathbb{P}(B_{U_1}^{(n)}(k) = (T, y), U_2 \notin B_{U_1}^{(n)}(2k)). \end{aligned}$$

We already know that $\mathbb{P}(B_{U_1}^{(n)}(k) = (T, y)) \rightarrow \mathbb{P}(\mathcal{T}(p, k) \simeq (T, y))$, so also

$$(2.2.28) \quad \mathbb{P}(B_{U_1}^{(n)}(k) = (T, y), U_2 \notin B_{U_1}^{(n)}(2k)) \rightarrow \mathbb{P}(\mathcal{T}(p, k) \simeq (T, y)).$$

EXERCISE 2.13 (Proof of no-overlap property in (2.2.28)). *Prove that $\mathbb{P}(B_{U_1}^{(n)}(k) = (T, y), U_2 \in B_{U_1}^{(n)}(2k)) \rightarrow 0$, and conclude that (2.2.28) holds.*

Therefore, it suffices to prove that also

$$(2.2.29) \quad \mathbb{P}(B_{U_2}^{(n)}(k) = (T, y) \mid B_{U_1}^{(n)}(k) = (T, y), U_2 \notin B_{U_1}^{(n)}(2k)) \rightarrow \mathbb{P}(\mathcal{T}(p, k) \simeq (T, y)).$$

For this, we note that, conditionally on $B_{U_1}^{(n)}(k) = (T, y), U_2 \notin B_{U_1}^{(n)}(k)$, the probability that $B_{U_2}^{(n)}(k) = (T, y)$ is the *same* as the probability that $B_{U_2}^{(n)}(k) = \mathcal{T}$ in $\text{CM}_{n'}(\mathbf{d}')$ which is obtained by removing the vertices in the first $k - 1$ generations of \mathcal{T} and adding one vertex with degree the total number of vertices in the k th generation (which is the same as the total degree of vertices in the $(k - 1)$ st generation). Thus, $n' = n - t_k + 1$ and \mathbf{d}' is the corresponding degree sequence. The whole point is that the degree distribution \mathbf{d}' still satisfies Conditions 1.6(a)-(b). Therefore, also

$$(2.2.30) \quad \mathbb{P}(B_{U_2}^{(n)}(k) = (T, y), U_2 \notin B_{U_1}^{(n)}(k) \mid B_{U_1}^{(n)}(k) = (T, y)) \rightarrow \mathbb{P}(\mathcal{T}(p, k) \simeq (T, y)),$$

and we have proved (2.2.24). Since $\mathbb{E}[N_n(T, y)]/n \rightarrow \mathbb{P}(\mathcal{T}(p, k) \simeq (T, y))$ and $\mathbb{E}[N_n(T, y)^2]/n^2 \rightarrow \mathbb{P}(\mathcal{T}(p, k) \simeq (T, y))^2$, it follows that $\text{Var}(\mathbb{E}[N_n(T, y)]/n) \rightarrow 0$, so that $\mathbb{E}[N_n(T, y)]/n$ is concentrated. Since also $\mathbb{E}[N_n(T, y)]/n \rightarrow \mathbb{P}(\mathcal{T}(p, k) \simeq (T, y))$, we obtain that $N_n(T, y)/n \xrightarrow{\mathbb{P}} \mathbb{P}(\mathcal{T}(p, k) \simeq (T, y))$, as required. \square

Lemma 2.13 completes the proof of Theorem 2.11. \square

2.2.3. Local weak convergence to unimodular trees for $\text{GRG}_n(\mathbf{w})$. We next investigate the locally tree-like nature of the generalized random graph. Our main result is as follows:

THEOREM 2.14 (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*

$$(2.2.31) \quad p_k = \mathbb{P}(D = k) = \mathbb{E} \left[e^{-W} \frac{W^k}{k!} \right].$$

This result also applies to $\text{NR}_n(\mathbf{w})$ and $\text{CL}_n(\mathbf{w})$ under the same conditions.

Of course, we could prove Theorem 2.14 directly. However, we choose to follow a different route, by relying on the local tree-like nature of $\text{CM}_n(\mathbf{d})$ proved in Theorem 2.11, and the relation between $\text{GRG}_n(\mathbf{w})$ and $\text{CM}_n(\mathbf{d})$ discussed in Section 1.3. 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 first note that $\text{GRG}_n(\mathbf{w})$ conditioned on its degrees $\mathbf{d} = (d_i)_{i \in [n]}$ is a uniform simple graph with degrees \mathbf{d} (recall (1.3.26)). Therefore, it has the same distribution as $\text{CM}_n(\mathbf{d})$ conditioned on simplicity. By (1.3.43), the probability of simplicity is uniformly positive when Conditions 1.6(a)-(c). Therefore, we conclude that also $N_n(T, y)/n \xrightarrow{\mathbb{P}} \mathbb{P}(\mathcal{T}(q, k) \simeq \mathcal{T})$ holds for $\text{CM}_n(\mathbf{d})$ conditioned on simplicity. If we would know that the degrees \mathbf{d} in $\text{GRG}_n(\mathbf{w})$ satisfy Conditions 1.6(a)-(c) in probability when \mathbf{w} satisfies Conditions 1.1(a)-(c), then the result for $\text{GRG}_n(\mathbf{w})$ would follow for weights satisfying Conditions 1.1(a)-(c). This is the content of the next result.

In the following theorem, we show that Condition 1.1 on the weight sequence \mathbf{w} implies that the degrees \mathbf{d} in $\text{GRG}_n(\mathbf{w})$ satisfy Condition 1.6. To avoid confusion, we now write d_i for the degree of vertex i in $\text{GRG}_n(\mathbf{w})$, which is a *random variable*. The reason is that Condition 1.6 is phrased in terms of D_n , which is the degree of a uniformly chosen vertex, and which causes confusion with the degree of vertex n , which is d_n :

THEOREM 2.15 (Regularity conditions weights and degrees). *Let d_i be the degree of vertex i in $\text{GRG}_n(\mathbf{w})$ defined in (1.3.19), and let $\mathbf{d} = (d_i)_{i \in [n]}$. Then, \mathbf{d} satisfies Conditions 1.6(a)-(c) in probability when \mathbf{w} satisfies Conditions 1.1(a)-(c), where*

$$(2.2.32) \quad \mathbb{P}(D = k) = \mathbb{E} \left[\frac{W^k}{k!} e^{-W} \right]$$

denotes the mixed-Poisson distribution with mixing distribution W having distribution function F in Condition 1.1(a).

This is [160, Theorem 7.19]. The proof of [160, Theorem 7.19] also reveals that \mathbf{d} satisfies Conditions 1.6(a)-(b) in probability when \mathbf{w} satisfies Conditions 1.1(a)-(b), which is useful in what follows. By the discussion above, Theorem 2.15 proves Theorem 2.14 under Conditions 1.1(a)-(c).

Extension to local weak convergence under Conditions 1.1(a)-(b). The extension of Theorem 2.14 under Conditions 1.1(a)-(b) can easily be obtained using an appropriate truncation argument of the weight distribution. This is worked out in detail in [161, Proof of Theorem 4.11], where it was used to identify the giant component in this setting. and we omit further details here.

Restriction to $\text{CM}_n(\mathbf{d})$. The above line of argument can be followed in many settings, and implies that once a result is proved under appropriate conditions for $\text{CM}_n(\mathbf{d})$, it can be relatively easily be extended to $\text{GRG}_n(\mathbf{w})$ under appropriate conditions on the weight distribution. This explains why, in the sequel, we focus mainly on the configuration model, and less on generalized or other inhomogeneous random graphs.

EXERCISE 2.14 (Local weak convergence of $\text{ER}_n(\lambda/n)$). *Use Theorem 2.14 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 .*

2.2.4. 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. This multitype branching process local weak limit has only been identified in the preferential attachment models with conditionally independent edges, and is established by Berger, Borgs, Chayes and Saberi [34] for the regular PAM, and by Dereich and Mörters [98, 99, 100] for the PAM with conditionally independent edges.

Local weak convergence of PAMs with fixed number of edges. The proof by Berger et al. [34] makes crucial use of the fact that the model with $\delta \geq 0$ can be obtained by choosing the i th vertex at time n according to degree with probability $\tilde{\alpha}_n(i)$, and uniform with probability $1 - \tilde{\alpha}_n(i)$. Here, $\tilde{\alpha}_n(i)$ is chosen appropriately as

$$(2.2.33) \quad \tilde{\alpha}_n(i) = \alpha \frac{2m(n-1)}{2m(n-1) + 2m\alpha + (1-\alpha)(i-1)} = \alpha + O(1/n),$$

where

$$(2.2.34) \quad \alpha = \frac{\delta}{2m + \delta}.$$

See also [160, Exercise 8.7] where this connection is explained in the setting where $m = 1$. The setting for $m \geq 2$ follows again by collapsing several vertices. The preferential attachment model investigated by Berger et al. [34] is one where no self-loops are allowed (coined $\text{PA}_t^{(m,\delta)}(b)$ in [160, Chapter 8]). It can be expected that the precise rules do not matter tremendously, so let us explain the result in the simplest possible setting.

Since choosing vertices according to degrees, as well as choosing them uniformly, both have nice exchangeability properties, Pólya urn schemes can be used to prove local weak convergence in this setting. Let the graph start 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 $\text{PA}_n^{(m,\delta)}(b)$ (described above) 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

$$(2.2.35) \quad u = \frac{\alpha}{1-\alpha} = \frac{\delta}{2m},$$

and write

$$(2.2.36) \quad \chi = \frac{1+2u}{1+u} = \frac{2(m+\delta)}{2m+\delta}, \quad \psi = \frac{1-\chi}{\chi} = \frac{1}{1+2u} = \frac{m}{m+\delta}.$$

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 $\text{PA}_t^{(m,\delta)}(b)$ 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. Recall that a Gamma distribution X with parameter r has density $f_X(x) = x^{r-1}e^{-x}/\Gamma(r)$ for $x \geq 0$ and 0 otherwise. Let X have a Gamma distribution with parameter $m + 2mu$, and let X^* be the size-biased version of X , which has a Gamma distribution with parameter $m + 2mu + 1$:

EXERCISE 2.15 (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 $r + 1$.*

Again, we see that the local weak limit of $\text{PA}_n^{(m,\delta)}(b)$ is 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 is in connected to, as well as the number of older 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 sequences

$$(2.2.37) \quad a = (\emptyset, a_1, a_2, \dots, a_l),$$

each carrying a label R or L , which are defined inductively as follows: The root \emptyset has a position $x_0 = y^x$, where y_0 is chosen uniformly at random in $[0, 1]$. In the induction step, we assume that $(\emptyset, a_1, a_2, \dots, a_l)$ and the corresponding variable $x_a \in [0, 1]$ have been chosen in a previous step. Define $(a, j) = (\emptyset, a_1, a_2, \dots, a_l, j)$, for $j \geq 1$, and set

$$(2.2.38) \quad m_-(a) = \begin{cases} m & \text{if } a \text{ is the root or of type } L, \\ m - 1 & \text{if } a \text{ is of type } R. \end{cases}$$

We then take

$$(2.2.39) \quad \gamma_a \sim \begin{cases} X & \text{if } a \text{ is the root or of type } R, \\ X^* & \text{if } a \text{ is of type } L, \end{cases}$$

independently of everything else. Let $(a, 1), \dots, (a, m_-(a))$ be the children of type L , and let $x_{(a,1)}, \dots, x_{(a,m_-(a))}$ be sampled i.i.d. and uniformly at random from $[0, x_a]$. Further, let $(x_{(a,m_-(a))+j})_{j \geq 1}$ be the (ordered) points of a Poisson point process with intensity

$$(2.2.40) \quad \rho_a(x) = \gamma_a \frac{\psi x^{\psi-1}}{x_a^\psi}$$

on $[x_a, 1]$, and the vertices $(a, m_-(a)+j)_{j \geq 1}$ have type R . The children of a are the vertices (a, j) of type L and R . Obviously, there are finitely many children of type L . Further, note

that $\psi = \frac{1}{1+2u} \in (0, 1]$, so the intensity ρ_a in (2.2.40) 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 $\{x_a\}_a$ the *Pólya-point process*. The Pólya-point tree is a multitype discrete-time branching process, where the type of a vertex a is equal to the pair (x_a, t_a) , where $x_a \in [0, 1]$ corresponds to the (rescaled) age of the vertex, and $t_a \in \{L, R\}$ 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 $\text{PA}_n^{(m,\delta)}(b)$, which is [34, Theorem 2.2]:

THEOREM 2.16 (Local weak convergence of preferential attachment models). *Fix $\delta \geq 0$ and $m \geq 1$. The preferential attachment model $\text{PA}_n^{(m,\delta)}(b)$ converges locally weakly in probability to the Pólya-point tree. The same is true for the variants of the PAM in which edges are added independently, and where edges are chosen independently conditioned on attaching to different vertices.*

Theorem 2.16 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 the result also applies to $\text{PA}_n^{(m,\delta)}$:

OPEN PROBLEM 2.1 (Local weak convergence of PAM with self-loops). *Extend the local weak convergence result in Theorem 2.16 to the model with self-loops $\text{PA}_n^{(m,\delta)}$. (This was done for $\delta = 0$ by Bollobás and Riordan [57], even though the limit was described differently.)*

Theorem 2.16 states local weak convergence in probability to the Pólya-point tree, while [34, Theorem 2.2] states local weak convergence in probability. Local weak convergence in probability can be deduced from the convergence in probability of subgraph counts in [34, Lemma 2.4]. We refrain from discussing this issue further.

We will not prove Theorem 2.16 completely. We explain the core aspects of its proof, which relies 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 $\text{PA}_n^{(m,\delta)}(b)$ that relies on independent random variables. We explain this now. We start by explaining

Pólya urn schemes. An important application of De Finetti's Theorem (see e.g., [161, Theorem 6.1]) 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

$$(2.2.41) \quad \frac{W_b(B_n)}{W_b(B_n) + W_r(R_n)}.$$

After drawing a ball, it is replaced together with a *second* ball of the same color. We shall denote this Pólya urn scheme by $((B_n, R_n))_{n \geq 1}$. Naturally, $B_n + R_n = b_0 + r_0 + n$.

In this section, we restrict to the case where there are $a_r, a_b \geq 0$ such that

$$(2.2.42) \quad W_b(k) = a_b + k, \quad W_r(k) = a_r + k,$$

i.e., both weight functions are linear with the same slope, but possibly a different intercept. The main result concerning such linear Pólya urn schemes is the following theorem:

THEOREM 2.17 (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 (2.2.42). Then, as $n \rightarrow \infty$,*

$$(2.2.43) \quad \frac{B_n}{B_n + R_n} \xrightarrow{a.s.} U,$$

where U has a Beta-distribution with parameters $a = b_0 + a_b$ and $b = r_0 + a_r$, and

$$(2.2.44) \quad \mathbb{P}(B_n = B_0 + k) = \mathbb{E} \left[\mathbb{P}(\text{Bin}(n, U) = k) \right].$$

Theorem 2.17 is highly convenient, not only since it describes the limiting relative ratios between the different types of balls, but also because it gives an explicit formula for the number of balls of a given type at each time, in terms of the limiting variable as well as binomial random variables. This will prove to be extremely useful in what follows.

Finite-graph Pólya version of $\text{PA}_t^{(m,\delta)}(b)$. We now explain the finite-graph Pólya version of $\text{PA}_t^{(m,\delta)}(b)$. We start by introducing the necessary notation. Let $(\psi_j)_{j \geq 1}$ be independent random variables with a Beta distribution with parameters $m + 2mu$, $(2j - 3)m + 2mu(j - 1)$, i.e.,

$$(2.2.45) \quad \psi_j \sim \beta(m + 2mu, (2j - 3)m + 2mu(j - 1)).$$

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

$$(2.2.46) \quad \varphi_j = \psi_j \prod_{i=j+1}^t (1 - \psi_i), \quad S_k = \sum_{j=1}^k \varphi_j.$$

Finally, let $I_k = [S_{k-1}, S_k)$. 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 $[0, S_{k-1}]$. 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 $\text{PA}_n^{(m,\delta)}(b)$ is as follows:

THEOREM 2.18 (Finite-graph Pólya version of $\text{PA}_n^{(m,\delta)}(b)$). *Fix $\delta \geq 0$ and $m \geq 1$. Then, the distribution of $\text{PA}_n^{(m,\delta)}(b)$ is the same as that of the finite-size Pólya graph of size n .*

The nice thing about Theorems 2.16–2.18 is that they also allow for an investigation of the degree distribution and various other quantities of interest in the PAM. We refer to Berger et al. [34] for details.

A related version of Theorem 2.18 for $\delta = 0$ was proved by Bollobás and Riordan [57] in terms of a pairing representation. This applies to the model $\text{PA}_n^{(m,\delta)}$.

Let us give some insight into the 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 (2.2.45) follows from tracing back the graph construction obtained in this way.

Let us now make this intuition more precise. We follow [34, Section 3.1] closely. 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

$$(2.2.47) \quad \begin{aligned} \tilde{\alpha}_n(i) \frac{1}{n-1} + (1 - \tilde{\alpha}_n(i)) \frac{d_k}{2m(n-2) + (i-1)} \\ = \frac{2m\alpha + (1-\alpha)d_k}{2m(n-1) + 2m\alpha + (1-\alpha)(i-1)}, \end{aligned}$$

while the probability that it is connected to a vertex in $[k-1]$ is equal to

$$(2.2.48) \quad \begin{aligned} \tilde{\alpha}_n(i) \frac{k-1}{n-1} + (1 - \tilde{\alpha}_n(i)) \frac{d_{<k}}{2m(n-2) + (i-1)} \\ = \frac{2m\alpha + (1-\alpha)d_{<k}}{2m(n-1) + 2m\alpha + (1-\alpha)(i-1)}, \end{aligned}$$

Thus, conditioned on connecting to $[k]$, the probability that the i th edge from n to $[n-1]$ is attached to k is $(2mu + d_k)/Z$, while the probability that the i th edge from n to $[n-1]$ is attached to $[k-1]$ is $(2mu(k-1) + d_{<k})/Z$, where $Z = 2muk + d_{<k+1}$ is the normalization constant. Note in particular that these probabilities do not depend on i , which explains the importance of $\tilde{\alpha}_n(i)$ in (2.2.33). 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 $\beta(m + 2mu, (2k-3)m + 2mu(k-1))$ distribution. We next use this to complete the proof of Theorem 2.18:

Completion of the proof of Theorem 2.18. Using the two urn process as an inductive input, we can now easily construct the Pólya graph defined in Theorem 2.18. Indeed, 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

$$(2.2.49) \quad \phi_j^{(k)} = \psi_j \prod_{i=j+1}^k (1 - \psi_i), \quad j = 1, \dots, k.$$

Here $\psi_k \sim B_k$ 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 $\phi_j^{(k)}$ can be expressed in terms of the random variables introduced in Theorem 2.18 as follows: by induction on k , it is easy to show that

$$(2.2.50) \quad S_k = \prod_{j=k+1}^n (1 - \psi_j).$$

This implies that $\phi_j^{(k)} = \psi_j / S_k$, which relates the strengths $\phi_j^{(k)}$ to the random variables defined right before Theorem 2.18, and shows that the process derived above is indeed the process given in the theorem. \square

The scale-free setting of $\delta \in (-m, 0)$. The above results only apply to $\delta \geq 0$. We close this section with an open problem, dealing with $\delta < 0$:

OPEN PROBLEM 2.2 (Local weak convergence of PAM with $\delta < 0$).
Determine the local weak limit of the preferential attachment model $\text{PA}_n^{(m, \delta)}(b)$ when $\delta \in (-m, 0)$.

We note that the variables u in (2.2.35) and χ, ψ in (2.2.36) are well defined even when $\delta \in (-m, 0)$. Further, $\psi > 1$ in this case, so that $\psi > 0$ remains to be true and the Poisson intensity in (2.2.40) remains integrable. However, now $\alpha = \delta / (2m + \delta)$ in (2.2.34) becomes negative, so we cannot think of α as the asymptotic probability of choosing a vertex uniformly at random. Note, however, that instead of attaching proportional to

the degree and uniformly, we could also attach to the degree minus m and uniformly. Possibly, this allows us to extend the above setting.

Indeed, in (2.2.47)–(2.2.48), define $d'_k = d_k - m$ to be the in-degree of vertex k at the time of the connection of the i th edge of vertex n , and $d'_{<k} = d_{<k} - (k-1)m$ the total in-degree of the vertices in $[k-1]$ at that time. Define

$$(2.2.51) \quad \tilde{\alpha}'_n(i) = \alpha \frac{m(n-1)}{m(n-1) + m\alpha + (1-\alpha)(i-1)} = \alpha + O(1/n),$$

Then the probabilities in (2.2.47)–(2.2.48) are replaced by

$$(2.2.52) \quad \begin{aligned} \tilde{\alpha}'_n(i) \frac{1}{n-1} + (1 - \tilde{\alpha}'_n(i)) \frac{d'_k}{m(n-2) + (1-\alpha)(i-1)} \\ = \frac{m\alpha + (1-\alpha)d'_k}{m(n-1) + m\alpha + (1-\alpha)(i-1)}, \end{aligned}$$

while the probability that it is connected to a vertex in $[k-1]$ is equal to

$$(2.2.53) \quad \begin{aligned} \tilde{\alpha}'_n(i) \frac{k-1}{n-1} + (1 - \tilde{\alpha}'_n(i)) \frac{d'_{<k}}{2m(n-2) + (1-\alpha)(i-1)} \\ = \frac{m\alpha + (1-\alpha)d'_{<k}}{m(n-1) + m\alpha + (1-\alpha)(i-1)}. \end{aligned}$$

Again the conditional probability of connecting to k conditioned on connecting to $[k]$, the probability that the i th edge from n to $[n-1]$ is attached to k is independent of i . Indeed, with $u' = \alpha'/(1-\alpha')$, the probability that the i th edge from n to $[n-1]$ is attached to k is $(mu' + d'_k)/Z$, while the probability that the i th edge from n to $[n-1]$ is attached to $[k-1]$ is $(mu'(k-1) + d'_{<k})/Z$, where $Z = mu'k + d'_{<k+1}$ is the normalization constant. Now the two urns start with 0 and $(k-1)m$ balls, respectively, so that the evolution of the two bins is a Pólya urn with strengths ψ'_k and $1 - \psi'_k$, where ψ'_k has the $\beta(m + mu', (k-1)m + mu'(k-1))$ distribution. This suggests that a local weak limit for $\text{PA}_n^{(m,\delta)}(b)$ as in Theorem 2.18 might extend to the setting where $\delta \in (-m, 0)$, with the above changes.

2.3. The giant component in random graphs

In this section, we investigate the connected components in the our random graphs. We start with the Erdős-Rényi random graph $\text{ER}_n(\lambda/n)$.

2.3.1. The phase transition for the Erdős-Rényi random graph. In this section, we identify when $\text{ER}_n(\lambda/n)$ has a giant component with high probability. As can be expected by Exercise 2.14, this condition has the interpretation that the branching process that describes the local weak limit of the $\text{ER}_n(\lambda/n)$ has a strictly positive survival probability. Let ζ_λ be the survival probability of the branching process with Poisson offspring with parameter λ . Then, we know that $\zeta_\lambda > 0$ precisely when $\lambda > 1$. Thus, the local weak limit of a random vertex also survives to generation k for every k fixed with probability ζ_λ . This suggests that also a giant component exists precisely when $\lambda > 1$, and that this giant component consist precisely of those vertices whose local neighborhood survive. Of course, this is not a direct consequence of the local weak convergence, as this does not rule the possibility out that all ‘surviving’ vertices are in connected components whose size is $o(n)$, as this is a global property rather than a local one. Our main theorem indeed

shows that almost all of the $n\zeta_\lambda(1 + o_{\mathbb{P}}(1))$ vertices whose local neighborhoods survive are in the *same* connected component, so, as Alon and Spencer [13] dub this, ‘there is no middle ground’:

THEOREM 2.19 (Phase transition in $\text{ER}_n(\lambda/n)$). *Let \mathcal{C}_{\max} and $\mathcal{C}_{(2)}$ be the largest and second largest components of $\text{ER}_n(\lambda/n)$.*

(a) *If $\lambda > 1$, then there exists $\zeta = \zeta_\lambda \in (0, 1]$ such that*

$$(2.3.1) \quad |\mathcal{C}_{\max}|/n \xrightarrow{\mathbb{P}} \zeta_\lambda,$$

while $|\mathcal{C}_{(2)}|/n \xrightarrow{\mathbb{P}} 0$ and $|E(\mathcal{C}_{(2)})|/n \xrightarrow{\mathbb{P}} 0$.

(b) *If $\lambda \leq 1$, then $|\mathcal{C}_{\max}|/n \xrightarrow{\mathbb{P}} 0$ and $|E(\mathcal{C}_{\max})|/n \xrightarrow{\mathbb{P}} 0$.*

By Theorem 2.1 and the fact that $\mathbb{E}[\text{Poi}(\lambda)] = \lambda$, we see that $\zeta_\lambda > 0$ precisely when $\lambda > 1$. Thus, we can summarize the convergence of $|\mathcal{C}_{\max}|/n$ in Theorem 2.19 succinctly by (2.3.1). Theorem 2.19 also implies that the giant component is *unique* in the supercritical regime.

Let us provide some intuition for Theorem 2.19. By the local weak convergence in Theorem 2.14, the Erdős-Rényi random graph converges locally weakly to a Poisson branching process with parameter λ . More precisely, let $N_k^{(n)}$ denote the number of vertices v in $[n]$ for which $|\mathcal{C}(v)| > k$. Then, by Theorem 2.14 (or Exercise 2.14),

$$(2.3.2) \quad \frac{N_k^{(n)}}{n} \xrightarrow{\mathbb{P}} \zeta(k) = \mathbb{P}(T > k),$$

and $\zeta(k) = \zeta_\lambda(k) \rightarrow \zeta_\lambda$. Thus, the upper bound in (2.3.1) immediately follows:

EXERCISE 2.16 (Upper bound on giant component $\text{ER}_n(\lambda/n)$). *Prove that $|\mathcal{C}_{\max}| \leq (\zeta_\lambda + \varepsilon)n$ whp as $n \rightarrow \infty$, by formalizing the above argument.*

The surprising effect that we see in Theorems 2.14 and 2.19 is that, for $\lambda > 1$, in fact *almost all* vertices $v \in [n]$ for which $|\mathcal{C}(v)| > k$ are in the *same* connected component. Alternatively phrased, and following Alon and Spencer [13], there is *no middle ground*.

We will not prove Theorem 2.19, but we will explain how it will follow from a rather general result about the phase transition for the configuration model. First, we will extend the analysis to inhomogeneous random graphs.

2.3.2. The phase transition for inhomogeneous random graphs. In this section, we extend Theorem 2.19 to the generalized random graph:

THEOREM 2.20 (Phase transition in generalized random graph). *Consider $\text{GRG}_n(\mathbf{w})$, where the weights \mathbf{w} satisfy Conditions 1.1(a)-(b). Let \mathcal{C}_{\max} and $\mathcal{C}_{(2)}$ be the largest and second largest components of $\text{GRG}_n(\mathbf{w})$. Then, there exists $\zeta \in [0, 1]$ such that*

$$(2.3.3) \quad |\mathcal{C}_{\max}|/n \xrightarrow{\mathbb{P}} \zeta,$$

while $|\mathcal{C}_{(2)}|/n \xrightarrow{\mathbb{P}} 0$ and $|E(\mathcal{C}_{(2)})|/n \xrightarrow{\mathbb{P}} 0$.

Theorem 2.20 is [161, Theorem 3.2]. We have chosen for a succinct description of the phase transition in Theorem 2.20. Let us next describe a condition that guarantees that $\zeta > 0$ in that setting. Define

$$(2.3.4) \quad \nu = \mathbb{E}[W^2]/\mathbb{E}[W].$$

While this may not be immediately obvious, we can interpret ν in (2.3.4) as the expected *forward* degree of a vertex. Indeed, in Theorem 2.14, the degree distribution of the root equals $\text{Poi}(W)$, where W is the limiting weight distribution. The size-biased version of this distribution equals $\text{Poi}(W^*)$, where W^* is the size-biased version of W :

EXERCISE 2.17 (Size-biased mixed Poisson random variable). *Let $X \sim \text{Poi}(W)$ be a mixed-Poisson random variable with mixing distribution W . Show that the size-biased distribution X^* of X has the same distribution as $\text{Poi}(W^*)$.*

Since

$$(2.3.5) \quad \nu = \mathbb{E}[W^*] = \mathbb{E}[W^2]/\mathbb{E}[W],$$

we thus see that the unimodular branching process that is the local weak limit of $\text{GRG}_n(\mathbf{w})$ has mean ν . We conclude that $\zeta > 0$ precisely when $\nu > 1$. That gives an explicit criterion when $\zeta > 0$ in Theorem 2.20. Again, we will not discuss the proof of Theorem 2.20, and refer instead to Theorem 2.22 below, where the relation to the phase transition for the configuration model is described.

Theorem 2.20 has been extended to general inhomogeneous random graphs by Bollobás, Janson and Riordan in [56]. There, however, the precise condition for existence of the giant component, i.e., for $\zeta > 0$, is given in terms of the operator norm of the reproduction operator, which is not always easy to obtain explicitly. We refrain from going into more details here, and refer the interested reader to [56] instead.

2.3.3. The phase transition for the configuration model. 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 1.6(a)-(b), and sometimes also Condition 1.6(c). We recall that D_n is the degree of a uniformly chosen vertex in $[n]$, i.e., $D_n = d_U$, where U is a uniformly chosen vertex from $[n]$. Equivalently,

$$(2.3.6) \quad \mathbb{P}(D_n = k) = n_k/n,$$

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 2.21 (Phase transition in $\text{CM}_n(\mathbf{d})$). *Suppose that Conditions 1.6(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 $\text{CM}_n(\mathbf{d})$.*

(a) *If $\nu = \mathbb{E}[D(D-1)]/\mathbb{E}[D] > 1$, then there exist $\eta^* \in [0, 1), \zeta \in (0, 1]$ such that*

$$(2.3.7) \quad |\mathcal{C}_{\max}|/n \xrightarrow{\mathbb{P}} \zeta,$$

$$(2.3.8) \quad v_k(\mathcal{C}_{\max})/n \xrightarrow{\mathbb{P}} p_k(1 - (\eta^*)^k) \quad \text{for every } k \geq 0,$$

$$(2.3.9) \quad |E(\mathcal{C}_{\max})|/n \xrightarrow{\mathbb{P}} \frac{1}{2}\mathbb{E}[D](1 - (\eta^*)^2).$$

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 $|E(\mathcal{C}_{\max})|/n \xrightarrow{\mathbb{P}} 0$.

EXERCISE 2.18 (Relation ξ and η^*). Show that $\zeta = \sum_{k \geq 0} p_k(1 - (\eta^*)^k) = 1 - G_D(\eta^*)$ by Theorem 2.21(a) and (b).

EXERCISE 2.19 (Relation ξ and η^* (Cont.)). Show that η^* is the extinction probability of a branching process with offspring distribution $(p_k^*)_{k \geq 0}$, while ζ is the survival probability of the unimodular Galton-Watson tree with degree distribution $(p_k)_{k \geq 0}$.

Theorem 2.21 is proved by Janson and Luczak in [185], extending work by Molloy and Reed [215, 216].

The local weak convergence in Theorem 2.11 suggests that the probability for a uniform vertex to have a large connected component is ζ . We cannot conclude this so directly, however, since 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. Further, also cycles might occur, even though they are rare and have a small probability.

Luckily, the number of available half-edges that we start with equals $\ell_n - 1$, which is very large when Conditions 1.6(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. This is again the idea behind the local weak convergence in Theorem 2.11, but for the giant component many more half-edges need to be paired to find its size of the giant. Thus, we cannot directly study the size of the giant component from local-weak convergence arguments. Let us discuss this branching process approximation, though, as well as the conditions on its survival, in more detail now.

In terms of this branching process, we can interpret ζ in Theorem 2.21 as the survival probability of the above unimodular Galton-Watson process, so that ζ satisfies

$$(2.3.10) \quad \zeta = \sum_{k=1}^{\infty} p_k(1 - (\eta^*)^k),$$

where η^* is the extinction probability of the branching process with offspring distribution $(p_k^*)_{k \geq 0}$, which satisfies

$$(2.3.11) \quad \eta^* = \sum_{k=0}^{\infty} p_k^*(\eta^*)^k.$$

Clearly, $\eta^* = 1$ precisely when

$$(2.3.12) \quad \nu = \sum_{k \geq 0} k p_k^* \leq 1.$$

Since $p_k^* = (k+1)p_{k+1}/\mathbb{E}[D]$, we can rewrite

$$(2.3.13) \quad \nu = \frac{1}{\mathbb{E}[D]} \sum_{k \geq 0} k(k+1)p_{k+1} = \mathbb{E}[D(D-1)]/\mathbb{E}[D],$$

which explains the condition on ν in Theorem 2.21(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 $1 - \eta^*$, so that the probability that at least one of them survives is close to $1 - (\eta^*)^k$. When one of the neighbors of the vertex of degree k survives, the vertex itself is part of the giant component, which explains why $v_k(\mathcal{C}_{\max})/n \xrightarrow{\mathbb{P}} p_k(1 - (\eta^*)^k)$. 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 $1 - (\eta^*)^2$. There are in total $\ell_n/2 = n\mathbb{E}[D_n]/2 \approx n\mathbb{E}[D]/2$ edges, which explains why $|E(\mathcal{C}_{\max})|/n \xrightarrow{\mathbb{P}} \frac{1}{2}\mathbb{E}[D](1 - (\eta^*)^2)$. Therefore, all results in Theorem 2.21 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 $[n]$.

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. We give three examples showing that quite different behaviors are possible when $p_2 = 1$, corresponding to the finer asymptotics of the number n_2 of vertices of degree 2.

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., [22]. This implies that $|\mathcal{C}_{\max}|/n$ converges in distribution to a non-degenerate distribution on $[0, 1]$ and not to any constant [22, 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:

EXERCISE 2.20 (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*

$$(2.3.14) \quad |\mathcal{C}(1)|/n \xrightarrow{d} T,$$

where $\mathbb{P}(T \leq x) = \sqrt{1-x}$.

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 $|\mathcal{C}_{\max}| = O_{\mathbb{P}}(n)$:

EXERCISE 2.21 (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*

$$(2.3.15) \quad |\mathcal{C}(1)|/n \xrightarrow{\mathbb{P}} 0.$$

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:

EXERCISE 2.22 (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*

$$(2.3.16) \quad |\mathcal{C}(1)|/n \xrightarrow{\mathbb{P}} 1.$$

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. This explains the condition $p_2 < 1$ in Theorem 2.21.

The phase transition for the rank-1 inhomogeneous random graphs. We can use the description of the giant component in Theorem 2.21 also to study the giant component in $\text{GRG}_n(\mathbf{w})$, $\text{CL}_n(\mathbf{w})$ and $\text{NR}_n(\mathbf{w})$ when \mathbf{w} satisfies Condition 1.1(a)-(b):

THEOREM 2.22 (Phase transition in $\text{GRG}_n(\mathbf{w})$). *Let \mathbf{w} satisfy Condition 1.1(a)-(b). Then, the results in Theorem 2.21 also hold for $\text{GRG}_n(\mathbf{w})$, $\text{CL}_n(\mathbf{w})$ and $\text{NR}_n(\mathbf{w})$.*

Theorem 2.22 is [161, Theorem 4.11]. It follows rather straightforwardly from Theorem 2.21, together with the claimed relation between $\text{GRG}_n(\mathbf{w})$ and $\text{CM}_n(\mathbf{d})$ in Theorem 1.5. The results for $\text{CL}_n(\mathbf{w})$ and $\text{NR}_n(\mathbf{w})$ follow by *asymptotic equivalence*, as discussed in more detail in [160, Sections 6.6-6.7].

2.3.4. The phase transition for the preferential attachment model. 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 independent edges as studied by Dereich and Mörters [98, 99, 100]. Here, we describe the existence of the giant component in their model.

THEOREM 2.23 (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*

$$(2.3.17) \quad \gamma \geq \frac{1}{2} \quad \text{or} \quad \beta > \frac{(1/2 - \gamma)^2}{1 - \gamma}.$$

The notation used by Dereich and Mörters [100] is slightly different from ours. They show, for example, in [98] they 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 2.21).

The more 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 [100]. We call a preferential attachment function $f: \{0, 1, 2, \dots\} \mapsto (0, \infty)$ *concave* when

$$(2.3.18) \quad f(0) \leq 1 \quad \text{and} \quad \Delta f(k) := f(k+1) - f(k) < 1 \text{ for all } k \geq 0.$$

By concavity the limit

$$(2.3.19) \quad \gamma := \lim_{n \rightarrow \infty} \frac{f(n)}{n} = \min_{k \geq 0} \Delta f(k)$$

exists. γ plays a crucial role in the analysis, as can already be observed in Theorem 2.23.

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

$$(2.3.20) \quad M(t) = \int_0^t e^{-s} \mathbb{E}[f(Z_s)] ds, \quad M(t) = \mathbb{E}[Z_t], \quad M^\tau(t) = \mathbb{E}[Z_t \mid \Delta Z_\tau = 1] - \mathbb{1}_{[\tau, \infty)}(t),$$

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

$$(2.3.21) \quad (A_\alpha g)(\tau) = \int_0^\infty g(t) e^{\alpha t} dM(t) + \int_0^\infty g(t) e^{-\alpha t} dM^\tau(t).$$

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 2.24 (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 [100, 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 multitype 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.

2.4. Connectivity of random graphs

In this section, we discuss when random graphs are fully connected. We will see that in inhomogeneous random graphs, one needs the average degree to tend to infinite for connectivity to occur, while for the configuration model, connectivity can already occur for degrees with finite mean. This makes the configuration model in many practical settings a more suitable model than its closely related inhomogeneous random graphs

relatives. For example, the Internet is by definition connected, since otherwise it could not support its services such as email. We also study the connectivity in the preferential attachment model, which also occurs with bounded average degrees as soon as the degrees are all at least 2.

2.4.1. Connectivity transition of the configuration model. A beautiful property of the configuration model is that it can lead to *connected graphs*, even in the sparse regime. This is not true for the generalized random graph, as can be easily seen:

EXERCISE 2.23 (No connectivity in $\text{GRG}_n(\mathbf{w})$). *Show that $\text{GRG}_n(\mathbf{w})$ where the weights \mathbf{w} satisfy Condition 1.1(a)-(b) is with high probability disconnected.*

For the configuration model, when all vertices have degree at least 3, this does not happen. Even when all degrees are at least 2, $\text{CM}_n(\mathbf{d})$ has a sizable probability of being connected, as the following theorem shows:

THEOREM 2.25 (Connectivity threshold for the configuration model). *Consider $\text{CM}_n(\mathbf{d})$, where \mathbf{d} satisfies Condition 1.6(a)-(b) with $p_2 = \mathbb{P}(D = 2) \in [0, 1)$. Then*

$$(2.4.1) \quad \lim_{n \rightarrow \infty} \mathbb{P}(\text{CM}_n(\mathbf{d}) \text{ is connected}) = \left(\frac{d - 2p_2}{d} \right)^{1/2}.$$

Moreover,

$$(2.4.2) \quad n - |\mathcal{C}_{\max}| \xrightarrow{d} X,$$

where $X = \sum_k kC_k$, and $(C_k)_{k \geq 1}$ are independent Poisson random variables such that

$$C_k \stackrel{d}{=} \text{Poi} \left(\frac{(2p_2)^k}{2kd^k} \right).$$

Finally,

$$(2.4.3) \quad \lim_{n \rightarrow \infty} \mathbb{E}[n - |\mathcal{C}_{\max}|] = \frac{p_2}{d - 2p_2}.$$

Theorem 2.25 is [131, Theorem 2.2], where also an extension is given for the setting where there are of the order \sqrt{n} vertices of degree 1. This is the *critical window for connectivity* of the configuration model.

Theorem 2.25 is quite relevant, as many real-world networks, such as the Internet, are in fact connected by nature. By Exercise 2.23, this does not occur for sparse generalized random graphs. For the Erdős-Rényi random graph, we need to take λ close to $\log n$ to obtain connectivity (see e.g., [123] or [160, Section 5.3]).

2.4.2. Connectivity transition of the preferential attachment model. In this section, we study the connectivity of the preferential attachment model. Clearly, when $m = 1$, $\text{PA}_n^{(m, \delta)}$ is a forest, and will be with high probability disconnected:

EXERCISE 2.24 (No connectivity in $\text{PA}_n^{(1, \delta)}$). *Show that $\text{PA}_n^{(1, \delta)}$ is with high probability disconnected.*

On the other hand, when $m \geq 2$, the probability that vertex t creates a new connected component is at most $O(t^{-m})$, which is summable, so this only occurs finitely often a.s. Further, later vertices can use their m edges to connect up previously disconnected components, so that the graph becomes more connected. This suggests that $\text{PA}_n^{(m,\delta)}$ is with high probability connected for large n , as stated in the next theorem:

THEOREM 2.26 (Connectivity for the preferential attachment model). *Fix $m \geq 2$ and $\delta > -m$. Then, $\text{PA}_n^{(m,\delta)}$ is with high probability connected for large n .*

EXERCISE 2.25 (Finite number of connected components in $\text{PA}_n^{(m,\delta)}$ for $m \geq 2$). *Fix $m \geq 2$. Prove that $\text{PA}_n^{(m,\delta)}$ has a tight number of connected components by making the argument before Theorem 2.26 precise.*

EXERCISE 2.26 (Connectivity in $\text{PA}_n^{(m,\delta)}$). *Prove Theorem 2.26 along the following strategy. Use Exercise 2.25. Since only finitely often a vertex creates a new connected component upon arrival, for $K \geq 1$ sufficiently large, there are at any time at most K connected components whp. Show that these connected components will be connected to each other whp as $n \rightarrow \infty$. Fill in the details in this argument.*

2.5. Small-world random graphs

In this section, we discuss the small-world nature of random graphs. We recall that $\text{dist}_G(U_1, U_2)$ denotes the graph distance between two vertices U_1 and U_2 chosen uniformly at random from $[n]$ in a graph G , where the graph distance between two vertices is the minimal number of edges in all paths connecting the vertices. It is possible that no path connecting the vertices exists, in which case, we define $\text{dist}_G(U_1, U_2) = +\infty$. We thus see that the size of the giant component, as studied in Sections 2.3–2.4, is highly relevant for graph distances, as $\mathbb{P}(\text{dist}_G(U_1, U_2) < \infty) \rightarrow \zeta^2$, where ζ denotes the size of the giant component:

EXERCISE 2.27 (Connectivity of uniformly chosen vertices). *Suppose we draw two vertices uniformly at random from $[n]$. Prove that the probability that the vertices are connected converges to ζ^2 when $|\mathcal{C}_{\max}|/n \xrightarrow{\mathbb{P}} \zeta > 0$, and $|\mathcal{C}_{(2)}| = o_{\mathbb{P}}(n)$.*

We conclude that it only makes sense to study small-world effects in settings where the giant component has substantial size, i.e., when $\zeta > 0$.

2.5.1. Small-world nature of inhomogeneous random graphs. In this section, we discuss typical distances in $\text{NR}_n(\mathbf{w})$. By Theorem 2.22,

$$(2.5.1) \quad \mathbb{P}(\text{dist}_{\text{NR}_n(\mathbf{w})}(U_1, U_2) = +\infty) \rightarrow 1 - \zeta^2 > 0,$$

since $\zeta < 1$ (see also Exercise 2.27). In particular, when $\zeta = 0$, which is equivalent to $\nu \leq 1$, $\mathbb{P}(\text{dist}_{\text{NR}_n(\mathbf{w})}(U_1, U_2) = +\infty) \rightarrow 1$. Therefore, in our main results, we shall condition on $\text{dist}_G(U_1, U_2) < \infty$. We will only investigate typical distances when there exists a giant component.

Distances in inhomogeneous random graphs with finite-variance weights.

Recall from Theorem 2.14 that $\text{NR}_n(\mathbf{w})$ converges locally-weakly to a unimodular branching process with mixed Poisson offspring distribution $\text{Poi}(W)$. This means that each vertex different from the root has a $\text{Poi}(W^*)$ offspring distribution, where W^* is the size-biased version of W . When $\mathbb{E}[\text{Poi}(W^*)] = \mathbb{E}[W^2]/\mathbb{E}[W] < \infty$, this offspring has finite expectation, so that generations grow exponentially. We start by investigating the behavior of $\text{dist}_{\text{NR}_n(\mathbf{w})}(U_1, U_2)$ in this case where the weights have finite variance:

THEOREM 2.27 (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 $\text{dist}_{\text{NR}_n(\mathbf{w})}(U_1, U_2) < \infty$,*

$$(2.5.2) \quad \text{dist}_{\text{NR}_n(\mathbf{w})}(U_1, U_2) / \log n \xrightarrow{\mathbb{P}} 1 / \log \nu.$$

The same result applies, under the same conditions, to $\text{CL}_n(\mathbf{w})$ and $\text{GRG}_n(\mathbf{w})$.

This is [161, Theorem 3.3]. This theorem was proved with van den Esker and Hooghiemstra in [126].

The intuition behind Theorem 2.27 is as follows. Theorem 2.14 implies 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 $(p_k^*)_{k \geq 0}$ given by $p_k^* = \mathbb{P}(\text{Poi}(W^*) = k)$ (recall Exercise 2.17). When $\nu = \sum_{k \geq 0} k p_k^* < \infty$, then the number of vertices at distance m is close to $M \nu^m$, where M is the martingale limit of Z_m / ν^m in Theorem 2.2. To know what $\text{dist}_{\text{NR}_n(\mathbf{w})}(U_1, U_2)$ 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 $Z_m \approx n$, which suggests that the relevant k is such that $\nu^m \approx n$, so that $m \approx \log_\nu n$.

While the above heuristic is quite convincing, the argument is fatally flawed. Indeed, by local weak convergence (recall Theorem 2.14), the neighborhoods of a uniform vertex are well-approximated by a branching process as long as the number of vertices found grows arbitrarily slowly with n . With considerable effort, this might be improved to a number of vertices that is much smaller than n . When the number of vertices found becomes of order n , the depletion-of-points effect already seriously kicks in. Therefore, the above approach is doomed to fail. One can, instead, prove Theorem 2.27 by using path counting arguments. See [161, Chapter 3] for more details. We will explain the intuition in more detail for the configuration model in the next section.

EXERCISE 2.28 (Typical distances in $\text{ER}_n(\lambda/n)$). *Prove that $\text{dist}_{\text{NR}_n(\mathbf{w})}(U_1, U_2) / \log n \xrightarrow{\mathbb{P}} 1 / \log \lambda$ in $\text{ER}_n(\lambda/n)$.*

Theorem 2.27 leaves open what happens when $\nu = \infty$. We can use Theorem 2.27 to show that $\text{dist}_{\text{NR}_n(\mathbf{w})}(U_1, U_2) = o_{\mathbb{P}}(\log n)$:

EXERCISE 2.29 (Typical distances when $\nu = \infty$). *Prove that $\text{dist}_{\text{NR}_n(\mathbf{w})}(U_1, U_2) / \log n \xrightarrow{\mathbb{P}} 0$ when $\nu = \infty$.*

We next investigate graph distances in rank-1 random graphs with infinite-variance degrees in more detail.

Distances in rank-1 random graphs with infinite-variance weights. We next 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 $\text{dist}_{\text{NR}_n(\mathbf{w})}(U_1, U_2) = o_{\mathbb{P}}(\log n)$ (recall Exercise 2.29). Many of our arguments also apply to the generalized random graph $\text{GRG}_n(\mathbf{w})$ and the Chung-Lu model $\text{CL}_n(\mathbf{w})$.

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 ,

$$(2.5.3) \quad c_1 x^{-(\tau-1+\delta)} \leq [1 - F_n](x) \leq c_2 x^{-(\tau-1-\delta)},$$

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 (2.5.3) 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 (2.5.3) 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 (2.5.3) 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* $\alpha \in (1/2, 1)$.

We now give simpler conditions for (2.5.3) in special cases:

EXERCISE 2.30 (Power-law tails in key example of deterministic weights). *Let \mathbf{w} be defined as $w_i = [1 - F]^{-1}(i/n)$ (recall (1.3.16)), and assume that F satisfies*

$$(2.5.4) \quad 1 - F(x) = x^{-(\tau-1)}L(x),$$

where the exponent satisfies $\tau \in (2, 3)$, and where $x \mapsto L(x)$ is slowly varying at infinity. Prove that (2.5.3) holds.

EXERCISE 2.31 (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 (2.5.4) with $\tau \in (2, 3)$, and where $x \mapsto L(x)$ is slowly varying. Prove that (2.5.3) holds whp.*

The main result on graph distances for $\tau \in (2, 3)$ is as follows:

THEOREM 2.28 (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 Condition 1.1(a)-(b) and (2.5.3). Then, conditionally on $\text{dist}_{\text{NR}_n(\mathbf{w})}(U_1, U_2) < \infty$,*

$$(2.5.5) \quad \frac{\text{dist}_{\text{NR}_n(\mathbf{w})}(U_1, U_2)}{\log \log n} \xrightarrow{\mathbb{P}} \frac{2}{|\log(\tau - 2)|}.$$

The same results apply, under the same conditions, to $\text{CL}_n(\mathbf{w})$ and $\text{GRG}_n(\mathbf{w})$.

This is [161, Theorem 3.4]. This theorem was first proved by Norros and Reittu in [226] in the setting of i.i.d. vertex weights. Theorem 2.28 implies that $\text{NR}_n(\mathbf{w})$ with \mathbf{w} as in (1.3.16), for $\tau \in (2, 3)$, is an ultra-small world when (2.5.3) is satisfied. Again, a proof can be given using path counting techniques.

2.5.2. Small-world nature of configuration models. We next discuss the small-world nature of configuration models. We start by discussing the case of finite-variance degrees.

Distances in configuration models with finite-variance degrees. In the CM with finite-variance degrees, we again see that the typical distances are logarithmic:

THEOREM 2.29 (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 Condition 1.6(a)-(c) with $\nu > 1$, conditionally on $\text{dist}_{\text{CM}_n(\mathbf{d})}(U_1, U_2) < \infty$,*

$$(2.5.6) \quad \text{dist}_{\text{CM}_n(\mathbf{d})}(U_1, U_2) / \log n \xrightarrow{\mathbb{P}} 1 / \log \nu.$$

Theorem 2.29 is [161, Theorem 5.1] and was first proved with Hooghiemstra and Van Mieghem in [162] for the special case of i.i.d. degrees. Theorem 2.29 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 2.27. We note that stronger results than Theorem 2.29 have been obtained in the literature. Indeed, in [162], it is proved that the re-centered graph distances $\text{dist}_{\text{CM}_n(\mathbf{d})}(U_1, U_2) - \log_\nu n$ form a tight sequence of random variables in the case of i.i.d. degrees.

Let us sketch the proof of Theorem 2.29 in some more detail. The local-weak convergence result in Theorem 2.11 can be extended to neighborhoods of radius m that grows with n . We let $k_n = \frac{1}{2} \log_\nu n$ (fingers crossed!), so that $|\partial B_{U_1}(k)| \approx M_1 \nu^{k_n} = M_1 \sqrt{n}$. Then we grow the neighborhood of U_2 up to the moment that it connects to $\partial B_{U_1}(m)$. We call such a half-edge a *collision half-edge*. The graph distance between U_1 and U_2 will then be $k + 1 + l$, where l is the graph distance between U_2 and the vertex incident to the collision half-edge. The question is thus how large such l typically is.

Each time that we pair a half-edge, we create an edge between $\partial B_{U_1}(k)$ and U_2 with probability roughly $|\partial B_{U_1}(k)| / \ell_n \approx M_1 / (\mathbb{E}[D] \sqrt{n})$. Thus, we need to pair of order \sqrt{n} half-edges to have a reasonable chance of success. Since $|\partial B_{U_2}(l)| \approx M_2 \nu^l$, this means that $l \approx \frac{1}{2} \log_\nu n$, so that the graph distance is roughly $k_n + l = \log_\nu n$, as indicated in Theorem 2.29. Of course, the above is not rigorous, since a rigorous argument requires detailed estimates on the growth of $|\partial B_{U_1}(k)|$ and $|\partial B_{U_2}(k)|$ for $k \approx \frac{1}{2} \log_\nu n$. However, this proof can be made rigorous, and even suggests that $\text{dist}_{\text{CM}_n(\mathbf{d})}(U_1, U_2) - \log_\nu n$ is a *tight* sequence of random variables. The latter is indeed proved in [162] for the special case of i.i.d. degrees. Interestingly, this tight random variable does *not* converge in distribution, giving a rare example of a tight sequence of random variables that does not converge weakly.

Distances in configuration models with 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 (2.5.3). 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 ,

$$(2.5.7) \quad c_1 x^{-(\tau-1+\delta)} \leq [1 - F_n](x) \leq c_2 x^{-(\tau-1-\delta)},$$

where the upper bound holds for every $x \geq 1$, while the lower bound is only required to hold for $1 \leq x \leq n^\alpha$ for some $\alpha > 1/2$. The typical distance of $\text{CM}_n(\mathbf{d})$ is identified in the following theorem:

THEOREM 2.30 (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 1.6(a)-(b) and (2.5.7). Then, conditionally on $\text{dist}_{\text{CM}_n(\mathbf{d})}(U_1, U_2) < \infty$,*

$$(2.5.8) \quad \frac{\text{dist}_{\text{CM}_n(\mathbf{d})}(U_1, U_2)}{\log \log n} \xrightarrow{\mathbb{P}} \frac{2}{|\log(\tau - 2)|}.$$

Theorem 2.30 is [161, Theorem 5.3]. Theorem 2.30 is similar in spirit to Theorem 2.28 for $\text{NR}_n(\mathbf{w})$.

Recently, in work with Komjáthy [170], a much stronger version of Theorem 2.30 has been proved. The extension is in the fact that the result also shows that $\text{dist}_{\text{CM}_n(\mathbf{d})}(U_1, U_2) - 2 \log \log n / |\log(\tau - 2)|$ is a *tight* sequence of random variables when (2.5.7) holds. Interestingly, this random variable does *not* converge in distribution. Further, in [170] versions of Theorem 2.30 are proved when the degrees are *truncated*, which means that (2.5.7) is true for all $x \leq n^{\beta_n}$, while $[1 - F_n](x) = 0$ for $x > n^{\beta_n}$. This means that the maximal degree is of order n^{β_n} rather than $n^{1/(\tau-1)+o(1)}$. This truncation affects the distances, in the sense that an extra term of order $1/\beta_n$ needs to be added. We refer the interested reader to [170] for more details.

Let us sketch the proof of Theorem 2.30 in some more detail. We follow the sketch of the proof of Theorem 2.29. The local-weak convergence result in Theorem 2.11 can be extended to neighborhoods of radius m that grow slowly with n . However, now the branching process that approximates the neighborhoods of U_1 and U_2 has *infinite mean*. Recall Theorem 2.4, which shows that under the condition (2.1.16), $\log |\partial B_{U_1}(k)| \approx Y_1(\tau - 2)^{-k}$ for some random variable Y_1 . Take $k_n = (\log \log \sqrt{n} + a) / |\log(\tau - 2)|$ for some a chosen later. Then $|\partial B_{U_1}(k)| \approx e^{e^a Y_1 \log \sqrt{n}}$, so taking $a = \log(1/Y_1)$ makes $|\partial B_{U_1}(k)| \approx \sqrt{n}$.

Then we again grow the neighborhood of U_2 up to the moment that it connects to $\partial B_{U_1}(k)$. The graph distance between U_1 and U_2 will then be $k + 1 + l$, where l is the graph distance between U_2 and the vertex incident to the collision half-edge. Now, we see that since also $\log |\partial B_{U_2}(k)| \approx Y_2(\tau - 2)^{-k}$, instead we have $l \approx \log \log \sqrt{n} / |\log(\tau - 2)|$, so that the typical distance is roughly $2 \log \log \sqrt{n} / |\log(\tau - 2)|$, as Theorem 2.30 proves. Again, this proof can be made rigorous, and even suggests that $\text{dist}_{\text{CM}_n(\mathbf{d})}(U_1, U_2) - 2 \log \log n / |\log(\tau - 2)|$ is a *tight* sequence of random variables. The latter is indeed again proved in [165] for the special case of i.i.d. degrees, and in [170] under more general restrictions. This tight random variable once more does *not* converge in distribution.

Let us present an alternative viewpoint for Theorem 2.30 that turns out to be quite helpful, also in understanding related result. By local-weak convergence (Theorem 2.11), $\partial B_{U_1}(k)$ can whp be successfully coupled to the individuals in the k th generation of an infinite-mean branching process. Let $D_{\max}(k)$ be the maximal degree of any vertex in $\partial B_{U_1}(k)$. Then, $D_{\max}(k) \xrightarrow{\mathbb{P}} \infty$ when k grows. This leads us to the following question. Suppose that we have a vertex v of degree d which is quite large. What is the maximal degree of any vertex to which v is connected?

For this question, we rely on the following lemma, which is of independent interest:

LEMMA 2.31 (Connectivity in $\text{CM}_n(\mathbf{d})$). *For any two sets of vertices $A, B \subseteq [n]$,*

$$(2.5.9) \quad \mathbb{P}(A \text{ not directly connected to } B) \leq e^{-d_A d_B / (2\ell_n)},$$

where, for any $A \subseteq [n]$,

$$(2.5.10) \quad d_A = \sum_{i \in A} d_i$$

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

$$(2.5.11) \quad 1 - \frac{d_B}{\ell_n - 2k + 1} \leq 1 - \frac{d_B}{\ell_n}.$$

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

$$(2.5.12) \quad \mathbb{P}(A \text{ not directly connected to } B) \leq \left(1 - \frac{d_B}{\ell_n}\right)^{d_A/2} \leq e^{-d_A d_B / (2\ell_n)},$$

where we have used that $1 - x \leq e^{-x}$. □

We apply the Connectivity Lemma 2.31 to investigate what the maximal degree is of a neighbor of a vertex v_d of large degree d is. Fix q and let $A = \{v_d\}$ and $B = \{v : d_v \geq q\}$. By assumption (2.5.7),

$$(2.5.13) \quad d_B \geq q \#\{v : d_v \geq q\} = qn[1 - F_n](q) \geq c_1 q n q^{-(\tau-1+\delta)} = c_1 n q^{-(\tau-2+\delta)}.$$

Thus,

$$(2.5.14) \quad \mathbb{P}(v_d \text{ not directly connected to } B) \leq e^{-d_A d_B / (2\ell_n)} \leq e^{-c_1 n d q^{-(\tau-2+\delta)} / [2\ell_n]} \approx e^{-c d q^{-(\tau-2+\delta)}},$$

which is quite small when $q \leq d^{1/(\tau-2+\delta)}$. Thus, a vertex of high degree d is whp connected to a vertex of degree roughly $d^{1/(\tau-2)}$. We conclude that after some k_n steps, the vertex of degree d is connected to a vertex of degree of order $d^{(\tau-2)^{-k_n}}$, which becomes of order n^a when $k_n \approx \log \log n / |\log(\tau-2)|$. Thus, a vertex of large degree d will be connected to the maximum-degree vertex in roughly $\log \log n / |\log(\tau-2)|$ steps. This is true for a high-degree vertex in $\partial B_{U_1}(k)$ as well as in $\partial B_{U_2}(k)$, and whp these sets contain vertices of quite large degree. This suggests that the graph distance is roughly $2 \log \log n / |\log(\tau-2)|$, as well as that shortest paths mainly pass through higher and higher degree vertices to go to the maximal-degree vertex, before again passing through smaller and smaller degree vertices to go to the destination. Of course, filling in the details is tricky, and we refer to [170] where this is worked out in detail.

Diameter in configuration models. We continue the discussion of distances in the configuration model by investigating the diameter in the model. Before stating the main result, we introduce some notation. Let $G_D^*(x)$ be defined as the probability generating function of $p^* = (p_k^*)_{k \geq 0}$ defined in (2.2.16). We recall that η^* is the extinction probability of the branching process with offspring distribution p^* defined in (2.3.11) and further define

$$(2.5.15) \quad \mu = G_D^*(\eta^*) = \sum_{k=1}^{\infty} k(\eta^*)^{k-1} p_k^*.$$

When $\eta^* < 1$, we also have that $\mu \leq 1$. In turn, $\eta^* < 1$ when $\nu > 1$, by the phase transition of the branching process. Then, the main result for the logarithmic asymptotics of the diameter of $\text{CM}_n(\mathbf{d})$ is as follows:

THEOREM 2.32 (Diameter of the configuration model). *Let Condition 1.6(a)-(b) hold, and suppose that $\nu > 1$. Assume that $n_1 = 0$ when $p_1 = 0$, and that $n_2 = 0$ when $p_2 = 0$. Then,*

$$(2.5.16) \quad \frac{\text{diam}(\text{CM}_n(\mathbf{d}))}{\log n} \xrightarrow{\mathbb{P}} \frac{1}{\log \nu} + \frac{(2 - \mathbb{1}_{\{p_1=0\}} - \mathbb{1}_{\{p_2=0\}})}{|\log \mu|}.$$

Theorem 2.32 was first proved by Fernholz and Ramachandran [132], see also [161, Theorem 5.2]. We note that, by Theorem 2.29 and Theorem 2.32, 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 2.25). We also remark that Theorem 2.32 applies not only to the finite-variance degree 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 2.32 implies that the diameter is $o_{\mathbb{P}}(\log n)$. Theorem 2.32 also applies to uniform random graphs with a given degree sequence in the finite-variance setting.

Interestingly, when $\tau \in (2, 3)$ and there are no vertices of degrees 1 and 2, then also the diameter in $\text{CM}_n(\mathbf{d})$ turns out to be doubly logarithmic:

THEOREM 2.33 (Diameter of the scale-free configuration model). *Let Condition 1.6(a)-(b) hold and assume that (2.5.7) holds. Assume further that $n_1 = n_2 = 0$. Then,*

$$(2.5.17) \quad \frac{\text{diam}(\text{CM}_n(\mathbf{d}))}{\log \log n} \xrightarrow{\mathbb{P}} \frac{2}{\log(d_{\min} - 1)} + \frac{2}{|\log(\tau - 2)|}.$$

Theorem 2.33 shows that $\text{CM}_n(\mathbf{d})$ for which (2.5.7) holds and that do not contain vertices of degree 1 and 2 are extremely small-world random graphs: each pair of vertices is $O(\log \log n)$ apart. Theorem 2.33 is proved with Caravenna, Garavaglia in [75].

2.5.3. Small-world nature of preferential attachment models. In this section, we discuss the small-world properties of preferential attachment models.

Distances in scale-free preferential attachment trees. We start by investigating distances in scale-free trees, arising for $m = 1$:

THEOREM 2.34 (Typical distance in scale-free trees). *Fix $m = 1$ and $\delta > -1$. Then*

$$(2.5.18) \quad \frac{\text{dist}_{\text{PA}_n^{(1,\delta)}}(U_1, U_2)}{\log n} \xrightarrow{\mathbb{P}} \frac{2(1+\delta)}{(2+\delta)},$$

while, with γ the non-negative solution of

$$(2.5.19) \quad \gamma + (1+\delta)(1+\log \gamma) = 0,$$

it holds

$$(2.5.20) \quad \frac{\text{diam}(\text{PA}_n^{(1,\delta)})}{\log n} \xrightarrow{\mathbb{P}} \frac{2(1+\delta)}{(2+\delta)\gamma}.$$

This is [161, Theorems 7.1 and 7.2].

Of course, the setting of scale-free trees is rather special. We next present the available results for the non-tree setting.

Logarithmic distances in preferential attachment 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 2.35 (A $\log t$ bound for distances in PAMs). *Fix $m \geq 1$ and $\delta > 0$. For $\text{PA}_n^{(m,\delta)}$ there exist $0 < a_1 < a_2 < \infty$ such that, as $n \rightarrow \infty$,*

$$(2.5.21) \quad \mathbb{P}(a_1 \log n \leq \text{dist}_{\text{PA}_n^{(m,\delta)}}(U_1, U_2) \leq a_2 \log n) = 1 - o(1).$$

Further, there exist $0 < b_1 < b_2 < \infty$ such that, as $n \rightarrow \infty$,

$$(2.5.22) \quad \mathbb{P}(b_1 \log n \leq \text{diam}(\text{PA}_n^{(m,\delta)}) \leq b_2 \log n) = 1 - o(1).$$

This is [161, Theorems 7.7 and 7.8]. The result is originally proved in work with Dommers and Hooghiemstra [111]. The precise asymptotics of the typical distances and diameter in $\text{PA}_n^{(m,\delta)}$ is unknown when $\delta > 0$:

OPEN PROBLEM 2.3 (Typical distances in $\text{PA}_n^{(m,\delta)}$ with $\delta > 0$). *Show that, for $\delta > 0$ and $m \geq 2$, there exist $0 < a < b < \infty$,*

$$(2.5.23) \quad \frac{\text{dist}_{\text{PA}_n^{(m,\delta)}}(U_1, U_2)}{\log n} \xrightarrow{\mathbb{P}} a, \quad \frac{\text{diam}(\text{PA}_n^{(m,\delta)})}{\log n} \xrightarrow{\mathbb{P}} b,$$

Distances in preferential attachment models with $m \geq 2$ and $\delta = 0$. For $\delta = 0$, $\tau = 3$, and distances turn out to grow as $\log n / \log \log n$:

THEOREM 2.36 (Diameter of $\text{PA}_n^{(m,\delta)}$ for $\delta = 0$). *Fix $m \geq 2$ and $\delta = 0$. As $n \rightarrow \infty$,*

$$(2.5.24) \quad \text{dist}_{\text{PA}_n^{(m,\delta)}}(U_1, U_2) \frac{\log \log n}{\log n} \xrightarrow{\mathbb{P}} 1,$$

and

$$(2.5.25) \quad \text{diam}(\text{PA}_n^{(m,\delta)}) \frac{\log \log n}{\log n} \xrightarrow{\mathbb{P}} 1.$$

Theorem 2.36 shows that distances for $\tau = 3$ are similar in $\text{PA}_n^{(m,\delta)}$ as in $\text{NR}_n(\mathbf{w})$ and $\text{CM}_n(\mathbf{d})$ for the appropriate degrees. 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 generally known, even though Dereich, Mörters and Mönch [97] provide some detailed results for the rank-1 setting as well as certain preferential attachment models.

Doubly logarithmic distances in preferential attachment 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 2.37 ($\log \log n$ asymptotics for the distances for $\delta < 0$). *Fix $m \geq 2$ and assume that $\delta \in (-m, 0)$. As $n \rightarrow \infty$,*

$$(2.5.26) \quad \frac{\text{dist}_{\text{PA}_n^{(m,\delta)}}(U_1, U_2)}{\log \log n} \xrightarrow{\mathbb{P}} \frac{4}{|\log(\tau - 2)|},$$

while

$$(2.5.27) \quad \frac{\text{diam}(\text{PA}_n^{(m,\delta)})}{\log \log n} \xrightarrow{\mathbb{P}} \frac{2}{\log m} + \frac{4}{|\log(\tau - 2)|}.$$

Theorem 2.37 is proved with Caravenna and Garavaglia in [75], extending earlier work with Dommers and Hooghiemstra [111] and by Dereich, Mönch and Mörters [96].

Interestingly, the term $4/|\log(\tau - 2)|$ appearing in Theorem 2.37 replaces the term $2/|\log(\tau - 2)|$ in Theorems 2.30 and 2.33 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 2.31). 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 the next exercises:

EXERCISE 2.32 (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,*

$$(2.5.28) \quad \lim_{n \rightarrow \infty} \mathbb{P}(\text{dist}_{\text{PA}_n^{(m,\delta)}}(v_i, v_j) \leq 2) = 1.$$

EXERCISE 2.33 (Early vertices are not at distance 2 when $\delta > 0$). Let $\delta > 0$ and $m \geq 2$. Show that

$$(2.5.29) \quad \limsup_{i,j \rightarrow \infty} \lim_{n \rightarrow \infty} \mathbb{P}(\text{dist}_{\text{PA}_n^{(m,\delta)}}(v_i, v_j) = 2) = 0.$$

Let us give some more details about this effect, as it is quite a bit more subtle than for $\text{CM}_n(\mathbf{d})$. Indeed, in $\text{CM}_n(\mathbf{d})$ for $\tau \in (2, 3)$, a vertex v_d of degree d with d large, is whp connected to a vertex of degree $d^{1/(\tau-2)+o(1)}$. Since $\tau \in (2, 3)$, also $1/(\tau-2) > 1$, so that vertices of high degree are connected to vertices of even higher degrees. For $\text{PA}_n^{(m,\delta)}$ with $\delta \in (-m, 0)$, a related estimate holds. Fix n large, and consider a vertex i satisfying $D_i(n/2) = d$ for some d large. Thus, the vertex not only has large degree at time n , but it already had large degree at half the time. Since degrees grow like powers of time, the degree $D_i(n/2)$ can be expected to be a constant times $D_i(n)$, so halving time should not matter a great deal. However, it gives us a great amount of independence. Indeed, consider the set $B = \{j: D_j(n/2) \geq q\}$ of vertices that have degree at least q at time $n/2$. We now aim to find out for which q , the vertex i is whp connected to a vertex in B . For this, we note that the first two edges of a vertex $j \in [n] \setminus [n/2]$ connects to *both* i and a vertex in B with probability at least

$$(2.5.30) \quad \frac{(D_i(n/2) - \delta)(D_B(n/2) - \delta|B|)}{[n(2m + \delta)]^2} = C \frac{dD_B(n/2)}{n^2},$$

where, for a set A , we define $D_A(n) = \sum_{a \in A} D_a(n)$. It remains to study how large $D_B(n/2)$ is. By Theorem 1.9, the degree distribution is close to $p_k \sim k^{-\tau}$. This suggests that $D_B(n/2) \sim nq^{-(\tau-2)}$. As a result,

$$(2.5.31) \quad \frac{(D_i(n/2) - \delta)(D_B(n/2) - \delta|B|)}{[n(2m + \delta)]^2} = C \frac{dq^{-(\tau-2)}}{n}.$$

Thus, the probability that *none* of the vertices in $j \in [n] \setminus [n/2]$ connects to *both* i and a vertex in B is roughly equal to

$$(2.5.32) \quad \left(1 - \frac{Cdq^{-(\tau-2)}}{n}\right)^{n/2} \leq e^{-Cdq^{-(\tau-2)}/2},$$

which is quite small when d is large and $q = d^{1/(\tau-2+\delta)}$. Thus, now vertices of degree d are whp in two steps connected to vertices of degree $d^{1/(\tau-2)+o(1)}$. Of course, the fact that we now need to deal with the degrees at time $n/2$ rather than those at time n is a nuisance, but can be dealt with. This explains why distances in $\text{PA}_n^{(m,\delta)}$ with $\delta < 0$ are about twice as large as those in $\text{CM}_n(\mathbf{d})$ with a similar degree distribution. This explains the extra factor 2 in Theorem 2.37 compared to Theorem 2.30.

While we know quite a bit about distances in preferential attachment models, virtually nothing is known about their fluctuations. This suggests the following open problem:

OPEN PROBLEM 2.4 (Fluctuations of typical distances and diameter of $\text{PA}_n^{(m,\delta)}$). For $m \geq 2$. How do the fluctuations of $\text{dist}_{\text{PA}_n^{(m,\delta)}}(U_1, U_2)$ and $\text{diam}(\text{PA}_n^{(m,\delta)})$ behave?

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. In the remainder of this text, such universality (or sometimes the lack of it) are at the center of our attention.

2.6. Further technical ingredients

In this section, we state and prove some useful results that we will frequently rely on. We start by discussing truncations of moments of power-law distributions in Section 2.6.1. We continue in Section 2.6.2 by discussing a highly useful *super-martingale inequality*, and the *size-biased re-orderings of the degree distribution* for $\text{CM}_n(\mathbf{d})$ in Section 2.6.3. We close in Section 2.6.4 by explaining a highly useful coupling of the forward degrees in the configuration model.

2.6.1. Truncated moments of power-law distributions. 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 2.38 (Truncated moments). *Let D be a random variable whose distribution function satisfies that for every $x \geq 1$,*

$$(2.6.1) \quad 1 - F_D(x) \leq C_D x^{-(\tau-1)}.$$

Then, there exists a constant C such that, for $a < \tau - 1$,

$$(2.6.2) \quad \mathbb{E}[D \mathbb{1}_{\{D > \theta\}}] \leq C \theta^{a-(\tau-1)}.$$

PROOF. We note that for any probability distribution $(q_k)_{k \geq 0}$ on the non-negative integers, we have the *partial summation identity*

$$(2.6.3) \quad \sum_{k \geq 0} q_k f(k) = f(0) + \sum_{\ell \geq 1} q_{\geq \ell} [f(\ell) - f(\ell - 1)],$$

provided that either $[f(\ell) - f(\ell - 1)]_{q_{\geq \ell}}$ is absolutely summable, or $k \mapsto f(k)$ is either non-decreasing or non-increasing. Indeed,

$$(2.6.4) \quad \sum_{k \geq 0} q_k f(k) = f(0) + \sum_{k=0}^{\infty} q_k [f(k) - f(0)] = f(0) + \sum_{k=0}^{\infty} q_k \sum_{\ell=1}^k [f(\ell) - f(\ell - 1)],$$

and the claim follows by interchanging the summation order, which is allowed by Fubini's Theorem for non-negative functions (see [151, Section 3.6, Theorem B]) when $k \mapsto f(k)$ is non-decreasing, and by Fubini's Theorem [151, Section 3.6, Theorem C] when $[f(\ell) - f(\ell - 1)] \mathbb{1}_{\{0 \leq \ell \leq k\}} q_k$ is absolutely summable, which, by non-negativity of q_k , is equivalent to the absolute summability of $[f(\ell) - f(\ell - 1)]_{q_{\geq \ell}}$.

When $D \geq 0$, using (2.6.1) and (2.6.3), for $a < \tau - 1$,

$$(2.6.5) \quad \begin{aligned} \mathbb{E} [D^a \mathbb{1}_{\{D > \ell\}}] &= [\ell]^a \mathbb{P}(D \geq \ell) + \sum_{k > \ell} [k^a - (k-1)^a] \mathbb{P}(D \geq k) \\ &\leq C_D [\ell]^{a-(\tau-1)} + aC_D \sum_{[\ell]+1}^{\infty} k^{a-1} (k+1)^{-(\tau-1)} \leq C_{a,\tau} \ell^{a-(\tau-1)}. \end{aligned}$$

as required. \square

We continue to discuss truncated higher moments:

LEMMA 2.39 (Truncated moments of D). *Let D be a random variable whose distribution function satisfies that for every $x \geq 1$ and some $\tau > 2$,*

$$(2.6.6) \quad 1 - F_D(x) \leq C_D x^{-(\tau-1)}.$$

Then, there exist constants $C_{a,\tau} = C_{a,\tau}(C_D) > 0$ such that, as $\ell \rightarrow \infty$,

$$(2.6.7) \quad \mathbb{E} [D^a \mathbb{1}_{\{D \leq \ell\}}] \leq \begin{cases} C_{a,\tau} \ell^{a-(\tau-1)} & \text{when } a > \tau - 1, \\ C_{\tau-1,\tau} \log \ell & \text{when } a = \tau - 1. \end{cases}$$

Finally, when (2.6.6) holds with $\tau = 5$ and, for all $x \geq 1$,

$$(2.6.8) \quad 1 - F_D(x) \geq c_D x^{-(\tau-1)}$$

then there exists a constant $c_{3,5} = c_{3,5}(c_D) > 0$ such that, as $\ell \rightarrow \infty$,

$$(2.6.9) \quad \mathbb{E} [D^4 \mathbb{1}_{\{D \leq \ell\}}] \geq c_{3,5} \log \ell.$$

PROOF. We start by upper bounding the truncated moments of D . We rewrite, using (2.6.3) and with $f(k) = k^a \mathbb{1}_{\{k \leq \ell\}}$,

$$(2.6.10) \quad \begin{aligned} \mathbb{E} [D^a \mathbb{1}_{\{D \leq \ell\}}] &= \sum_{k=0}^{\infty} f(k) \mathbb{P}(D = k) = \sum_{k=1}^{\infty} [f(k) - f(k-1)] \mathbb{P}(D \geq k) \\ &\leq \sum_{k=1}^{[\ell]} [k^a - (k-1)^a] \mathbb{P}(D \geq k). \end{aligned}$$

Using $k^a - (k-1)^a = a \int_{k-1}^k x^{a-1} dx \leq ak^{a-1}$, we arrive at

$$(2.6.11) \quad \mathbb{E} [D^a \mathbb{1}_{\{D \leq \ell\}}] \leq aC_D \sum_{k=1}^{[\ell]} k^{a-1} k^{-(\tau-1)} \leq aC_D \sum_{k=1}^{[\ell]+1} k^{a-\tau}.$$

Note that $k \mapsto k^{a-\tau}$ is either increasing or decreasing. Hence,

$$(2.6.12) \quad \sum_{k=1}^{[\ell]+1} k^{a-\tau} \leq \int_1^{\ell+2} k^{a-\tau} dk.$$

For $a > \tau - 1$,

$$(2.6.13) \quad \int_1^{\ell+2} k^{a-\tau} dk \leq \frac{2}{a+2-\tau} \ell^{a-(\tau-1)},$$

whereas for $a = \tau - 1$,

$$(2.6.14) \quad \int_1^{\ell+2} k^{a-\tau} dk \leq 2 \log \ell.$$

For (2.6.9), we compute with $f(k) = k^4$,

$$(2.6.15) \quad \begin{aligned} \mathbb{E} [D^4 \mathbb{1}_{\{D \leq \ell\}}] &= \sum_{k=1}^{\infty} [f(k) - f(k-1)] \sum_{l=k}^{\ell} \mathbb{P}(D = l) \\ &\leq \sum_{k=3}^{\infty} [k^4 - (k-1)^4] \sum_{l=k}^{\ell} \mathbb{P}(D = l). \end{aligned}$$

We bound this from below by

$$(2.6.16) \quad \mathbb{E} [D^4 \mathbb{1}_{\{D \leq \ell\}}] \geq \sum_{k=0}^{\sqrt{\ell}} [k^4 - (k-1)^4] [\mathbb{P}(D \geq k) - \mathbb{P}(D \geq \ell)].$$

For $\tau = 5$, the contribution due to $\mathbb{P}(D \geq \ell)$ is at most

$$(2.6.17) \quad \ell^2 \mathbb{P}(D \geq \ell) \leq C_D \ell^{-3/2} = o(1),$$

while the contribution due to $\mathbb{P}(D \geq k)$ and using $[k^4 - (k-1)^4] \geq k^3$ for every $k \geq 4$, is at least

$$(2.6.18) \quad c_D \sum_{k=4}^{\sqrt{\ell}} k^{-1} \geq c_D \int_4^{\sqrt{\ell}+1} \frac{dx}{x} = c_D [\log(\sqrt{\ell} + 1) - \log 4],$$

which proves the claim by choosing the constant $c_{3,5}$ correctly. \square

2.6.2. Martingale techniques. In this text, we often rely on martingale techniques. We state a useful *super-martingale inequality*:

LEMMA 2.40 (Super-martingale inequality [235, Lemma 2.54.5]). *For any super-martingale $(M(t))_{t \geq 0}$, with $M(0) = 0$,*

$$(2.6.19) \quad \mathbb{P}\left(\sup_{s \leq t} |M(s)| > 3\epsilon\right) \leq 3\mathbb{E}[|M(t)|] \leq 3\left(|\mathbb{E}[M(t)]| + \sqrt{\text{Var}(M(t))}\right).$$

Lemma 2.40 allows us to prove *concentration result* for various stochastic processes used in this book. We will mainly employ Lemma 2.40 in settings where $M(t) = M_n(t)$, where $M_n(t)$ is a (super)-martingale with $\mathbb{E}[M_n(t)]$, $\text{Var}(M_n(t)) = o(1)$ for every $t > 0$, in which case it implies that also $\sup_{t \leq T} |M_n(t)| \xrightarrow{\mathbb{P}} 0$ for every $T > 0$.

2.6.3. Size-biased reordering and general cluster weights. In this section, we investigate the vertices found in an exploration of the CM in more detail. This will be useful throughout the remainder of this text. We note that, as long as we do not close cycles, the vertices are found in a *size-biased order*. Suppose that we start with a vertex U^* that is chosen such that $\mathbb{P}(U^* = j) = d_j / \ell_n$. After this, we construct the neighborhoods as in the previous section. This means that we see vertices in an order that is random, and vertices that have a larger degree are more likely to be found earlier. This is formalized in the following definition:

DEFINITION 2.41 (Size-biased reordering). *Given a sequence of weights $\mathbf{d} = (d_i)_{i \in [n]}$, the size-biased reordering of $[n]$ is the random vector of vertices $(v_i)_{i \in [n]}$ such that*

$$(2.6.20) \quad \mathbb{P}(v_i = j \mid v_1, \dots, v_{i-1}) = \frac{d_j}{\sum_{\ell \notin \{v_1, \dots, v_{i-1}\}} d_\ell}.$$

The following exercise proves that, when the degrees are not all equal, $\mathbb{E}[d_{v_1}] > \mathbb{E}[D_n]$. Since also

$$(2.6.21) \quad \sum_{i \in [n]} d_{v_i} = \sum_{i \in [n]} d_i = \ell_n,$$

this suggests that the earlier vertices have a relatively high degree, while the later vertices have a relatively high degree.

EXERCISE 2.34 (Expected degree of first vertex in size-biased reordering). *Let $(v_i)_{i \in [n]}$ be the size-biased reordering of the vertices. Show that*

$$(2.6.22) \quad \mathbb{E}[d_{v_1}] = \frac{\mathbb{E}[D_n^2]}{\mathbb{E}[D_n]},$$

and conclude that $\mathbb{E}[d_{v_1}] > \mathbb{E}[D_n]$, except when $d_i = d$ for all $i \in [n]$ and some $d \geq 1$.

The following proposition investigates properties of the degrees of vertices found in the size-biased reordering:

PROPOSITION 2.42. *Suppose $f: \mathbb{N} \rightarrow \mathbb{R}$ is a function with $n^{-1} \sum_{i \in [n]} f(d_i) d_i \rightarrow \sigma_f$, where $(d_i)_{i \in [n]}$ satisfies Conditions 1.6(a)-(c). Suppose $(v_i)_{i \in [n]}$ is the size-biased ordering of $[n]$ with size $(d_i/\ell_n)_{i \in [n]}$. Consider $H_n(u) := n^{-\alpha} \sum_{i=1}^{\lfloor un^\alpha \rfloor} f(d_{v_i})$ for some $\alpha \in (0, \frac{2}{3}]$ and $u > 0$. Then,*

$$(2.6.23) \quad \sup_{u \leq t_n} \left| H_n(u) - \frac{\sigma_f}{\mu} u \right| = O_{\mathbb{P}}(a_n \vee b_n \vee n^{-1/3}),$$

where $a_n = (t_n n^{-\alpha} \max_{i \in [n]} f(d_i))^{1/2}$ and $b_n = d_{\max} t_n^2 n^{\alpha-1}$, as long as $t_n = o(n^{\beta/2})$ for some β satisfying $\max\{\alpha/2, 2\alpha - 1\} \leq \beta < \alpha$.

PROOF. We use a clever randomization trick, together with a martingale argument. Consider n independent exponential random variables $T_i \sim \exp(d_i/\ell_n)$. Define

$$(2.6.24) \quad \tilde{H}_n(u) = n^{-\alpha} \sum_{i \in [n]} f(d_i) \mathbb{1}_{\{T_i \leq n^\alpha u\}}.$$

Therefore, $\tilde{H}_n(u) = \sum_{i=1}^{N(n^\alpha u)} f(d_i) = H_n(N(n^\alpha u))$ where $N(u) := \#\{j: T_j \leq un^\alpha\}$.

Consider $Y_0(s) = n^{-\beta} (N(sn^\alpha) - sn^\alpha)$ for some $\max\{\alpha/2, 2\alpha - 1\} \leq \beta < \alpha$. Define $\mathcal{V}'_s := \{j: T_j \leq sn^\alpha\}$. We have

$$(2.6.25) \quad \begin{aligned} \mathbb{E}[Y_0(u) \mid \mathcal{F}_s] &= Y_0(s) + \mathbb{E}[Y_0(u) - Y_0(s) \mid \mathcal{F}_s] \\ &= Y_0(s) + \frac{1}{n^\beta} [\mathbb{E}[\#\{j: T_j \in ([sn^\alpha], [un^\alpha]]\} \mid \mathcal{F}_s] - (u-s)n^\alpha] \\ &= Y_0(s) + \frac{1}{n^\beta} \left[\sum_{j \notin \mathcal{V}'_s} (1 - \exp(-d_j(t-s)n^\alpha \ell_n^{-1})) - (u-s)n^\alpha \right] \leq Y_0(s), \end{aligned}$$

where the last step follows because $1 - e^{-x} \leq x$. Therefore, $(Y_0(s))_{s \geq 0}$ is a super-martingale. Also noting that $\mathbb{E}[Y_0(0)] = 0$ and $e^{-x} \leq 1 - x + x^2/2$,

$$\begin{aligned}
|\mathbb{E}[Y_0(u)]| &= -\mathbb{E}[Y_0(u)] = \frac{1}{n^\beta} \left[un^\alpha - \sum_{i \in [n]} (1 - \exp(-un^\alpha d_i \ell_n^{-1})) \right] \\
&= \frac{1}{n^\beta} \left[\sum_{i \in [n]} (un^\alpha d_i \ell_n^{-1} - (1 - \exp(-un^\alpha d_i \ell_n^{-1}))) \right] \leq n^{-\beta} \frac{u^2 n^{2\alpha}}{2} \frac{\sum_{i \in [n]} d_i^2}{\ell_n^2} \\
(2.6.26) \quad &= \frac{u^2}{2} n^{2\alpha-1-\beta} \frac{\frac{1}{n} \sum_{i=1}^n d_i^2}{\left(\frac{1}{n} \sum_{i=1}^n d_i\right)^2},
\end{aligned}$$

and

$$\begin{aligned}
\text{Var}(Y_0(u)) &= n^{-2\beta} \text{var}(N(un^\alpha)) = n^{-2\beta} \sum_{i \in [n]} \mathbb{P}(T_j \leq un^\alpha) (1 - \mathbb{P}(T_j \leq un^\alpha)) \\
(2.6.27) \quad &\leq n^{-2\beta} \sum_{i \in [n]} \frac{d_j un^\alpha}{\ell_n} = un^{\alpha-2\beta}.
\end{aligned}$$

By the super-martingale inequality in Lemma 2.40,

$$(2.6.28) \quad \sup_{u \leq t_n} |n^{-\alpha} N(un^\alpha) - u| = O_{\mathbb{P}}(t_n^{-2} n^\beta).$$

This, in particular, implies that $N(2t_n n^\alpha) \geq t_n n^\alpha$ whp when $t_n = o(n^{\beta/2})$. Therefore,

$$\begin{aligned}
\sup_{u \leq t_n} \left| H_n(u) - \frac{\sigma_f u}{\mu} \right| &\leq \sup_{u \leq 2t_n} \left| n^{-\alpha} \sum_{i=1}^{N(un^\alpha)} f(d_{v_i}) - \frac{\sigma_f}{\mu} n^{-\alpha} N(n^\alpha u) \right| \\
(2.6.29) \quad &\leq \sup_{u \leq 2t_n} \left| n^{-\alpha} \tilde{H}_n(u) - \frac{\sigma_f u}{\mu} \right| + \frac{\sigma_f}{\mu} \sup_{u \leq 2t_n} |n^{-\alpha} N(un^\alpha) - u|.
\end{aligned}$$

Define $Y_1(u) = n^{-\alpha} \sum_{i=1}^{N(un^\alpha)} f(d_{v_i}) - \sigma_f^{(n)} u$, where $\sigma_f^{(n)} = \sum_{i \in [n]} f(d_i) d_i / \ell_n = \sigma_f / \mu + o(1)$ and hence

$$\begin{aligned}
\mathbb{E}[Y_1(t) | \mathcal{F}_s] &= Y_1(s) + \mathbb{E}[Y_1(t) - Y_1(s) | \mathcal{F}_s] \\
(2.6.30) \quad &= Y_1(s) + \frac{1}{n^\alpha} \sum_{j \notin \mathcal{V}'_s} f(d_j) (1 - \exp(-d_j(t-s)n^{2/3} \ell_n^{-1})) - (t-s) \sigma_f^{(n)} \leq Y_1(s).
\end{aligned}$$

Thus, $(Y_1(u))_{u \geq 0}$ is also a super-martingale and by noting that $\mathbb{E}[Y_1(0)] = 0$,

$$\begin{aligned}
|\mathbb{E}[Y_1(t)]| &= -\mathbb{E}[Y_1(t)] = \sigma_f^{(n)} t - n^{-\alpha} \sum_{i \in [n]} f_n(i) (1 - \exp(-tn^\alpha d_i \ell_n^{-1})) \\
&= n^{-\alpha} \sum_{i \in [n]} f_n(i) (\exp(-tn^\alpha d_i \ell_n^{-1}) - 1 + tn^\alpha d_i \ell_n^{-1}) \leq n^{-\alpha} \frac{t^2}{2} n^{2\alpha} \frac{\sum_{i \in [n]} f_n(i) d_i^2}{\ell_n^2} \\
(2.6.31) \quad &= \frac{t^2}{2} n^{\alpha-2/3} \frac{d_{\max}}{n^{1/3}} \frac{\frac{1}{n} \sum_{i \in [n]} f_n(i) d_i}{\left(\frac{\ell_n}{n}\right)^2},
\end{aligned}$$

where the last step follows by the conditions on f_n, d_i and the fact that $\alpha \leq 2/3$. Also,

$$\begin{aligned}
 \text{Var}(Y_1(t)) &= n^{-2\alpha} \sum_{i \in [n]} f^2(d_i) \exp\left(\frac{-tn^\alpha d_i}{\ell_n}\right) \left(1 - \exp\left(\frac{-tn^\alpha d_i}{\ell_n}\right)\right) \\
 &\leq n^{-2\alpha} \max_{i \in [n]} (f(d_i)) tn^\alpha \frac{\sum_{i \in [n]} f_n(i) d_i}{\ell_n} \\
 (2.6.32) \quad &= n^{-\alpha} \max_{i \in [n]} (f(d_i)) \frac{t \sum_{i \in [n]} f(d_i) d_i}{\ell_n}.
 \end{aligned}$$

Recalling the assumptions on f , another application of the super-martingale inequality in Lemma 2.40 yields

$$(2.6.33) \quad \sup_{u \leq t} \left| n^{-\alpha} \tilde{H}_n(u) - \frac{\sigma f u}{\mu} \right| = O_{\mathbb{P}}(a_n \vee b_n),$$

where a_n, b_n are as stated in Proposition 2.42. Thus, (2.6.29) together with (2.6.33) and (2.6.28) completes the proof. \square

2.6.4. Coupling forward degrees in configuration models to i.i.d. degrees.

In this section, we state a coupling result about the sequence of the forward degrees $(\Delta_m^{(n)})_{m=2}^{n^\rho}$ in $\text{CM}_n(\mathbf{d})$ to i.i.d. random variables that will prove quite useful along the way:

PROPOSITION 2.43 ([41, Proposition 2.1]). *Consider $\text{CM}_n(\mathbf{D})$, where the degrees are i.i.d. random variables with finite mean. There exists $0 < \rho < 1$ such that the random vector $(\Delta_m^{(n)})_{m=2}^{n^\rho}$ can be coupled to an independent sequence of random variables $(\Delta)_{m=2}^{n^\rho}$ with probability mass function given in (2.2.16) and $(\Delta_m^{(n)})_{m=2}^{n^\rho} = (\Delta_m)_{m=2}^{n^\rho}$ w.h.p. The same result holds for $\text{CM}_n(\mathbf{d})$ when there exists an $\delta > 0$ such that $d_{\text{TV}}(F_{D_n^* - 1}, F_{D^*}) \leq n^{-\delta}$.*

PROOF. See [41, Proposition 4.5] and the proof of Proposition 4.7 in [41, Appendix A.2]. The extension to $\text{CM}_n(\mathbf{d})$ for which $d_{\text{TV}}(F_{D_n^* - 1}, F_{D^*}) \leq n^{-\delta}$ follows from the proof of [41, Proposition 2.1]. Indeed, there it is shown that $d_{\text{TV}}(F_{D_n^* - 1}, F_{D^*}) \leq n^{-\delta}$ whp for some $\delta > 0$ when the degrees are i.i.d. \square

2.7. Notes and discussion for Chapter 2

Notes on Section 2.1. Branching processes play a crucial role in this text. The unimodular Galton-Watson tree, which we will see occurs naturally as the local limit of the neighborhood of a uniform vertex in random graphs that can be well-approximated by branching process, has been given several other names, such as *delayed* branching process as well as *two-stage* branching process. In this text, we stick to unimodular Galton-Watson tree as it explains the reason why it arises in the context of neighborhoods of uniform vertices in random graphs.

Notes on Section 2.2. We learned of local weak convergence from Justin Salez, who has used it to great effect in some highly non-trivial problems. Examples include the densest subgraph problem [19] (which actually is not quite a continuous functional in the local weak convergence topology), as well as various concave graph parameters in [236]. The latter include the independence number, the maximum cut size, the logarithm of the Tutte polynomial, and the free energy of the anti-ferromagnetic Ising and Potts models.

Notes on Section 2.3. Bollobás, Janson and Riordan [56] extend Theorem 2.20 to rather general inhomogeneous random graphs.

Notes on Section 2.4. The connectivity threshold for inhomogeneous random graphs has attracted some attention as well. In general, there the average degree needs to be of order $\log n$ for connectivity to occur. See the work by Erdős and Rényi in [123], or [160, Section 5.3] and the references therein, for results for Erdős-Rényi random graph and Devroye and Fraiman [102] for results on general inhomogeneous random graphs.

In work with Janson and Luczak [166], Theorem 2.21 was extended to the *barely supercritical* regime, where the giant component contains a vanishing proportion of the vertices, but is still close to deterministic as well as unique. See Chapter 4, where we study the critical behavior of percolation on the configuration model, for more details.

Notes on Section 2.5. There is also some work on distances of $\text{CM}_n(\mathbf{d})$ when the degrees are i.i.d. with *infinite expectation*, i.e., when $\tau \in (1, 2)$. See the work with van den Esker, Hooghiemstra and Znamenski [127]. Distances for $\text{CM}_n(\mathbf{d})$ for which $\tau = 3$ (and with some carefully chosen slowly-varying functions) are studied by Dereich, Mönch and Mörters in [97]. This proves related results to those for $\text{PA}_n^{(m,\delta)}$ with $\delta = 0$ in Theorem 2.36.

Notes on Section 2.6. Proposition 2.42 is taken from work on the critical behavior of the configuration model with Dhara, van Leeuwen and Sen [103].

First passage percolation on random graphs

Abstract

In this chapter, we investigate first-passage percolation on random graphs, which models the effect of independent and identically distributed edge weights on the metric structure of the graph. For this, we investigate the asymptotic properties of smallest-weight paths between vertices in complete and random graphs, both in terms of their weight-structure as well as in terms of the number of edges along optimal paths. We start by motivating the problem.

3.1. First passage percolation on general graphs and motivation

In first-passage percolation, we study *random geometries*, where the edges in a graph have *edge weights* associated with them. These edge weights change the geometry of graph distances, and here we will be interested in properties of geodesics or smallest-weight paths between vertices. First-passage percolation has many interpretations. For example, when we interpret the weight of an edge as the *time* it takes a disease to be spread between the vertices on either end of the end, first-passage percolation gives rise to a caricature model of an epidemic. When we think of two competing species trying to explore new territory, first-passage percolation can be used to describe which species wins the most of the available territory. When we think of the edge weight as corresponding to the *cost* of using that edge, first-passage percolation describes the structure of cheapest routes of connections.

Let us start by describing the basic model. Let G be a connected graph on n vertices. Assign independent and identically distributed (i.i.d.) random edge *weights* or *lengths* to the edges of the graph. These random edge-weights generate geodesics on the graph. Think of the graph as a disordered random system carrying flow between pairs of vertices in the graph via shortest paths between them. Choose two vertices in the graph uniformly at random amongst the n vertices. We will call these two vertices “typical” vertices. Two functionals of interest are the minimal weight W_n of a path between the two vertices and the number of edges H_n on the minimal path, often referred to as the *hopcount*. We assume that the common distribution of the edge weights is continuous, so that the optimal paths are a.s. unique and one can talk about objects such as the number of edges in *the* optimal path. This model has been studied intensively, largely in the context of the integer lattice $[-N, N]^d$ (see e.g. Kesten [195], Hammersley [152], Howard [175], and Smythe and Wierman [241], or the recent overview on 50 years of first-passage percolation by Auffinger, Damron and Hanson [24]). For the power of this model to analyze more complicated interacting particle systems, see Liggett [206] and Durrett [115]

and the references therein. Key results for first-passage percolation on \mathbb{Z}^d include the *shape theorem*, showing that the set of vertices $B(t)$ that are reachable within time t satisfies that $B(t)/t \xrightarrow{a.s.} \mathcal{B}$ for some (convex) set \mathcal{B} . The shape of this set \mathcal{B} is generally unknown, and it is believed that it is not always *strictly* convex. Also, many recent results exist on critical exponents for first-passage percolation on \mathbb{Z}^d . See in particular [24] for more details.

In this chapter, we focus on first-passage percolation on *finite graphs*. Below, we will specialize even further to complete and random graphs. Of course, in such settings, shape theorems do not make much sense, and instead we focus on the geometry of geodesics or smallest-weight paths. Let us describe smallest-weight routing on graphs formally now.

Formal definition of the problem. Let $G = (V(G), E(G))$ be a graph with vertex set $V(G)$ and edge set $E(G)$. The graphs that we work on are finite, so that we may write $V(G) = [n]$. Let $(Y_e)_{e \in E(G)}$ denote the edge-weights on our graph G , where we assume that $Y_e > 0$ a.s. Then, for $i, j \in [n]$, we let $\mathcal{C}_n(i, j)$ denote the minimal weight over all paths from i to j , i.e.,

$$(3.1.1) \quad \mathcal{C}_n(i, j) = \min_{\pi: i \rightarrow j} \sum_{e \in \pi} Y_e.$$

Denote the optimal path realizing the minimum in (3.1.1) by $\pi_n(i, j)$. When we assume that $(Y_e)_{e \in E(G)}$ are i.i.d. random variables with a continuous distribution, which we assume henceforth, then $\pi_n(i, j)$ is unique. We define the hopcount between i and j to be

$$(3.1.2) \quad H_n(i, j) = |\pi_n(i, j)|,$$

where we recall that $|\pi|$ denotes the number of edges in π . The weight distribution and hopcount of a graph are given by

$$(3.1.3) \quad \mathcal{C}_n = \mathcal{C}_n(U_1, U_2), \quad H_n = H_n(U_1, U_2),$$

where U_1, U_2 are two vertices chosen uniformly at random, independently of each other. When the graph is disconnected, we will often condition on $H_n < \infty$. Note that the *typical distance* $\text{dist}_{\text{CM}_n(d)}(U_1, U_2)$ as introduced in Section 2.5 is the special example of (3.1.3) where $Y_e \equiv 1$ for every $e \in E(G)$. Further objects of interest are the *flooding time* of vertex $i \in [n]$ given by $\max_{j \in [n]} \mathcal{C}_n(i, j)$ and *weight diameter* $\max_{i, j \in [n]} \mathcal{C}_n(i, j)$.

When we think about $Y_{(i,j)}$ as the amount of time it takes for a rumor to be transferred from person i to person j , then $\mathcal{C}_n(i, j)$ is the time for the rumor started by person i to reach person j and $\max_{j \in [n]} \mathcal{C}_n(i, j)$ is the time for the whole population to hear the rumor. This is why this is sometimes also referred to as the *flooding time*. The weight-diameter $\max_{i, j \in [n]} \mathcal{C}_n(i, j)$ has the interpretation as the time it takes all rumors to reach the entire population when each person starts with its own unique rumor, which may sound less realistic. When thinking about transportation costs, however, the diameter is the cost of traveling between the pair of vertices in the network with the highest travel cost. We will mainly focus on the limiting properties of the random variables \mathcal{C}_n and H_n , such as their leading order asymptotics as well as their fluctuations.

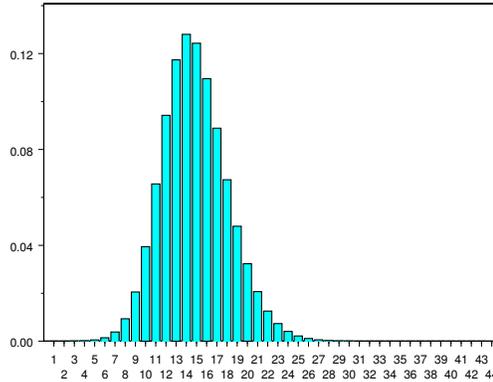


FIGURE 6. Internet hopcount data. Data courtesy of H. Tangmunarunkit.

Applications for weighted real-world networks. In the modern context, first-passage percolation takes on an added significance. Many real-world networks (such as the Internet at the router level or various road and rail networks) are entrusted with carrying flow between various parts of the network. These networks have both a graph theoretic structure as well as weights on edges that could represent congestion weights. For example, in Internet, routing is (at least partially) performed by smallest-weight routing. While we do not know which weights are used, in 2000, CISCO advised its customers to use the inverse of the bandwidth as edge weights. In turn, the empirical properties of bandwidths of cables in Internet are unknown, so that we resort to assuming that edge weights are independent and identically distributed random variables. This leads to first-passage percolation on networks. When we model real-world networks with random graphs, we are lead to studying first-passage percolation on them.

In applied settings, understanding properties of both the number of edges in, and weight of, and the optimal path are crucial. Indeed, while routing is done via smallest-weight paths, the actual time delay experienced by users scales like the hopcount (the number of “hops” a message has to perform in getting from its source to its destination). In this chapter, we investigate properties of smallest-weight paths, in particular the properties of minimal weight paths between vertices, focussing on their weight and the number of edges in them.

An instance of such distances are the *hopcount in Internet*, see Figure 6. In Internet, the hopcount is the number of routers traversed by an e-mail message between two uniformly chosen routers. Measurements of the hopcount make use of `traceroute` which allows one to record the number of routers traversed by e-mails. Thus, with a sufficiently large set of routers, we can measure the hopcount between any pair of them, thus yielding an estimate of the hopcount in Internet. See Shavit and Shir [239] for more details on Internet data. We emphasize that obtaining such measurements is a highly non-trivial task, as it is unclear how to obtain a representative set of routers. Further, uncleaned `traceroute`-data contains many errors, since routers may appear with distinct names, etc. In fact, such errors lie at the basis of the controversy about the Internet topology, for which we refer to Willinger et al. in [253], or to [160, Chapter 1], for more details.

An alternative interpretation of first-passage percolation is that of a model for rumor spread or an epidemic. The edge weights then correspond to the random times needed to spread the rumor across an edge. Epidemics on random graphs have attracted tremendous attention in the literature. See the recent overview by Pastor-Satorras, Castellano, Van Mieghem and Vespignani [227]. First-passage percolation corresponds to the Susceptible-Infected or SI model. We will study the related problem of Susceptible-Infected-Recovered or SIR model in Chapter 4, and the more-involved setting of Susceptible-Infected-Susceptible or SIS model in Section 6.2. The SIS model also goes under the name of *contact process*. See also the book by Draief and Massoulié [114] for a description of many of the rigorous results.

Organization of Chapter 3. This chapter is organized as follows. In Section 3.2, we start with the simplest possible setting, the complete graph with exponential edge weights. Just as the Erdős-Rényi random graph is a natural starting point for studying random graphs, and thus serves as a benchmark model, the complete graph with exponential edge weights serves as a benchmark model for weighted random graphs. Of course, the complete graph is not a realistic model for any real-world network, and in Section 3.3, we investigate the configuration model with exponential edge weights on random graphs instead. In Section 3.4, we introduce continuous-time branching processes, which serve as a main tool in the remainder of this chapter. We continue in Section 3.5 by showing that first-passage percolation on the configuration model with finite-variance degrees shows a remarkable degree of universality, in that the limiting behavior of both the weight and the number of edges in the minimal-weight path between two uniform connected vertices hardly depend on the specific edge-weight and degree distribution. In Section 3.6, we discuss scale-free settings in which the asymptotic degree distribution has infinite variance. There, it is unclear what the universality classes are precisely. We discuss that there are settings where the weight of the smallest-weight path \mathcal{C}_n between two uniform vertices remains uniformly bounded, and converges in distribution. We also discuss a case where \mathcal{C}_n grows with n . We continue in Section 3.7 by describing related results on routing on complete and random graphs. We close in Section 3.8 with notes and discussion.

3.2. Fixing ideas: first-passage percolation on the complete graph

Assign to every edge ij of the complete graph a random weight Y_{ij} . By convention, we let $Y_{ji} = Y_{ij}$. We assume that the $\binom{n}{2}$ weights $(Y_{ij})_{1 \leq i < j \leq n}$, are i.i.d. exponential random variables with parameter 1, so that $\mathbb{P}(Y_{ij} > x) = e^{-x}$ for $x \geq 0$. This setting was first studied by Janson [179]. The first main theorem of this section is a set of three different asymptotic results for $\mathcal{C}_n(i, j)$:

THEOREM 3.1 (One, two and three times $\log n/n$). *For the complete graphs with exponential edge weights, as $n \rightarrow \infty$,*

(i) *for any fixed i and j ,*

$$(3.2.1) \quad \frac{\mathcal{C}_n(i, j)}{\log n/n} \xrightarrow{\mathbb{P}} 1;$$

(ii) *for any fixed i ,*

$$(3.2.2) \quad \frac{\max_{j \leq n} \mathcal{C}_n(i, j)}{\log n/n} \xrightarrow{\mathbb{P}} 2;$$

(iii)

$$(3.2.3) \quad \frac{\max_{i,j \leq n} \mathcal{C}_n(i,j)}{\log n/n} \xrightarrow{\mathbb{P}} 3.$$

Hence, whp, $\mathcal{C}_n(i,j)$ is about $\log n/n$ for any fixed (or random) pair of vertices, but there are pairs of vertices for which it is larger: up to $2 \log n/n$ if i is fixed and j varies, and up to $3 \log n/n$ globally. Theorem 3.1(i),(ii) may alternatively be stated in terms of first-passage percolation on the complete graph (the time to reach a given vertex is about $\log n/n$ and the time to reach all is $2 \log n/n$). In the following exercise, we investigate properties of the minimal edge weight from a given vertex, as these will be useful to interpret the results in Theorem 3.1:

EXERCISE 3.1 (Minimal weights from a given vertex). *Let $X_i = \min_{j \neq i} Y_{ij}$. Show that, as $n \rightarrow \infty$,*

$$(3.2.4) \quad nX_i \xrightarrow{d} \text{Exp}(1), \quad n^2 \min_{i \in [n]} X_i = n^2 \min_{i,j \in [n]: i \neq j} Y_{ij} \xrightarrow{d} \text{Exp}(2),$$

while

$$(3.2.5) \quad \frac{\max_{i \in [n]} X_i}{\log n/n} \xrightarrow{\mathbb{P}} 1.$$

Using Exercise 3.1, we can give a simple informal interpretation of the three parts of Theorem 3.1 as follows, interpreting the weights as travel times. Most vertices are connected by efficient highways, which take you to almost any other vertex within about $\log n/n$ (but rarely much quicker). Some vertices, however, are remote villages, from which it takes up to $\log n/n$ to get to any other vertex at all. Hence, starting at a typical vertex, most travel times are about $\log n/n$, but it takes an extra $\log n/n$ (just for the final step in the path) to reach a few remote vertices. Similarly, if we start at one of the very remote vertices, then it takes about $\log n/n$ to get to any other vertex at all, $2 \log n/n$ to get to most other vertices and $3 \log n/n$ to get to the other very remote vertices.

Proof of Theorem 3.1. For parts (i) and (ii), we may assume that $i = 1$. We adopt the first-passage percolation viewpoint, so we regard vertex 1 as initially infected, and assume that the infection spreads along each edge with an $\text{Exp}(1)$ -distributed waiting time. We first study when the other vertices get infected, considering them in order of infection and ignoring their labels.

Since there are $n - 1$ neighbors of the initially infected vertex, the time T_1 until the second vertex is infected is exponentially distributed with expectation $1/(n-1)$. Recall the *memoryless property* of the exponential distribution, which states that, conditionally on an exponential random variable E satisfying $E > x$, the excess $E - x$ is again exponentially distributed with the same parameter. This can be extended in a convenient way as follows. Let $(E_i)_{i=1}^k$ be i.i.d. exponential random variables with parameter 1. Then, $\min_{i \in [k]} E_i$ has an exponential distribution with parameter k (and thus mean $1/k$). Further, let i_k be the index i for which $E_{i_k} = \min_{i \in [k]} E_i$, then the random variables $(E_j - E_{i_k})_{j \in [k] \setminus \{i_k\}}$ is a collection of $k - 1$ i.i.d. exponential random variables with parameter 1.

Let us use the above realization in the first-passage percolation setting on the complete graph. We conclude that when $k < n$ vertices have been infected, there are $k(n - k)$ edges connecting the infected and non-infected vertices, and the excess edge-weights along these

edges are i.i.d. exponential random variables with mean 1. We conclude that the time T_k until the next vertex is infected is $\text{Exp}(k(n-k))$. Moreover, this time is independent of T_1, \dots, T_{k-1} . In other words, the time S_m until m with $m \geq 1$ vertices have become infected can be written as

$$(3.2.6) \quad S_m = \sum_{k=1}^{m-1} T_k,$$

where T_1, \dots, T_{m-1} are independent with T_k having an $\text{Exp}(k(n-k))$ distribution.

The times $(S_m)_{m=2}^n$ are just the minimal path weights $(\mathcal{C}_n(1, j))_{j=2}^n$, arranged in increasing order. In particular,

$$(3.2.7) \quad \max_{j \in [n]} \mathcal{C}_n(1, j) = S_n = \sum_{k=1}^{n-1} T_k.$$

Hence, using that

$$(3.2.8) \quad \frac{1}{k(n-k)} = \frac{1}{n} \left(\frac{1}{k} + \frac{1}{n-k} \right),$$

we arrive at

$$(3.2.9) \quad \begin{aligned} \mathbb{E}[\max_{j \in [n]} \mathcal{C}_n(1, j)] &= \sum_{k=1}^{n-1} \mathbb{E}[T_k] = \sum_{k=1}^{n-1} \frac{1}{k(n-k)} = \frac{1}{n} \sum_{k=1}^{n-1} \left(\frac{1}{k} + \frac{1}{n-k} \right) \\ &= \frac{2}{n} \sum_{k=1}^{n-1} \frac{1}{k} = \frac{2 \log n}{n} + O(1/n), \end{aligned}$$

and, similarly,

$$(3.2.10) \quad \begin{aligned} \text{Var}(\max_{j \in [n]} \mathcal{C}_n(1, j)) &= \sum_{k=1}^{n-1} \text{Var}(T_k) = \sum_{k=1}^{n-1} \left(\frac{1}{k(n-k)} \right)^2 \\ &\leq 2 \sum_{k=1}^{n/2} \frac{1}{k^2(n-k)^2} \leq \frac{8}{n^2} \sum_{k=1}^{n/2} \frac{1}{k^2} = O(1/n^2). \end{aligned}$$

Part (ii) now follows by Chebyshev's inequality.

For part (i), fix $j = 2$. Observe that, if N is the number of vertices found by the flow before vertex 2 is found, then

$$(3.2.11) \quad \mathcal{C}_n(1, 2) = S_{N+1} = \sum_{k=1}^N T_k,$$

where, by exchangeability, N is uniformly distributed over $1, \dots, n-1$ and independent of T_1, \dots, T_{n-1} . We rewrite this equation as

$$(3.2.12) \quad \mathcal{C}_n(1, 2) = \sum_{k=1}^{n-1} \mathbb{1}_{\{N \geq k\}} T_k,$$

using indicator functions to eliminate the random summation limit. This rewrite, together with the independence between N and $(T_k)_{k=1}^{n-1}$, allows us to compute the first moment of

$\mathcal{C}_n(1, 2)$ as

$$\begin{aligned}
\mathbb{E}[\mathcal{C}_n(1, 2)] &= \sum_{k=1}^{n-1} \mathbb{E}[\mathbb{1}_{\{N \geq k\}} T_k] = \sum_{k=1}^{n-1} \mathbb{P}(N \geq k) \mathbb{E}[T_k] \\
&= \sum_{k=1}^{n-1} \frac{n-k}{n-1} \frac{1}{k(n-k)} = \sum_{k=1}^{n-1} \frac{1}{k(n-1)} \\
(3.2.13) \quad &= \frac{\log n}{n} + O(1/n).
\end{aligned}$$

In order to estimate the variance, we further rewrite the sum as

$$\begin{aligned}
\mathcal{C}_n(1, 2) &= \sum_{k=1}^N (T_k - \mathbb{E}[T_k]) + \sum_{k=1}^N \frac{1}{n} \left(\frac{1}{k} + \frac{1}{n-k} \right) \\
(3.2.14) \quad &= \sum_{k=1}^N (T_k - \mathbb{E}[T_k]) + \frac{1}{n} (\log N + \log n - \log(n-N)) + O(1/n).
\end{aligned}$$

We consider the three terms on the right-hand side separately. Since N, T_1, \dots, T_{n-1} are independent,

$$\begin{aligned}
\text{Var} \left(\sum_{k=1}^N (T_k - \mathbb{E}[T_k]) \right) &= \mathbb{E} \left[\left(\sum_{k=1}^N (T_k - \mathbb{E}[T_k]) \right)^2 \right] = \mathbb{E} \left[\sum_{k=1}^N \text{Var}(T_k) \right] \\
&\leq \sum_{k=1}^{n-1} \text{Var}(T_k) = \sum_{k=1}^{n-1} \frac{1}{k^2(n-k)^2} \\
(3.2.15) \quad &\leq \sum_{k=1}^{n/2} \frac{4}{k^2 n^2} + \sum_{k=n/2}^{n-1} \frac{4}{n^2(n-k)^2} = O(1/n^2).
\end{aligned}$$

For the second term, we observe that, as $n \rightarrow \infty$,

$$(3.2.16) \quad \mathbb{E}[(\log N - \log(n-1))^2] = \mathbb{E} \left[\left(\log \frac{N}{n-1} \right)^2 \right] \rightarrow \int_0^1 (\log x)^2 dx < \infty.$$

Hence $\text{Var}(\log N) = \text{Var}(\log(n-N)) = O(1)$, and it follows that the variance of the second term in (3.2.14) is also $O(1/n^2)$. The same is trivially true for the third term. Consequently, $\text{Var}(\mathcal{C}_n(1, 2)) = O(1/n^2)$, which together with (3.2.13) yields part (i).

The proof of (iii) is divided into two parts, considering upper and lower bounds separately. We start with the upper bound, and rely on an exponential Chebychev inequality. First, by (3.2.7), for $-\infty \leq t < 1 - 1/n$.

$$(3.2.17) \quad \mathbb{E}[e^{nt \max_{j \in [n]} \mathcal{C}_n(1, j)}] = \prod_{k=1}^{n-1} \mathbb{E}[e^{nt T_k}] = \prod_{k=1}^{n-1} \left(1 - \frac{nt}{k(n-k)} \right)^{-1}.$$

Hence, for every $a > 0$, where we assume without loss of generality that $n \geq 3$,

$$\begin{aligned}
\mathbb{P}(n \max_{j \in [n]} \mathcal{C}_n(1, j) > a \log n) &\leq \mathbb{E}[e^{nt \max_{j \in [n]} \mathcal{C}_n(1, j) - ta \log n}] \\
&= e^{-ta \log n} \prod_{k=1}^{n-1} \left(1 - \frac{nt}{k(n-k)}\right)^{-1} \\
(3.2.18) \quad &= \left(1 - \frac{nt}{n-1}\right)^{-2} \exp\left(-ta \log n - \sum_{k=2}^{n-2} \log\left(1 - \frac{nt}{k(n-k)}\right)\right).
\end{aligned}$$

Using the Taylor expansion $-\log(1-x) = x + O(x^2)$, and choosing $t = 1 - 1/\log n$, this leads us to

$$\begin{aligned}
\mathbb{P}(n \max_{j \in [n]} \mathcal{C}_n(1, j) > a \log n) \\
&\leq \left(1 - \frac{nt}{n-1}\right)^{-2} \exp\left(-ta \log n + \sum_{k=2}^{n-2} \left(\frac{nt}{k(n-k)} + \left(\frac{nt}{k(n-k)}\right)^2\right)\right) \\
(3.2.19) \quad &= (1 - t + O(1/n))^{-2} \exp(-ta \log n + 2t \log n + O(1)) = O(n^{2-a} \log^2 n).
\end{aligned}$$

This evidently implies that

$$\begin{aligned}
(3.2.20) \quad \mathbb{P}(\max_{i, j \in [n]} \mathcal{C}_n(i, j) > a \log n/n) &= \mathbb{P}(\max_i \max_{j \in [n]} \mathcal{C}_n(i, j) > a \log n/n) \\
&\leq n \mathbb{P}(\max_{j \in [n]} \mathcal{C}_n(1, j) > a \log n/n) = O(n^{3-a} \log^2 n),
\end{aligned}$$

which tends to 0 as $n \rightarrow \infty$, for every fixed $a > 3$. This establishes the required upper bound.

The lower bound on $\max_{i, j \in [n]} \mathcal{C}_n(i, j)$ makes use of two steps. We first show that whp there are vertices $i \in [n]$ whose minimal edge weight $X_i = \min_{j \neq i} Y_{ij}$ is at least $(1 - \varepsilon) \log n/n$. In the second step, we show that most pairs of such vertices have a smallest weight at least $3(1 - \varepsilon) \log n/n$. We start by investigating the number of vertices with minimal edge weight $X_i \geq (1 - \varepsilon) \log n/n$.

We let $\varepsilon > 0$ to be determined later on, and define the vertices with minimal edge weight at least $(1 - \varepsilon) \log n/n$ to be

$$(3.2.21) \quad \mathcal{N} = \{i: X_i \geq (1 - \varepsilon) \log n/n\},$$

where $X_i = \min_{j \neq i} Y_{ij}$ is the minimal edge-weight from vertex i as studied in Exercise 3.1.

We apply a second moment method on $|\mathcal{N}|$. Since $X_i \stackrel{d}{=} \text{Exp}(n-1)$,

$$(3.2.22) \quad \mathbb{E}[|\mathcal{N}|] = n \mathbb{P}(nX_1 \geq (1 - \varepsilon) \log n) = ne^{-(1-\varepsilon)\frac{n-1}{n} \log n} = n^\varepsilon(1 + o(1)).$$

Further,

$$(3.2.23) \quad \mathbb{E}[|\mathcal{N}|(|\mathcal{N}| - 1)] = n(n-1) \mathbb{P}(nX_1 \geq (1 - \varepsilon) \log n, nX_2 \geq (1 - \varepsilon) \log n),$$

where

$$(3.2.24) \quad \mathbb{P}(nX_1 \geq (1 - \varepsilon) \log n, nX_2 \geq (1 - \varepsilon) \log n) = e^{-(1-\varepsilon)\frac{(2n-3)}{n} \log n}.$$

Therefore,

$$(3.2.25) \quad \mathbb{E}[|\mathcal{N}|(|\mathcal{N}| - 1)] \leq \mathbb{E}[|\mathcal{N}|]^2 e^{(1-\varepsilon) \log n/n}.$$

We conclude that $\text{Var}(|\mathcal{N}|) = o(\mathbb{E}[|\mathcal{N}|]^2)$, so that by the Chebychev inequality

$$(3.2.26) \quad n^{-\varepsilon}|\mathcal{N}| \xrightarrow{\mathbb{P}} 1.$$

This completes the first step in the proof.

We next show that $n\mathcal{C}_n(i, j) \geq 3(1 - \varepsilon) \log n$ for most pairs of vertices $i, j \in \mathcal{N}$. For this, we let

$$(3.2.27) \quad \mathcal{P} = \{(i, j) \in \mathcal{N}^2: nY_{ij} \geq 2(1 - \varepsilon) \log n, n\mathcal{C}_n(i, j) \geq 3(1 - \varepsilon) \log n\}$$

denote the set of pairs (i, j) in \mathcal{N} for which the weight of the edge between i and j is at least $2(1 - \varepsilon) \log n/n$, and the minimal weight is at least $3(1 - \varepsilon) \log n/n$.

Our aim is to show that $n^{-2\varepsilon}|\mathcal{P}| \xrightarrow{\mathbb{P}} 1$, which proves the lower bound. For this, we start by noticing that

$$(3.2.28) \quad 0 \leq |\mathcal{N}|(|\mathcal{N}| - 1) - |\mathcal{P}| \leq N_1 + N_2,$$

where

$$(3.2.29) \quad N_1 = |\{(i, j) \in \mathcal{N}^2: nY_{ij} < 2(1 - \varepsilon) \log n\}|,$$

$$(3.2.30) \quad N_2 = |\{(i, j) \in \mathcal{N}^2: nY_{ij} \geq 2(1 - \varepsilon) \log n, n\mathcal{C}_n(i, j) < 3(1 - \varepsilon) \log n\}|.$$

We know that $n^{-2\varepsilon}|\mathcal{N}|(|\mathcal{N}| - 1) \xrightarrow{\mathbb{P}} 1$. We perform a first moment method on N_1 and N_2 to show that $n^{-2\varepsilon}N_i \xrightarrow{\mathbb{P}} 0$ for $i = 1, 2$, which then shows that indeed $n^{-2\varepsilon}|\mathcal{P}| \xrightarrow{\mathbb{P}} 1$. We compute

$$(3.2.31) \quad \begin{aligned} \mathbb{E}[N_1] &= \mathbb{E}|\{(i, j) \in \mathcal{N}^2: nY_{ij} < 2(1 - \varepsilon) \log n\}| \\ &\leq n(n - 1)[1 - e^{-2(1 - \varepsilon) \log n/n}]e^{-2(1 - \varepsilon) \frac{n-1}{n} \log n} = o(n^{2\varepsilon}), \end{aligned}$$

so that by the Markov inequality indeed $n^{-2\varepsilon}N_1 \xrightarrow{\mathbb{P}} 0$. For N_2 , we compute

$$(3.2.32) \quad \begin{aligned} \mathbb{E}[N_2] &= n(n - 1)\mathbb{P}(1, 2 \in \mathcal{N}, nY_{12} \geq 2(1 - \varepsilon) \log n, n\mathcal{C}_n(1, 2) < 3(1 - \varepsilon) \log n) \\ &= n(n - 1)\mathbb{P}(1, 2 \in \mathcal{N}, nY_{12} \geq 2(1 - \varepsilon) \log n) \\ &\quad \times \mathbb{P}(n\mathcal{C}_n(1, 2) < 3(1 - \varepsilon) \log n \mid 1, 2 \in \mathcal{N}, nY_{12} \geq 2(1 - \varepsilon) \log n). \end{aligned}$$

A minor adaptation of (3.2.24) yields that

$$(3.2.33) \quad n^{2(1 - \varepsilon)}\mathbb{P}(1, 2 \in \mathcal{N}, nY_{12} \geq 2(1 - \varepsilon) \log n) \rightarrow 1,$$

so that

$$(3.2.34) \quad \mathbb{E}[N_2] = n^{2\varepsilon}\mathbb{P}(n\mathcal{C}_n(1, 2) < 3(1 - \varepsilon) \log n \mid 1, 2 \in \mathcal{N}, nY_{12} \geq 2(1 - \varepsilon) \log n)(1 + o(1)).$$

We are left to investigate the conditional probability. For this, we note that conditionally on $Y_{ij} > x$, the distribution of $Y_{ij} - x$ is again $\text{Exp}(1)$. Therefore, the information that $1, 2 \in \mathcal{N}, nY_{12} \geq 2(1 - \varepsilon) \log n$ implies that $Y_{12} - 2(1 - \varepsilon) \log n$, $(Y_{1j} - (1 - \varepsilon) \log n/n)_{j \geq 3}$ and $(Y_{2j} - (1 - \varepsilon) \log n/n)_{j \geq 3}$ are a collection of $2n - 3$ independent $\text{Exp}(1)$ random variables. The law of all other weights, i.e., Y_{ij} with both $i \neq 1$ and $j \neq 2$ are unaffected by the conditioning on $1, 2 \in \mathcal{N}, nY_{12} \geq 2(1 - \varepsilon) \log n$.

Let π be a path from 1 to 2. When π contains one edge, its conditional weight given $1, 2 \in \mathcal{N}, nY_{12} \geq 2(1 - \varepsilon) \log n$ has distribution $2(1 - \varepsilon) \log n/n + E_{12}$ where E_{12} has an $\text{Exp}(1)$ distribution. On the other hand, when π contains at least two edges, then both the edge incident to 1 and that incident to 2 have edge have weight $E_{1\pi(1)} + (1 - \varepsilon) \log n/n$ and

$E_{\pi(|\pi|-1)\pi(|\pi|)} + (1 - \varepsilon) \log n/n$ respectively. As a result, conditionally on $1, 2 \in \mathcal{N}, nY_{12} \geq 2(1 - \varepsilon) \log n$, the weight of a path π from 1 to 2 has distribution

$$(3.2.35) \quad \sum_{e \in \pi} E_e + 2(1 - \varepsilon) \log n/n,$$

for all paths π from 1 to 2. This distribution is equal to $2(1 - \varepsilon) \log n/n$ plus the *unconditional* distribution of the smallest-edge weight between vertices 1 and 2, so that

$$(3.2.36) \quad \begin{aligned} \mathbb{P}(nC_n(1, 2) < 3(1 - \varepsilon) \log n \mid 1, 2 \in \mathcal{N}, nY_{12} \geq 2(1 - \varepsilon) \log n) \\ = \mathbb{P}(nC_n(1, 2) < (1 - \varepsilon) \log n) = o(1), \end{aligned}$$

by part (i). This completes the lower bound in part (iii). \square

EXERCISE 3.2 (Alternative proof of lower bound $\max_{j \in [n]} \mathcal{C}_n(1, j)$). *Adapt the proof of the lower bound on $\max_{i, j \in [n]} \mathcal{C}_n(i, j)$ to give an alternative proof of the fact that $n \max_{j \in [n]} \mathcal{C}_n(1, j) \geq 2(1 - \varepsilon) \log n$ whp.*

Limit distributions of path weights. We next study the *fluctuations* of the flooding $\max_{j \in [n]} \mathcal{C}_n(i, j)$ and $\mathcal{C}_n(i, j)$. The stochastic description in (3.2.6) and (3.2.10) also yields their asymptotic distributions:

THEOREM 3.2 (Asymptotic distributions of minimal weights). *As $n \rightarrow \infty$, for every $i \neq j$ with $i, j \in [n]$ fixed,*

$$(3.2.37) \quad n\mathcal{C}_n(i, j) - \log n \xrightarrow{d} \Lambda_1 + \Lambda_2 - \Lambda_{12},$$

and, for every $i \in [n]$ fixed,

$$(3.2.38) \quad n \max_{j \in [n]} \mathcal{C}_n(i, j) - 2 \log n \xrightarrow{d} \Lambda_1 + \Lambda_2,$$

where $\Lambda_1, \Lambda_2, \Lambda_{12}$ are independent Gumbel random variables, i.e., $\mathbb{P}(\Lambda_i \leq x) = e^{-e^{-x}}$ for all $x \in \mathbb{R}$.

PROOF. We write $A_n \approx B_n$ to mean that $\mathbb{E}[(A_n - B_n)^2] = o(1)$ as $n \rightarrow \infty$. Equation (3.2.11) implies that with $(E_i)_{i \geq 1}$ denoting independent $\text{Exp}(1)$ random variables and N a discrete uniform random variable in the set $[n - 1]$ independent of $(E_i)_{i \geq 1}$,

$$(3.2.39) \quad \begin{aligned} n\mathcal{C}_n(i, j) &\stackrel{d}{=} \sum_{k=1}^N \frac{n}{k(n-k)} E_k = \sum_{k=1}^N \frac{n}{k(n-k)} (E_k - 1) + \sum_{k=1}^N \left(\frac{1}{k} + \frac{1}{n-k} \right) \\ &\approx \sum_{k=1}^N (E_k - 1)/k + \log N + \gamma + \log n - \log(n - N) \\ &\approx \sum_{k=1}^{\infty} (E_k - 1)/k + \log \left(\frac{N/n}{1 - N/n} \right) + \log n + \gamma, \end{aligned}$$

where γ is the Euler-Mascheroni constant.

Similarly, with E'_k denoting i.i.d. exponentials with mean 1 independent of $(E_k)_{k \geq 1}$, (3.2.7) implies that

$$\begin{aligned}
n\text{Flood}_n(i) &\stackrel{d}{=} \sum_{k=1}^{n-1} \frac{n}{k(n-k)} E_k = \sum_{k=1}^{n-1} \frac{n}{k(n-k)} (E_k - 1) + 2 \sum_{k=1}^{n-1} \frac{1}{k} \\
&\approx \sum_{k=1}^{\lfloor n/2 \rfloor} (E_k - 1)/k + \sum_{k=\lfloor n/2 \rfloor + 1}^{n-1} \frac{1}{n-k} (E_k - 1) + 2 \log n + 2\gamma \\
(3.2.40) \quad &\stackrel{d}{=} \sum_{k=1}^{\lfloor n/2 \rfloor} (E_k - 1)/k + \sum_{k=1}^{\lfloor n/2 \rfloor - 1} (E'_k - 1)/k + 2 \log n + 2\gamma.
\end{aligned}$$

Since $N/n \xrightarrow{d} U$, where U has a uniform distribution on $[0, 1]$, the random variable $L = \log(U/(1-U))$ has the logistic distribution, i.e., for every $x \in \mathbb{R}$,

$$(3.2.41) \quad \mathbb{P}(L \leq x) = \mathbb{P}(\log(U/(1-U)) \leq x) = \mathbb{P}(U \leq e^x/(1+e^x)) = e^x/(1+e^x).$$

Therefore,

$$(3.2.42) \quad n\mathcal{C}_n(i, j) - \log n \xrightarrow{d} \sum_{k=1}^{\infty} (E_k - 1)/k + \gamma + L,$$

and

$$(3.2.43) \quad n \max_{j \in [n]} \mathcal{C}_n(i, j) - 2 \log n \xrightarrow{d} \sum_{k=1}^{\infty} (E_k - 1)/k + \sum_{k=1}^{\infty} (E'_k - 1)/k + 2\gamma.$$

We are left to show that

$$(3.2.44) \quad \Lambda = \sum_{k=1}^{\infty} (E_k - 1)/k + \gamma$$

has a Gumbel distribution, and that L has the same distribution as $\Lambda_1 - \Lambda_2$.

We see that

$$(3.2.45) \quad \Lambda = \sum_{k=1}^{\infty} \frac{1}{k} (E_k - 1) + \gamma = \lim_{n \rightarrow \infty} \sum_{k=1}^n E_k/k - \log n.$$

Further, by the memoryless property of the exponential distribution, $\sum_{k=1}^n E_k/k$ has the same distribution as $\max_{i \in [n]} E_i$:

EXERCISE 3.3 (Order statistics of i.i.d. exponentials). *Let $(E_i)_{i \in [n]}$ be a sequence of i.i.d. exponential distributions with parameter 1. Let $(E_{(i:n)})_{i \in [n]}$ be the order statistics of $(E_i)_{i \in [n]}$, i.e., the reordering of $(E_i)_{i \in [n]}$ in increasing order. Show that $(E_{(i:n)})_{i \in [n]}$ has the same distribution as $(\sum_{k=n}^{n-i+1} E_k/k)_{i \in [n]}$. Conclude that $E_{(n:n)} = \max_{i \in [n]} E_i \stackrel{d}{=} \sum_{k=1}^n E_k/k$.*

By Exercise 3.3 and (3.2.45),

$$\begin{aligned}
(3.2.46) \quad \mathbb{P}(\Lambda \leq x) &= \lim_{n \rightarrow \infty} \mathbb{P}(\max_{i=1}^n E_i - \log n \leq x) = \lim_{n \rightarrow \infty} \mathbb{P}(E_1 \leq x + \log n)^n \\
&= \lim_{n \rightarrow \infty} [1 - e^{-(x+\log n)}]^n = [1 - e^{-x}/n]^n = e^{-e^{-x}},
\end{aligned}$$

so that Λ indeed has the Gumbel distribution. Further, we compute the distribution function of $\Lambda_1 - \Lambda_2$ as

$$(3.2.47) \quad \mathbb{P}(\Lambda_1 - \Lambda_2 \leq x) = \mathbb{E}[\mathbb{P}(\Lambda_1 \leq x + \Lambda_2 \mid \Lambda_2)] = \mathbb{E}[e^{-e^{-(x+\Lambda_2)}}] = \mathbb{E}[e^{-e^{-\Lambda_2}e^{-x}}].$$

When Λ has a Gumbel distribution, $E = e^{-\Lambda}$ has an exponential distribution (see Exercise 3.4), so that

$$(3.2.48) \quad \mathbb{P}(\Lambda_1 - \Lambda_2 \leq x) = \mathbb{E}[e^{-Ee^{-x}}] = (1 + e^{-x})^{-1} = e^x/(1 + e^x),$$

so that $\Gamma_1 - \Gamma_2$ has a logistic distribution. This completes the proof. \square

EXERCISE 3.4 (Distribution of $e^{-\Lambda}$). *Let Λ have a Gumbel distribution. Show that $E = e^{-\Lambda}$ has an exponential distribution with parameter 1.*

EXERCISE 3.5 (Integral representation of Euler-Mascheroni). *Use the fact that*

$$(3.2.49) \quad \sum_{k=1}^{\infty} \frac{1}{k} (E_k - 1) + \gamma$$

has a Gumbel distribution to conclude the following integral representation for the Euler-Mascheroni constant γ :

$$(3.2.50) \quad \gamma = \int_{-\infty}^{\infty} x e^{-x} e^{-e^{-x}} dx = \int_0^{\infty} e^{-y} \log y dy.$$

Lengths of minimal paths: the hopcount. So far, we have studied the *weights* of the minimal paths, but one might also ask how long they are, disregarding their weights, that is, how many edges they contain. Let $H_n(i, j)$ be the length of the path between i and j that has minimal weight, and recall from (3.1.3) that $H_n = H_n(U_1, U_2)$ where U_1, U_2 are two independent draws uniformly from $[n]$. Note that, conditionally on $U_1 \neq U_2$, H_n has the same distribution as $H_n(i, j)$. Here we will be primarily interested in the fluctuations of $H_n(i, j)$ and we show a central limit theorem for it:

THEOREM 3.3 (Central limit theorem for the hopcount). *Consider first-passage percolation on the complete graph K_n with i.i.d. exponential edge weights. As $n \rightarrow \infty$, for any fixed $i, j \in [n]$,*

$$(3.2.51) \quad \frac{H_n(i, j) - \log n}{\sqrt{\log n}} \xrightarrow{d} Z,$$

where Z is standard normal.

If we compare Theorem 3.1 and 3.3, then we see that the fluctuations of $H_n(i, j)$ are much larger than those of $n\mathcal{C}_n(i, j)$. The asymptotic normal distribution with asymptotically equal mean and variance reminds us of Figure 6, where the empirical mean and variance of the number of edges in the smallest-weight paths are close.

PROOF. The proof of Theorem 3.1 shows that the collection of minimal weight paths from a given vertex, say i , form a tree (rooted at i) that can be constructed as follows: Begin with a single root and add $n - 1$ vertices one by one, each time joining the new vertex to a (uniformly) randomly chosen old vertex. This type of random tree is known as a *random recursive tree*. Let H'_n denote the height of a random vertex in $[n]$, so that H_n has the same distribution as $H_n(i, U)$, where U is uniform in $[n]$. Further, conditionally

on $U \neq i$, $H_n(i, U)$ has the same distribution as $H_n(i, j)$, so that a central limit theorem for H'_n implies an identical one for $H_n(i, j)$. To investigate H'_n , we use the following characterization of its distribution:

LEMMA 3.4 (Distribution of the hopcount). *Fix $n \geq 2$. Let H'_n be the height in a random recursive tree of a uniformly chosen vertex in $[n]$. Then H'_n has the same distribution as $G_n = \sum_{i=2}^n I_i$, where I_i are independent Bernoulli random variables with success probability $1/i$.*

PROOF. We prove this lemma by induction on n . When $n = 2$, we have that $H'_2 \in \{0, 1\}$ each with probability $1/2$, and also $G_2 = I_2 \in \{0, 1\}$ with equal probability. This initializes the induction hypothesis.

To advance the induction hypothesis, we let V_n denote a random vertex in $[n]$. We condition on $V_n = n$ or not, to obtain

$$(3.2.52) \quad \mathbb{P}(H'_n = k) = \frac{1}{n}\mathbb{P}(H'_n = k \mid V_n = n) + \left(1 - \frac{1}{n}\right)\mathbb{P}(H'_n = k \mid V_n \neq n).$$

Now, $\mathbb{P}(H'_n = k \mid V_n \neq n) = \mathbb{P}(H'_{n-1} = k)$, since conditionally on $V_n \neq n$, V_n has a uniform distribution on $[n-1]$. Further, when $V_n = n$, we have that H'_n is the height of the last added vertex, which is equal to 1 plus the height of a uniformly chosen vertex in $[n-1]$ since the parent of vertex n is a uniform vertex in $[n-1]$. Therefore, $\mathbb{P}(H'_n = k \mid V_n = n) = \mathbb{P}(H'_{n-1} = k - 1)$, so that we arrive at the recursion relation

$$(3.2.53) \quad \mathbb{P}(H'_n = k) = \frac{1}{n}\mathbb{P}(H'_{n-1} = k) + \left(1 - \frac{1}{n}\right)\mathbb{P}(H'_{n-1} = k - 1).$$

We conclude that H'_n has the same distribution as $H'_{n-1} + I_n$, where I_n has a Bernoulli distribution with success probability $1/n$ independently of H'_{n-1} . This advances the induction hypothesis, and completes the proof of Lemma 3.4. \square

Theorem 3.3 follows immediately from Lemma 3.4, together with the fact that

$$(3.2.54) \quad \frac{\sum_{i=2}^n I_i - \log n}{\sqrt{\log n}} \xrightarrow{d} Z,$$

where Z is standard normal (see also Exercise 3.8 below). \square

EXERCISE 3.6 (Mean hopcount). *Show that $\mathbb{E}[H_n] = \sum_{k=2}^n 1/k$, which is the harmonic series.*

EXERCISE 3.7 (Variance hopcount). *Show that $\text{Var}(H_n) = \sum_{k=2}^n 1/k - \sum_{k=2}^n 1/k^2 = \mathbb{E}[H_n] - \zeta(2) + O(1/n)$, where $\zeta(2) = \sum_{k=1}^{\infty} 1/k^2$ is equal to the zeta-function at 2.*

EXERCISE 3.8 (CLT for sums of independent indicators). *Let $(I_i)_{i \geq 1}$ be a collection of independent $\text{Ber}(p_i)$ random variables, and assume that*

$$(3.2.55) \quad \sum_{i=1}^m p_i \rightarrow \infty, \quad \sum_{i=1}^m p_i^2 = o\left(\sum_{i=1}^m p_i\right).$$

Show that, as $m \rightarrow \infty$,

$$(3.2.56) \quad \frac{\sum_{i=1}^m (I_i - p_i)}{\sqrt{\sum_{i=1}^m p_i}} \xrightarrow{d} Z,$$

where Z is standard normal. Use this to prove (3.2.54).

EXERCISE 3.9 (Joint convergence). *Show that the convergence of $H_n(i, j)$ and $\mathcal{C}_n(i, j)$ in Theorems 3.2 and 3.3, respectively, occurs jointly, i.e.,*

$$(3.2.57) \quad \left(\frac{H_n(i, j) - \log n}{\sqrt{\log n}}, n\mathcal{C}_n(i, j) - \log n \right) \xrightarrow{d} (Z, \Lambda_1 + \Lambda_2 - \Lambda_{12}),$$

where Z is standard normal and $\Lambda_1, \Lambda_2, \Lambda_{12}$ are three independent Gumbel variables independent of Z .

Extremal functionals for FPP on the complete graph. We continue by describing some more recent results about first-passage percolation on the complete graph. We start with the fluctuations of the weight-diameter for first-passage percolation on the complete graph as derive with Bhamidi [39]. Let us first define the limiting weighted random graph that arises in the description of our main result. The vertex set of this graph is the set of positive integers $\mathbb{Z}_+ = \{1, 2, \dots\}$. Let PP be a Poisson process on \mathbb{R} with intensity measure having density

$$(3.2.58) \quad \lambda(y) = \exp(-y), \quad y \in \mathbb{R}.$$

Exercise 3.10 implies that $\Xi' = \max\{x : x \in \text{PP}\} < \infty$ a.s. Thus, we can order the points in PP as $Y_1 > Y_2 > \dots$. Let $\{\Lambda_{st} : s, t \in \mathbb{Z}_+, s < t\}$ be a family of independent standard Gumbel random variables. Define

$$(3.2.59) \quad \Xi = \max\{Y_s + Y_t - \Lambda_{st} : s, t \in \mathbb{Z}_+, s < t\}.$$

It can be shown that $\Xi < \infty$ almost surely:

EXERCISE 3.10 (The random variables Ξ, Ξ' are finite almost surely). *Show that $\Xi' < \infty$ and $\Xi < \infty$ almost surely.*

The main result concerning the weight-diameter for first-passage percolation on the complete graph is as follows:

THEOREM 3.5 (Fluctuations of the weight diameter). *For first-passage percolation on the complete graph with exponential edge weights with mean 1, as $n \rightarrow \infty$,*

$$(3.2.60) \quad n \max_{i, j \in [n]} \mathcal{C}_n(i, j) - 3 \log n \xrightarrow{d} \Xi.$$

The proof of Theorem 3.5 is a close adaptation of the proof of the upper bound in (3.2.3) in Theorem 3.1. For $i \in [n]$, let $X_i = \min_{j \in [n]} E_{ij}$ be the minimal edge weight from vertex i , and let $X_{(i)}$ be the order statistics of $(X_i)_{i \in [n]}$, so that $X_{(1)} < X_{(2)} < \dots < X_{(n)}$. Further, let V_i be the vertex corresponding to $X_{(i)}$. Then, obviously,

$$(3.2.61) \quad \max_{i, j \in [n]} \mathcal{C}_n(i, j) = \max_{i, j \in [n]} \mathcal{C}_n(V_i, V_j).$$

The point process $(nX_{(i)} - \log n)_{i \in [n]}$ converges to the point-process with intensity λ in (3.2.58). For $i < j$ fixed, we can write

$$(3.2.62) \quad \mathcal{C}_n(V_i, V_j) \stackrel{d}{=} X_{(i)} + X_{(j)} + \mathcal{C}_n(\{V_i, U_i\}, \{V_j, U_j\}),$$

where $\mathcal{C}_n(A, B)$ denotes the minimal weight between the subsets $A, B \subseteq [n]$ and U_i is such that $E_{V_i U_i} = X_{(i)}$ (and we have ignored the case where $E_{V_i V_j} = X_{(i)}$ or $E_{ij} = X_{(j)}$ for simplicity). The ideas used in Theorem 3.1 can be extended to deduce that

$$(3.2.63) \quad \mathcal{C}_n(\{i, n\}, \{j, n-1\}) - \log n \xrightarrow{d} \Lambda_i + \Lambda_j - \Lambda_{ij},$$

and $(\Lambda_i)_{i \geq 1}$ and $(\Lambda_{ij})_{1 \leq i < j}$ are independent random variables, with Λ_{ij} having a Gumbel distribution and

$$(3.2.64) \quad \Lambda_i = \lim_{m \rightarrow \infty} \sum_{k=2}^m E_k/k - \log m.$$

This suggests that

$$(3.2.65) \quad n \max_{i,j \in [n]} \mathcal{C}_n(V_i, V_j) - 3 \log n \xrightarrow{d} \max_{1 \leq i < j < \infty} (Y_i + Y_j + \Lambda_i + \Lambda_j - \Lambda_{ij}),$$

which is close to what we wish to prove, except that the point process $(Y_i)_{i \geq 1}$ is replaced by the point process $(Y_i + \Lambda_i)_{i \geq 1}$. Interestingly, these have the same distribution, which explains Theorem 3.5.

We continue by investigating the lengths of the maximal paths. Indeed, recall that $H_n(i, j)$ is the length of the optimal path between vertices $i, j \in [n]$, following Addario-Berry, Broutin and Lugosi [7]. The next theorem identifies the first order asymptotics of $\max_{j \in [n]} H_n(1, j)$ and $\max_{i,j \in [n]} H_n(i, j)$:

THEOREM 3.6 (Hopcount diameter). *For first-passage percolation on the complete graph with exponential edge-weights with mean 1, as $n \rightarrow \infty$,*

$$(3.2.66) \quad \frac{\max_{j \in [n]} H_n(1, j)}{\log n} \xrightarrow{\mathbb{P}} e, \quad \frac{\max_{i,j \in [n]} H_n(i, j)}{\log n} \xrightarrow{\mathbb{P}} \alpha^*,$$

where $\alpha^* \approx 3.5911$ is the unique solution of the equation $\alpha \log \alpha - \alpha = 1$.

EXERCISE 3.11 (Equation for the height). *Note that e is the unique solution to the equation $\alpha \log \alpha - \alpha = 0$.*

EXERCISE 3.12 (Height of URT). *Show that $\max_{j \in [n]} H_n(1, j)$ is the same as the height of the uniform recursive tree. Conclude that*

$$(3.2.67) \quad \frac{\max_{j \in [n]} H_n(1, j)}{\log n} \xrightarrow{\mathbb{P}} e.$$

We now give some intuition for the results in Theorem 3.6, relying on the analysis in the proof of Theorem 3.3. We only prove the upper bound. First note that, by Boole's inequality,

$$(3.2.68) \quad \mathbb{P}(\max_{i,j \in [n]} H_n(i, j) \geq \alpha \log n) \leq n \mathbb{P}(\max_{j \in [n]} H_n(1, j) \geq \alpha \log n).$$

Again by Boole's inequality,

$$(3.2.69) \quad \mathbb{P}(\max_{j \in [n]} H_n(1, j) \geq \alpha \log n) \leq n \mathbb{P}(H_n(1, 2) \geq \alpha \log n),$$

so that, for the upper bound, it suffices to bound the tails of the distribution of $H_n(1, 2)$. By Lemma 3.4, for every $t \geq 0$,

$$(3.2.70) \quad \mathbb{P}(H_n(1, 2) \geq \alpha \log n) \leq e^{-t\alpha \log n} \mathbb{E}[e^{tH_n(1,2)}] = e^{-t\alpha \log n} \prod_{i=2}^n \left(1 - \frac{1}{i} + e^t \frac{1}{i}\right).$$

We further bound, using $1 + x \leq e^x$ and $\sum_{i=2}^n 1/i \leq \int_1^n dx/x = \log n$,

$$(3.2.71) \quad \prod_{i=2}^n \left(1 - \frac{1}{i} + e^t \frac{1}{i}\right) = \prod_{i=2}^n \left(1 + (e^t - 1) \frac{1}{i}\right) \leq e^{\sum_{i=1}^n (e^t - 1) \frac{1}{i}} \leq e^{(e^t - 1) \log n}.$$

This leads to

$$(3.2.72) \quad \mathbb{P}(H_n(1, 2) \geq \alpha \log n) \leq e^{\log n(-t\alpha + e^t - 1)}.$$

Optimizing over $t \geq 0$ yields $t = \log \alpha$, so that

$$(3.2.73) \quad \mathbb{P}(H_n(1, 2) \geq \alpha \log n) \leq n^{-(\alpha \log \alpha - \alpha + 1)}.$$

Thus,

$$(3.2.74) \quad \mathbb{P}(\max_{j \in [n]} H_n(1, j) \geq \alpha \log n) = o(1) \quad \text{when} \quad \alpha \log \alpha - \alpha + 1 > 1,$$

which occurs when $\alpha > e$. Further,

$$(3.2.75) \quad \mathbb{P}(\max_{i, j \in [n]} H_n(i, j) \geq \alpha \log n) = o(1) \quad \text{when} \quad \alpha \log \alpha - \alpha + 1 > 2,$$

which occurs when $\alpha > \alpha^*$, α^* being the solution of $\alpha \log \alpha - \alpha = 1$. We conclude that the upper bounds in Theorem 3.6 hold. The respective lower bounds are much harder to prove, as one has to show that whp there really exists a vertex i for which $H_n(1, i) \geq \alpha \log n$ for any $\alpha < e$, and that whp there really exist two vertices i, j for which $H_n(i, j) \geq \alpha \log n$ for any $\alpha < \alpha^*$.

The lower bound on $\max_{i, j \in [n]} H_n(i, j)$ in (3.2.66) is the hard part, and is proved in [7] using a delicate conditional second moment method. The basic starting point is that if a path $\pi: v_0 \rightarrow v_k$ has a weight $w(\pi) = \sum_{e \in \pi} E_e$ that satisfies $(1 - \varepsilon) \log n \leq w(\pi) \leq (1 + \varepsilon)w(\pi)$, then whp it is the shortest path (see [7, Corollary 5]). In turn, let us call $\mathcal{P}_{k, \varepsilon} = \mathcal{P}_{k, \varepsilon}(n)$ be the set of paths with k edges and weight at most $(1 - \varepsilon) \log n$. Then it is not hard to show that

$$(3.2.76) \quad \mathbb{E}[\mathcal{P}_{k, \varepsilon}(n)] \sim n^{k+1} \left(\frac{(1 - \varepsilon) \log n}{n} \right)^k \frac{1}{k!} \sim \frac{n}{\sqrt{2\pi k}} \left(\frac{e \log n}{k} \right)^k.$$

This tends to infinity when $\alpha < \alpha^*$ and $\varepsilon > 0$ is sufficiently small. It would thus be tempting to use a second moment directly on $\mathcal{P}_{k, \varepsilon}(n)$, but there is too much dependence between these paths. Instead, Addario-Berry, Broutin and Lugosi [7] use a second moment method on a subclass of paths in $\mathcal{P}_{k, \varepsilon}(n)$, which has a high enough mean (comparable to that of $\mathcal{P}_{k, \varepsilon}(n)$), yet for which the second moment can be controlled.

Theorem 3.6 only studies *first-order* asymptotic of extremal hopcounts for first-passage percolation on the complete graph with exponential edge-weights, we lack any results about their fluctuations:

OPEN PROBLEM 3.1 (Fluctuations of extremal hopcount). *How do the fluctuations of $\max_{j \in [n]} H_n(1, j) - e \log n$ and $\max_{i, j \in [n]} H_n(i, j) - \alpha^* \log n$ in Theorem 3.6 behave?*

Conclusion of first-passage percolation on the complete graph. In this section, we have studied first-passage percolation on the complete graph with i.i.d. exponential edge-weights. We have concluded that $\mathcal{C}_n(i, j)$ is roughly $(\log n + \Lambda_1 + \Lambda_2 - \Lambda_{12})/n$ for three i.i.d. Gumbel distributions $\Lambda_1, \Lambda_2, \Lambda_{12}$. We can interpret $1/n$ as corresponding to order of the minimal edge-weight from each vertex, which can also be seen as the inverse of the degree of a vertex.

Of course, the exponential edge-weights are rather special, as this makes the growth of the smallest-weight tree and the smallest-weights along it completely independent. It is worthwhile studying first-passage percolation on the complete graph with different edge-weights. Of course, what really matters on the complete graph are *small* weights. When these edge-weights have a finite and strictly positive density in 0, one can expect that the scaling behavior is identical to that with exponential edge-weights. One can even expect the hopcount H_n to behave in a similar way. Janson [179] indeed proves this. However, when the random edge-weights have rather different behavior close to zero, then it is not so clear how first-passage percolation on the complete graph behaves. We will discuss the available results in more detail in Section 3.7.1.

First-passage percolation on random graphs. In what follows, we will focus on first-passage percolation on random graphs, where the average degree is bounded. Then, the smallest weight between two random vertices cannot be expected to vanish as in Theorem 3.1, but rather the factor $1/n$ should be replaced with a constant instead. This suggests that $\mathcal{C}_n(U_1, U_2)$ grows like $\log n$ with fluctuations that are tight random variables. Our results below will show that this is indeed the case, at least in the case where the random graphs have finite-variance degrees. It even turns out that the properly re-centered version of $\mathcal{C}_n(U_1, U_2)$ converges to a limit that bears close resemblance to the limit $\Lambda_1 + \Lambda_2 - \Lambda_{12}$ for the complete graph. How $H_n(U_1, U_2)$ behaves is less clear from the above discussion, and we will again see central limit theorems results for it very alike those in Theorem 3.3. We start with the setting of the configuration model with finite-variance degrees and exponential edge-weights, where again the memoryless property can be cleverly used to reduce first-passage percolation on the random graph to a *Markovian* problem.

Some first-passage percolation problems can also be analysed and understood in terms of *local weak convergence*. In fact, the objective method of Aldous and Steele [12] was to a large extent inspired by problems of weighted random (and in particular complete) graphs. Let us explain this in some detail now. For first-passage percolation on the complete graph, we can introduce a limiting metric space that describes the smallest-weight distances to the neighbors starting from a uniform vertex. Indeed, rescale every edge-weight by a factor n , so that the edge-weights are i.i.d. exponential random variables with mean n . Then, the weights of the edges to the smallest-weight neighbors of vertex 1 are close to sums of i.i.d. exponential random variables with mean 1. This is a rate 1 Poisson process. Now, when starting from any neighbor of vertex 1, we can again explore its closest neighbors in the ed-weight metric, and this is again close to a rate 1 Poisson process. This leads to a metric space in which every vertex has infinitely many neighbors, and the edge-weight to these neighbors form i.i.d. rate 1 Poisson processes. This metric structure is called the *Poisson-weighted infinite tree* or PWIT.

The setting is slightly different from that studied in Section 2.2, as there we had settings in mind where every vertex has a random but almost surely *bounded* number

of neighbors. However, when we are only interested in the structure of neighborhoods up to a given metric distance, then only finite relevant connections remain, and thus we can interpret the metric space as having ‘almost finite degrees’. In the above light, first-passage percolation on the complete graph then, locally, is closely related to first-passage percolation on the PWIT. This turns out to be a *continuous-time branching process*, which goes under the name of the *Yule process*. We discuss continuous-time branching processes in detail in Section 3.4, and crucially use relations to such processes in the sequel. For example, the general first-passage percolation problem on configuration models with finite-variance degrees could be interpreted in terms of its local-weak limit (including the edge-weights) being an age-dependent branching process, which is a continuous-time branching process in which the birth-times of individuals are drawn from an i.i.d. distribution. The PWIT will return in several related problems, such as the minimal-spanning tree on the complete graph.

Let us close this discussion by highlighting the form of the limiting random variable $\Lambda_1 + \Lambda_2 - \Lambda_{12}$ in (3.2.37). For first-passage percolation on the PWIT, which is a Yule process, we will see that the number of vertices that can be reached within time t grows like Ee^t , where E is an exponential random variable with parameter 1. The random variable E acts as a measure of how long it takes to escape the neighborhood of the root. A large E signifies that many vertices are close by, while a small E signifies that it takes a significant amount of time to reach some neighbors. We can thus interpret $\Lambda_1 + \Lambda_2 - \Lambda_{12} = \log(1/E_1) + \log(1/E_2) - \Lambda_{12}$, where $\log(1/E_1), \log(1/E_2)$ quantify how quickly the flow starts finding neighbors from the respective starting and ending points, while Λ_{12} measures how quickly the smallest-weight graphs find each other. We will see that this is a rather general structure for smallest-weight paths between uniformly chosen vertices.

3.3. Markovian flow on the configuration model

In this section, we investigate first-passage percolation on the configuration model with exponential edge-weights. Our main aim is to describe the asymptotics in distribution of H_n and \mathcal{C}_n on $\text{CM}_n(\mathbf{d})$ where the degree distribution has finite variance. The main result of this section is the following theorem:

THEOREM 3.7 (Hopcount and minimal path weight for CM with exponential edge-weights). *Let $\text{CM}_n(\mathbf{d})$ satisfy Condition 1.6(a)-(c) and $\nu = \mathbb{E}[D(D-1)]/\mathbb{E}[D] > 1$. Further, assume that $\nu_n = \mathbb{E}[D_n(D_n - 1)]/\mathbb{E}[D_n] = \nu + o(1/\log n)$ and that there exists $\varepsilon > 0$ such that*

$$(3.3.1) \quad \lim_{n \rightarrow \infty} \mathbb{E}[D_n^{2+\varepsilon}] = \mathbb{E}[D^{2+\varepsilon}].$$

Let the edge-weights be given by i.i.d. exponential random variables with parameter 1. Then, there exists a random variable Q such that, conditionally on $\mathcal{C}_n < \infty$,

$$(3.3.2) \quad \frac{H_n - \alpha \log n}{\sqrt{\alpha \log n}} \xrightarrow{d} Z, \quad \mathcal{C}_n - \frac{1}{\lambda} \log n \xrightarrow{d} Q,$$

where Z has a standard normal distribution, and

$$(3.3.3) \quad \alpha = \nu/(\nu - 1), \quad \lambda = \nu - 1.$$

Theorem 3.7 is proved in work with Bhamidi and Hooghiemstra [40] in the setting of i.i.d. degrees (for which $\nu_n = \mathbb{E}_n[D_n(D_n - 1)]/\mathbb{E}_n[D_n] = \nu + o(1/\log n)$ can be shown

to hold, where $\mathbb{E}_n[D_n(D_n - 1)]$ and $\mathbb{E}_n[D_n]$ correspond to average of such i.i.d. random variables). By Theorem 2.29, we know that the graph distance $d_{\text{CM}_n(\mathbf{d})}(U_1, U_2)$ between two uniformly chosen vertices satisfies

$$(3.3.4) \quad d_{\text{CM}_n(\mathbf{d})}(U_1, U_2) / \log n \xrightarrow{\mathbb{P}} 1 / \log \nu.$$

By Theorem 3.7, $H_n / \log n \xrightarrow{\mathbb{P}} \alpha = \nu / (\nu - 1) > 1 / \log \nu$, so that the smallest-weight path is substantially longer than the shortest path realizing the graph distance. In turn,

$$(3.3.5) \quad \mathcal{C}_n / \log n \xrightarrow{\mathbb{P}} 1 / (\nu - 1) < 1 / \log \nu,$$

while the expected weight of any shortest path has the same asymptotics as in (3.3.4), so that the smallest-weight path has a significantly smaller weight. Thus, the topology of the configuration model $\text{CM}_n(\mathbf{d})$ is significantly altered by adding i.i.d. weights along the edges.

We next discuss the fluctuations of H_n and \mathcal{C}_n in Theorem 3.7. By the remark below Theorem 2.29, the centered graph distance $d_{\text{CM}_n(\mathbf{d})}(U_1, U_2) - \log n / \log \nu$ is a tight sequence of random variables that does *not* converge in distribution. The latter is due to the discrete nature of $d_{\text{CM}_n(\mathbf{d})}(U_1, U_2)$. In Theorem 3.7, H_n instead satisfies a CLT with asymptotic mean and variance equal to $\alpha \log n$, where $\alpha = \nu / (\nu - 1)$. Thus, it has much larger fluctuations than the graph distance. The centered smallest-weight $\mathcal{C}_n - \log n / \lambda$ is a continuous random variable and thus does not suffer from the discreteness as $d_{\text{CM}_n(\mathbf{d})}(U_1, U_2)$ does. Theorem 3.7 shows that this is reflected in the fact that it converges in distribution.

In order to intuitively understand this asymptotics, we see that, for first-passage percolation on $\text{CM}_n(\mathbf{d})$, the number of vertices that can be reached with weight at most t grows proportional to $e^{\lambda t}$. As a result, when we draw an alive vertex uniformly at random from all the alive vertices at time t , then its life-time will be a tight random variable (in fact, it has an asymptotically exponential distribution with parameter $1/\lambda$). This is a hint at why the fluctuations of \mathcal{C}_n converge in distribution. Instead, if we draw a uniform alive vertex at time t and look at its generation, then it has an approximate normal distribution, which suggests that the same must be true for H_n . We discuss this matter in more detail below.

The limiting random variable Q in (3.3.2) can be identified as

$$(3.3.6) \quad Q = Q_1 + Q_2 - \frac{\Lambda}{\nu - 1} + \frac{\log \mu(\nu - 1)}{\nu - 1},$$

where Q_1, Q_2 are two independent copies of a certain limiting random variable and Λ has a Gumbel distribution independent from Q_1, Q_2 , and where we write, throughout this section, $\mu = \mathbb{E}[D]$.

We will not give a full proof of Theorem 3.7, but rather give a sketch. We start by describing the key steps in the proof of Theorem 3.7:

Local structure of first-passage percolation on $\text{CM}_n(\mathbf{d})$. Theorem 2.11 shows that the local structure around a uniform vertex in $\text{CM}_n(\mathbf{d})$ is close to a two-stage branching process, where the root has offspring distribution D appearing in Condition 1.6(a), while all other individuals have offspring $D^* - 1$, where D^* is the size-biased version of D . This leads us to study first-passage percolation on Galton-Watson trees. Due to the memoryless property of the exponential distribution, this process has a Markovian structure that is closely related to, but slightly more involved than, that on the complete graph

as in the proofs of Theorems 3.1, 3.2 and 3.3. We start by describing this construction on *deterministic trees* where the i th explored vertex has degree d_i . Later, we will apply this to $d_1 = D$, and $d_i = D_i^* - 1$ for $i \geq 2$, where these random variables are independent. However, since the construction is true for *any* tree, in fact we may also apply it to situations where the distributions of the d_i are weakly dependent, which will turn out to be highly useful in the sequel.

We now present the details. Suppose we have positive (non-random) integers d_1, d_2, \dots . Consider the following construction of a branching tree:

CONSTRUCTION 3.1 (Flow from root of tree and split-times). *The smallest-weight graph and split times on a tree with degrees $(d_i)_{i \geq 1}$ are obtained as follows:*

- (1) *Start with the root which dies immediately giving rise to d_1 alive offspring;*
- (2) *Each alive offspring lives for a random amount of time, which has an exponential distribution with parameter 1, independent of all other randomness involved;*
- (3) *When the m th vertex dies, it leaves behind d_m alive offspring.*

In Construction 3.1, the number of offspring d_i is fixed once and for all. It is crucial that there always is at least one alive vertex. This occurs precisely when $s_i = d_1 + \dots + d_i - (i - 1) \geq 1$ for every $i \geq 1$:

EXERCISE 3.13 (The number of alive vertices). *Show that the number of alive vertices after the i th step in Construction 3.1 is equal to*

$$(3.3.7) \quad s_i = d_1 + \dots + d_i - (i - 1).$$

Thus, this construction continues forever precisely when $s_i \geq 1$ for every $i \geq 1$.

The split-times (or death-times) of the process in Construction 3.1 are denoted by $(T_i)_{i \geq 0}$, where, by convention, $T_0 = 0$ and T_1 has an exponential distribution with parameter d_1 . We refer to i as *time*. Note that Construction 3.1 is equivalent to this process, observed at the discrete times $(T_i)_{i \geq 0}$. The fact that the chosen alive vertex is chosen uniformly at random follows from the memoryless property of the exponential random variables that compete to become the minimal one, and makes the process Markovian.

Markovian first-passage percolation on Galton-Watson trees. We continue to investigate the *generation* and *weight* to the m th chosen alive vertex in our first-passage percolation problem on the tree:

PROPOSITION 3.8 (Smallest-weight paths on a tree). *Pick an alive vertex at time $m \geq 1$ uniformly at random among all vertices alive at this time. Then,*

(a) *the weight of the smallest-weight path between the root of the tree and the vertex chosen at time m is equal in distribution to*

$$(3.3.8) \quad T_m \stackrel{d}{=} \sum_{i=1}^m E_i / s_i,$$

where $(E_i)_{i \geq 1}$ are i.i.d. exponential random variables with mean 1 and $s_i = d_1 + \dots + d_i - (i - 1)$ denotes the number of alive individuals at time i ;

(b) *the generation of the vertex chosen at time m is equal in distribution to*

$$(3.3.9) \quad G_m \stackrel{d}{=} \sum_{i=1}^m I_i,$$

where $(I_i)_{i \geq 1}$ are independent Bernoulli random variables with

$$(3.3.10) \quad \mathbb{P}(I_i = 1) = d_i/s_i.$$

Proposition 3.8(a) is a generalization of the distribution of the time to find the m th vertex for first-passage on the complete graph in (3.2.6). Further, Proposition 3.8(b) is a generalization of the distribution of the hopcount for first-passage on the complete graph in Lemma 3.4. We see that the proof of Proposition 3.8 also bears close resemblance to the proof for the corresponding results on the complete graph:

PROOF. We start by proving part (a). At each time m , and conditionally on $(T_i)_{i=1}^{m-1}$, the memoryless property of the exponential distribution implies that $T_m - T_{m-1}$ has an exponential distribution with parameter s_m . Here we use that the minimum of s_i independent exponential 1 random variables has an exponential distribution with parameter s_i , and is hence equal in distribution to E_i/s_i . Therefore, conditionally on $(T_i)_{i=1}^{m-1}$,

$$(3.3.11) \quad T_m - T_{m-1} \stackrel{d}{=} E_m/s_m,$$

where E_m is independent from $(T_i)_{i=1}^{m-1}$. This proves the claim in part (a).

We prove part (b) by induction on $m \geq 1$. The statement is trivial for $m = 1$, since then $G_1 = 1$ a.s., while $s_1 = d_1$, so that also $I_1 = 1$ a.s. This initializes the induction hypothesis.

We next assume that (3.3.9) holds for m , where $(I_i)_{i=1}^m$ are independent Bernoulli random variables satisfying (3.3.10). We then advance the induction hypothesis to $m + 1$, by showing that G_{m+1} has the distribution in (3.3.9).

Let G_{m+1} denote the generation of the randomly chosen vertex at time $m + 1$, and consider the event $\{G_{m+1} = k\}$ for $1 \leq k \leq m$. If randomly choosing one of the alive vertices at time $m + 1$ results in one of the d_{m+1} newly added vertices, then, in order to obtain generation k , the previous uniform choice, i.e., the choice of the vertex which was the last one to die, must have been a vertex from generation $k - 1$. On the other hand, if a uniform pick is conditioned on not taking one of the d_{m+1} newly added vertices, then this choice is a uniform alive vertex from generation k . Hence, we obtain, for $1 \leq k \leq m$,

$$(3.3.12) \quad \mathbb{P}(G_{m+1} = k) = \frac{d_{m+1}}{s_{m+1}} \mathbb{P}(G_m = k - 1) + \left(1 - \frac{d_{m+1}}{s_{m+1}}\right) \mathbb{P}(G_m = k).$$

As a result, $G_{m+1} \stackrel{d}{=} G_m + I_{m+1}$, where I_{m+1} is a Bernoulli variable with success probability d_{m+1}/s_{m+1} . The proof of part (a) is now immediate from the induction hypothesis. \square

EXERCISE 3.14 (Gumbel law for deterministic trees). Let $d_1 = r - 2$ and $d_i = r - 1$ for every $i \geq 2$. Use Proposition 3.8 to conclude that

$$(3.3.13) \quad T_m - \frac{\log m}{r - 2} \xrightarrow{d} \Lambda/(r - 2),$$

where Λ has a Gumbel distribution.

EXERCISE 3.15 (CLT for height of uniform point in deterministic tree). Let $d_1 = r - 2$ and $d_i = r - 1$ for every $i \geq 2$. Use Proposition 3.8 to conclude that

$$(3.3.14) \quad \frac{G_m - \frac{r-1}{r-2} \log m}{\sqrt{\frac{r-1}{r-2} \log m}} \xrightarrow{d} Z,$$

where Z has a standard normal distribution.

Back to Markovian first-passage percolation on $\text{CM}_n(\mathbf{d})$. We next intuitively relate the above result to our setting. Start from vertex U_1 , and iteratively choose the half-edge with minimal additional weight attached to the SWG so far. As mentioned before, because of the memoryless-property of the exponential distribution, the half-edge with minimal additional weight can be considered to be picked uniformly at random from all half-edges incident to the SWG at that moment.

With high probability, this half-edge is paired to a half-edge that is not incident to the SWG. Let $\Delta_i^{(n)}$ denote the *forward degree* (i.e., the degree minus 1) of the vertex incident to the half-edge to which the i th half-edge is paired. By the results in Chapter 2, see in particular Theorem 2.11, $(\Delta_i^{(n)})_{i \geq 2}$ are close to being i.i.d., and have distribution given by $D^* - 1$. Therefore, we are lead to studying random variables of the form (3.3.9)–(3.3.10), where $(\Delta_i)_{i \geq 1}$ are independent random variables with $D_1 \stackrel{d}{=} D$ and $\Delta_i \stackrel{d}{=} D^* - 1$. Thus, we study the *unconditional* law of G_m in (3.3.9), in the setting where the vector $(d_i)_{i \geq 1}$ is replaced by a sequence of independent random variables $(\Delta_i)_{i \geq 1}$. We first state a CLT for G_m and a limit result for T_m in this setting.

Let us recall the definitions of T_m and G_m in this setting, and introduce some notation. Let $S_i^{(\text{ind})} = D_1 + \Delta_2 + \cdots + \Delta_i - (i - 1)$, where $D_1 \stackrel{d}{=} D$ and $\Delta_i \stackrel{d}{=} D_i^* - 1$, and $(\Delta_i)_{i \geq 1}$ are independent. Then, define

$$(3.3.15) \quad T_m = \sum_{i=1}^m E_i / S_i^{(\text{ind})}, \quad G_m \stackrel{d}{=} \sum_{i=1}^m I_i,$$

where, conditionally on $(S_i^{(\text{ind})})_{i \geq 1}$, $(I_i)_{i \geq 1}$ are independent Bernoulli random variables with

$$(3.3.16) \quad \mathbb{P}(I_i = 1) = \Delta_i / S_i^{(\text{ind})}.$$

Then, Proposition 3.8 implies the following limiting results for T_m and G_m :

COROLLARY 3.9 (Asymptotics for smallest-weight paths on trees). *Let $(\Delta_i)_{i \geq 1}$ be a sequence of independent and identically distributed non-degenerate integer-valued random variables satisfying $\mathbb{E}[\Delta_i^{1+\varepsilon}] < \infty$ for some $\varepsilon > 0$. Denote $\nu = \mathbb{E}[\Delta_1]$. Then, conditionally on $S_i^{(\text{ind})} = D_1 + \Delta_2 + \cdots + \Delta_i - (i - 1) \geq 1$ for all $i \geq 1$,*

(a) *for T_m given by (3.3.15), there exists a random variable X such that*

$$(3.3.17) \quad T_m - \frac{1}{\nu - 1} \log m \xrightarrow{d} X.$$

(b) *for G_m given in (3.3.15)–(3.3.16), with $\alpha = \nu / (\nu - 1) \geq 1$, as $m \rightarrow \infty$,*

$$(3.3.18) \quad \frac{G_m - \alpha \log m}{\sqrt{\alpha \log m}} \xrightarrow{d} Z,$$

where Z is a standard normal variable.

We next identify the limit of $T_m - (\log m) / \lambda$ where $\lambda = \nu - 1$. We start from (3.3.8), and split

$$(3.3.19) \quad T_m = \sum_{i=1}^m \frac{E_i - 1}{S_i^{(\text{ind})}} + \sum_{i=1}^m \left(\frac{1}{S_i^{(\text{ind})}} - \frac{1}{(\nu - 1)i} \right) + \frac{1}{\nu - 1} \sum_{i=1}^m \frac{1}{i}.$$

On the event that $S_i^{(\text{ind})} \geq 1$ for every $i \geq 1$, by the Strong Law of Large Numbers,

$$(3.3.20) \quad S_i^{(\text{ind})}/i \xrightarrow{\text{a.s.}} \nu - 1,$$

As a result, there are a.s. only finitely many values for which $S_i \leq (1 - \varepsilon)(\nu - 1)i$. Denote $M_m = \sum_{i=1}^m (E_i - 1)/S_i^{(\text{ind})}$, and note that, conditionally on $(S_i^{(\text{ind})})_{i \geq 0}$, $(M_m)_{m \geq 1}$ is a martingale with conditional second moment

$$(3.3.21) \quad \mathbb{E}[M_m^2 \mid (S_i^{(\text{ind})})_{i \geq 0}] = \sum_{m=1}^{\infty} 1/(S_i^{(\text{ind})})^2,$$

which is a.s. bounded on the events that $S_i^{(\text{ind})} \geq 1$ for every $i \geq 1$ and that $S_i^{(\text{ind})} \leq (1 - \varepsilon)(\nu - 1)i$ except for finitely many i . As a result, for a.e. $(S_i^{(\text{ind})})_{i \geq 0}$, M_m and converges a.s. to its limit, which equals $M_\infty = \sum_{i=1}^{\infty} (E_i - 1)/S_i^{(\text{ind})}$. For the second term in (3.3.19), we can rewrite

$$(3.3.22) \quad \sum_{i=1}^m \left(\frac{1}{S_i^{(\text{ind})}} - \frac{1}{(\nu - 1)i} \right) = \sum_{i=1}^m \left(\frac{1}{S_i^{(\text{ind})}} - \frac{1}{\mathbb{E}[S_i^{(\text{ind})}]} \right) + \sum_{i=1}^{\infty} \left(\frac{1}{(\mu + (\nu - 1)(i - 1))} - \frac{1}{(\nu - 1)i} \right),$$

and

$$(3.3.23) \quad \sum_{i=1}^{\infty} \left(\frac{1}{(\mu + (\nu - 1)(i - 1))} - \frac{1}{(\nu - 1)i} \right) = \frac{1}{\nu - 1} \sum_{i=1}^{\infty} \frac{\nu - \mu - 1}{i(\mu + (\nu - 1)(i - 1))},$$

which is finite. Further, on the event that $S_i^{(\text{ind})} \geq 1$ for every $i \geq 1$ and that $S_i^{(\text{ind})} \leq (1 - \varepsilon)(\nu - 1)i$ except for finitely many i ,

$$(3.3.24) \quad \begin{aligned} \sum_{i=1}^m \left(\frac{1}{S_i^{(\text{ind})}} - \frac{1}{\mathbb{E}[S_i^{(\text{ind})}]} \right) &= - \sum_{i=1}^m \frac{S_i^{(\text{ind})} - \mathbb{E}[S_i^{(\text{ind})}]}{S_i^{(\text{ind})} \mathbb{E}[S_i^{(\text{ind})}]} \\ &= \Theta_{\mathbb{P}}(1) \sum_{i=1}^m \frac{|S_i^{(\text{ind})} - \mathbb{E}[S_i^{(\text{ind})}]|}{i^2}. \end{aligned}$$

Since $S_i^{(\text{ind})} = D_1 + \sum_{j=2}^i \Delta_j + (i - 1)$ and the summands have a $(1 + \varepsilon)$ bounded moment, $\mathbb{E}[|S_i^{(\text{ind})} - \mathbb{E}[S_i^{(\text{ind})}]|] = O(i^{1/(1+\varepsilon)})$, which implies that the random variable on the r.h.s. of (3.3.24) has a finite mean. Finally,

$$(3.3.25) \quad \sum_{i=1}^m 1/i - \log m \rightarrow \gamma,$$

where γ is Euler-Mascheroni's constant. This identifies the limit of X_1 of $T_m - (\log m)/(\nu - 1)$ as

$$(3.3.26) \quad X_1 = \sum_{i=1}^{\infty} (E_i - 1)/S_i^{(\text{ind})} + \sum_{i=1}^{\infty} (1/S_i^{(\text{ind})} - 1/[(\nu - 1)i] + \gamma/(\nu - 1)).$$

In turn, (3.3.26) identifies the limit law Q_1 appearing in the limit law of $\mathcal{C}_n - \gamma \log n$ in (3.3.6).

The above results are highly suggestive, but of course do not lead to a rigorous proof of Theorem 3.7. Indeed, the forward degrees are not quite independent, and thus we have to take an error into account. More importantly, the smallest-weight graph from a vertex can only be expected to be well described by a tree when it has size *much smaller* than n .

As a result, we cannot grow the smallest-weight graph from vertex U_1 up to the moment that it contains U_2 , as this would take time that is $\Theta_{\mathbb{P}}(n)$. Both these effects make the proof for the configuration model considerably more involved. Let us briefly mention how such problems can be resolved, postponing a more rigorous analysis to Section 3.5, where we indicate a proof for a much more general setting.

Impurities in the coupling to a branching process. It turns out that as long as we do not grow the smallest-weight graph from U_1 too large, we can handle the impurities in the degree distribution quite well. Indeed, the size-biased reordering and an i.i.d. sample are not too far away from each other, which can be used effectively. Special care is needed to handle cycles in the smallest-weight graphs, but these can be resolved by creating some artificial vertices and relying on the memoryless property of the exponentially-distributed edge-weights. We refrain from giving more details.

Growing two smallest-weight graphs. Rather than growing the smallest-weight graph only from vertex U_1 , we grow two such smallest-weight graphs from both vertices U_1 and U_2 . We grow the one from vertex U_1 up to size \sqrt{n} , and the one from U_2 for as long as it needs to find the smallest-weight path from U_1 to U_2 . The nice thing is that the exponential distribution is *memoryless*, so that the remaining life-times of the half-edges incident to smallest-weight graph from U_1 again have i.i.d. exponential distributions with parameter 1. Thus, we can instead think of the weight of any connecting edge between the smallest-weight graph from U_1 and that of U_2 as being associated to the half-edge that is incident to a vertex closer to U_2 than to U_1 . Thus, when we ‘find’ a half-edge, it will pair to a *uniform* half-edge that is still available, so that each time a connection between the smallest-weight graphs from U_1 and U_2 is created with probability equal to the number of unpaired half-edges incident to the smallest-weight graph from U_1 divided by the total number of available half-edges. However, this does imply that we need to study the dynamics of the first connection between the two smallest-weight graphs, which makes the analysis a little more tricky. Let us give some more details of this analysis.

The connecting edge. Let $a_n = \sqrt{n}$ and let $\text{SWG}_{a_n}^{(1)}$ denote the smallest-weight graph of vertex U_1 when it has size $a_n = \sqrt{n}$. Here, such a smallest-weight graph consists of all the a_n vertices that are closest to \mathbb{Z}_1^d as being the *dead* individuals, whereas the half-edges incident to them that are yet to be paired are the *alive* individuals. It is through these half-edges that the smallest-weight path can be realized. Further, let $\text{SWG}_m^{(2)}$ denote the smallest-weight graph of vertex U_2 when it has size m , for $m \geq 1$. As described above, we grow the two smallest-weight graphs until the first half-edge with minimal weight incident to $\text{SWG}_m^{(2)}$ is paired to a half-edge incident to $\text{SWG}_{a_n}^{(1)}$. We call the created edge linking the two smallest-weight graphs the *connecting edge*. More precisely, let the *connection time* be given by

$$(3.3.27) \quad C_n = \min\{m \geq 0: \text{SWG}_{a_n}^{(1)} \cap \text{SWG}_m^{(2)} \neq \emptyset\}.$$

Then C_n is the first time that $\text{SWG}_{a_n}^{(1)}$ and $\text{SWG}_m^{(2)}$ share a vertex. When $m = 0$, this means that $U_2 \in \text{SWG}_{a_n}^{(1)}$ (which happens with small probability of order $a_n/n \approx 1/\sqrt{n}$), while when $m \geq 1$, this means that the m th half-edge of $\text{SWG}^{(2)}$ that is chosen and then paired, is paired to a half-edge from $\text{SWG}_{a_n}^{(1)}$. The path found then is the smallest-weight path between vertices 1 and 2, since $\text{SWG}_{a_n}^{(1)}$ and $\text{SWG}_m^{(2)}$ precisely consists of the closest vertices to the roots U_i , for $i = 1, 2$, respectively.

For $i \in \{1, 2\}$, let $H_n^{(i)}$ denote the length of the smallest-weight path between vertex U_i and the common vertex in $\text{SWG}_{a_n}^{(1)}$ and $\text{SWG}_{C_n}^{(2)}$, so that

$$(3.3.28) \quad H_n = H_n^{(1)} + H_n^{(2)}.$$

Since at time C_n we have found the smallest-weight path,

$$(3.3.29) \quad (H_n^{(1)}, H_n^{(2)}) \stackrel{d}{=} (G_{a_n+1}^{(1)} - 1, G_{C_n}^{(2)}),$$

where $(G_m^{(1)})_{m \geq 1}$ and $(G_m^{(2)})_{m \geq 1}$ are copies of the process in (3.3.9). Indeed, at the time of the connecting edge, a uniform half-edge of $\text{SWG}_m^{(2)}$ is drawn, and it is paired to a uniform alive half-edge of $\text{SWG}_{a_n}^{(1)}$. The number of hops in $\text{SWG}_{a_n}^{(1)}$ to the end of the attached edge is therefore equal in distribution to $G_{a_n+1}^{(1)}$. The -1 in (3.3.29) arises since the connecting edge is counted twice in $G_{a_n+1}^{(1)} + G_{C_n}^{(2)}$. The processes $(G_m^{(1)})_{m \geq 1}$ and $(G_m^{(2)})_{m \geq 1}$ are close to being independent, due to the coupling to the unimodular branching process of neighborhoods of uniform vertices.

Further, because of the way the weight of the potential connecting edges has been distributed over the two half-edges out of which the connecting edge is comprised,

$$(3.3.30) \quad C_n = T_{a_n}^{(1)} + T_{C_n}^{(2)},$$

where $(T_m^{(1)})_{m \geq 1}$ and $(T_m^{(2)})_{m \geq 1}$ are two copies of the process $(T_m)_{m \geq 1}$ in (3.3.8).

The connection time. We now intuitively explain why the leading order asymptotics of C_n is given by $a_n = \sqrt{n}$. For this, we must know how many allowed half-edges there are, i.e., we must determine how many half-edges there are incident to the union of the two SWGs at any time. Denote the number of allowed half-edges in the SWG from vertex i at time m by $S_m^{(i;n)}$. Here, a half-edge is called *allowed* when the vertex incident to it is in the smallest-weight graph, but the edge that it is part of is not. The total number of allowed half-edges incident to $\text{SWG}_{a_n}^{(1)}$ is $S_{a_n}^{(1;n)}$, while the number incident to $\text{SWG}_m^{(2)}$ is equal to $S_m^{(2;n)}$, and where

$$(3.3.31) \quad S_m^{(i)} = D_i + \sum_{l=2}^m (\Delta_l^{(i;n)} - 1).$$

Here, the forward degrees $(\Delta_l^{(1;n)})_{l \geq 2}$ and $(\Delta_l^{(2;n)})_{l \geq 2}$ are approximately i.i.d. random variables, independent of each other. The time to connection now has the following asymptotics:

PROPOSITION 3.10 (The time to connection). *For $i \in \{1, 2\}$, and with $\alpha = \nu/(\nu - 1)$,*

$$(3.3.32) \quad \left(\frac{G_{a_n+1}^{(1)} - \alpha \log a_n}{\sqrt{\alpha \log a_n}}, \frac{G_{C_n}^{(2)} - \alpha \log a_n}{\sqrt{\alpha \log a_n}} \right) \xrightarrow{d} (Z_1, Z_2),$$

where Z_1, Z_2 are two independent standard normal random variables. Moreover, there exist random variables X_1, X_2 two independent copies of the random variable X in (3.3.26) and an independent exponential random variable E such that,

$$(3.3.33) \quad \left(T_{a_n}^{(1)} - \frac{1}{\nu - 1} \log a_n, T_{C_n}^{(2)} - \frac{1}{\nu - 1} \log a_n \right) \xrightarrow{d} \left(X_1, X_2 + \frac{1}{\nu - 1} \log E + \frac{1}{\nu - 1} \log(\mu/(\nu - 1)) \right).$$

PROOF. Conditionally on $\text{SWG}_{a_n}^{(1;n)}$ and $(S_l^{(2;n)})_{l=1}^{m-1}$, and assuming that m and $S_m^{(i;n)}$ satisfy appropriate bounds,

$$(3.3.34) \quad \mathbb{P}(\mathbf{C}_n = m \mid \mathbf{C}_n > m - 1) \approx \frac{S_{a_n}^{(1;n)}}{\ell_n}.$$

We use this heuristic in order to identify the limit of \mathbf{C}_n/a_n :

LEMMA 3.11 (Weak convergence of connection time). *Under the conditions of Theorem 3.7, with $a_n = \sqrt{n}$,*

$$(3.3.35) \quad S_{a_n}^{(1;n)}/a_n \xrightarrow{\mathbb{P}} \nu - 1.$$

Consequently,

$$(3.3.36) \quad \mathbf{C}_n/a_n \xrightarrow{d} E\mu/(\nu - 1),$$

where E has an exponential distribution with mean 1.

We only sketch the proof of Lemma 3.11. The details are not very difficult, but somewhat tedious:

Sketch of proof of Lemma 3.11. We have that $S_{a_n}^{(1;n)} \approx \sum_{i=1}^{a_n} (\Delta_i - 1)$, where $(\Delta_i)_{i=1}^{a_n}$ is a collection of i.i.d. random variables with distribution $D^* - 1$. Since $\mathbb{E}[(D^*)^{1+\varepsilon}] < \infty$

$$(3.3.37) \quad \frac{1}{a_n} \sum_{i=1}^{a_n} (\Delta_i - 1) \xrightarrow{a.s.} \nu - 1.$$

Naturally, $S_{a_n}^{(1;n)}$ is not quite a sum of i.i.d. random variables, and, therefore, the above argument misses its rigorous details.

Each time m that we pair a half-edge in $\text{SWG}_m^{(2)}$, we have a probability close to $S_{a_n}^{(1;n)}/\ell_n \approx (\nu - 1)/(\mu\sqrt{n})$ of drawing one that is incident to $\text{SWG}_{a_n}^{(1)}$. Since this probability is close to constant and quite small, the time it takes until we first draw one is close to a Geometric random variable Q_n with parameter $(\nu - 1)/(\mu\sqrt{n})$. The conclusion follows since $Q_n(\nu - 1)/(\mu\sqrt{n}) \xrightarrow{d} E$, where E is an exponential random variable. \square

Now we are ready to complete the proof of Proposition 3.10:

Proof of Proposition 3.10. We first complete the proof of (3.3.33). It is not hard to prove from (3.3.30) that

$$(3.3.38) \quad \left(T_{a_n}^{(1)} - \frac{1}{\nu - 1} \log a_n, T_{\mathbf{C}_n}^{(2)} - \frac{1}{\nu - 1} \log \mathbf{C}_n \right) \xrightarrow{d} (X_1, X_2),$$

where (X_1, X_2) are two independent random variables with distribution given by (3.3.26). By Lemma 3.11,

$$(3.3.39) \quad \log \mathbf{C}_n - \log a_n = \log (\mathbf{C}_n/a_n) \xrightarrow{d} \log E + \frac{1}{\nu - 1} \log(\mu/(\nu - 1)).$$

Also, the two limits are independent, since the limit in (3.3.38) is independent of the limit of \mathbf{C}_n/a_n . This completes the proof for the weight of the shortest path in (3.3.33). The proof for (3.3.32) is similar. \square

Proof of Theorem 3.7. The statement for \mathcal{C}_n in (3.3.2), and the identification of the limiting random variable Q in (3.3.6), follows by (3.3.30), (3.3.33) in Proposition 3.10 and the fact that $-\log E = \log(1/E)$ has a Gumbel distribution. This leads to

$$(3.3.40) \quad \mathcal{C}_n - \frac{1}{\nu - 1} \xrightarrow{d} X_1 + X_2 + \frac{1}{\nu - 1} \log E + \frac{1}{\nu - 1} \log(\mu/(\nu - 1)),$$

and thus identifies Q in (3.3.2) as $X_1 + X_2 + \frac{1}{\nu - 1} \log E + \frac{1}{\nu - 1} \log(\mu/(\nu - 1))$.

By (3.3.29) and Proposition 3.10, with Z_1, Z_2 denoting independent standard normal random variables, and with $Z = (Z_1 + Z_2)/\sqrt{2}$, which is again standard normal,

$$(3.3.41) \quad \begin{aligned} H_n &\stackrel{d}{=} G_{a_{n+1}}^{(1)} + G_{\mathcal{C}_n}^{(2)} - 1 \\ &= 2\alpha \log a_n + Z_1 \sqrt{\alpha \log a_n} + Z_2 \sqrt{\alpha \log a_n} + O_{\mathbb{P}}(\sqrt{\log n}) \\ &= 2\alpha \log a_n + Z \sqrt{2\alpha \log a_n} + O_{\mathbb{P}}(\sqrt{\log n}) = \alpha \log n + Z \sqrt{\alpha \log n} + O_{\mathbb{P}}(\sqrt{\log n}). \end{aligned}$$

This completes the proof of (3.3.2) for the hopcount. □

Concluding remarks. The limit X in (3.3.17), which appears in the limiting random variable Q in (3.3.2) through (3.3.40), is a bit hard to interpret. Recall (3.3.26), where its explicit distributional form is given. We will see below that, as in Section 3.2, it can be interpreted as describing the speed with which vertices are found close to the respective vertices U_1, U_2 . In order to describe this in detail, however, we need to investigate the exploration of first-passage percolation in Galton-Watson trees with more general edge-weight distributions. Such processes are *continuous-time branching processes*, and are discussed in quite some detail in the following section.

3.4. The trick of the trade: Continuous-time branching processes

In the previous section, we have relied upon the memoryless property of the exponential distribution to describe the limiting behavior of the weight of the smallest-weight path, as well as to decouple it from the number of edges in this path. When dealing with first-passage percolation with *general* edge-weights, as we will do in the next section, the memoryless property no longer holds. It turns out that many of such settings can be described in terms of continuous-time branching processes.

In this section, we discuss such continuous-time branching processes, which are simple models for the evolution of a population in continuous time, contrary to discrete-time branching processes or Galton-Watson processes (recall Section 2.1), which describe the generations in populations instead. In a continuous-time branching process, each individual has offspring that is independent and identically distributed. The times at which the offspring of an individual are born form a point process or counting process in continuous time, and we are interested in the evolution of $|\text{BP}(t)|$, which is the number of individuals alive at a given time $t \geq 0$, as well as many related properties of the continuous-time branching process $\text{BP}(t)$ when viewed as a tree. Here one can think of the heights of vertices or the residual life-times.

We now formally define continuous-time branching processes. Let $(\mathcal{P}(t))_{t \geq 0}$ be a birth process. This means that $(\mathcal{P}(t))_{t \geq 0}$ is a counting process taking values in \mathbb{N} , and $\mathcal{P}(t)$ denotes the number of births in the time interval $[0, t]$. Recall that \mathcal{T}_∞ denotes the infinite

Ulam-Harris tree of words, so that $v \in \mathcal{T}_\infty$ can be represented as $v = v_1 \cdots v_k$, where $v_i \in \mathbb{N}$ and k is the generation of the word $v \in \mathcal{T}_\infty$. Each vertex $v \in \mathcal{T}_\infty$ has its own birth process $(\mathcal{P}^{(v)}(t))_{t \geq 0}$ indicating the birth-times of its children, where the birth processes for different v are independent and identically distributed birth processes. The fact that these birth processes are i.i.d. for different v makes the process a branching process, just like the number of children are i.i.d. in Galton-Watson processes.

At time 0, the birth process of the root \emptyset , which is denoted as $(\mathcal{P}^{(\emptyset)}(t))_{t \geq 0}$, starts running. Each time that an individual v is born, the birth process of v denoted by $(\mathcal{P}^{(v)}(t))_{t \geq 0}$ starts running, and individuals are born into this process. We denote the set of alive vertices by $\mathbf{BP}(t)$ and its cardinality or size by $|\mathbf{BP}(t)|$. Thus, $\mathbf{BP}(t)$ consists of those $v \in \mathcal{T}_\infty$ that are alive at time t . Sometimes, births of the children of a vertex v occur when the parent $p(v)$ of v dies, sometimes the individuals remain on living forever. Thus, a continuous-time branching process $(\mathbf{BP}(t))_{t \geq 0}$ is characterized by its birth process together with the death rules of the individuals.

We next discuss two examples that relate to the first-passage percolation processes studied in the previous two sections:

3.4.1. Yule process. Let the birth process $(\mathcal{P}(t))_{t \geq 0}$ be a Poisson point process $(\mathbf{PP}(t))_{t \geq 0}$, so that $\mathbf{PP}[a, b] = \#\{i: Y_i \in [a, b]\}$, where $Y_i = E_1 + \dots + E_i$ and $(E_i)_{i \geq 1}$ are i.i.d. exponential random variables with parameter 1. Individuals live forever. The process $(\mathbf{BP}(t))_{t \geq 0}$ where $\mathbf{BP}(t)$ is the set of alive individuals at time t is called a Yule process.

We first investigate the link to first-passage percolation on the complete graph in Section 3.2. Let $T_m = \inf\{t: |\mathbf{BP}(t)| = m\}$ denote the birth time of the m th individual in the Yule process, where by convention $T_1 = 0$ so that the ancestor of the continuous-time branching process is born instantaneously. Then, by the memoryless property of the exponential distribution,

$$(3.4.1) \quad T_{m+1} \stackrel{d}{=} \sum_{k=1}^m E_k/k.$$

Thus, T_m has the same distribution as the limit as $n \rightarrow \infty$ of the birth time of the m th individual for first-passage percolation on the complete graph with exponential edge weights. Further, for the Yule process and again by the memoryless property of the exponential distribution, the parent of the m th individual is a uniform vertex in $[m-1]$. Thus, the tree of the first m individuals is a uniform recursive tree of size m , so that the distribution of the height of the m th vertex in Theorem 3.3 satisfies a CLT with asymptotic mean and variance equal to $\log m$.

We next investigate the number of alive particles in the Yule process at time t . Let $p_m(t) = \mathbb{P}(|\mathbf{BP}(t)| = m)$. Then, since the birth process of the root is a Poisson point process,

$$(3.4.2) \quad p_1(t) = \mathbb{P}(\mathbf{PP}[0, t] = 0) = e^{-t}.$$

Further, for $|\mathbf{BP}(t)| = m$ to occur, the m th birth should occur at some time $u \in [0, t]$, while the $(m+1)$ st birth should occur after time t . Conditionally on $|\mathbf{BP}(u)| = m-1$, the rate of birth of the m th individual is equal to $m-1$. This implies the following relation

between $p_m(t)$ and $p_{m-1}(t)$:

$$(3.4.3) \quad p_m(t) = \int_0^t p_{m-1}(u)(m-1)e^{-m(t-u)} du.$$

The solution to (3.4.3) is

$$(3.4.4) \quad p_m(t) = e^{-t}(1 - e^{-t})^{m-1}, \quad m \geq 1, t \geq 0,$$

which we now prove by induction. The claim for $m = 1$ is already proved in (3.4.2) above. To advance the induction hypothesis, we can rewrite

$$(3.4.5) \quad p_m(t) = (m-1)e^{-mt} \int_0^t e^{mu} p_{m-1}(u) du.$$

We use induction, which yields that

$$(3.4.6) \quad (m-1)e^{mu} p_{m-1}(u) = (m-1)e^u (e^u - 1)^{m-2} = \frac{d}{du} (e^u - 1)^{m-1},$$

so that

$$(3.4.7) \quad p_m(t) = e^{-mt} \int_0^t (m-1)e^{mu} p_{m-1}(u) du = e^{-mt} (e^t - 1)^{m-1} = e^{-t} (1 - e^{-t})^{m-1},$$

as required.

By (3.4.4), the distribution of $|\text{BP}(t)|$ is Geometric with parameter e^{-t} . A geometric random variable with small success probability p is close to an exponential random variable:

EXERCISE 3.16 (Geometric with small success probability). *Let X_p have a geometric distribution with parameter p . Show that, as $p \searrow 0$,*

$$(3.4.8) \quad pX_p \xrightarrow{d} E,$$

where E has an exponential distribution.

By Exercise 3.16, as $t \rightarrow \infty$,

$$(3.4.9) \quad e^{-t}|\text{BP}(t)| \xrightarrow{d} E.$$

In fact, we can improve this result to convergence almost surely using a martingale argument:

EXERCISE 3.17 (Yule process martingale). *Show that the process $M(t) = e^{-t}|\text{BP}(t)|$ is a continuous-time martingale, and conclude that the convergence in (3.4.9) occurs a.s.*

3.4.2. Bellman-Harris processes. In a Bellman-Harris process, the initial individual dies instantaneously, giving birth to a random number of children Δ . Each of these individuals has a life-time that is an exponential random variable with mean 1, and upon dying, produces a random number of children independently of and having the same distribution as Δ . Each of these children again has an exponential life-time with mean 1, etc. This is a continuous-time branching process with birth process

$$(3.4.10) \quad \mathcal{P}[a, b] = \Delta \mathbb{1}_{\{E \in [a, b]\}}.$$

Life-times and number of offspring across individuals are independent. When the offspring distribution is i.i.d., we retrieve first-passage percolation on a Galton-Watson tree with offspring have the same distribution as Δ and with exponential passage times.

To analyze a Bellman-Harris continuous-time branching process, we let $S_0 = 1$, and define $(S_i)_{i \geq 0}$ recursively by

$$(3.4.11) \quad S_i = S_{i-1} + \Delta_i - 1,$$

as long as $S_{i-1} \geq 1$. Here $(\Delta_i)_{i \geq 1}$ is an i.i.d. sequence of random variables whose law is called the *offspring distribution*. Let T_m denote the time of death of the $(m + 1)$ st individual, so that $T_1 = 0$. Then, when $S_i \geq 1$ for all $i \in [m - 1]$,

$$(3.4.12) \quad T_m = \sum_{j=1}^m E_j / S_j,$$

where $(E_j)_{j \geq 1}$ is a sequence of i.i.d. exponential random variables with mean 1. When $S_i = 0$ for some $i \in [m - 1]$, we define $T_m = \infty$. With $|\text{BP}(t)|$ equal to the number of alive vertices at time t ,

$$(3.4.13) \quad |\text{BP}(T_m)| = S_{T_m},$$

so that $|\text{BP}(t)| = S_{T(t)}$, where $T(t) = \max\{m : T_m \leq t\}$.

Let us first make the relation to the first-passage percolation on $\text{CM}_n(\mathbf{d})$ discussed in Section 3.3. There, we had that $S_1 = D$, while, for $i > 1$, the recursion relation in (3.4.11) is satisfied. We arrive at a process that is a Bellman-Harris continuous-time branching process, apart from the fact that the offspring distribution at the root is D instead of $\Delta = D^* - 1$.

Naturally, continuous-time branching processes allow us to study settings where the number of children of the root is different from that of any other individual. We call such a continuous-time branching process a *unimodular* continuous-time branching process. Indeed, let the number of children of the root be equal to D and the offspring of the i th individual to be equal to Δ_i where $(\Delta_i)_{i \geq 2}$ is an i.i.d. sequence whose law is given by $\Delta = D^* - 1$. Then, we have the relation that

$$(3.4.14) \quad |\text{BP}(t)| = \sum_{i=1}^D |\text{BP}_i(t - E_i)|,$$

where $(E_i)_{i \geq 1}$ are i.i.d. exponentials, and $(|\text{BP}_i(t)|)_{t \geq 0}$ are i.i.d. Bellman-Harris processes with i.i.d. offsprings. By convention, $|\text{BP}_i(t)| = 0$ when $t < 0$.

In Section 3.3, it was proved that, with T_m the death time of the m th individual and G_m its height,

$$(3.4.15) \quad T_m - \frac{1}{\lambda} \log m \xrightarrow{d} X, \quad \frac{G_m - \frac{\lambda}{\lambda-1} \log m}{\sqrt{\frac{\lambda}{\lambda-1} \log m}} \xrightarrow{d} Z,$$

where $\lambda = \mathbb{E}[\Delta]$ is the expected offspring, Z is standard normal and

$$(3.4.16) \quad X = \sum_{j=1}^{\infty} (E_j - 1) / S_j + \sum_{j=1}^{\infty} [1/S_j - 1/(\lambda j)] + \gamma / \lambda,$$

where γ is the Euler-Mascheroni constant and $\lambda = \nu - 1 = \mathbb{E}[\Delta] - 1$. Therefore, $me^{-\lambda T_m} \xrightarrow{d} e^{-\lambda X}$, which immediately implies that

$$(3.4.17) \quad e^{-\lambda t} |\text{BP}(t)| \xrightarrow{d} W = e^{-\lambda X}.$$

By the Markov property of this process, the random variable $M(t) = e^{-\lambda t}|\mathbf{BP}(t)|$ is a continuous-time martingale, and therefore the convergence in (3.4.17) even holds almost surely. We conclude that the limiting random variable X in (3.3.17) has the interpretation as $(1/\lambda) \log(1/W)$, where W is the martingale limit of $e^{-\lambda t}|\mathbf{BP}(t)|$. Note that the Gumbel variable Λ_1 in, e.g., (3.2.37), plays a similar role.

From a first-passage percolation point of view, it may look more natural to let $|\mathbf{BP}(t)|$ denote the number of vertices that can be found by the flow before time t , instead of the number of neighbors of vertices found by the flow. However, with the current setup, the process $|\mathbf{BP}(t)|$ is a continuous-time Markov process. In the alternative formulation, this is *not* the case, since even when we know the number of vertices found, the next vertex is found at rate proportional to the number of neighbors, which is random.

3.4.3. General continuous-time branching processes. In Sections 3.4.1 and 3.4.2, we let $|\mathbf{BP}(t)|$ denote the number of alive individuals, and we see that there exists a λ and a limiting random variable W such that

$$(3.4.18) \quad e^{-\lambda t}|\mathbf{BP}(t)| \xrightarrow{d} W.$$

Further, there are α, β such that

$$(3.4.19) \quad \frac{G_m - \alpha \log m}{\sqrt{\beta \log m}} \xrightarrow{d} Z,$$

where Z is standard normal. These results have been obtained in a direct way in Sections 3.2-3.3. We next turn to an example where such a direct computation is much more difficult:

EXAMPLE 3.12 (Age-dependent branching processes). Let $(\mathbf{BP}(t))_{t \geq 0}$ denote the continuous-time branching process where the birth process \mathcal{P} is given by

$$(3.4.20) \quad \mathcal{P}[a, b] = \Delta \mathbb{1}_{\{Y \in [a, b]\}}.$$

Here the birth-time Y is a non-negative continuous random variable having distribution function F_Y .

At time $t = 0$, we start with one individual which we refer to as the original ancestor or the root of the branching process. This individual immediately dies giving rise to Δ_1 alive children. Each new individual v in the branching process lives for a random amount of time which has distribution F_Y , and then dies. At the time of death again the individual gives birth to Δ_v children, where $(\Delta_v)_v$ is a sequence of i.i.d. random variables with the same distribution as Δ_1 . Life-times and number of offspring across individuals are independent.

We let $|\mathbf{BP}(t)|$ denote the number of alive individuals at time t . When the offspring distribution is i.i.d., we retrieve first-passage percolation on a Galton-Watson tree with offspring distribution Δ and passage-time distribution F_Y . When the life-time distribution is exponential, we retrieve the Bellman-Harris process. Here, we make the following observation:

REMARK 3.13 (Connection age-dependent branching process and first-passage percolation). The *alive* individuals at time t in the age-dependent branching process correspond to the individuals that are *waiting to be born* in first-passage percolation, i.e., those individuals whose parents have been found by first-passage percolation at time t , but they

themselves have not yet been. The *dead* individuals correspond to those that have been found by first-passage percolation at time t .

When the life-time distribution Y is not exponential, the process $(|\mathbf{BP}(t)|)_{t \geq 0}$ is *not* Markovian. As a result, this process is much more difficult to study. Still, one would guess that $|\mathbf{BP}(t)|$ again grows exponentially in t . We will study such questions below.

In order to study general continuous-time branching processes, we define the following *Malthusian parameter*:

DEFINITION 3.14 (Malthusian parameter). *Let $(\mathbf{BP}(t))_{t \geq 0}$ be a continuous-time branching process with birth process $(\mathcal{P}(t))_{t \geq 0}$. The Malthusian parameter λ is the solution to the equation*

$$(3.4.21) \quad \mathbb{E} \left[\int_0^\infty e^{-\lambda t} \mathcal{P}(dt) \right] = 1.$$

The Malthusian parameter is the solution to (3.4.21) when this exists. The necessary and sufficient condition for this to hold is that $\mathbb{E}[|\mathcal{P}[0, \infty)|] > 1$ and $\lim_{\lambda \rightarrow \infty} \mathbb{E} \left[\int_0^\infty e^{-\lambda t} \mathcal{P}(dt) \right] < 1$. The following theorem, proved by Jagers and Nerman [178, Theorem 5.3], shows that λ indeed describes the exponential growth of the continuous-time branching process:

THEOREM 3.15 (Exponential growth continuous-time branching process). *Assume the $X \log X$ condition in the form*

$$(3.4.22) \quad \mathbb{E}[\hat{X}(\lambda) \log(\hat{X}(\lambda))_+] < \infty,$$

where the random variable $\hat{X}(\lambda)$ is defined by

$$(3.4.23) \quad \hat{X}(\lambda) = \int_0^\infty e^{-\lambda t} \mathcal{P}(dt).$$

Assume that $\mathbb{E}[\hat{X}(0)] > 1$ and $\lim_{\lambda \rightarrow \infty} \mathbb{E}[\hat{X}(\lambda)] < 1$. Then, the Malthusian parameter $\lambda > 0$ given in Definition 3.14 exists and is unique, and there exists a random variable W such that

$$(3.4.24) \quad e^{-\lambda t} |\mathbf{BP}(t)| \xrightarrow{d} W.$$

Furthermore, W is positive if and only if the survival event $\{|\mathbf{BP}(t)| > 0 \forall t \geq 0\}$ occurs.

Let us now describe some examples of the above theorem:

EXAMPLE 3.16 (Examples of exponential growth continuous-time branching process). We now give three examples:

(a) For the Yule process, we can compute that

$$(3.4.25) \quad \mathbb{E} \left[\int_0^\infty e^{-\lambda t} \mathcal{P}(dt) \right] = \sum_{i \geq 1} \mathbb{E}[e^{-\lambda Y_i}],$$

since a Poisson point process \mathbf{PP} can be written as $\mathbf{PP} = \{Y_i\}_{i \geq 1}$ with $Y_i = E_1 + \dots + E_i$ and $(E_i)_{i \geq 1}$ are i.i.d. exponential random variables. Since Y_i are i.i.d., we can further

compute

$$(3.4.26) \quad \begin{aligned} \sum_{i \geq 1} \mathbb{E}[e^{-\lambda Y_i}] &= \sum_{i \geq 1} \mathbb{E}[e^{-\lambda E_1}]^i \\ &= \sum_{i \geq 1} \left(\frac{1}{\lambda + 1}\right)^i = 1/\lambda, \end{aligned}$$

so that $\lambda = 1$. Thus, (3.4.24) reduces to (3.4.9), apart from the identification of the limiting random variable. For this, it is useful to use the branching property to deduce that W satisfies the following distributional relation:

$$(3.4.27) \quad W \stackrel{d}{=} \sum_{i=1}^{\infty} e^{-Y_i} W_i,$$

where again $Y_i = E_1 + \dots + E_i$ and $(E_i)_{i \geq 1}$ are i.i.d. exponential random variables. Note that e^{-E_i} has a uniform distribution on $(0, 1)$. Since $(e^{-Y_i} W_i)_{i \geq 1}$ has the same distribution as $e^{-E_1} (e^{-Y_{i-1}} W_i)_{i \geq 1}$, we thus obtain that

$$(3.4.28) \quad W \stackrel{d}{=} U(W_1 + W_2),$$

where W_1 and W_2 are i.i.d. copies of W . It can be shown that the unique solution of this stochastic recurrence relation is the exponential distribution.

(b) For Bellman-Harris processes with $\mathbb{E}[\Delta] \in (1, \infty)$, we compute that

$$(3.4.29) \quad \mathbb{E} \left[\int_0^{\infty} e^{-\lambda t} \mathcal{P}(dt) \right] = \mathbb{E}[\Delta] \mathbb{E}[e^{-\lambda E}] = \frac{\mathbb{E}[\Delta]}{\lambda + 1},$$

so that $\lambda = \mathbb{E}[\Delta] - 1 > 0$.

(c) For general age-dependent branching processes with $\mathbb{E}[\Delta] \in (1, \infty)$, instead

$$(3.4.30) \quad \mathbb{E} \left[\int_0^{\infty} e^{-\lambda t} \mathcal{P}(dt) \right] = \mathbb{E}[\Delta] \mathbb{E}[e^{-\lambda Y}],$$

where Y has the edge weight distribution. The equation $\mathbb{E}[\Delta] \mathbb{E}[e^{-\lambda Y}] = 1$ always has a solution, since $\mathbb{E}[\Delta] \in (1, \infty)$ and $\mathbb{E}[e^{-\lambda Y}] \searrow 0$ as $\lambda \nearrow \infty$ since $Y > 0$ due to the dominated convergence theorem. Thus, Theorem 3.15 identifies the exponential growth of such age-dependent branching processes.

We next investigate the $X \log X$ -condition in the case of Yule processes and Bellman-Harris processes:

EXERCISE 3.18 ($X \log X$ -condition Yule process). *Show that the $X \log X$ condition holds for a Yule process. Hint: Show that $\hat{X}(1) = \sum_{i=1}^{\infty} e^{-Y_i}$, where $Y_i = E_1 + \dots + E_i$ are the points of a Poisson process and $(E_i)_{i \geq 1}$ are i.i.d. exponential random variables with parameter 1. Then prove that $\mathbb{E}[\hat{X}(1)^2] < \infty$, which is stronger than the $X \log X$ condition.*

EXERCISE 3.19 ($X \log X$ -condition Bellman-Harris branching process). *Show that the $X \log X$ condition holds for a Bellman-Harris process precisely when $\mathbb{E}[\Delta \log(\Delta)_+] < \infty$.*

EXERCISE 3.20 ($X \log X$ -condition age-dependent branching process). *Show that the $X \log X$ condition holds for age-dependent branching process precisely when $\mathbb{E}[\Delta \log(\Delta)_+] < \infty$.*

SKETCH OF PROOF OF THEOREM 3.15. We will not give a full proof of Theorem 3.15. Instead, in the case of age-dependent branching process, we prove that there exists an $A \in (0, \infty)$ such that

$$(3.4.31) \quad e^{-\lambda t} \mathbb{E}[|\text{BP}(t)|] \rightarrow A.$$

While this is significantly weaker than Theorem 3.15, it is a highly suggestive partial result.

We assume that the expected offspring $\mathbb{E}[\Delta] = \nu$ satisfies $\nu > 1$ and that $\mathbb{E}[\Delta^2] < \infty$. Further, the life-time distribution function F_Y has to be non-lattice, plus some small additional condition for the almost sure convergence (see Harris [153] for the proof).

Age-dependent branching processes: intuition for the number of alive individuals. Using Remark 3.13 as well as Wald's identity, we deduce that

$$(3.4.32) \quad \mathbb{E}[|\text{BP}(t)|] = \sum_{k \geq 0} \mathbb{E}[Z_k] \mathbb{P}(Y_1 + \dots + Y_k \leq t, Y_1 + \dots + Y_{k+1} > t),$$

since $\mathbb{E}[Z_k]$ measures the number of individuals in generation k , and any such individual is alive when its parent is dead while it itself is not yet dead. This means that the sum of edge-weights up to its parent is at most t , while that including its own life-time is larger than t . When Y has a continuous distribution, we can condition on $Y_1 + \dots + Y_k$, denoting the density of $Y_1 + \dots + Y_k$ by $f_{Y_1 + \dots + Y_k}(u)$, to obtain

$$(3.4.33) \quad \mathbb{E}[|\text{BP}(t)|] = \sum_{k \geq 0} \mathbb{E}[Z_k] \int_0^t f_{Y_1 + \dots + Y_k}(u) [1 - F_Y](t - u) du.$$

We use that $\mathbb{E}[Z_k] = \nu^k$, and multiply through by $e^{-\lambda t}$ to obtain

$$(3.4.34) \quad e^{-\lambda t} \mathbb{E}[|\text{BP}(t)|] = \sum_{k \geq 0} \nu^k e^{-\lambda t} \int_0^t f_{Y_1 + \dots + Y_k}(u) [1 - F_Y](t - u) du.$$

Denote the density $f_{\bar{Y}}$ by

$$(3.4.35) \quad f_{\bar{Y}}(s) = \nu e^{-\lambda s} f_Y(s),$$

and note that

$$(3.4.36) \quad \nu^k e^{-\lambda u} f_{Y_1 + \dots + Y_k}(u) = f_{\bar{Y}_1 + \dots + \bar{Y}_k}(u),$$

to arrive at

$$(3.4.37) \quad e^{-\lambda t} \mathbb{E}[|\text{BP}(t)|] = \sum_{k \geq 0} \int_0^t f_{\bar{Y}_1 + \dots + \bar{Y}_k}(u) e^{-\lambda(t-u)} [1 - F_Y](t - u) du.$$

Equation (3.4.37) is highly suggestive. Indeed, due to the factor $e^{-\lambda(t-u)}$, the integral is dominated by values of u for which $u \leq t$ but u is close to t . By the central limit theorem for $\bar{Y}_1 + \dots + \bar{Y}_k$, we see that k must then be such that $\bar{\nu}k \approx t$, so that $k \approx t/\bar{\nu}$. Here $\bar{\nu} = \mathbb{E}[\bar{Y}]$ is the mean of the stable-age distribution. Both these observations are correct, and we will next explain and prove these results in more detail. For this, we rely on *renewal theory* to make the above analysis precise.

Age-dependent branching processes and renewal theory. Age-dependent branching processes are intimately connected to *renewal theory*. This is immediate when we demand that the offspring distribution is degenerated at 1, in which case we deal with a renewal process with inter-arrival distribution F_Y . However, renewal theory also plays a role when F_Y is non-degenerate and $\nu > 1$. To see this, define the probability generating function $F(s, t)$ of the number of alive individuals $|\text{BP}(t)|$ at time t , by

$$(3.4.38) \quad F(s, t) = \sum_{k=0}^{\infty} s^k \mathbb{P}(|\text{BP}(t)| = k).$$

The function $F(s, t)$ satisfies the equation

$$(3.4.39) \quad F(s, t) = s(1 - F_Y(t)) + \int_0^t G_{\Delta}(F(s, t - u)) F_Y(du),$$

where $G_{\Delta}(s) = \mathbb{E}[s^{\Delta}]$ is the generating function of the offspring Δ . Equation (3.4.39) follows by conditioning on the life-time and number of offspring of the first individual. In particular $\mathbb{E}[|\text{BP}(t)|]$ satisfies the renewal-like equation

$$(3.4.40) \quad \mathbb{E}[|\text{BP}(t)|] = 1 - F_Y(t) + \nu \int_0^t \mathbb{E}[|\text{BP}(t - u)|] F_Y(du),$$

where $\nu = \mathbb{E}[\Delta]$. To make (3.4.40) into a *genuine* renewal equation, we define the *stable-age distribution* as follows:

DEFINITION 3.17 (Stable-age distribution). *Let the stable-age distribution be the distribution function \bar{F}_Y given by*

$$(3.4.41) \quad \bar{F}_Y(y) = \mathbb{E} \left[\int_0^y e^{-\lambda t} \mathcal{P}(dt) \right].$$

Let \bar{Y} be a random variable with distribution function \bar{F}_Y .

Multiplying both sides of (3.4.40) by $e^{-\lambda t}$, where λ is the Malthusian parameter, we obtain

$$(3.4.42) \quad e^{-\lambda t} \mathbb{E}[|\text{BP}(t)|] = e^{-\lambda t} [1 - F_Y(t)] + e^{-\lambda u} \nu \int_0^t e^{-\lambda(t-u)} \mathbb{E}[|\text{BP}(t - u)|] F_Y(du),$$

or, equivalently, using that by (3.4.41) $\bar{F}_Y(du) = \nu e^{-\lambda u} F_Y(du)$,

$$(3.4.43) \quad K(t) = f(t) + \int_0^t K(t - u) \bar{F}_Y(du),$$

where

$$(3.4.44) \quad K(t) = e^{-\lambda t} \mathbb{E}[|\text{BP}(t)|], \quad f(t) = e^{-\lambda t} [1 - F_Y(t)].$$

Remco: Add ref for renewal theorem!

The Key-Renewal Theorem applies to such convolution equations. Indeed, it states that when there exist a function f and probability measure m on the non-negative reals such that

$$(3.4.45) \quad K(t) = f(t) + \int_0^t K(t - u) m(du),$$

where m is non-lattice (i.e., there does not exist a d such that $\sum_n m(nd) = 1$) and f is directly integrable, then

$$(3.4.46) \quad \lim_{t \rightarrow \infty} K(t) = \frac{1}{\mu} \int_0^\infty f(u) du,$$

where $\mu = \int_0^\infty um(du)$ is the mean of the probability measure m . The following exercises give an idea of why (3.4.46) could be true:

EXERCISE 3.21 (Key-renewal theorem (1)). For $\alpha > 0$, let $\hat{K}(\alpha) = \int_0^\infty e^{-\alpha t} K(t) dt$, where $(K(t))_{t \geq 0}$ is the solution to (3.4.45). Prove that

$$(3.4.47) \quad \hat{K}(\alpha) = \frac{\int_0^\infty e^{-\alpha t} f(t) dt}{1 - \int_0^\infty e^{-\alpha t} m(dt)}.$$

Conclude that $\alpha \hat{K}(\alpha) \rightarrow \int_0^\infty f(u) du / \int_0^\infty um(du)$ when $\alpha \searrow 0$.

EXERCISE 3.22 (Key-renewal theorem (2)). Use the previous exercise to show that if $K(t) \rightarrow A$ when $t \rightarrow \infty$, then A must satisfy $A = \int_0^\infty f(u) du / \int_0^\infty um(du)$.

Applying the Key-Renewal Theorem to (3.4.43) yields

$$(3.4.48) \quad \lim_{t \rightarrow \infty} e^{-\lambda t} \mathbb{E}[|\text{BP}(t)|] = \lim_{t \rightarrow \infty} K(t) = \int_0^\infty f(y) dy / \bar{\nu},$$

where $\bar{\nu} = \int_0^\infty [1 - \bar{F}_Y(t)] dt$ equals the mean of $\bar{F}_Y(t)$ and $K(t)$ and $f(t)$ are defined in (3.4.44).

An easy computation using (3.4.44) verifies that (3.4.48) implies that

$$(3.4.49) \quad \lim_{t \rightarrow \infty} e^{-\lambda t} \mathbb{E}[|\text{BP}(t)|] = A,$$

or $\mathbb{E}[|\text{BP}(t)|] = Ae^{\lambda t}(1 + o(1))$, where

$$(3.4.50) \quad A = \frac{\nu - 1}{\lambda \nu \bar{\nu}} = \frac{\nu - 1}{\lambda \nu^2 \int_0^\infty ye^{-\lambda y} F_Y(dy)}.$$

This proves (3.4.31) for age-dependent branching processes.

The proof of (3.4.24) in the more general case is more involved, and will be omitted here. \square

EXERCISE 3.23 (Second moment for continuous-time branching process*). Assume that $\mathbb{E}[\Delta^2] < \infty$ where Δ is the random amount of offspring of our continuous-time branching process. Show that

$$(3.4.51) \quad \lim_{t \rightarrow \infty} e^{-2\lambda t} \mathbb{E}[|\text{BP}(t)|^2] = \frac{A^2 a}{1 - \eta},$$

where

$$(3.4.52) \quad A = \lim_{t \rightarrow \infty} e^{-\lambda t} \mathbb{E}[|\text{BP}(t)|], \quad \eta = \int_0^\infty e^{-\alpha y} d\bar{F}_Y(y) < 1, \quad a = \eta \mathbb{E}[\Delta(\Delta - 1)] / \mathbb{E}[\Delta].$$

Hint: Adapt (3.4.39) to get a recursion formula for $\mathbb{E}[|\text{BP}(t)|^2]$ by conditioning on the first generation.

In the following, we derive some results on the number of individuals in a continuous-time branching process satisfying certain properties. We investigate the residual life-time distribution and the heights of alive individuals in a continuous-time branching process. In what follows, we restrict to age-dependent branching processes.

The residual life-time distribution. We let $|\text{BP}[t, t + s]|$ denote the number of individuals in the continuous-time branching process at time t and with residual life-time at most s . These are precisely the alive individuals that will die before time $t + s$. In order to state the result, we define the *residual life-time distribution* F_R to have density f_R given by

$$(3.4.53) \quad f_R(x) = \frac{\int_0^\infty e^{-\lambda y} f_Y(x + y) dy}{\int_0^\infty e^{-\lambda y} [1 - F_Y(y)] dy},$$

where f_Y is the density of the life time distribution F_Y .

EXERCISE 3.24 (Residual life-time distribution). *Show that f_R in (3.4.53) is a density on $[0, \infty)$.*

The main result of this section is the following theorem:

THEOREM 3.18 (The residual life-time distribution). *Let $(\text{BP}(t))_{t \geq 0}$ be an age-dependent branching process with offspring Δ . Assume that the $X \log X$ condition holds, i.e., $\mathbb{E}[\Delta \log(\Delta)_+] < \infty$. Then, with $A = (\nu - 1)/\lambda\nu\bar{\nu}$ and for all $s \geq 0$,*

$$(3.4.54) \quad \lim_{t \rightarrow \infty} e^{-\lambda t} \mathbb{E}[|\text{BP}[t, t + s]|] = AF_R(s).$$

PROOF. We adapt the proof of (3.4.31) in Theorem 3.15. We note that (3.4.40) is now replaced by

$$(3.4.55) \quad \mathbb{E}[|\text{BP}[t, t + s]|] = F_Y(t + s) - F_Y(t) + \nu \int_0^t \mathbb{E}[|\text{BP}[t - u, t + s - u]|] dF_Y(u).$$

Multiplying both sides of (3.4.55) by $e^{-\lambda t}$, where λ is the Malthusian parameter, we obtain

$$(3.4.56) \quad e^{-\lambda t} \mathbb{E}[|\text{BP}[t, t + s]|] = e^{-\lambda t} [F_Y(t + s) - F_Y(t)] + \nu \int_0^t e^{-\lambda u} e^{-\lambda(t-u)} \mathbb{E}[|\text{BP}[t - u, t + s - u]|] F_Y(du),$$

or, equivalently,

$$(3.4.57) \quad K(t) = f(t) + \int_0^t K(t - u) \bar{F}_Y(du),$$

where again $\bar{F}_Y(du) = \nu e^{-\lambda u} F_Y(du)$ and now

$$(3.4.58) \quad K(t) = e^{-\lambda t} \mathbb{E}[|\text{BP}[t, t + s]|], \quad f(t) = f_s(t) = e^{-\lambda t} [F_Y(t + s) - F_Y(t)].$$

By the Key-renewal theorem,

$$(3.4.59) \quad \lim_{t \rightarrow \infty} e^{-\lambda t} \mathbb{E}[|\text{BP}[t, t + s]|] = \frac{1}{\mu} \int_0^\infty f(u) du = \frac{\int_0^\infty e^{-\lambda y} [F_Y(y + s) - F_Y(y)] dy}{\int_0^\infty u \bar{F}_Y(du)}.$$

Inserting the definition $A = (\nu - 1)/(\lambda\nu\bar{\nu})$ and rewriting the above yields the result:

EXERCISE 3.25 (Completion proof Theorem 3.18.). *Use (3.4.59) and Theorem 3.18 and (3.4.53) to complete the proof of Theorem 3.18.*

□

The height of a random vertex. The growth of the continuous-time branching process is related to the size of the tree at a given time, which we will later relate to the weight-distance in random graphs. Instead, the hopcount is related to the heights of vertices in the continuous-time branching process, which we investigate now. For this, we let $|\text{BP}_j(t)|$ denote the number of individuals alive at time t in generation j , and write

$$(3.4.60) \quad |\text{BP}_{\leq k}(t)| = \sum_{j=0}^k |\text{BP}_j(t)|.$$

We know that $|\text{BP}(t)|$ grows like $Ae^{\lambda t}$. Thus, we may think of $e^{-\lambda t}|\text{BP}_{\leq k}(t)|$ as the *height profile* of the continuous-time branching process. We will show that this height profile is close to a normal distribution:

THEOREM 3.19 (CLT for vertex heights in continuous-time branching process). *Let $(\text{BP}(t))_{t \geq 0}$ be an age-dependent branching process with offspring Δ . Assume that $\mathbb{E}[\Delta \log(\Delta)_+] < \infty$, so that the $X \log X$ condition holds. Then, with $A = (\nu - 1)/\lambda\nu\bar{\nu}$ and for all $x \in \mathbb{R}$,*

$$(3.4.61) \quad \lim_{t \rightarrow \infty} e^{-\lambda t} \mathbb{E}[|\text{BP}_{\leq k_t(x)}(t)|] = A\Phi(x),$$

where

$$(3.4.62) \quad k_t(x) = \frac{t}{\bar{\nu}} + x\sqrt{t\frac{\bar{\sigma}^2}{\bar{\nu}^3}},$$

and $\bar{\nu}$ and $\bar{\sigma}^2$ are the mean, respectively variance, of the stable-age distribution \bar{F}_Y .

Theorem 3.19 suggests that a *random* alive individual at time t has a height that is close to normal with mean $t/\bar{\nu}$ and variance $t\bar{\sigma}^2/\bar{\nu}^3$.

PROOF. Conditioning on the life-time (with distribution function F_Y) of the first individual, after which the individual dies and splits in a random number offspring with mean ν ,

$$(3.4.63) \quad \mathbb{E}[|\text{BP}_j[t, t+s)|] = \nu \int_0^t \mathbb{E}[|\text{BP}_{j-1}(t-y)|] dF_Y(y).$$

As before, we multiply by $e^{-\lambda t}$ and define

$$(3.4.64) \quad |\bar{\text{BP}}_j(t)| = e^{-\lambda t} |\text{BP}_j(t)|.$$

Rewriting (3.4.63), we obtain the recursion

$$(3.4.65) \quad \mathbb{E}[|\bar{\text{BP}}_j(t)|] = \int_0^t \mathbb{E}[|\bar{\text{BP}}_{j-1}(t-y)|] d\bar{F}_Y(y),$$

where, as before $\bar{F}_Y(du) = \nu e^{-\lambda u} F_Y(du)$ is the stable-age distribution. Hence, if we continue to iterate, then we get

$$(3.4.66) \quad \mathbb{E}[|\bar{\text{BP}}_j(t)|] = \int_0^t \mathbb{E}[|\bar{\text{BP}}(t-y)|] d\bar{F}_Y^{*j}(y),$$

where \bar{F}_Y^{*j} is the j -fold convolution of \bar{F}_Y , and hence the distribution function of the independent sum of j copies of a random variable each having c.d.f. \bar{F}_Y . This is the point

where we use the classical central limit theorem for i.i.d. random variables. For fixed $t > 0$ and $m \geq 0$, we define

$$(3.4.67) \quad |\overline{\text{BP}}_{>m}(t)| = \sum_{j=m+1}^{\infty} |\overline{\text{BP}}_j(t)|.$$

By Theorem 3.15,

$$(3.4.68) \quad \lim_{t \rightarrow \infty} \mathbb{E}[|\overline{\text{BP}}(t)|] = \lim_{t \rightarrow \infty} \sum_{j=0}^{\infty} \mathbb{E}[|\overline{\text{BP}}_j(t)|] = A.$$

Hence, (3.4.61) follows if we show that

$$(3.4.69) \quad \mathbb{E}[|\overline{\text{BP}}_{>k_t(x)}(t)|] \rightarrow A - A\Phi(x) = A\Phi(-x).$$

Note that

$$(3.4.70) \quad \mathbb{E}[|\overline{\text{BP}}_{>k_t(x)}(t)|] = \int_0^t \mathbb{E}[|\overline{\text{BP}}_0(t-u)|] d\bar{F}_Y^{*k_t(x)}(u).$$

Take an arbitrary $\varepsilon > 0$ and take t_0 so large so that for $t > t_0$,

$$(3.4.71) \quad |\mathbb{E}[|\overline{\text{BP}}(t)|] - A| \leq \varepsilon.$$

Then,

$$(3.4.72) \quad \begin{aligned} & \left| \mathbb{E}[|\overline{\text{BP}}_{>k_t(x)}(t)|] - A\Phi(-x) \right| \\ & \leq \varepsilon \bar{F}_Y^{*k_t(x)}(t) + A \left| \bar{F}_Y^{*k_t(x)}(t) - \Phi(-x) \right| + \int_{t-t_0}^t \mathbb{E}[|\overline{\text{BP}}(t)|] d\bar{F}_Y^{*k_t(x)}(u). \end{aligned}$$

The last term vanishes since $\mathbb{E}[|\overline{\text{BP}}(t)|]$ is uniformly bounded and $\bar{F}_Y^{*k_t(x)}(t) - \bar{F}_Y^{*k_t(x)}(t-t_0) = o(1)$. Furthermore, with $m = k_t(x) \rightarrow \infty$,

$$(3.4.73) \quad k_t(x) \sim \frac{t}{\bar{\nu}} + x \sqrt{t \frac{\bar{\sigma}^2}{\bar{\nu}^3}} \quad \text{precisely when} \quad t \sim m\bar{\nu} - x\bar{\sigma}\sqrt{m}.$$

As a result, by the central limit theorem and the fact that $\bar{\nu}$ and $\bar{\sigma}^2$ are the mean and the variance of the distribution function \bar{F}_Y ,

$$(3.4.74) \quad \lim_{t \rightarrow \infty} \bar{F}_Y^{*k_t(x)}(t) = \Phi(-x).$$

Together with (3.4.72), this proves the claim in (3.4.69), and hence Theorem 3.19. \square

EXERCISE 3.26 (Joint residual life-time and height convergence). *Adapt the proof of Theorem 3.19 to show that*

$$(3.4.75) \quad \lim_{t \rightarrow \infty} e^{-\lambda t} \mathbb{E}[|\text{BP}_{\leq k_t(x)}[t, t+s)|] = A\Phi(x)F_R(s),$$

where $|\text{BP}_{\leq k}[t, t+s)|$ is the number of alive individuals with residual life-time at most s and with height at most k . Thus, the asymptotic height and residual life-time are close to independent.

Distributional convergence of alive vertices with residual life-time and height. Theorems 3.18-3.19 investigate the mean number of vertices with given residual life-time and height. Since we can interpret $A = \lim_{t \rightarrow \infty} e^{-\lambda t} \mathbb{E}[|\mathbf{BP}(t)|]$, this suggests that on the event of survival, we also have that $e^{-\lambda t} |\mathbf{BP}[t, t+s]| \xrightarrow{d} W F_R(s)$, as well as $e^{-\lambda t} |\mathbf{BP}_{\leq k_t(x)}(t)| \xrightarrow{d} W \Phi(x)$. However, Theorems 3.18-3.19 only investigate first moment, and thus certainly do not imply this convergence in distribution. The asymptotics of the number of alive individuals and their heights and residual life-times are investigated in the following theorem:

THEOREM 3.20 (Residual life-time and heights in a continuous-time branching process). *Let $(\mathbf{BP}(t))_{t \geq 0}$ be an age-dependent branching process with offspring Δ . Assume that $\mathbb{E}[\Delta \log(\Delta)_+] < \infty$ so that the $X \log X$ condition holds. Then, with W denoting the limit in Theorem 3.15, and for all $x \in \mathbb{R}, s \geq 0$,*

$$(3.4.76) \quad e^{-\lambda t} |\mathbf{BP}_{\leq k_t(x)}[t, t+s]| \xrightarrow{d} \Phi(x) F_R(s) W,$$

and these limits also hold jointly for different x and s .

We will not give a proof of Theorem 3.20. Instead, we argue why the residual life-time and height of individuals are approximately independent from the asymptotic growth of the continuous-time branching process that is described by W . Note that W is primarily determined by what happens early on in the continuous-time branching process, since a fast or slow growth initially will be felt throughout the entire future. Thus, it is this initial growth that determines W . On the other hand, the majority of individuals counted in $|\mathbf{BP}(t)|$ were born in a time that is close to t . Thus, the heights and residual life-times of most of the individuals in $\mathbf{BP}(t)$ are described by what happens close to time t . This explains why these random influences are close to independent.

Universality of heights and birth-time scaling in continuous-branching processes. Theorem 3.20 is an extremely powerful result. Unfortunately, the central limit theorem for the height of alive individuals in a continuous-time branching process is not known in the most general setting, which is why we have only stated it for age-dependent branching processes. This is also the most useful result for us, as first-passage percolation on the configuration model gives rise to age-dependent branching processes. Recall that the continuous-time branching process has a Malthusian parameter when $\mathbb{E}[\Delta] < \infty$, and its rescaled limit exists when $\mathbb{E}[\Delta \log(\Delta)_+] < \infty$. Theorem 3.20 shows that in this case, the heights of vertices generally have a Gaussian profile, and the residual life-time has a well-defined limit. These results are true rather universally. Theorem 3.20 thus suggests that rather universal results might be expected in random graphs when the limiting degree distribution satisfies $\mathbb{E}[\Delta \log(\Delta)_+] < \infty$ and converges sufficiently nicely. In the following section, we investigate such settings more closely.

3.5. Universality for first-passage percolation on configuration models

In this section, we use the general theory of age-dependent branching processes as discussed in Section 3.4 in order to show that Theorem 3.7 extends to first-passage percolation with *general* continuous edge-weight distributions. We start by introducing the necessary notation. We investigate the configuration model $\text{CM}_n(\mathbf{d})$ under the degree-regularity condition in Condition 1.6(a)-(c). We need to extend Condition 1.6(c) slightly

in order to have that the branching process approximation of local neighborhoods in $\text{CM}_n(\mathbf{d})$ satisfies the $X \log X$ condition. We assume a *uniform $X^2 \log X$ condition* for the degrees of $\text{CM}_n(\mathbf{d})$, stating that

$$(3.5.1) \quad \limsup_{n \rightarrow \infty} \mathbb{E}[D_n^2 \log(D_n)_+] = \mathbb{E}[D^2 \log(D)_+] < \infty.$$

The condition in (3.5.1) implies that the $X \log X$ -condition is satisfied for the limiting continuous-time branching process with offspring distribution $D^* - 1$, where D^* is the size-biased version of D .

Now let $(\text{BP}(t))_{t \geq 0}$ denote the following continuous-time branching process:

- (a) At time $t = 0$, we start with one individual, which we refer to as the original ancestor or the root of the branching process. Generate D having the asymptotic degree-distribution F_D in Condition 1.6(a). This individual immediately dies giving rise to D children.
- (b) Each new individual v in the branching process lives for a random amount of time which has distribution F_Y , i.e., the edge weight distribution, and then dies. At the time of death again the individual gives birth to $D_v^* - 1$ children, where $D_v^* \sim F^*$. Life-times and numbers of offspring across individuals are independent.

Note that in the above construction, by Condition 1.6(b), if we let $X_v = D_v^* - 1$ be the number of children of an individual then the expected number of children satisfies

$$(3.5.2) \quad \mathbb{E}[X_v] = \mathbb{E}[D_v^* - 1] = \nu > 1,$$

The continuous-time branching process defined above is a *splitting* process, with life-time distribution F_Y and offspring distribution $D^* - 1$, except for the root, which has offspring distribution D . Thus, this is a *unimodular* continuous-time branching process. Naturally, the Malthusian parameter of this unimodular continuous-time branching process is equal to the one with offspring distribution $D^* - 1$. By (3.4.30), the Malthusian parameter λ of the branching process $\text{BP}(\cdot)$ is the unique solution of the equation

$$(3.5.3) \quad \nu \int_0^\infty e^{-\lambda t} dF_Y(t) = 1.$$

Since $\nu > 1$, we obtain that $\lambda \in (0, \infty)$. We also let λ_n be the solution to (3.5.3) with ν replaced with $\nu_n = \mathbb{E}[D_n(D_n - 1)]/\mathbb{E}[D_n]$. Clearly, $\lambda_n \rightarrow \lambda$, when Condition 1.6(c) holds, and $|\lambda_n - \lambda| = O(|\nu_n - \nu|)$.

By Theorem 3.15, there exists a random variable W such that

$$(3.5.4) \quad e^{-\lambda t} |\text{BP}(t)| \xrightarrow{d} W.$$

Recall that \bar{F}_Y denotes the stable-age distribution in Definition 3.17, and that $\bar{\nu}$ and $\bar{\sigma}^2$ are the mean and variance of \bar{F}_Y . Then $\bar{\nu}, \bar{\sigma}^2 \in (0, \infty)$, since $\lambda > 0$. We also define $\bar{F}_{Y,n}$ to be the stable-age distribution in Definition 3.17 with ν and λ replaced with ν_n and λ_n , and we let $\bar{\nu}_n$ and $\bar{\sigma}_n^2$ be its mean and variance. The main result in this section is the following theorem:

THEOREM 3.21 (Joint convergence hopcount and weight). *Consider the configuration model $\text{CM}_n(\mathbf{d})$ with degrees satisfying Condition 1.6(a)-(c) and (3.5.1), and with i.i.d. edge weights distributed according to the continuous distribution F_Y . Then, there exist constants $\alpha, \lambda, \beta \in (0, \infty)$ and α_n, λ_n with $\alpha_n \rightarrow \alpha, \lambda_n \rightarrow \lambda$, such that the hopcount H_n*

and weight \mathcal{C}_n of the optimal path between two uniformly selected vertices conditioned on being connected, satisfy

$$(3.5.5) \quad \left(\frac{H_n - \alpha_n \log n}{\sqrt{\beta \log n}}, \mathcal{C}_n - \frac{1}{\lambda_n} \log n \right) \xrightarrow{d} (Z, Q),$$

as $n \rightarrow \infty$, where

(a)

$$(3.5.6) \quad \alpha_n = \frac{1}{\lambda_n \bar{\nu}_n}, \quad \beta = \frac{\bar{\sigma}^2}{\bar{\nu}^3 \lambda},$$

(b) Z and Q are independent and Z has a standard normal distribution, while Q has a continuous distribution given by

$$(3.5.7) \quad Q = \frac{1}{\lambda} (\log(1/W^{(1)}) + \log(1/W^{(2)}) - \Lambda + c),$$

where $\mathbb{P}(\Lambda \leq x) = e^{-e^{-x}}$, so that Λ is a standard Gumbel random variable, $W^{(1)}, W^{(2)}$ are two independent copies of the variable W in (3.5.4), also independent from Λ , and c is the constant

$$(3.5.8) \quad c = \log(\mathbb{E}[D](\nu - 1)^2 / (\nu \alpha \bar{\nu})).$$

Theorem 3.21 implies that also the random variable Q is remarkably universal, in the sense that it always involves two martingale limit variables corresponding to the flow problem, and a Gumbel distribution.

REMARK 3.22 (Asymptotic mean). We can replace λ_n and α_n by their limits λ and $\alpha = 1/(\lambda \bar{\nu})$ in (3.5.5) precisely when $\alpha_n = \alpha + o(1/\sqrt{\log n})$ and $\lambda_n = \lambda + o(1/\log n)$. Since $|\bar{\nu}_n - \bar{\nu}| = O(|\nu_n - \nu|)$ and $|\alpha_n - \alpha| = O(|\nu_n - \nu|)$, these conditions are equivalent to $\nu_n = \nu + o(1/\log n)$ and $\nu_n = \nu + o(1/\sqrt{\log n})$, respectively.

Organization of this section. The remainder of this section is organized as follows. In Section 3.5.1, we first use Theorem 3.21 to show that identical results also apply to related random graph models, such as uniform random graphs with a prescribed degree sequence and rank-1 inhomogeneous random graphs. There, we also give some examples and match these up to related results on distances. Then, in Section 3.5.2, we explain the key ingredients in the proof of Theorem 3.21.

3.5.1. Extensions and examples. Theorem 3.21 extends easily to *uniform* random graphs with a prescribed degree sequence:

THEOREM 3.23 (Extension to uniform random graphs with prescribed degrees). *Under the conditions of Theorem 3.21 the results in Theorem 3.21 apply to uniform random graphs with prescribed degree sequence $\text{UG}_n(\mathbf{d})$.*

The proof of Theorem 3.23 follows rather directly from that of Theorem 3.21, by conditioning on simplicity. By [160, Theorem 7.12], recalling (1.3.43), under Condition 1.6(a-c),

$$(3.5.9) \quad \lim_{n \rightarrow \infty} \mathbb{P}(\text{CM}_n(\mathbf{d}) \text{ simple}) = e^{-\nu/2 - \nu^2/4}.$$

The proof of Theorem 3.21 reveals that in order to find the minimal weight path between vertices U_1, U_2 , we only need to investigate of order \sqrt{n} edges. Therefore, the event of

simplicity of the configuration model will be mainly determined by the *uninspected* edges, and is therefore asymptotically independent of (H_n, \mathcal{C}_n) . This explains Theorem 3.23.

We next extend Theorem 3.21 to rank-1 inhomogeneous random graphs:

THEOREM 3.24 (Extension to rank-1 inhomogeneous random graphs). *Let \mathbf{w} satisfy Condition 1.1(a)-(c), and further assume that*

$$(3.5.10) \quad \lim_{n \rightarrow \infty} \mathbb{E}[W_n^2 \log(W_n)_+] = \mathbb{E}[W^2 \log(W)_+].$$

Then, the results in Theorem 3.21 also hold for $\text{GRG}_n(\mathbf{w})$, $\text{CL}_n(\mathbf{w})$ and $\text{NR}_n(\mathbf{w})$.

The proof of Theorem 3.24 is similar to the proof of Theorem 2.22 in Section 2.3. Indeed, we already know that Condition 1.1(a)-(c) implies that the degree sequence of $\text{GRG}_n(\mathbf{w})$ satisfies Condition 1.6(a)-(c). Therefore, the only thing left is to prove that (3.5.10) implies (3.5.1), which we omit here.

EXERCISE 3.27 (Example of exponential weights). *Show that Theorem 3.21 implies Theorem 3.7 in the case of exponential weights.*

EXAMPLE 3.25 (Exponential weights plus a large constant.). We next study what happens when $X_e = 1 + E_e/k$, where $(E_e)_{e \in E_n}$ are i.i.d. exponentials with mean 1, and k is a large constant. This setting is, apart from a trivial time-rescaling, identical to the setting where $X_e = k + E_e$. In this case, one would expect that for large k , H_n is close to the graph distance between a pair of uniformly chosen vertices in $[n]$, conditioned to be connected. An extension of Theorem 2.29, proved with Hooghiemstra and Van Mieghem in [162], shows that, for i.i.d. degrees with finite variance, $(H_n - \log_\nu n)_{n \geq 1}$ is a tight sequence of random variables. This suggests that, as $k \rightarrow \infty$,

$$(3.5.11) \quad \lambda \rightarrow \log \nu, \quad \bar{\nu} \rightarrow 1, \quad \frac{\bar{\sigma}^2}{\bar{\nu}^3 \lambda} \rightarrow 0.$$

We now check this intuitive argument. Indeed,

$$(3.5.12) \quad \nu \int_0^\infty e^{-\lambda x} dF_Y(x) = \nu k \int_1^\infty e^{-\lambda x} e^{-k(x-1)} dx = \frac{\nu k}{\lambda + k} e^{-\lambda} = 1.$$

While solving this equation *explicitly* is hard, it is not too difficult to see that $k \rightarrow \infty$ implies that $\lambda \rightarrow \log \nu$. We next investigate the stable-age distribution in this setting:

EXERCISE 3.28 (Stable-age distribution). *Show that for $Y = 1 + E/k$ and E an exponential with parameter 1, the stable-age distribution in Definition 3.17 is equal to $1 + \text{Exp}(k + \lambda)$.*

By Exercise 3.28, the stable-age distribution is equal to $1 + \text{Exp}(k + \lambda)$, so that $\bar{\nu} = 1 + 1/(k + \lambda)$, while $\bar{\sigma}^2 = 1/(k + \lambda)^2 \rightarrow 0$. Therefore, $\bar{\nu} \sim 1$, which in turn also implies that $\lambda \bar{\nu} \rightarrow \log \nu$. Further,

$$(3.5.13) \quad \frac{\bar{\sigma}^2}{\bar{\nu}^3 \lambda} = k^{-2} (\log \nu)^{-1} (1 + o(1)) \rightarrow 0.$$

This shows that the two settings of graph distances and first-passage percolation with weights $Y \stackrel{d}{=} 1 + \text{Exp}(1)/k$ match up nicely when $k \rightarrow \infty$.

3.5.2. Overview of proof of Theorem 3.21. To understand the smallest-weight path between two vertices U_1 and U_2 , think of water percolating through the network at rate one, started simultaneously from the two vertices. For any $t \geq 0$, the set of vertices first seen by the flow from U_i will often be referred to the flow cluster or the smallest-weight graph of vertex U_i . When the two flows collide or create prospective collision edges, then these generate prospective smallest-weight paths.

Recall the construction of the smallest-weight graphs from vertices U_1 and U_2 in Section 3.3. There, these graphs were constructed in *discrete* steps. Here we will rather grow them in continuous time, and we will write $\text{SWG}_t^{(i)}$ for these graphs at time t . Also, $\text{SWG}_t^{(i)}$ will contain more information, for example, not only the number of alive half-edges but also their residual life-times.

In Section 3.3, we have performed these flows by first growing $(\text{SWG}_t^{(1)})_{t \geq 0}$ until it contains $a_n = \sqrt{n}$ vertices, followed by the growth of $(\text{SWG}_t^{(2)})_{t \geq 0}$ until an edge was created linking $\text{SWG}_t^{(2)}$ to the smallest-weight graph of vertex U_1 . This construction crucially depends on the memoryless property of the exponential distribution, which implies that the residual life-times of all edges incident to the smallest-weight graph of vertex U_1 again have an exponential distribution. For general edge-weights, however, this is not the case and we need to revise our strategy. The strategy we now choose is that we grow $(\text{SWG}_t^{(1)})_{t \geq 0}$ and $(\text{SWG}_t^{(2)})_{t \geq 0}$ *simultaneously*, and each time we find a vertex, we check whether any of its half-edges are paired to half-edges in the other smallest-weight graph. When this is the case, however, this edge has not been completely filled by the fluid, so that it could create the smallest-weight path, but it might also not. Indeed, this depends on the residual life-time of the edge in question, as well as all other edges linking to two smallest-weight graphs.

Construction of flow clusters from two sources. Let us now give a precise mathematical formulation to the above description. We grow two flow clusters (i.e. two stochastic processes in continuous time) from U_1 and U_2 , simultaneously. We keep track of the the alive set $A(t)$ of half-edges that are incident to a vertex in the two flow clusters, but who have not yet been completely filled. The alive set $A(t)$ only changes at random times $T_0 = 0 < T_1 < T_2 < \dots$ and therefore we give the definition recursively. At time $t = T_0 = 0$, the vertices U_1 and U_2 die instantaneously, and give rise to d_{U_1} and d_{U_2} children. These children correspond to half-edges incident to U_1 and U_2 .

We start by testing whether any of the half-edges incident to U_1 are paired to one another. If so, then we remove both half-edges from the total set of d_{U_1} half-edges. We then define $X_0^{(1)}$ to be the number of unpaired half-edges after the self-loops incident to U_1 are removed. We next continue with the d_{U_2} half-edges incident to U_2 , and check whether they are paired to one of the $X_0^{(1)}$ remaining half-edges incident to U_1 or any of the d_{U_2} half-edges incident to U_2 . When such a half-edge is paired to one of the d_{U_2} sibling half-edges, a self-loop is formed. When such a half-edge is paired to one of the $X_0^{(1)}$ remaining half-edges incident to vertex U_1 , a so-called *collision edge* is formed. A collision edge possibly yields the path with minimal weight between U_1 and U_2 . We let $X_0^{(2)}$ denote the number of unpaired half-edges after the tests for collision edges and cycles have been performed. Note that, by construction, each of the $X_0^{(i)}$ half-edges incident to the vertices U_i , where $i \in \{1, 2\}$, are paired to *new* vertices, i.e., vertices distinct from U_1 and U_2 .

For the moment we collect the collision edges at time $T_0 = 0$, together with the weights of the connecting edge between U_1 and U_2 that are chosen i.i.d. from the weight distribution F_Y , and continue with the description of the flow clusters. All half-edges that are not paired to one of the other $d_{U_1} + d_{U_2} - 1$ half-edges incident to either U_1 or U_2 are called *alive*, and together form the set of alive half-edges $\mathbf{A}(0)$ at time 0. For $y \in \mathbf{A}(0)$, we define $I(y) = i$ if the half-edge y is incident to U_i , $i = 1, 2$, and we define $(R_0(y))_{y \in \mathbf{A}(0)}$ as an i.i.d. sequence of life-times having distribution function F_Y .

We denote the set of alive half-edges at time t by $\mathbf{A}(t)$. For $y \in \mathbf{A}(t)$, we record its label $I(y)$, which is the index $i \in \{1, 2\}$ to which U_i the half-edge is connected, and we let $H(y)$ denote the height of y , i.e., the number of edges in the path from the vertex V_y incident to y to $U_{I(y)}$. This height equals 0 for $y \in \mathbf{A}(0)$. When we introduce new half-edges in $\mathbf{A}(t)$ at later times we will specify the heights and labels of these half-edges. Now define $T_1 = \min_{y \in \mathbf{A}(0)} R_0(y)$ and denote by y_0^* the half-edge equal to the argument of this minimum, hence $R_0(y_0^*) = \min_{y \in \mathbf{A}(0)} R_0(y)$. Since life-times have a continuous distribution, y_0^* is a.s. *unique*. Now set $\mathbf{A}(t) = \mathbf{A}(0)$ for all $0 \leq t < T_1$, i.e., the set of alive half-edges remains the same during the interval $[0, T_1)$. Next, define the flow cluster $\text{SWG}(t)$, for $0 \leq t < T_1$, by

$$(3.5.14) \quad \text{SWG}(t) = \{y, I(y), H(y), R_t(y)\}_{y \in \mathbf{A}(0)},$$

where $I(y)$ and $H(y)$ are defined above and $R_t(y) = R_0(y) - t$, for $0 \leq t \leq T_1$, denotes the remaining life-time of half-edge y . This concludes the initial step in the recursion, where we defined $\mathbf{A}(t)$ and $\text{SWG}(t)$ during the random interval $[0, T_1)$.

We continue recursively, by defining $\mathbf{A}(t)$ and $\text{SWG}(t)$ during the random interval $[T_k, T_{k+1})$, given that the processes are defined on $[0, T_k)$. For a half-edge y , we recall that P_y is the half-edge to which it is paired, and V_y is the vertex to which it is incident. At time $t = T_k$, we let y_{k-1}^* be the argument of $\min_{y \in \mathbf{A}(t)} R_t(y)$, and we remove y_{k-1}^* from the set $\mathbf{A}(t-)$. We then pair y_{k-1}^* to a uniform available half-edge, which we denote by z_k . By construction, we know that $z_k \equiv P_{y_{k-1}^*} \notin \mathbf{A}(t-)$, so that V_{z_k} is not a vertex that has been reached by the flow at time t . Then, for each of the $d_{V_{z_k}} - 1$ other half-edges incident to vertex V_{z_k} we test whether it is part of a self-loop or paired to a half-edge from the set $\mathbf{A}(t-)$. All half-edges incident to V_{z_k} that are part of a self-loop or are paired to a half-edge incident to $\mathbf{A}(t-)$ are removed from vertex V_{z_k} . We also remove the involved half-edges from the set $\mathbf{A}(t-)$. We will discuss the role of the half-edges incident to V_{z_k} that are paired to half-edges in $\mathbf{A}(t-)$ in more detail in the next paragraph below.

We call a half-edge x a *sibling half-edge* of half-edge y when $V_x = V_y$, i.e., the half-edges are incident to the same vertex. For all the remaining siblings of z_k we do the following: Let x be one such half-edge incident to V_{z_k} , then x is added to $\mathbf{A}(T_k)$, and we define $I(x) = I(y_{k-1}^*)$, $H(x) = H(y_{k-1}^*) + 1$, while $R_{T_k}(x)$ is an i.i.d. life-time with distribution F_Y . We now set $\mathbf{A}(t) = \mathbf{A}(T_k)$ for $T_k \leq t < T_{k+1}$, where $T_{k+1} = T_k + \min_{y \in \mathbf{A}(T_k)} R_{T_k}(y)$, and where the minimizing half-edge is called y_k^* . Furthermore, for $t \in [T_k, T_{k+1})$, we can define $\text{SWG}(t)$ by (3.5.14), where $R_t(y) = R_{T_k}(y) - (t - T_k)$. Finally, we denote the number of the $d_{V_{z_k}} - 1$ other half-edges incident to vertex V_{z_k} that do not form a self-loop and that are not paired to a half-edge from the set $\mathbf{A}(t-)$ by X_k . Later, it will also be convenient to introduce $\Delta_k = d_{V_{z_k}} - 1$. Let $S_k = |\mathbf{A}(T_k)|$, so that $S_0 = X_0^{(1)} + X_0^{(2)}$, while S_k satisfies the recursion

$$(3.5.15) \quad S_k = S_{k-1} + X_k - 1.$$

This describes the evolution of $(\text{SWG}(t))_{t \geq 0}$.

Cycle edges and collision edges. At the times T_k , $k \geq 1$, we find the half-edge y_{k-1}^* which is paired to $z_k = P_{y_{k-1}^*}$, and for each of the other half-edges x incident to V_{z_k} , we check whether or not $P_x \in \mathbf{A}(T_k-)$, where P_x is the half-edge to which x is paired. The half-edges paired to alive half-edges in $\mathbf{A}(T_k-)$ are special. Indeed, the edge (x, P_x) creates a cycle when $I(x) = I(P_x)$ while (x, P_x) completes a path between U_1 and U_2 when $I(x) = 3 - I(P_x)$. Precisely the latter edges can create the smallest-weight path between U_1, U_2 . Let us describe these collision edges in more detail.

At time T_k and when we create a collision edge consisting of x_k and P_{x_k} , then we record

$$(3.5.16) \quad \left((T_k, I(z_k), H(z_k), H(P_{x_k}), R_{T_k}(P_{x_k})) \right)_{k \geq 0}.$$

It is possible that multiple half-edges incident to V_{z_k} create collision edges, and if so, we collect all of them in the list in (3.5.16). In this definition it is tempting to write $I(x_k)$ and $H(x_k)$, but note that $x_k \notin \mathbf{A}(T_k)$, whereas its sibling half-edge does satisfy $z_k \in \mathbf{A}(T_k)$, and, moreover, x_k and z_k have the same ancestor and the same height.

With some abuse of notation we denote the i th collision edge by (x_i, P_{x_i}) . Here P_{x_i} is an alive half-edge and x_i the half-edge which pairs to P_{x_i} , further z_i is the sibling of x_i paired with the minimal edge y^* found by the flow. Let $T_i^{(\text{col})}$ be the time of creation of the i th collision edge. The weight of the (unique) path between U_1 and U_2 that passes through the edge consisting of x_i and P_{x_i} equals $2T_i^{(\text{col})} + R_{T_i^{(\text{col})}}(P_{x_i})$, so that the smallest weight of all paths between U_1 and U_2 equals

$$(3.5.17) \quad \mathcal{C}_n = \min_{i \geq 0} [2T_i^{(\text{col})} + R_{T_i^{(\text{col})}}(P_{x_i})].$$

Let I^* denote the minimizer of $i \mapsto 2T_i^{(\text{col})} + R_{T_i^{(\text{col})}}(P_{x_i})$, then

$$(3.5.18) \quad H_n = H(z_{I^*}) + H(P_{x_{I^*}}) + 1.$$

The equalities (3.5.17) and (3.5.18) need a proof, which we give now:

Proof that \mathcal{C}_n given by (3.5.17) yields the minimal weight, and that (3.5.18) follows. Observe that each path between U_1 and U_2 has a weight \mathcal{C} that can be written in the form $2T_i^{(\text{col})} + R_{T_i^{(\text{col})}}(P_{x_i})$ for some $i \geq 0$. Indeed, let $(i_0 = U_1, i_1, i_2, \dots, i_k = U_2)$ form a path with weight \mathcal{C} , and denote the weight on $i_{j-1}i_j$ by Y_{e_j} for $1 \leq j \leq k$. For $k = 1$, we obviously find $Y_{e_1} = 2T_0 + Y_{e_1}$. For general $k \geq 1$, take the maximal $j \geq 0$ such that $Y_{e_1} + \dots + Y_{e_j} \leq \mathcal{C}/2$. Then, we write

$$(3.5.19) \quad \mathcal{C} = \begin{cases} 2 \sum_{s=1}^j Y_{e_s} + [\sum_{s=j+1}^k Y_{e_s} - \sum_{s=1}^j Y_{e_s}], & \text{when } \sum_{s=1}^j Y_{e_s} < \sum_{s=j+1}^k Y_{e_s}, \\ 2 \sum_{s=j+1}^k Y_{e_s} + [\sum_{s=1}^j Y_{e_s} - \sum_{s=j+1}^k Y_{e_s}], & \text{when } \sum_{s=1}^j Y_{e_s} > \sum_{s=j+1}^k Y_{e_s}, \end{cases}$$

which in either case is of the form $\mathcal{C} = 2T_m + R_{T_m}(y)$, for some $m \geq 0$ and some half-edge y . Note that in the construction of the flow clusters, instead of putting weight on the edges, we have given weights to half-edges instead. In the representation (3.5.17), the full edge weight is given to the active half-edges and weight 0 to the ones with which they are paired. When the collision edge has been found we give the full weight to the half-edge P_x that is part of the collision edge (x, P_x) . Thus, in fact, by the redistribution of the weights in (3.5.17) is an equality in distribution. This completes the proof of the claim

in (3.5.17). The above argument shows that (3.5.17) produces the smallest-weight path, and thus (3.5.18) also follows. \square

Basic constructions and properties. To state our main technical result concerning the appearance of collision edges, we need to introduce some notation. We start by defining a rescaled version of the point process corresponding to the points in (3.5.16). Let us first set up some notation. For $i \in \{1, 2\}$ and $t \geq 0$, we let

$$(3.5.20) \quad |\text{SWG}(t)| = \#\{y \in \mathbf{A}(t)\}, \quad |\text{SWG}^{(i)}(t)| = \#\{y \in \mathbf{A}(t) : I(y) = i\},$$

be the number of alive half-edges at time t , as well as those that are closest to vertex i . By construction, since we check whether the half-edges form a cycle or a collision edge when the half-edges are born, $\text{SWG}^{(1)}(t)$ and $\text{SWG}^{(2)}(t)$ are *disjoint*.

Further, we will often work conditionally on the flow at time s . For this, we introduce the filtration $(\mathcal{F}_s)_{s \geq 0}$ with $\mathcal{F}_s = \sigma((\text{SWG}(t))_{t \in [0, s]})$ denoting the sigma-algebra generated by the smallest-weight graph up to time s . This informally means that $\mathcal{F}_s = \sigma((\text{SWG}(t))_{t \in [0, s]})$ contains all the information of the flow up to time s .

Fix a deterministic sequence $s_n \rightarrow \infty$ that will be chosen later on. Now let

$$(3.5.21) \quad t_n = \frac{1}{2\lambda_n} \log n, \quad \bar{t}_n = \frac{1}{2\lambda_n} \log n - \frac{1}{2\lambda_n} \log(W_{s_n}^{(1)} W_{s_n}^{(2)}),$$

where, for $s \geq 0$,

$$(3.5.22) \quad W_s^{(i)} = e^{-\lambda_n s} |\text{SWG}^{(i)}(s)|.$$

Note that $e^{\lambda_n t_n} = \sqrt{n}$, so that at time t_n , both $|\text{SWG}^{(i)}(s)|$ are of order \sqrt{n} ; consequently the variable t_n denotes the typical time at which collision edges start appearing, and the time \bar{t}_n incorporates for stochastic fluctuations in the size of the SWGs. We choose $s_n \rightarrow \infty$ such that $\text{SWG}^{(i)}(t)$ for $t \leq s_n$ can be coupled with two independent two-stage branching processes $\text{BP}^{(i)}(t)$ such that $\{\text{BP}(t) = \text{SWG}(t) \forall t \leq s_n\}$ whp (see (3.5.43) below).

Recall from (3.4.53) that the *residual life-time distribution* F_R has density f_R given by

$$(3.5.23) \quad f_R(x) = \frac{\int_0^\infty e^{-\lambda y} f_Y(x+y) dy}{\int_0^\infty e^{-\lambda y} [1 - F_Y(y)] dy}.$$

Recall further that the i th collision edge is given by (x_i, P_{x_i}) , where P_{x_i} is an alive half-edge and x_i the half-edge which pairs to P_{x_i} and that is incident to a vertex in the other smallest-weight graph. In terms of the above definitions, we define

$$(3.5.24) \quad \bar{T}_i^{(\text{col})} = T_i^{(\text{col})} - \bar{t}_n, \quad \bar{H}_i^{(\text{or})} = \frac{H(x_i) - t_n/\bar{\nu}_n}{\sqrt{\bar{\sigma}^2 t_n/\bar{\nu}^3}}, \quad \bar{H}_i^{(\text{de})} = \frac{H(P_{x_i}) - t_n/\bar{\nu}_n}{\sqrt{\bar{\sigma}^2 t_n/\bar{\nu}^3}},$$

and write the random variables $(\Xi_i)_{i \geq 1}$ with $\Xi_i \in \mathbb{R} \times \{1, 2\} \times \mathbb{R} \times \mathbb{R} \times [0, \infty)$, by

$$(3.5.25) \quad \Xi_i = (\bar{T}_i^{(\text{col})}, I(x_i), \bar{H}_i^{(\text{or})}, \bar{H}_i^{(\text{de})}, R_{T_i}(P_{x_i})).$$

Then, for sets A in the Borel σ -algebra of the space $\mathcal{S} = \mathbb{R} \times \{1, 2\} \times \mathbb{R} \times \mathbb{R} \times [0, \infty)$, we define the point process

$$(3.5.26) \quad \Pi_n(A) = \sum_{i \geq 1} \delta_{\Xi_i}(A),$$

where δ_x gives measure 1 to the point x and $A \subset \mathcal{S}$.¹ In the theorem, we let Φ denote the distribution function of a standard normal random variable. The main result in the proof of Theorem 3.21 is the following Poisson point process limit result for the appearance of the collision edges, as well as its characteristics:

THEOREM 3.26 (Poisson point process limit of collision edges). *Consider the distribution of the point process $\Pi_n \in \mathcal{M}(\mathcal{S})$ defined in (3.5.26) conditional on $(\text{SWG}(s))_{s \in [0, s_n]}$ such that $W_{s_n}^{(1)} > 0$ and $W_{s_n}^{(2)} > 0$. Then Π_n converges in distribution as $n \rightarrow \infty$ to a Poisson point process Π with intensity measure*

$$(3.5.27) \quad \begin{aligned} & \lambda(dt \times i \times dx \times dy \times dr) \\ &= \frac{2\nu f_R(0)}{\mathbb{E}[D]} e^{2\lambda t} dt \otimes \{1/2, 1/2\} \otimes \Phi(dx) \otimes \Phi(dy) \otimes F_R(dr). \end{aligned}$$

Theorem 3.26 is equivalent to the statement that $\Pi_n(A) \xrightarrow{d} \Pi(A)$ for every $A \subseteq \mathcal{S}$, where $\Pi(A)$ is a Poisson random variable with mean $\int_A \lambda(dt \times i \times dx \times dy \times dr)$. The product structure of the intensity measure $\lambda(\cdot)$ in (3.5.27) implies that the limits of arrival times, index, and the two heights are asymptotically *independent*, which is closely related to the independence of the two limits in (3.5.5).

Completion of the proof of Theorem 3.21 subject to Theorem 3.26. Let us now prove Theorem 3.21 subject to Theorem 3.26. First of all, by (3.5.24), (3.5.17) and (3.5.18),

$$(3.5.28) \quad \left(\frac{H_n - \frac{1}{\lambda_n \bar{\nu}} \log n}{\sqrt{\frac{\bar{\sigma}^2}{\bar{\nu}^3 \lambda} \log n}}, \mathcal{C}_n - \frac{1}{\lambda_n} \log n \right),$$

is a continuous function of the point process Π_n , and, therefore, by the continuous mapping theorem, the above random variable converges in distribution to some limiting random variables (Z, Q) that we will determine now.

Recall that I^* denotes the minimizer of $i \mapsto 2T_i^{(\text{col})} + R_{T_i^{(\text{col})}}(P_{x_i})$. By (3.5.17), the weight \mathcal{C}_n as well as the value of I^* , are functions of the first and the last coordinates of Π_n . The hopcount H_n is a function of the third and the fourth, instead. By the product form of the intensity in (3.5.27), we obtain that the limits (Z, Q) are independent. Therefore, it suffices to study their marginals.

We start with the limiting distribution of the hopcount. By (3.5.24),

$$(3.5.29) \quad \frac{H_n - \frac{1}{\lambda_n \bar{\nu}_n} \log n}{\sqrt{\frac{\bar{\sigma}^2}{\bar{\nu}^3 \alpha} \log n}} = \frac{1}{2} \sqrt{2} \bar{H}_{I^*}^{(\text{or})} + \frac{1}{2} \sqrt{2} \bar{H}_{I^*}^{(\text{de})} + O_{\mathbb{P}}(1).$$

By Theorem 3.26, the random variables $(\bar{H}_{I^*}^{(\text{or})}, \bar{H}_{I^*}^{(\text{de})})$ converge in distribution to two *independent* standard normals, so that also the left-hand side of (3.5.29) converges in distribution to a standard normal.

¹Let $\mathcal{M}(\mathcal{S})$ denote the space of all simple locally finite point processes on \mathcal{S} equipped with the vague topology (see e.g. [191]). On this space one can naturally define the notion of weak convergence of a sequence of random point processes $\Pi_n \in \mathcal{M}(\mathcal{S})$. This is the notion of convergence referred to in Theorem 3.26.

The limiting distribution of the weight \mathcal{C}_n is slightly more involved. By (3.5.21), (3.5.17) and (3.5.24),

$$\begin{aligned}
 (3.5.30) \quad \mathcal{C}_n - \frac{1}{\lambda_n} \log n &= \mathcal{C}_n - 2t_n = \mathcal{C}_n - 2\bar{t}_n - \frac{1}{\lambda_n} \log(W_{s_n}^{(1)}W_{s_n}^{(2)}) \\
 &= -\frac{1}{\lambda_n} \log(W_{s_n}^{(1)}W_{s_n}^{(2)}) + \min_{i \geq 1} [2T_i^{(\text{col})} + R_{T_i^{(\text{col})}}(P_{x_i})] - 2\bar{t}_n \\
 &= -\frac{1}{\lambda_n} \log(W_{s_n}^{(1)}W_{s_n}^{(2)}) + \min_{i \geq 1} [2\bar{T}_i^{(\text{col})} + R_{T_i^{(\text{col})}}(P_{x_i})].
 \end{aligned}$$

By (3.5.43) below, $(W_{s_n}^{(1)}, W_{s_n}^{(2)}) \xrightarrow{d} (W^{(1)}, W^{(2)})$, which are two independent copies of the random variable W in (3.4.24) in Theorem 3.15. Hence,

$$(3.5.31) \quad \mathcal{C}_n - \frac{1}{\lambda_n} \log n \xrightarrow{d} -\frac{1}{\lambda} \log(W^{(1)}W^{(2)}) + \min_{i \geq 1} [2P_i + R_i],$$

where $(P_i)_{i \geq 1}$ form a PPP with intensity $\frac{2\nu f_R(0)}{\mu} e^{2\lambda t} dt$, and $(R_i)_{i \geq 1}$ are i.i.d. random variables with distribution function F_R independently of $(P_i)_{i \geq 1}$. The distribution of the first point of the Poisson point process with intensity $2ce^{2\lambda t}$ is a rescaled Gumbel distribution:

EXERCISE 3.29 (Minimum of a PPP with exponential intensity). *Let (P_i) be a Poisson point process with intensity measure $\lambda(t) = 2ce^{2\lambda t}$. Show that $\min_{i \geq 1} [2P_i] \stackrel{d}{=} -\frac{1}{\lambda}\Lambda + \frac{1}{\lambda} \log(c/\lambda)$, where Λ has a Gumbel distribution.*

Comparing to Exercise 3.29, however, we need to compute $M = \min_{i \geq 1} [2P_i + R_i]$, where $(P_i)_{i \geq 1}$ is a PPP with intensity $\frac{2\nu f_R(0)}{\mathbb{E}[D]} e^{2\lambda t} dt$ and $(R_i)_{i \geq 1}$ are i.i.d. random variables with distribution function F_R .

Interestingly, the *same* Gumbel variable appears in this analysis, but the constant is changed. Indeed, let us identify the distribution of $M = \min_{i \geq 1} [2P_i + R_i]$. First, $(2P_i)_{i \geq 1}$ forms a Poisson process with intensity $\frac{\nu f_R(0)}{\mathbb{E}[D]} e^{\lambda t} dt$. According to Resnick [232, Example 3.3] the point process $(2P_i + R_i)_{i \geq 1}$ is a non-homogeneous Poisson process with mean-measure equal to the convolution of $\mu(-\infty, x] = \int_{-\infty}^x \frac{\nu f_R(0)}{\mathbb{E}[D]} e^{\lambda t} dt$ and F_R . Hence $\mathbb{P}(M \geq x)$ equals the Poisson probability of 0, where the parameter of the Poisson distribution is $(\mu * F_R)(x)$, so that

$$(3.5.32) \quad \mathbb{P}(M \geq x) = \exp \left\{ -\frac{\nu f_R(0)}{\mathbb{E}[D]} e^{\lambda x} \int_0^\infty F_R(z) e^{-\lambda z} dz \right\}.$$

Let Λ have a Gumbel distribution, i.e., $\mathbb{P}(\Lambda \leq x) = e^{-e^{-x}}$, $x \in \mathbb{R}$, then

$$(3.5.33) \quad \mathbb{P}(-a\Lambda + b \geq x) = e^{-e^{x/a} e^{-b/a}}.$$

Solving for a and b in

$$(3.5.34) \quad \frac{\nu f_R(0)}{\mathbb{E}[D]} e^{\lambda x} \int_0^\infty F_R(z) e^{-\lambda z} dz = e^{x/a} e^{-b/a},$$

we obtain

$$(3.5.35) \quad a = 1/\lambda \quad \text{and} \quad b = -\lambda^{-1} \log \left((\nu f_R(0)/\mathbb{E}[D]) \int_0^\infty F_R(z) e^{-\lambda z} dz \right).$$

From this we conclude that

$$(3.5.36) \quad \min_{i \geq 1} (2P_i + R_i) \stackrel{d}{=} -\lambda^{-1}\Lambda - \lambda^{-1} \log(\nu f_R(0)q/\mathbb{E}[D]),$$

with $q = \int_0^\infty F_R(z)e^{-\lambda z} dz$. In the following lemma, we simplify the constants q and $f_R(0)$:

LEMMA 3.27 (The constant). *The constants $q = \int_0^\infty F_R(z)e^{-\lambda z} dz$ and $f_R(0)$ are given by*

$$(3.5.37) \quad q = \bar{\nu}/(\nu - 1), \quad f_R(0) = \lambda/(\nu - 1).$$

Consequently, the constant c in the limit variable (3.5.7) equals

$$(3.5.38) \quad c = -\log(\nu f_R(0)q/\mathbb{E}[D]) = \log(\mathbb{E}[D](\nu - 1)^2/(\lambda\nu\bar{\nu})).$$

PROOF. We start by computing $f_R(0)$, for which we note that by (3.5.23) and the definition of the Malthusian parameter in Definition 3.14,

$$(3.5.39) \quad f_R(0) = \frac{\int_0^\infty e^{-\lambda y} f_Y(y) dy}{\int_0^\infty e^{-\lambda y} [1 - F_Y(y)] dy} = \left(\nu \int_0^\infty e^{-\lambda y} [1 - F_Y(y)] dy \right)^{-1}.$$

Further, by partial integration,

$$(3.5.40) \quad \begin{aligned} \int_0^\infty e^{-\lambda y} [1 - F_Y(y)] dy &= \left[-\frac{1}{\lambda} e^{-\lambda y} [1 - F_Y(y)] \right]_{y=0}^\infty - \frac{1}{\lambda} \int_0^\infty e^{-\lambda y} f_Y(y) dy \\ &= \frac{1}{\lambda} - \frac{1}{\lambda\nu} = \frac{\nu - 1}{\lambda\nu}, \end{aligned}$$

where we again use the definition of the Malthusian parameter in Definition 3.14. Combining both equalities yields $f_R(0) = \lambda/(\nu - 1)$.

For q , we again use partial integration, followed by the substitution of (3.5.23). This yields

$$(3.5.41) \quad \begin{aligned} q &= \int_0^\infty F_R(z)e^{-\lambda z} dz = \frac{1}{\lambda} \int_0^\infty f_R(z)e^{-\lambda z} dz \\ &= \frac{\nu}{\nu - 1} \int_0^\infty e^{-\lambda z} \int_0^\infty e^{-\lambda y} f_Y(y + z) dy dz, \end{aligned}$$

by (3.5.40). The final integral can be computed using

$$(3.5.42) \quad \begin{aligned} &\int_{-\infty}^\infty e^{-\lambda z} \mathbb{1}_{\{z \geq 0\}} \int_{-\infty}^\infty e^{-\lambda y} f_Y(y + z) \mathbb{1}_{\{y \geq 0\}} dy dz \\ &= \int_0^\infty s f_Y(s) e^{-\lambda s} ds = \frac{1}{\nu} \int_0^\infty s \bar{F}_Y(ds) = \bar{\nu}/\nu. \end{aligned}$$

This completes the proof of Theorem 3.21 subject to Theorem 3.26. \square

Overview of the proof of Theorem 3.26. We next informally explain how to prove Theorem 3.26. Recall the smallest-weight graph or flow cluster $\text{SWG}(t)$ defined in the previous section as well as the associated filtration $(\mathcal{F}_t)_{t \geq 0}$. The proof relies on a coupling of these flow clusters from two points with $(\text{BP}(t))_{t \geq 0}$ where $\text{BP}(t) = (\text{BP}^{(1)}(t), \text{BP}^{(2)}(t))$ are two independent unimodular continuous-time branching processes where the root offspring distribution D and all other vertices offspring $D^* - 1$, in such a way that for some $s_n \rightarrow \infty$

$$(3.5.43) \quad \mathbb{P}\left((\text{SWG}(s))_{s \in [0, s_n]} = (\text{BP}(s))_{s \in [0, s_n]} \right) = 1 - o(1).$$

EXERCISE 3.30 (Perfect coupling for increasing times). *Prove the perfect coupling statement in (3.5.43) using the local weak convergence for $\text{CM}_n(\mathbf{d})$ in Theorem 2.11.*

By (3.5.43), with $W_{s_n}^{(i)} = e^{-\lambda_n s_n} |\text{SWG}^{(i)}(s_n)|$,

$$(3.5.44) \quad \liminf_{\varepsilon \downarrow 0} \liminf_{n \rightarrow \infty} \mathbb{P} \left(W_{s_n}^{(1)} \in [\varepsilon, 1/\varepsilon], W_{s_n}^{(2)} \in [\varepsilon, 1/\varepsilon] \mid W_{s_n}^{(1)} > 0, W_{s_n}^{(2)} > 0 \right) = 1.$$

Further, the coupling satisfies that, conditionally on \mathcal{F}_{s_n} ,

$$(3.5.45) \quad \mathbb{P} \left(|\text{SWG}(t_n + B^{(n)}) \Delta \text{BP}_{(n)}(t_n + B^{(n)})| \geq \varepsilon_n \sqrt{n} \mid \mathcal{F}_{s_n} \right) \xrightarrow{\mathbb{P}} 0.$$

For $i \in \{1, 2\}$, $k \geq 0$, and $t \geq 0$, we define

$$(3.5.46) \quad |\text{SWG}_{\leq k}^{(i)}[t, t+s]| = \#\{y \in \mathbf{A}(t) : I(y) = i, H(y) \leq k, R_t(y) \in [0, s]\},$$

as the number of alive half-edges at time t that (a) are in the SWG of vertex U_i , (b) have height at most k , and (c) have remaining life-time at most s . To formulate the CLT for the height of vertices, we recall (3.4.62). Finally, for a half-edge $y \in \mathbf{A}(t)$, we let $X_y^* = d_{V_y} - 1$.

We will argue that, for $\mathcal{I} = [a, b) \times \{j\} \times (-\infty, x] \times (-\infty, y] \times [0, s]$ a subset of \mathcal{S} ,

$$(3.5.47) \quad \mathbb{P}(\Pi_n(\mathcal{I}) = 0 \mid \mathcal{F}_{s_n}) \xrightarrow{\mathbb{P}} \exp \left\{ - \int_a^b \frac{2\nu f_R(0)}{\mathbb{E}[D]} e^{2\alpha t} \Phi(x) \Phi(y) F_R(s) dt \right\}.$$

By Kallenberg [191, Theorem 4.7], this proves the claim.

We associate the weight of an edge to its half-edge that becomes alive first. Below, we will say that a half-edge is *found* at time t by the flow when the weight of the smallest-weight path between U_1 and U_2 and the vertex incident to the half-edge we consider, together with its own weight, is at most t .

We assume that $b = a + \varepsilon$, where $\varepsilon > 0$ is small, so that we need to show that

$$(3.5.48) \quad \mathbb{P}(\Pi_n(\mathcal{I}) \geq 1 \mid \mathcal{F}_{s_n}) \approx \varepsilon \frac{2\nu f_R(0)}{\mathbb{E}[D]} e^{2\alpha a} \Phi(x) \Phi(y) F_R(s).$$

The number of half-edges z in $\text{SWG}^{(j)}(\bar{t}_n + a)$ with $I(z) = j \in \{1, 2\}$ that is found by the flow in the interval $[\bar{t}_n + a, \bar{t}_n + b]$ is close to $|\text{SWG}^{(j)}[\bar{t}_n + a, \bar{t}_n + b]|$. These half-edges are in turn paired to other half-edges whose siblings can also be found in the time interval of interest $[\bar{t}_n + a, \bar{t}_n + b]$. When $b = a + \varepsilon$ with $\varepsilon > 0$ small, this number is negligible, and we will ignore this effect. In order for the half-edge z that is found in the time interval $[\bar{t}_n + a, \bar{t}_n + b)$ to create a collision edge, it needs to be paired to the half-edge P_z for which one of the sibling half-edges pairs to a half-edge incident to the *other* SWG. On average, there are $\nu |\text{SWG}^{(j)}[\bar{t}_n + a, \bar{t}_n + b]|$ sibling half-edges to P_z , and each of them pairs to a half-edge incident to $\text{SWG}^{(3-j)}(\bar{t}_n + a)$ with probability close to

$$\frac{|\text{SWG}^{(3-j)}(\bar{t}_n + a)|}{\ell_n}.$$

Therefore, the probability that at least one collision edge is created in the interval $[\bar{t}_n + a, \bar{t}_n + b]$ is close to

$$\frac{\nu}{\ell_n} |\text{SWG}^{(j)}[\bar{t}_n + a, \bar{t}_n + b]| |\text{SWG}^{(3-j)}(\bar{t}_n + a)| \approx \frac{\nu}{\ell_n} F_R(b-a) |\text{SWG}^{(j)}(\bar{t}_n + a)| |\text{SWG}^{(3-j)}(\bar{t}_n + a)|,$$

where we use Theorem 3.18. We can approximate $F_R(b-a) \approx (b-a)f_R(0) = \varepsilon f_R(0)$. Also, replacing $|\text{SWG}^{(j)}(\bar{t}_n + a)| \approx W_{s_n}^{(j)} e^{\lambda_n(\bar{t}_n + a)}$, as suggested by Theorem 3.15, leads us to

$$\frac{\nu}{\ell_n} |\text{SWG}^{(j)}[\bar{t}_n + a, \bar{t}_n + b]| |\text{SWG}^{(3-j)}(\bar{t}_n + a)| \approx \frac{\varepsilon f_R(0)}{\ell_n} W_{s_n}^{(j)} W_{s_n}^{(3-j)} e^{2\lambda_n(\bar{t}_n + a)}.$$

By the definition of \bar{t}_n in (3.5.21),

$$(3.5.49) \quad W_{s_n}^{(j)} W_{s_n}^{(3-j)} e^{2\lambda_n(\bar{t}_n + a)} = W_{s_n}^{(1)} W_{s_n}^{(2)} e^{2\lambda_n(\bar{t}_n + a)} = n e^{2\lambda_n a}.$$

We conclude that the probability that at least one collision edge is created in the interval $[\bar{t}_n + a, \bar{t}_n + b]$ is close to

$$(3.5.50) \quad \frac{\nu \varepsilon f_R(0)}{\ell_n} n e^{2\lambda_n a} = \frac{\nu \varepsilon f_R(0)}{\mathbb{E}[D_n]} e^{2\lambda_n a} \rightarrow \frac{\nu \varepsilon f_R(0)}{\mathbb{E}[D]} e^{2\lambda a},$$

as required. Further, by Exercise 3.26 (which is based on Theorems 3.18 and 3.19) the height of the two half-edges that form the collision edge are close to normal, and the residual life-time of the collision half-edge has distribution close to F_R . This explains Theorem 3.26. We omit further details.

3.6. First-passage percolation on scale-free configuration models

In this section, we discuss first-passage percolation on random graphs with infinite-variance degrees. We start by discussing the behavior of their continuous-time branching process approximations.

3.6.1. Explosion vs. conservation of infinite-mean age-dependent processes.

So far, we have only dealt with continuous-time branching processes satisfying the $X \log X$ -condition. For age-dependent branching processes with offspring $\Delta = D^* - 1$, this in particular implies that $\mathbb{E}[\Delta] < \infty$. We now investigate the case where $\mathbb{E}[\Delta] = \infty$, in which the continuous-time branching process may explode in finite time. We will make the stronger assumption that $\mathbb{E}[\Delta^{1-\delta}] = \infty$ for some $\delta > 0$. We call a continuous-time branching process for which $\lim_{m \rightarrow \infty} T_m = T_\infty < \infty$ a.s. *explosive*, and one where $\lim_{m \rightarrow \infty} T_m = \infty$ a.s. *conservative*. The main goal of this section is to investigate when continuous-time branching processes are explosive and conservative, and then study the properties of first-passage percolation on $\text{CM}_n(\mathbf{D})$ when they are. We stick to the setting of i.i.d. degrees in this section, since we will rely on Proposition 2.43 that produces a very strong coupling result of the size-biased reordering of the degrees in $\text{CM}_n(\mathbf{D})$ to i.i.d. random variables in this setting.

Exponential edge-weights: explosion of Bellman-Harris processes. The results are simplest in the case of exponential life-times, with which we start. In Section 3.3, we have seen that the m th split time T_m for a Bellman-Harris process is given by

$$(3.6.1) \quad T_m = \sum_{j=1}^m E_j / S_j,$$

where $S_j = \sum_{i=1}^j \Delta_i - (j-1)$ and $(E_j)_{j \geq 1}$ are i.i.d. exponential random variables with mean 1. There, we have assumed that $\mathbb{E}[\Delta^{1+\varepsilon}] < \infty$ for some $\varepsilon > 0$. In Theorem 3.15, the condition on Δ was weakened to $\mathbb{E}[\Delta \log(\Delta)_+] < \infty$ (see also Theorem 3.20). We now investigate cases where $\mathbb{E}[\Delta] = \infty$, for which it is possible that $\lim_{m \rightarrow \infty} T_m = T_\infty < \infty$

a.s. This means that the continuous-time branching process *explodes in finite time*, i.e., after a finite amount of time, there are infinitely many alive individuals. The following theorem gives a precise condition when this can occur:

THEOREM 3.28 (Finite explosion time). *A Bellman-Harris process with offspring Δ almost surely explodes in finite time with explosion time given by*

$$(3.6.2) \quad T_\infty = \sum_{j=1}^{\infty} E_j/S_j < \infty \quad a.s.$$

precisely when $\mathbb{P}(\Delta \geq 1) = 1$ and the probability generating function G_Δ of the offspring distribution Δ satisfies that there exists $\delta > 0$ such that

$$(3.6.3) \quad \int_0^\delta \frac{1}{1 - \mathbb{E}[e^{-t(\Delta-1)}]} ds < \infty.$$

The Bellman-Harris process with offspring Δ satisfying (3.6.3) can also have a finite explosion time when $\mathbb{P}(\Delta \geq 1) < 1$, but then only on the event of survival. Theorem 3.28 also has implications for age-dependent branching processes:

EXERCISE 3.31 (Explosion time for age-dependent branching process). *Show that if the life-time Y satisfies that there exists an $a > 0$ such that $Y \preceq aE$, then also the age-dependent branching process with life-time Y explodes in finite time when (3.6.3) holds.*

Proof of Theorem 3.28. Note that (3.6.3) and $\mathbb{P}(\Delta \geq 1) = 1$ imply that $\mathbb{P}(\Delta \geq 2) > 0$. Indeed, when $\mathbb{P}(\Delta = 1) = 1$, then $\mathbb{E}[e^{-t(\Delta-1)}] = 1$, and thus (3.6.3) fails to hold.

The birth times $m \mapsto T_m$ are increasing in m , and thus it suffices to prove that

$$(3.6.4) \quad \mathbb{E} \left[\sum_{j=1}^{\infty} E_j/S_j \right] < \infty.$$

For this, we compute that by independence of E_j and S_j ,

$$(3.6.5) \quad \mathbb{E} \left[\sum_{j=1}^{\infty} E_j/S_j \right] = \sum_{j=1}^{\infty} \mathbb{E}[1/S_j].$$

We rewrite, for any $\delta > 0$,

$$(3.6.6) \quad \mathbb{E}[1/S_j] = \int_0^\delta \mathbb{E}[e^{-tS_j}] dt + \mathbb{E}[e^{-\delta S_j}/S_j].$$

Using $S_j = 1 + \sum_{i=1}^j (\Delta_i - 1) \geq 1$, the contribution of the second term is bounded by

$$(3.6.7) \quad \begin{aligned} \sum_{j=1}^{\infty} \mathbb{E}[e^{-\delta S_j}/S_j] &\leq \sum_{j=1}^{\infty} \mathbb{E}[e^{-\delta S_j}] \leq \sum_{j=1}^{\infty} (\mathbb{E}[e^{-\delta(\Delta-1)}])^j \\ &\leq 1/(1 - \mathbb{E}[e^{-\delta(\Delta-1)}]) < \infty, \end{aligned}$$

since $\mathbb{P}(\Delta \geq 1) = 1$ implies that $S_j \geq 1$ a.s. for every $j \geq 1$. Further, $\mathbb{P}(\Delta \geq 1) = 1$ and $\mathbb{P}(\Delta \geq 2) > 0$ imply that $\mathbb{E}[e^{-\delta(\Delta-1)}] < 1$.

In a similar way, again using that $S_j = 1 + \sum_{i=1}^j (\Delta_i - 1)$, we compute the contribution of the first term as

$$(3.6.8) \quad \begin{aligned} \sum_{j=1}^{\infty} \int_0^{\delta} \mathbb{E}[e^{-tS_j}] dt &= \int_0^{\delta} \sum_{j=1}^{\infty} e^{-t} (\mathbb{E}[e^{-t(\Delta-1)}])^j dt \\ &= \int_0^{\delta} \frac{e^{-t} \mathbb{E}[e^{-t(\Delta-1)}]}{1 - \mathbb{E}[e^{-t(\Delta-1)}]} dt. \end{aligned}$$

By assumption, this is finite. Since $\mathbb{E}[T_{\infty}] < \infty$, certainly $T_{\infty} < \infty$ a.s. \square

General edge-weights: explosion of age-dependent processes. We next investigate two related theorems for age-dependent branching processes by Grey [141]. There, also examples are presented of explosive settings. Interestingly, age-dependent branching process with offspring distribution with power-law tails with exponent $\tau - 2$ (as we will be interested in below) are explosive for some life-time distribution Y for *all* $\tau \in (2, 3)$ or for *none*. There are, however, also life-time distributions for which the continuous-time branching process is not explosive:

THEOREM 3.29 (Infinite explosion time age-dependent branching processes). *For every Δ with $\mathbb{E}[\Delta] = \infty$, there exists a non-negative random variable Y such that the age-dependent branching process with offspring Δ and life-time Y does not explode in finite time.*

Of course, the easiest example is when there exists a $\varepsilon > 0$ such that $\mathbb{P}(Y \leq \varepsilon) = 0$:

EXERCISE 3.32 (Conservative age-dependent branching processes with infinite mean). *Show that if the life-time Y satisfies that there exists an $\varepsilon > 0$ such that $\mathbb{P}(Y \leq \varepsilon) = 0$, then the age-dependent branching process is conservative.*

We next investigate the precise condition on the behavior of $F_Y(t)$ for small t that guarantee explosion in the case of infinite-mean continuous-time branching processes where even a moment smaller than 1 is infinite. Grey [141] gives the following examples. When $F_Y(t) = e^{-k/t^{\beta}}$ for $k, \beta > 0$, the continuous-time branching process is explosive. When, on the other hand, $F_Y(t) = e^{-ke^{1/t}}$ for $k > 0$, the continuous-time branching process is conservative. We see that the tails of $F_Y(t)$ for small t need to be pretty small for the continuous-time branching process to become conservative!

We next discuss some more recent results that show more clearly where the boundary between explosion and conservation lies. We follow Amini, Devroye, Griffiths and Olver [18]. They observed that there is a rather obvious necessary condition for explosion, of which we will see that it is also close to optimal. Let Y_k be the minimum weight edge at level k in the tree. Then the sum of weights along any infinite path is certainly at least $\sum_{k \geq 0} Y_k$. We say that a fixed weighted tree is *min-summable* if $\sum_{k \geq 0} Y_k$ is bounded. We conclude that when a tree is not min-summable, it cannot have an exploding path and thus is conservative. This turns out to be close to an *if and only if* statement, which is quite surprising. We introduce some further notation to state the main result from [18].

We call an offspring distribution Δ *plump* when $\mathbb{P}(\Delta \geq m^{1+\varepsilon}) \geq 1/m$ for some $\varepsilon > 0$ and all m sufficiently large. This is closely related to the fact that $\mathbb{E}[\Delta^{1-\delta}] = \infty$ for some $\delta > 0$:

EXERCISE 3.33 (Plump and strongly infinite mean). *Show that Δ is plump when $\mathbb{E}[\Delta^{1-\delta}] = \infty$ for some $\delta > 0$.*

The main result of Amini, Devroye, Griffiths and Olver [18] is as follows:

THEOREM 3.30 (Explosion vs. conservation of age-dependent branching processes). *Let Δ be a plump offspring distribution, and assume that m_0 is such that $\mathbb{P}(\Delta \geq m^{1+\epsilon}) \geq 1/m$ for all $m \geq m_0$. Let h_k be defined by $h_1 = m_0$, and*

$$(3.6.9) \quad h_{k+1} = [1 - F_B]^{-1}(1/h_k).$$

Then, the age-dependent branching process with life-time distribution Y and offspring distribution Δ is explosive if and only if

$$(3.6.10) \quad \sum_{k \geq 1} F_Y^{-1}(1/h_k) < \infty.$$

In the case where $[1 - F_\Delta](x) \sim cx^{-\alpha}$ (for example, for $\Delta = D^* - 1$ in the configuration model, where the degree distribution satisfies $1 - F_D(x) \sim cx^{-(\tau-1)}$, for which $\alpha = \tau - 2$), we see that $h_k \sim \alpha^{-k}$. Then, it is not hard to see that (3.6.10) occurs precisely when

$$(3.6.11) \quad \int_{1/\epsilon}^{\infty} F_Y(e^{-t}) \frac{1}{t} dt < \infty.$$

See the exercise below. In particular, this condition does not depend on the precise power-law α , as already observed by Komjáthy [200], to which we refer for a much more extensive discussion:

EXERCISE 3.34 (Integral condition explosion). *Show that (3.6.10) is equivalent to (3.6.11) when Δ satisfies $[1 - F_\Delta](x) \sim cx^{-\alpha}$.*

Let us give some intuition behind the proof of Theorem 3.30. Let the offspring distribution Δ satisfy the conditions in Theorem 2.4, and assume that $\mathbb{P}(\Delta \geq 1) = 1$. Then, $\alpha^k \log(Z_k)$ converges a.s. to a positive limit. Thus, we can only have explosion when $\sum_{k \geq 1} \min_{i \leq \alpha^{-k}} Y_{k,i} < \infty$, where $(Y_{n,i})_{i,n \geq 1}$ is an infinite array of i.i.d. copies of the weight distribution Y . Amini, Devroye, Griffiths and Olver [18] show that this sum is finite precisely when $\sum_{k \geq 1} F_Y^{-1}(\alpha^{-k}) < \infty$, since $h_k \approx \alpha^{-k}$. The finiteness of this sum actually does not depend on the precise value of α .

We can construct a ‘greedy’ path from the maximal degree vertex in generation l for some large l by each time taking the vertex with the maximal degree amongst the offspring of the present vertex. Let $Z'_k \leq Z_k$ denote the offspring of the individual in the $(k - 1)$ st generation constructed this way. Then also $\log(Z'_k)$ is of the order α^{-k} (recall the intuition around (2.5.14) for this). By a minor adaptation of this argument, we can see that we can also construct a path along vertices with degree b^k as well as very small weight of the order $F_Y^{-1}(b^{-k})$, both for some $b > 1$. This sum is again finite precisely when (3.6.10) holds. See Amini, Devroye, Griffiths and Olver [18, Algorithm FINDPATH on Page 1878] for the precise ‘greedy’ construction of an exploding path.

EXERCISE 3.35 (Example explosion). *Assume that $\mathbb{P}(\Delta \geq 1) = 1$, and let $[1 - F_\Delta](x) \sim cx^{-\alpha}(1 + o(1))$ as $x \rightarrow \infty$ for some $\alpha \in (0, 1)$. Show that when $F_Y(t) = e^{-k/t^\beta}$ for $k, \beta > 0$, the continuous-time branching process is explosive.*

EXERCISE 3.36 (Example conservation). *Assume that $\mathbb{P}(\Delta \geq 1) = 1$, and let $[1 - F_\Delta](x) \sim cx^{-\alpha}(1 + o(1))$ as $x \rightarrow \infty$ for some $\alpha \in (0, 1)$. For which $k, \beta > 0$ is the age-dependent branching process with edge-weight distribution $F_Y(t) = e^{-ke^{1/t^\beta}}$ conservative?*

3.6.2. Scale-free configuration models with exponential edge-weights. Having studied continuous-time branching processes in the range where the offspring distribution has infinite $(1 - \varepsilon)$ th moment, we now turn to random graphs. The relevant setting is when $\tau \in (2, 3)$, as then any p th moment of $D^* - 1$ with $p > \tau - 2$ is infinite. We start with the configuration model and exponentially distributed edge-weights, where the results are the most precise, as proved with Bhamidi and Hooghiemstra [41]:

THEOREM 3.31 (Precise asymptotics for $\tau \in (2, 3)$). *Consider the configuration model $\text{CM}_n(\mathbf{D})$ with i.i.d. degrees having distribution F_D satisfying $\mathbb{P}(D = 1) = 0$, and satisfying that there exists $\tau \in (2, 3)$ and $0 < c_1 \leq c_2 < \infty$ such that, for all $x \geq 0$,*

$$(3.6.12) \quad c_1 x^{-(\tau-1)} \leq 1 - F_D(x) \leq c_2 x^{-(\tau-1)}.$$

Then, there exists a random variable Q such that

$$(3.6.13) \quad \mathcal{C}_n \xrightarrow{d} Q, \quad \frac{H_n - \alpha \log n}{\sqrt{\alpha \log n}} \xrightarrow{d} Z,$$

where $\alpha = \frac{2(\tau-2)}{\tau-1} \in (0, 1)$, Z has a standard normal distribution and

$$(3.6.14) \quad Q = Q_1 + Q_2,$$

where Q_1, Q_2 are two independent copies of a random variable which is the explosion time of the infinite-mean Markovian continuous-time branching process.

Theorem 3.31 shows that we can travel between pairs of vertices with a *bounded* travel time, irrespective of how large the graph is. Thus, the flow is extremely efficient. The number of edges in the smallest-weight path is much larger than typical shortest paths, which by Theorem 2.30 consist of $2 \log \log n / |\log(\tau - 2)|$ edges instead of the $\alpha \log n$ edges in Theorem 3.31. Thus, it pays off to make a long detour to avoid edges with high weights.

Let us now describe the random variables Q_1 and Q_2 in some more detail, using Theorem 3.28. Define S_i recursively by $S_0 = D$ and $S_i = S_{i-1} + D_i^* - 2$, where $(D_i^*)_{i \geq 1}$ are i.i.d. Then, $(S_i)_{i \geq 0}$ is a random walk starting in D with step distribution $D^* - 2 \geq 0$. In particular, when (3.6.12) holds for some $\tau \in (2, 3)$, then D^* is in the domain of attraction of a $(\tau - 2)$ -stable random variable. Therefore, $\mathbb{E}[(D^*)^p] = \infty$ for all $p \geq \tau - 2$. In terms of this random variable, we have the description

$$(3.6.15) \quad Q_1 = \sum_{j \geq 1} E_j / S_j.$$

When D^* is in the domain of attraction of a $(\tau - 2)$ -stable random variable, $\mathbb{E}[Q_1] < \infty$. Indeed, then it can be shown that $\mathbb{E}[S_j] \sim j^{1/(\tau-2)}$ and $1/(\tau - 2) > 1$, recall also (3.6.4). Comparing to Theorem 3.28, we see that Q_i is the explosion time of a two-stage Markov continuous-time branching process, where the offspring in the first generation is equal to D and the offspring in all other generations is equal to $D^* - 1$.

Theorem 3.31, combined with the result on graph distances in Theorem 2.30, raises the question what the *universality classes* are for first-passage percolation on $\text{CM}_n(\mathbf{D})$

with infinite-variance degrees. Indeed, contrary to the setting of finite-variance degrees in Theorem 3.21, we can show that there are continuous edge-weight distributions for which the hopcount and weight-distance scale like $\log \log n$:

EXERCISE 3.37 (Infinite variance degrees and another universality class). *Consider the configuration model $\text{CM}_n(\mathbf{d})$ where the degrees satisfy the conditions in Theorem 2.30. Let the edge weights be equal to $(1 + E_e)_{e \in E(\text{CM}_n(\mathbf{d}))}$. Use Theorem 2.30 to show that, whp, $\mathcal{C}_n \leq 4(1 + \varepsilon) \log \log n / |\log(\tau - 2)|$. Use this, in turn, to show that $H_n \leq 4(1 + \varepsilon) \log \log n / |\log(\tau - 2)|$.*

Exercise 3.37 shows that there are at least two universality classes for first-passage percolation on the configuration model with $\tau \in (2, 3)$. We now explore such universality classes in more detail, also significantly improving upon the result in Exercise 3.37.

3.6.3. Explosion vs. conservation in scale-free configuration models. We next consider the general first-passage percolation problem on $\text{CM}_n(\mathbf{D})$ with i.i.d. degrees having infinite variance, and investigate the behavior of the weight-distance between two uniformly chosen vertices in the explosive and conservative cases, respectively. We follow the work with Barono and Komjáthy [29], whose main result is the following:

THEOREM 3.32 (Explosion vs. conservation for scale-free configuration models with general edge-weights). *Consider the configuration model $\text{CM}_n(\mathbf{D})$ with i.i.d. degrees having distribution F_D satisfying $\mathbb{P}(D = 1) = 0$, and satisfy that there exists $\tau \in (2, 3)$ and $0 < c_1 \leq c_2 < \infty$ such that (3.6.12) holds.*

(a) *When the age-dependent branching process with life-time distribution Y and offspring distribution $D^* - 1$ is explosive,*

$$(3.6.16) \quad \mathcal{C}_n \xrightarrow{d} Q_1 + Q_2,$$

where Q_1, Q_2 are two independent copies of the explosion time of the age-dependent unimodular branching process with distribution D .

(a) *When, on the other hand, the age-dependent branching process with life-time distribution Y and offspring distribution $D^* - 1$ is conservative,*

$$(3.6.17) \quad \mathcal{C}_n \xrightarrow{\mathbb{P}} \infty.$$

Theorem 3.32 is proved with Baroni and Komjáthy in [29]. We next sketch some intuition behind the proof of Theorem 3.32. Recall that Proposition 2.43 shows that whp we can couple n^ρ forward degrees $\Delta_m^{(n)}$ to n^ρ i.i.d. random variables having distribution $\Delta = D^* - 1$.

For $i \in \{1, 2\}$, let $T_m^{(i,n)}$ denote the time for vertex U_i to find the $(i + 1)$ st closest vertex to it (so that $T_m^{(0,n)} = 0$). The above coupling implies that $T_m^{(i,n)}$ in the graph can be coupled whp to $T_m^{(i)}$ for the continuous-time branching process. Now, when $m = n^\rho$ for ρ small, the two smallest-weight graphs will whp be disjoint. Indeed, recall Lemma 2.12, where a similar statement is proved for the graph-distance balls of size $m = o(n/d_{\max}) = o(n^{(\tau-2)/(\tau-1)})$ (recall (2.2.21)). Thus, taking $\rho < (\tau - 2)/(\tau - 1)$ and $m_n = n^\rho$, this proves that the stated disjointness occurs whp.

The coupling result immediately implies the lower bounds. Indeed, when the two smallest-weight graphs are disjoint, and whp,

$$(3.6.18) \quad \mathcal{C}_n \geq T_{m_n}^{(1,n)} + T_{m_n}^{(2,n)} = T_{m_n}^{(1)} + T_{m_n}^{(2)},$$

the latter by the successful coupling. In the conservative setting, $T_{m_n}^{(1)} \xrightarrow{\mathbb{P}} \infty$, proving the claim. Further, in the explosive setting, $(T_{m_n}^{(1)}, T_{m_n}^{(2)}) \xrightarrow{\mathbb{P}} (Q_1, Q_2)$, which are two i.i.d. copies of the explosion times of the continuous-time branching process approximation of the smallest-weight graph.

What remains is the upper bound on \mathcal{C}_n in terms of $T_{m_n}^{(1,n)} + T_{m_n}^{(2,n)}$. We prove in [29] that whp there exists a path connecting the two smallest-weight graphs of weight at most ε for any $\varepsilon > 0$. This can be understood by noting that the maximal forward degree of the two smallest-weight graphs grows like $m_n^{1/(\tau-2)}$, which is quite large. We can perform a *percolation argument*, by only considering paths that use edges with edge-weight at most ε . We discuss such percolation models in more detail in the next chapter. For now, it suffices to mention that such a percolated configuration model is again a configuration model with slightly smaller degrees. When $\varepsilon > 0$, we keep a positive proportion of the edges of high-degree vertices, so that even after percolation, the maximal degree is still of the order $m_n^{1/(\tau-2)}$. From such a vertex, we can reach the vertex with maximal percolated degree (the degree of which is of order $n^{1/(\tau-1)}$) in a *bounded* number of steps (recall the argument right below Lemma 2.31). Thus, its weight will be at most a bounded multiple times ε , as required. This explains how Theorem 3.32 is proved.

In the conservative case, we do not know how the weight-distance \mathcal{C}_n scales precisely:

OPEN PROBLEM 3.2 (Determine the scaling of the weight-distance in conservative setting). *How does the weight-distance \mathcal{C}_n between two uniformly chosen vertices scale for infinite-variance configuration models with conservative edge weights?*

Amini, Devroye, Griffiths and Olver [18, Page 1871], combined with the ideas above, do prove the lower bound that, whp,

$$(3.6.19) \quad \mathcal{C}_n \geq 2 \sum_{k=1}^n F_Y^{-1}(e^{-2^k}).$$

The power 2 is irrelevant here, as any power gives the same asymptotics in n . Also, we can replace n by any power n^a with $a > 0$, at the expense of a small correction. The lower bound in (3.6.19) follows directly from (3.6.18). We expect that (3.6.19) is sharp up to leading order (for which a proof is so far missing). Open Problem 3.2 is more about the random *fluctuations* in \mathcal{C}_n in the conservative setting.

Theorem 3.31 contains results about the weight-distance \mathcal{C}_n , but not on the number of edges in the optimal path H_n . In general, we do not know how the number of edges in the optimal path H_n scales for the configuration model with infinite-variance degrees:

OPEN PROBLEM 3.3 (Determine the scaling of the number of edges in optimal paths). *How does the number of edges H_n in the smallest-weight path between two uniformly chosen vertices scale for configuration models with infinite-variance degrees? When does H_n satisfy a central limit theorem?*

Note that we lack results both in the explosive as well as in the conservative settings. The only example where we do know how H_n scales is the case of exponential edge-weights, which is explosive. Indeed, by (3.6.13) in Theorem 3.31, we know that H_n scales like $\log n$ with variance also of order $\log n$ for exponential edge weights. Does this extend to general edge weights with a positive density at 0? Since optimal paths in this case mainly use edges with rather small edge weights (as this corresponds to the explosive setting), one might believe that indeed a central limit theorem holds for such edge weight. For other edge-weights, on the other hand, we do not know.

With Baroni and Komjáthy [29], we also study a particular example of conservative continuous-time branching processes, namely, a case where $Y = 1 + X$ for some non-negative random variable X for which $\text{infsupp}(X) = 0$ (recall Exercise 3.37). There, it is also shown that

$$(3.6.20) \quad \frac{\mathcal{C}_n}{\log \log n} \xrightarrow{\mathbb{P}} \frac{2}{|\log(\tau - 2)|},$$

which is quite surprising. Indeed, apparently there are so many paths containing roughly $2 \log \log n / |\log(\tau - 2)|$ edges, as for the shortest path, that one of these paths is such that it also has a small weight. We can thus view results as in (3.6.20) as quantifying the proliferation of shortest paths. We next investigate this case further, and also investigate the number of edges in the smallest-weight path H_n . Recall that both Theorem 3.31 and Theorem 3.21 show central limit theorems for the number of edges in optimal paths. We now discuss that such a central limit theorem does *not* hold for $Y = 1 + X$.

3.6.4. Scale-free configuration models conservative edge-weights: an example. We understand the setting of conservative edge weights much less well. As a result, we follow an *example-based approach*, where we investigate certain special cases. We already briefly discussed the setting where $Y = 1 + X$ and $\text{infsupp}(X) = 0$ in [29]. With Baroni and Komjáthy, we have taken his problem up again in [28], where a rather precise scaling result was proved as summarized in the following theorem:

THEOREM 3.33 (Weight-distances in conservative configuration model with $\tau \in (2, 3)$). *Consider the configuration model $\text{CM}_n(\mathbf{D})$ with i.i.d. degrees having distribution F_D satisfying (3.6.12) for some $\tau \in (2, 3)$ and $0 < c_1 \leq c_2 < \infty$, and that $\mathbb{P}(D = 1) = 0$. Let the edge distribution be given by $Y = 1 + X$, where X has a continuous distribution on $[0, \infty)$ for which $\text{infsupp}(X) = 0$.*

- (a) *When the age-dependent branching process with life-time distribution X and offspring distribution $D^* - 1$ is explosive,*

$$(3.6.21) \quad \mathcal{C}_n - \frac{2 \log \log n}{|\log(\tau - 2)|}$$

is a tight sequence of random variables.

- (a) *When, on the other hand, the age-dependent branching process with life-time distribution X and offspring distribution $D^* - 1$ is conservative,*

$$(3.6.22) \quad \mathcal{C}_n - \frac{2 \log \log n}{|\log(\tau - 2)|} \xrightarrow{\mathbb{P}} \infty.$$

The conservative result is quite easy to see. Indeed, denoting \mathcal{C}_n^x for the weight-distance between U_1 and U_2 when the edge-weights are i.i.d. from the distribution X , it

holds (see Exercise 3.38 below)

$$(3.6.23) \quad \mathcal{C}_n^Y = \mathcal{C}_n^{1+X} \geq \mathcal{C}_n^1 + \mathcal{C}_n^X = \text{dist}_{\text{CM}_n(d)}(U_1, U_2) + \mathcal{C}_n^X.$$

EXERCISE 3.38 (Edge-weights that are sums). *Show that $\mathcal{C}_n^{X+Y} \geq \mathcal{C}_n^X + \mathcal{C}_n^Y$ when X, Y are non-negative edge weights.*

Recall Theorem 2.30, and the discussion below it about the recent work with Komjáthy [170]. The latter proves that $\text{dist}_{\text{CM}_n(d)}(U_1, U_2) - 2 \log \log n / |\log(\tau - 2)|$ is tight, and, by assumption, $\mathcal{C}_n^X \xrightarrow{\mathbb{P}} \infty$. This shows (3.6.22).

The statement in (3.6.21) is much more subtle. It heavily relies on the precise condition for explosion of age-dependent branching processes in Theorem 3.30 (recall also (3.6.11)). For an edge e consisting of half-edges x and y , we bound $Y_e = 1 + X_e \leq 1 + X_x + X_y$, where $X_x = X_e$ and X_y is an independent copy of X . Thus, we can think of each edge consisting of 1 plus edge-weights along each of the half-edges. The argument makes crucial use of a *degree-dependent percolation* argument. For a half-edge x incident to a vertex of degree d , we keep the half-edge when $X_x \leq \xi(d)$, where the threshold $\xi(d)$ is chosen appropriately. We choose $\xi(d)$ such that the degree distribution *after removal of half-edges x with edge-weight $X_x \leq \xi(d)$ incident to vertices of degree d* , still satisfies (3.6.12) with the same τ . We know that there exists a nearly shortest-path for which the degree of the k th vertex is roughly $(\tau - 2)^{-k}$. This informally suggests that an upper bound of the form

$$(3.6.24) \quad \mathcal{C}_n^Y \leq \text{dist}_{G'}(U_1, U_2) + 2 \sum_k \xi((\tau - 2)^{-k})$$

holds, where G' is the degree-percolated graph. Since the degree distribution of this graph still satisfies (3.6.12) with the same τ , again $\text{dist}_{G'}(U_1, U_2) - 2 \log \log n / |\log(\tau - 2)|$ is tight. When (3.6.21) holds, we can also choose $\xi(d)$ such that $\xi((\tau - 2)^{-k})$ is summable. Together, these two ingredients explain (3.6.21). We refer to [29] for more details.

3.7. Further results for first-passage percolation

In this section, we review some further results for first-passage percolation on the complete graph as well as on random graphs. We will heuristically explain the results, but will not give their complete proofs. In Section 3.7.1 we discuss further results for first-passage percolation on the complete graph, and in Section 3.7.2 we discuss further results for first-passage percolation on random graphs.

3.7.1. Related results for weighted structures on the complete graph.

In this section, we discuss a number of extensions of the results in Section 3.2. We start by discussing what is known when the edge-weight distributions are not exponential.

The complete graph with positive powers of exponential edge weights. We start by studying a particular example of first-passage percolation on the complete graph with edge-weights having a distribution *different* from the exponential distribution. We work on the complete graph with vertex set $[n] \equiv \{1, \dots, n\}$ and edge set $E_n = \{ij : i, j \in [n], i \neq j\}$. Each edge e is given weight $l_e = (E_e)^s$ for some fixed $s > 0$, where $(E_e)_{e \in E_n}$ are i.i.d. exponential random variables with mean 1. For a fixed $s \in \mathbb{R}^+$, we are interested in statistics of the optimal path, in particular, in the asymptotics for the weight and hopcount of the optimal path as $n \rightarrow \infty$. When $s = 1$, we retrieve the exponential

weights studied in great detail before, and our interest lies in investigating whether the change in edge-weight distribution changes the minimal-weight topology of the graph.

To state the results, we start by setting up some notation. Let $(E_j)_{j \geq 1}$ be i.i.d. mean 1 exponential random variables. Define the random variables Y_i by the equation

$$(3.7.1) \quad P_i = \left(\sum_{j=1}^i E_j \right)^s.$$

Let PP be the above point process, i.e.,

$$(3.7.2) \quad \text{PP} = (P_1, P_2, \dots).$$

Now consider the continuous-time branching process where at time $t = 0$ we start with one vertex (called the root or the original ancestor), each vertex v lives forever, and has an offspring distribution $\text{PP}^{(v)} \sim \text{PP}$ as in (3.7.2) independent of every other vertex. Let $(\text{BP}_t)_{t \geq 0}$ denote the continuous-time branching process with the above offspring distribution.

Let us first investigate some of the properties of this continuous-time branching process. Let $\lambda = \lambda(s)$ be the *Malthusian parameter* in Definition 3.14, which determines the rate of exponential growth of this model. Indeed, by Theorem 3.15, there exists a strictly positive random variable W such that

$$(3.7.3) \quad e^{-\lambda t} |\text{BP}(t)| \xrightarrow{a.s.} W.$$

The constant λ satisfies the equation

$$(3.7.4) \quad \mathbb{E} \left[\int_0^\infty e^{-\lambda t} \text{PP}(dt) \right] = \sum_{i=1}^\infty \mathbb{E}(e^{-\lambda P_i}) = 1.$$

In this case, since $P_i^{1/s}$ has a Gamma-distribution with parameter i , we can explicitly compute that

$$(3.7.5) \quad \begin{aligned} \sum_{i=1}^\infty \mathbb{E}(e^{-\lambda P_i}) &= \sum_{i=1}^\infty \int_0^\infty \frac{y^{i-1}}{(i-1)!} e^{-y} e^{-\lambda y^s} dy \\ &= \int_0^\infty e^{-\lambda y^s} \left(\sum_{i=1}^\infty \frac{y^{i-1}}{(i-1)!} e^{-y} \right) dy \\ &= \int_0^\infty e^{-\lambda y^s} dy = \lambda^{-1/s} \Gamma(1 + 1/s), \end{aligned}$$

so that

$$(3.7.6) \quad \lambda = \lambda(s) = \Gamma(1 + 1/s)^s.$$

In the next theorem, we investigate H_n and \mathcal{C}_n in this setting:

THEOREM 3.34 (The weight of and edges in the smallest-weight path). *Consider first-passage percolation on the complete graph with i.i.d. edge-weights with distribution E^s , where E has an exponential distribution with mean 1 and $s > 0$. Then, as $n \rightarrow \infty$,*

$$(3.7.7) \quad \left(\frac{H_n - s \log n}{\sqrt{s^2 \log n}}, n^s \mathcal{C}_n - \frac{1}{\lambda} \log n \right) \xrightarrow{d} (Z, Q),$$

where Z has a standard normal distribution, and

$$(3.7.8) \quad Q \stackrel{d}{=} \frac{1}{\lambda} (-\log W^{(1)} - \log W^{(2)} - \Lambda - \log(1/s)),$$

where Λ is a standard Gumbel random variable independent of $W^{(1)}$ and $W^{(2)}$, and $W^{(1)}$ and $W^{(2)}$ are two independent copies of the random variable W appearing in (3.7.3).

Theorem 3.34 shows that the behavior of H_n and \mathcal{C}_n on the complete graph is quite universal. Indeed, H_n always satisfies a central limit theorem with mean and variance proportional to $\log n$. Further, let $u_n = F_Y^{-1}(1/n)$ denote the typical size of a minimal edge-weight from a vertex, then \mathcal{C}_n/u_n is of order $\log n$ with proportionality constant $1/\lambda$, and λ the Matusian parameter of the corresponding continuous-time branching process approximation of local neighborhoods, and the difference $\mathcal{C}_n/u_n - (1/\lambda) \log n$ converges in distribution. The limiting random variable even has the same shape as for first-passage percolation on the configuration model in (3.5.7) in Theorem 3.21. These results suggest that if a continuous-time branching process approximation of the local neighborhood exploration holds, then the limiting distributions of the hopcount and weight of the smallest-weight path behave rather universally.

The proof of Theorem 3.34 relies on a continuous-time branching process approximation of the neighborhoods of the starting and end vertices U_1 and U_2 . Similar ideas play a role as in the proof of Theorem 3.21.

The complete graph with negative powers of exponential edge-weights. It may seem from the above analysis that first-passage percolation on the complete graph *always* leads to a continuous-time branching process problem. We next study an example where this is not the case. We study the complete graph with i.i.d. edge weights E_{ij}^{-s} , $1 \leq i \leq j \leq n$. Thus, compared to the discussion in the previous section, we have replaced s there by $-s$ here, and we study the $s > 0$ regime. For fixed $s > 0$, define the function

$$(3.7.9) \quad g_s(x) = \frac{x^{s+1}}{(x-1)^s}, \quad x \geq 2.$$

Observe that, for $0 < s \leq 1$, the function $g_s(x)$, $x \geq 2$, is increasing, while for $s > 1$, the function is strictly convex with unique minimum at $x = s + 1$. We are interested in minimizing this function only on the space \mathbb{Z}_+ , the set of positive integers. Then, there is a sequence of values $s = s_j$, $j \geq 2$, for which the minimum integer of g_s is *not* unique. From the equation $g_s(j) = g_s(j+1)$ and the bounds $j-1 < s < j$, it is not hard to verify that

$$(3.7.10) \quad s_j = \frac{\log(1+j^{-1})}{\log(1+(j^2-1)^{-1})} \in (j-1, j), \quad j = 2, 3, \dots$$

We will need to deal with these special points separately. When $s \notin \{s_2, \dots\}$, then there is a unique integer which minimizes the function $g_s(x)$ on \mathbb{Z}_+ . Let $\mathcal{S} = \{s_2, \dots\}$.

THEOREM 3.35 (Hopcount and weight asymptotics). *Consider first-passage percolation on the complete graph with i.i.d. edge weights with distribution E^{-s} , where E has an exponential distribution with mean 1 and $s > 0$.*

For $s \notin \mathcal{S}$, let $k = k^(s) \in \{\lfloor s+1 \rfloor, \lceil s+1 \rceil\}$ denote the unique integer that minimizes the function defined in (3.7.9). Then, the hopcount $H_n = H_n(s) \xrightarrow{\mathbb{P}} k^*(s)$ as $n \rightarrow \infty$,*

and the optimal weight W_n , properly normalized, converges in distribution to a Gumbel distribution. More precisely, as $n \rightarrow \infty$, there exist constants b_k such that

$$(3.7.11) \quad \mathbb{P} \left(\frac{k-1}{s g_s(k)} (\log n)^{s+1} \left(W_n - \frac{g_s(k)}{(\log n)^s} \right) + \frac{k-1}{2} \log \log n + b_k > t \right) \rightarrow e^{-e^t}.$$

When $s \in \mathcal{S}$, $\mathbb{P}(H_n \in \{\lceil s+1 \rceil, \lfloor s+1 \rfloor\}) \rightarrow 1$ and again a linear transformation of \mathcal{C}_n converges in distribution to a Gumbel random variable.

Theorem 3.35 states that the hopcount H_n converges to the optimal value of the function $x \mapsto g_s(x)$ defined in (3.7.9), while the rescaled and re-centered minimal weight \mathcal{C}_n converges in distribution to a Gumbel distribution. We can intuitively understand this as follows. For fixed k , the minimal path of length k is similar to an *independent* minimum of copies of sums of k random variables E^{-s} . The number of independent copies is equal to the number of disjoint paths between vertices 1 and n , which is close to n^{k-1} . While on the complete graph, the appearing paths do *not* have independent weights, the paths that are particularly short are almost independent, which forms the crux of the proof. Due to the fact that the random variables are close to independent, extreme value ideas can be applied.

The following series of exercises shows that H_n is tight:

EXERCISE 3.39 (Minimum of independent inverse powers of exponentials). *Let $(E_i)_{i \in [n]}$ be a sequence of i.i.d. exponential random variables with parameter 1. Show that*

$$(3.7.12) \quad (\log n)^s \min_{i \in [n]} E_i^{-s} \xrightarrow{\mathbb{P}} 1.$$

Conclude that the minimal edge weight Y_{\min} for first-passage percolation on the random graph with i.i.d. edge weights with distribution E^{-s} satisfies $(\log n)^s Y_{\min} \xrightarrow{\mathbb{P}} 2^{-s}$.

EXERCISE 3.40 (Minimum of independent sums of inverse powers of exponentials). *Let $Y_i = E_{1i}^{-s} + E_{i2}^{-s}$ denote the edge weight of the two-step path between vertices 1 and 2 through the vertex i . Show that $(Y_i)_{i=3}^n$ are i.i.d. Further, show that $(\log n)^s \min_{i=3}^n Y_i \xrightarrow{\mathbb{P}} 2^{s+1}$.*

EXERCISE 3.41 (Hopcount for inverse powers of exponentials). *Use Exercises 3.39 and 3.40 to show that $\mathbb{P}(H_n \leq 2^{2s+1}) \rightarrow 1$.*

Similar results as in Theorem 3.35 were proved in work with Eckhoff, Goodman and Nardi [117], but now in a setting where the edge-weights satisfy are equal to $E_{ij}^{s_n}$ for some $s_n \rightarrow 0$. Theorem 3.35 makes it clear that there is not a unique universality class for first-passage percolation on the complete graph. In the following paragraph, we make this picture more precise and state further open problems for this.

Heavy-tailed edge-weights and relation to Poisson-weighted infinite tree.

In recent work with Eckhoff, Goodman and Nardi [118, 119], we investigated the problem of universality for first-passage percolation on the complete graph in a more fundamental way. Let us start by introducing some notation.

Suppose that the i.i.d. edge weights Y_e have distribution function F_Y with no atoms. Let $u \mapsto F_Y^{-1}(u)$ denote the inverse of the distribution function F_Y . The main assumption made in [118, 119] is that $u \mapsto u(F_Y^{-1})'(u)/F_Y^{-1}(u)$ is regularly varying with index $-\alpha$ as

$u \nearrow 0$, i.e.,

$$(3.7.13) \quad \frac{u(F_Y^{-1})'(u)}{F_Y^{-1}(u)} = u^{-\alpha}L(1/u) \quad \text{for all } u \in (0, 1),$$

where $t \mapsto L(t)$ is slowly varying as $t \rightarrow \infty$. We can interpret $u(F_Y^{-1})'(u)/F_Y^{-1}(u)$ as the derivative of $\log(F_Y^{-1}(u))$ with respect to $\log u$. Further, let $u_n = F_Y^{-1}(1/n)$. For such edge-weights with Eckhoff, Goodman and Nardi [118, 119] we prove the following result:

THEOREM 3.36 (Weight and hopcount – regularly-varying logarithmic derivatives). *Suppose that the edge weights $(Y_e)_{e \in E(K_n)}$ follow an n -independent distribution F_Y that satisfies (3.7.13). Define*

$$(3.7.14) \quad s_n = n^\alpha L(n)$$

and assume that $s_n/\log \log n \rightarrow \infty$ and $s_n = o(n^{1/3})$. Then there exist sequences $(\lambda_n)_n$ and $(\phi_n)_n$ with $\phi_n/s_n \rightarrow 1$, $\lambda_n u_n \rightarrow e^{-\gamma}$, where γ is Euler's constant, such that

$$(3.7.15) \quad nF_Y\left(C_n - \frac{1}{\lambda_n} \log(n/s_n^3)\right) \xrightarrow{d} M^{(1)} \vee M^{(2)},$$

$$(3.7.16) \quad \frac{H_n - \phi_n \log(n/s_n^3)}{\sqrt{s_n^2 \log(n/s_n^3)}} \xrightarrow{d} Z.$$

Here Z is standard normal, and $M^{(1)}, M^{(2)}$ are i.i.d. random variables for which $\mathbb{P}(M^{(j)} \leq x)$ is the survival probability of a Poisson Galton–Watson branching process with mean x .

We realize that the condition in (3.7.13) is a little hard to parse. In the following exercise, an example is discussed:

EXERCISE 3.42 (Example of Theorem 3.36). *Let $\rho > 0, \alpha \in (0, 1)$. Take $Y_e \stackrel{d}{=} \exp(-\rho e^{\alpha E}/\alpha)$, i.e., $F_Y(y) = \exp(-\frac{1}{\alpha} \log(\frac{\alpha}{\rho} \log(1/y)))$. Show that (3.7.13) holds and that s_n in (3.7.14) satisfies $s_n = \rho n^\alpha$.*

Exercise 3.42 highlights what kind of weights we should think about in Theorem 3.36. These are the weights that vary substantially, in the sense that the minimal weight incident to a vertex is *much* smaller than the second-smallest weight incident to it. Theorem 3.36 is proved in [119], while [118] shows that the local neighborhoods of vertices weakly converge to invasion percolation on the Poisson-weighted infinite tree. In invasion percolation, we add edges one by one according by taking the edge with minimal edge-weight incident to the current vertices. This is Prim's algorithm for the minimal-spanning tree and is discussed in more detail in Section 4.8.6 due to its close relation to percolation on the complete graph.

In [119], we further conjecture that the universality classes of first-passage percolation on the complete graph are precisely given by settings as in (3.7.13). Further, it is proved that the setting where $Y_e \stackrel{d}{=} E^{s_n}$ behave the same, at least when $s_n = o(n^{1/3})$. Note, in this respect, the similarity between the scaling behavior for H_n in Theorem 3.36 (where $s_n \rightarrow \infty$) and that in Theorem 3.34 (where $s_n = s$ is fixed instead). Indeed, when $Y_e \stackrel{d}{=} E^{s_n}$, it is proved that we can take $\phi_n = s_n$. The cases where $s_n \rightarrow 0$ should correspond to those studied in [117], fixed s to those in Theorem 3.34, while $s_n \rightarrow \infty$ with

$s_n = o(n^{1/3})$ to Theorem 3.36. When $s_n \gg n^{1/3}$, it is argued that the problem is even closer to the minimal-spanning tree, but rigorous results are missing:

OPEN PROBLEM 3.4 (Universality classes of first-passage percolation on the complete graph). *Determine the universality classes of first-passage percolation on the complete graph by completing the program in [119].*

Multicast and Steiner trees on the complete graph. We next discuss properties of the minimal-weight tree of a single vertex. For $m \geq 1$, let SWG_m denote the minimal-weight tree between vertex 1 and m other vertices in $[n] \setminus 1$ chosen uniformly at random conditioned to be distinct. Let $\mathcal{C}_n(m)$ denote its weight, and $H_n(m)$ denote the number of edges in it. Here we focus on aspects of $\mathcal{C}_n(m)$. With Hooghiemstra and Van Mieghem [163], we have identified the mean of the weight of the multicast tree as

$$(3.7.17) \quad \mathbb{E}[\mathcal{C}_n(m)] = \sum_{j=1}^m \frac{1}{n-j} \sum_{k=j}^{n-1} \frac{1}{k}.$$

In particular,

$$(3.7.18) \quad \mathbb{E}[\mathcal{C}_n(n-1)] = \sum_{k=1}^{n-1} \frac{1}{k^2}.$$

In a follow-up work [164], we extended the result is extended to a central limit theorem of the form

$$(3.7.19) \quad \sqrt{n}(\mathcal{C}_n(n-1) - \zeta(2)) \xrightarrow{d} \sigma Z,$$

where $\sigma^2 = 4\zeta(3) \simeq 4.80823$. A crucial ingredient in this analysis is the computation of the variance of $\mathcal{C}_n(n-1)$, which is explicitly given by

$$(3.7.20) \quad \text{Var}(\mathcal{C}_n(n-1)) = \frac{4}{n} \sum_{k=1}^{n-1} \frac{1}{k^3} + 4 \sum_{j=1}^{n-1} \frac{1}{j^3} \sum_{k=1}^j \frac{1}{k} - 5 \sum_{j=1}^{n-1} \frac{1}{j^4}.$$

The fact that $n\text{Var}(\mathcal{C}_n(n-1)) \rightarrow 4\zeta(3)$ is non-trivial and amounts to carefully compute the arising sums.

The Steiner tree problem is the task of finding a minimum weight subtree containing all these vertices, where the weight of a tree is the sum of the weights of the edges it contains. The Steiner tree of size 2 is the smallest-weight path between the two vertices. For $m \geq 3$, however, the Steiner tree can be different from the multicast tree, since the Steiner tree minimizes the weight of the tree and is thus not necessarily equal to the union of smallest-weight paths. Let $\mathcal{C}_n^{(\text{Ste})}(m)$ denote the weight of the Steiner tree of size m on the complete graph with exponential edge weights. Bollobás, Gamarnik, Riordan and Sudakov [54], show that, with $w_n(m) = (m-1)(\log(n/m))/n$

$$(3.7.21) \quad \frac{1}{w_n(m)} \mathcal{C}_n^{(\text{Ste})}(m) \xrightarrow{\mathbb{P}} 1.$$

With Hooghiemstra and Van Mieghem [163], we show the following analogous result for the multicast tree, namely, for all $m = O(n^a)$ with $a < 1$,

$$(3.7.22) \quad \mathbb{E}[\mathcal{C}_n(m)] = \frac{m}{n} \log \frac{n}{m+1} + O(n^{2(a-1)} \log n).$$

EXERCISE 3.43 (Comparison Steiner and multicast trees). *Show that (3.7.21) and (3.7.22) imply that also*

$$(3.7.23) \quad \frac{1}{w_n(m)} W_n(m) \xrightarrow{\mathbb{P}} 1.$$

Remco: Discuss recent results by Davidson!

3.7.2. Related results for first-passage percolation on random graphs. In this section, we discuss related results for first-passage percolation on random graphs. We start by reviewing results on extremal functionals for first-passage percolation on random graphs with exponential edge weights.

Extremal functionals for configuration model with exponential edge-weights.

We investigate extremal functionals on weighted configuration models. We study the limits of the weight-flooding and -diameter, as well as the hopcount-flooding and -diameter in the case of random regular graphs. We start with the weight-flooding and -diameter.

In order to state our main result, we need the following notation. Recall from (2.2.16) that $p_k^* = \mathbb{P}(D^* - 1 = k)$ denotes the asymptotic forward degree distribution in $\text{CM}_n(\mathbf{d})$. Let η^* denote the extinction probability of the branching process with offspring distribution $D^* - 1$, and recall that $\eta^* < 1$ precisely when $\mathbb{E}[D^* - 1] = \nu > 1$. Let $\mu^* = \mathbb{E}[(D^* - 1)(\eta^*)^{D^* - 2}] = G'_{D^* - 1}(\eta^*)$. The crucial object to describe the weight-flooding and -diameter in $\text{CM}_n(\mathbf{d})$ is the following quantity:

$$(3.7.24) \quad \Lambda(d) = d \mathbb{1}_{\{d \geq 3\}} + 2(1 - q_1) \mathbb{1}_{\{d=2\}} + (1 - \mu^*) \mathbb{1}_{\{d=1\}}.$$

Recall that $d_{\min} = \min_{v \in [n]} d_v$ is the minimal degree. In terms of $\Gamma(d)$, the weight-flooding and -diameter in $\text{CM}_n(\mathbf{d})$ can be quantified as follows:

THEOREM 3.37 (Weight-diameter and flooding on CM). *Let $\text{CM}_n(\mathbf{d})$ satisfy Condition 1.6(a)-(c), assume that $p_{d_{\min}} > 0$ and that $\limsup_{n \rightarrow \infty} \mathbb{E}[D_n^{2+\varepsilon}] < \infty$ for some $\varepsilon > 0$. Equip $\text{CM}_n(\mathbf{d})$ with i.i.d. exponential edge weights with mean 1. Then, as $n \rightarrow \infty$,*

$$(3.7.25) \quad \frac{\max_j \mathcal{C}_n(1, j)}{\log n} \xrightarrow{\mathbb{P}} \frac{1}{\nu - 1} + \frac{1}{\Lambda(d_{\min})},$$

and

$$(3.7.26) \quad \frac{\max_{i,j} \mathcal{C}_n(i, j)}{\log n} \xrightarrow{\mathbb{P}} \frac{1}{\nu - 1} + \frac{2}{\Lambda(d_{\min})},$$

where $\Lambda(d)$ is defined in (3.7.24).

Theorem 3.37 is proved by Amini [16], see also Amini, Draief and Lelarge [14].

Theorem 3.37 can be best understood when $d_{\min} \geq 3$. In this case, $\Lambda(d_{\min}) = d_{\min}$, and we see that the flooding is $1/d_{\min}$ larger than the typical weight as described in Theorem 3.7, while the diameter is $2/d_{\min}$ larger than the typical weight. This is similar to the situation for the complete graph in Theorem 3.1. Indeed, for $i \in [n]$, let $e \in E(\text{CM}_n(\mathbf{d}))$

denote the edges in $\text{CM}_n(\mathbf{d})$ and define $X_i = \min_{j: ij \in E(\text{CM}_n(\mathbf{d}))} E_{ij}$, where $(E_{ij})_{ij \in E(\text{CM}_n(\mathbf{d}))}$ are the exponential edge weights with mean 1. Then, for every $t > 0$,

$$(3.7.27) \quad \mathbb{P}(X_i \geq t) = e^{-td_i},$$

so that, for every $a > 0$,

$$(3.7.28) \quad \mathbb{P}(X_i \geq a \log n) = n^{-ad_i}.$$

Therefore, the expected number of vertices i of degree d with $X_i \geq a \log n$ is equal to $n_d n^{-ad}$. As a result, for every d such that $p_d > 0$, whp there exist vertices i having degree d and minimal edge-weight $X_i \geq a \log n$ when $a < 1/d$, but not when $a > 1/d$. The largest possible a arises when $d = d_{\min}$ is as small as possible. We conclude that, under the conditions of the theorem, whp there exist vertices i such that $X_i \geq a \log n$ when $a < 1/d_{\min}$, but not when $a > 1/d_{\min}$. As a result, the weight-flooding is achieved by the minimal-weight between one of such vertices and vertex 1, while the weight-diameter is achieved by the minimal-weight between two of such vertices.

When $d_{\min} \leq 2$, the above arguments still apply. However, in this case, paths with even larger weight can be found. Indeed, we note that $\Lambda(d) < d$ when $d = 1, 2$. For $d_{\min} = 2$, such paths can be realized by long thin paths. In order to prove Theorem 3.37 in this case, it is instructive to consider a Markov branching process starting from k individuals having offspring distribution Δ . Denote the time it takes the continuous-time branching process to reach n vertices by T_n^k . Then,

$$(3.7.29) \quad - \lim_{n \rightarrow \infty} \frac{1}{\log n} \log \mathbb{P}\left(\left(\frac{1}{\mathbb{E}[\Delta]} + x\right) T_n^k < \infty\right) = xg(\Delta_{\min}, k),$$

where $B_{\min} = \min\{k: \mathbb{P}(\Delta = k) > 0\}$. Note that $\Lambda(d_{\min}) = g((D^* - 1)_{\min}, d_{\min})$.

It would be of great interest to also investigate the *fluctuations* of these weight-extremal functionals:

OPEN PROBLEM 3.5 (Fluctuations weight-flooding and -diameter).
Determine the fluctuations of the weight-flooding and -diameter in Theorem 3.37. How are these related to those for the complete graph as determined in Theorems 3.2 and 3.5?

We next study the maximal hopcount, both from a fixed vertex as well as between all pairs in the graph. We restrict our attention to the random d -regular graph. We define the function $f: \mathbb{R}_+ \mapsto \mathbb{R}$ by

$$(3.7.30) \quad f(\alpha) = \alpha \log \left(\frac{d-2}{d-1} \alpha \right) - \alpha + \frac{1}{d-2}.$$

THEOREM 3.38 (Hopcount-diameter on random regular graphs). *Fix $d_i = d \geq 3$ for every $i \in [n]$. Consider the random-regular graph $\text{CM}_n(\mathbf{d})$ equipped with i.i.d. exponential edge weights with mean 1. Then, as $n \rightarrow \infty$,*

$$(3.7.31) \quad \frac{\max_j H_n(1, j)}{\log n} \xrightarrow{\mathbb{P}} \alpha,$$

and

$$(3.7.32) \quad \frac{\max_{i,j} H_n(i, j)}{\log n} \xrightarrow{\mathbb{P}} \alpha^*,$$

where α is the solution to $f(\alpha) = 0$ and α^* is the solution to $f(\alpha) = 1$.

Theorem 3.38 is proved by Amini and Peres [17].

Open problems for first-passage percolation on configuration models. So far, for extremal functionals, we have focussed on the exponential edge-weight case. It would be highly interesting to extend this to their extremal functionals:

OPEN PROBLEM 3.6 (Asymptotics weight-flooding and -diameter). *Determine the asymptotics of the weight-flooding and -diameter in Theorem 3.37 for general edge-weight distributions.*

The exponential distribution is rather special, as the scale of maximal minimal edge-weights (i.e., the maximal values of $X_i = \min_{j: ij \in E(\text{CM}_n(d))} E_{ij}$) are on the *same* scale as the weight of typical smallest paths. This is quite different when we consider edge-weight with a rather different density close to zero. A test-case could be edge-weights of the form $Y_e = E^s$ for some fixed s . Then, also the minimal edge-weights are transformed to X_i^s , where $X_i = \min_{j: ij \in E(\text{CM}_n(d))} E_{ij}$. This, the maximal minimal edge-weights are of the order $(\log n)^s$, which is much smaller than the typical weight when $s < 1$ and much larger than the typical weights when $s > 1$. In this setting, it is reasonable to assume that the weight-flooding and -diameter are of the form $a(\log n)^s + b \log n$ for some a and b , the fluctuations being of smaller order.

The behavior of the hopcount-diameter on random graphs is even harder:

OPEN PROBLEM 3.7 (Asymptotics hopcount-flooding and -diameter). *Determine the asymptotics of the hopcount-flooding and -diameter in Theorem 3.38 for general degree distributions. Can this be extended to general edge-weights?*

First-passage percolation on related geometries. There are many more results on first-passage percolation on random and non-random finite graphs, and here we list some of them. Kolossvary and Komjáthy [199] study first-passage percolation on a class of inhomogeneous random graphs, where the expected degrees are constant. The results are similar to those in Theorem 3.7. Komjáthy and Vadon [201] study first-passage percolation on the Newman-Strogatz small-world model.

A major open problem is to investigate first-passage percolation on preferential attachment models:

OPEN PROBLEM 3.8 (First-passage percolation on preferential attachment models). *Determine the asymptotics first-passage percolation on the preferential attachment model.*

Open Problem 3.8 has one aspect that is highly similar to that on other random graphs. Indeed, one can study the asymptotics of $\mathcal{C}_n = \mathcal{C}_n(U_1, U_2)$ when $n \rightarrow \infty$. This raises the question whether the results for the configuration model also hold in this case, which is unclear.

One of the exciting aspects of Open Problem 3.8 is that of the *dynamics* of smallest-weight paths. Indeed, fix two vertices i, j and investigate what the smallest-weight $\mathcal{C}_n(i, j)$ is. It is not hard to show that $\mathcal{C}_n(i, j) \xrightarrow{\mathbb{P}} 0$ for $\text{PA}_n^{(m, \delta)}$ when $\delta \in (-m, 0)$:

EXERCISE 3.44 (Vanishing weight-distance for $\delta < 0$). *Fix $m \geq 2$ and consider $\text{PA}_n^{(m, \delta)}$ when $\delta \in (-m, 0)$. Fix i, j . Show that $\mathcal{C}_n(i, j) \xrightarrow{\mathbb{P}} 0$ as $n \rightarrow \infty$.*

Clearly, $n \mapsto \mathcal{C}_n(i, j)$ is decreasing for all parameter values of m and δ . It can be expected that $\mathcal{C}_n(i, j)$ converges to a non-negative random variable when $\delta > 0$.

Remco: Add results for FPP on hypercube!

3.8. Notes and discussion of first-passage percolation on random graphs

Notes on Section 3.2. The presentation in Section 3.2 follows Janson [179]. The restrictions on the weight distribution Y_{ij} in [179] are slightly weaker than assumed here. Assume that Y_{ij} are non-negative and that their distribution function $\mathbb{P}(Y_{ij} \leq t) = t + o(t)$ as $t \searrow 0$, the main examples being the uniform $U(0, 1)$ and the exponential $\text{Exp}(1)$ distributions.

We see from Theorem 3.1(ii) and (iii) that $\max_{j \in [n]} \mathcal{C}_n(i, j)$ typically is about $2 \log n/n$, but that it is larger for a few vertices with $\max_j \mathcal{C}_n(i, j)$ being about $3 \log n/n$. A companion result by Janson [179] shows that, in contrast, $\max_{j \in [n]} \mathcal{C}_n(i, j)$ is not significantly smaller than $2 \log n/n$ for any vertex i :

THEOREM 3.39 (Minimal-maximal weight). *As $n \rightarrow \infty$,*

$$(3.8.1) \quad \frac{\min_{i \in [n]} \max_{j \in [n]} \mathcal{C}_n(i, j)}{\log n/n} \xrightarrow{\mathbb{P}} 2.$$

For the case of exponentially-distributed Y_{ij} , the proof of Theorem 3.1 shows that the collection of minimal weight paths from a given vertex, 1 say, form a random recursive tree (rooted at 1) which can be constructed as follows. Begin with a single root and add $n - 1$ vertices one by one, each time joining the new vertex to a (uniformly) randomly chosen old vertex. Distances between vertices in random recursive trees as in Theorem 3.3 have been studied by Moon [218], and the diameter by Pittel [229] as $n \rightarrow \infty$. See also the survey by Smythe and Mahmoud [240].

The first result in Theorem 3.6 is by Devroye [101] and Pittel [228], the second is the main result of Addario-Berry, Broutin and Lugosi [7]. In [7], it is proved further that $\max_{i, j} H_n(i, j) \leq \alpha^* \log n + \omega(n)$ whp for any $\omega(n) \rightarrow \infty$, while $\max_{i, j} H_n(i, j) \geq \alpha^* \log n + L \log \log n$ whp for some constant $L > 0$. The precise fluctuations of $\max_j H_n(1, j)$ and $\max_{i, j} H_n(i, j)$ are unknown.

Theorem 3.5 is proved with Bhamidi in [39] and solves [179, Problem 1]. Theorem 3.34 is proved with Bhamidi in [47], where, erroneously, $-\Lambda$ in (3.7.8) was replaced by Λ .

Notes on Section 3.3. Theorem 3.7 is proved in [40] under slightly more restrictive conditions on the degrees as the ones stated. Indeed, in [40], it is assumed that the degrees are i.i.d. random variables with a distribution D satisfying $\mathbb{P}(D \geq 2)$ a.s. The present version can be obtained from [42] under slightly weaker conditions. We follow the proof with Bhamidi and Hooghiemstra in [40].

Proposition 3.8 is proved first by Bühler [70]. We present the probabilistic proof with Bhamidi and Hooghiemstra in [40], since there is some confusion between the definition s_i given here, and the definition of s_i given in [70, below Equation (3.1)]. More precisely, in [70], s_i is defined as $s_i = d_1 + \dots + d_i - i$, which is our $s_i - 1$.

Notes on Section 3.4. A comprehensive study of continuous-time branching processes can be found in Harris [153], Jagers [177] and Athreya and Ney [23]. We follow Harris [153]. In particular, [153, Chapter V] deals with Markov branching processes, in which the evolution of $(|\text{BP}(t)|)_{t \geq 0}$ is Markovian. Examples are Bellman-Harris processes. [153, Chapter VI] deals with general age-dependent branching processes. The most general results on continuous-time branching processes can be found in [178]. The convergence in (3.4.31) is [153, Theorem 17.1]. Jagers and Nerman [178] have proved some of the strongest results for general continuous-time branching processes. In particular, Theorem 3.15 is [178, Theorem 5.3].

Notes on Section 3.5. Theorem 3.21 is proved in work with Bhamidi and Hooghiemstra [42].

Notes on Section 3.6. Theorem 3.40 is due to Grey [141]. He also gives an if and only if criterion for explosion, but we find the min-summability conditions simpler. We state the condition by Grey here. In its statement, we let $G_\Delta(s) = \mathbb{E}[s^\Delta]$ be the probability generating function of the offspring $\Delta = D^* - 1$:

THEOREM 3.40 (Finite explosion time age-dependent branching processes). *An age-dependent branching process with offspring Δ and life-time Y a.s. explodes in finite time precisely when $\mathbb{P}(\Delta \geq 1) = 1$ and the probability generating function G_Δ of the offspring distribution Δ satisfies that there exists $\delta > 0$ such that, for all $0 \leq t \leq \delta$,*

$$(3.8.2) \quad 1 - G_\Delta^{-1}(t) \leq \int_0^t G_Y(s) ds < \infty.$$

This result is proved in Grey [141]. Komjáthy [200] not only treats age-dependent processes, but also other forms of epidemic processes. She gives conditions for explosion for such epidemics in the presence of contagious intervals $[I, C]$ on every individual in the continuous-time branching process. The effect of this is that only the children with birth-times that fall in this interval of the parent are being kept. Komjáthy shows the surprising result that the distribution of the end C of the contagious interval does not matter in terms of explosion, while the distribution of I does. Indeed, in terms of I , the epidemic explodes if and only if the two age-dependent branching processes with the original birth-times and birth-times I explode. Related results can be found in the master thesis of Gulikers [148].

Theorem 3.31 is proved in work with Bhamidi and Hooghiemstra [41]. In [40], we further study first-passage percolation with exponential edge weights on $\text{CM}_n(\mathbf{D})$ with $\tau \in (1, 2)$, and it is shown that both \mathcal{C}_n and H_n converge in distribution. Interestingly, the distributional limits of \mathcal{C}_n and H_n depend on whether we work with the original

configuration model (where we keep all multiple edges) or the erased configuration model (where all multiple edges are merged into a single edge).

Notes on Section 3.7. It is yet unclear what the universality classes of first-passage percolation on the complete graph are (recall Open Problem 3.4). Recall Theorem 3.36, where the setting was described for some set of particular examples. In [119], there is a much more extensive discussion. There, it is proved that the behavior of first-passage percolation on the complete graph with such weights is closely related to that with weights equal to E^{s_n} , where E is exponential with mean 1 and s_n depends on n . Particularly the case where $s_n \rightarrow \infty$, sometimes also called the *strong disorder phase* (see Braunstejn et al. [67, 154], and Sreenivasan et al. [245]), is highly interesting. Remarkably, this setting can both arise from n -dependent weights as in E^{s_n} , as well as from n -independent weights. The case where $s_n \rightarrow 0$ is handled in [119], partial cases where $s_n \rightarrow \infty$ in [119]. In the latter case, a CLT is proved with mean $s_n \log(n/s_n^3)$ and variance $s_n^2 \log(n/s_n^3)$ as long as $s_n = o(n^{1/3})$. The strong disorder phase of first-passage percolation on the complete graph is closely related to the properties of the minimal spanning tree, as investigated in [5, 8, 135] (see Section 4.8.6). There, the graph distance between vertices grows like $n^{1/3}$, which suggests that first-passage percolation and the minimal spanning tree can be in the *same* universality class.

Theorem 3.37 is proved by Amini [16], see also Amini, Draief and Lelarge [14]. The result in (3.7.29), which is crucial in the understanding of Theorem 3.37, is proved by Amini and Lelarge [15]. The special case of d -random regular graphs with $d \geq 3$ is also studied by Ding, Kim, Lubetzky and Peres [105], where it was used to investigate the diameter of near-critical supercritical Erdős-Rényi random graphs. Theorem 3.38 is proved by Amini and Peres [17].

Percolation on random graphs

Abstract

In this chapter, we discuss percolation on random graphs. Percolation is a model for random damage to a network, and it is one of the simplest models that displays a phase transition: when the network is severely damaged, it falls apart in many small connected components, while if the damage is light, a giant component remains on existing. We study the location of the phase transition on random graphs. We further study the connectivity structure close to criticality, where components display intricate scaling behavior such that a typical cluster has a bounded size, while the average and maximal cluster sizes grow like powers of the network size.

4.1. Percolation on general graphs and motivation

In percolation, edges and/or vertices are randomly removed from a connected graph $G = (V, E)$. Often, percolation is studied on infinite graphs, such as the hypercubic lattice. See the books by Grimmett [144], Bollobás and Riordan [58] and Kesten [194] for general backgrounds on percolation. The beauty of the model is that percolation on infinite graphs is the simplest imaginable model that displays a *phase transition*. Suppose edges in the graph are removed independently with some probability $\pi \in [0, 1]$.¹ We think of the retention probability π as the proportion of edges that escapes being damaged. When π is small, the damage is substantial, and the graph falls apart into many small and finite clusters. In particular, while G had an infinite connected component, after percolation, for π small, no infinite connected structure remains. On the other hand, when the damage is very little and π is close to 1, then the connected graph remains an infinite structure even after the damage is inflicted. By monotonicity in the edge-retention probability, there is a *unique* π_c above which infinite structures remain, while below it, percolation with edge-retention probability π has no infinite components. We could interpret percolation as a special example of first-passage percolation as studied in Chapter 3, where $Y_e = 0$ with probability π , while $Y_e = +\infty$ with probability $1 - \pi$. Of course, this is not allowed in Chapter 3, and the fact that edge-weights can take values zero and infinity imply that we are directly lead to studying connectivity properties of the resulting graph instead.

¹Traditionally, p is used as the edge-retention probability. Due to the potential conflicts with other usages of p in this book, such as the degree distribution, we choose for π instead.

Many more questions have been investigated in the setting where G is infinite and transitive, meaning that all vertices are the same. Examples are the uniqueness of the infinite component, the scaling of large clusters close to the critical value π_c , the behavior of random walk on percolation clusters, etc. We refer to [58, 144, 194] for more background on percolation, and to the recent book with Heydenreich [157] for recent progress on percolation in the high-dimensional setting, which is the setting that is most closely related to ours. The simplest question is whether π_c is non-trivial, meaning that $\pi_c \in (0, 1)$. For the hypercubic lattice, this turns out to be true for $d \geq 2$:

EXERCISE 4.1 ($\pi_c = 1$ for \mathbb{Z}). *Show that $\pi_c = 1$ for percolation on the integers \mathbb{Z} and the nearest-neighbor edge set $E(\mathbb{Z}) = \{\{x, y\} : |x - y| = 1\}$.*

EXERCISE 4.2 ($\pi_c > 0$ for regular graphs). *Show that $\pi_c \geq 1/d$ for percolation on a transitive graph where each vertex has degree d .*

Motivation for percolation on finite graphs. Finite graphs are used as network models. Such networks can be affected by vertices and edges being down or attacked, rendering them out of service. This is sometimes called the *attack* or *failure vulnerability*. Of course, this process does not need to be deliberate, but could also arise from vertices or edges randomly malfunctioning. When we model the attack as being a random process, then the question is how much connectivity is preserved while vertices or edges are randomly knocked out. We will often assume that this is a truly random process, in the sense that the edges or vertices fail with equal probabilities independently of one another. Such an attack is sometimes called a *random attack*. It could also be that the attacker has some knowledge about the network, and can deliberately knock out specific vertices, in a *deliberate attack*. Whether these settings are highly different is of practical interest, as it quantifies the effect of a knowledgeable opponent. We discuss such problems from an abstract perspective.

Percolation on finite graphs. Let G be a finite graph with vertex set $[n]$. Let π be the edge-retention probability. For $u, v \in [n]$, we write that $u \longleftrightarrow v$ when there exists a path of edges that are retained, and let $\mathcal{C}_v(\pi) = \{u : u \longleftrightarrow v\}$ be the collection of all vertices that are connected to v . Percolation on finite graphs is inherently different from percolation on infinite transitive graphs is that a unique critical value, such as present on many infinite transitive graphs, cannot exist. Let \mathbb{P}_π denote the percolation measure on the finite graph G . Then, the dependence on the edge-retention probability π is smooth, and $\mathbb{P}_\pi(\mathcal{E})$ is a polynomial of finite degree for every event \mathcal{E} :

EXERCISE 4.3 (Continuity in the edge-retention probability on finite graphs). *Let \mathcal{E} be an event. Show that $\mathbb{P}_\pi(\mathcal{E})$ is a polynomial in π of degree at most $|E(G)|$.*

As a result, there cannot be a sharp phase transition, in the sense that if π is close to critical and δ very small, then also $\pi + \delta$ will be close to critical. Further, on a finite graph, no infinite components can exist. This means that we have to redefine the notion of a critical value. For this, let us start by introducing some notation.

A *local* way of quantifying the critical behavior is to take a *uniform* vertex U , and to investigate the component $\mathcal{C}_U(\pi)$ of that vertex after percolation with percolation parameter π . When $\limsup_{k \rightarrow \infty} \limsup_{n \rightarrow \infty} \mathbb{P}(|\mathcal{C}_U(\pi_n)| \geq k) = 0$, then the clusters in the percolated graph are small and we can think of $(\pi_n)_{n \geq 1}$ as being subcritical, while

if $\liminf_{k \rightarrow \infty} \liminf_{n \rightarrow \infty} \mathbb{P}(|\mathcal{C}_U(\pi_n)| \geq k) > 0$, the sequence $(\pi_n)_{n \geq 1}$ is supercritical. Of course, the boundary between these two settings may not be sharp.

Percolation on random graphs. In this chapter, we focus on percolation on random graphs, which poses an additional difficulty due to the fact that the graphs are *random*, and thus we deal with *double randomness* that is present throughout this book. As a result, we can think of the stochastic models on random graphs as processes in a disordered medium. Especially for percolation, this is somewhat paradoxical, since percolation often provides the random medium on which other processes, in particular random walks, act.

Let us now continue by studying the local definition of a critical value in the random graph setting. The random graphs that we are interested in are locally tree-like in the large n limit, and this limit does have a sharp phase transition. This allows us to identify the critical value, at least to first order, as the critical value of percolation on the random tree. The nice thing about this is that percolation on a branching process tree again gives rise to a branching process tree. In the context of random graphs that have a unimodular Galton-Watson tree as a local limit, as the configuration model and rank-1 random graphs (recall Theorems 2.11 and 2.14), it is not hard to see that the critical value of the limiting tree equals $\pi_c = 1/\nu$, where ν is the expected value of the forward degree. Indeed, for this value, the expected offspring *after percolation* equals 1, and thus the unimodular branching process is critical (recall Theorem 2.1). This suggests that the critical behavior of percolation on such random graphs occurs close to this value, and will constitute one of the main results in this chapter. We next discuss the critical behavior in some more detail, by highlighting some critical exponents for critical percolation on random graphs.

Critical exponents for percolation on random graphs. Let $(G_n)_{n \geq 1}$ be some graph sequence that converges in the local-weak convergence sense to some limit (G, o) . The sequence $(G_n)_{n \geq 1}$ could be a sequence of random graphs. Suppose that π_c is some (asymptotic) critical value of percolation on the limiting infinite graph (G, o) . In many situations (though certainly not in all, see Grimmett [144] for details), we know that at the critical value π_c , the percolated cluster of o after performing percolation on (G, o) is *finite* almost surely. This is closely related to key question of continuity of the percolation function in general percolation, which is one of the major questions in percolation theory, we also refer to the text with Heydenreich [157] for more details. One may then expect that, for k large, $\mathbb{P}(|\mathcal{C}_o(\pi_c)| > k)$ to decay when $k \rightarrow \infty$, but not too quickly. In general, this decay is believed to be characterized by a critical exponent $\delta \geq 1$ in the form that

$$(4.1.1) \quad \mathbb{P}(|\mathcal{C}_o(\pi_c)| > k) \sim k^{-1/\delta},$$

where \sim can denote any form of asymptotics. Let us explore this further in the context of percolation on the configuration model (the rank-1 inhomogeneous random graph setting being highly similar). There, we have already argued informally that $\pi_c = 1/\nu$. The local-weak limit (G, o) in this setting is a unimodular Galton-Watson tree. After percolation, the limit is still a unimodular Galton-Watson tree:

EXERCISE 4.4 (Percolation on a unimodular Galton-Watson tree). *Perform percolation with percolation threshold π on a unimodular Galton-Watson tree (G, o) with root offspring distribution $(p_k)_{k \geq 0}$. Show that the result is again a unimodular Galton-Watson*

tree (G, o) with root offspring distribution $(p_k(\pi))_{k \geq 0}$ given by

$$(4.1.2) \quad p_k(\pi) = \sum_{\ell \geq k} p_\ell \mathbb{P}(\text{Bin}(\ell, \pi) = k).$$

As a result, we may as well consider a *critical* unimodular Galton-Watson tree (G, o) , i.e., those unimodular Galton-Watson tree (G, o) for which $\nu = \sum_{k \geq 0} k p_k^* = 1$. Recall Theorem 2.3, where the total progeny of a branching process was investigated. For simplicity, we ignore the fact that the root has offspring distribution $(p_k)_{k \geq 0}$ rather than $(p_k^*)_{k \geq 0}$ (see also Exercise 4.5 below). Then, by the distribution of the total progeny derived in Theorem 2.3,

$$(4.1.3) \quad \mathbb{P}(|\mathcal{C}_o(\pi_c)| = k) = \frac{1}{k} \mathbb{P}(X_1 + \cdots + X_k = k - 1),$$

where $(X_i)_{i=1}^k$ are i.i.d. random variables with probability mass function $(p_k^*)_{k \geq 0}$. Note that X_i has mean 1. When X_i also has finite variance, then we can expect a local central limit theorem to be valid, so that²

$$(4.1.4) \quad \mathbb{P}(|\mathcal{C}_o(\pi_c)| = k) = \frac{1}{k} \mathbb{P}(X_1 + \cdots + X_k = k - 1) = \frac{1}{\sqrt{2\pi\sigma^2 k^3}} (1 + o(1)),$$

where $\sigma^2 = \text{Var}(X_i)$. A simple computation then shows that

$$(4.1.5) \quad \mathbb{P}(|\mathcal{C}_o(\pi_c)| > k) = \frac{c}{\sqrt{k}} (1 + o(1)),$$

so that $\delta = 2$. Note that X_i also has finite variance precisely when $(p_k)_{k \geq 0}$ has a finite third moment. When X_i does not have finite variance, for example when $\sum_{\ell > k} p_\ell^* \sim k^{-(\tau-2)}$, then a local limit theorem, now for asymptotically stable variable, gives

$$(4.1.6) \quad \mathbb{P}(|\mathcal{C}_o(\pi_c)| = k) = \frac{1}{k} \mathbb{P}(X_1 + \cdots + X_k = k - 1) = \frac{c'}{k^{1+1/(\tau-2)}} (1 + o(1)),$$

implying that

$$(4.1.7) \quad \mathbb{P}(|\mathcal{C}_o(\pi_c)| > k) = \frac{c}{k^{1/(\tau-2)}} (1 + o(1)),$$

so that $\delta = \tau - 2$. This shows that the power-law exponent τ plays a crucial role in the critical behavior of percolation on random graphs.

EXERCISE 4.5 (Critical exponent δ for a unimodular Galton-Watson tree). *Show that δ for a unimodular Galton-Watson tree takes on the same value as δ for a homogeneous branching process with offspring distribution $(p_k^*)_{k \geq 0}$.*

²For simplicity, we assume here that the distribution of X_1 is non-lattice. When it is lattice, which can only occur when $\mathbb{P}(X_1 \in \{0, 2\}) = 1$ in the critical case where $\mathbb{E}[X_1] = 1$) extra factors appears and some terms may be equal to zero.

The largest components in percolation on random graphs. We let $(|\mathcal{C}_{(i)}(\pi)|)_{i \geq 1}$ denote the ordered collection of connected components, where we break ties (if they exist) arbitrarily. Thus, in particular,

$$(4.1.8) \quad |\mathcal{C}_{\max}(\pi)| = |\mathcal{C}_{(1)}(\pi)| = \max_{i \geq 1} |\mathcal{C}_{(i)}(\pi)|$$

denotes the largest or maximal connected component. Often, we remove π from the notation when its value is clear. Note that the vector $(|\mathcal{C}_{(i)}(\pi)|)_{i \geq 1}$ has a *random* number of coordinates, which equals the number of disjoint clusters. Sometimes, it will be convenient to extend this finite vector to an infinite vector by appending infinitely many zeros at the end of the vector.

In the above, we have defined critical behavior in terms of the local limit of the connected component of a random vertex after percolation. This does not give us much information though about the largest connected component of the graph after percolation. This largest connected component is what describes the remaining connectivity structure of graphs after they are randomly destroyed. The difficulty of dealing with the largest connected component is that it is an inherently *global* quantity, making it a difficult object to get our hands upon. We now discuss the percolation critical value from this global perspective.

We say that $(\pi_n)_{n \geq 1}$ (possibly depending on the network size n) is *asymptotically supercritical* when there exists $\varepsilon > 0$ independent of n such that

$$(4.1.9) \quad \liminf_{n \rightarrow \infty} \mathbb{P}(|\mathcal{C}_{\max}(\pi_n)| \geq \varepsilon n) = 1,$$

and *asymptotically subcritical* otherwise. Thus, for an asymptotically supercritical sequence $(\pi_n)_{n \geq 1}$ the maximal component contains, with high probability, a positive proportion of the vertices. Clearly, this is a monotone definition in the sense that if $(\pi_n)_{n \geq 1}$ is an asymptotically supercritical sequence and $\pi'_n \geq \pi_n$ for every n , then also $(\pi'_n)_{n \geq 1}$ is an asymptotically supercritical sequence. Similarly, if $(\pi_n)_{n \geq 1}$ is an asymptotically subcritical sequence and $\pi'_n \leq \pi_n$ for every n , then also $(\pi'_n)_{n \geq 1}$ is an asymptotically subcritical sequence. This can be easily seen by a useful *coupling* of all percolation measures on the same probability space, which goes under the name of the *Harris coupling*. For this, let $(U_e)_{e \in E(G)}$ denote a sequence of i.i.d. uniform random variables. Then we declare the edge e π -retained when $U_e \leq \pi$. Clearly, the π -retained edges have distribution \mathbb{P}_π , while if an edge is π -retained and $\pi' \geq \pi$, then the edge is also π' -retained.

EXERCISE 4.6 (Monotonicity in retention probability). *Use the Harris coupling to show that if $(\pi_n)_{n \geq 1}$ is an asymptotically supercritical sequence and $\pi'_n \geq \pi_n$ for every n , then also $(\pi'_n)_{n \geq 1}$ is an asymptotically supercritical sequence.*

Critical value in large graph limits. In many cases, we see that

$$(4.1.10) \quad \frac{|\mathcal{C}_{\max}(\pi)|}{n} \xrightarrow{\mathbb{P}} \zeta(\pi),$$

for some $\zeta(\pi) \in [0, 1]$. Then, we can define the asymptotical percolation critical value also in terms of $\zeta(\pi)$ as

$$(4.1.11) \quad \pi_c = \inf\{\beta: \zeta(\beta) > 0\}.$$

In what follows, we will also be interested in finer properties of $|\mathcal{C}_{\max}(\pi)|$ for π close to the critical value π_c .

Critical behavior in large graph limits. While the notion of asymptotically supercritical sequences is useful to determine whether, after percolation, a giant component still exists, it is less useful to describe the critical behavior for percolation on finite graphs. Indeed, for critical sequences $(\pi_n)_{n \geq 1}$, we would expect that the scaling of largest percolation clusters is quite intricate. A large body of work has appeared recently in identifying the critical behavior of percolation on various random and non-random finite graphs. Typically, an explicit sequence $(\pi_n)_{n \geq 1}$ is then chosen, and the behavior of the largest percolation clusters determined. In each of these case, the sequences involved are asymptotically subcritical. From this large body of work, we distill the following general picture of what it means to be barely subcritical, critical and supercritical: We say that a sequence $(\pi'_n)_{n \geq 1}$ is *barely subcritical* when it is asymptotically subcritical and there exists an $\varepsilon > 0$ such that

$$(4.1.12) \quad \liminf_{n \rightarrow \infty} \mathbb{P}_{\pi_n} (|\mathcal{C}_{(2)}(\pi_n)|/|\mathcal{C}_{(1)}(\pi_n)| \geq \varepsilon) = 1,$$

while we say that a sequence $(\pi_n)_{n \geq 1}$ is *barely supercritical* when it is asymptotically subcritical and

$$(4.1.13) \quad \frac{|\mathcal{C}_{(2)}(\pi_n)|}{|\mathcal{C}_{(1)}(\pi_n)|} \xrightarrow{\mathbb{P}} 0.$$

Finally, we say that $(\pi_n)_{n \geq 1}$ is *in the critical window* when it is asymptotically subcritical and there exist $0 < a < b < \infty$, such that

$$(4.1.14) \quad 0 < \liminf_{n \rightarrow \infty} \mathbb{P}_{\pi_n} (|\mathcal{C}_{(2)}(\pi_n)|/|\mathcal{C}_{(1)}(\pi_n)| \in [a, b]) \\ \leq \limsup_{n \rightarrow \infty} \mathbb{P}_{\pi_n} (|\mathcal{C}_{(2)}(\pi_n)|/|\mathcal{C}_{(1)}(\pi_n)| \in [a, b]) < 1.$$

A few remarks are in place here. In most cases, much stronger results are being proved in the literature. Often, in the barely subcritical regime, explicit sequences $(a_n)_{n \geq 1}$ are identified such that $|\mathcal{C}_{(i)}(\pi_n)|/a_n \xrightarrow{\mathbb{P}} c_i > 0$ for every $i \geq 1$, which certainly implies (4.1.12). In the barely supercritical regime, explicit sequences $(a_n)_{n \geq 1}$ are identified such that $|\mathcal{C}_{(1)}(\pi_n)|/a_n \xrightarrow{\mathbb{P}} c > 0$, while $|\mathcal{C}_{(2)}(\pi_n)|/a_n \xrightarrow{\mathbb{P}} 0$. Again this certainly implies (4.1.12). Finally, in the critical window, often results are proved that there exists a sequence $(a_n)_{n \geq 1}$ such that $a_n^{-1} (|\mathcal{C}_{(i)}(\pi_n)|)_{i \geq 1}$ converges in distribution to a certain (infinite) random vector in a suitable topology. We will see quite a few of such sharper results in what follows. Since the precise limits and sequences $(a_n)_{n \geq 1}$ depend on the precise setting considered, we have generalized these results so that at least we have a clear idea what the three different regimes mean.

We could have defined the three regimes of barely subcritical, critical window and barely supercritical $(\pi_n)_{n \geq 1}$ also in terms of related random graph quantities. A particularly interesting one is the *average cluster size* or *susceptibility*

$$(4.1.15) \quad X_n(\pi) = \frac{1}{n} \sum_{v \in [n]} |\mathcal{C}_v(\pi)| = \frac{1}{n} \sum_{i \geq 1} |\mathcal{C}_{(i)}(\pi)|^2,$$

and let $\chi_n(\pi) = \mathbb{E}_\pi[X_n(\pi)]$ denote the average susceptibility. Then, the critical window could be defined in terms of those sequences of edge-retention probabilities $(\pi_n)_{n \geq 1}$ for which $X_n(\pi)/\chi_n(\pi)$ does not converge in probability to 1. The barely subcritical sequences $(\pi_n)_{n \geq 1}$ are then the sequences that are not in the critical window and are below it, while

the supercritical sequences $(\pi_n)_{n \geq 1}$ are then the sequences that are not in the critical window and are above it. However, this definition is often harder to work with. Since we will be mainly interested in the scaling of the *largest* clusters, we decided for (4.1.12)–(4.1.14) instead. It is of interest to investigate other definitions of the barely sub- and supercritical regimes and the critical window, see e.g., Nachmias and Peres [219] for an extensive discussion, and Janson and Warnke [188] for the example of the complete graph.

Organisation of Chapter 4. This chapter is organised as follows. In Section 4.2, we fix ideas by studying percolation on the complete graph, which is the same as the Erdős-Rényi random graph. There, we also discuss one of the main tools used in this section, namely, the Martingale Functional Central Limit Theorem. In Section 4.3, we discuss the asymptotic phase transition of percolation on the configuration model. In Section 4.4, we discuss the critical window of the configuration model when the degrees have finite third moments. In Section 4.5, we study the critical window of the configuration model with power-law degree sequences with in finite third moments. In Section 4.6, we discuss the notion of metric space convergence of critical clusters, a topic of fierce research in recent years. In Section 4.7, we discuss percolation on the preferential attachment model, which turns out to behave rather differently compared to the configuration model. We close with further results in Section 4.8, also discussing the relation between critical percolation and minimal spanning trees on graphs, and notes and discussion in Section 4.9.

4.2. Fixing ideas: complete graph and Erdős-Rényi

In this section, we discuss percolation on the complete graph, which is the Erdős-Rényi random graph. Of course, if we take $ER_n(\lambda/n)$ and perform percolation on it with edge-retention probability π , then we simply obtain $ER_n(\pi\lambda/n)$, and thus, there is not much to prove. This is reflected in the following exercise, which identifies the asymptotically subcritical and supercritical sequences $(\pi_n)_{n \geq 1}$:

EXERCISE 4.7 (Asymptotic phase transition supercritical Erdős-Rényi random graph). *Use Theorem 2.19 to show that any $\pi > 1/\lambda$ is asymptotically supercritical, while any $\pi \leq 1/\lambda$ is an asymptotically subcritical. We conclude that $\pi_c = 1/\lambda$ is asymptotically critical.*

However, while asymptotically supercriticality and subcriticality are not so interesting for the Erdős-Rényi random graph, the critical window as well as the barely sub- and supercritical regimes *are* highly interesting. We focus on the scaling limits of the cluster sizes in the critical window of the Erdős-Rényi random graph, as identified in the beautiful work of David Aldous [9]. Aldous' Theorem can be seen as concluding a long line of research, mainly in the probabilistic combinatorics community, on scaling behavior of the Erdős-Rényi random graph at, and close to, the critical point. This work was initiated by Erdős and Rényi [122, 123], and explored to full detail in works by Bollobás, Łuczak, Knuth, Pittel and Wierman. See the book of Bollobás [51] for results up to 1984, and [184, 210] for references to subsequent work. Bollobás [49] and Łuczak [208] were the first to identify the *scaling window* of the Erdős-Rényi random graph. For results on the Erdős-Rényi random graph we further refer to the monographs by Bollobás [52], Janson, Łuczak and Rucenski [186] and Alon and Spencer [13] and the references therein. Also [160,

Chapters 4 and 5] describes many results about the phase transition of the Erdős-Rényi random graph.

Let us start by describing the main result in an informal way. We will look at the scaling limit of the ordered cluster sizes within the critical window. As mentioned before, for the convergence of an infinite vector, the *topology* in which convergence takes place is highly relevant. We will be quite careful with topologies in this chapter, and will later also consider convergence of random metric spaces for which topologies are quite delicate.

Take $\pi_\lambda = (1 + \lambda n^{-1/3})/n$, and recall that $\mathcal{C}_{(i)}$ denotes the i th largest component. It is convenient to make the dependence on λ explicit, so that from now on we write $\mathcal{C}_{(i)}(\pi_\lambda)$. In this section, we informally describe the scaling limit of $n^{-2/3}(|\mathcal{C}_{(i)}(\pi_\lambda)|)_{i \geq 1}$, stating some of the main results as well as highlight the ideas behind the proof.

To formulate the convergence result, define ℓ_{\searrow}^2 to be the set of infinite sequences $x = (x_i)_{i \geq 1}$ with $x_1 \geq x_2 \geq \dots \geq 0$ and $\sum_{i \geq 1} x_i^2 < \infty$, and define the ℓ_{\searrow}^2 metric by

$$(4.2.1) \quad d(x, y) = \sqrt{\sum_{i \geq 1} (x_i - y_i)^2}.$$

For fixed $\lambda \in \mathbb{R}$, consider the inhomogeneous Brownian motion, or *Brownian motion minus a parabola*, as

$$(4.2.2) \quad B^\lambda(t) = B(t) + \lambda t - \frac{1}{2}t^2,$$

where B is standard Brownian motion. The reflected version of this process is denoted by R^λ , and can be obtained as

$$(4.2.3) \quad R^\lambda(t) = B^\lambda(t) - \min_{0 \leq u \leq t} B^\lambda(u).$$

Aldous [9] shows that the excursions of B^λ from 0 can be ranked in increasing order as, say, $\gamma_1(\pi_\lambda) > \gamma_2(\pi_\lambda) > \dots$. The main result for the cluster sizes of the Erdős-Rényi random graph within the scaling window is as follows:

THEOREM 4.1 (Cluster sizes Erdős-Rényi random graph within the scaling window [9]). *For the Erdős-Rényi random graph with $\pi = (1 + \lambda n^{-1/3})/n$, as $n \rightarrow \infty$,*

$$(4.2.4) \quad (n^{-2/3}|\mathcal{C}_{(i)}(\pi_\lambda)|)_{i \geq 1} \xrightarrow{d} (\gamma_i(\pi_\lambda))_{i \geq 1},$$

in distribution and with respect to the ℓ_{\searrow}^2 topology.

We will fully prove Theorem 4.1 below, and we start by explaining the ideas leading up to it. The proof relies on an *exploration* of the components of the Erdős-Rényi random graph. To set this exploration up, we successively explore clusters, starting from the cluster of a single vertex. In the exploration, we keep track of the number of active and neutral vertices. Here we call a vertex *active* when it has been found to be part of the cluster that is currently explored, but its neighbors in the random graph have not yet been identified. We call a vertex *neutral* when it has not been active yet.

This exploration is described in terms of a stochastic process $(S_n(i))_{i \geq 0}$, which encodes the cluster sizes as well as their structure. To describe the exploration process, we let $S_n(0) = 0, N_n(0) = n - 1, R_n(0) = 1$, and let $S_n(i)$, for $i \geq 1$, satisfy the recursion

$$(4.2.5) \quad S_n(i) = S_n(i - 1) + X_n(i) - 1,$$

where $X_n(i) \sim \text{Bin}(N_n(i-1), p)$. Here $N_n(i-1)$ denotes the number of neutral vertices after we have explored $i-1$ vertices. In this exploration, $-\inf_{j \in [i]} S_n(j)$ denotes the number of disjoint clusters that have been fully explored after the i th exploration, while, for $i \geq 1$,

$$(4.2.6) \quad R_n(i) = S_n(i) - \inf_{j \in [i-1]} S_n(j) + 1$$

denotes the number of active vertices at time i in the cluster that we are currently exploring. We see that, for $i \geq 1$,

$$(4.2.7) \quad N_n(i) = n - i - R_n(i)$$

denotes the number of neutral vertices after the i th exploration.

We can view $(R_n(i))_{i \geq 0}$ as the reflection of the process $(S_n(i))_{i \geq 0}$. The exploration can perhaps be more easily understood in terms $(R_n(i))_{i \geq 0}$. When $R_n(i) = 0$, we have fully explored a cluster. In particular, when we explore \mathcal{C}_{v_1} , then

$$(4.2.8) \quad |\mathcal{C}_{v_1}| = \inf\{i > 0: R_n(i) = 0\}.$$

After having explored $\mathcal{C}(v_1)$, we explore $\mathcal{C}(v_2)$ for some $v_2 \notin \mathcal{C}_{v_1}$, and obtain that

$$(4.2.9) \quad |\mathcal{C}_{v_2}| = \inf\{i > |\mathcal{C}_{v_1}|: R_n(i) = 0\} - |\mathcal{C}_{v_1}|.$$

Iterating this procedure, we see that

$$(4.2.10) \quad |\mathcal{C}_{v_j}| = \inf\{i > |\mathcal{C}_{v_1}| + \dots + |\mathcal{C}_{v_{j-1}}|: R_n(i) = 0\} - |\mathcal{C}_{v_1}| - \dots - |\mathcal{C}_{v_{j-1}}|.$$

Inspecting (4.2.6), we thus see that a cluster is fully explored when $S_n(i) - \inf_{j \in [i-1]} S_n(j) + 1 = 0$, which is the same as saying that $S_n(i) = \inf_{j \in [i]} S_n(j)$ for the first time. Thus, the total number of clusters that are fully explored up to time i indeed equals $-\inf_{j \in [i]} S_n(j)$.

It turns out to be more convenient to deal with the process $(S_n(i))_{i \geq 0}$ than with $(R_n(i))_{i \geq 0}$, since the scaling limit of $(S_n(i))_{i \geq 0}$ can be more easily described:

THEOREM 4.2 (Cluster exploration Erdős-Rényi random graph within the scaling window [9]). *For the Erdős-Rényi random graph with $p = (1 + \lambda n^{-1/3})/n$, as $n \rightarrow \infty$,*

$$(4.2.11) \quad (n^{-1/3} S_n(tn^{2/3}))_{t \geq 0} \xrightarrow{d} (B^\lambda(t))_{t \geq 0}.$$

Theorem 4.2, together with (4.2.6) and the continuity of the reflection mapping, implies that

$$(4.2.12) \quad (n^{-1/3} R_n(tn^{2/3}))_{t \geq 0} \xrightarrow{d} (R^\lambda(t))_{t \geq 0}, \quad \text{where} \quad R^\lambda(t) = B^\lambda(t) - \inf_{s \in [0, t]} B^\lambda(s).$$

Since the excursions of $(R^\lambda(t))_{t \geq 0}$ describe the component sizes, (4.2.12) can be seen to imply Theorem 4.1 and describes the limiting law of the largest connected components.

Before giving a proof of Theorem 4.2, let us next informally describing the proof of Theorem 4.2. Since $n^{-1/3} S_n(tn^{2/3})$ will be seen to converge in distribution, and we are investigating the exploration process at times $tn^{2/3}$, we simplify our lives considerably, and approximate (4.2.7) to $N_n(i) \approx n - i$. This means that the random variables $(X_n(i))_{i \geq 1}$ in (4.2.5) are close to being independent with

$$(4.2.13) \quad X_n(i) \approx \text{Bin}(n - i, p) = \text{Bin}(n - i, (1 + \lambda n^{-1/3})/n).$$

Here, and in what follows, \approx denotes an uncontrolled approximation.

It is well known that a Binomial random variable with parameters n and success probability μ/n is close to a Poisson random variable with mean μ . In our case,

$$(4.2.14) \quad \mu = (n - i)p = (n - i)(1 + \lambda n^{-1/3})/n \approx 1 + \lambda n^{-1/3} - i/n.$$

Note that when $i = tn^{2/3}$, both correction terms are of the same order in n . Thus, we approximate

$$(4.2.15) \quad S_n(i) \approx \sum_{j=1}^i (Y_j - 1),$$

where $Y_j \sim \text{Poi}(1 + \lambda n^{-1/3} - j/n)$ are independent. Since sums of independent Poisson variables are Poisson again, we thus obtain that

$$(4.2.16) \quad S_n(i) \approx S_n^*(i),$$

where

$$(4.2.17) \quad S_n^*(i) \sim \text{Poi}\left(\sum_{j=1}^i (1 + \lambda n^{-1/3} - j/n)\right) - i = \text{Poi}\left(i + i\lambda n^{-1/3} - \frac{i^2}{2n}\right) - i,$$

and $(S_n^*(i) - S_n^*(i - 1))_{i \geq 1}$ are independent. Now we multiply by $n^{-1/3}$ and take $i = tn^{2/3}$ to obtain

$$(4.2.18) \quad n^{-1/3} S_n^*(tn^{2/3}) \sim n^{-1/3} \left(\text{Poi}(tn^{2/3} + t\lambda n^{1/3} - \frac{1}{2}t^2 n^{1/3}) - tn^{2/3} \right),$$

Since a Poisson process is to leading order deterministic, we can approximate

$$(4.2.19) \quad n^{-1/3} S_n^*(tn^{2/3}) \sim n^{-1/3} (\text{Poi}(tn^{2/3}) - tn^{2/3}) + t\lambda - \frac{1}{2}t^2 \xrightarrow{d} B(t) + t\lambda - \frac{1}{2}t^2.$$

This informally explains the proof of Theorem 4.2. To make this proof rigorous, one typically resorts to Martingale Functional Central Limit Theorems, see for example how Aldous does this nicely in [9]. We explain this in more detail in the next section.

4.2.1. The trick of the trade: Martingale Functional CLT. Even though the above approximations may seem quite convincing, they are far from a proof. Indeed, there is substantial *dependence* between the random variables $(X_n(i))_{i \geq 1}$, and, of course, binomial random variables with small success probabilities can be well-approximated by Poisson random variables, but one needs to keep track of the error made along the way, and these errors might accumulate. Therefore, we need a highly *robust* proof, particularly when we wish to extend the results also to other settings, as we will do in the remainder of this chapter.

Such a robust tool can be found in the form of *martingale functional central limit theorems* (MFCLT), see e.g., the classic book on the subject by Ethier and Kurtz [128, Section 7]. In this section, we study such limit theorems in the special case in which the limit process is a standard Brownian motion. For a thorough overview of this important case, see Whitt [252]. Recall that if $\mathbf{M} = (M(t))_{t \geq 0}$ is a square-integrable martingale with respect to a filtration $(\mathcal{F}_t)_{t \geq 0}$, the *predictable quadratic variation process* associated with $M(t)$ is the unique nondecreasing, nonnegative, predictable, integrable process $V(t)$ such that $M^2(t) - V(t)$ is a martingale with respect to $(\mathcal{F}_t)_{t \geq 0}$. Here, we use the boldface notation \mathbf{X} for a time-dependent stochastic process $(X(t))_{t \geq 0}$. Further, unless stated otherwise, $\mathbb{C}[0, t]$ denotes the set of all continuous functions from $[0, t]$ to \mathbb{R} equipped

with the topology induced by sup-norm $\|\cdot\|_t$. Similarly, $\mathbb{D}[0, t]$ (resp. $\mathbb{D}[0, \infty)$) denotes the set of all càdlàg functions from $[0, t]$ (resp. $[0, \infty)$) to \mathbb{R} equipped with the Skorokhod J_1 topology. The MFCLT is the following result:

THEOREM 4.3 (Martingale functional central limit theorem). *Let $(\mathcal{F}_n)_{n \in \mathbb{N}}$ be an increasing filtration and $(\bar{M}_n(t))_{t \geq 0}$ be a sequence of continuous-time, real-valued, square-integrable martingales, each with respect to \mathcal{F}_n , such that $\bar{M}_n(0) = 0$. Assume that $(\bar{V}_n(t))_{t \geq 0}$, the predictable quadratic variation process associated with $(\bar{M}_n(t))_{t \geq 0}$, and $(\bar{M}_n(t))_{t \geq 0}$ satisfy the following conditions:*

- (a) $\bar{V}_n(t) \xrightarrow{\mathbb{P}} \sigma^2 t, \quad \forall t \in [0, \infty];$
- (b) $\lim_{n \rightarrow \infty} \mathbb{E}[\sup_{t \leq \bar{t}} |\bar{V}_n(t) - \bar{V}_n(t^-)|] = 0, \quad \forall \bar{t} \in \mathbb{R}^+;$
- (c) $\lim_{n \rightarrow \infty} \mathbb{E}[\sup_{t \leq \bar{t}} |\bar{M}_n(t) - \bar{M}_n(t^-)|^2] = 0, \quad \forall \bar{t} \in \mathbb{R}^+.$

Then, as $n \rightarrow \infty$, $(\bar{M}_n(t))_{t \geq 0}$ converges in distribution in $\mathcal{D}([0, \infty), \mathbb{R})$ to a centered Brownian motion with variance $\sigma^2 t$.

In this chapter, we often rely on the martingale functional central limit theorem for various explorations of critical clusters in random graphs, and we give an example of its usage in the case of the Erdős-Rényi random graph, where its use is the simplest. In order to apply the martingale functional central limit theorem, we apply a Doob-type decomposition to the process $S_n(k)$ to write it as the sum of a martingale—which will converge to the Brownian motion—and an appropriate drift term:

$$\begin{aligned} S_n(k) &= \sum_{i=1}^k (X_n(i) - \mathbb{E}[X_n(i) | \mathcal{F}_{i-1}]) + \sum_{i=1}^k (\mathbb{E}[X_n(i) | \mathcal{F}_{i-1}] - 1) \\ (4.2.20) \quad &=: M_n(k) + C_n(k), \end{aligned}$$

with $(\mathcal{F}_i)_{i \geq 1}$ the filtration generated by $(X_n(k))_{k \geq 1}$, i.e. $\mathcal{F}_i = \sigma((X_n(k))_{k=1}^i)$. Another useful Doob decomposition is

$$(4.2.21) \quad M_n^2(k) = Z_n(k) + V_n(k),$$

with $Z_n(k)$ a martingale and $V_n(k)$ the discrete-time predictable quadratic variation of the process $M_n(k)$. Note that for every fixed n and k , $|M_n(k)|$ is bounded and thus has a finite second moment. Therefore, $V_n(k)$ exists and is given by

$$\begin{aligned} V_n(k) &= \sum_{i=1}^k \mathbb{E}[(X_n(i) - \mathbb{E}[X_n(i) | \mathcal{F}_{i-1}])^2 | \mathcal{F}_{i-1}] \\ (4.2.22) \quad &= \sum_{i=1}^k (\mathbb{E}[X_n(i)^2 | \mathcal{F}_{i-1}] - \mathbb{E}[X_n(i) | \mathcal{F}_{i-1}]^2). \end{aligned}$$

To see this, we rewrite

$$\begin{aligned} M_n^2(k) &= \sum_{i=1}^k (X_n(i) - \mathbb{E}[X_n(i) | \mathcal{F}_{i-1}])^2 + \sum_{\substack{i, j \leq k \\ i \neq j}} (X_n(i) - \mathbb{E}[X_n(i) | \mathcal{F}_{i-1}]) (X_n(j) - \mathbb{E}[X_n(j) | \mathcal{F}_{j-1}]) \\ (4.2.23) \quad &=: \sum_{i=1}^k (X_n(i) - \mathbb{E}[X_n(i) | \mathcal{F}_{i-1}])^2 + L_n(k). \end{aligned}$$

By developing $L_n(k)$ one can easily see that it is a martingale:

EXERCISE 4.8 (Martingale $(L_n(k))_{k \geq 1}$). *Show that the process $(L_n(k))_{k \geq 1}$ indeed is a martingale.*

The decomposition (4.2.21) then follows from

$$\begin{aligned} Z_n(k) &:= \sum_{i=1}^k (X_n(i) - \mathbb{E}[X_n(i)|\mathcal{F}_{i-1}])^2 - \sum_{i=1}^k \mathbb{E}[(X_n(i) - \mathbb{E}[X_n(i)|\mathcal{F}_{i-1}])^2 | \mathcal{F}_{i-1}] \\ &\quad + L_n(k) \\ (4.2.24) \quad V_n(k) &:= \sum_{i=1}^k \mathbb{E}[(X_n(i) - \mathbb{E}[X_n(i)|\mathcal{F}_{i-1}])^2 | \mathcal{F}_{i-1}]. \end{aligned}$$

Note that $Z_n(k)$ is the sum of two martingales and is thus indeed itself a martingale.

4.2.2. Application Martingale FCLT to the Erdős-Rényi random graph.

The general idea to prove Theorem 4.2 is to first show the weak convergence of the (rescaled) process $(S_n(k))_{k \geq 0}$ defined in (4.2.5) and then to deduce the convergence of the reflected process $(R_n(k))_{k \geq 0}$ exploiting the following continuous-mapping theorem. In fact, this is a special case of a general technique known as the *continuous-mapping approach* (Whitt in [250] and [251] gives a detailed description of this technique). Through the continuous-mapping approach one reduces the problem of establishing convergence of random objects to one of continuity of suitable functions:

THEOREM 4.4 (Continuous Mapping Theorem). *If $\mathbf{X}_n \xrightarrow{d} \mathbf{X}$ and f is continuous almost surely with respect to the distribution of \mathbf{X} , then $f(\mathbf{X}_n) \xrightarrow{d} f(\mathbf{X})$.*

Suppose we have shown that $n^{-1/3}S_n(\cdot n^{2/3}) \xrightarrow{d} B^\lambda(\cdot)$. To prove Theorem 4.1, we are left to use this convergence to obtain the weak convergence of the cluster sizes. This is done in two key steps, namely, (a) the convergence of the reflected process; and (b) the convergence of the excursion from it.

(a) Convergence of the reflected process. Here we study the convergence

$$(4.2.25) \quad n^{-1/3}R_n(\cdot n^{2/3}) \xrightarrow{d} R^\lambda(\cdot) \quad \text{subject to} \quad n^{-1/3}S_n(\cdot n^{2/3}) \xrightarrow{d} B^\lambda(\cdot).$$

It suffices to prove that the function $\phi: \mathcal{D} \rightarrow \mathcal{D}$ given by

$$(4.2.26) \quad \phi: f(\cdot) \mapsto f(\cdot) - \inf_{y \leq \cdot} f^-(y)$$

is continuous almost surely with respect to the distribution of $B^\lambda(\cdot)$. For this note that $\mathbb{P}(B^\lambda(\cdot) \in \mathcal{C}) = 1$, where $\mathcal{C} = \mathcal{C}([0, \infty), \mathbb{R}) \subset \mathcal{D}$ denotes the space of continuous functions from $[0, \infty)$ to \mathbb{R} . Then by Whitt [250, Theorem 4.1] and [250, Theorem 6.1], ϕ is continuous almost surely with respect to the distribution of $B^\lambda(\cdot)$. Thus, the convergence $n^{-1/3}R_n(\cdot n^{2/3}) \xrightarrow{d} R^\lambda(\cdot)$ follows from $n^{-1/3}S_n(\cdot n^{2/3}) \xrightarrow{d} B^\lambda(\cdot)$.

(b) Convergence of the excursions of the reflected process. We again apply the Continuous Mapping Theorem 4.4. For this, we note that, by Aldous [9, Lemma 25] and the Cameron-Martin theorem, almost surely, the excursions of the paths of \mathbf{B}^λ can be rearranged in decreasing order of length and the ordered excursion lengths can be considered as a vector in ℓ_1^2 . Let $\gamma^\lambda = (|\gamma_j^\lambda|)_{j \geq 1}$ be the ordered excursion lengths of $\mathbf{B}_{\mu, \eta}^\lambda$. Since the reflection as well as the excursions form a continuous functional of the path, we obtain that the excursions of \mathbf{R}_n , after rescaling, also converge to those of \mathbf{R}^λ . This explains why it suffices to prove Theorem 4.2 in order to prove Theorem 4.1. We next discuss the proof of Theorem 4.2.

Proof of Theorem 4.2. Recall that $S_n(k)$ can be decomposed as $S_n(k) = M_n(k) + C_n(k)$, where $M_n(k)$ is a martingale and $C_n(k)$ is the drift term. Moreover, $M_n^2(k)$ was also written as $M_n^2(k) = Z_n(k) + V_n(k)$ with $Z_n(k)$ the Doob martingale and $V_n(k)$ its drift. The proof then consists of verifying the following conditions:

$$\begin{aligned} \text{MCLT(a)} \quad & \sup_{t \leq \bar{t}} |n^{-1/3} C_n(tn^{2/3}) - \lambda t + \frac{1}{2} t^2| \xrightarrow{\mathbb{P}} 0, \quad \forall \bar{t} \in \mathbb{R}^+; \\ \text{MCLT(b)} \quad & n^{-2/3} V_n(tn^{2/3}) \xrightarrow{\mathbb{P}} \sigma^2 t, \quad \forall t \in \mathbb{R}^+; \\ \text{MCLT(c)} \quad & \lim_{n \rightarrow \infty} n^{-2/3} \mathbb{E}[\sup_{t \leq \bar{t}} |V_n(tn^{2/3}) - V_n(tn^{2/3}-)|] = 0, \quad \forall \bar{t} \in \mathbb{R}^+; \\ \text{MCLT(d)} \quad & \lim_{n \rightarrow \infty} n^{-2/3} \mathbb{E}[\sup_{t \leq \bar{t}} |M_n(tn^{2/3}) - M_n(tn^{2/3}-)|^2] = 0, \quad \forall \bar{t} \in \mathbb{R}^+. \end{aligned}$$

We next verify these conditions in the reverse order, which is also the order of increasing difficulty:

Verification of MCLT(d). Recall (4.2.5) to see that

$$(4.2.27) \quad M_n(k) - M_n(k-1) = X_n(k) - \mathbb{E}[X_n(k) \mid \mathcal{F}_{k-1}] = X_n(k) - N_n(k-1)(1 + \lambda n^{-1/3})/n,$$

so that

$$(4.2.28) \quad (M_n(k) - M_n(k-1))^2 \leq X_n(k)^2 + N_n(k-1)^2(1 + \lambda n^{-1/3})^2/n^2.$$

The second term is uniformly bounded by $(1 + \lambda n^{-1/3})^2 = o(n^{2/3})$, as required. Thus,

$$(4.2.29) \quad n^{-2/3} \mathbb{E}[\sup_{t \leq \bar{t}} |M_n(tn^{2/3}) - M_n(tn^{2/3}-)|^2] \leq n^{-2/3} \mathbb{E}[\max_{k \leq \bar{t}n^{2/3}} X_n(k)^2] + o(1).$$

Since, conditionally on \mathcal{F}_{k-1} , $X_n(k) \preceq Y_n(k) \stackrel{d}{=} \text{Bin}(n, \pi)$, with $(Y_n(k))_{k \geq 1}$ being i.i.d. we obtain that

$$(4.2.30) \quad \max_{k \leq \bar{t}n^{2/3}} X_n(k)^2 \preceq \max_{k \leq \bar{t}n^{2/3}} Y_n(k)^2,$$

where $(Y_n(k))_{k \geq 1}$ are i.i.d. $\text{Bin}(n, (1 + \lambda n^{-1/3})/n)$ random variables. We compute that, for every $x, t \geq 0$

$$(4.2.31) \quad \mathbb{P}(Y_n(k) \geq x) = \mathbb{P}(e^{tY_n(k)} \geq e^{tx}) \leq e^{-tx} \mathbb{E}[e^{tY_n(k)}] = e^{-tx} [1 + (e^t - 1)(1 + \lambda n^{-1/3})/n]^n.$$

Using the bound that $1 + x \leq e^x$, valid for all $x \in \mathbb{R}$, we arrive at

$$(4.2.32) \quad \mathbb{P}(Y_n(k) \geq x) \leq e^{-tx + (e^t - 1)(1 + \lambda n^{-1/3})}.$$

Taking $t = 1$, and using that $(e^t - 1)(1 + \lambda n^{-1/3}) \leq 2$ for n sufficiently large, this gives rise to

$$(4.2.33) \quad \mathbb{P}(Y_n(k) \geq x) \leq e^2 e^{-x}.$$

As a result,

$$(4.2.34) \quad \mathbb{P}\left(\max_{k \leq \bar{t}n^{2/3}} Y_n(k) \geq 4 \log n\right) \leq \frac{e^2}{n^4}.$$

We conclude that

$$(4.2.35) \quad \begin{aligned} \mathbb{E}[\sup_{t \leq \bar{t}} |M_n(tn^{2/3}) - M_n(tn^{2/3}-)|^2] &\leq o(n^{2/3}) + (4 \log n)^2 + n^2 \mathbb{P}\left(\max_{k \leq \bar{t}n^{2/3}} Y_n(k) \geq 4 \log n\right) \\ &\leq o(n^{2/3}) + (4 \log n)^2 + \frac{e^2}{n^2} = o(n^{2/3}), \end{aligned}$$

as required.

Verification of MCLT(c). Recall (4.2.5) to see that, conditionally on \mathcal{F}_{k-1} , $X_n(k) \sim \text{Bin}(n, (1 + \lambda n^{-1/3})/n)$. Therefore, since $\text{Var}(\text{Bin}(n, p)) = np(1-p)$,

$$(4.2.36) \quad \begin{aligned} V_n(k) - V_n(k-1) &= \mathbb{E}[X_n(k)^2 | \mathcal{F}_{k-1}] - \mathbb{E}[X_n(k) | \mathcal{F}_{k-1}]^2 = \text{Var}(X_n(k)^2 | \mathcal{F}_{k-1}) \\ &= N_n(k-1)(1 + \lambda n^{-1/3})/n [1 - (1 + \lambda n^{-1/3})/n] \leq (1 + \lambda n^{-1/3}). \end{aligned}$$

As a result, MCLT(c) is trivially satisfied.

Verification of MCLT(b). As above, we obtain that

$$(4.2.37) \quad V_n(k) = \sum_{i=1}^k \frac{N_n(i-1)}{n} (1 + \lambda n^{-1/3}) [1 - (1 + \lambda n^{-1/3})/n].$$

We proceed to prove upper and lower bounds. The upper bound follows from the almost sure bound in (4.2.36), together with the fact that

$$(4.2.38) \quad V_n(k) = \sum_{i=1}^k \frac{N_n(i-1)}{n} (1 + \lambda n^{-1/3}) [1 - (1 + \lambda n^{-1/3})/n] \leq k(1 + \lambda n^{-1/3}),$$

so that $n^{-2/3}V_n(tn^{2/3}) \leq t + o(1)$, as required. For the lower bound, we need to prove a lower bound on $N_n(i-1)$. Note that $i \mapsto N_n(i)$ is decreasing, so that

$$(4.2.39) \quad n^{-2/3}V_n(tn^{2/3}) \geq t \frac{N_n(tn^{2/3})}{n} (1 + o(1)).$$

We recall that

$$(4.2.40) \quad N_n(tn^{2/3}) = n - \lceil tn^{2/3} \rceil - R_n(tn^{2/3}).$$

Further, as in the proof of MCLT(d), with $(Y_n(i))_{i \geq 1}$ i.i.d. such that $Y_n(i) \stackrel{d}{=} \text{Bin}(n, (1 + \lambda/n^{1/3})/n)$,

$$(4.2.41) \quad S_n(k) \leq \sum_{i=1}^k X_n(i) \leq \sum_{i=1}^k Y_n(i) \leq k(\log n)^4,$$

the latter bound holding whp. Therefore, also

$$(4.2.42) \quad R_n(k) = S_n(k) - \inf_{j \in [k-1]} S_n(j) + 1 \leq S_n(k) + 1 \leq k(\log n)^4 + 1.$$

We conclude that whp, and uniformly in $t \leq \bar{t}$,

$$(4.2.43) \quad n^{-2/3}V_n(tn^{2/3}) \geq t(1 - tn^{-1/3} - tn^{-1/3}(\log n)^4)(1 + o(1)) = t - o_{\mathbb{P}}(1),$$

as required.

Verification of MCLT(a). The convergence of the drift term in MCLT(a) is the hardest. We use (4.2.20) to write

$$\begin{aligned}
 C_n(k) &= \sum_{i=1}^k (\mathbb{E}[X_n(i) - 1 \mid \mathcal{F}_{i-1}] - 1) = \sum_{i=1}^k \left[\frac{N_n(i-1)}{n} (1 + \lambda n^{-1/3}) - 1 \right] \\
 (4.2.44) \quad &= k\lambda n^{-1/3} - \sum_{i=1}^k \frac{i-1}{n} (1 + \lambda n^{-1/3}) - \sum_{i=1}^k \frac{R_n(i-1)}{n} (1 + \lambda n^{-1/3}).
 \end{aligned}$$

We again work with upper and lower bounds. For the upper bound, obviously since $R_n(i-1) \geq 0$,

$$(4.2.45) \quad n^{-1/3} C_n(tn^{2/3}) \leq \lambda t - \frac{1}{2} t^2 (1 + \lambda n^{-1/3}) = \lambda t - \frac{1}{2} t^2 + o(1),$$

where the error term is deterministic and uniform in t . This establishes the uniform upper bound. For the lower bound, instead, we use $R_n(k) \leq S_n(k) + 1 \leq 1 + \sum_{i=1}^k (Y_n(i) - 1)$. Then, the ideas around (4.2.42), together with the fact that $\mathbb{E}[Y_n(i)] = 1 + o(1)$, can be used to prove that whp, for $k \geq \sqrt{n}$, there exists $\varepsilon_n = o(1)$ such that

$$(4.2.46) \quad \mathbb{P}\left(\sum_{i=1}^k (Y_n(i) - 1) \geq \varepsilon_n k\right) = o(n^{-2}),$$

so that the same bound also holds whp for *all* $k \leq n$ simultaneously. Consequently, the same bound also holds for $S_n(k)$, as we use next. Using further that (4.2.42) shows that whp $R_n(i-1) \leq (i-1)(\log n)^4$ for $i \leq \sqrt{n}$,

$$\begin{aligned}
 n^{-1/3} \sum_{i=1}^{tn^{2/3}} \frac{R_n(i-1)}{n} (1 + \lambda n^{-1/3}) &= n^{-4/3} (1 + o(1)) \left[\sum_{i=1}^{\sqrt{n}} [(i-1)(\log n)^4 + 1] + \sum_{i=\sqrt{n}}^{tn^{2/3}} \varepsilon_n i \right] \\
 (4.2.47) \quad &\leq n^{-4/3} (1 + o(1)) [n(\log n)^4 + \varepsilon_n (tn^{2/3})^2 + 1] = o(1),
 \end{aligned}$$

as required.

EXERCISE 4.9 (Tail bound $S_n(k)$). *Use the ideas around (4.2.42) to prove that (4.2.46) holds.*

Having established MCLT(a)-MCLT(d), we have obtained the proof of Theorem 4.2. \square

The Martingale functional central limit theorem in Theorem 4.3 is a highly robust tool, and has been the major workhorse for proving scaling limit results for critical random graphs, where considerable complications may arise due to the strong inhomogeneities in the graphs. We will see this in more detail in Section 4.4, where these ideas are applied to the configuration model. Further applications, with fewer details of the proofs, can be found in Section 4.8.

4.2.3. Further results on critical Erdős-Rényi random graphs. There are many recent extensions to the above scaling limit results. For example, already in his original paper, Aldous also investigates the *surplus* of the largest connected components. The surplus of a graph is the minimal number of edges that need to be removed in order to obtain a tree, and is related to the number of cycles in the graph. We leave this as an exercise:

EXERCISE 4.10 (Surplus critical Erdős-Rényi clusters). *Prove, using Theorem 4.2 and (4.2.12), that the number of cycle edges between vertices found in the time interval $[t_1 n^{2/3}, t_2 n^{2/3}]$ has an approximate Poisson distribution with random parameter*

$$(4.2.48) \quad \int_{t_1}^{t_2} R^\lambda(t) dt.$$

The multiplicative coalescent. Interestingly, Aldous gives a second interpretation of the scaling limit of clusters in terms of *multiplicative coalescents*. Indeed, when we interpret θ as a time variable and use the Harris coupling of percolation (see page 153), we see that within the time interval $(\theta, \theta + \varepsilon\theta)$, two clusters of size $|\mathcal{C}_{(i)}(\pi_\lambda)| = xn^{2/3}$ and $|\mathcal{C}_{(j)}(\pi_\lambda)| = yn^{2/3}$ merge with probability close to $\frac{\varepsilon\theta}{n^{4/3}}(xn^{2/3})(yn^{2/3}) = \varepsilon\theta xy$ to create a cluster of size $(x + y)n^{2/3}$. When rescaling the cluster sizes by $n^{-2/3}$, this dynamics is called the *multiplicative coalescent*. More formally, define a Markov process $\mathbb{X} := (\mathbf{X}(\lambda))_{-\infty < \lambda < \infty}$ on $\mathbb{D}((-\infty, \infty), \ell_{\downarrow}^2)$, called the *multiplicative coalescent process*. Think of $\mathbf{X}(\lambda)$ as a collection of masses of some objects in a system at time λ . Thus the i object has mass $X_i(\lambda)$ at time λ . The evolution of the system take place according to the following rule at time λ : At rate $X_i(\lambda)X_j(\lambda)$, objects i and j merge into a new component of mass $X_i(\lambda) + X_j(\lambda)$. This process has been extensively studied in by Aldous and Limic [9, 10]. In particular, Aldous, in [9, Proposition 5] showed that this is a Feller process. We can interpret the limiting cluster sizes $(\gamma_i(\pi_\lambda))_{i \geq 1}$ in Theorem 4.1 as the ordered distribution of masses at time λ . Thus, Theorem 4.1 can be seen as the convergence of the process as a function of λ at a *single* time. We will later get back to this issue, and discuss convergence *as a process*.

More recent work of Addario-Berry, Broutin and Goldschmidt [5] (see also [4]) shows that the largest critical clusters, viewed as metric spaces, converge in distribution in an appropriate topology to some limiting graphs. The components $\mathcal{C}_{(i)}$ are considered as metric spaces using the graph distance rescaled by $n^{-1/3}$. The authors prove that the sequence of rescaled components converges to a sequence of continuous compact metric spaces. These limiting continuum graphs are *almost* trees, but they contain a finite number of cycles corresponding to the surpluses of the clusters (recall Exercise 4.10). The convergence is in the Gromov-Hausdorff-Prokhorov topology on metric spaces. This result has several interesting consequences, such as the convergence in distribution of the diameter of $\mathcal{C}_{(i)}$, rescaled by $n^{-1/3}$ and of the array of distances between uniformly chosen vertices in the cluster. We discuss these results in more detail in Section 4.6 below.

In Section 4.9, we discuss other results for the Erdős-Rényi random graph, including tail estimates on the largest connected component in the scaling window, and the structure of the largest connected component in the barely super-critical regime. We also comment more extensively on the history of the literature on the (critical) Erdős-Rényi random graph.

4.3. Percolation on the configuration model: Janson's construction

In the following four sections, we describe the critical behavior of percolation on the configuration model. These results can be seen as a blueprint for several results on the critical behavior of random graphs, which is why we explain it in quite some detail. We refer to Section 4.8 for a detailed explanation of such results.

In this section, we discuss the percolation phase transition on the configuration model. Remarkably, percolation on the configuration model can again be interpreted as another configuration model with a different degree distribution, just like percolation on the Erdős-Rényi random graph is again an Erdős-Rényi random graph with an updated edge probability. This is called *Janson's construction*, and we explain this in detail here. It will also be a major tool in the sections to follow.

Let us start by discussing our setting. We assume that Conditions 1.6(a)-(b) hold, and that $\text{CM}_n(\mathbf{d})$ is supercritical, so that

$$(4.3.1) \quad \nu_n = \frac{\sum_{i \in [n]} d_i(d_i - 1)}{\sum_{i \in [n]} d_i} \rightarrow \nu = \frac{\mathbb{E}[D(D - 1)]}{\mathbb{E}[D]} > 1.$$

When (4.3.1) fails, then, by Theorem 2.21, there is no giant component. As a result, even when we do not apply percolation, there is no large connected component, so that in particular, there cannot be an interesting percolation phase transition. In our first main result, which is due to Janson [181], we identify the percolation phase transition. In its statement, we write $G_D(s) = \mathbb{E}[s^D]$ for the probability generating function of the asymptotic degree distribution, as well as $\text{CM}_n(\mathbf{d}; \pi)$ for the result of percolation on $\text{CM}_n(\mathbf{d})$ with edge-retention probability π :

THEOREM 4.5 (Percolation phase transition in $\text{CM}_n(\mathbf{d})$). *Suppose that Conditions 1.6(a)-(b) hold and consider the random graph $\text{CM}_n(\mathbf{d})$, letting $n \rightarrow \infty$. Assume that (4.3.1) holds, so that $\text{CM}_n(\mathbf{d})$ is supercritical. Let $\mathcal{C}_{\max}(\pi)$ and $\mathcal{C}_{(2)}(\pi)$ be the largest and second largest components of $\text{CM}_n(\mathbf{d}; \pi)$ (breaking ties arbitrarily when necessary).*

(a) *If $\pi > 1/\nu = \mathbb{E}[D]/\mathbb{E}[D(D - 1)]$, then there exist $\eta^*(\pi) \in [0, 1)$, $\zeta(\pi) \in (0, 1]$ such that*

$$\begin{aligned} |\mathcal{C}_{\max}(\pi)|/n &\xrightarrow{\mathbb{P}} 1 - G_D(1 - \sqrt{\pi} + \eta^*(\pi)\sqrt{\pi}), \\ |E(\mathcal{C}_{\max}(\pi))|/n &\xrightarrow{\mathbb{P}} \frac{1}{2}\mathbb{E}[D](1 - \eta^*(\pi)^2), \end{aligned}$$

where $\eta^*(\pi)$ is the unique positive solution to

$$(4.3.2) \quad \eta^*(\pi) = (1 - \sqrt{\pi}) + \sqrt{\pi}G_{D^*-1}(1 - \sqrt{\pi} + \eta^*(\pi)\sqrt{\pi}).$$

while $|\mathcal{C}_{(2)}(\pi)|/n \xrightarrow{\mathbb{P}} 0$ and $|E(\mathcal{C}_{(2)}(\pi))|/n \xrightarrow{\mathbb{P}} 0$.

(b) *If $\pi \leq 1/\nu = \mathbb{E}[D]/\mathbb{E}[D(D - 1)]$, then $|\mathcal{C}_{\max}(\pi)|/n \xrightarrow{\mathbb{P}} 0$ and $|E(\mathcal{C}_{\max}(\pi))|/n \xrightarrow{\mathbb{P}} 0$.*

Theorem 4.5 identifies the percolation critical value as $\pi_c = 1/\nu = \mathbb{E}[D]/\mathbb{E}[D(D - 1)]$, in the sense that any $\pi < \pi_c$ is an asymptotically subcritical sequence, while any $\pi > \pi_c$ is an asymptotically supercritical sequence. In particular, $\pi_c = 0$ when $\mathbb{E}[D^2] = \infty$, a property that is sometimes called *instantaneous percolation* or *robustness*:

EXERCISE 4.11 (Robust phase transition for infinite-variance degrees). *Show that $\pi_c = 0$ when Conditions 1.6(a)-(b) holds with $\mathbb{E}[D^2] = \infty$. This means that, however small the edge-retention probability π is, there is always a giant percolation cluster remaining, so that $|\mathcal{C}_{\max}(\pi)|/n$ remains strictly positive in the large n limit.*

The property in Exercise 4.11 is sometimes called *robustness* of the underlying random graph. Robustness is a desirable property, since it implies that whichever the proportion of edges removed is, there always remains some connectivity in the remainder of the graph. Robustness does not happen on any graph with uniformly bounded degree:

EXERCISE 4.12 (Percolation on graphs with fixed degrees). *Let $(G_n)_{n \geq 1}$ be a graph sequence, where G_n has vertex set $[n]$. Show that $\pi_c \geq 1/(d-1)$ when G_n is such that every vertex has degree d . Extend this result to the statement that $\pi_c \geq 1/(d-1)$ when G_n is such that every vertex has degree at most d .*

EXERCISE 4.13 (Giant component cannot be all after percolation). *In Theorem 4.5, show that $\eta^*(\pi) > 0$ and $\zeta(\pi) < 1$ whenever $\pi < 1$. Thus, non-trivial percolation destroys a positive proportion of the giant.*

The proof of the percolation phase transition as described in Theorem 4.5 is quite pretty. It relies on the beautiful realization by Janson [181] that percolation on $\text{CM}_n(\mathbf{d})$ can again be described in terms of another configuration model. We call this *Janson's construction for percolation on the configuration model*. An application of Theorem 2.21 then proves the claim. Let us continue by describing Janson's construction:

Janson's construction for percolation on the configuration model. On the Erdős-Rényi random graph $\text{ER}_n(\lambda/n)$, when we perform percolation with parameter π , we obtain another Erdős-Rényi random graph, now with parameter $\pi\lambda/n$ instead. When the Erdős-Rényi random graph is supercritical so that $\lambda > 1$, we thus obtain that $\pi_c = 1/\lambda$. Janson's construction for percolation on the configuration model extends this observation to the configuration model, for which it is a little more involved.

Fix a degree distribution \mathbf{d} . For $v \in [n]$, let $d_v(\pi) = \text{Bin}(d_v, \sqrt{\pi})$, where $(\text{Bin}(d_v, \sqrt{\pi}))_{v \in [n]}$ are independent binomial random variables. Further, let

$$(4.3.3) \quad N^+ = N^+(\pi) = \sum_{v \in [n]} [d_v - \text{Bin}(d_v, \sqrt{\pi})],$$

and let $N = N(\pi) = n + N^+$. For $v \in [N] \setminus [n]$, set $d_v(\pi) = 1$ and color these vertices *red*. This has the following interpretation: We split a vertex of degree d_v into a single vertex of degree $d_v(\pi) = \text{Bin}(d_v, \sqrt{\pi})$, and $d_v - d_v(\pi)$ red vertices of degree 1. This is called the *explosion* of vertex v . We can think of the red vertices as being *artificial*, and they will later need to be removed. After this, an edge is retained when both of its half-edges are retained, which occurs with probability $\sqrt{\pi} \times \sqrt{\pi} = \pi$, as it should be. Since all edge-retention decisions are made independently, this is thus equivalent to percolation on $\text{CM}_n(\mathbf{d})$. We will give a formal proof of this fact in Lemma 4.6 below. Let $\text{CM}_N(\mathbf{d}(\pi))$ denote the configuration model with vertex set $[N]$ and degree distribution $\mathbf{d}(\pi) = (d_v(\pi))_{v \in [N]}$. Janson's construction is summarized in the following lemma:

LEMMA 4.6 (Janson's construction for percolation on the configuration model). *Percolation on the configuration model $\text{CM}_n(\mathbf{d}; \pi)$ has the same distribution as $\text{CM}_N(\mathbf{d}(\pi))$*

followed by the removal of the $N - n = N^+$ red vertices with degree 1. Here, we refer to the removal of a vertex of degree 1 as the act of removal of the vertex from the vertex set, together with the removal of its (single) edge from the edge set.

PROOF. By construction, decisions for different edges are made independently, so that it suffices to prove that each edge is kept with probability π . For this, we see that the pairing of the half-edges in $\text{CM}_N(\mathbf{d}(\pi))$ is a uniform matching, just as in $\text{CM}_n(\mathbf{d})$. Now, consider an edge $e = xy$, where x and y are the half-edges that e consists of, and assume that x is incident to vertex u while y is incident to vertex v . The half-edge x remains incident to vertex u with probability $\sqrt{\pi}$, and is made incident to another vertex $u' \in [N] \setminus [n]$ with probability $1 - \sqrt{\pi}$. Similarly, the half-edge y remains incident to vertex v with probability $\sqrt{\pi}$, and is made incident to another vertex $v' \in [N] \setminus [n]$ with probability $1 - \sqrt{\pi}$. Thus, the edge $e = xy$ between u and v is retained with probability π . When it is not retained, then one or both of its half-edges is incident to a red vertex in $[N] \setminus [n]$, which occurs with probability $1 - \pi$. In the final step of Janson's construction, it is therefore removed from the graph, irrespective of how many of the vertices the edge is incident to are red. This means that the artificial vertex of degree 1 that is created in the explosion of vertex u is later removed, together with its edge, and the same is true for the artificial vertex created through the explosion of v . This completes the proof. \square

REMARK 4.7 (Removing a random number of degree 1 vertices). By symmetry, the removal of the $N - n = N^+$ red vertices with degree 1 is, in distribution, the same as the removal of a collection of $N - n = N^+$ vertices of degree 1, uniformly at random from the collection of all vertices of degree 1 in $\text{CM}_N(\mathbf{d}(\pi))$. This is the construction that we will use in the sequel.

EXERCISE 4.14 (Janson's construction for site percolation). *Extend Janson's construction in Lemma 4.6 to site-percolation, where every vertex is removed with probability $1 - \pi$.*

PROOF OF THEOREM 4.5. We aim to use Theorem 2.21 on the configuration model $\text{CM}_N(\mathbf{d}(\pi))$. By Janson's construction in Lemma 4.6, it suffices to study the effect of the vertex explosions and the removal of red vertices of degree 1. We start by analyzing the degree distribution of $\text{CM}_N(\mathbf{d}(\pi))$ in order to verify that Conditions 1.6(a)-(b) hold.

Conditions 1.6(a)-(b) for the degree distribution of $\text{CM}_N(\mathbf{d}(\pi))$. Recall that $\ell_n = \sum_{v \in [n]} d_v$ denotes the total degree in $\text{CM}_n(\mathbf{d})$. The red vertices add $N - n = N^+$ of degree 1 vertices, and we compute that

$$(4.3.4) \quad \frac{N^+}{N} = \frac{\text{Bin}(\ell_n, 1 - \sqrt{\pi})}{n + \text{Bin}(\ell_n, 1 - \sqrt{\pi})} = \frac{\frac{1}{n} \text{Bin}(\ell_n, 1 - \sqrt{\pi})}{1 + \frac{1}{n} \text{Bin}(\ell_n, 1 - \sqrt{\pi})} \xrightarrow{\mathbb{P}} \frac{\mathbb{E}[D](1 - \sqrt{\pi})}{1 + \mathbb{E}[D](1 - \sqrt{\pi})}.$$

Since $\ell_n/n \rightarrow \mathbb{E}[D]$, we obtain that

$$(4.3.5) \quad \frac{N}{n} = \frac{n + \text{Bin}(\ell_n, 1 - \sqrt{\pi})}{n} = 1 + \frac{1}{n} \text{Bin}(\ell_n, 1 - \sqrt{\pi}) \xrightarrow{\mathbb{P}} 1 + \mathbb{E}[D](1 - \sqrt{\pi}).$$

Therefore, for every $k \neq 1$,

$$(4.3.6) \quad \mathbb{P}(D_N(\pi) = k) = \frac{1}{N} \sum_{v \in [N]} \mathbb{1}_{\{d_v(\pi) = k\}} = \frac{1}{N} \sum_{v \in [n]} \mathbb{1}_{\{d_v(\pi) = k\}} = \frac{n}{N} \frac{1}{n} \sum_{v \in [n]} \mathbb{1}_{\{d_v(\pi) = k\}},$$

where we use the fact that only the vertices in $[n]$ can have degree $k \neq 1$. We split

$$(4.3.7) \quad \frac{1}{n} \sum_{v \in [n]} \mathbb{1}_{\{d_v(\pi) = k\}} = \sum_{\ell \geq k} \frac{1}{n} \sum_{v \in [n]} \mathbb{1}_{\{d_v(\pi) = k, d_v = \ell\}}.$$

By the independence of $(d_v(\pi))_{v \in [n]}$,

$$(4.3.8) \quad \sum_{v \in [n]} \mathbb{1}_{\{d_v(\pi) = k, d_v = \ell\}} \stackrel{d}{=} \text{Bin}(n_\ell, b_{k,\ell}(\sqrt{\pi})),$$

where n_ℓ is the number of vertices in $[n]$ for which $d_v = \ell$ and

$$(4.3.9) \quad b_{k,\ell}(q) = \mathbb{P}(\text{Bin}(\ell, q) = k)$$

is the probability that a binomial random variable with success probability q and ℓ trials equals k . Since $n_\ell/n \rightarrow p_\ell$, we thus obtain that

$$(4.3.10) \quad \frac{1}{n} \sum_{v \in [n]} \mathbb{1}_{\{d_v(\pi) = k, d_v = \ell\}} = \frac{n_\ell}{n} \frac{1}{n_\ell} \sum_{v \in [n]} \mathbb{1}_{\{d_v(\pi) = k, d_v = \ell\}} \xrightarrow{\mathbb{P}} p_\ell b_{k,\ell}(\sqrt{\pi}).$$

As a result,

$$(4.3.11) \quad \mathbb{P}(D_N(\pi) = k) \xrightarrow{\mathbb{P}} \frac{1}{1 + \mathbb{E}[D](1 - \sqrt{\pi})} \sum_{\ell \geq k} p_\ell b_{k,\ell}(\sqrt{\pi}) \equiv p_k(\pi),$$

which proves that Condition 1.6(a) holds for every $k \neq 1$. Further, $\mathbb{P}(D_N(\pi) = 1) = 1 - \sum_{k \neq 1} \mathbb{P}(D_N(\pi) = k)$, which therefore also converges to some $p_1(\pi)$. We conclude that Condition 1.6(a) holds. The case where $k = 1$ is special, and, for later reference, let us compute that

$$(4.3.12) \quad p_1(\pi) = \frac{\mathbb{E}[D](1 - \sqrt{\pi})}{1 + \mathbb{E}[D](1 - \sqrt{\pi})} + \frac{1}{1 + \mathbb{E}[D](1 - \sqrt{\pi})} \sum_{\ell \geq 1} p_\ell b_{1,\ell}(\sqrt{\pi}).$$

Here, the first term corresponds to the proportion of red vertices as identified in (4.3.4), the second to the non-red vertices of degree 1 after percolation as identified in (4.3.10) for $k = 1$.

For Condition 1.6(b), we note that

$$(4.3.13) \quad \begin{aligned} \mathbb{E}_N[D_N(\pi)] &= \frac{1}{N} \sum_{v \in [N]} d_v(\pi) = \frac{1}{N} [N^+ + \sum_{v \in [n]} d_v(\pi)] = \frac{1}{N} [N^+ + \text{Bin}(\ell_n, \sqrt{\pi})] \\ &\xrightarrow{\mathbb{P}} \frac{\mathbb{E}[D](1 - \sqrt{\pi}) + \mathbb{E}[D]\sqrt{\pi}}{1 + \mathbb{E}[D](1 - \sqrt{\pi})} = \frac{\mathbb{E}[D]}{1 + \mathbb{E}[D](1 - \sqrt{\pi})}, \end{aligned}$$

as required. Thus, Conditions 1.6(a)-(b) hold.

The supercriticality condition for $\text{CM}_N(\mathbf{d}(\pi))$ and concentration inequalities.

Let

$$(4.3.14) \quad \nu_N(\pi) = \frac{\frac{1}{N} \sum_{v \in [N]} d_v(\pi)(d_v(\pi) - 1)}{\frac{1}{N} \sum_{v \in [N]} d_v(\pi)}$$

denote the ν_n parameter of $\text{CM}_N(\mathbf{d}(\pi))$, with n replaced with N and the degree distribution \mathbf{d} by $\mathbf{d}(\pi)$. In order to verify whether $\text{CM}_N(\mathbf{d}(\pi))$ is supercritical or subcritical, we need to verify whether $\lim_{n \rightarrow \infty} \nu_N(\pi) > 1$ or not. By Condition 1.6(b) as proved

in (4.3.13), we already know that the denominator converges to the right hand side of (4.3.13). We are left to deal with the numerator. For this, we will make use of a useful concentration inequality for sums of functions of independent variables, that we state now:

LEMMA 4.8 (Concentration for Lipschitz functions of i.i.d. random variables). *Let Z_1, Z_2, \dots, Z_N be independent random variables with Z_i taking values in Λ_i and $f: \prod_{i=1}^N \Lambda_i \rightarrow \mathbb{R}$ satisfies the following: If two vectors $z, z' \in \prod_{i=1}^N \Lambda_i$ differ only in the i th coordinate, then $|f(z) - f(z')| \leq c_i$ for some constant c_i . Then, for any $t > 0$, the random variable $X = f(Z_1, Z_2, \dots, Z_N)$ satisfies*

$$(4.3.15) \quad \mathbb{P}\left(|X - \mathbb{E}[X]| > t\right) \leq 2 \exp\left(-\frac{t^2}{2 \sum_{i=1}^N c_i^2}\right).$$

PROOF. This is Janson, Łuczak and Rucinski [186, Corollary 2.27]. \square

We apply Lemma 4.8 to $\sum_{v \in [n]} d_v(\pi)(d_v(\pi) - 1) = \sum_{v \in [n]} d_v(\pi)(d_v(\pi) - 1)$. Let $\mathbf{I} = (I_i)_{i \in [\ell_n]}$ denote the collection of Bernoulli variables needed to evaluate $(d_v(\pi))_{v \in [n]}$. This is the collection of independent random variables to which we apply Lemma 4.8. Write $f(\mathbf{I}) = \sum_{v \in [n]} d_v(\pi)(d_v(\pi) - 1)$, and compute that

$$(4.3.16) \quad \mathbb{E}[f(\mathbf{I})] = \sum_{v \in [n]} \mathbb{E}[d_v(\pi)(d_v(\pi) - 1)] = \pi \sum_{v \in [n]} d_v(d_v - 1) = n\pi \mathbb{E}[D_n(D_n - 1)],$$

since $\mathbb{E}[\text{Bin}(m, p)(\text{Bin}(m, p) - 1)] = m(m - 1)p^2$.

Further, let \mathbf{I}' be equal to \mathbf{I} except in the i th coordinate. Then,

$$(4.3.17) \quad |f(\mathbf{I}') - f(\mathbf{I})| \leq d_{v_i},$$

where v_i is the vertex to which the i th half-edge is incident, so that $c_i = d_{v_i}$. We compute that

$$(4.3.18) \quad \sum_{i \in [\ell_n]} c_i^2 = \sum_{v \in [n]} d_v^3 = n \mathbb{E}[D_n^3].$$

Therefore, we obtain

$$(4.3.19) \quad \mathbb{P}\left(|f(\mathbf{I}) - \mathbb{E}[f(\mathbf{I})]| > t\right) \leq 2 \exp\left(-\frac{t^2}{2 \sum_{v \in [n]} d_v^3}\right).$$

We now make a split depending on whether Condition 1.6(c) holds or not. When Condition 1.6(c) holds, $d_{\max} = \max_{v \in [n]} d_v = o(n^{1/2})$, so (4.3.19) implies that $f(\mathbf{I}) - \mathbb{E}[f(\mathbf{I})] = O_{\mathbb{P}}(n^{3/4})$, so that

$$(4.3.20) \quad \frac{f(\mathbf{I})}{n} = \frac{1}{n} \sum_{v \in [N]} d_v(\pi)(d_v(\pi) - 1) \xrightarrow{\mathbb{P}} \pi \mathbb{E}[D(D - 1)].$$

We continue with the case where Condition 1.6(c) does not hold. We use an appropriate truncation argument to deal with the few vertices of large degree. In this case, denote

$$(4.3.21) \quad f_2(\mathbf{I}) = \sum_{v \in [n]} d_v(\pi)(d_v(\pi) - 1) \mathbb{1}_{\{d_v \leq n^{1/2-\varepsilon}\}},$$

The above argument can be extended to show that, for any $\varepsilon > 0$,

$$(4.3.22) \quad \frac{f_2(\mathbf{I})}{\pi \sum_{v \in [n]} d_v(d_v - 1) \mathbb{1}_{\{d_v \leq n^{1/2-\varepsilon}\}}} \xrightarrow{\mathbb{P}} 1.$$

Further, for any ε and v with $d_v \geq n^{1/2-\varepsilon}$,

$$(4.3.23) \quad \mathbb{P}(|d_v(\pi) - \sqrt{\pi}d_v| \geq d_v^{(1+\varepsilon)/2}) \leq e^{-cn^\varepsilon},$$

so that even every percolated degree of vertices with high degrees are concentrated. Therefore, also

$$(4.3.24) \quad \frac{\sum_{v \in [n]} d_v(\pi)(d_v(\pi) - 1) \mathbb{1}_{\{d_v > n^{1/2-\varepsilon}\}}}{\pi \sum_{v \in [n]} d_v(d_v - 1) \mathbb{1}_{\{d_v > n^{1/2-\varepsilon}\}}} \xrightarrow{\mathbb{P}} 1.$$

EXERCISE 4.15 (Concentration binomial). *Use Lemma 4.8 to prove (4.3.23).*

EXERCISE 4.16 (Contribution of high-degree vertices). *Use a union bound and (4.3.23) to prove (4.3.24).*

We conclude again that

$$(4.3.25) \quad \frac{\sum_{v \in [n]} d_v(\pi)(d_v(\pi) - 1)}{\pi \sum_{v \in [n]} d_v(d_v - 1)} \xrightarrow{\mathbb{P}} 1,$$

so that

$$(4.3.26) \quad \begin{aligned} \nu_N(\pi) &= \frac{n \frac{1}{n} \sum_{v \in [N]} d_v(\pi)(d_v(\pi) - 1)}{N \frac{1}{N} \sum_{v \in [N]} d_v(\pi)} \\ &\xrightarrow{\mathbb{P}} \frac{1}{1 + \mathbb{E}[D](1 - \sqrt{\pi})} \pi \frac{\mathbb{E}[D(D-1)]}{\frac{\mathbb{E}[D]}{1 + \mathbb{E}[D](1 - \sqrt{\pi})}} = \pi/\nu, \end{aligned}$$

by combining (4.3.5), (4.3.13) and (4.3.20)-(4.3.25).

By Theorem 2.21, when $\pi > 1/\nu$, $\nu_N(\pi) \xrightarrow{\mathbb{P}} \pi/\nu > 1$, and $\text{CM}_N(\mathbf{d}(\pi))$ will have a giant component, while if $\pi \leq 1/\nu$, $\nu_N(\pi) \xrightarrow{\mathbb{P}} \pi/\nu \leq 1$, and $\text{CM}_N(\mathbf{d}(\pi))$ will not have a giant component. Since the giant component of $\text{CM}_n(\mathbf{d}; \pi)$ is always at most that of $\text{CM}_N(\mathbf{d}(\pi))$ this completes the proof of the subcritical results in Theorem 4.5(b).

We have now established the subcritical results in Theorem 4.5(b), and we move on to the supercritical results in Theorem 4.5(a). Having related percolation on $\text{CM}_n(\mathbf{d})$ with edge-retention probability π to $\text{CM}_N(\mathbf{d}(\pi))$, this proof proceeds in two key steps. In the first step, we analyze the giant component of $\text{CM}_N(\mathbf{d}(\pi))$ for $\pi > 1/\nu$. After this, we analyze the effect of removing the artificial red vertices, so as to obtain the size of the giant component of $\text{CM}_n(\mathbf{d}; \pi)$. We start with the giant of $\text{CM}_N(\mathbf{d}(\pi))$:

The size of the giant for $\text{CM}_N(\mathbf{d}(\pi))$ for $\pi > 1/\nu$. We apply Theorem 2.21 to $\text{CM}_N(\mathbf{d}(\pi))$, for which we use the asymptotic degree distribution in (4.3.11). Let $\mathcal{C}'_{\max}(\pi)$ denote the giant component of $\text{CM}_N(\mathbf{d}(\pi))$. By (2.3.8), for every $k \geq 0$,

$$(4.3.27) \quad v_k(\mathcal{C}'_{\max}(\pi))/N \xrightarrow{\mathbb{P}} p_k(\pi)(1 - \eta^*(\pi)^k),$$

where $\eta^*(\pi)$ satisfies

$$(4.3.28) \quad \eta^*(\pi) = \sum_{k \geq 0} (\eta^*(\pi))^k p_k^*(\pi),$$

with $p_k(\pi)$ defined on the right-hand side of (4.3.11), and

$$(4.3.29) \quad p_k^*(\pi) = \frac{(k+1)p_{k+1}(\pi)}{\sum_{\ell \geq 0} \ell p_\ell(\pi)}.$$

Let us analyze $\eta^*(\pi)$ in more detail. By (4.3.13),

$$(4.3.30) \quad \sum_{\ell \geq 0} \ell p_\ell(\pi) = \mathbb{E}[D(\pi)] = \frac{\mathbb{E}[D]}{1 + \mathbb{E}[D](1 - \sqrt{\pi})}.$$

Further, by (4.3.30) and (4.3.11)-(4.3.12),

$$(4.3.31) \quad \begin{aligned} \sum_{k \geq 0} \eta^k p_k^*(\pi) &= \frac{1}{\mathbb{E}[D(\pi)]} \sum_{k \geq 0} \eta^k (k+1) p_{k+1}(\pi) \\ &= (1 - \sqrt{\pi}) + \frac{1}{\mathbb{E}[D]} \eta^k (k+1) \sum_{\ell \geq k+1} p_\ell b_{k+1, \ell}(\sqrt{\pi}) \\ &= (1 - \sqrt{\pi}) + \sum_{\ell \geq 0} \frac{p_\ell}{\mathbb{E}[D]} \sum_{0 \leq k+1 \leq \ell} b_{k+1, \ell}(\sqrt{\pi}) \eta^k (k+1). \end{aligned}$$

Here, the first term is due to the first term in (4.3.12), and we again use (4.3.30). Note that

$$(4.3.32) \quad \begin{aligned} \sum_{0 \leq k+1 \leq \ell} b_{k+1, \ell}(\sqrt{\pi}) \eta^k (k+1) &= \sqrt{\pi} \sum_{0 \leq k \leq \ell-1} \binom{\ell}{k+1} (k+1) (\eta \sqrt{\pi})^k (1 - \sqrt{\pi})^{\ell-k-1} \\ &= \sqrt{\pi} \ell \sum_{0 \leq k \leq \ell-1} \binom{\ell-1}{k} (\eta \sqrt{\pi})^k (1 - \sqrt{\pi})^{\ell-k-1} \\ &= \sqrt{\pi} \ell (1 - \sqrt{\pi} + \eta \sqrt{\pi})^{\ell-1}. \end{aligned}$$

We conclude that

$$(4.3.33) \quad \begin{aligned} \sum_{k \geq 0} \eta^k p_k^*(\pi) &= (1 - \sqrt{\pi}) + \sqrt{\pi} \sum_{\ell \geq 0} \frac{\ell p_\ell}{\mathbb{E}[D]} (1 - \sqrt{\pi} + \eta \sqrt{\pi})^{\ell-1} \\ &= (1 - \sqrt{\pi}) + \sqrt{\pi} \sum_{\ell \geq 0} p_{\ell-1}^* (1 - \sqrt{\pi} + \eta \sqrt{\pi})^{\ell-1} \\ &= (1 - \sqrt{\pi}) + \sqrt{\pi} G_{D^*-1} (1 - \sqrt{\pi} + \eta \sqrt{\pi}), \end{aligned}$$

so that $\eta^*(\pi)$ solves (4.3.2). As a result,

$$(4.3.34) \quad \begin{aligned} \frac{|\mathcal{C}'_{\max}(\pi)|}{N} &\xrightarrow{\mathbb{P}} \zeta(\pi) = 1 - G_{D(\pi)}(\eta^*(\pi)) \\ &= 1 - \eta^*(\pi) \frac{\mathbb{E}[D](1 - \sqrt{\pi})}{1 + \mathbb{E}[D](1 - \sqrt{\pi})} - \frac{G_D(1 - \sqrt{\pi} + \eta^*(\pi)\sqrt{\pi})}{1 + \mathbb{E}[D](1 - \sqrt{\pi})}, \end{aligned}$$

where the last equality is proved in the following exercise:

EXERCISE 4.17 (Generating function $G_{D(\pi)}$). *Show that $G_{D(\pi)}(s) = G_D(1 - \sqrt{\pi} + s\sqrt{\pi})$.*

By Theorem 2.21, we conclude that

$$\begin{aligned}
\frac{|\mathcal{C}'_{\max}(\pi)|}{n} &= \frac{N}{n} \frac{|\mathcal{C}'_{\max}(\pi)|}{N} \\
&\xrightarrow{\mathbb{P}} (1 + \mathbb{E}[D](1 - \sqrt{\pi})) \left[1 - \eta^*(\pi) \frac{\mathbb{E}[D](1 - \sqrt{\pi})}{1 + \mathbb{E}[D](1 - \sqrt{\pi})} - \frac{G_D(1 - \sqrt{\pi} + \eta^*(\pi)\sqrt{\pi})}{1 + \mathbb{E}[D](1 - \sqrt{\pi})} \right] \\
(4.3.35) \quad &= 1 + (1 - \eta^*(\pi))\mathbb{E}[D](1 - \sqrt{\pi}) - G_D(1 - \sqrt{\pi} + \eta^*(\pi)\sqrt{\pi}).
\end{aligned}$$

Further, we conclude that $|\mathcal{C}'_{(2)}(\pi)|/n \xrightarrow{\mathbb{P}} 0$, where $\mathcal{C}'_{(2)}(\pi)$ denotes the second largest component of $\text{CM}_N(\mathbf{d}(\pi))$. Since clusters in $\text{CM}_N(\mathbf{d}(\pi))$ can only be larger than those in $\text{CM}_n(\mathbf{d}; \pi)$, this also proves that $|\mathcal{C}_{(2)}(\pi)|/n \xrightarrow{\mathbb{P}} 0$, as stated in Theorem 4.5(a). We conclude that $\mathcal{C}_{\max}(\pi)$ is obtained by removing the red vertices from $\mathcal{C}'_{\max}(\pi)$. In the next step, we derive the limit of $|\mathcal{C}_{\max}(\pi)|/n$, for which we need to investigate the effect of removing the red vertices.

The size of the giant for $\text{CM}_n(\mathbf{d}; \pi)$ for $\pi > 1/\nu$: removing red vertices. We write

$$(4.3.36) \quad \frac{|\mathcal{C}_{\max}(\pi)|}{n} = \frac{N}{n} \frac{|\mathcal{C}'_{\max}(\pi)|}{N} - \frac{\rho_n(\pi)}{n},$$

where $\rho_n(\pi)$ is the number of red vertices removed from $\mathcal{C}'_{\max}(\pi)$. We note that there are $Np_1(\pi)(1 + o_{\mathbb{P}}(1))$ degree-one vertices in the entire graph, and, by (4.3.27), $Np_1(\pi)(1 - \eta^*(\pi))(1 + o_{\mathbb{P}}(1))$ degree-one vertices in $\mathcal{C}'_{\max}(\pi)$. We need to remove $N - n = \mathbb{E}[D](1 - \sqrt{\pi})(1 + o_{\mathbb{P}}(1))$ degree-one vertices from the graph. By Remark 4.7, we can choose these degree-one vertices uniformly at random from the entire collection of degree-one vertices in the graph. As a result, and conditionally on N , $\rho_n(\pi)$ has a hypergeometric distribution with $N - n$ trials, $M = Np_1(\pi)(1 - \eta^*(\pi))$ objects of type 1 out of a total of $T = Np_1(\pi)$ objects. By concentration of the hypergeometric distribution,

$$(4.3.37) \quad \frac{\rho_n(\pi)}{n} = (N - n) \frac{M}{T} (1 + o_{\mathbb{P}}(1)) = \mathbb{E}[D](1 - \sqrt{\pi})(1 - \eta^*(\pi))(1 + o_{\mathbb{P}}(1)).$$

We arrive at

$$(4.3.38) \quad \frac{|\mathcal{C}_{\max}(\pi)|}{n} = \frac{N}{n} \frac{|\mathcal{C}'_{\max}(\pi)|}{N} - \frac{\rho_n(\pi)}{n} \xrightarrow{\mathbb{P}} 1 - G_D(1 - \sqrt{\pi} + \eta^*(\pi)\sqrt{\pi}),$$

as required. Similarly, since removing every red vertex removes a single edge,

$$(4.3.39) \quad \frac{E(\mathcal{C}_{\max}(\pi))}{n} = \frac{N}{n} \frac{E(\mathcal{C}'_{\max}(\pi))}{N} - \frac{\rho_n(\pi)}{n}.$$

By Theorem 2.21,

$$(4.3.40) \quad \frac{E(\mathcal{C}'_{\max}(\pi))}{N} \xrightarrow{\mathbb{P}} \frac{1}{2} \mathbb{E}[D(\pi)](1 - \eta^*(\pi)^2)$$

By (4.3.13),

$$\begin{aligned}
\frac{E(\mathcal{C}'_{\max}(\pi))}{n} &\xrightarrow{\mathbb{P}} \frac{1}{2} \mathbb{E}[D(\pi)](1 - \eta^*(\pi)^2)(1 + \mathbb{E}[D](1 - \sqrt{\pi})) \\
(4.3.41) \quad &= \frac{1}{2} \mathbb{E}[D](1 - \eta^*(\pi)^2).
\end{aligned}$$

We use (4.3.37) to conclude that

$$\begin{aligned}
 \frac{E(\mathcal{C}_{\max}(\pi))}{n} &= \frac{N}{n} \frac{E(\mathcal{C}'_{\max}(\pi))}{N} - \frac{\rho_n(\pi)}{n} \\
 &\xrightarrow{\mathbb{P}} \frac{1}{2} \mathbb{E}[D](1 - \eta^*(\pi)^2) - \mathbb{E}[D](1 - \sqrt{\pi})(1 - \eta^*(\pi)) \\
 (4.3.42) \quad &= \sqrt{\pi} \mathbb{E}[D](1 - \eta^*(\pi)) - \frac{1}{2} \mathbb{E}[D](1 - \eta^*(\pi))^2.
 \end{aligned}$$

□

EXERCISE 4.18 (Proof identity). *Prove the last identity in (4.3.42).*

Extensions. In the previous, we have considered *bond* percolation only. We could instead also focus on *site* percolation, where we remove all vertex (and all of their edges) independently and with probability π . In this case, the construction in Lemma 4.6 can be extended. Indeed, for site percolation, we explode each vertex v with probability π , and replace it by d_v red vertices having degree 1. Thus, in this setting, the number of red vertices equals $N^+ = N^+(\pi) = \sum_{v \in [n]} d_v [1 - \text{Ber}_v(\pi)]$, where $(\text{Ber}_v(\pi))_{v \in [n]}$ denote independent Bernoulli random variables with success probability π . Apart from this, little changes. You are asked to investigate this setting in the following exercises:

EXERCISE 4.19 (Site percolation). *Extend Theorem 4.5 to site percolation by proving that $|\mathcal{C}_{\max}^{(\text{si})}(\pi)|/n \xrightarrow{\mathbb{P}} \zeta^{(\text{si})}(\pi)$ for all π . Here $|\mathcal{C}_{\max}^{(\text{si})}(\pi)|$ is the size of the largest connected component after performing site percolation with probability π . Show that $\zeta^{(\text{si})}(\pi) > 0$ precisely when $\pi > 1/\nu$.*

EXERCISE 4.20 (Site percolation (cont.)). *Prove that $\zeta^{(\text{si})}(\pi) = \pi \zeta(\pi)$, where ζ is the limit for bond percolation in Theorem 4.5.*

4.4. Critical percolation on CMs with finite third-moment degrees

In this section, we study *critical percolation* on the configuration model, where we assume that the third moment of the degrees is finite. The results here are obtained in work with Dhara, van Leeuwen and Sen [103], and extend and improve various results obtained in the literature by Joseph [189] and Riordan [234].

For percolation to be critical, we need to take the edge-retention probability π to be sufficiently close to $\pi_c = 1/\nu$, and how close it needs to be is determined by the *scaling window*. Also, finite-size effects may play a role (as π_c might actually depend on n when tuning in to the critical behavior more closely). It turns out that the width of the scaling window and the size of the largest critical clusters are determined by the number of finite moments of the degree distribution. In Theorem 4.5, we have already seen that when $\nu = \infty$, percolation on the configuration model is always supercritical (see also Exercise 4.11). This can be understood by the fact that when we explore a cluster in the configuration model, the degrees of vertices that we encounter during the exploration asymptotically have the size-biased distribution D_n^* (recall Theorem 2.11). When $\mathbb{E}[D] < \infty$, yet $\mathbb{E}[D^2] = \infty$, we have that $\mathbb{E}[D_n^*] \rightarrow \infty$. Thus, the local weak limit is an infinite-mean branching process tree, for which the critical percolation parameter equals $\pi_c = 0$:

EXERCISE 4.21 (Robustness of the phase transition for infinite-mean branching processes). *Let \mathcal{T} be a Galton-Watson tree with offspring distribution X with $\mathbb{E}[X] = \infty$. Perform percolation on this tree. Show that $\pi_c = 0$, in the sense that whenever $\pi > 0$, percolation on \mathcal{T} always has (infinitely many) infinite components.*

When $\mathbb{E}[D^2] < \infty$ and $\mathbb{E}[D_n^*] \rightarrow \mathbb{E}[D^*] < \infty$, the critical nature of the branching process will be highly dependent on whether the *variance* of the offspring distribution is finite or not. This is closely related to the random walk representation of branching processes (recall Section 2.1.3), and the fact that the configuration model locally weakly converges to a unimodular Galton-Watson tree (recall Theorem 2.11). Recall the discussion around (4.1.5).

The variance of D_n^* converges precisely when the third moment of the degrees converges, which is what we will assume in this section. In the next section, we will treat the case where $\text{Var}(D_n^*) \rightarrow \infty$ and the degrees obey an asymptotic power law. Before stating our main result, we take an excursion into topological issues of weak convergence, which will be crucial in order to identify the scaling limit of the excursion of critical clusters for percolation on the configuration model. Before we state our formal result, let us intuitively argue that in this setting, the largest critical clusters are of order $n^{2/3}$.

Order of largest critical clusters. Let us start by giving an intuitive explanation of why critical clusters are of order $n^{2/3}$ when $\mathbb{E}[(D_n^*)^2] \rightarrow \mathbb{E}[(D^*)^2] < \infty$. By Janson's Construction in Section 4.3, critical percolation on the configuration model gives rise to a critical configuration model, so we might as well restrict the argument to this setting. This is also how the proof proceeds. Rather than computing the precise scaling limit as we will do below, let us merely aim to find the right order of magnitude of the largest critical clusters. Fix $k \geq 1$ and define the random variables

$$(4.4.1) \quad Z_{\geq k} = \sum_{v \in [n]} \mathbb{1}_{\{|\mathcal{C}_v| \geq k\}},$$

where we recall that $|\mathcal{C}_v|$ denotes the size of the connected component of $v \in [n]$ in $\text{CM}_n(\mathbf{d})$. We note that $|\mathcal{C}_{\max}| \geq k$ precisely when $Z_{\geq k} \geq k$. Thus, by Markov's inequality,

$$(4.4.2) \quad \mathbb{P}(|\mathcal{C}_{\max}| \geq k) = \mathbb{P}(Z_{\geq k} \geq k) \leq \frac{1}{k} \mathbb{E}[Z_{\geq k}] = \frac{n}{k} \mathbb{P}(|\mathcal{C}_U| \geq k),$$

where $U \in [n]$ is chosen uniformly at random. Recall the local weak convergence result in Theorem 2.11, and let $\mathcal{C}_o(\pi_c)$ denote the cluster of the root of the unimodular Galton-Watson tree of the local weak limit. We take a leap of faith to approximate $\mathbb{P}(|\mathcal{C}_U| \geq k) \approx \mathbb{P}(|\mathcal{C}_o(\pi_c)| > k)$ for all values of k of interest, which is allowed for *fixed* k by Theorem 2.11. However, we will apply this approximation for a $k = k_n$ that tends to infinity with n , which is clearly not what Theorem 2.11 allows.

Recall (4.1.5), which suggests that

$$(4.4.3) \quad \mathbb{P}(|\mathcal{C}_{\max}| \geq k) = \mathbb{P}(Z_{\geq k} \geq k) \leq \frac{1}{k} \mathbb{E}[Z_{\geq k}] = O\left(\frac{n}{k^{3/2}}\right).$$

Now we take $k = k_n = An^{2/3}$ for some large A (fingers crossed!), to arrive at

$$(4.4.4) \quad \mathbb{P}(|\mathcal{C}_{\max}| \geq An^{2/3}) = O\left(\frac{1}{A^{3/2}}\right),$$

which is small when $A > 0$ is large. A second moment argument can be used to suggest that

$$(4.4.5) \quad \mathbb{P}(|\mathcal{C}_{\max}| < n^{2/3}/A) = \mathbb{P}(Z_{\geq n^{2/3}/A} = 0) = O(1/A),$$

again under some unproven assumptions of independence of different clusters etc. [Such an argument was performed for in a rigorous way for critical rank-1 random graphs in [159], as well as for the critical Erdős-Rényi random graph in [160, Chapter 5], see also [157].]

This suggests that $|\mathcal{C}_{\max}|n^{-2/3}$ is a tight sequence of random variables, and suggests the anomalous scaling exponent $2/3$ at the critical percolation threshold π_c . This result closely resembles the scaling for Erdős-Rényi random graph as stated in Theorem 4.1, as well as highlights the role of the finite-third moment assumption of the degrees that was imperative to conclude (4.1.5). Of course, the above is no proof, as it has used (4.1.5) for $k = \Theta(n^{2/3})$, which is clearly not allowed. Below, we make this connection precise, not only by showing that indeed $n^{2/3}$ is the right order for the largest critical clusters, but even that the scaling limit of those clusters closely resembles that derived for Erdős-Rényi random graph in Theorem 4.1. In order to be able to state these results, we start with some notation.

Weak convergence of cluster sizes and surpluses. Recall the definition of ℓ_{\downarrow}^2 in Section 1.3.1. We extend this topology now to also include a strong notion of convergence for the surplus edges. By $\ell_{\downarrow}^2 \times \mathbb{N}^{\infty}$, we denote the product topology of ℓ_{\downarrow}^2 and \mathbb{N}^{∞} , where \mathbb{N}^{∞} denotes the collection of sequences on \mathbb{N} , endowed with the product topology. Define also

$$(4.4.6) \quad \mathbb{U}_{\downarrow} := \left\{ ((x_i, y_i))_{i=1}^{\infty} \in \ell_{\downarrow}^2 \times \mathbb{N}^{\infty} : \sum_{i=1}^{\infty} x_i y_i < \infty \text{ and } y_i = 0 \text{ whenever } x_i = 0 \forall i \right\},$$

endowed with the metric

$$(4.4.7) \quad d_{\mathbb{U}}((\mathbf{x}_1, \mathbf{y}_1), (\mathbf{x}_2, \mathbf{y}_2)) := \left(\sum_{i=1}^{\infty} (x_{1i} - x_{2i})^2 \right)^{1/2} + \sum_{i=1}^{\infty} |x_{1i} y_{1i} - x_{2i} y_{2i}|.$$

Further, we introduce $\mathbb{U}_{\downarrow}^0 \subset \mathbb{U}_{\downarrow}$ as

$$(4.4.8) \quad \mathbb{U}_{\downarrow}^0 := \left\{ ((x_i, y_i))_{i=1}^{\infty} \in \mathbb{U}_{\downarrow} : \text{if } x_k = x_m, k \leq m, \text{ then } y_k \geq y_m \right\}.$$

The $\mathbb{U}_{\downarrow}^0$ -topology will allow us to state joint convergence of cluster sizes and surpluses of clusters in a strong topology.

The critical window and finite-third moment degrees. Consider bond percolation on $\text{CM}_n(\mathbf{d})$ with probability π , yielding $\text{CM}_n(\mathbf{d}; \pi)$. Fix

$$(4.4.9) \quad \pi_n = \pi_n(\lambda) := \frac{1}{\nu_n} \left(1 + \frac{\lambda}{n^{1/3}} \right),$$

for some $\lambda \in \mathbb{R}$. Here we recall that $\nu_n = \mathbb{E}[D_n(D_n - 1)]/\mathbb{E}[D_n]$ as defined in (4.3.1) is close to $\pi_c = 1/\nu$, which identifies as being the asymptotic critical value in Theorem 4.5. The value π_n as in (4.4.9) rather than $\pi_c = 1/\nu$ takes the finite-size corrections of the critical behavior more closely into account. The value of $\lambda \in \mathbb{R}$ is similar to that for the Erdős-Rényi random graph in Theorem 4.2. Note that $\pi_n(\lambda)$ in (4.4.9) is non-negative for n sufficiently large even when $\lambda < 0$.

To describe the scaling limit, we again rely on Janson's construction. As in Section 4.3, suppose that $d_i(\pi_n) \sim \text{Bin}(d_i, \sqrt{\pi_n})$, $N^+ := \sum_{i \in [n]} (d_i - d_i(\pi_n))$ and $N = n + N^+$. Again consider the degree sequence $\mathbf{d}(\pi_n)$ consisting of $d_i(\pi_n)$ for $i \in [n]$ and N^+ additional vertices of degree 1, i.e., $d_i(\pi_n) = 1$ for $i \in [N] \setminus [n]$. We have already seen that the degree $D_N(\pi)$ of a random vertex from this degree sequence satisfies Conditions 1.6(a)-(c) in probability for some random variable $D(\pi)$ with $\mathbb{E}[D(\pi)^2] < \infty$. It is not hard to extend this to $D_N(\pi_n)$ with π_n as defined in (4.4.9):

EXERCISE 4.22 (Finite-size corrections have no effect on asymptotic degree distribution). *Verify that the argument in Section 4.3 showing that the random variable $D_N(\pi)$ satisfies Conditions 1.6(a)-(c) in probability can be extended to $D_N(\pi_n)$ for π_n in (4.4.9).*

We will later improve this to

$$(4.4.10) \quad \mathbb{E}_N[D_N(\pi_n)^3] = \frac{1}{N} \sum_{i \in [N]} d_i(\pi)^3 \xrightarrow{\mathbb{P}} \mathbb{E}[D(\pi)^3] < \infty,$$

when $\mathbb{E}[D_n^3] \rightarrow \mathbb{E}[D^3]$ holds.

The main result in this section will investigate the scaling limit of the ordered clusters of percolation on $\text{CM}_n(\mathbf{d})$ with edge-retention probability π_n in (4.4.9). Let us now define the objects whose scaling limit we consider in detail. Recall that $(|\mathcal{C}_{(j)}(\pi)|)_{j \geq 1}$ denote the ordered cluster sizes of $\text{CM}_n(\mathbf{d}; \pi)$. Let $\text{SP}(\mathcal{C})$ denote the surplus of the connected component \mathcal{C} . The objects of interest in this section are given by

$$(4.4.11) \quad \mathbf{Z}_n(\lambda) = \left((n^{-2/3} |\mathcal{C}_{(j)}(\pi_n)|, \text{SP}(\mathcal{C}_{(j)}(\pi_n))) \right)_{j \geq 1},$$

which denotes the vector in \mathbb{U}_\downarrow^0 obtained by rearranging critical percolation clusters according to the cluster sizes and their corresponding surplus edges.

Let us continue by defining the necessary objects so as to be able to describe the scaling limit of $\mathbf{Z}_n(\lambda)$. Define the Brownian morion with negative parabolic drift by

$$(4.4.12) \quad B_{\mu, \eta}^\lambda(s) = \frac{\sqrt{\eta}}{\mu} B(s) + \lambda s - \frac{\eta s^2}{2\mu^3},$$

where $\mathbf{B} = (B(s))_{s \geq 0}$ is a standard Brownian motion, and $\mu > 0$, $\eta > 0$ and $\lambda \in \mathbb{R}$ are constants to be determined later on. Define the reflected version of $\mathbf{B}_{\mu, \eta}^\lambda$ as

$$(4.4.13) \quad R_{\mu, \eta}^\lambda(s) = B_{\mu, \eta}^\lambda(s) - \min_{0 \leq t \leq s} B_{\mu, \eta}^\lambda(t).$$

In terms of the above quantities, define γ_j^λ to be the ordered excursions of the inhomogeneous Brownian motion $\mathbf{B}_{\mu, \eta}^\lambda$ with the parameters

$$(4.4.14) \quad \mu = \mathbb{E}[D(\pi_c)], \quad \eta = \mathbb{E}[D(\pi_c)^3] \mathbb{E}[D(\pi_c)] - \mathbb{E}[D(\pi_c)^2]^2, \quad \beta = 1/\mathbb{E}[D(\pi_c)],$$

where $\pi_c = 1/\nu$ is the asymptotic critical value as defined in Theorem 4.5. Denote the j th largest cluster of $\text{CM}_n(\mathbf{d}; \pi_n(\lambda))$ by $\mathcal{C}_{(j)}(\lambda)$. We will also identify the scaling limit of the *surplus edges* of the large critical clusters. For this, define the counting process of marks $\mathbf{N}^\lambda = (N^\lambda(s))_{s \geq 0}$ to be a process that has intensity $\beta R_{\mu, \eta}^\lambda(s)$ at time s conditional

on $(R_{\mu,\eta}^\lambda(u))_{u \leq s}$, so that

$$(4.4.15) \quad N_{\mu,\eta}^\lambda(s) - \int_0^s \beta R_{\mu,\eta}^\lambda(u) du$$

is a martingale (see Aldous [9]). Alternatively, we can define \mathbf{N}^λ to be a mixed-Poisson process where $N^\lambda(s)$ is Poisson with parameter $\int_0^s R_{\mu,\eta}^\lambda(u) du$.

For an excursion γ starting at time $l(\gamma)$ and ending at time $r(\gamma)$, let $N(\gamma)$ denote the number of marks in the interval $[l(\gamma), r(\gamma)]$. Then, $((|\gamma_j^\lambda|, N(\gamma_j^\lambda)))_{j \geq 1}$ can be ordered as an element of \mathbb{U}_\downarrow^0 almost surely by Bhamidi, Budhiraja and Wang [36, Theorem 4.1]. Denote this element of \mathbb{U}_\downarrow^0 by $\mathbf{Z}(\lambda) = ((Y_j^\lambda, N_j^\lambda))_{j \geq 1}$ obtained from $((|\gamma_j^\lambda|, N(\gamma_j^\lambda)))_{j \geq 1}$. This will be the description of the scaling limit of our cluster sizes and their surpluses. The main result of this section is then the following weak convergence result:

THEOREM 4.9 (Critical clusters for percolation on configuration models with finite third-moment degrees). *Assume that Conditions 1.6(a)-(c) hold for the degree sequence and that $\mathbb{E}[D_n^3] \rightarrow \mathbb{E}[D^3] < \infty$. Further, assume that $\text{CM}_n(\mathbf{d})$ is super-critical, i.e., $\nu_n \rightarrow \nu > 1$. Fix $\pi_n(\lambda)$ as in (4.4.9). Then,*

$$(4.4.16) \quad \mathbf{Z}_n(\lambda) \xrightarrow{d} \mathbf{Z}(\lambda)$$

with respect to the \mathbb{U}_\downarrow^0 topology.

Next we consider the percolation cluster for multiple values of λ . There is a natural way to couple $(\text{CM}_n(\mathbf{d}; \pi_n(\lambda)))_{\lambda \in \mathbb{R}}$ described as follows: for $\lambda < \lambda'$, perform bond-percolation on $\text{CM}_n(\mathbf{d}, \pi_n(\lambda'))$ with probability $\pi_n(\lambda)/\pi_n(\lambda')$. The resulting graph is distributed as $\text{CM}_n(\mathbf{d}, \pi_n(\lambda))$. This can be used to couple $(\text{CM}_n(\mathbf{d}, \pi_n(\lambda_i)))_{i=0}^{k-1}$ for any fixed $k \geq 1$. The next theorem shows that the convergence of the component sizes holds jointly in finitely many locations within the critical window, under the above described coupling:

THEOREM 4.10 (Multiple times convergence). *Under the hypotheses of Theorem 4.9 and with $\mathbf{C}_n(\lambda) = (n^{-2/3} |\mathcal{C}_{(j)}(\lambda)|)_{j \geq 1}$, for any fixed $k \in \mathbb{N}$ and $-\infty < \lambda_0 < \lambda_1 < \dots < \lambda_{k-1} < \infty$,*

$$(4.4.17) \quad (\mathbf{C}_n(\lambda_0), \mathbf{C}_n(\lambda_1), \dots, \mathbf{C}_n(\lambda_{k-1})) \xrightarrow{d} \sqrt{\pi}(\gamma^{\lambda_0}, \gamma^{\lambda_1}, \dots, \gamma^{\lambda_{k-1}})$$

with respect to the $(\ell_\downarrow^2)^k$ topology where $\pi = 1/\nu$.

REMARK 4.11 (Informal relation to multiplicative coalescent). The coupling for the limiting process in Theorem 4.10 is given by the multiplicative coalescent process described in Section 4.2. This will become more clear when we describe the ideas of the proof. An intuitive picture is that as we change the value of the percolation parameter from $\pi_n(\lambda)$ to $\pi_n(\lambda + d\lambda)$, exactly one edge is added to the graph and the two endpoints i, j are chosen approximately proportional to the number of half-edges of i and j that were not retained in $\text{CM}_n(\mathbf{d}; \pi_n(\lambda))$. Define the *degree deficiency* \mathcal{D}_i of a component \mathcal{C}_i to be the total number of half-edges in a component that were not retained in percolation. Think of \mathcal{D}_i as the mass of \mathcal{C}_i . By the above heuristics, \mathcal{C}_i and \mathcal{C}_j merge at rate proportional to $\mathcal{D}_i \mathcal{D}_j$ and creates a cluster of mass $\mathcal{D}_i + \mathcal{D}_j - 2$. The proof with Dhara, van Leeuwen and Sen [103] shows that the degree deficiency of a component is approximately proportional

to the component size. Therefore, the component sizes merge *approximately* like the multiplicative coalescent over the critical scaling window.

REMARK 4.12 (Conditioning on simplicity). Theorems 4.9 and 4.10 also hold for configuration models conditioned on simplicity. We do not give a proof here. The arguments by Joseph in [189, Section 7] can be followed verbatim to obtain a proof of this fact. We do not give the details of this argument.

Discussion of the proofs of Theorems 4.9 and 4.10. We will not prove Theorem 4.10 in detail, but rather informally discuss how it can be obtained from Theorem 4.9 by using the coupling of the graphs $(\text{CM}_n(\mathbf{d}, \pi_n(\lambda_i))_{i=0}^{k-1})$. The proof of Theorem 4.9 follows by again using Janson's Construction in Lemma 4.6. This proof, alike the proof of Theorem 4.5, consists of three key steps. The first key step is the statement and proof of the corresponding result for the configuration model *without* percolation under appropriate conditions on the degree sequence. The second key step is to show that the percolated degree sequence in Janson's construction satisfies the conditions posed in the theorem for the critical configuration model. The final and third step is to describe the effect of removing the red vertices. The proof of the statement for the configuration model, as stated in Theorem 4.13 below, will be the main content of this section.

Cluster scaling for critical configuration models. Let us start by stating the results on the cluster distributions for critical configuration models and its conditions. In its statement, we let \mathbf{Z}_n denote the rescaled cluster sizes ordered in size, and the corresponding surpluses of these clusters, as defined in (4.4.11) and in the absence of percolation (i.e., $\pi = 1$). While there will be a parameter λ appearing in the result, we do not write its dependence explicitly to avoid confusion with $\mathbf{Z}_n(\lambda)$ and $\mathbf{Z}(\lambda)$ in Theorem 4.9, where λ appears as in (4.4.9). Let $\sigma_r = \mathbb{E}[D^r]$ and consider the reflected Brownian motion, the excursions, and the counting process \mathbf{N}^λ as defined in (4.4.12) with parameters

$$(4.4.18) \quad \mu := \sigma_1, \quad \eta := \sigma_3\mu - \sigma_2^2, \quad \beta := 1/\mu.$$

Note that (4.4.18) reduces to (4.4.14) when $\pi = 1$ (for which $D(\pi) = D$). The main result is as follows:

THEOREM 4.13 (Critical configuration models with finite third-moment degrees). *Assume that Conditions 1.6(a)-(c) hold for the degree sequence with $p_2 \in [0, 1)$, and that $\mathbb{E}[D_n^3] \rightarrow \mathbb{E}[D^3] < \infty$. Further, assume that $\text{CM}_n(\mathbf{d})$ is critical, i.e., for some $\lambda \in \mathbb{R}$,*

$$(4.4.19) \quad \nu_n := \frac{\mathbb{E}[D_n(D_n - 1)]}{\mathbb{E}[D_n]} = 1 + \lambda n^{-1/3} + o(n^{-1/3}),$$

Then,

$$(4.4.20) \quad \mathbf{Z}_n \xrightarrow{d} \mathbf{Z}$$

with respect to the \mathbb{U}_\downarrow^0 topology. Further, for every $i, k \geq 1$,

$$(4.4.21) \quad \frac{v_k(\mathcal{C}_{(i)}(\pi_n(\lambda)))}{|\mathcal{C}_{(i)}(\pi_n(\lambda))|} \xrightarrow{\mathbb{P}} p_k^*.$$

The criticality condition in (4.4.19) is close in spirit to that for percolation on $\text{CM}_n(\mathbf{d})$ in (4.4.9), apart from the fact that now the $\text{CM}_n(\mathbf{d})$ is critical itself, rather than that it is supercritical and becomes critical due to percolation acting on it.

Theorem 4.13 is interesting in its own right, as it describes the critical behavior of configuration models. In the light of Janson’s Construction for percolation on the configuration model, one could also see Theorem 4.9 as a way to create, from super-critical configuration models, degrees sequences that take a specific value inside the scaling window. Aside from using percolation, this is not so easy. Further, percolation on the configuration model also allows one to see how the critical clusters emerge, by studying the behavior of $\text{CM}_n(\mathbf{d}; \pi_n(\lambda))$ when λ varies, as in Theorem 4.10.

We continue as follows. In Section 4.4.1, we prove Theorem 4.13. In Section 4.4.2, we show that Theorem 4.9 follows from Theorem 4.13.

4.4.1. Critical configuration models with finite-third moment degrees: proof of Theorem 4.13. In this section, we prove Theorem 4.13. We start by describing a convenient way to *explore* clusters in the configuration model.

The configuration model exploration process. Let us explore the graph sequentially using a natural approach outlined by Riordan [234] and also used in the work with Dhara, van Leeuwen and Sen [103]. At step k , divide the set of half-edges into three groups: sleeping half-edges \mathcal{S}_k , active half-edges \mathcal{A}_k , and dead half-edges \mathcal{D}_k . The depth-first exploration process can be summarized in the following algorithm. In its statement, the active half-edges in \mathcal{A}_k will be ordered, and this ordering will be updated as time proceeds:

ALGORITHM 4.14 (DFS exploration). At $k = 0$, \mathcal{S}_k contains all the half-edges and $\mathcal{A}_k, \mathcal{D}_k$ are empty. While ($\mathcal{S}_k \neq \emptyset$ or $\mathcal{A}_k \neq \emptyset$) we do the following at stage $k + 1$:

- (S1) If $\mathcal{A}_k \neq \emptyset$, then take the smallest half-edge a from \mathcal{A}_k .
- (S2) Take the half-edge b from \mathcal{S}_k that is paired to a . Suppose b is attached to a vertex w (which is necessarily not discovered yet). Declare w to be discovered, let $r = d_w - 1$ and $b_{w1}, b_{w2}, \dots, b_{wr}$ be the half-edges of w other than b . Declare $b_{w1}, b_{w2}, \dots, b_{wr}, b$ to be smaller than all other half-edges in \mathcal{A}_k . Also order the half-edges of w among themselves as $b_{w1} > b_{w2} > \dots > b_{wr} > b$. Now identify $\mathcal{B}_k \subset \mathcal{A}_k \cup \{b_{w1}, b_{w2}, \dots, b_{wr}\}$ as the collection of all half-edges in \mathcal{A}_k paired to one of the b_{wi} ’s and the corresponding b_{wi} ’s. Similarly identify $\mathcal{C}_k \subset \{b_{w1}, b_{w2}, \dots, b_{wr}\}$ as the collection of self-loops incident to w . Finally, declare $\mathcal{A}_{k+1} = \mathcal{A}_k \cup \{b_{w1}, b_{w2}, \dots, b_{wr}\} \setminus (\mathcal{B}_k \cup \mathcal{C}_k)$, $\mathcal{D}_{k+1} = \mathcal{D}_k \cup \{a, b\} \cup \mathcal{B}_k \cup \mathcal{C}_k$ and $\mathcal{S}_{k+1} = \mathcal{S}_k \setminus (\{b\} \cup \{b_{w1}, b_{w2}, \dots, b_{wr}\})$. Go to stage $k + 2$.
- (S3) If $\mathcal{A}_k = \emptyset$ for some k , then take out one half-edge a from \mathcal{S}_k uniformly at random and identify the vertex v incident to it. Declare v to be discovered. Let $r = d_v - 1$ and assume that $a_{v1}, a_{v2}, \dots, a_{vr}$ are the half-edges of v other than a and identify the collection of half-edges involved in a self-loop \mathcal{C}_k as in Step 2 (since we now explore a new cluster, there cannot yet be any multiple edges). Order the half-edges of v as $a_{v1} > a_{v2} > \dots > a_{vr} > a$. Set $\mathcal{A}_{k+1} = \{a, a_{v1}, a_{v2}, \dots, a_{vr}\} \setminus \mathcal{C}_k$, $\mathcal{D}_{k+1} = \mathcal{D}_k \cup \mathcal{C}_k$, and $\mathcal{S}_{k+1} = \mathcal{S}_k \setminus \{a, a_{v1}, a_{v2}, \dots, a_{vr}\}$. Go to stage $k + 2$.

In words, we explore a new vertex at each stage and throw away all the half-edges involved in a loop/multiple edge/cycle with the vertex set already discovered before proceeding to the next stage. The ordering of the half-edges is such that the connected components of $\text{CM}_n(\mathbf{d})$ are explored in the depth-first way. We call the half-edges of $\mathcal{B}_k \cup \mathcal{C}_k$ *cycle* half-edges because they create loops, cycles or multiple edges in the graph. Let

$$(4.4.22) \quad A_k := |\mathcal{A}_k|, \quad c_{(k+1)} := (|\mathcal{B}_k| + |\mathcal{C}_k|)/2, \quad U_k := |\mathcal{S}_k|.$$

Let $d_{(j)}$ be the degree of the j th explored vertex and define the exploration process $(S_n(i))_{i \geq 0}$ by $S_n(0) = 0$ and

$$(4.4.23) \quad S_n(i) = \sum_{j=1}^i (d_{(j)} - 2 - 2c_{(j)}).$$

The exploration in (4.4.23) is highly similar to that for the Erdős-Rényi random graph in (4.2.5), apart from the fact that $(S_n(i))_{i \geq 0}$ records active half-edges whereas its Erdős-Rényi random graph counterpart records active vertices. It is still the case that connected components are explored between times that $(S_n(i))_{i \geq 0}$ attains a novel minimum. Due to the fact that we filter out the cycles, each time we perform a step in Algorithm 4.14 we find a *new* vertex. Critical components contain few cycles, as we will prove later on, so that in (4.4.23), we can ignore the cycle contribution $c_{(j)}$ in the leading order. By construction, the sequence $(d_{(j)})_{j \geq 1}$ are the degrees in a size-biased reordering of the vertices, for which Proposition 2.42 will prove to be extremely useful.

Let us make the above observations more precise and make a start with the proof. The process $\mathbf{S}_n = (S_n(i))_{i \in [n]}$ “encodes the component sizes as lengths of path segments above past minima” as discussed by Aldous [9]. Suppose $\tilde{\mathcal{C}}_k$ is the k th connected component explored by the above exploration process.³ Define

$$(4.4.24) \quad \sigma_k = \inf \{i : S_n(i) = -2k\}.$$

Then $\tilde{\mathcal{C}}_k$ is discovered between the times $\sigma_{k-1} + 1$ and σ_k and

$$(4.4.25) \quad |\tilde{\mathcal{C}}_k| = \sigma_k - \sigma_{k-1}.$$

Here we note that since we throw away all the half-edges involved in a loop/multiple edge/cycle, each time we pair a half-edge, we connect it to a *new* vertex. The following exercise asks you to prove the crucial identity (4.4.25) in some special cases:

EXERCISE 4.23 (Cluster exploration). *Let \mathcal{C} be the graph consisting of one vertex of degree 2 with a self-loop. Show that $S_n(0) = 0$ and $S_n(1) = -2$, so that (4.4.25) is correct.*

EXERCISE 4.24 (Cluster exploration (cont.)). *Let \mathcal{C} be the graph consisting of two vertices of degree 1 with one edge in between them. Show that $S_n(0) = 0$, $S_n(1) = -1$ and $S_n(2) = -2$, so that (4.4.25) is correct.*

EXERCISE 4.25 (Cluster exploration (cont.)). *Let \mathcal{C} be the graph consisting of one vertex of degree 2 connected to two vertices of degree 1. Show that $S_n(0) = S_n(1) = 0$, $S_n(2) = -1$ and $S_n(3) = -2$, so that (4.4.25) is correct.*

³We avoid the notation \mathcal{C}_k for the k th cluster, as \mathcal{C}_k has previously been used to denote the cluster of vertex $k \in [n]$.

EXERCISE 4.26 (Cluster exploration (cont.)). *Prove (4.4.25) for the exploration of the first cluster. Conclude that it is true for any cluster.*

Key ingredient proof of Theorem 4.13: convergence of cluster exploration.

The key ingredient in the proof of Theorem 4.13 is the statement that the cluster exploration process \mathbf{S}_n , after appropriate rescaling, converges in distribution to the Brownian motion with negative parabolic drift as defined in (4.4.12) with the parameters as in (4.4.18). Let us set up the notation for this. For any $\mathbb{D}[0, \infty)$ -valued process \mathbf{X}_n define

$$(4.4.26) \quad \bar{X}_n(u) := n^{-1/3} X_n(\lfloor n^{2/3} u \rfloor) \quad \text{and} \quad \bar{\mathbf{X}}_n := (\bar{X}_n(u))_{u \geq 0}.$$

The following result is the main ingredient for proving Theorem 4.13. Recall the definition of $\mathbf{B}_{\mu, \eta}^\lambda$ from (4.4.12) with parameters given in (4.4.18):

THEOREM 4.15 (Convergence of the exploration process). *Suppose that Conditions 1.6(a)-(c) hold, that $\mathbb{E}[D_n^3] \rightarrow \mathbb{E}[D^3] < \infty$ and that the criticality condition (4.4.19) holds for ν_n . Then, as $n \rightarrow \infty$,*

$$(4.4.27) \quad \bar{\mathbf{S}}_n \xrightarrow{d} \mathbf{B}_{\mu, \eta}^\lambda$$

with respect to the Skorokhod J_1 topology.

In (4.4.23), we see that the size-biased degrees appear prominently. We will ignore the contributions due to cycles, and focus on $s_n(i) = \sum_{j=1}^i (d_{(j)} - 2)$ for the time being. We note that

$$(4.4.28) \quad \mathbb{E}[d_{(1)} - 2] = \sum_{v \in [n]} (d_v - 2) \frac{d_v}{\ell_n} = \nu_n - 1.$$

Thus, since $\nu_n = 1 + \lambda n^{-1/3} + o(n^{-1/3})$ by (4.4.19), the summands in $s_n(i) = \sum_{j=1}^i (d_{(j)} - 2)$ are close to having mean zero. However, $\mathbb{E}[d_{(j)} - 2]$ can be expected to decrease slowly when j increases since vertices with high degrees are more likely to have been chosen early on, so that the remaining vertices has somewhat small degrees on average. This makes the mean of the summands more and more deviating from zero. Bear in mind that we take j to be of the order $n^{2/3}$ (recall (4.4.26)), so this effect could be pronounced. In the next section, we investigate this effect precisely, after which we can explain Theorem 4.15 informally.

Size-biased exploration. The vertices are explored in a size-biased manner with sizes proportional to their degrees, i.e., if $v_{(i)}$ denotes the i th explored vertex in Algorithm 4.14 and $d_{(i)}$ the degree of $v_{(i)}$, then

$$(4.4.29) \quad \mathbb{P}(v_{(i)} = j \mid v_{(1)}, v_{(2)}, \dots, v_{(i-1)}) = \frac{d_j}{\sum_{k \notin \mathcal{V}_{i-1}} d_k} = \frac{d_j}{\sum_{k \in [n]} d_k - \sum_{k=1}^{i-1} d_{(k)}}, \quad \forall j \in \mathcal{V}_{i-1},$$

where \mathcal{V}_i denotes the first i vertices to be discovered in the above exploration process. The following lemma will be used in the proof of Theorem 4.13:

LEMMA 4.16 (Sums of size-biased degrees). *Suppose that Conditions 1.6(a)-(c) hold, that $\mathbb{E}[D_n^3] \rightarrow \mathbb{E}[D^3] < \infty$ and that the criticality condition (4.4.19) holds for ν_n . Denote*

$\sigma_r = \mathbb{E}[D^r]$ and $\mu = \mathbb{E}[D]$. Then for all $t > 0$ and $r \in \{1, 2\}$, as $n \rightarrow \infty$,

$$(4.4.30) \quad \sup_{u \leq t} \left| n^{-2/3} \sum_{i=1}^{\lfloor n^{2/3}u \rfloor} d_{(i)}^r - \frac{\sigma_{1+ru}}{\mu} \right| \xrightarrow{\mathbb{P}} 0.$$

Lemma 4.16 follows from Proposition 2.42 and some standard estimates that we state now for completeness:

LEMMA 4.17. *Conditions 1.6(a)-(c) and $\mathbb{E}[D_n^3] \rightarrow \mathbb{E}[D^3] < \infty$ imply that*

$$(4.4.31) \quad \lim_{k \rightarrow \infty} \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j \in [n]} \mathbb{1}_{\{d_j > k\}} d_j^r = 0, \quad r = 1, 2, 3.$$

For $r = 3$, in particular, this implies $d_{\max}^3 = o(n)$.

Intuitive explanation of Theorem 4.15. Before launching into the proof, we give an informal explanation of the scaling limit result in Theorem 4.15. We let, with $\mathcal{F}_j = \sigma(v_{(1)}, v_{(2)}, \dots, v_{(j)})$ denoting the σ -algebra generated by the first j vertices found,

$$(4.4.32) \quad \nu_n(j) = \mathbb{E}[d_{(j)} - 1 \mid v_{(1)}, v_{(2)}, \dots, v_{(j-1)}].$$

We then split

$$(4.4.33) \quad s_n(i) = \sum_{j=1}^i (d_{(j)} - 2) = s_n(i) = \sum_{j=1}^i (d_{(j)} - 1 - \nu_n(j)) + \sum_{j=1}^i (\nu_n(j) - 1).$$

Now, the random variables $(d_{(j)} - 1 - \nu_n(j))_{j \geq 1}$ all have mean zero (by construction), and, by Lemma 4.16, their variance converges to some limiting value that we will denote by η . We investigate the rescaled version of $(s_n(j))_{j \geq 1}$ given by

$$(4.4.34) \quad n^{-1/3} s_n(un^{2/3}) = n^{-1/3} \sum_{j=1}^{un^{2/3}} (d_{(j)} - 1 - \nu_n(j)) + n^{-1/3} \sum_{j=1}^{un^{2/3}} (\nu_n(j) - 1),$$

where one can expect the first part to converge in distribution to $B_u \sqrt{\eta}/\mu$, where $(B_u)_{u \geq 0}$ is a standard Brownian motion. Here we can informally compute the variance, also using that $\nu_n \approx 1$, as

$$(4.4.35) \quad \begin{aligned} \text{Var}(d_{(j)} - 1 - \nu_n(j)) &= \mathbb{E}[(d_{(j)} - 1 - \nu_n(j))^2] \approx \mathbb{E}[(d_{(1)} - 1 - \nu_n)^2] \\ &\approx \mathbb{E}[(d_{(1)} - 2)^2] = \frac{1}{\ell_n} \sum_{v \in [n]} (d_v - 2)^2 = \frac{\mathbb{E}[D_n(D_n - 2)^2]}{\mathbb{E}[D_n]} \\ &\rightarrow \frac{\mathbb{E}[D(D - 2)^2]}{\mathbb{E}[D]} = \frac{\eta}{\mu^2}, \end{aligned}$$

when we use that

$$(4.4.36) \quad \mathbb{E}[D]\mathbb{E}[D(D - 2)^2] = \mathbb{E}[D]\mathbb{E}[D^2(D - 2)] = \mu\sigma_3 - 2\sigma_2\mu = \mu\sigma_3 - \sigma_2^2,$$

since $\nu = 1$ implies that $\sigma_2/\mu = \nu + 1 = 2$.

We are left to deal with the second term in (4.4.34), for which we note that $\nu_n(j) - 1$ is random with a non-zero mean. Any contribution to $\nu_n(j) - 1$ that is of order $n^{-1/3}$ will

contribute in the limit in (4.4.34), by the factor $n^{-1/3}$ and since there are $n^{2/3}$ terms in the sum. Writing $\nu_n(j) - 1 = \nu_n(j) - \nu_n + \nu_n - 1$ and using (4.4.29), we obtain

$$\nu_n(j) - 1 = \frac{\sum_{k \in [n]} d_k(d_k - 2) - \sum_{k=1}^{j-1} d_{(k)}(d_{(k)} - 2)}{\sum_{k \in [n]} d_k - \sum_{k=1}^{j-1} d_{(k)}} - \frac{\sum_{k \in [n]} d_k(d_k - 2)}{\sum_{k \in [n]} d_k} + \lambda n^{-1/3} + o(1),$$

where we use (4.4.19). It turns out that we can ignore the difference in the denominators (see also Exercise 4.27 below), so that

$$(4.4.37) \quad \nu_n(j) - 1 = -\frac{\sum_{k=1}^{j-1} d_{(k)}(d_{(k)} - 2)}{n\mathbb{E}[D_n]} + \lambda n^{-1/3} + o(1)$$

By Lemma 4.16 and using that $\mathbb{E}[D_n] = \mathbb{E}[D] + o(1)$,

$$\nu_n(sn^{2/3}) - 1 = n^{-1/3} \left(-\frac{s(\sigma_3 - 2\sigma_2)}{\mu^2} + \lambda + o_{\mathbb{P}}(1) \right).$$

We conclude that

$$(4.4.38) \quad n^{-1/3} \sum_{j=1}^{un^{2/3}} (\nu_n(j) - 1) = -\frac{u^2(\sigma_3 - 2\sigma_2)}{2\mu^2} + \lambda u.$$

It is not hard to show that the constant can be rewritten as $(\sigma_3 - 2\sigma_2)/\mu^2 = \eta/\mu^3$, where η is defined in (4.4.18). This suggests that

$$(4.4.39) \quad \begin{aligned} n^{-1/3} s_n(un^{2/3}) &\approx \frac{\sqrt{\eta}}{\mu} B_u + \int_0^u \left(-\frac{s(\sigma_3 - 2\sigma_2)}{\mu^2} + \lambda \right) ds \\ &= \frac{\sqrt{\eta}}{\mu} B_u + \lambda u - \frac{(\sigma_3 - 2\sigma_2)}{2\mu^2} u^2 = B_{\mu, \eta}^\lambda(u), \end{aligned}$$

by (4.4.18). Apart from the variance computation, this explains how the constants in (4.4.18) arise, as well as the importance of the finite third moment of the degree distribution.

EXERCISE 4.27 (Sharper version of (4.4.38)). *Argue that (4.4.38) remains to be true when we do not ignore the difference in the denominators in (4.4.37). [Hint: Use that $\nu_n = 1 + o(1)$, so that $\sum_{v \in [n]} d_v(d_v - 2) = o(n)$.]*

Estimating the cycle half-edges. The following lemma gives an estimate of the number of cycle half-edges created up to time t . This result is proved by Riordan in [234] for bounded degrees. In our case, it follows from Lemma 4.16 as we show below:

LEMMA 4.18 (Estimates on self-loops and cycles). *Consider Algorithm 4.14, and recall from (4.4.22) that $A_k = |\mathcal{A}_k|$ and $U_k = |\mathcal{S}_k|$. Define further $B_k := |\mathcal{B}_k|$ and $C_k := |\mathcal{C}_k|$. Then*

$$(4.4.40) \quad \mathbb{E}[B_k | \mathcal{F}_k] = (1 + o_{\mathbb{P}}(1)) \frac{2A_k}{U_k} + O_{\mathbb{P}}(n^{-2/3})$$

and

$$(4.4.41) \quad \mathbb{E}[C_k | \mathcal{F}_k] = O_{\mathbb{P}}(n^{-1})$$

uniformly for $k \leq tn^{2/3}$ and any $t > 0$, where \mathcal{F}_k is the sigma-field generated by the information revealed up to stage k .

PROOF. First note that, by (4.4.30),

$$(4.4.42) \quad \frac{U_k}{n} = \frac{1}{n} \sum_{j \in [n]} d_j - \frac{1}{n} \sum_{j=1}^k d_{(j)} = \mathbb{E}[D] + o_{\mathbb{P}}(1)$$

uniformly over $k \leq tn^{2/3}$. Let a be the half-edge that is being explored at stage $k+1$. Each of the $(A_k - 1)$ half-edges of $\mathcal{A}_k \setminus \{a\}$ is equally likely to be paired with a half-edge of $v_{(k+1)}$, thus creating two elements of \mathcal{B}_k . Also, given \mathcal{F}_k and $v_{(k+1)}$, the probability that a half-edge of $\mathcal{A}_k \setminus \{a\}$ is paired to one of the half-edges of $v_{(k+1)}$ is $(d_{(k+1)} - 1)/(U_k - 1)$. Therefore,

$$(4.4.43) \quad \mathbb{E}[B_k \mid \mathcal{F}_k, v_{(k+1)}] = 2(A_k - 1) \frac{d_{(k+1)} - 1}{U_k - 1} = 2(d_{(k+1)} - 1) \frac{A_k}{U_k - 1} - 2 \frac{d_{(k+1)} - 1}{U_k - 1}.$$

Hence,

$$(4.4.44) \quad \mathbb{E}[B_k \mid \mathcal{F}_k] = 2\mathbb{E}[d_{(k+1)} - 1 \mid \mathcal{F}_k] \frac{A_k}{U_k - 1} - 2 \frac{\mathbb{E}[d_{(k+1)} - 1 \mid \mathcal{F}_k]}{U_k - 1}.$$

By (4.4.30),

$$(4.4.45) \quad \mathbb{E}[d_{(k+1)} - 1 \mid \mathcal{F}_k] = \frac{\sum_{j \notin \mathcal{V}_k} d_j (d_j - 1)}{\sum_{j \notin \mathcal{V}_k} d_j} = \frac{\sum_{j \in [n]} d_j^2}{\sum_{j \in [n]} d_j} - 1 + o_{\mathbb{P}}(1) = 1 + o_{\mathbb{P}}(1).$$

uniformly over $k \leq tn^{2/3}$, where the last step follows from the criticality condition (4.4.19) for ν_n . Further, since $p_0 < 1$, $U_k \geq c_0 n$ uniformly over $k \leq o(n)$. Thus, (4.4.44) gives (4.4.40). The fact that all the $O_{\mathbb{P}}$, $o_{\mathbb{P}}$ are uniform in $k \leq tn^{2/3}$ follows from $\sum_{j \in [n]} d_j^r - kd_{\max}^2 \leq \sum_{j \notin \mathcal{V}_k} d_j^r \leq \sum_{j \in [n]} d_j^r$ for $r = 1, 2$, together with $d_{\max} = o(n^{1/3})$. To prove (4.4.41), note that

$$(4.4.46) \quad \mathbb{E}[C_k \mid \mathcal{F}_k, v_{(k+1)}] = 2(d_{(k+1)} - 2) \frac{d_{(k+1)} - 1}{U_k - 1}.$$

By Conditions 1.6(a)-(c), $\mathbb{E}[D_n^3] \rightarrow \mathbb{E}[D^3] < \infty$ and (4.4.30),

$$(4.4.47) \quad \mathbb{E}[d_{(k+1)}^2 \mid \mathcal{F}_k] = \frac{\sum_{j \notin \mathcal{V}_k} d_j^3}{\sum_{j \notin \mathcal{V}_k} d_j} \leq \frac{\sum_{j \in [n]} d_j^3}{\sum_{j \in [n]} d_j + o_{\mathbb{P}}(n^{2/3})} = O_{\mathbb{P}}(1),$$

uniformly for $k \leq tn^{2/3}$. Therefore,

$$(4.4.48) \quad \mathbb{E}[C_k \mid \mathcal{F}_k] = O_{\mathbb{P}}(n^{-1})$$

uniformly over $k \leq tn^{2/3}$. □

Ignoring cycles. We prove Theorem 4.15 by approximating \mathbf{S}_n by a simpler process defined as

$$(4.4.49) \quad s_n(0) = 0, \quad s_n(i) = \sum_{j=1}^i (d_{(j)} - 2).$$

This idea was first used by Joseph [189]. Note that the difference between the processes \mathbf{S}_n and \mathbf{s}_n is due to the cycles, loops, and multiple-edges encountered during the exploration. Following the approach of Joseph in [189], it will be enough to prove the following:

PROPOSITION 4.19. *Suppose that Conditions 1.6(a)-(c) hold, that $\mathbb{E}[D_n^3] \rightarrow \mathbb{E}[D^3] < \infty$ and that the criticality condition (4.4.19) holds for ν_n . Then, as $n \rightarrow \infty$,*

$$(4.4.50) \quad \bar{s}_n \xrightarrow{d} \mathbf{B}_{\mu,\eta}^\lambda$$

with respect to the Skorohod J_1 topology.

REMARK 4.20. It will be shown that Proposition 4.19 implies Theorem 4.15 by showing that the distributions of $\bar{\mathbf{S}}_n$ and \bar{s}_n are very close as $n \rightarrow \infty$. This is achieved by proving that we do not see *too many* cycle half-edges up to the time $\lfloor n^{2/3}u \rfloor$ for any fixed $u > 0$.

From here onwards we look at the continuous versions of the processes $\bar{\mathbf{S}}_n$ and \bar{s}_n , obtained by linearly interpolating between the values at the jump points. We denote this continuous version with the same notation. It is easy to see that these continuous versions differ from their càdlàg versions by at most $n^{-1/3}d_{\max} = o(1)$ uniformly on $[0, T]$, for any $T > 0$. Therefore, the convergence in law of the continuous versions implies the convergence in law of the càdlàg versions and vice versa. Before proceeding to show that Theorem 4.15 is a consequences of Proposition 4.19, we bound the difference of these two processes in a suitable way. In the following lemma, which is Joseph [189, Lemma 6.1], we recall the definition of $c_{(k+1)} := (B_k + C_k)/2$ from (4.4.22):

LEMMA 4.21 (Dealing with cycle half-edges). *Fix $t > 0$ and $M > 0$ (large). Define*

$$(4.4.51) \quad E_n(t, M) := \left\{ \max_{s \leq t} \{ \bar{s}_n(s) - \min_{u \leq s} \bar{s}_n(u) \} < M \right\}.$$

Then

$$(4.4.52) \quad \limsup_{n \rightarrow \infty} \sum_{k \leq tn^{2/3}} \mathbb{E}[c_{(k)} \mathbb{1}_{E_n(t, M)}] < \infty.$$

PROOF. Note that, for all large n , $A_k \leq Mn^{1/3}$ on $E_n(t, M)$, because

$$(4.4.53) \quad \begin{aligned} A_k &= S_n(k) - \min_{j \leq k} S_n(j) + 2 = s_n(k) - 2 \sum_{j=1}^k c_{(j)} - \min_{j \leq k} S_n(j) + 2 \\ &\leq s_n(k) - \min_{j \leq k} s_n(j) + 2, \end{aligned}$$

where the last step follows by noting that $\min_{j \leq k} s_n(j) \leq \min_{j \leq k} S_n(j) + 2 \sum_{j=1}^k c_{(j)}$. By Lemma 4.18,

$$(4.4.54) \quad \mathbb{E}[c_{(k)} \mathbb{1}_{E_n(t, M)}] \leq \frac{Mn^{1/3}}{\mu n} + o(n^{-2/3}) = \frac{M}{\mu} n^{-2/3} + o(n^{-2/3}),$$

uniformly for $k \leq tn^{2/3}$. Summing over $1 \leq k \leq tn^{2/3}$ and taking the lim sup completes the proof. \square

PROOF OF THEOREM 4.15 SUBJECT TO PROPOSITION 4.19. This is a standard argument that we include for completeness (see Joseph [189, Section 6.2]). For $t > 0$ and $M > 0$ define the event $E_n(t, M)$ as in Lemma 4.21. Also, denote by $\mathbb{C}_L[0, t]$ the set of all bounded Lipschitz functions from $[0, t]$ to \mathbb{R} . Since $\mathbb{C}_L[0, t]$ separates the points in $[0, t]$,

and $[0, t]$ is compact, $\mathbb{C}_L[0, t]$ is dense in $\mathbb{C}[0, t]$. Therefore, to prove Theorem 4.15, it suffices to prove that, as $n \rightarrow \infty$,

$$(4.4.55) \quad \mathbb{E}[f(\bar{\mathbf{S}}_n)] - \mathbb{E}[f(\bar{\mathbf{s}}_n)] \rightarrow 0,$$

for all $f \in \mathbb{C}_L[0, t]$. Choose and fix $b > 0$ such that

$$(4.4.56) \quad z, z_1, z_2 \in \mathbb{C}[0, t] \implies |f(z)| \leq b \quad \text{and} \quad |f(z_1) - f(z_2)| \leq b \|z_1 - z_2\|_\infty.$$

Now,

$$(4.4.57) \quad \begin{aligned} & |\mathbb{E}[f(\bar{\mathbf{S}}_n)] - \mathbb{E}[f(\bar{\mathbf{s}}_n)]| \\ & \leq |(\mathbb{E}[f(\bar{\mathbf{S}}_n) \mathbb{1}_{E_n(t, M)}] - \mathbb{E}[f(\bar{\mathbf{s}}_n) \mathbb{1}_{E_n(t, M)}])| \\ & \quad + |(\mathbb{E}[f(\bar{\mathbf{S}}_n) \mathbb{1}_{E_n(t, M)^c}] - \mathbb{E}[f(\bar{\mathbf{s}}_n) \mathbb{1}_{E_n(t, M)^c}])| \\ & \leq b \mathbb{E}[|\bar{\mathbf{S}}_n - \bar{\mathbf{s}}_n|_t \mathbb{1}_{E_n(t, M)}] + 2b \mathbb{P}(\max_{s \leq t} \{\bar{s}_n(s) - \min_{u \leq s} \bar{s}_n(u)\} \geq M) \\ & \leq 2bn^{-1/3} \sum_{k \leq tn^{2/3}} \mathbb{E}[c_{(k)} \mathbb{1}_{E_n(t, M)}] + 2b \mathbb{P}(\max_{s \leq t} \{\bar{s}_n(s) - \min_{u \leq s} \bar{s}_n(u)\} \geq M). \end{aligned}$$

The first term in the above sum tends to zero as $n \rightarrow \infty$, by Lemma 4.21. The reflection of a process is a continuous map from $\mathbb{D}([0, \infty), \mathbb{R})$ to $\mathbb{D}([0, \infty), \mathbb{R})$ (see Whitt [251, Theorem 13.5.1]). Therefore, Proposition 4.19 implies that

$$(4.4.58) \quad (\bar{s}_n(s) - \min_{u \leq s} \bar{s}_n(u))_{s \geq 0} \xrightarrow{d} \mathbf{R}_{\mu, \eta}^\lambda,$$

where $\mathbf{R}_{\mu, \eta}^\lambda$ was defined in (4.4.13) (recall also (4.4.12)) with parameters given in (4.4.18). By the Portmanteau lemma,

$$(4.4.59) \quad \limsup_{n \rightarrow \infty} \mathbb{P}(\max_{s \leq t} \{\bar{s}_n(s) - \min_{u \leq s} \bar{s}_n(u)\} \geq M) \leq \mathbb{P}(\max_{s \leq t} R_{\mu, \eta}^\lambda(s) \geq M),$$

for any $t > 0$. The proof follows by taking the limit $M \rightarrow \infty$. \square

From here onwards, the main focus of this section is to prove Proposition 4.19. We use the martingale functional central limit theorem in a similar manner as Aldous in [9], as described in some detail in Section 4.2.2.

PROOF OF PROPOSITION 4.19. Let $\{\mathcal{F}_i\}_{i \in [n]}$ be the natural filtration defined in Lemma 4.18. Recall the definition of $s_n(i)$ from (4.4.49). By the Doob-Meyer decomposition (see Karatzas and Shreve [192, Theorem 4.10]), we can write

$$(4.4.60a) \quad s_n(i) = M_n(i) + C_n(i),$$

$$(4.4.60b) \quad M_n^2(i) = Z_n(i) + V_n(i),$$

where

$$(4.4.61a) \quad M_n(i) = \sum_{j=1}^i (d_{(j)} - \mathbb{E}[d_{(j)} | \mathcal{F}_{j-1}]),$$

$$(4.4.61b) \quad C_n(i) = \sum_{j=1}^i \mathbb{E}[d_{(j)} - 2 | \mathcal{F}_{j-1}],$$

$$(4.4.61c) \quad V_n(i) = \sum_{j=1}^i (\mathbb{E}[d_{(j)}^2 | \mathcal{F}_{j-1}] - \mathbb{E}^2[d_{(j)} | \mathcal{F}_{j-1}]).$$

This decomposition is alike the one for $\text{ER}_n(\lambda/n)$ in (4.2.20)–(4.2.21), with $X_n(i)$ replaced with $d_{(j)}$ and $X_n(i) - 1$ in $C_n(i)$ replaced by $d_{(j)} - 2$.

Recall that for a discrete time process $(X_n(i))_{i \geq 1}$, we write $\bar{X}_n(t) = n^{-1/3} X_n(\lfloor tn^{2/3} \rfloor)$. Our result follows from the martingale functional central limit theorem as formulated in Theorem 4.3 if we can prove the four conditions in MCLT(a)–MCLT(d), as formulated on page 161. The validation of these conditions are given separately in the subsequent part of this section.

Verification of MCLT(b). Denote by $\sigma_r(n) = \frac{1}{n} \sum_{i \in [n]} d_i^r$, $r = 2, 3$ and $\mu(n) = \frac{1}{n} \sum_{i \in [n]} d_i$. To prove MCLT(b), it is enough to prove that

$$(4.4.62) \quad n^{-2/3} V_n(\lfloor un^{2/3} \rfloor) \xrightarrow{\mathbb{P}} \frac{\sigma_3 \mu - \sigma_2^2}{\mu^2} u.$$

Recall that $\mathbb{E}[d_{(i)}^2 | \mathcal{F}_{i-1}] = \sum_{j \notin \mathcal{V}_{i-1}} d_j^3 / \sum_{j \notin \mathcal{V}_{i-1}} d_j$. Furthermore, since $d_{\max} = o(n^{1/3})$, uniformly over $i \leq un^{2/3}$,

$$(4.4.63) \quad \sum_{j \notin \mathcal{V}_{i-1}} d_j = \sum_{j \in [n]} d_j + O_{\mathbb{P}}(d_{\max} i) = \ell_n + o_{\mathbb{P}}(n).$$

Assume that, without loss of generality, $j \mapsto d_j$ is non-increasing. Then, uniformly over $i \leq un^{2/3}$,

$$(4.4.64) \quad \left| \sum_{j \notin \mathcal{V}_{i-1}} d_j^3 - n \sigma_3(n) \right| \leq \sum_{j=1}^{un^{2/3}} d_j^3.$$

For each fixed k ,

$$(4.4.65) \quad \begin{aligned} \frac{1}{n} \sum_{j=1}^{un^{2/3}} d_j^3 &\leq \frac{1}{n} \sum_{j=1}^{un^{2/3}} \mathbb{1}_{\{d_j \leq k\}} d_j^3 + \frac{1}{n} \sum_{j \in [n]} \mathbb{1}_{\{d_j > k\}} d_j^3 \\ &\leq k^3 un^{-1/3} + \frac{1}{n} \sum_{j \in [n]} \mathbb{1}_{\{d_j > k\}} d_j^3 = o(1), \end{aligned}$$

where we first let $n \rightarrow \infty$ and then $k \rightarrow \infty$ and use Lemma 4.17. Therefore, the right-hand side of (4.4.64) is $o(n)$ and we conclude that, uniformly over $i \leq un^{2/3}$,

$$(4.4.66) \quad \mathbb{E}[d_{(i)}^2 | \mathcal{F}_{i-1}] = \frac{\sigma_3}{\mu} + o_{\mathbb{P}}(1),$$

as required. A similar argument gives

$$(4.4.67) \quad \mathbb{E}[d_{(i)} | \mathcal{F}_{i-1}] = \frac{\sigma_2}{\mu} + o_{\mathbb{P}}(1),$$

and MCLT(b) follows by noting that the error term is $o_{\mathbb{P}}(1)$, since we are summing $n^{2/3}$ terms, scaling by $n^{-2/3}$ and using the uniformity of errors over $i \leq un^{2/3}$. \square

Verification of MCLT(c) and MCLT(d). For MCLT(c), we bound

$$\begin{aligned}
\mathbb{E}\left[\sup_{s \leq u} |\bar{M}_n(s) - \bar{M}_n(s-)|^2\right] &= n^{-2/3} \mathbb{E}\left[\sup_{k \leq un^{2/3}} |M_n(k) - M_n(k-1)|^2\right] \\
&= n^{-2/3} \mathbb{E}\left[\sup_{k \leq un^{2/3}} |d_{(k)} - \mathbb{E}[d_{(k)} | \mathcal{F}_{k-1}]|^2\right] \\
&\leq n^{-2/3} \mathbb{E}\left[\sup_{k \leq un^{2/3}} d_{(k)}^2\right] + n^{-2/3} \mathbb{E}\left[\sup_{k \leq un^{2/3}} \mathbb{E}^2[d_{(k)} | \mathcal{F}_{k-1}]\right] \\
(4.4.68) \quad &\leq 2n^{-2/3} d_{\max}^2.
\end{aligned}$$

Similarly,

$$\begin{aligned}
n^{-1/3} \mathbb{E}\left[\sup_{s \leq u} |\bar{V}_n(s) - \bar{V}_n(s-)|^2\right] &= n^{-2/3} \mathbb{E}\left[\sup_{k \leq un^{2/3}} |V_n(k) - V_n(k-1)|\right] \\
&= n^{-2/3} \mathbb{E}\left[\sup_{k \leq un^{2/3}} \text{Var}(d_{(k)} | \mathcal{F}_{k-1})\right] \\
(4.4.69) \quad &\leq 2n^{-2/3} d_{\max}^2,
\end{aligned}$$

and MCLT(c)-MCLT(d) follow from Lemma 4.17 using $d_{\max} = o(n^{1/3})$. \square

Verification of MCLT(a). As for the Erdős-Rényi random graph, proving MCLT(a) requires some more work. Note that

$$\begin{aligned}
\mathbb{E}[d_{(i)} - 2 | \mathcal{F}_{i-1}] &= \frac{\sum_{j \notin \mathcal{V}_{i-1}} d_j (d_j - 2)}{\sum_{j \notin \mathcal{V}_{i-1}} d_j} \\
&= \frac{\sum_{j \in [n]} d_j (d_j - 2)}{\sum_{j \in [n]} d_j} - \frac{\sum_{j \in \mathcal{V}_{i-1}} d_j (d_j - 2)}{\sum_{j \in [n]} d_j} + \frac{\sum_{j \notin \mathcal{V}_{i-1}} d_j (d_j - 2) \sum_{j \in \mathcal{V}_{i-1}} d_j}{\sum_{j \notin \mathcal{V}_{i-1}} d_j \sum_{j \in [n]} d_j} \\
(4.4.70) \quad &= \frac{\lambda}{n^{1/3}} - \frac{\sum_{j \in \mathcal{V}_{i-1}} d_j^2}{\sum_{j \in [n]} d_j} + \frac{\sum_{j \notin \mathcal{V}_{i-1}} d_j^2 \sum_{j \in \mathcal{V}_{i-1}} d_j}{\sum_{j \notin \mathcal{V}_{i-1}} d_j \sum_{j \in [n]} d_j} + o(n^{-1/3}),
\end{aligned}$$

where the last step follows from the criticality condition for ν_n in (4.4.19). Therefore,

$$\begin{aligned}
C_n(k) &= \sum_{i=1}^k \mathbb{E}[d_{(i)} - 2 | \mathcal{F}_{i-1}] \\
(4.4.71) \quad &= \frac{k\lambda}{n^{1/3}} - \sum_{i=1}^k \frac{\sum_{j \in \mathcal{V}_{i-1}} d_j^2}{\sum_{j \in [n]} d_j} + \sum_{i=1}^k \frac{\sum_{j \notin \mathcal{V}_{i-1}} d_j^2 \sum_{j \in \mathcal{V}_{i-1}} d_j}{\sum_{j \notin \mathcal{V}_{i-1}} d_j \sum_{j \in [n]} d_j} + o(kn^{-1/3}).
\end{aligned}$$

The following lemma estimates the sums on the right-hand side of (4.4.71):

LEMMA 4.22. *For all $u > 0$ and $r = 1, 2$, as $n \rightarrow \infty$,*

$$(4.4.72) \quad \sup_{s \leq u} \left| n^{-1/3} \sum_{i=1}^{\lfloor sn^{2/3} \rfloor} \sum_{j=1}^{i-1} \frac{d_{(j)}^r}{\ell_n} - \frac{\sigma_{r+1} s^2}{2\mu^2} \right| \xrightarrow{\mathbb{P}} 0.$$

Consequently,

$$(4.4.73) \quad \sup_{s \leq u} \left| n^{-1/3} \sum_{i=1}^{\lfloor sn^{2/3} \rfloor} \frac{\sum_{j \notin \mathcal{V}_{i-1}} d_j^2 \sum_{j \in \mathcal{V}_{i-1}} d_j}{\sum_{j \notin \mathcal{V}_{i-1}} d_j \sum_{j \in [n]} d_j} - \frac{\sigma_2^2 s^2}{2\mu^3} \right| \xrightarrow{\mathbb{P}} 0.$$

The proof of Lemma 4.22 is a straightforward consequence of Lemma 4.16, and is left as an exercise:

EXERCISE 4.28 (Proof of Lemma 4.22). *Prove that Lemma 4.22 follows from Lemma 4.16.*

The proof of MCLT(a) follows by using Lemma 4.22 in (4.4.71). \square

The proof of MCLT(a)-MCLT(d) completes the proof of Proposition 4.19, and thus that of Theorem 4.15. \square

Finite-dimensional convergence of the ordered component sizes. The convergence of the exploration process in Theorem 4.15 implies that, for any large $T > 0$, the k -largest components explored up to time $Tn^{2/3}$ converge to the k -largest excursions above past minima of $\mathbf{B}_{\mu,\eta}^\lambda$ up to time T . Therefore, we can conclude the finite-dimensional convergence of the ordered components sizes in the whole graph if we can show that the large components are explored early by the exploration process. The following lemma formalizes the above statement:

LEMMA 4.23 (No large components are found late in the exploration). *Let $\mathcal{C}_{\max}^{\geq T}$ denote the largest component that is started exploring after time $Tn^{2/3}$. Then, for any $\delta > 0$,*

$$(4.4.74) \quad \lim_{T \rightarrow \infty} \limsup_{n \rightarrow \infty} \mathbb{P}(|\mathcal{C}_{\max}^{\geq T}| > \delta n^{2/3}) = 0.$$

Let us first state the two main ingredients to complete the proof of Lemma 4.23:

LEMMA 4.24 (Path counting by Janson [183, Lemma 5.2]). *Consider $\text{CM}_n(\mathbf{d})$ with $\nu_n < 1$ and let $\mathcal{C}(V_n)$ denote the component containing the vertex V_n where V_n is a random vertex chosen independently of the graph. Then,*

$$(4.4.75) \quad \mathbb{E}[|\mathcal{C}(V_n)|] \leq 1 + \frac{\mathbb{E}[D_n]}{1 - \nu_n}.$$

LEMMA 4.25 (Depletion-of-points effect for configuration model). *Define, $\nu_{n,i} = \sum_{j \notin \mathcal{V}_{i-1}} d_j(d_j - 1) / \sum_{j \notin \mathcal{V}_{i-1}} d_j$. There exists some constant $C_0 > 0$ such that for any $T > 0$,*

$$(4.4.76) \quad \nu_{n,Tn^{2/3}} = \nu_n - C_0 T n^{-1/3} + o_{\mathbb{P}}(n^{-1/3}).$$

PROOF. Using a similar split up as in (4.4.70), we have

$$(4.4.77) \quad \nu_{n,i} = \nu_n + \frac{\sum_{j \in \mathcal{V}_{i-1}} d_j(d_j - 1)}{\ell_n} - \frac{\sum_{j \notin \mathcal{V}_{i-1}} d_j(d_j - 1) \sum_{j \in \mathcal{V}_{i-1}} d_j}{\ell_n \sum_{j \notin \mathcal{V}_{i-1}} d_j}.$$

Now, (4.4.30) gives that, uniformly over $i \leq Tn^{2/3}$,

$$(4.4.78a) \quad \frac{\sum_{j \notin \mathcal{V}_{i-1}} d_j(d_j - 1)}{\sum_{j \notin \mathcal{V}_{i-1}} d_j} = \frac{\sum_{j \in [n]} d_j(d_j - 1) + o_{\mathbb{P}}(n^{2/3})}{\sum_{j \in [n]} d_j + o_{\mathbb{P}}(n^{2/3})} = 1 + o_{\mathbb{P}}(n^{-1/3}),$$

$$(4.4.78b) \quad \sum_{j \in \mathcal{V}_{i-1}} d_j(d_j - 2) = \left(\frac{\sigma_3}{\mu} - 2 \right) (i - 1) + o_{\mathbb{P}}(n^{2/3}).$$

Further, note that $\sigma_3 - 2\mu = \mathbb{E}[D(D-1)(D-2)] + \mathbb{E}[D(D-2)] > 0$, by Conditions 1.6(a)-(c) and $p_2 \in [0, 1)$, together with the criticality condition on ν_n in (4.4.19). Therefore, (4.4.77) gives (4.4.76) for some constant $C_0 > 0$. \square

PROOF OF LEMMA 4.23. Let $i_T := \inf\{i \geq Tn^{2/3} : S_n(i) = \inf_{j \leq i} S_n(j)\}$. Thus, i_T denotes the first time at which we finish exploring a component after time $Tn^{2/3}$. Note that, conditionally on the explored vertices up to time i_T , the remaining graph $\bar{\mathcal{G}}$ is still a configuration model. Let $\bar{\nu}_n = \sum_{i \in \bar{\mathcal{G}}} d_i(d_i - 1) / \sum_{i \in \bar{\mathcal{G}}} d_i$ be the criticality parameter of $\bar{\mathcal{G}}$. Then, using (4.4.76) in Lemma 4.25, we conclude that

$$(4.4.79) \quad \bar{\nu}_n \leq \nu_n - C_0 T n^{-1/3} + O_{\mathbb{P}}(n^{-1/3}).$$

Take $T > 0$ such that $\lambda - C_0 T < 0$. Thus, with high probability, $\bar{\nu}_n < 1$. Denote the component corresponding to a randomly chosen vertex from $\bar{\mathcal{G}}$ by $\mathcal{C}^{\geq T}(V_n)$, and the i th largest component of $\bar{\mathcal{G}}$ by $\mathcal{C}_{(i)}^{\geq T}$. Also, let $\bar{\mathbb{P}}$ denote the probability measure conditioned on \mathcal{F}_{i_T} , and $\bar{\nu}_n < 1$ and let $\bar{\mathbb{E}}$ denote the corresponding expectation. Now, for any $\delta > 0$,

$$(4.4.80) \quad \begin{aligned} \bar{\mathbb{P}}\left(|\mathcal{C}_{\max}^{\geq T}| > \delta n^{2/3}\right) &\leq \bar{\mathbb{P}}\left(\sum_{i \geq 1} |\mathcal{C}_{(i)}^{\geq T}|^2 > \delta n^{4/3}\right) \leq \frac{1}{\delta n^{4/3}} \sum_{i \geq 1} \bar{\mathbb{E}}(|\mathcal{C}_{(i)}^{\geq T}|^2) \\ &= \frac{1}{\delta n^{1/3}} \bar{\mathbb{E}}(|\mathcal{C}^{\geq T}(V_n)|) \leq \frac{1}{\delta(-\lambda + C_0 T + O_{\mathbb{P}}(1))}, \end{aligned}$$

where the second step follows from the Markov inequality and the last step follows by combining Lemma 4.24 and (4.4.79). Noting that $\bar{\nu}_n < 1$ whp, we get

$$(4.4.81) \quad \limsup_{n \rightarrow \infty} \mathbb{P}\left(|\mathcal{C}_{\max}^{\geq T}| > \delta n^{2/3}\right) \leq \frac{C}{\delta T},$$

for some constant $C > 0$ and the proof follows by taking T large. \square

By Theorem 4.15 and Lemma 4.23, the convergence in Theorem 4.13 holds with respect to the product topology. We have to work slightly harder to extend this to convergence in the ℓ_{\downarrow}^2 topology, which we defer to later. We first investigate the number of vertices of degree k in the explored clusters, which will be helpful to prove (4.4.21) in Theorem 4.13.

Vertices of degree k . We now compute the number of vertices of degree k in each connected component at criticality. This will be needed to obtain control over the number of degree one vertices, which in turn are needed when we remove the red vertices in the percolated configuration model at criticality. Such an estimate was proved by Janson and Luczak in [185, Theorem 2.4] for supercritical graphs under a stronger moment assumption.

LEMMA 4.26. *Denote by $N_k(t)$ the number of vertices of degree k discovered up to time t . For any $t > 0$,*

$$(4.4.82) \quad \sup_{u \leq t} \left| n^{-2/3} N_k(un^{2/3}) - \frac{kn_k}{\ell_n} u \right| = O_{\mathbb{P}}((kn^{1/3})^{-1}).$$

PROOF. By setting $\alpha = 2/3$ and $f(d_i) = \mathbb{1}_{\{d_i=k\}}$ in Proposition 2.42, we can directly conclude that

$$(4.4.83) \quad \sup_{u \leq t} \left| n^{-2/3} N_k(un^{2/3}) - \frac{kn_k}{\ell_n} u \right| \xrightarrow{\mathbb{P}} 0.$$

However, if we repeat the same arguments leading to the proof of Proposition 2.42 (see in particular (2.6.28)) we get

$$(4.4.84) \quad \mathbb{P}\left(\sup_{u \leq t} \left| n^{-2/3} N_k(un^{2/3}) - \frac{kn_k}{\ell_n} u \right| > \frac{A}{kn^{1/3}}\right) \leq \frac{3\left(k^3 s^2 \frac{p_k^{(n)}}{(\mathbb{E}[D])^2} + \sqrt{s \frac{k^3 p_k^{(n)}}{\mathbb{E}[D]}}\right)}{A} + o(1),$$

where we write $p_k^{(n)} = n_k/n$ for the proportion of vertices of degree k in $\text{CM}_n(\mathbf{d})$. Now, we can use the finite third-moment assumption to conclude that the numerator in the right hand side can be taken to be uniform over k . Thus, the proof follows. \square

Define $v_k(G) :=$ the number of vertices of degree k in the connected graph G . As a corollary to Lemma 4.26 and (4.4.74), we can deduce that

$$(4.4.85) \quad v_k(\mathcal{C}_{(j)}) = p_k^* |\mathcal{C}_{(j)}| + O_{\mathbb{P}}((k^{-1}n^{1/3})),$$

which proves (4.4.21) in Theorem 4.13. \square

Completion of the proof of Theorem 4.13. The proof of Theorem 4.13 follows using similar argument as by Aldous in [9, Section 3.3]. However, the proof is a bit tricky since the components are explored in a size-biased manner with sizes being the total degree in the component (not the component sizes as in [9]). For a sequence of random variables $\mathbf{Y} = (Y_i)_{i \geq 1}$ satisfying $\sum_{i \geq 1} Y_i^2 < \infty$ almost surely, define $\boldsymbol{\xi} := (\xi_i)_{i \geq 1}$ such that $\xi_i | \mathbf{Y} \sim \text{Exp}(Y_i)$ and the coordinates of $\boldsymbol{\xi}$ are independent conditional on \mathbf{Y} . For $a \geq 0$, let $\mathcal{S}(a) := \sum_{\xi_i \leq a} Y_i$. Then the *size-biased point process* is defined to be the random collection of points $\Xi := \{(\mathcal{S}(\xi_i), Y_i)\}$ (see Aldous [9, Section 3.3]). We use in particular Aldous [9, Lemma 8, Lemma 14 and Proposition 15].

Let $\mathfrak{C} := \{\mathcal{C} : \mathcal{C} \text{ is a component of } \text{CM}_n(\mathbf{d})\}$. Consider the collection $\boldsymbol{\xi} := (\xi(\mathcal{C}))_{\mathcal{C} \in \mathfrak{C}}$ such that conditionally on the values $(\sum_{k \in \mathcal{C}} d_k, |\mathcal{C}|)_{\mathcal{C} \in \mathfrak{C}}$, $\xi(\mathcal{C})$ has an exponential distribution with rate $n^{-2/3} \sum_{k \in \mathcal{C}} d_k$ independently over \mathcal{C} . Then the order in which Algorithm 4.14 explores the components can be obtained by ordering the components according to their ξ -value. Recall that \mathcal{C}_i denotes the i th explored component by Algorithm 4.14 and let $D_i := \sum_{k \in \mathcal{C}_i} d_k$. Define the size-biased point process

$$(4.4.86) \quad \Xi_n := \left(n^{-2/3} \sum_{j=1}^i D_j, n^{-2/3} D_i \right)_{i \geq 1}.$$

Also define the point processes

$$(4.4.87) \quad \Xi'_n := \left(n^{-2/3} \sum_{j=1}^i |\mathcal{C}_j|, n^{-2/3} |\mathcal{C}_i| \right)_{i \geq 1},$$

$$(4.4.88) \quad \Xi_\infty := \{(l(\gamma), |\gamma|) : \gamma \text{ an excursion of } \mathbf{B}_{\mu, \eta}^\lambda\},$$

where we recall that $l(\gamma)$ are the left endpoints of the excursions of $\mathbf{B}_{\mu, \eta}^\lambda$ and $|\gamma|$ is the length of the excursion γ . Note that Ξ'_n is not a size-biased point process. However, applying [9, Lemma 8] and Theorem 4.15, we get $\Xi'_n \xrightarrow{d} \Xi_\infty$. We claim that

$$(4.4.89) \quad \Xi_n \xrightarrow{d} 2\Xi_\infty.$$

To verify the claim, note that (4.4.30) and Conditions 1.6(a)-(c), together with the criticality condition ν_n in (4.4.19), imply

$$(4.4.90) \quad \sup_{u \leq t} \left| n^{-2/3} \sum_{i=1}^{\lfloor un^{2/3} \rfloor} d_{(i)} - \frac{\sigma_2}{\mu} u \right| = \sup_{u \leq t} \left| n^{-2/3} \sum_{i=1}^{\lfloor un^{2/3} \rfloor} d_{(i)} - 2u \right| \xrightarrow{\mathbb{P}} 0,$$

for any $t > 0$ since $\sigma_2/\mu = \mathbb{E}[D^2]/\mathbb{E}[D] = 2$. Therefore,

$$(4.4.91) \quad \sum_{\xi(\mathcal{C}) \leq s} D(\mathcal{C}) - 2 \sum_{\xi(\mathcal{C}) \leq s} |\mathcal{C}| = O_{\mathbb{P}}(n^{2/3}).$$

Thus, (4.4.89) follows using (4.4.90) and (4.4.91). Now, the point process $2\Xi_{\infty}$ satisfies all the conditions of [9, Proposition 15] as shown by Aldous. Thus, [9, Lemma 14] gives

$$(4.4.92) \quad \{D_{(i)}\}_{i \geq 1} \text{ is tight in } \ell_{\downarrow}^2.$$

This implies that $(n^{-2/3} |\mathcal{C}_{(i)}|)_{i \geq 1}$ is tight in ℓ_{\downarrow}^2 by simply observing that $|\mathcal{C}_i| \leq \sum_{k \in \mathcal{C}_i} d_k + 1$. Therefore, the proof of the convergence in ℓ_{\downarrow}^2 in Theorem 4.13 is complete using the convergence in finite-dimensional distributions that we have already proved. \square

Having completed the proof of convergence in ℓ_{\downarrow}^2 in Theorem 4.13, we are still missing the convergence of the surplus of clusters alongside the cluster sizes. The proof of Theorem 4.13 is completed in two separate lemmas. In Lemma 4.27 we first show that the convergence in Theorem 4.13 holds with respect to the $\ell_{\downarrow}^2 \times \mathbb{N}^{\infty}$ topology. The tightness of $(\mathbf{Z}_n)_{n \geq 1}$ with respect to the $\mathbb{U}_{\downarrow}^0$ topology is ensured in Lemma 4.28 and Theorem 4.13 follows.

LEMMA 4.27 (Surplus edges found in exploration). *Let $N_n^{\lambda}(k)$ be the number of surplus edges discovered up to time k and $\bar{N}_n^{\lambda}(u) = N_n^{\lambda}(\lfloor un^{2/3} \rfloor)$. Then, as $n \rightarrow \infty$,*

$$(4.4.93) \quad \bar{\mathbf{N}}_n^{\lambda} \xrightarrow{d} \mathbf{N}^{\lambda},$$

where \mathbf{N}^{λ} is defined in (4.4.15).

PROOF. Recall the definitions of a , b , \mathcal{A}_k , \mathcal{B}_k , \mathcal{C}_k , \mathcal{S}_k in Algorithm 4.14. Recall also that $A_k := |\mathcal{A}_k|$, $B_k := |\mathcal{B}_k|$, $C_k := |\mathcal{C}_k|$, $U_k := |\mathcal{S}_k|$, $c_{(k+1)} := (|\mathcal{B}_k| + |\mathcal{C}_k|)/2$ as in (4.4.22). We have $A_k = S_n(k) - \min_{j \leq k} S_n(j) + 2$. From Lemma 4.18 we can conclude that

$$(4.4.94) \quad \mathbb{E}[c_{(k+1)} | \mathcal{F}_k] = \frac{A_k}{\mu n} + o_{\mathbb{P}}(n^{-1}).$$

The counting process \mathbf{N}_n^{λ} has conditional intensity (conditioned on \mathcal{F}_{k-1}) given by (4.4.94). Writing the conditional intensity in (4.4.94) in terms of $\bar{\mathbf{S}}_n$, we get that the conditional intensity of the re-scaled process $\bar{\mathbf{N}}_n^{\lambda}$ is given by

$$(4.4.95) \quad \frac{1}{\mu} [\bar{S}_n(u) - \min_{\tilde{u} \leq u} \bar{S}_n(\tilde{u})] + o_{\mathbb{P}}(1).$$

Denote by $\bar{R}_n(u) := \bar{S}_n(u) - \min_{\tilde{u} \leq u} \bar{S}_n(\tilde{u})$ which is the reflected version $\bar{\mathbf{S}}_n$. By Theorem 4.15,

$$(4.4.96) \quad \bar{\mathbf{R}}_n \xrightarrow{d} \mathbf{R}_{\mu, \eta}^{\lambda},$$

where $\mathbf{R}_{\mu,\eta}^\lambda$ is as defined in (4.4.13). Therefore, we can assume that there exists a probability space such that $\bar{\mathbf{R}}_n \rightarrow \mathbf{R}_{\mu,\eta}^\lambda$ almost surely. Using the continuity of the sample paths of $\mathbf{B}_{\mu,\eta}^\lambda$ in [207, Theorem 1; Chapter 5.3], we conclude that

$$(4.4.97) \quad \bar{\mathbf{N}}_n^\lambda \xrightarrow{d} \mathbf{N}_{\mu,\eta}^\lambda,$$

where $\mathbf{N}_{\mu,\eta}^\lambda$ is defined in (4.4.15). \square

LEMMA 4.28 (Tightness in \mathbb{U}_\downarrow^0). *The vector $(\mathbf{Z}_n)_{n \geq 1}$ is tight with respect to the \mathbb{U}_\downarrow^0 topology.*

The proof of Lemma 4.28 makes use of the following crucial technical estimate of the probability that a component with small size has very large number of surplus edges:

LEMMA 4.29 (Small clusters do not have many surplus edges). *Assume that $\lambda < 0$. Let V_n denote a vertex chosen uniformly at random, independent of the graph $\text{CM}_n(\mathbf{d})$ and let $\mathcal{C}(V_n)$ denote the component containing V_n . Let $\delta_k = \delta k^{-0.12}$. Then, for $\delta > 0$ (small),*

$$(4.4.98) \quad \mathbb{P}\left(\text{SP}(\mathcal{C}(V_n)) \geq K, |\mathcal{C}(V_n)| \in (\delta_K n^{2/3}, 2\delta_K n^{2/3})\right) \leq \frac{C\sqrt{\delta}}{n^{1/3}K^{1.1}}.$$

where C is a fixed constant independent of n, δ, K .

We will not prove Lemma 4.29 here, and complete the proof of Theorem 4.13 by showing the tightness in Lemma 4.28:

PROOF OF LEMMA 4.28. To simplify the notation, we write $Y_i^n = n^{-2/3}|\mathcal{C}_{(i)}|$ and $N_i^n = \#\{\text{surplus edges in } \mathcal{C}_{(i)}\}$. Let Y_i, N_i denote the distributional limits of Y_i^n and N_i^n respectively. The process $\mathbf{Z}(\lambda)$ is almost surely \mathbb{U}_\downarrow^0 -valued. Using the definition of $d_{\mathbb{U}}$ from (4.4.7) and Lemma 4.27, the proof of Lemma 4.28 is complete if we can show that, for any $\eta > 0$

$$(4.4.99) \quad \lim_{\varepsilon \rightarrow 0} \limsup_{n \rightarrow \infty} \mathbb{P}\left(\sum_{Y_i^n \leq \varepsilon} Y_i^n N_i^n > \eta\right) = 0.$$

First, consider the case $\lambda < 0$. For every $\eta, \varepsilon > 0$ sufficiently small and with $i_{k,\varepsilon} = \log_2(1/(k^{0.12}\varepsilon))$,

$$\begin{aligned} \mathbb{P}\left(\sum_{Y_i^n \leq \varepsilon} Y_i^n N_i^n > \eta\right) &\leq \frac{1}{\eta} \mathbb{E}\left[\sum_{i=1}^{\infty} Y_i^n N_i^n \mathbb{1}_{\{Y_i^n \leq \varepsilon\}}\right] = \frac{n^{-2/3}}{\eta} \mathbb{E}\left[\sum_{i=1}^{\infty} |\mathcal{C}_{(i)}| N_i^n \mathbb{1}_{\{|\mathcal{C}_{(i)}| \leq \varepsilon n^{2/3}\}}\right] \\ &= \frac{n^{1/3}}{\eta} \mathbb{E}[\text{SP}(\mathcal{C}(V_n)) \mathbb{1}_{\{|\mathcal{C}(V_n)| \leq \varepsilon n^{2/3}\}}] \\ &= \frac{n^{1/3}}{\eta} \sum_{k=1}^{\infty} \sum_{i \geq i_{k,\varepsilon}} \mathbb{P}(\text{SP}(\mathcal{C}(V_n)) \geq k, |\mathcal{C}(V_n)| \in (2^{-(i+1)}k^{-0.12}n^{2/3}, 2^{-i}k^{-0.12}n^{2/3}]) \end{aligned}$$

(4.4.100)

$$\leq \frac{C}{\eta} \sum_{k=1}^{\infty} \frac{1}{k^{1.1}} \sum_{i \geq \log_2(1/(k^{0.12}\varepsilon))} 2^{-(1/2)i} \leq \frac{C}{\eta} \sum_{k=1}^{\infty} \frac{\sqrt{\varepsilon}}{k^{1.04}} = O(\sqrt{\varepsilon}),$$

where the last-but-one step follows from Lemma 4.29. Therefore, (4.4.99) holds when $\lambda < 0$. Now consider the case $\lambda > 0$. For $T > 0$ (large), let

$$(4.4.101) \quad \mathcal{K}_n := \{i : Y_i^n \leq \varepsilon, \mathcal{C}_{(i)} \text{ is explored before } Tn^{2/3}\}.$$

Then, by applying the Cauchy-Schwarz inequality,

$$(4.4.102) \quad \begin{aligned} \sum_{i \in \mathcal{K}_n} Y_i^n N_i^n &\leq \left(\sum_{i \in \mathcal{K}_n} (Y_i^n)^2 \right)^{1/2} \times \left(\sum_{i \in \mathcal{K}_n} (N_i^n)^2 \right)^{1/2} \\ &\leq \left(\sum_{i \in \mathcal{K}_n} (Y_i^n)^2 \right)^{1/2} \times (\# \text{ surplus edges explored before } Tn^{2/3}) \end{aligned}$$

Using similar ideas as the proof of Lemma 4.23, we can run the exploration process till $Tn^{2/3}$ and the unexplored graph becomes a configuration model with negative criticality parameter for large $T > 0$, by (4.4.76). Thus, the proof can be completed using (4.4.102), the ℓ_1^2 convergence of the component sizes proved before, and Lemma 4.27. \square

4.4.2. Proof of Theorem 4.9 subject to Theorem 4.13. In this section, we reduce the proof of Theorem 4.9 to Theorem 4.13. In order to inspect that Theorem 4.13 is indeed useful for us, let us start by checking that the degree sequence of the percolated graph in Janson's construction satisfy the conditions of Theorem 4.13.

The degree conditions for percolation on the configuration model. The main result is the following lemma:

LEMMA 4.30 (Percolated degrees in Janson's construction). *Fix $\pi_n(\lambda)$ as in (4.4.9) and let $\mathbf{d}(\pi_n(\lambda)) = (d_i(\pi_n(\lambda)))_{i \in [N]}$ denote the degree sequence of $\text{CM}_n(\mathbf{d}; \pi_n(\lambda))$ appearing through Janson's Construction. Assume that Conditions 1.6(a)-(c) hold for the degree sequence \mathbf{d} and that $\mathbb{E}[D_n^3] \rightarrow \mathbb{E}[D^3] < \infty$. Further, assume that $\text{CM}_n(\mathbf{d})$ is super-critical, i.e., $\nu_n \rightarrow \nu > 1$. Then, Conditions 1.6(a)-(c) hold for the degree sequence $\mathbf{d}(\pi_n(\lambda))$ and*

$$(4.4.103) \quad \nu_N(\pi_n(\lambda)) = \frac{\mathbb{E}_N[D_N(\pi_n(\lambda))(D_N(\pi_n(\lambda)) - 1)]}{\mathbb{E}_N[D_N(\pi_n(\lambda))]} = 1 + \lambda n^{-1/3} + O_{\mathbb{P}}(n^{-1/2}).$$

Further, with $\pi_c = 1/\nu$,

$$(4.4.104) \quad \mathbb{E}_N[D_N(\pi_n(\lambda))^3] \xrightarrow{\mathbb{P}} \mathbb{E}[D(\pi_c)^3].$$

PROOF. We have already shown that $\mathbb{E}_N[D_N(\pi_n(\lambda))] \xrightarrow{\mathbb{P}} \mathbb{E}[D(\pi_c)]$ (recall (4.3.13)), that $\mathbb{E}_N[D_N(\pi_n(\lambda))^2] \xrightarrow{\mathbb{P}} \mathbb{E}[D(\pi_c)^2]$ (recall (4.3.20)), and that $D_N(\pi_n(\lambda)) \xrightarrow{d} D(\pi_c)$ in (4.3.11). We are left to prove the critical window result in (4.4.103) and the convergence of the third moment in (4.4.104).

For (4.4.103), we recall (4.3.19), which implies that

$$(4.4.105) \quad \begin{aligned} &\mathbb{P}\left(\left| \sum_{v \in [n]} d_v(\pi_n(\lambda))(d_v(\pi_n(\lambda)) - 1) - \sum_{v \in [n]} \mathbb{E}[d_v(\pi_n(\lambda))(d_v(\pi_n(\lambda)) - 1)] \right| > t\right) \\ &\leq 2 \exp\left(-\frac{t^2}{2n\mathbb{E}[D_n^3]}\right). \end{aligned}$$

When $\mathbb{E}[D_n^3] \rightarrow \mathbb{E}[D^3] < \infty$, we thus obtain that

$$(4.4.106) \quad \sum_{v \in [n]} d_v(\pi_n(\lambda))(d_v(\pi_n(\lambda)) - 1) = \sum_{v \in [n]} \mathbb{E}[d_v(\pi_n(\lambda))(d_v(\pi_n(\lambda)) - 1)] + O_{\mathbb{P}}(\sqrt{n}).$$

By (4.3.16),

$$(4.4.107) \quad \begin{aligned} \sum_{v \in [n]} \mathbb{E}[d_v(\pi_n(\lambda))(d_v(\pi_n(\lambda)) - 1)] &= \pi_n(\lambda) \sum_{v \in [n]} d_v(d_v - 1) \\ &= n\pi_n(\lambda)\mathbb{E}[D_n(D_n - 1)]. \end{aligned}$$

Further,

$$(4.4.108) \quad \sum_{v \in [N]} d_v(\pi_n(\lambda)) = \ell_n = n\mathbb{E}[D_n],$$

which is deterministic. Thus,

$$(4.4.109) \quad \begin{aligned} \nu_N(\pi) &= \frac{\sum_{v \in [N]} d_v(\pi_n(\lambda))(d_v(\pi_n(\lambda)) - 1)}{\sum_{v \in [N]} d_v(\pi_n(\lambda))} \\ &= \pi_n(\lambda) \frac{\mathbb{E}[D_n(D_n - 1)]}{\mathbb{E}[D_n]} + O_{\mathbb{P}}(n^{-1/2}) \\ &= 1 + \lambda n^{-1/3} + O_{\mathbb{P}}(n^{-1/2}), \end{aligned}$$

by the choice of $\pi_n(\lambda)$ in (4.4.9). This proves (4.4.103).

For (4.4.104), by the above results, it suffices to prove that

$$(4.4.110) \quad \mathbb{E}_N[D_N(\pi_n(\lambda))(D_N(\pi_n(\lambda)) - 1)(D_N(\pi_n(\lambda)) - 2)] \xrightarrow{\mathbb{P}} \mathbb{E}[D(\pi)(D(\pi) - 1)(D(\pi) - 2)].$$

For this, in turn, since $N/n \xrightarrow{\mathbb{P}} 1 + \mathbb{E}[D](1 - \sqrt{\pi})$ by (4.3.5), it suffices to prove that

$$(4.4.111) \quad \sum_{v \in [n]} d_v(\pi_n(\lambda))(d_v(\pi_n(\lambda)) - 1)(d_v(\pi_n(\lambda)) - 2) = \sqrt{\pi_n(\lambda)}^3 \sum_{v \in [n]} d_v(d_v - 1)(d_v - 2) + O_{\mathbb{P}}(n).$$

We aim to use Lemma 4.8. Define

$$(4.4.112) \quad f_3(\mathbf{I}) = \sum_{v \in [n]} d_v(\pi)(d_v(\pi) - 1)(d_v(\pi) - 2).$$

Let \mathbf{I}' be equal to \mathbf{I} except in the i th coordinate. Then,

$$(4.4.113) \quad |f_3(\mathbf{I}') - f_3(\mathbf{I})| \leq d_{v_i}^2 \equiv c_i,$$

where v_i is the vertex to which the i th half-edge is incident, so that $c_i = d_{v_i}^2$. We compute that

$$(4.4.114) \quad \sum_{i \in [\ell_n]} c_i^2 = \sum_{v \in [n]} d_v^5.$$

By Lemma 4.8,

$$(4.4.115) \quad \mathbb{P}\left(|f_3(\mathbf{I}) - \mathbb{E}[f_3(\mathbf{I})]| > t\right) \leq 2 \exp\left(-\frac{t^2}{2 \sum_{v \in [n]} d_v^5}\right).$$

When $\mathbb{E}[D_n^3] \rightarrow \mathbb{E}[D^3]$, we have that $d_{\max} = o(n^{1/3})$. Thus,

$$(4.4.116) \quad \sum_{v \in [n]} d_v^5 = o(n^{2/3}) \sum_{v \in [n]} d_v^3 = o(n^{5/3}).$$

We conclude that

$$(4.4.117) \quad f_3(\mathbf{I}) - \mathbb{E}[f_3(\mathbf{I})] = O_{\mathbb{P}}(n^{5/6}),$$

which, together with the fact that

$$(4.4.118) \quad \mathbb{E}[f_3(\mathbf{I})] = \sqrt{\pi_n(\lambda)}^3 \sum_{v \in [n]} d_v(d_v-1)(d_v-2) = n\sqrt{\pi}^3 \mathbb{E}[D(D-1)(D-2)](1+o(1)),$$

completes the claim. \square

Removing the red vertices. Denote by $\mathcal{C}'_{(i)}(\pi_n(\lambda))$ the connected components of $\text{CM}_N(\mathbf{d}(\pi))$ ordered by size. By Lemma 4.30, the degrees $\mathbf{d}(\pi_n(\lambda))$ satisfy the conditions in Theorem 4.13. We note that the surplus of $\mathcal{C}'_{(i)}(\pi_n(\lambda))$ is the same as the surplus of $\mathcal{C}_{(i)}(\pi_n(\lambda))$, since we only remove vertices of degree 1, which cannot break any cycles. Thus, we only need to investigate the effect of the removal of the red vertices on the cluster sizes. This is the content of the following lemma:

LEMMA 4.31 (Number of red vertices in critical configuration models). *The number of red vertices $\rho_n(\mathcal{C}'_{(i)}(\pi_n(\lambda)))$ in $\mathcal{C}'_{(i)}(\pi_n(\lambda))$ satisfies*

$$(4.4.119) \quad \frac{\rho_n(\mathcal{C}'_{(i)}(\pi_n(\lambda)))}{|\mathcal{C}'_{(i)}(\pi_n(\lambda))|} \xrightarrow{\mathbb{P}} \frac{\mathbb{E}[D](1 - \sqrt{\pi})}{1 + \mathbb{E}[D](1 - \sqrt{\pi})}.$$

PROOF. By (4.4.21) in Theorem 4.13,

$$(4.4.120) \quad \frac{v_1(\mathcal{C}'_{(i)}(\pi_n(\lambda)))}{|\mathcal{C}'_{(i)}(\pi_n(\lambda))|} \xrightarrow{\mathbb{P}} p_1^*(\pi),$$

where, by (4.3.12)–(4.3.13),

$$(4.4.121) \quad p_1^*(\pi) = (1 - \sqrt{\pi}) + \frac{1}{\mathbb{E}[D]} \sum_{\ell \geq 1} p_\ell b_{1,\ell}(\sqrt{\pi}).$$

The total number of removed vertices is equal to N^+ , which, by (4.3.4)–(4.3.5), equals

$$(4.4.122) \quad N^+ = n\mathbb{E}[D](1 - \sqrt{\pi}).$$

The total number of vertices of degree 1 is, by (4.3.12) and (4.3.5), equal to

$$(4.4.123) \quad Np_1(\pi)(1 + o_{\mathbb{P}}(1)) = n \left[\mathbb{E}[D](1 - \sqrt{\pi}) + p_\ell b_{1,\ell}(\sqrt{\pi}) \right].$$

Therefore, the number of removed vertices is

$$(4.4.124) \quad \begin{aligned} v_1(\mathcal{C}'_{(i)}(\pi_n(\lambda))) \frac{N^+}{Np_1(\pi)} (1 + o_{\mathbb{P}}(1)) &= |\mathcal{C}'_{(i)}(\pi_n(\lambda))| p_1^*(\pi) \frac{N^+}{Np_1(\pi)} (1 + o_{\mathbb{P}}(1)) \\ &= |\mathcal{C}'_{(i)}(\pi_n(\lambda))| \frac{\mathbb{E}[D](1 - \sqrt{\pi})}{1 + \mathbb{E}[D](1 - \sqrt{\pi})}, \end{aligned}$$

as required. \square

Proof of Theorem 4.9. To prove Theorem 4.9, we rely on Janson's construction in Lemma 4.6. We first investigate the scaling of the critical clusters for $\text{CM}_N(\mathbf{d}(\pi))$ for $\pi = \pi_n(\lambda)$

as in (4.4.9). By (4.4.103) in Lemma 4.30, as well as the convergence of the third moment of the degrees, Theorem 4.13 can be applied. It yields the convergence $\mathbf{Z}_n \xrightarrow{d} \mathbf{Z}$, where

$$(4.4.125) \quad \mathbf{Z}_N(\lambda) = \left((N^{-2/3} |\mathcal{C}'_{(j)}(\pi_n)|, \text{SP}(\mathcal{C}'_{(j)}(\pi_n))) \right)_{j \geq 1},$$

where we should bear in mind that the asymptotics now is for $N \rightarrow \infty$ and we use $\mathcal{C}'_{(j)}(\pi_n)$ for the ordered clusters in $\text{CM}_N(\mathbf{d}(\pi))$ for $\pi = \pi_n(\lambda)$. Note that

$$(4.4.126) \quad \nu_N(\pi_n(\lambda)) = 1 + \lambda' N^{-1/3} + O_{\mathbb{P}}(n^{-1/2}),$$

with $\lambda' = \kappa^{-1/3} \lambda$ by (4.3.5) and writing $\kappa = (1 + \mathbb{E}[D](1 - \sqrt{\pi_c}))^{-1}$ for the limit in probability of n/N . Then, the constants in the limiting Brownian motion with a negative parabolic drift are given by (4.4.14) (recall (4.4.18)), with λ replaced by λ' in (4.4.12).

Removing the red vertices has no effect on the surpluses since $\text{SP}(\mathcal{C}'_{(j)}(\pi_n)) = \text{SP}(\mathcal{C}_{(j)}(\pi_n))$, so we are left to inspect the cluster sizes. For this, we rewrite

$$(4.4.127) \quad N^{-2/3} |\mathcal{C}'_{(j)}(\pi_n)| = (n/N)^{2/3} n^{-2/3} |\mathcal{C}'_{(j)}(\pi_n)| = \kappa^{-2/3} n^{-2/3} |\mathcal{C}'_{(j)}(\pi_n)| (1 + o_{\mathbb{P}}(1)).$$

By Lemma 4.31,

$$(4.4.128) \quad |\mathcal{C}'_{(i)}(\pi_n(\lambda))| = |\mathcal{C}'_{(i)}(\pi_n(\lambda))| - \rho_n(\mathcal{C}'_{(i)}(\pi_n(\lambda))) = \kappa |\mathcal{C}'_{(i)}(\pi_n(\lambda))| (1 + o_{\mathbb{P}}(1)).$$

Therefore, the convergence of the ordered cluster sizes in Theorem 4.9 follows from the one proved in Theorem 4.13 with λ being replaced by $\lambda' = \kappa^{1/3} \lambda$ and the cluster sizes having an additional multiplicative factor $\kappa^{1/3}$. \square

4.5. Percolation on configuration models with infinite third-moment degrees

In the previous section, we have focused on the configuration model with finite third-moment degrees, and have shown that the critical behavior of percolation on such graphs is in the same universality class as the Erdős-Rényi random graph. This can be understood by the fact that the number of vertices found in one exploration is $d_{(i)}$, where $d_{(i)} = d_{v_i}$ with $(v_i)_{i \in [n]}$ the size-biased reordering of the vertices. Therefore,

$$(4.5.1) \quad \text{Var}(d_{(1)}) = \frac{\mathbb{E}[D_n^3]}{\mathbb{E}[D_n]} - \left(\frac{\mathbb{E}[D_n^2]}{\mathbb{E}[D_n]} \right)^2 \rightarrow \frac{\mathbb{E}[D^3]}{\mathbb{E}[D]} - \left(\frac{\mathbb{E}[D^2]}{\mathbb{E}[D]} \right)^2 < \infty$$

when Condition 1.6(a) holds and $\mathbb{E}[D_n^3] \rightarrow \mathbb{E}[D^3]$. Thus, the exploration process has *finite variance*, so that it is reasonable to expect that its scaling limit is closely related to Brownian motion. In this section, we will look at the heavy-tailed setting, where Conditions 1.6(a)-(b) hold, but $\mathbb{E}[D^3] = \infty$. We follow the work with Dhara, van Leeuwen and Sen [104] closely. Let us start by explaining informally what the size of the large critical connected components are, and on what scale the exploration process runs.

Informal explanation of the scaling exponents. We extend the analysis around (4.4.1). Let us again for simplicity assume that we are in the situation of a critical configuration model. Recall that

$$(4.5.2) \quad \mathbb{P}(|\mathcal{C}_{\max}(\pi_c)| \geq k) = \mathbb{P}(Z_{\geq k} \geq k) \leq \frac{1}{k} \mathbb{E}[Z_{\geq k}] = \frac{n}{k} \mathbb{P}(|\mathcal{C}_U(\pi_c)| \geq k),$$

where $Z_{\geq k}$ denotes the number of vertices in components of size at least k . Using the scaling for the branching process total progeny tail in (4.1.7), implying that $\mathbb{P}(|\mathcal{C}_o(\pi_c)| >$

$k) = \frac{c}{k^{1/(\tau-2)}}(1 + o(1))$, leads to

$$(4.5.3) \quad \mathbb{P}(|\mathcal{C}_{\max}(\pi_c)| \geq k) = O\left(\frac{n}{k^{(\tau-1)/(\tau-2)}}\right).$$

Now we take $k = k_n = An^{(\tau-2)/(\tau-1)}$ for some large A (again fingers crossed!), to arrive at

$$(4.5.4) \quad \mathbb{P}(|\mathcal{C}_{\max}(\pi_c)| \geq An^{(\tau-2)/(\tau-1)}) = O\left(\frac{1}{A^{(\tau-1)/(\tau-2)}}\right),$$

which is small when $A > 0$ is large. This suggests that $|\mathcal{C}_{\max}(\pi_c)| = O_{\mathbb{P}}(n^{(\tau-2)/(\tau-1)})$, as we will derive in more detail in this section.

Let us extend the above argument to the scaling of the exploration process. Recall from Theorem 4.15 (see also (4.4.26)) that the cluster exploration process of the configuration model with finite-third degrees runs on the scale $n^{1/3}$ and the relevant time scale is $n^{2/3}$. The scale $n^{2/3}$ is equal to the cluster size, as the largest excursions have the same order of magnitude and excursions correspond to the complete exploration of clusters, while $n^{1/3}$ should be thought of as the typical order of the number of unexplored or active half-edges during the exploration.

When $\tau \in (3, 4)$, the maximal degree has size $\Theta(n^{1/(\tau-1)})$, which is much larger than $n^{1/3}$. When such a high-degree vertex is found in the exploration, the exploration process makes a jump of that same order. This suggests that for $\tau \in (3, 4)$, the exploration process runs on scale $n^{1/(\tau-1)}$ instead, and makes a macroscopic jump precisely when one of the vertices of maximal degrees is being found. The time needed for the exploration process to discover one of the vertices of degree $n^{1/(\tau-1)}$ is equal to the inverse of the probability of pairing to a vertex of degree of order $n^{1/(\tau-1)}$, which equals $n/n^{1/(\tau-1)} = n^{(\tau-2)/(\tau-1)}$, confirming the guess for the largest cluster size derived above.

This intuition gives us a clear idea that the exploration of clusters in the regime $\tau \in (3, 4)$ is entirely governed by the vertices of maximal degree, and this will be the intuition behind the formal results that we will state now. We start by introducing some notation.

Notation and limiting processes. Consider a decreasing sequence $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots) \in \ell_{\downarrow}^3 \setminus \ell_{\downarrow}^2$. Denote by $\mathcal{I}_i(s) := \mathbb{1}_{\{\xi_i \leq s\}}$ where $\xi_i \sim \text{Exp}(\theta_i/\mu)$ independently, and $\text{Exp}(r)$ denotes the exponential distribution with rate r . Consider the process

$$(4.5.5) \quad \bar{S}_{\infty}^{\lambda}(t) = \sum_{i=1}^{\infty} \theta_i (\mathcal{I}_i(t) - (\theta_i/\mu)t) + \lambda t,$$

for some $\lambda \in \mathbb{R}$, $\mu > 0$ and define the reflected version of $\bar{S}_{\infty}^{\lambda}(t)$ by

$$(4.5.6) \quad \bar{R}_{\infty}^{\lambda}(t) = \bar{S}_{\infty}^{\lambda}(t) - \min_{0 \leq u \leq t} \bar{S}_{\infty}^{\lambda}(u).$$

Processes of the form (4.5.5) were termed *thinned Lévy process* in work with Bhamidi and van Leeuwaarden [44], since the summands are thinned version of Poisson processes. Indeed, we can interpret $\mathcal{I}_i(t) = \mathbb{1}_{\{N_i(t) \geq 1\}}$, where $(N_i(t))_{i \geq 1}$ are independent Poisson processes of rate θ_i/μ . The process

$$(4.5.7) \quad \bar{L}_{\infty}^{\lambda}(t) = \sum_{i=1}^{\infty} \theta_i (N_i(t) - (\theta_i/\mu)t) + \lambda t$$

is a Lévy process:

EXERCISE 4.29 (Lévy process). *Show that the process in (4.5.7) is a Lévy process.*

Obviously, $\bar{S}_\infty^\lambda(t) \leq \bar{L}_\infty^\lambda(t)$ holds. In what follows, we will explain that the indicator processes $(\mathcal{I}_i(t))_{i \geq 1}$ describe whether the vertex with the i th largest degree has been found by the exploration or not, and then $\theta_i \mathcal{I}_i(t)$ describes the jump made by the exploration process when i has been found. Clearly, in the graph context, any vertex can only be found *once* for the first time, which explains why the process $(\mathcal{I}_i(t))_{i \geq 1}$ is relevant rather than the Poisson process $(N_i(t))_{i \geq 1}$. Thus, the fact that we deal with a thinned Lévy process is directly related to the depletion-of-points effect present in exploring random graphs.

Let us continue to describe the necessary notion for the scaling limits of cluster sizes for critical percolation on $\text{CM}_n(\mathbf{d})$ in the infinite third-moment degree setting. For any function $f \in \mathbb{D}[0, \infty)$, define $\underline{f}(x) = \inf_{y \leq x} f(y)$. $\mathbb{D}_+[0, \infty)$ is the subset of $\mathbb{D}[0, \infty)$ consisting of functions with positive jumps only. Note that \underline{f} is continuous when $f \in \mathbb{D}_+[0, \infty)$. An *excursion* of a function $f \in \mathbb{D}_+[0, T]$ is an interval (l, r) such that

$$(4.5.8) \quad \min\{f(l-), f(l)\} = \underline{f}(l) = \underline{f}(r) = \min\{f(r-), f(r)\} \quad \text{and} \\ f(x) > \underline{f}(r), \quad \forall x \in (l, r) \subset [0, T].$$

Excursions of a function $f \in \mathbb{D}_+[0, \infty)$ are defined similarly. We use γ to denote an excursion, as well as the length of the excursion γ , to simplify notation.

Also, define the counting process \mathbf{N} to be the Poisson process that has intensity $\bar{R}_\infty^\lambda(t)$ at time t conditional on $(S_\infty^\lambda(u))_{u \leq t}$. Formally, \mathbf{N} is characterized as the counting process for which

$$(4.5.9) \quad N(t) - \int_0^t \bar{R}_\infty^\lambda(u) du$$

is a martingale (recall (4.4.15) where a similar process was defined in the finite-third moment setting). We use the notation $N(\gamma)$ to denote the number of marks in the interval γ .

Finally, we define a Markov process $(\mathbf{Z}(s))_{s \in \mathbb{R}}$ on $\mathbb{D}(\mathbb{R}, \mathbb{U}_+^0)$, called the *augmented multiplicative coalescent* (AMC) process. Think of a collection of particles in a system with $\mathbf{X}(s)$ describing their masses and $\mathbf{Y}(s)$ describing an additional attribute at time s . Let $K_1, K_2 > 0$ be constants. The evolution of the system takes place according to the following rule at time s :

- ▷ For $i \neq j$, at rate $K_1 X_i(s) X_j(s)$, the i th and j th components merge and create a new component of mass $X_i(s) + X_j(s)$ and attribute $Y_i(s) + Y_j(s)$.
- ▷ For any $i \geq 1$, at rate $K_2 X_i^2(s)$, $Y_i(s)$ increases to $Y_i(s) + 1$.

Of course, at each event time, the indices are re-organized to give a proper element of \mathbb{U}_+^0 . This process was first introduced by Bhamidi, Budhiraja and Wang in [36] to study the joint behavior of the component sizes and the surplus edges over the critical window for so-called bounded size rules processes. Broutin and Marckert [69] introduced it for the Erdős-Rényi random graph. Bhamidi, Budhiraja and Wang [36] extensively study the properties of the standard version of AMC, i.e., the case $K_1 = 1, K_2 = 1/2$ and showed in [36, Theorem 3.1] that this is a (nearly) Feller process.

Main results for heavy-tailed configuration models. Throughout this section we use the following shorthand notation

$$(4.5.10a) \quad \alpha = 1/(\tau - 1), \quad \rho = (\tau - 2)/(\tau - 1), \quad \eta = (\tau - 3)/(\tau - 1),$$

$$(4.5.10b) \quad a_n = n^\alpha L(n), \quad b_n = n^\rho (L(n))^{-1}, \quad c_n = n^\eta (L(n))^{-2},$$

where $\tau \in (3, 4)$ and $L(\cdot)$ is a slowly-varying function. These sequences turn out to describe the sizes of the maximal degrees, the sizes of the largest connected components, and the width of the scaling window, respectively. Here, we will assume that the degree distribution obeys a power-law with exponent $\tau \in (3, 4)$, and with slowly-varying corrections that are related to $n \mapsto L(n)$.

Before stating our main result, let us define the necessary conditions on the degrees under which it applies. Assume that there exists $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots) \in \ell_\downarrow^3 \setminus \ell_\downarrow^2$ such that, for any fixed $i \geq 1$,

$$(4.5.11a) \quad \frac{d_i}{a_n} \rightarrow \theta_i,$$

$$(4.5.11b) \quad \lim_{K \rightarrow \infty} \limsup_{n \rightarrow \infty} a_n^{-3} \sum_{i=K+1}^n d_i^3 = 0,$$

Equation (4.5.11a) shows that a_n in (4.5.10b) describes the order of the largest degrees, while (4.5.11b) shows that the largest degrees provide the main contribution to the third moments of the degrees.

The main result in this section is the following theorem, that describes scaling of the largest critical connected components for percolation on the configuration model with heavy-tailed degrees:

THEOREM 4.32 (Critical percolation on heavy-tailed configuration models). *Consider $\text{CM}_n(\mathbf{d})$ with the degrees satisfying Condition 1.6(a)-(c) with $\nu > 1$. Further, assume that (4.5.11a)–(4.5.11b) hold for any fixed $i \geq 1$, where $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots) \in \ell_\downarrow^3 \setminus \ell_\downarrow^2$. Assume that the percolation parameter $\pi_n(\lambda)$ satisfies*

$$(4.5.12) \quad \pi_n(\lambda) := \frac{1}{\nu_n} (1 + \lambda c_n^{-1} + o(c_n^{-1}))$$

for some $\lambda \in \mathbb{R}$. Let $\tilde{\mathbf{S}}_\infty^\lambda$ denote the process in (4.5.5) with θ_i replaced by $\theta_i/\sqrt{\nu}$. Let

$$(4.5.13) \quad \mathbf{Z}_n(\lambda) := (b_n^{-1} |\mathcal{C}_{(j)}(\pi_n(\lambda))|, \text{SP}(\mathcal{C}_{(j)}(\pi_n(\lambda)))_{j \geq 1}), \quad \mathbf{Z}(\lambda) := (\nu^{1/2} \tilde{\gamma}_i(\lambda), N(\tilde{\gamma}_i(\lambda)))_{i \geq 1},$$

where $\tilde{\gamma}_i(\lambda)$ is the largest excursion of $\tilde{\mathbf{S}}_\infty^\lambda$. Then, for any $\lambda \in \mathbb{R}$, as $n \rightarrow \infty$,

$$(4.5.14) \quad \mathbf{Z}_n(\lambda) \xrightarrow{d} \mathbf{Z}(\lambda)$$

with respect to the \mathbb{U}_\downarrow^0 topology.

Note that $\tau \in (3, 4)$, so that $b_n = n^{\rho+o(1)}$ with $\rho = (\tau - 2)/(\tau - 1) \in (\frac{1}{2}, \frac{2}{3})$. Thus, the largest critical connected components of random graphs with *larger* degrees become smaller. This appears to be somewhat counterintuitive. It is related to the statement that the survival probability of a critical branching process with infinite-variance offspring is *smaller* than that of a branching process with finite variance offspring. These connections

are studied in quite some detail (and under much weaker conditions on the degrees) in work with Janson and Luczak [166].

We will not give all the details of the proof of Theorem 4.32, and mainly remark on the differences with the proof of Theorem 4.9 in the finite third moment case. As in the proof of Theorem 4.9, where we first have reduced the proof of Theorem 4.9 to Theorem 4.13 describing critical percolation clusters in the configuration model setting, we again formulate a main result for critical configuration models with heavy-tailed degrees:

THEOREM 4.33 (Heavy-tailed configuration models at criticality). *Let $\text{SP}(\mathcal{C}_{(i)})$ denote the number of surplus edges in $\mathcal{C}_{(i)}$ and let $\mathbf{Z}_n := \text{ord}(b_n^{-1}|\mathcal{C}_{(i)}|, \text{SP}(\mathcal{C}_{(i)}))_{i \geq 1}$ and $\mathbf{Z} := \text{ord}(\gamma_i(\lambda), N(\gamma_i))_{i \geq 1}$. Consider $\text{CM}_n(\mathbf{d})$ where the degrees satisfy Conditions 1.6(a)-(c) with $\nu_n = 1 + \lambda c_n^{-1} + o(c_n^{-1})$. Further, assume that (4.5.11a)-(4.5.11b) hold. Then, as $n \rightarrow \infty$,*

$$(4.5.15) \quad \mathbf{Z}_n \xrightarrow{d} \mathbf{Z}$$

with respect to the \mathbb{U}_\downarrow^0 topology, where \mathbf{N} is defined in (4.5.9).

The reduction of Theorem 4.32 to Theorem 4.33 is a straightforward adaptation of the reduction of the proof of Theorem 4.9 to Theorem 4.13, and will be omitted here. Rather, let us start by discussing its relevance.

Obviously, Theorem 4.33 is interesting in its own right, so before continuing with the proof, let us discuss its history in a little more detail. There are three early papers investigating the critical behaviour of random graphs with heavy-tailed degrees. I started this line of research in [159] by investigating the scaling behavior of critical rank-1 inhomogeneous random graphs.⁴ There, it became obvious that the setting of $\tau \in (3, 4)$ is different from that for finite third-moment degrees. With Bhamidi and van Leeuwen [44], we looked at scaling limits of the maximal cluster sizes, proving a theorem alike Theorem 4.33, but slightly weaker as the AMC was not involved, for rank-1 inhomogeneous random graphs in the heavy-tailed degrees case. A related result was proved by Joseph [189], where he analyzed critical configuration models with i.i.d. degrees. As it turns out, the scaling limit in this setting is different compared to that in Theorem 4.33, due to the randomness in the degrees. Let us discuss this in some more detail:

Configuration model with i.i.d. degrees. A natural setting to investigate the role of heavy-tailed degrees is to start with i.i.d. random degrees having a power-law distribution with $\tau \in (3, 4)$. This is the setting investigated by Joseph [189]. We now relate this to the setting investigated so far. When dealing with i.i.d. degrees, it is helpful to *couple* the degrees in the configuration model for different values of n to one another. Indeed, otherwise we cannot ever hope that (4.5.11a)-(4.5.11b) hold for any fixed $i \geq 1$. To couple the degrees, we use the fact that when $(E_i)_{i \geq 1}$ are i.i.d. exponentials, and $\Gamma_i = E_1 + \dots + E_i$ is the sum of the first i of them, then the vector $(\Gamma_i/\Gamma_{n+1})_{i \in [n]}$ has the same distribution as the order statistics of $(U_i)_{i \in [n]}$:

EXERCISE 4.30 (Uniform order statistics and sums of exponentials). *Prove that the vector $(\Gamma_i/\Gamma_{n+1})_{i \in [n]}$ has the same distribution as the order statistics of $(U_i)_{i \in [n]}$.*

⁴The publication dates are slightly misleading, as [159] was completed early 2009, but only appeared in 2013.

Let F_D be the distribution function of the i.i.d. degrees $(D_i)_{i \in [n]}$, so that D_i has the same distribution as $[1 - F_D]^{-1}(U_i)$. Let $(D_{(i)})_{i \in [n]}$ be the order statistics, ordered such that $D_{(1)} \leq D_{(2)} \leq \dots \leq D_{(n)}$. Then, we use that

$$(4.5.16) \quad (D_{(n-i+1)})_{i \in [n]} \stackrel{d}{=} \left([1 - F_D]^{-1}(U_{(i)}) \right)_{i \in [n]} \stackrel{d}{=} \left([1 - F_D]^{-1}(\Gamma_i/\Gamma_{n+1}) \right)_{i \in [n]}.$$

Equation (4.5.16) provides a highly powerful *coupling* between the degree sequences for random graphs of different sizes, which are such that in particular the *maximal degrees* are closely related. This mimics the setting in (4.5.11a)-(4.5.11b), and we now show that the convergence result in it holds almost surely in this setting.

For this, we note that, by the Strong Law of Large Numbers,

$$(4.5.17) \quad \frac{1}{n} \Gamma_{n+1} \xrightarrow{a.s.} 1,$$

so that we may replace Γ_{n+1} in (4.5.16) by n . Further, assume that D has a power-law distribution, in the sense that, when $x \rightarrow \infty$, there exist $\tau \in (3, 4)$ and a constant c_D such that

$$(4.5.18) \quad [1 - F_D](x) = c_D x^{-(\tau-1)}(1 + o(1)).$$

In this case, for $u \in (0, 1)$ and as $u \searrow 0$,

$$(4.5.19) \quad [1 - F_D]^{-1}(u) = (c_D/u)^{-1/(\tau-1)}(1 + o(1)).$$

Combining these ingredients leads us to the statement that, almost surely and for every i ,

$$(4.5.20) \quad n^{-1/(\tau-1)} D_{(n-i+1)} \xrightarrow{a.s.} (c_D/\Gamma_i)^{1/(\tau-1)},$$

thus proving the convergence statement in (4.5.11a) with $\theta_i = (c_D/\Gamma_i)^{1/(\tau-1)}$. Since $\Gamma_i/i \xrightarrow{a.s.} 1$ and $\tau \in (3, 4)$, we also have that $(\theta_i)_{i \geq 1} \in \ell_3 \setminus \ell_2$ almost surely, as required. This links the critical behavior of $\text{CM}_n(\mathbf{D})$ to that of $\text{CM}_n(\mathbf{d})$ with deterministic degrees.

Joseph [189] proves that the exploration process in the configuration model with i.i.d. degrees has *independent increments*. This is not true for our thinned Lévy process in (4.5.5). This process describes the critical clusters under the *conditional measure* \mathbb{P}_n where we condition on the degrees in the range where (4.5.20) holds. Apparently, this means that when substituting $\theta_i = (c_D/\Gamma_i)^{1/(\tau-1)}$ and considering the process under the *unconditional measure* $\mathbb{P}(\cdot) = \mathbb{E}[\mathbb{P}_n(\cdot)]$, where the expectation is with respect to the randomness in the degrees, the thinned Lévy process in (4.5.5) has independent increments. That is the content of the following open problem:

OPEN PROBLEM 4.1 (Relation CM with i.i.d. and deterministic degrees). *Unravel the relation between critical configuration models with power-law degrees with exponent $\tau \in (3, 4)$ by showing that the thinned Lévy process in (4.5.5) has, under the measure $\mathbb{P}(\cdot) = \mathbb{E}[\mathbb{P}_n(\cdot)]$, independent increments.*

We have tried to prove this, but due to the highly involved way how the randomness of the degrees enters into the limiting excursion process (4.5.5), we have not been able to do so.

Scaling limit critical heavy-tailed CM: ingredients proof of Theorem 4.33.

We now discuss some of the ingredients to the proof of Theorem 4.33. This proof again relies on an appropriate *exploration* of the critical components of $\text{CM}_n(\mathbf{d})$ in the setting where the degrees satisfy Condition 1.6(a)-(c) with $\nu_n = 1 + \lambda c_n^{-1} + o(c_n^{-1})$ and (4.5.11a)–(4.5.11b). However, the properties of this scaling limit are quite different. Indeed, the exploration process describing the number of active half-edges at any time within the exploration of a large cluster now consists of infinitely many jumps, most of them being extremely small. These jumps correspond to the sum over $i \geq 1$ in (4.5.5). Informally, this exploration makes a jump when the active vertex being explored connects to vertex i having degree d_i . Since $d_i \approx \theta_i n^\alpha$ with $\alpha = 1/(\tau - 1)$, the jump is then of size $a_n = n^{\alpha+o(1)}$. When $\tau \in (3, 4)$, $\alpha = 1/(\tau - 1) \in (\frac{1}{2}, \frac{1}{3})$, so that the number of unpaired half-edges is much larger than the $n^{1/3}$ in the finite-third moment case (recall Theorem 4.15 and in particular, (4.4.26)). As a result, the exploration process is characterized by its discovery of the *high-degree vertices*, where all of them contribute, but unequally, the main jumps coming from the few vertices of largest degrees. We now move forward to make this picture precise.

Recall Algorithm 4.14 for the depth-first exploration of the components of the configuration model. We now adapt this exploration algorithm somewhat in the context of the current set-up:

ALGORITHM 4.34 (Exploring the graph). Consider the configuration model $\text{CM}_n(\mathbf{d})$. The algorithm carries along vertices that can be alive, active, exploring and killed and half-edges that can be alive, active or killed. We sequentially explore the graph as follows:

- (S0) At stage $i = 0$, all the vertices and the half-edges are *alive* but none of them are *active*. Also, there are no *exploring* vertices.
- (S1) At each stage i , if there is no active half-edge at stage i , then choose a vertex v proportional to its degree among the alive (not yet killed) vertices and declare all its half-edges to be *active* and declare v to be *exploring*. If there is an active vertex but no exploring vertex, then declare the *smallest* vertex to be exploring.
- (S2) At each stage i , take an active half-edge e of an exploring vertex v and pair it uniformly to another alive half-edge f . Kill e, f . If f is incident to a vertex v' that has not been discovered before, then declare all the half-edges incident to v' active, except f (if any). If $d_{v'} = 1$ (i.e., the only half-edge incident to v' is f) then kill v' . Otherwise, declare v' to be active and larger than all other vertices that are alive. After killing e , if v does not have another active half-edge, then kill v also.
- (S3) Repeat from (S1) at stage $i + 1$ if not all half-edges are already killed.

Algorithm 4.34 gives a *breadth-first* exploration of the connected components of $\text{CM}_n(\mathbf{d})$, rather than the depth-first in Algorithm 4.14. Define the exploration process by

$$(4.5.21) \quad S_n(0) = 0, \quad S_n(l) = S_n(l-1) + d_{(l)} J_l - 2,$$

where J_l is the indicator that a new vertex is discovered at time l and $d_{(l)}$ is the degree of the new vertex chosen at time l when $J_l = 1$. Suppose \mathcal{C}_k is the k th connected component

explored by the above exploration process and define $\tau_k = \inf \{i: S_n(i) = -2k\}$. Then \mathcal{C}_k is discovered between the times $\tau_{k-1} + 1$ and τ_k , and $\tau_k - \tau_{k-1} - 1$ gives the total number of edges in \mathcal{C}_k . Here we note that while Algorithm 4.14 assures that each time a step is made within the exploration a *new* vertex is found, Algorithm 4.34 does not do this, which explains the difference between (4.5.21) and its companion in the finite third moment case in (4.4.23). There is a crucial difference between these two settings. When the degrees have finite third moment, the contribution to the exploration process of single vertices is small, and even when we count the half-edges incident to a vertex several times during the exploration, we would still get the same scaling limit. In the infinite third moment setting, this is far from true, as the exploration process is governed by the high-degree vertices. Thus, we need to be much more careful when exploring vertices several times.

Call a vertex *discovered* if it is either active or killed. Let \mathcal{V}_l denote the set of vertices discovered up to time l and $\mathcal{I}_i^n(l) := \mathbb{1}_{\{i \in \mathcal{V}_l\}}$. Note that

$$(4.5.22) \quad S_n(l) = \sum_{i \in [n]} d_i \mathcal{I}_i^n(l) - 2l = \sum_{i \in [n]} d_i \left(\mathcal{I}_i^n(l) - \frac{d_i}{\ell_n} l \right) + (\nu_n(\lambda) - 1) l.$$

Recall the notation in (4.5.10b). Define the re-scaled version $\bar{\mathbf{S}}_n$ of \mathbf{S}_n by

$$(4.5.23) \quad \bar{S}_n(t) = a_n^{-1} S_n(\lfloor b_n t \rfloor).$$

Then, by (4.5.11a),

$$(4.5.24) \quad \bar{S}_n(t) = a_n^{-1} \sum_{i \in [n]} d_i \left(\mathcal{I}_i^n(tb_n) - \frac{d_i}{\ell_n} tb_n \right) + \lambda t + o(1).$$

Note the similarity between the expressions in (4.5.5) and (4.5.24). We will prove the following theorem describing the scaling limit of the exploration process for $\text{CM}_n(\mathbf{d})$ in the setting where the degree distribution is an asymptotic power-law with exponent $\tau \in (3, 4)$:

THEOREM 4.35 (Scaling limit of exploration process CM with heavy-tailed degrees). *Consider the process $\bar{\mathbf{S}}_n := (\bar{S}_n(t))_{t \geq 0}$ defined in (4.5.24) and recall the definition of $\bar{\mathbf{S}}_\infty := (\bar{S}_\infty(t))_{t \geq 0}$ from (4.5.5). Then,*

$$(4.5.25) \quad \bar{\mathbf{S}}_n \xrightarrow{d} \bar{\mathbf{S}}_\infty$$

with respect to the Skorohod J_1 topology.

Since $\mathcal{I}_i^n(tb_n)$ denotes the indicator whether vertex $i \in [n]$ has been explored, we see that indeed the vertices of high degrees contribute the majority to the exploration process. The proof will make this connection even more transparent. Since the excursions of $\bar{\mathbf{S}}_n$ describe the sizes of the connected components through the relation that the number of edges in \mathcal{C}_k equals $\tau_{k+1} - \tau_k - 1$, where $\tau_k = \inf \{i: S_n(i) = -2k\}$, the convergence of the exploration process implies the convergence of the number of edges in the largest connected components, which is a key step towards proving Theorem 4.33. Further, it is relatively easy to deduce that the number of vertices satisfies the same scaling, since the surplus of connected components remains tight. As a result, Theorem 4.33 follows relatively easily from Theorem 4.35. We omit these details, and rather spend some more time on the proof of the scaling limit of the exploration process.

Scaling limit exploration processes critical heavy-tailed configuration models. Let us now proceed by giving some insight in how the limiting process (4.5.5) arises. In the reformulation of the exploration process in (4.5.24), we see that the only randomness arises in the random indicator processes $(\mathcal{I}_i^n(t))_{i \in [n], t \geq 0}$, where $\mathcal{I}_i^n(tb_n) = \mathcal{I}_i^n(\lfloor tb_n \rfloor)$.

Fix $K \geq 1$ to be large. Define the truncated exploration process

$$(4.5.26) \quad \bar{S}_n^{\leq K}(t) = a_n^{-1} \sum_{i \in [K]} d_i \left(\mathcal{I}_i^n(tb_n) - \frac{d_i}{\ell_n} tb_n \right) + \lambda t,$$

and let

$$(4.5.27) \quad \bar{S}_n^{>K}(t) = \bar{S}_n(t) - \bar{S}_n^{\leq K}(t) = a_n^{-1} \sum_{i=K+1}^n d_i \left(\mathcal{I}_i^n(tb_n) - \frac{d_i}{\ell_n} tb_n \right) + \lambda t.$$

We will show that (a) $\bar{S}_n^{>K}(t)$ is uniformly small; and (b) $\bar{S}_n^{\leq K}(t)$ converges in distribution to the truncated version of the limiting process $\bar{S}_\infty^{\leq K}(t)$ given by

$$(4.5.28) \quad S^{\leq K}(t) = \sum_{i=1}^K \theta_i (\mathcal{I}_i(t) - \theta/\mu t) + \lambda t,$$

where we recall that $\mathcal{I}_i(s) := \mathbb{1}_{\{\xi_i \leq s\}}$ where $\xi_i \sim \text{Exp}(\theta_i/\mu)$. We start with the statement that $\bar{S}_n^{>K}(t)$ is uniformly small, for which we use a martingale argument:

LEMMA 4.36 (High-degree vertices run the show). *Fix $\varepsilon > 0$. Then,*

$$(4.5.29) \quad \lim_{K \rightarrow \infty} \limsup_{n \rightarrow \infty} \mathbb{P} \left(\sup_{l \leq Tb_n} |S_n^{>K}(l)| > \varepsilon \right) = 0.$$

PROOF. We start by introducing some notation. Let \mathcal{F}_l be the sigma-field containing the information generated up to time l by Algorithm 4.34. Also, let Υ_l denote the set of time points up to time l when a component was discovered and $v_l = |\Upsilon_l|$. Note that we have lost $2(l - v_l)$ half-edges by time l . Thus, on the set $\{\mathcal{I}_i^n(l) = 0\}$,

$$(4.5.30) \quad \mathbb{P}(\mathcal{I}_i^n(l+1) = 1 \mid \mathcal{F}_l) = \begin{cases} \frac{d_i}{\ell_n - 2(l - v_l) - 1} & \text{if } l \notin \Upsilon_l, \\ \frac{d_i}{\ell_n - 2(l - v_l)} & \text{otherwise,} \end{cases}$$

and, uniformly over $l \leq Tb_n$,

$$(4.5.31) \quad \mathbb{P}(\mathcal{I}_i^n(l+1) = 1 \mid \mathcal{F}_l) \geq \frac{d_i}{\ell_n} \quad \text{on the set } \{\mathcal{I}_i^n(l) = 0\}.$$

Therefore,

$$(4.5.32) \quad \begin{aligned} \mathbb{E}[S_n^{>K}(l+1) - S_n^{>K}(l) \mid \mathcal{F}_l] &= \mathbb{E} \left[\sum_{i=K+1}^n a_n^{-1} d_i \left(\mathcal{I}_i^n(l+1) - \mathcal{I}_i^n(l) - \frac{d_i}{\ell_n} \right) \mid \mathcal{F}_l \right] \\ &= \sum_{i=K+1}^n a_n^{-1} d_i \left(\mathbb{E}[\mathcal{I}_i^n(l+1) \mid \mathcal{F}_l] \mathbb{1}_{\{\mathcal{I}_i^n(l)=0\}} - \frac{d_i}{\ell_n} \right) \geq 0. \end{aligned}$$

Thus $(S_n^{>K}(l))_{l=1}^{Tb_n}$ is a sub-martingale. Further, (4.5.30) implies that, uniformly for all $l \leq Tb_n$,

$$(4.5.33) \quad \mathbb{P}(\mathcal{I}_i^n(l) = 0) \geq \left(1 - \frac{d_i}{\ell_n} \right)^l,$$

where we abbreviate $\ell'_n = \ell_n - 2Tb_n - 1$. Thus, (4.5.11b) gives

$$\begin{aligned}
|\mathbb{E}[S_n^{>K}(l)]| &= a_n^{-1} \sum_{i=K+1}^n d_i \left(\mathbb{P}(\mathcal{I}_i^n(l) = 1) - \frac{d_i l}{\ell_n} \right) \\
&\leq a_n^{-1} \sum_{i=K+1}^n d_i \left(1 - \left(1 - \frac{d_i}{\ell'_n} \right)^l - \frac{d_i l}{\ell'_n} \right) + a_n^{-1} l \sum_{i \in [n]} d_i^2 \left(\frac{1}{\ell'_n} - \frac{1}{\ell_n} \right) \\
&\leq \frac{l^2}{2\ell_n^2 a_n} \sum_{i=K+1}^n d_i^3 + o(1) \\
(4.5.34) \quad &\leq \frac{T^2 n^{2\rho} n^{3\alpha} L(n)^3}{2\ell_n^2 L(n)^2 n^\alpha L(n)} \left(a_n^{-3} \sum_{i=K+1}^n d_i^3 \right) + o(1) = C \left(a_n^{-3} \sum_{i=K+1}^n d_i^3 \right) + o(1),
\end{aligned}$$

for some constant $C > 0$, where we have used the fact that

$$(4.5.35) \quad a_n^{-1} l \sum_{i \in [n]} d_i^2 \left(\frac{1}{\ell'_n} - \frac{1}{\ell_n} \right) = O(n^{2\rho+1-\alpha-2}/L(n)^3) = O(n^{(\tau-4)/(\tau-1)}/L(n)^3) = o(1),$$

uniformly for $l \leq Tb_n$. Therefore, uniformly over $l \leq Tb_n$,

$$(4.5.36) \quad \lim_{K \rightarrow \infty} \limsup_{n \rightarrow \infty} |\mathbb{E}[S_n^{>K}(l)]| = 0.$$

Now, note that for any (x_1, x_2, \dots) , $0 \leq a + b \leq x_i$ and $a, b > 0$ one has

$$\prod_{i=1}^R (1 - a/x_i)(1 - b/x_i) \geq \prod_{i=1}^R (1 - (a+b)/x_i).$$

Thus, by (4.5.30), for all $l \geq 1$ and $i \neq j$,

$$(4.5.37) \quad \mathbb{P}(\mathcal{I}_i^n(l) = 0, \mathcal{I}_j^n(l) = 0) \leq \mathbb{P}(\mathcal{I}_i^n(l) = 0)\mathbb{P}(\mathcal{I}_j^n(l) = 0),$$

so that $\mathcal{I}_i^n(l)$ and $\mathcal{I}_j^n(l)$ are negatively correlated. Observe also that, uniformly over $l \leq Tb_n$,

$$(4.5.38) \quad \text{Var}(\mathcal{I}_i^n(l)) \leq \mathbb{P}(\mathcal{I}_i^n(l) = 1) \leq \sum_{l_1=1}^l \mathbb{P}(\text{vertex } i \text{ is first discovered at stage } l_1) \leq \frac{ld_i}{\ell'_n}.$$

Therefore, using the negative correlation in (4.5.37), uniformly over $l \leq Tb_n$,

$$(4.5.39) \quad \text{Var}(S_n^{>K}(l)) \leq a_n^{-2} \sum_{i=K+1}^n d_i^2 \text{Var}(\mathcal{I}_i^n(l)) \leq \frac{l}{\ell'_n a_n^2} \sum_{i=K+1}^n d_i^3 \leq C a_n^{-3} \sum_{i=K+1}^n d_i^3,$$

for some constant $C > 0$. Thus, by (4.5.11b),

$$(4.5.40) \quad \lim_{K \rightarrow \infty} \limsup_{n \rightarrow \infty} \text{Var}(S_n^{>K}(l)) = 0,$$

uniformly for $l \leq Tb_n$. Using (4.5.36), (4.5.39), and (2.6.19) in the super-martingale inequality in Lemma 2.40, together with the fact that $(-S_n^{>K}(l))_{l=1}^{Tb_n}$ is a super-martingale, we arrive at (4.5.29), for any $\varepsilon > 0$. \square

We next prove the weak convergence of the truncated process $(\bar{S}_n^{\leq K}(t))_{t \geq 0}$:

LEMMA 4.37 (Convergence of indicator processes). *Fix any $K \geq 1$. As $n \rightarrow \infty$,*

$$(4.5.41) \quad (\mathcal{I}_i^n(tb_n))_{i \in [K], t \geq 0} \xrightarrow{d} (\mathcal{I}_i(t))_{i \in [K], t \geq 0}.$$

PROOF. By noting that $(\mathcal{I}_i^n(tb_n))_{t \geq 0}$ are indicator processes, it is enough to show that

$$(4.5.42) \quad \mathbb{P}(\mathcal{I}_i^n(t_i b_n) = 0, \forall i \in [K]) \rightarrow \mathbb{P}(\mathcal{I}_i(t_i) = 0, \forall i \in [K]) = \exp\left(-\sum_{i=1}^K \theta_i t_i / \mu\right).$$

for any $t_1, \dots, t_K \in \mathbb{R}$. Now,

$$(4.5.43) \quad \mathbb{P}(\mathcal{I}_i^n(m_i) = 0, \forall i \in [K]) = \prod_{l=1}^{\infty} \left(1 - \sum_{i \leq K: l \leq m_i} \frac{d_i}{\ell_n - \Theta(l)}\right),$$

where the $\Theta(l)$ term arises from the expression in (4.5.30) and noting that $v_l = |\Upsilon_l| \leq l$ since at most one vertex can be found each time a half-edge is paired in the exploration.

Taking logarithms on both sides of (4.5.43) and using the fact that $l \leq \max m_i = \Theta(b_n)$, we get

$$(4.5.44) \quad \begin{aligned} \mathbb{P}(\mathcal{I}_i^n(m_i) = 0 \forall i \in [K]) &= \exp\left(-\sum_{l=1}^{\infty} \sum_{i \leq K: l \leq m_i} \frac{d_i}{\ell_n} + o(1)\right) \\ &= \exp\left(-\sum_{i \in [K]} \frac{d_i m_i}{\ell_n} + o(1)\right). \end{aligned}$$

Putting $m_i = t_i b_n$, (4.5.11a) gives

$$(4.5.45) \quad \frac{m_i d_i}{\ell_n} = \frac{\theta_i t_i}{\mu} (1 + o(1)).$$

Hence, (4.5.44) and (4.5.45) complete the proof of Lemma 4.37. □

With Lemmas 4.36 and 4.37 in hand, the proof of the weak convergence in Theorem 4.35 is left to the reader as an exercise:

EXERCISE 4.31 (Convergence of the heavy-tailed exploration process). *Prove that Theorem 4.35 follows from Lemmas 4.36 and 4.37.*

Having concluded the proof of Theorem 4.35, the proof of the scaling of the largest critical components in Theorem 4.33 follows. □

4.6. Metric space convergence of critical percolation on configuration models

In this section, we study the *metric convergence* of the large critical clusters, viewed as metric spaces. Naturally, the notion of convergence of such metric spaces is rather tricky, and requires careful attention. We thus start by discussing convergence of metric spaces.

Convergence of metric spaces. Let us define the topology used in the statement of the metric space convergence, which is convergence in the Gromov-Hausdorff-Prokhorov topology. We refer to Abraham, Delmas and Hoscheit [1], Burago, Burago and Ivanov [71], Greven, Pfaffelhuber and Winter [140] and Gromov [147] for more information. Let us start by describing the Gromov-Hausdorff-Prokhorov topology.

In this section, all metric spaces under consideration will be compact metric spaces with associated probability measures. Let us first recall the Gromov-Hausdorff distance d_{GH} between metric spaces. Fix two metric spaces (X_1, d_1) and (X_2, d_2) . For a subset $C \subseteq X_1 \times X_2$, the *distortion* of C is defined as

$$(4.6.1) \quad \text{dis}(C) := \sup \{ |d_1(x_1, y_1) - d_2(x_2, y_2)| : (x_1, x_2), (y_1, y_2) \in C \}.$$

A *correspondence* C between X_1 and X_2 is a measurable subset of $X_1 \times X_2$ such that for every $x_1 \in X_1$ there exists at least one $x_2 \in X_2$ such that $(x_1, x_2) \in C$ and vice-versa. The *Gromov-Hausdorff distance* between the two metric spaces (X_1, d_1) and (X_2, d_2) is defined as

$$(4.6.2) \quad d_{\text{GH}}(X_1, X_2) = \frac{1}{2} \inf \{ \text{dis}(C) : C \text{ is a correspondence between } X_1 \text{ and } X_2 \}.$$

Suppose (X_1, d_1) and (X_2, d_2) are two metric spaces and $p_1 \in X_1$, and $p_2 \in X_2$. Then the *pointed Gromov-Hausdorff distance* between $\mathbf{X}_1 := (X_1, d_1, p_1)$ and $\mathbf{X}_2 := (X_2, d_2, p_2)$ is given by

$$(4.6.3) \quad d_{\text{GH}}^{\text{point}}(\mathbf{X}_1, \mathbf{X}_2) = \frac{1}{2} \inf \{ \text{dis}(C) : C \text{ is correspondence between } X_1 \text{ and } X_2 \text{ and } (p_1, p_2) \in C \}.$$

We will need a metric that also keeps track of associated measures on the corresponding spaces. A compact measured metric space (X, d, μ) is a compact metric space (X, d) with an associated probability measure μ on the Borel sigma algebra $\mathcal{B}(X)$. Given two compact measured metric spaces (X_1, d_1, μ_1) and (X_2, d_2, μ_2) and a measure π on the product space $X_1 \times X_2$, the *discrepancy* of π with respect to μ_1 and μ_2 is defined as

$$(4.6.4) \quad D(\pi; \mu_1, \mu_2) := \|\mu_1 - \pi_1\| + \|\mu_2 - \pi_2\|,$$

where π_1, π_2 are the marginals of π and $\|\cdot\|$ denotes the total variation distance between probability measures. Then the *Gromov-Hausdorff-Prokhorov distance* between X_1 and X_2 is defined as

$$(4.6.5) \quad d_{\text{GHP}}(X_1, X_2) := \inf \left\{ \max \left(\frac{1}{2} \text{dis}(C), D(\pi; \mu_1, \mu_2), \pi(C^c) \right) \right\},$$

where the infimum is taken over all correspondences C and measures π on $X_1 \times X_2$.

Similar to (4.6.3), we can define a “*pointed Gromov-Hausdorff-Prokhorov distance*”, $d_{\text{GHP}}^{\text{point}}$ between two metric measure spaces X_1 and X_2 having two distinguished points p_1 and p_2 respectively by taking the infimum in (4.6.5) over all correspondences C and measures π on $X_1 \times X_2$ such that $(p_1, p_2) \in C$.

Write \mathcal{S} for the collection of all measured metric spaces (X, d, μ) . The function d_{GHP} is a pseudometric on \mathcal{S} , and defines an equivalence relation $X \sim Y \Leftrightarrow d_{\text{GHP}}(X, Y) = 0$ on \mathcal{S} . Let $\bar{\mathcal{S}} := \mathcal{S}/\sim$ be the space of isometry equivalent classes of measured compact metric spaces and \bar{d}_{GHP} the induced metric. Then by Abraham, Delmas and Hoscheit [1], $(\bar{\mathcal{S}}, \bar{d}_{\text{GHP}})$ is a complete separable metric space. To ease notation, we will continue to use

$(\mathcal{S}, d_{\text{GHP}})$ instead of $(\bar{\mathcal{S}}, \bar{d}_{\text{GHP}})$ and $X = (X, d, \mu)$ to denote both the metric space and the corresponding equivalence class.

We now state the main result from Bhamidi and Sen [46]:

THEOREM 4.38 (Metric scaling limits with finite third-moment degrees). *Fix $\lambda \in \mathbb{R}$ and consider the critical configuration model $\text{CM}_n(\mathbf{d})$. Assume that $\mathbb{E}[D_n^3] \rightarrow \mathbb{E}[D^3] < \infty$. Then, there exists an appropriate limiting sequence of random compact metric measure spaces $\mathbf{M}_\infty(\lambda) := (M_i(\lambda))_{i \geq 1}$ such that the components in the critical regime satisfy*

$$(4.6.6) \quad \frac{1}{n^{1/3}} \mathbf{M}_n(\lambda) \xrightarrow{d} \mathbf{M}_\infty(\lambda), \quad \text{as } n \rightarrow \infty.$$

Here convergence is with respect to the product topology on $\mathcal{S}^{\mathbb{N}}$ induced by the Gromov-Hausdorff-Prokhorov metric on each coordinate \mathcal{S} . For each $i \geq 1$, the limiting metric spaces have the following properties:

- (a) $M_i(\lambda)$ is random compact metric measure space obtained by taking a random **real tree** $\mathcal{T}_i(\lambda)$ and identifying a random (finite) number of pairs of points (thus creating shortcuts);
- (b) The Hausdorff dimension equals 2 almost surely, i.e., $\dim_h(M_i(\lambda)) = 2$ a.s.

Let us discuss some consequences of Theorem 4.38. First, since the edges are rescaled by a factor $n^{-1/3}$, the typical distances between vertices in large critical clusters is of the order $n^{1/3}$. Theorem 4.38 implies that the matrix of distances between a collection of k uniformly chosen vertices converges to that in the limiting metric space. Furthermore, since the limiting metric space is *compact*, also the maximal distance given by the diameter of $\mathbf{M}_n(\lambda)$ converges in distribution to the diameter of the limiting metric space $\mathbf{M}_\infty(\lambda)$. This shows that typical distances are of order $n^{1/3}$ while typical cluster sizes are of size $n^{2/3}$ as observed previously. The fact that the Hausdorff dimension equals 2 can also be interpreted this way. Within distance of order $n^{1/3}$, one typically finds of the order $n^{2/3} = (n^{1/3})^2$ vertices suggesting that the Hausdorff dimension indeed equals 2. This fact was first proved for the critical Erdős-Rényi random graph by Addario-Berry, Broutin and Goldschmidt [4, 5], who also constructed the limiting object in Theorem 4.38.

Here is some discussion of the history of this problem. Bhamidi, Broutin, Sen and Wang [35] obtained the result under the assumption that there exists $a > 0$ such that $\sum_{n \geq 1} \mathbb{E}[e^{aD_n}] < \infty$. With more work, Bhamidi, Broutin, Sen and Wang [35] can improve the condition to a finite moment of some kind. Bhamidi and Sen [46] then improved it to the finite third-moment condition that is necessary for the scaling of component sizes and surpluses in Theorem 4.13, which is optimal. Interestingly, Bhamidi and Sen [46] also investigated this for other types of random graphs, including the vacant set left by a random walker. See Section 4.8.4 for a discussion of this setting.

We next extend the above setting to the case of infinite-third moment degrees. We combine the papers with Bhamidi, Dhara and Sen [38, 37], see also the work with Bhamidi and Sen [45]:

THEOREM 4.39 (Scaling limits with degree exponent $\tau \in (3, 4)$). *Fix $\lambda \in \mathbb{R}$ and consider the critical configuration model $\text{CM}_n(\mathbf{d})$. Assume that the degree distribution satisfies that D_n^* is stochastically dominated by a random variable Y with $\mathbb{E}[Y] = 1$.*

Then, there exists an appropriate limiting sequence of random compact metric measure spaces $\mathbf{M}_\infty(\lambda) := (M_i(\lambda))_{i \geq 1}$ such that the components in the critical regime satisfy

$$(4.6.7) \quad \frac{1}{n^{(\tau-3)/(\tau-1)}} \mathbf{M}_n(\lambda) \xrightarrow{d} \mathbf{M}_\infty(\lambda), \quad \text{as } n \rightarrow \infty.$$

Here convergence is with respect to the product topology on $\mathcal{S}^{\mathbb{N}}$ induced by the Gromov-Hausdorff-Prokhorov metric on each coordinate \mathcal{S} . For each $i \geq 1$, the limiting metric spaces have the following properties:

- (a) $M_i(\lambda)$ is random compact metric measure space obtained by taking a random real tree $\mathcal{T}_i(\lambda)$ and identifying a random (finite) number of pairs of points (thus creating shortcuts).
- (b) Call a point $u \in \mathcal{T}_i(\lambda)$ a **hub** point if deleting the u results in infinitely many disconnected components of $\mathcal{T}_i(\lambda)$. Then $\mathcal{T}_i(\lambda)$ has infinitely many hub points which are everywhere dense on the tree $\mathcal{T}_i(\lambda)$.
- (c) The box-counting or Minkowski dimension of $M_i(\lambda)$ satisfies

$$(4.6.8) \quad \dim(M_i(\lambda)) = \frac{\tau - 2}{\tau - 3} \quad \text{a.s.}$$

Consequently, the Hausdorff dimension satisfies the bound $\dim_h(M_i(\lambda)) \leq (\tau - 2)/(\tau - 3)$ a.s.

The metric space limit of the critical $\text{CM}_n(\mathbf{d})$ in the heavy-tailed setting was first derived in work with Bhamidi and Sen [45] for rank-1 inhomogeneous random graphs, where also the Minkowski dimension was determined. In [38], we determined convergence in the Gromov weak sense under the weakest possible conditions as also formulated in Theorem 4.33. This implies, for example, that the matrix of typical distances between a uniformly chosen collection of vertices converges to that in the limiting metric space. This was extended to convergence in the Gromov-Hausdorff-Prokhorov sense in [38] under the stronger condition on the degrees that D_n^* is stochastically dominated by a random variable Y with $\mathbb{E}[Y] = 1$. This property is crucially used to bound the probability that the critical cluster of a uniformly chosen vertex survives for a long time. Under the stochastic domination condition, we can relate this to the survival probability of a critical branching process.

The value $(\tau - 2)/(\tau - 3)$ for the Minkowski dimension can again intuitively be understood by a counting argument. Typical distances grow like $n^{(\tau-3)/(\tau-1)}$, while cluster sizes like $n^{(\tau-2)/(\tau-1)} = (n^{(\tau-3)/(\tau-1)})^{(\tau-2)/(\tau-3)}$. This suggests that we need $\varepsilon^{-(\tau-2)/(\tau-3)}$ macroscopic boxes of radius ε to cover a large critical cluster.

Again, we believe that the extra degree condition in Theorem 4.39 is superfluous:

OPEN PROBLEM 4.2 (Universal metric convergence for configuration models with finite third-moment degrees). *Extend Theorem 4.39 to the setting of Theorem 4.33 where (4.5.11a)–(4.5.11b) hold.*

4.7. Percolation on the PAM with conditionally independent edges

In this section, we describe some results for percolation on the preferential attachment model with conditionally independent edges as derived by Dereich and Mörters [100] and Eckhoff and Mörters [120]. Recall the definition of the preferential attachment model with conditionally independent edges in Section 1.3.6. Dereich and Mörters [100] study the existence of a phase transition for bond percolation, proving that when the power-law exponent $\tau \in (2, 3)$, the giant component remains for all bond percolation parameters $\pi \in (0, 1]$, while for $\tau > 3$, there is a critical value $\pi_c \in (0, 1)$ such that the giant component remains when $\pi > \pi_c$, while it is removed when $\pi \leq \pi_c$. These results are very alike the results in Section 4.3 (see in particular Theorem 4.5), and are discussed in some more detail below.

Eckhoff and Mörters [120] investigate further properties of the robust setting for $\tau \in (2, 3)$. Indeed, fix $\varepsilon > 0$, and remove the oldest εn vertices. This is a much more damaging operation than merely removed a proportion ε of the vertices, since the old vertices are the vertices with the highest possible degrees. Then, they show that site percolation on the resulting graph *does* have a phase transition at some critical value $\pi_c(\varepsilon)$, satisfying that $\pi_c(\varepsilon) \asymp 1/\log(1/\varepsilon)$.

For the configuration model, the corresponding statement is considerably simpler to prove, and is left as an exercise:

EXERCISE 4.32 (Removing vertices of high degrees in configuration models). *Consider $\text{CM}_n(\mathbf{d})$ with power-law degrees with exponent $\tau \in (2, 3)$. Let $\text{CM}_n^\varepsilon(\mathbf{d})$ be the graph that results from removing the εn vertices of highest degrees. Show that $\text{CM}_n^\varepsilon(\mathbf{d})$ has a critical percolation parameter $\pi_c(\varepsilon)$. What is the asymptotics of $\pi_c(\varepsilon)$ as $\varepsilon \searrow 0$?*

Percolation on preferential attachment models with conditionally independent edges. The following theorem describes the percolation critical value in preferential attachment models with conditionally independent edges for affine preferential attachment functions:

THEOREM 4.40 (Percolation on preferential attachment models with conditionally independent edges). *Suppose $f(k) = \gamma k + \beta$. Then percolation on the preferential attachment model with conditionally independent edges has critical value*

$$(4.7.1) \quad \pi_c = \left(\frac{1}{2\gamma} - 1\right) \left(\sqrt{1 + \frac{\gamma}{\beta}} - 1\right).$$

We see that $\pi_c = 0$ precisely when $\gamma \geq \frac{1}{2}$, in which case the degree distribution has exponent in $(2, 3)$. Recall from (2.3.18) that a preferential attachment function f is called *concave* $f(0) \leq 1$ and $\Delta f(k) := f(k + 1) - f(k) < 1$ for all $k \geq 0$. Recall that we call a giant component *robust* when it has $\pi_c = 0$. We next study the robustness of the general setting in some more detail:

THEOREM 4.41 (Percolation on preferential attachment models with conditionally independent edges). *Suppose f is an arbitrary concave attachment rule and recall the definition of the parameter γ from (2.3.19). Then the giant component in the preferential attachment network with attachment rule f is robust if and only if $\gamma \geq \frac{1}{2}$.*

Let \mathcal{I} be the set of parameters α for which A_α is a well-defined (and therefore also compact) linear operator. Then \mathcal{I} is a (possibly empty) subinterval of $(0, 1)$. Dereich and Mörters [100] identify the percolation critical value as

$$(4.7.2) \quad \pi_c = 1 / (\min_{\alpha \in \mathcal{I}} \rho(A_\alpha)).$$

This result is highly similar to $\pi_c = 1/\nu$ for the configuration model (see Theorem 4.5).

Of course, the results of Dereich and Mörters [100] and Eckhof and Mörters [120] do not apply to the preferential attachment model with a *fixed* number of edges. However, the main tool in [100, 120] involves the study of the branching process approximation to the neighborhoods of uniform vertices. Of course, for $(\text{PA}_t^{(m,\delta)})_{t \geq 1}$, progress has been made in that direction, recall Theorem 2.16. This only extends to settings where $\tau \geq 0$, unfortunately, recall Open Problem 2.2. The following open problem is about an extension to that setting:

OPEN PROBLEM 4.3 (Critical percolation threshold on preferential attachment models). *Investigate the percolation phase transition on the preferential attachment model $(\text{PA}_t^{(m,\delta)})_{t \geq 1}$. Are the results identical to the ones for the preferential attachment model with conditionally independent edges? Is the critical value π_c that of the multitype branching process appearing as the local-weak limit of in Theorem 2.16? Is $\pi_c = 0$ for $\delta \in (-m, 0)$, while $\pi_c > 0$ for $\delta > 0$?*

It can be expected that many of the technical tools from [100, 120] can be extended to this setting. In particular, we conjecture that $\pi_c = 0$ for $(\text{PA}_t^{(m,\delta)})_{t \geq 1}$ with $\delta \in (-m, 0)$, for which the degree power-law exponent $\tau = 3 + \delta/m$ satisfies $\tau \in (2, 3)$. Eckhof and Mörters [120] prove this statement for preferential attachment models with conditionally independent edges. However, dealing with a fixed number of edges is considerably harder.

For preferential attachment models, the critical regime is poorly understood. This leads us to the following open problem:

OPEN PROBLEM 4.4 (Critical behavior percolation on preferential attachment models). *Investigate the percolation phase transition on the preferential attachment model $(\text{PA}_t^{(m,\delta)})_{t \geq 1}$ with $\delta > 0$, so that $\tau = 3 + \delta/m$ satisfies $\tau > 3$. What is the scaling of the largest critical cluster sizes, and what is the width of the scaling window.*

We believe that percolation on the preferential attachment model behaves rather differently from that on other random graphs such as the configuration model. This is due to the more pronounced hierarchy present in the network, which is due to the age of the vertices present. This difference is clearly visible in how shortest paths are being formed, particularly comparing the configuration model with $\tau \in (2, 3)$, and the preferential attachment model with $\delta \in (-m, 0)$, for which $\tau = 3 + \delta/m \in (2, 3)$. Since the local neighborhood of $(\text{PA}_t^{(m,\delta)})_{t \geq 1}$ with $\delta > 0$ are described in terms of (multitype) branching

processes, it is a good idea to start describing the percolation phase transition in more detail there. See also Section 4.8.2, where such a problem is stated in the context of general inhomogeneous random graphs. It would be of great interest to investigate percolation on such preferential attachment models in more detail.

4.8. Further results on percolation on random graphs

In this section, we discuss some further results on percolation on random graphs. We study percolation on rank-1 inhomogeneous random graphs in Section 4.8.1. We broaden the discussion to general inhomogeneous random graphs in Section 4.8.2. In Section 4.8.3, we further investigate a special example of such inhomogeneous random graphs, uniformly grown random graph, which displays rather different scaling behavior of the giant component close to the critical value compared to, say, configuration models. We continue with the phase transition on high-dimensional transitive graphs in Section 4.8.5. We close in Section 4.8.6 by discussing the relationship between critical percolation on a finite graph and its minimal spanning tree.

4.8.1. Scaling limits of cluster sizes in rank-1 inhomogeneous random graph.

Scaling limits for critical percolation on random graphs have been obtained in many settings, in particular when the graphs themselves are quite *inhomogeneous*. Recall the definition of inhomogeneous random graphs in Section 1.3.2. The nice thing about inhomogeneous random graphs is that percolation on them gives rise to another inhomogeneous random graph. Thus, it suffices to study the critical behavior of inhomogeneous random graphs themselves. This is similar to the situation for the configuration due to Janson's construction in Lemma 4.6, but here it is even simpler. Let us illustrate this by discussing percolation on rank-1 inhomogeneous random graphs, where the edge probabilities are moderated by vertex weights $(w_i)_{i \in [n]}$. For example, for the Chung-Lu model or random graph with prescribed expected degrees,

$$(4.8.1) \quad p_{ij} = \min \left\{ \frac{w_i w_j}{\sum_{k \in [n]} w_k}, 1 \right\} = \frac{w_i w_j}{\sum_{k \in [n]} w_k},$$

when we assume for simplicity that $\max_{i \in [n]} w_i^2 \leq \sum_{k \in [n]} w_k \equiv \ell_n$. When applying percolation with percolation probability π on this model, we obtain that the edges are still independent, but now with edge probabilities πp_{ij} instead. However, obviously, this is the same as the Chung-Lu model with vertex weights $(\pi w_i)_{i \in [n]}$ instead:

EXERCISE 4.33 (Percolation on Chung-Lu model). *Consider $\text{CL}_n(\mathbf{w})$ with $\max_{i \in [n]} w_i^2 \leq \sum_{k \in [n]} w_k \equiv \ell_n$. Verify that percolation with percolation π gives rise to a Chung-Lu model with vertex weights $(\pi w_i)_{i \in [n]}$.*

Historically, the critical behavior of rank-1 inhomogeneous random graphs was the first for inhomogeneous models to be considered in the mathematics literature (see [159, 43, 44]). Indeed, there it is shown that most of the results proved for the configuration model also hold for such rank-1 inhomogeneous random graph models under appropriate conditions on the weight sequence (such as formulated in the Regularity Condition 1.1). However, having the results for the configuration model in hand, it is simpler to use those to prove this. Indeed, by Remark 4.12, the configuration model results also hold when conditioning on simplicity. By Theorem 1.4, the generalized random graph conditioned on its degrees is a uniform random graph with those degrees, while by Theorem 1.5, the

degrees in the generalized random graph satisfy the necessary conditions for e.g., Theorem 4.9 in probability. Here, we be careful with the convergence of the third moment of the degrees $\mathbb{E}[D_n^3]$ that is assumed in Theorem 4.9, but it is not too hard to show that when $\mathbb{E}[W_n^3] \rightarrow \mathbb{E}[W^3]$, then also $\mathbb{E}[D_n^3] \rightarrow \mathbb{E}[D^3]$, with $\mathbb{P}(D = k) = \mathbb{E}[e^{-W} W^k / k!]$. We refrain from a further discussion of this topic.

4.8.2. Scaling limit of cluster sizes in general inhomogeneous random graph.

The setting of percolation on general inhomogeneous random graphs is to a large extent open. Indeed, it is known what the condition is the existence of a giant component (recall Theorem 2.20), but this condition is quite functional analytic and not always easy to work with. Indeed, with T the operator describing the expected offspring operator of the kernel κ , we know that there exists a giant component when $\|T_\kappa\| > 1$, while there is none when $\|T_\kappa\| \leq 1$. However, explicitly computing $\|T\|$ is in general hard. The notable exception is the case of rank-1 IRGs, where we can explicitly compute the norm of the expected offspring distribution as $\nu_n = \mathbb{E}[W_n^2] / \mathbb{E}[W_n]$. For general inhomogeneous random graphs, this is much harder. This leads us to the following open problem:

OPEN PROBLEM 4.5 (Critical behavior general IRGs). *What is the critical behavior of general IRGs, and how is this related to the properties of the kernel κ ?*

We do know that inhomogeneous random graphs with a kernel that has a *finite* number of types is in the same universality as the Erdős-Rényi random graph, in the sense that the scaling limit of critical clusters is as in Theorem 4.1. This is proved by Bhamidi, Broutin, Sen and Wang [35]. In general, one would expect this result to also be true when the kernel κ is bounded. When there are infinitely many types and κ is unbounded, this question is much harder. Let us give a simple example. Take, for some $\alpha > 0$,

$$(4.8.2) \quad \kappa(x, y) = \frac{c}{(x + y)^\alpha}, \quad x, y \in [0, 1].$$

Then, for $\alpha < 1$, the kernel κ is L^2 , so that it has a bounded norm. It is even unclear what the precise conditions on κ are to make T_κ have a bounded norm. It is less difficult to investigate the degree distribution of the corresponding inhomogeneous random graph:

EXERCISE 4.34 (Percolation on Chung-Lu model). *Consider the inhomogeneous random graph with $\kappa: [0, 1] \times [0, 1] \mapsto (0, \infty)$ given by (4.8.2). Let D_n be the degree of a uniform vertex. Show that*

$$(4.8.3) \quad D_n \xrightarrow{d} D,$$

where the distribution of D is given by

$$(4.8.4) \quad \mathbb{P}(D = k) = \mathbb{E}\left[\mathbb{P}(\text{Poi}(\lambda(U)) = k)\right], \quad \text{where} \quad \lambda(u) = \int_0^1 \frac{c}{(u + y)^\alpha} dy,$$

and U has a uniform distribution on $[0, 1]$.

It would be of great interest to investigate whether the scaling limits of critical clusters in the example (4.8.2) show similar behavior as for the generalized random graph (or, equivalently, the configuration model). In particular, is the scaling limit the same as

for Erdős-Rényi random graph critical behavior fails? Or is it again the same as for configuration model with $\tau \in (3, 4)$ as in Theorem 4.32 and 4.33.

Bollobás, Janson and Riordan [56] do discuss this issue to some extent. Indeed, they discuss the related question when the size of the giant component, is, close to the critical point, *linear*. Let $\|T_\kappa\| = 1 + \varepsilon$, where $\varepsilon > 0$ is small. Then, the size of the giant component $\zeta(\varepsilon)$ satisfies $\zeta(\varepsilon) > 0$. Bollobás, Janson and Riordan [56] conjecture that $\zeta(\varepsilon) \sim c\varepsilon$ precisely when the unique positive eigenfunction ϕ_κ of T_κ is in L^3 . We conjecture that this should also be the condition for Erdős-Rényi random graph critical behavior:

OPEN PROBLEM 4.6 (Critical behavior general inhomogeneous random graphs). *Show that inhomogeneous random graph have Erdős-Rényi random graph critical behavior precisely when unique positive eigenfunction ϕ_κ of T_κ is in L^3 , i.e., when $\int_0^1 \phi_\kappa^3(u) du < \infty$. In the case when $\int_0^1 \phi_\kappa^3(u) du = \infty$, while $\int_0^1 \phi_\kappa^2(u) du < \infty$, define*

$$(4.8.5) \quad \tau = 1 + \sup \left\{ p > 2 : \int_0^1 \phi_\kappa^p(u) du < \infty \right\}.$$

Does τ for general inhomogeneous random graphs play the same role as it does for the configuration model?

We are quite far from understanding the general critical behavior of inhomogeneous random graphs. It would be good to analyse some examples, away from the rank-1 setting, where explicit computations can be done. We investigate one of such examples, where rather different behavior compared to the configuration model is observed, in the following section.

4.8.3. Phase transition on uniformly grown random graph. The uniformly grown random graph was proposed by Callaway, Hopcroft, Kleinberg, Newman, and Strogatz [72]. 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 [116] discusses this model.

We will discuss a variant of this model proposed and analyzed by Bollobás, Janson and Riordan [55]. 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

$$(4.8.6) \quad p_{ij} = \frac{c}{\max\{i, j\}},$$

all edge statuses being independent random variables. Equivalently, we can view this random graph as arising dynamically, where vertex t connects to a vertex $s < t$ with probability $1/t$ independently for all $s \in [t - 1]$. Kalikow and Weiss [190] show that the

critical value of this model equals $c = \frac{1}{4}$. Riordan [233], building on the work by Bollobás, Janson and Riordan [55], shows that, for $c = \frac{1}{4} + \varepsilon$,

$$(4.8.7) \quad \frac{|\mathcal{C}_{\max}|}{n} \xrightarrow{\mathbb{P}} \rho(\varepsilon),$$

where

$$(4.8.8) \quad \rho(\varepsilon) = \exp\left(-\frac{\pi}{2\sqrt{\varepsilon}} + O(\log \varepsilon)\right).$$

See also [56, Sections 16.1 and 16.3] and the references therein for more details on both these models.

Bollobás, Janson and Riordan [55] further show that for any $c < \frac{1}{4}$, the expected size of the component of vertex 1 containing is $O(n^{\beta(c)})$, where $\beta(c) = \frac{1}{2} - \sqrt{\frac{1}{4} - c}$, while for any $c < \frac{1}{4}$, whp the component of containing the vertex 1 has size at least $n^{c-\varepsilon}$ for any $\varepsilon > 0$. Of course, the size of the largest critical cluster is highly non obvious:

OPEN PROBLEM 4.7 (Critical behavior uniformly grown random graph). *Is there a choice of $c \approx \frac{1}{4}$ for which the uniformly grown random graph displays critical behavior? If so, then what is the critical window?*

4.8.4. Scaling limits of vacant set left by random walker. Random walk percolation was introduced by Dembo and Sznitman [95] as a model for how termites eat through a beam of wood. Consider a graph G , and start a random walk on it. Run it for some amount of time $u|V(G)|$. Remove all vertices that have been visited by the random walker. Thus, we are left with the *vacant set*, i.e., the set of vertices that have escaped the random walker so far. The question is how large u should be for the vacant set to become highly disconnected. See the recent survey by Černý and Teixeira [76]. We will focus on the setting where the underlying graph is a random graph, more precisely, a configuration model. Cooper and Frieze [87] show that the critical value on a random d -regular graph is equal to

$$(4.8.9) \quad u^* = \frac{d(d-1)\log(d-1)}{(d-2)^2}.$$

Compare this to the critical value for bond percolation, which is $1/(d-1)$. Their proof makes crucial use of a clever construction that states that after the random walk has walked for time t , and conditionally on the degrees of the vacant set at that time, the resulting graph is again uniform on the collection of all graphs with that degree sequence. That is a lot like Janson's Construction (Lemma 4.6) for bond percolation on the configuration model. Černý, Teixeira and Windisch [78] prove sharper bounds on the size of the giant component of the vacant set at time un for supercritical u satisfying $u < u^*$ as well as subcritical u satisfying $u > u^*$. Beware that the vacant set *decreases* when u increases, which is a little confusing when used to the increasing nature in p of the connected components in regular percolation. Černý and Teixeira [77] take this a step further, by identifying the critical window as corresponding to u such that $n^{1/3}|u - u^*|$ is bounded, and showing that $n^{-2/3}|\mathcal{C}_{\max}|$ and $n^{2/3}/|\mathcal{C}_{\max}|$ are tight random variables for

such u . One of the key results for this, is the following nice description of the degree distribution of the vacant set. Let

$$(4.8.10) \quad p_{\text{vac}} = e^{-\frac{d \log(d-1)}{d-2}},$$

and let the limiting degree sequence be given by $D_{\text{vac}} = 0$ with probability $1 - p_{\text{vac}}$, and $D_{\text{vac}} = \text{Bin}(d, 1/(d-1))$ with probability p_{vac} . Thus, p_{vac} is the proportion of vertices that are in the vacant set. When a vertex is not in the vacant set, then its remaining degree within the vacant set equals 0, while otherwise any of its neighbors is visited by the random walk independently with probability $1/(d-1)$.

Bhamidi and Sen [46] extend this to metric space convergence as in Theorem 4.38, but again only for the random-regular case (as well as its regular configuration model equivalent). It is yet unclear what the critical value is for the configuration model with arbitrary degrees:

OPEN PROBLEM 4.8 (Critical value random-walk percolation). *Identify the critical value u^* for the vacant set left by random walker for general configuration models satisfying Condition 1.6(a)-(b). Is this critical value $u^* = 0$ when $\mathbb{E}[D_n^2] \rightarrow \infty$?*

An even more daunting question is to derive the metric scaling limit of the vacant set within the scaling window, as also conjectured by Bhamidi and Sen [46, Conjecture 2.2]:

OPEN PROBLEM 4.9 (Metric scaling limit for critical vacant set left by random walker). *Extend the metric space limits in Theorems 4.38–4.39 to the vacant set left by a random walker for general configuration models satisfying Condition 1.6(a)-(c). Is there a regime where Theorem 4.39 holds?*

The reason why it is not obvious that there a regime where Theorem 4.39 holds is that the stationary distribution of random walk is $\pi_v = d_v/\ell_n$. Thus, vertices with higher degrees will be seen more frequently. The mixing time of random walk on $\text{CM}_n(\mathbf{d})$ is proportional to $\log n$, so that it may be that after running random walk for time un , no vertices of very high degree are left, and the degree distribution has finite third moment. In this case, we can only expect that the scaling limit in Theorem 4.38 arises for appropriate values of u . Bhamidi and Sen [46, Conjecture 2.2(c)] do believe that such a setting arises, and that it arises precisely when the original degree distribution obeys a power-law with exponent τ satisfying $\tau \in (3, 4)$.

4.8.5. Scaling limits for percolation on critical high-dimensional tori. There are many results for percolation on critical high-dimensional tori. See [157] for the state of the art. We do not describe the full range of results here, as this is done in quite some detail in [157]. Let us summarize that there are many situations where it is known that within the critical scaling window, clusters have size that is comparable to the number of vertices raised to the power $2/3$, just as for the Erdős-Rényi random graph. See [60, 61, 62, 155, 156, 173, 172] for some examples, including high-dimensional tori of fixed

dimension larger than 6 as well as the hypercube in the limit where the dimension tends to infinity. For the hypercube, recent works with Nachmias have shown that there is a well defined phase transition, and various properties (including the size of the barely supercritical giant component) have been derived [173, 172].

As we have seen in the above description, scaling limits of critical random graphs are usually obtained by applying weak convergence results, such as Martingale Functional Central Limit Theorems, to the exploration processes that describe the exploration of such graphs. Here, we are helped by the fact that the models are sufficiently *mean field*, meaning that vertices interact with one another in roughly the same way. For example, for the Erdős-Rényi random graph, all vertices are the same, which makes that the depletion-of-points-effect comes out in a really simple form as the $-t^2/2$ term in (4.2.2) and as the $N_i = n - i - S_i - \inf_{j \in [i-1]} S_j \approx n - i$ in (4.2.7).

Unfortunately, on lattices, the depletion-of-points-effect is far less homogeneous. Some vertices that we explore still have many neighbors that are neutral, while others have far fewer than expected. On average, after a long time, most active vertices should have of the order 1 minus something small neighbors that become active after exploration. For example, on the hypercube, the scaling window is of width $2^{-d/3}/d$, so that, after of the order $2^{2d/3}$ explorations, we can expect vertices to have on average $[1 + O(2^{-d/3}/d)]/p_c$ neutral neighbors. This number is extremely close to d , and order 1 of them will be found to be neighbors of the vertex that is currently explored. One cannot expect to achieve this by the simple exploration process that was used for the Erdős-Rényi random graph. However, we *do* believe that the scaling limits are the same:

OPEN PROBLEM 4.10 (Scaling limit critical high-dimensional tori).
Consider the nearest-neighbor $\mathbb{T}_{r,d}$ with $d > 6$ and fix $p = p_c(\mathbb{Z}^d)(1 + \lambda V^{-1/3})$. Prove that

$$(4.8.11) \quad V^{-2/3}(|\mathcal{C}_{(i)}|)_{i \geq 1} \xrightarrow{d} (\chi_i(\pi_\lambda))_{i \geq 1} \quad \text{as } r \rightarrow \infty,$$

where $(\chi_i(\pi_\lambda))_{i \geq 1}$ are minor adaptations of the scaling limit on the Erdős-Rényi random graph in Theorem 4.1. Indeed, one would expect a translation of time and rescaling of χ_i as well as λ to be necessary to make $(\chi_i(\pi_\lambda))_{i \geq 1}$ have the same law as the scaling limit on the Erdős-Rényi random graph $(\gamma_i(\pi_\lambda))_{i \geq 1}$ in Theorem 4.1.

Open Problem 4.10 is way out of our reach currently. We believe that a rescaling in time in Open Problem 4.10 may be necessary, since we do not exactly know where $p_c(\mathbb{Z}^d)$ lies inside the scaling window. Possibly, the setting on the hypercube (which already is quite out of reach for the moment, see [173, 172]) is a more realistic starting for such an analysis than high-dimensional tori of fixed dimension. We refer to the recent text with Heydenreich [157] for an extensive discussion of high-dimensional percolation, including the hypercube, where also many open problems have been stated.

4.8.6. Minimal spanning trees. Imagine a finite and connected graph $G = (V, E)$, where each edge has an associated edge-weight. The *minimal spanning tree* of G is the spanning tree that minimizes the total weight. More precisely, given weights $(U_e)_{e \in E}$, the

MST is the spanning tree that minimizes the sum of the weights of all the edges in the tree $\sum_{e \in \mathcal{T}} U_e$. There are two algorithms that allow one to obtain the minimal spanning tree, implying that its computation (contrary to, for example the Steiner tree) is not computationally hard. The first is *Prim's algorithm* for the construction of the minimal spanning tree, which makes use of a dynamics that also goes under the name of *invasion percolation*. In this dynamics, we start from a given vertex, and successively add the minimal-weight edge that (a) is incident to the current vertex set; and (b) does not close a cycle. When started from the single vertex, the result of this invasion dynamics is a minimal spanning tree. The second algorithm is *Kruskal's algorithm*, which orders the edges according to their weight from small to large. We then process the edges one by one in the order of their edge-weight, accepting an edge when it does not close a cycle, and rejecting it otherwise. Again, the result is a minimal spanning tree.

The minimal spanning tree can be investigated on any finite graph and for any edge-weight distribution. In fact, whenever the edge-weight distribution does not have atoms, the distribution of the edges chosen in the minimal spanning tree is the same. Further, in this case, the minimal spanning tree is also *unique*, a property that can obviously fail when the edge weights do have atoms:

EXERCISE 4.35 (Non-uniqueness of minimal spanning tree). *Find a simple example where the minimal spanning tree is not unique.*

EXERCISE 4.36 (Uniqueness of minimal spanning tree for continuous edge-weights). *Show that the minimal spanning tree is unique when the weights are i.i.d. with a continuous distribution.*

Of course, the actual *weight* of the minimal spanning tree, highly depends on the edge-weight distribution. It turns out to be particularly convenient to assume that the edge weights $(U_e)_{e \in E}$ are i.i.d. exponential random variables with parameter 1, and we denote this setting by writing E_e instead of U_e for the edge weight of edge e . For the time being, we focus on the mean-field situation, where G is the complete graph.

Let us start by introducing some notation to formalize the notion of the minimal spanning tree. For a spanning tree \mathcal{T} , we let $W(\mathcal{T})$ denote its weight, i.e.,

$$(4.8.12) \quad W(\mathcal{T}) = \sum_{e \in \mathcal{T}} E_e.$$

Here we assume that $(E_e)_{e \in E}$ are i.i.d. exponential distributions with parameter 1. Then, the minimal spanning tree \mathbb{T}_n is the minimizer of $W(\mathcal{T})$ over all spanning trees. When the edge weights are all different, this minimizer is unique. Since we assume that the edge weight distribution E_e is *continuous*, the minimal spanning tree is a.s. unique (recall also Exercise 4.36).

We discuss two important results on the minimal spanning tree on the complete graph. We start with Frieze's result from [135] showing that the weight of the minimal spanning tree is asymptotically equal to $\zeta(3)$, where $\zeta(s)$ is the zeta-function given by $\zeta(s) = \sum_{n \geq 1} n^{-s}$:

THEOREM 4.42 (Weight of the minimal spanning tree [135]). *Let $(E_e)_{e \in E(K_n)}$ be i.i.d. exponential random variables with parameter 1 and let \mathbb{T}_n denote the minimal spanning*

tree on the complete graph K_n , i.e., the tree that minimizes $\sum_{e \in \mathcal{T}} E_e$ over all spanning trees of K_n . Then,

$$(4.8.13) \quad \lim_{n \rightarrow \infty} \mathbb{E}[W(\mathbb{T}_n)] = \zeta(3).$$

PROOF. We follow Addario-Berry in [2]. Rather than Prim's algorithm for computing the minimal spanning tree, we rely on *Kruskal's algorithm*. In Kruskal's algorithm, we grow the minimal spanning tree by going through the edges in increasing edge weight, adding each edge as long as it does not create a cycle. Thus, we grow a forest, and, after having added the n th edge, we have obtained a spanning tree. The fact that this is the MST is not hard, and a nice exercise:

EXERCISE 4.37 (Kruskal's algorithm). *Show that the spanning tree obtained in Kruskal's algorithm equals the minimal spanning tree.*

We write $N = \binom{n}{2}$ for the total number of edges in the complete graph, and order the exponential edge weights as $(E_{(m)})_{m \in [N]}$ with $E_{(1)}$ being the smallest weight, and call this the *weight sequence*. Further, let $e_{(m)}$ denote the edge that corresponds to the edge weight $E_{(m)}$. The resulting *edge sequence* $(e_{(m)})_{m \in [N]}$ is a sequence of uniform draws without replacement from $[n]$. Since weights have been assigned in an i.i.d. manner, the resulting edge sequence is *independent* from the weight sequence $(E_{(m)})_{m \in [N]}$. As a result, for any $M \leq N$, $(e_{(m)})_{m \in [M]}$ is a uniform draw from the edges in the complete graph, so that $(e_{(m)})_{m \in [M]}$ is the Erdős-Rényi random graph with a fixed number of edges. By Kruskal's algorithm,

$$(4.8.14) \quad W(\mathbb{T}_n) = \sum_{m \in [n]} E_{(m)} \mathbb{1}_{\{e_{(m)} \text{ does not create a cycle}\}}.$$

We can now take the expectation and use the independence between edge sequence and weight sequence to obtain

$$(4.8.15) \quad \mathbb{E}[W(\mathbb{T}_n)] = \sum_{m \in [n]} \mathbb{E}[E_{(m)}] \mathbb{P}(e_{(m)} \text{ does not create a cycle}).$$

The problem has decoupled, and we can investigate each of the terms separately. By the memoryless property of the exponential distribution,

$$(4.8.16) \quad \mathbb{E}[E_{(m)}] = \sum_{s=1}^m \frac{1}{N - s + 1}.$$

This was the easier part of the two. For the second term, let

$$(4.8.17) \quad \chi_n(\lambda) = \mathbb{E}_\lambda[|\mathcal{C}(1)|]$$

be the expected cluster size after adding $m = \lambda N / (n - 1) = \lambda n / 2$ edges, where we recall that $N = \binom{n}{2}$ denotes the total number of edges. This is the susceptibility of the Erdős-Rényi random graph when adding m edges, which we call the *combinatorial* Erdős-Rényi random graph model (in contrast to the earlier introduced *binomial* Erdős-Rényi random graph model, where edges are inserted with probability $p = \lambda/n$). The binomial model can be obtained from the combinatorial model by taking $m \sim \text{Bin}(N, \lambda/n)$. Morally, they should be the same, since a $\text{Bin}(N, \lambda/n)$ distribution is very close to its mean $\lambda(n - 1)/2 \approx \lambda n / 2 = m$ for n large.

It is not too hard to show that

$$(4.8.18) \quad \frac{\chi_n(\lambda)}{n} \rightarrow \theta(\lambda)^2,$$

where $\theta(\lambda)$ is the survival probability of a branching process with a Poisson offspring distribution with parameter λ . In particular, $\theta(\lambda) = 0$ when $\lambda \leq 1$, while, for $\lambda > 1$, $\theta(\lambda)$ is the largest solution to

$$(4.8.19) \quad \theta(\lambda) = 1 - e^{-\lambda\theta(\lambda)}.$$

The proof of (4.8.18) is left as an exercise:

EXERCISE 4.38 (Susceptibility of Erdős-Rényi random graph). *Consider the binomial Erdős-Rényi random graph with edge probability $p = \lambda/n$. Prove (4.8.18).*

Then, the important fact is that, with $\lambda = nm/N$ and using (4.8.17) for the combinatorial model,

$$(4.8.20) \quad \mathbb{P}(e_{(m+1)} \text{ does not create a cycle}) = 1 - \frac{\chi_n(\lambda) - 1}{n - 1 - \lambda}.$$

To see (4.8.20), we condition on the cluster when we have added m edges. Let \mathcal{F}_m denote the σ -field describing the first m choices of the edges. Then,

$$(4.8.21) \quad \mathbb{P}(e_{(m+1)} \text{ creates a cycle} \mid \mathcal{F}_m) = \sum_{i \geq 1} \frac{|\mathcal{C}_{(i)}|(|\mathcal{C}_{(i)}| - 1)}{2(N - m)}.$$

Indeed, we have already chosen m edges, so there are $N - m$ edges left. Conditionally on the structure of the clusters formed up to the addition of the m th edge, we create a cycle precisely when we choose both endpoints of the next uniform edge inside the *same* connected component. There are precisely $|\mathcal{C}_{(i)}|(|\mathcal{C}_{(i)}| - 1)/2$ different ways to choose an edge in the i th largest cluster. We complete the proof of (4.8.20) by noting that

$$(4.8.22) \quad \sum_{i \geq 1} |\mathcal{C}_{(i)}|(|\mathcal{C}_{(i)}| - 1) = \sum_{v \in [n]} |\mathcal{C}(v)| - n,$$

so we arrive at

$$(4.8.23) \quad \mathbb{P}(e_{(m+1)} \text{ does not create a cycle}) = 1 - \sum_{v \in [n]} \frac{\mathbb{E}[|\mathcal{C}(v)|] - n}{n(n - 1) - 2m} = 1 - \frac{\chi_n(\lambda) - 1}{n - 1 - \lambda}.$$

Since the Erdős-Rényi random graph with $\lambda = (1 + \varepsilon) \log n$ is whp connected for any $\varepsilon > 0$ (see the original paper by Rényi [231] or [160, Section 5.3]), we can restrict to $\lambda = \Theta(\log n)$. Thus, we can approximate

$$(4.8.24) \quad \mathbb{P}(e_{(m+1)} \text{ does not create a cycle}) \approx 1 - \frac{\chi_n(\lambda)}{n} \approx 1 - \theta(\lambda)^2.$$

We see that all the action happens when $\lambda \approx 1$. Further, when $\lambda \leq 1$, almost *all* edges are accepted, since then $\theta(\lambda) = 0$, while as soon as $\lambda > 1$, a larger and larger positive proportion of the edges will be rejected. For λ fixed and $m = \lambda n/2$, we obtain that $\mathbb{E}[E_{(m)}] \approx m/N = 2m/(n(n - 1)) = \lambda/(n - 1) \approx \lambda/n$, so that, with the convention that $m = \lambda n/2$,

$$(4.8.25) \quad \mathbb{E}[W(\mathbb{T}_n)] \approx \sum_{m \in [n]} \frac{\lambda}{n} [1 - \theta(\lambda)^2] \approx \frac{1}{2} \int_0^\infty \lambda [1 - \theta(\lambda)^2] d\lambda.$$

We are left to show that the integral equals $2\zeta(3)$. This was first done by Aldous and Steele in [12], and is a cute exercise in calculus. It is remarkable that even though we do not have a nice description of $\lambda \mapsto \theta(\lambda)$, we can still compute the integral $\int_0^\infty \lambda[1 - \theta(\lambda)^2]d\lambda$.

For this, we first use partial integration to write

$$(4.8.26) \quad \int_0^\infty \lambda[1 - \theta(\lambda)^2]d\lambda = \int_0^\infty \lambda^2\theta'(\lambda)\theta(\lambda)d\lambda = \int_1^\infty \lambda^2\theta'(\lambda)\theta(\lambda)d\lambda,$$

since $\theta(\lambda) = 0$ for $\lambda \in [0, 1]$. Then, we use (4.8.19) to write

$$(4.8.27) \quad \lambda = -\frac{\log[1 - \theta(\lambda)]}{\theta(\lambda)},$$

to rewrite the integral as

$$(4.8.28) \quad \begin{aligned} \int_0^\infty \lambda[1 - \theta(\lambda)^2]d\lambda &= \int_1^\infty \frac{(\log[1 - \theta(\lambda)])^2}{\theta(\lambda)^2}\theta'(\lambda)\theta(\lambda)d\lambda = \int_1^\infty \frac{(\log[1 - \theta(\lambda)])^2}{\theta(\lambda)}\theta'(\lambda)d\lambda \\ &= \int_0^1 \frac{(\log[1 - \theta])^2}{\theta}d\theta, \end{aligned}$$

using that $\lambda \mapsto \theta(\lambda)$ is bijective on $[0, 1]$. Now using the final change of variables $u = \log(1 - \theta)$, for which $\theta = e^{-u}/(1 - e^{-u})$, we arrive at

$$(4.8.29) \quad \int_0^\infty \lambda[1 - \theta(\lambda)^2]d\lambda = \int_0^\infty u^2 \frac{e^{-u}}{1 - e^{-u}}du = \sum_{k=1}^\infty \int_0^\infty u^2 e^{-ku}du = \sum_{k=1}^\infty \frac{2}{k^3},$$

since $\int_0^\infty u^2 e^{-ku}du = 2/k^3$. This completes the proof. \square

We continue to discuss the scaling of the minimal spanning tree as a tree. It turns out that, when properly rescaled, the minimal spanning tree converges towards a scaling limit, which is a so-called *real tree*. A real tree is a continuum object that has many of the properties that we know for finite trees. As we have already seen above, there are deep relations between the minimal spanning tree and Erdős-Rényi random graphs. This connection is particularly tight for the scaling limit, as shown in a sequence of papers by Addario-Berry, Broutin and Goldschmidt [4, 5], the final piece and ‘pièce de résistance’ being jointly with Miermont [6]. In turn, the latter paper is a follow-up on earlier work of Addario-Berry, Broutin and Reed [8]. We focus on the results in [6]. We interpret a discrete tree \mathcal{T} along with the graph metric on \mathcal{T} as a metric space. Further, for a tree \mathcal{T} , we write $a\mathcal{T}$ for the tree where graph distances are rescaled by a factor a . In particular, this changes the unit length of an edge in \mathcal{T} to length a in $a\mathcal{T}$. Rescaled discrete trees are real trees themselves. The main result proved by Addario-Berry, Broutin, Goldschmidt and Miermont [6] is the following:

THEOREM 4.43 (Metric limit of the minimal spanning tree [6]). *Let \mathbb{T}_n denote the minimal spanning tree on the complete graph K_n . Then, the real tree $n^{-1/3}\mathbb{T}_n$ converges in distribution to some limiting real tree \mathbb{T} . In particular, this implies that the diameter of minimal spanning tree is of order $n^{1/3}$, i.e.,*

$$(4.8.30) \quad n^{-1/3}\text{diam}(\mathbb{T}_n) \xrightarrow{d} \text{diam}(\mathbb{T}).$$

The Minkowski dimension of $\text{diam}(\mathbb{T})$ equals 3.

The topology is crucial for weak convergence. The weak convergence in Theorem 4.43 is weak convergence in the *Gromov-Hausdorff-Prokhorov* topology as already discussed in Section 4.6. The proof of Theorem 4.43 makes strong use of the scaling limit of critical clusters as identified by Addario-Berry, Broutin and Goldschmidt in [4, 5]. Loosely speaking, the shape of the scaling limit of the minimal spanning tree is determined within the scaling window of the Erdős-Rényi random graph. This is remarkable, since these critical clusters have size $n^{2/3}$, while the minimal spanning tree has size n . This discrepancy appears prominently in the fact that the Minkowski dimension of the scaling limit of large critical clusters is 2, while the Minkowski dimension of the scaling limit of the minimal spanning tree is 3 [6]. The latter can again be understood by noting that the typical distances in the minimal spanning tree grow like $n^{1/3}$, while it contains $n = (n^{1/3})^3$ vertices, thus suggesting that its dimension is 3.

The proofs of Theorems 4.42 and 4.43 clearly highlight the relation between the minimal spanning tree and critical behavior of percolation on the underlying graph, in this case the complete graph K_n . We know, however, that critical percolation on $\text{CM}_n(\mathbf{d})$ has similar scaling limits (recall Theorem 4.9), which suggests that also the scaling limit of the minimal spanning tree are the same as that on the complete graph in Theorem 4.43. It would be of great interest to extend such results further, for example to configuration models with asymptotic degrees distributions D satisfying $\mathbb{P}(D \geq 3) = 1$, for which $\text{CM}_n(\mathbf{d})$ is whp connected:

OPEN PROBLEM 4.11 (Minimal spanning tree on configuration model). *Let the asymptotic degree distribution D satisfy $\mathbb{P}(D \geq 3) = 1$. Extend Theorem 4.43 to the $\text{CM}_n(\mathbf{d})$ under the conditions of Theorem 4.13.*

While Theorem 4.9 suggests a close connection between the minimal spanning tree on $\text{CM}_n(\mathbf{d})$, Theorems 4.33 and 4.39 instead suggest that for power-law degree sequences with $\tau \in (3, 4)$, the situation is rather different:

OPEN PROBLEM 4.12 (Minimal spanning tree on heavy-tailed configuration models). *Let the asymptotic degree distribution D satisfy $\mathbb{P}(D \geq 3) = 1$. Identify the scaling limit of the minimal spanning tree on $\text{CM}_n(\mathbf{d})$ under the conditions of Theorem 4.33.*

When $\tau \in (3, 4)$, the vertices of maximal degree play a crucial role. This makes it easier to locate the largest clusters. In fact, the proof by Addario-Berry, Broutin, Goldschmidt and Miermont [6] in the complete graph setting shows that the minimal spanning tree on K_n can be obtained by taking studying percolation at $\pi_n(\lambda) = (1 + \lambda n^{-1/3})/n$ for some $\lambda > 0$ large, and considering the largest connected component. This will contain a few cycles, that will be broken arbitrarily. After this, all the other vertices are attached to it by glueing them on in the form of small trees. In this argument, it is helpful to know that the largest connected component for λ large does not jump around, and basically only grows. This is called the *leader problem*, where the large components are competing

to be the largest. We would like to know that eventually one leader will win. Progress on that problem in the context of random graphs was made by Addario-Berry, Bhamidi and Sen in [3], where both the finite and infinite third-moment degrees were considered. In [3, Conjecture 3.2], Addario-Berry, Bhamidi and Sen further conjecture that the connected component of the maximal-degree vertex is the largest for all $\pi \in [0, 1]$ is strictly positive. We know that this occurs whp both above and below the scaling window, but in the scaling window it may not be true.

We can also make a conjecture about the Minkovski dimension of $\text{diam}(\mathbb{T})$ when $\tau \in (3, 4)$. Indeed, the graph distances grow like $n^{(\tau-3)/(\tau-1)}$, which suggests that Minkovski dimension of $\text{diam}(\mathbb{T})$ equals $(\tau-1)/(\tau-3)$. Notice that this dimension become arbitrarily large when τ approaches the boundary value $\tau = 3$, suggesting that the fractal nature of the minimal spanning tree becomes more and more involved.

4.9. Notes and discussion on percolation on random graphs

Notes on Section 4.2. There are many more results about percolation on the complete graph, for which we give part of the history. The critical window was first identified by Bollobás [49] (except for a logarithmic term, which was removed by Łuczak [208]). See also the impressive works by Janson, Knuth, Łuczak and Pittel [184], as well as that by Łuczak, Pittel and Wierman [210]. Janson and Spencer [187] give a point process description of the sizes and number of components of size $\varepsilon n^{2/3}$. Pittel [230] derives an explicit, yet involved, description for the distribution of the limit of $|\mathcal{C}_{\max}|n^{-2/3}$. The proof makes use of generating functions, and the relation between the largest connected component and the number of labeled graphs with a given surplus. Here, the complexity of a graph is its number of edges minus its number of vertices. Relations between Erdős-Rényi random graphs and the problem of counting the number of labeled graphs has received considerable attention in various works of, amongst others, Bollobás, Spencer and Wright, see e.g. [50, 174, 209, 243, 254, 255] and the references therein. Consequences of the result by Pittel [230] are for example that the probability that $|\mathcal{C}_{\max}|n^{-2/3}$ exceeds a for large a decays as $e^{-a^{3/8}}$ (in fact, the asymptotics are much finer than this!), and for very small $a > 0$, the probability that $|\mathcal{C}_{\max}|n^{-2/3}$ is smaller than a decays as $e^{-ca^{-3/2}}$ for some explicit constant $c > 0$.

Nachmias and Peres [221] use clever martingale techniques to prove extremely sharp bounds on the tails of the largest connected component within the critical window, which we summarize here:

THEOREM 4.44 (Cluster tail bounds in critical window). *Set $\pi_n(\lambda) = (1 + \lambda n^{-1/3})/n$ for some $\lambda \in \mathbb{R}$ and consider percolation on K_n with edge probability $\pi_n(\lambda)$. For $\lambda > 0$ and $A > 2\lambda + 3$, for large enough n ,*

$$(4.9.1) \quad \mathbb{P}(|\mathcal{C}_v(\pi_n(\lambda))| \geq An^{2/3}) \leq \left(\frac{4\theta}{1 - e^{-4\theta}} + 16 \right) n^{-1/3} e^{-((A-1)^2/2 - (A-1)\theta - 2)^2/(4A)},$$

and

$$(4.9.2) \quad \mathbb{P}(|\mathcal{C}_{\max}(\pi_n(\lambda))| \geq An^{2/3}) \leq \left(\frac{4\theta}{A(1 - e^{-4\theta})} + \frac{16}{A} \right) e^{-((A-1)^2/2 - (A-1)\theta - 2)^2/(4A)}.$$

Theorem 4.44 gives bounds that are remarkably close to the truth, as derived by Pittel [230]. Finally, with Kager and Müller [167] we derive an explicit local limit theorem for a finite number of the largest connected components inside the critical window.

We have already referred to the work by Addario-Berry, Broutin and Goldschmidt [5] on metric convergence of the Erdős-Rényi random graph. There is also related work on the metric structure in the barely supercritical case. The *two-core* of a connected component is the maximal part in which all vertices have degree at least two. This can be obtained iteratively by removing vertices of degree 1. Ding, Kim, Lubetzky and Peres [105, 106] show that the *kernel*, which is the two-core where all paths of vertices having degree precisely two are contracted to single edges, is close in law to a configuration model with a nice degree distribution. When $p = (1 + \varepsilon_n)/n$ with $n^{-1/3} \ll \varepsilon_n \ll n^{-1/4}$, this kernel is a random three-regular graph of size of order $\varepsilon_n^3 n$. The length of the contracted edges in the barely supercritical Erdős-Rényi random graph are close to i.i.d. exponential random variables with mean ε_n^{-2} , thus leading to first-passage percolation on the configuration model with exponential edge-weights, as studied in Section 3.3.

Notes on Section 4.3. There is substantial work on percolation on the configuration model. Fountoulakis [133] has an alternative construction to Lemma 4.6 for percolation on the configuration model, using the fact that the edges that are kept constitute a uniformly chosen sample. As explained, percolation on the configuration model is closely related to the configuration model itself, and there is substantial work on the giant component and critical value of the configuration model dating back to Molloy and Reed [215, 216]. Riordan [234] investigates the critical setting of the configuration model with bounded degrees, but also investigates the near-critical behavior. Barely supercritical results are also derived in work with Janson and Luczak [166]. When $\nu < 1$, Janson [180] shows that the largest connected component is equal to $d_{\max}/(1 - \nu)(1 + o_{\mathbb{P}}(1))$. When the degrees obey a power law, then the largest subcritical component is thus much larger than its corresponding cluster in the Erdős-Rényi random graph. This can be understood as follows. When $\nu < 1$, most components are small and are well approximated by a subcritical branching process. This is in particular true for the components attached to the d_{\max} half-edges that are incident to the maximal degree vertex. Further, one can expect that the sizes of these trees are close to being independent. A law of large number argument then predicts that the component of the vertex with maximal degree has size close to d_{\max} times the expected value of the total progeny a subcritical branching process tree that a single half-edge is part of. This expected value is approximately equal to $1/(1 - \nu)$. Janson [180] makes this intuitive argument rigorous.

Notes on Section 4.4. We follow the work with Dhara, van Leeuwaarden and Sen [103], and let us here comment on related work. Nachmias and Peres [220] study the case of critical percolation on the random regular graph. Barely supercritical results are also derived in work with Janson and Luczak [166]. Both the barely sub- as well as supercritical regimes had already been carefully analyzed by Riordan in [234], in the case of uniformly bounded degrees. However, when the degrees are bounded, many interesting phenomena are absent. Let us describe one interesting open problem in the context of subcritical configuration models with finite third-moment degrees. By Theorem 4.9, see also Theorem 4.13, the scaling of the maximal clusters inside the scaling or critical window agrees with that for the Erdős-Rényi random graph in Theorem 4.1. How about the barely

subcritical setting? We already mentioned the result by Janson [180] that in the strictly subcritical regime, the largest connected component has size $d_{\max}/(1-\nu)(1+o_{\mathbb{P}}(1))$. One could expect that when $\nu_n \rightarrow 1$ sufficiently slowly (in particular so that $\varepsilon_n = \nu_n - 1 \ll -n^{-1/3}$), this behavior persists when d_{\max} is sufficiently large. Thus, we predict that the maximal cluster has size $d_{\max}/\varepsilon_n(1+o_{\mathbb{P}}(1))$ when $\varepsilon_n \ll -n^{-1/3}$ tends to zero sufficiently slowly. When ε_n is very close to $-n^{-1/3}$ though, one would believe that the behavior in $\text{CM}_n(\mathbf{d})$ is close to that in the Erdős-Rényi random graph, where the largest barely subcritical component has size of order $\varepsilon_n^{-2} \log(n\varepsilon_n^3)$. This suggests that a *phase transition* occurs within the subcritical regime for $\text{CM}_n(\mathbf{d})$ in the finite third-moment degree setting, as formalized in the following open problem:

OPEN PROBLEM 4.13 (Phase transition in subcritical configuration models with finite third-moment degrees). *Identify the size of the largest connected component in the barely subcritical regime of $\text{CM}_n(\mathbf{d})$ when $\mathbb{E}[D_n^3] \rightarrow \mathbb{E}[D^3]$. Show that the largest connected component has a size of order $\max\left\{d_{\max}/\varepsilon_n, \varepsilon_n^{-2} \log(n\varepsilon_n^3)\right\}$, thus confirming that this setting has a phase transition.*

It would also be of interest to investigate the scaling of the second largest component in the supercritical regime, which the work with Janson and Luczak [166] does not do. Since the high-degree vertices are highly likely to reside in the unique largest component, it can be expected that this second largest connected component behaves as in the Erdős-Rényi random graph. Does the discrete duality principle still hold?

Notes on Section 4.5. Earlier work on the critical behavior of the configuration model with infinite third-moment degrees was obtained by Joseph in [189], in the context of i.i.d. degrees with a power-law degree distribution. The barely supercritical results in work with Janson and Luczak [166] also apply to this setting. We next discuss the scaling of the largest subcritical clusters, as derived in work with Bhamidi, Dhara and Sen [38]. There, it is shown that indeed only the high-degree vertices matter, and the size of the i th largest connected component is equal to $d_i/\varepsilon_n(1+o_{\mathbb{P}}(1))$, where i is the vertex with the i th largest degree. This shows that the phase transition as conjectured in Open Problem 4.13 does not occur there.

Notes on Section 4.6. The work by Bhamidi, Broutin, Sen and Wang [35] describing the metric structure of critical clusters in $\text{CM}_n(\mathbf{d})$ with finite third-moment degrees in fact extends to many more related settings, including general inhomogeneous random graphs with a finite number of types, dynamical models such as Bohman-Frieze models and bounded-size rules and rank-1 configuration models. We refer to [35] for the details. See also Section 4.8 where related settings are discussed in some detail.

Notes on Section 4.7.

Notes on Section 4.8. The results on the leader problem by Addario-Berry, Bhamidi and Sen in [3] apply to various other random graphs as well. Examples are rank-1 random graphs, as well as bounded-size rules etc.

CHAPTER 5

Ising models on random graphs

Abstract

In this chapter, we discuss Ising models on random graphs. The Ising model is one of the simplest models of random variables that are not independent, but rather positively correlated through a network structure, thus possibly giving rise to order. Here, with order we mean that without any outside preferences, the variables line up together. In the Ising model, each vertex has a random *spin* associated to it, the values being either -1 or $+1$. Nearest-neighbor vertices are more likely to have the same spin than opposing spins, and this is moderated by a coupling constant β . We are interested in the *total spin*, which equals the sum over all the vertices of the spin variables and is a measure of whether the Ising interaction forces the spins to create order on a macroscopical scale. The Ising model has a *phase transition*, much alike that in percolation, for the total spin in the model. For β small, the interaction is so weak that the spins are close to i.i.d. variables, while for β large, the spins are correlated even at large distances. We study such phenomena on complete and random graphs. We start by motivating the Ising model.

5.1. The Ising model on general graphs and motivation

The Ising model is a paradigmatic model in statistical physics to explain *magnetism*. It is one of the simplest models in statistical mechanics that shows a non-trivial phase transition when the graphs become infinitely large, which is called the *thermodynamic limit* in physics. In the Ising model, vertices are equipped with *spins* that take values in $\{-1, 1\}$. The idea is that a spin with value $+1$ points upwards, while a spin with value -1 points downwards. The key property of the Ising model is that neighboring spins are more likely to point in the same direction. With an external magnetic field, also a preference for spins to point up or rather down can be introduced. The phase transition arises in what is called the spontaneous magnetization. We take a very large graph, and give it a slight external magnetic field pointing upwards, so that the material becomes magnetized in the sense that the average spin is positive. We then take the thermodynamic limit where the system size tends to infinity. After this, we let the external field tend to zero, and we expect whether the magnetization persists. It turns out that many systems are such that there is a *phase transition* in the temperature, such that when the temperature is high, there is so much energy and movement in the particles that the magnetization is

lost, while at low temperatures, the fact that neighboring spins tend to align keeps the magnetization positive even when the external field vanishes. Similar phase transitions can also be observed in the real world, and the Ising model provides the simplest explanation of this effect. See the books by Ellis [121], Bovier [64] and Contucci and Giardinà [86] for introductions to statistical mechanics from various perspectives.

Recently, the Ising model has gained in popularity since it gives a very simple model for *consensus reaching* in populations. Indeed, when we assume that friends are more likely to have the same opinion rather than opposing opinions, then we can use the Ising model on the friendship network to model which opinion will prevail in a population and under which circumstances. As a result, the model has become popular amongst economists and social scientists as well, and later found further applications in neuroscience, etc. See for example Contucci, Gallo and Menconi [85] for an example in social sciences, Kohring [198] for an application in social impact, Fraiman, Balenzuela, Foss and Chialvo [134] for an application to the brain and Bornholdt and Wagner [63] for an example in economics.

We start by defining Ising models on finite graphs. Consider a graph sequence $(G_n)_{n \geq 1}$, where $G_n = (V_n, E_n)$, with vertex set $V_n = [n]$ and some random edge set E_n . To each vertex $i \in [n]$ we assign an Ising spin $\sigma_i = \pm 1$. A configuration of spins is denoted by $\vec{\sigma} = \{\sigma_i : i \in [n]\}$. The *Ising model on G_n* is then defined by the Boltzmann distribution

$$(5.1.1) \quad \mu_n(\vec{\sigma}) = \frac{1}{Z_n(\beta, \underline{B})} \exp \left\{ \beta \sum_{(i,j) \in E_n} \sigma_i \sigma_j + \sum_{i \in [n]} B_i \sigma_i \right\}.$$

Here, $\beta \geq 0$ is the inverse temperature and $\underline{B} = \{B_i : i \in [n]\} \in \mathbb{R}^n$ is the vector of external magnetic fields. We will write B instead of \underline{B} for a uniform external field, i.e., $B_i = B$ for all $i \in [n]$. The partition function $Z_n(\beta, \underline{B})$ is a normalization factor given by

$$(5.1.2) \quad Z_n(\beta, \underline{B}) = \sum_{\vec{\sigma} \in \{-1, +1\}^n} \exp \left\{ \beta \sum_{(i,j) \in E_n} \sigma_i \sigma_j + \sum_{i \in [n]} B_i \sigma_i \right\}.$$

We let $\langle \cdot \rangle_{\mu_n}$ denote the expectation with respect to the Ising measure μ_n , i.e., for every bounded function $f : \{-1, +1\}^n \rightarrow \mathbb{R}$,

$$(5.1.3) \quad \langle f(\sigma) \rangle_{\mu_n} = \sum_{\vec{\sigma} \in \{-1, +1\}^n} f(\sigma) \mu_n(\sigma).$$

The main quantity we shall study is the *pressure* per particle, which is defined as

$$(5.1.4) \quad \psi_n(\beta, B) = \frac{1}{n} \log Z_n(\beta, B),$$

in the thermodynamic limit of $n \rightarrow \infty$. It turns out that the pressure characterizes the behavior in the Ising model, and the phase transition and related quantities can be retrieved from it by taking appropriate derivatives with respect to the parameters B and β . For example, the *magnetization* $M_n(\beta, B)$ is given by

$$(5.1.5) \quad M_n(\beta, B) = \left\langle \frac{1}{n} \sum_{i \in [n]} \sigma_i \right\rangle_{\mu_n} = \frac{\partial}{\partial B} \psi_n(\beta, B).$$

Other relevant quantities are the *internal energy* $U_n(\beta, B)$ given by

$$(5.1.6) \quad U_n(\beta, B) = -\frac{1}{n} \sum_{(i,j) \in E_n} \langle \sigma_i \sigma_j \rangle_{\mu_n} = \frac{\partial}{\partial \beta} \psi_n(\beta, B),$$

and the *susceptibility* $\chi_n(\beta, B)$, which is given by

$$(5.1.7) \quad \chi_n(\beta, B) = \frac{1}{n} \sum_{(i,j) \in E_n} (\langle \sigma_i \sigma_j \rangle_{\mu_n} - \langle \sigma_i \rangle_{\mu_n} \langle \sigma_j \rangle_{\mu_n}) = \frac{\partial M_n}{\partial B}(\beta, B) = \frac{\partial^2}{\partial B^2} \psi_n(\beta, B).$$

We finally define the *specific heat* $C_n(\beta, B)$, which is given by

$$(5.1.8) \quad C_n(\beta, B) = \frac{\partial^2}{\partial \beta^2} \psi_n(\beta, B).$$

Thus, the pressure per particle and its derivatives give us a wealth of information about the behavior of the Ising model. The fact that much is known about the convexity and concavity properties of these quantities when considered as functions of (β, B) turns out to be extremely useful. These properties can be translated in terms of *correlation inequalities* that go under the name of GHS and GKS inequalities, and that will be discussed in detail below. These rather generally imply that convergence of the pressure implies the convergence of the magnetization, internal energy and susceptibilities as well. This leads us to the topic of *convergence* of the pressure and its derivatives, which goes under the name of *thermodynamic limits*, and will be discussed next.

Large n limits: thermodynamics and phase transition. We are interested in Ising models on random graphs, which generally are *finite*. Finite graphs do not have phase transitions, but they may appear when the graph becomes large. Such phase transitions are highly relevant, as they give us information about how robust the behavior of the Ising model on a finite graph is when the graph is quite large. Thus, we will investigate whether the thermodynamic limit of the pressure

$$(5.1.9) \quad \varphi(\beta, B) = \lim_{n \rightarrow \infty} \psi_n(\beta, B)$$

exists, and whether it is nicely differentiable as a function of (β, B) . Hopefully, then also the thermodynamic limits of the magnetization, internal energy and susceptibility also exist, and are equal to appropriate derivatives of $\varphi(\beta, B)$. When this is the case, we may investigate whether there is a phase transition by inspecting the behavior of $(\beta, B) \mapsto \varphi(\beta, B)$. In particular, the phase transition appears in the limit of small external field. Indeed, let

$$(5.1.10) \quad M(\beta, B) = \lim_{n \rightarrow \infty} M_n(\beta, B)$$

denote the thermodynamic limit of the magnetization. When $B > 0$, we can expect that $M(\beta, B) > 0$, but, by obvious symmetry, $M(\beta, 0) = 0$:

EXERCISE 5.1 (Magnetization vanishes in the absence of a magnetic field). *Prove that $M(\beta, 0) = 0$ always holds.*

While $M(\beta, 0) = 0$ always holds, $B \mapsto M(\beta, B)$ may still have discontinuities for B small that persist when $B \searrow 0$. Define the *spontaneous magnetization* by

$$(5.1.11) \quad M(\beta) = \lim_{B \searrow 0} M(\beta, B) \equiv M(\beta, 0+).$$

Here, the order of the limits of first taking $n \rightarrow \infty$ for $B > 0$, followed by $B \searrow 0$, is essential (recall Exercise 5.1). In many cases, $\beta \mapsto M(\beta)$ displays a phase transition, in the sense that there exists a β_c such that $M(\beta) = 0$ for $\beta < \beta_c$, while $M(\beta) > 0$ for $\beta > \beta_c$. This leads us to the following definition of the Ising critical value as

$$(5.1.12) \quad \beta_c = \inf\{\beta : M(\beta) > 0\}.$$

Note the close similarity to the definition of the limiting percolation critical value in (4.1.11). However, contrary to the setting for percolation where the size of the giant component is *monotone* in the percolation parameter, it is not obvious that β_c in (5.1.12) is well defined. It turns out that the Ising model has several highly non-trivial monotonicity properties that go under the name of *correlation inequalities* that will be discussed in detail in Section 5.2 below. With these correlation inequalities, one can show that β_c in (5.1.12) is indeed well defined.

The thermodynamic limits of the other relevant quantities, $U_n(\beta, B) \rightarrow U(\beta, B)$, $\chi_n(\beta, B) \rightarrow \chi(\beta, B)$ and $C_n(\beta, B) \rightarrow C(\beta, B)$, are defined similarly. Again, it is highly non-obvious that these limits exist.

Critical behavior: critical exponents and non-classical limit theorems. Having identified the asymptotic critical value of the Ising model on finite graphs, we now turn to the critical behavior. We will be interested in *critical exponents*, that quantify the singularity of the function $(\beta, B) \mapsto \varphi(\beta, B)$ for $(\beta, B) \approx (\beta_c, 0)$, as well as the behavior of the total spin

$$(5.1.13) \quad S_n = \sum_{i \in [n]} \sigma_i$$

for $(\beta, B) \approx (\beta_c, 0)$. Critical exponents are defined in the following definition. In its statement, we write $f(x) \asymp g(x)$ if the ratio $f(x)/g(x)$ is bounded away from 0 and infinity for the specified limit:

DEFINITION 5.1 (Critical exponents). The critical exponents $\beta, \delta, \gamma, \gamma'$ are defined by

$$(5.1.14) \quad M(\beta, 0^+) \asymp (\beta - \beta_c)^\beta, \quad \text{for } \beta \searrow \beta_c;$$

$$(5.1.15) \quad M(\beta_c, B) \asymp B^{1/\delta}, \quad \text{for } B \searrow 0;$$

$$(5.1.16) \quad \chi(\beta, 0^+) \asymp (\beta_c - \beta)^{-\gamma}, \quad \text{for } \beta \nearrow \beta_c;$$

$$(5.1.17) \quad \chi(\beta, 0^+) \asymp (\beta - \beta_c)^{-\gamma'}, \quad \text{for } \beta \searrow \beta_c;$$

$$(5.1.18) \quad C(\beta, 0^+) \asymp (\beta_c - \beta)^{-\alpha}, \quad \text{for } \beta \nearrow \beta_c;$$

$$(5.1.19) \quad C(\beta, 0^+) \asymp (\beta - \beta_c)^{-\alpha'}, \quad \text{for } \beta \searrow \beta_c.$$

Of course, the existence of such critical exponents on *any* graph is highly non-trivial, and identifying them has proved to be a major source of inspiration in the past decades. The definitions above are meaningful since for $\beta > \beta_c$ one has $M(\beta, 0^+) \neq M(\beta, 0) = 0$, i.e., the magnetization of the low-temperature phase is discontinuous in $B = 0$. We emphasize that there is a difference between the symbol β for the inverse temperature and the bold symbol β for the critical exponent in (5.1.14). Both notations are standard in the literature, so we decided to follow both of them and distinguish them by the font style.

Sometimes the above definitions are stated in a somewhat weaker form. For example, normally the critical exponent β is defined by requiring that

$$(5.1.20) \quad M(\beta) = M(\beta, 0^+) = (\beta - \beta_c)^{\beta + o(1)},$$

where $o(1)$ is a function tending to zero for $\beta \searrow \beta_c$.

Since the above critical exponents are obtained by *first* taking the thermodynamic limit $n \rightarrow \infty$, and only *then* taking $(\beta, B) \rightarrow (\beta_c, 0)$, it is not obvious that they tell us anything about the critical behaviour of the total spin. Its critical behavior is similar in spirit to the critical behavior of $|\mathcal{C}_{\max}(\pi_n)|$ for π_n inside the critical window. In general, one could hope for a statement of the kind that there exists a critical exponent λ and a limiting random variable X such that

$$(5.1.21) \quad \frac{S_n}{n^\lambda} \xrightarrow{d} X.$$

The above critical exponents provide the inspiration for the remainder of this chapter, where we will prove the existence of some of them for the Ising model on complete and random graphs.

Ising model as the stationary distribution of Glauber dynamics. We mainly focus on *static* aspects of the Ising model on finite graphs. However, the Ising model gains its popularity to a large extent due to its *dynamical definition*. For $\vec{\sigma} \in \{-1, +1\}^n$ and $j \in [n]$, we define

$$(5.1.22) \quad \sigma_i^j = \begin{cases} +\sigma_i & \text{for } i \neq j, \\ -\sigma_i & \text{for } i = j, \end{cases}$$

which is the version of σ where the spin at j is reversed. Then, Glauber dynamics on the set of spins is defined as a Markov chain $\vec{\sigma}(t)$ on $\{-1, +1\}^n$ where we choose $j \in [n]$ uniformly at random, and let $\vec{\sigma}(t+1)$ be $\vec{\sigma}^j(t)$ with probability

$$(5.1.23) \quad \max \left\{ e^{-[H(\vec{\sigma}^j(t)) - H(\vec{\sigma}(t))]}, 1 \right\}.$$

Here the *Hamiltonian* H is defined as

$$(5.1.24) \quad H(\vec{\sigma}) = H_{\beta, B}(\vec{\sigma}) = -\beta \sum_{(i, j) \in E_n} \sigma_i \sigma_j - \sum_{i \in [n]} B \sigma_i.$$

Thus, a move for which the Hamiltonian decreases is always accepted, while one that increases the Hamiltonian is accepted with a probability that is exponentially small in the change in Hamiltonian, thus favoring $\vec{\sigma}$ for which $H(\vec{\sigma})$ is small. This means that $\vec{\sigma}$ should align with the field B , and is such that σ_i is more likely to be equal to σ_j for every edge $(i, j) \in E_n$. The measure μ_n in (5.1.1)–(5.1.2) turns out to be the stationary distribution for this Glauber dynamics:

EXERCISE 5.2 (Ising model as stationary distribution of Glauber dynamics). *Show that the Ising measure μ_n in (5.1.1)–(5.1.2) is the stationary distribution of the dynamics in (5.1.22)–(5.1.23).*

The dynamical properties of the Ising model have attracted enormous attention, mainly for dynamics of the Ising model on cubes in \mathbb{Z}^d .

Motivation for Ising models on random graphs. When modeling a magnet, it is reasonable to take the underlying graph as a (part of a) hypercubic lattice, as real magnets live in three (or possibly two) dimensions. In the current societal setting, though, the Ising model lives within a population, where the dynamic properties of the Ising model could be interpreted as a way of consensus finding in a population. When considering random graphs as models for social networks, we thus arrive at the Ising model on a random graph. This will be the main focus of attention within this chapter.

Ising model on random graphs: quenched and annealed setting. In the remainder of this chapter, we focus on the Ising model on random graphs, where we have *double randomness*. Indeed, we have the randomness originating from the random graph as well as that from the Ising model. This means that we can consider several different settings, related to how we deal with the randomness imposed by the underlying network. The *random quenched* setting describes the Ising model on the random graph without taking the average with respect to the random graph. From an applied perspective, this may be the most natural setting. Indeed, even though the connections within the network may have arisen due to a random process, we consider them fixed and study the properties of the Ising model on them in the large graph limit. In this setting, when computing an expectation with respect to the Ising measure in we deal with a ratio of random variables (since they are functions of the random graph), and both numerator and denominator grow exponentially in the large graph limit. Typically, such problems can be difficult. Taking the expectation of the ratio over the graph, giving rise to the *averaged quenched* setting, does not resolve this difficulty unfortunately, as one still has to deal with the ratio of large random variables. A common approach is to take the average over the randomness both in numerator as well as denominator, giving rise to the *annealed* measure. In the annealed measure, the measure is replaced by

$$(5.1.25) \quad \mu_n^{\text{an}}(\vec{\sigma}) = \frac{1}{\mathbb{E}[Z_n(\beta, \underline{B})]} \mathbb{E} \left[\exp \left\{ \beta \sum_{(i,j) \in E_n} \sigma_i \sigma_j + \sum_{i \in [n]} B_i \sigma_i \right\} \right].$$

Here \mathbb{E} is the expectation over the randomness due to the random graph. In percolation, there is no partition function, so that the averaged quenched and annealed measure are the same. For the Ising model, they are quite different. It is even possible that the critical value β_c is different for the quenched and annealed settings (even though it is not hard to prove that β_c is the same in the random and averaged quenched setting). However, the general belief in physics is that the annealed and quenched cases have the same critical exponent, and are thus in the same universality class. We will see examples of such settings later in this chapter.

Organisation of Chapter 5. This chapter is organised as follows. In Section 5.2, we fix ideas by studying the Ising model on the complete graph, which is called the *Curie-Weiss model*. There, we also discuss one of the main tools used in this section, namely, *correlation inequalities*. In Section 5.3, we discuss the thermodynamic limit of the Ising model on the configuration model. The thermodynamic limit describes what happens when the graph size tends to infinity. The results immediately extend to *locally-tree like random graphs*, the prime example of which is the configuration model, but also generalized random graphs are in this class of random graphs. In Section 5.5, we discuss the critical exponents of the Ising model on the configuration model. In Section 5.6, we study the central limit theorem of the total spin away from the critical point, or,

alternatively speaking, in the uniqueness regime. In Section 5.7, we discuss non-classical limit theorems for the total spin at the critical point. We close with further results in Section 5.8, and notes and discussion in Section 5.9.

5.2. Fixing ideas: complete graph and Curie-Weiss model

In this section, we study the Ising model on the complete graph, where it is also called the *Curie-Weiss model*. On the complete graph, it turns out that a major simplification occurs that makes the model analytically tractable. Let us explain the setting first. On the complete graph, every vertex is connected to every other vertex, which makes the sum over edges very large. Recall (5.1.1)–(5.1.2) and note that $\sum_{i,j \in [n]} \sigma_i \sigma_j$ can be of order n^2 , which is much larger than the contribution of the external field $\sum_{i \in [n]} \sigma_i$, which is always of order n . In order to make sure that these terms act on the same scale, we renormalize β and replace it with β/n . We also make all the external field variables equal, i.e., $B_i = B$ for all $i \in [n]$. This gives rise to the Curie-Weiss measure given by

$$(5.2.1) \quad \mu_n(\vec{\sigma}) = \frac{1}{Z_n(\beta, B)} \exp \left\{ \frac{\beta}{n} \sum_{i < j \in [n]} \sigma_i \sigma_j + \sum_{i \in [n]} B_i \sigma_i \right\},$$

where

$$(5.2.2) \quad Z_n(\beta, B) = \sum_{\vec{\sigma} \in \{-1, +1\}^n} \exp \left\{ \frac{\beta}{n} \sum_{i < j \in [n]} \sigma_i \sigma_j + B \sum_{i \in [n]} \sigma_i \right\}.$$

The simplification that arises in the Curie-Weiss model, and that does not arise in any other model, is that $\mu_n(\sigma)$ can be described explicitly in terms of the magnetization

$$(5.2.3) \quad m_n(\vec{\sigma}) = \frac{1}{n} \sum_{i \in [n]} \sigma_i,$$

which is sometimes also called an *order parameter*. This rewrite yields

$$(5.2.4) \quad \mu_n(\vec{\sigma}) = \frac{1}{Z_n(\beta, B)} \exp \left\{ n\beta m_n(\vec{\sigma})^2/2 + Bn m_n(\vec{\sigma}) \right\},$$

where

$$(5.2.5) \quad Z_n(\beta, B) = \sum_{\vec{\sigma} \in \{-1, +1\}^n} \exp \left\{ n\beta m_n(\vec{\sigma})^2/2 + Bn m_n(\vec{\sigma}) \right\}.$$

EXERCISE 5.3 (Rewrite Curie-Weiss measure in terms of magnetization). *Prove (5.2.4)–(5.2.5).*

Thus, rather than needing to know *all* the spin variables $(\sigma_i)_{i \in [n]}$, it suffices to know their *sum*. This simplifies the problem tremendously, thus allowing for an explicit computation of various quantities.

Thermodynamic limit of the pressure. We start by analysing the thermodynamic limit of the pressure per particle:

THEOREM 5.2 (Thermodynamic limit of pressure for Curie-Weiss). *For the Curie-Weiss model, for all $0 \leq \beta < \infty$ and all $B \in \mathbb{R}$, the thermodynamic limit of the pressure exists and equals*

$$(5.2.6) \quad \lim_{n \rightarrow \infty} \psi_n(\beta, B) = \varphi(\beta, B),$$

where, for $B < 0$, $\varphi(\beta, B) = \varphi(\beta, -B)$, $\varphi(\beta, 0) = \lim_{B \downarrow 0} \varphi(\beta, B)$ and, for $B > 0$,

$$(5.2.7) \quad \varphi(\beta, B) = \log 2 + \log \cosh(z^*(\beta, B)\sqrt{\beta} + B) - z^*(\beta, B)^2/2,$$

where, for $B \neq 0$, $z^*(\beta, B)$ satisfies

$$(5.2.8) \quad z^*(\beta, B) = \sqrt{\beta} \tanh(z^*(\beta, B)\sqrt{\beta} + B).$$

For $B = 0$, the same formula holds with $z^*(\beta, B)$ replaced by $z^*(\beta) = \lim_{B \searrow 0} z^*(\beta, B)$.

We next prove Theorem 5.2, after which we discuss some of its consequences.

Proof of Theorem 5.2. Recall that

$$(5.2.9) \quad \psi_n(\beta, B) = \frac{1}{n} \log Z_n(\beta, B),$$

where the partition function $Z_n(\beta, B)$ is given in (5.2.5). While the formula in (5.2.5) is much simpler than the general formula for the partition function for the Ising model on a finite graph, it is still *quadratic* in the magnetization variable M_n , which makes it a difficult formula. When the formula would be *linear* instead, we could explicitly perform the sum over $(\sigma_i)_{i \in [n]}$, and we might be in business. We use a clever trick that turns the exponent into a *linear* function in M_n , using what is sometimes called the *Stratonovic-Hubbard* transformation. The main idea is that the moment generating function of a standard normal distribution Z satisfies $\mathbb{E}[e^{tZ}] = e^{t^2/2}$:

EXERCISE 5.4 (Moment generating function of standard normal). *Let Z have a standard normal distribution. Show that $\mathbb{E}[e^{tZ}] = e^{t^2/2}$.*

Applying this to $t = n\sqrt{\beta}m_n(\vec{\sigma})$, we thus obtain that

$$(5.2.10) \quad Z_n(\beta, B) = \sum_{\vec{\sigma} \in \{-1, +1\}^n} \mathbb{E} \left[\exp \left\{ (Z\sqrt{n\beta} + Bn)m_n(\vec{\sigma}) \right\} \right],$$

where the expectation is over Z . Now the exponent is *linear* in $m_n(\vec{\sigma})$, and, as a result, we can perform the sum over $m_n(\vec{\sigma})$. For this, we interchange the summations by Fubini, and compute the sum over $\sigma \in \{-1, +1\}^n$ to obtain

$$(5.2.11) \quad Z_n(\beta, B) = 2^n \mathbb{E} \left[\cosh \left(Z\sqrt{\beta/n} + B \right)^n \right].$$

EXERCISE 5.5 (Rewrite Curie-Weiss partition function). *Prove (5.2.11).*

So far, the computation has been exact. Now, we perform asymptotics on (5.2.11), by explicitly performing the integral over the density of Z and applying the Laplace method on the arising integral. There, we will see that the dominant contributions come from Z

that are of the order \sqrt{n} , so that the terms $Z\sqrt{\beta/n}$ and B are of similar magnitude. This leads to

$$(5.2.12) \quad Z_n(\beta, B) = 2^n (2\pi)^{-1/2} \int_{-\infty}^{\infty} e^{n \log \cosh(x\sqrt{\beta/n+B}) - x^2/2} dx.$$

Rescale $x = z\sqrt{n}$ to arrive at

$$(5.2.13) \quad Z_n(\beta, B) = 2^n \sqrt{\frac{n}{2\pi}} \int_{-\infty}^{\infty} e^{n[\log \cosh(z\sqrt{\beta+B}) - z^2/2]} dz.$$

Now we are in the classical setting where the Laplace method can be used. Laplace's method states that the integral in (5.2.13) is dominated, for n large, by those values of z that are close to the maximizer of $\log \cosh(x\sqrt{\beta} + B) - z^2/2$. To set this us, let

$$(5.2.14) \quad z^*(\beta, B) = \operatorname{argmax}[\log \cosh(z\sqrt{\beta} + B) - z^2/2],$$

so that $z^*(\beta, B)$ satisfies

$$(5.2.15) \quad z^*(\beta, B) = \sqrt{\beta} \tanh(z^*(\beta, B)\sqrt{\beta} + B).$$

It is not hard to see that $z^*(\beta, B)$ for $B > 0$ is *unique*:

EXERCISE 5.6 (Uniqueness of maximizer). *Show that, for $B > 0$, the maximizer in (5.2.14) is unique. Also show that, for $B < 0$, the maximizer is unique and satisfies $z^*(\beta, B) = -z^*(\beta, -B)$.*

EXERCISE 5.7 (Uniqueness of maximizer (cont.)). *Show also that (5.2.15) can have several solutions for $B > 0$.*

For $B = 0$, however, there may be more solutions. Fix $B \neq 0$ for the moment. By symmetry, we may then assume that $B > 0$. Then, Laplace's method gives that there exists a constant $C > 0$ such that, with $z^*(\beta, B)$ as in (5.2.14)–(5.2.15),

$$(5.2.16) \quad Z_n(\beta, B) = 2^n C e^{n[\log \cosh(z^*(\beta, B)\sqrt{\beta+B}) - z^*(\beta, B)^2/2]} (1 + o(1)).$$

EXERCISE 5.8 (Application Laplace method to Curie-Weiss partition function). *Fix $B > 0$. Prove (5.2.16) and compute C .*

Taking the log and dividing by n , followed by taking the limit $n \rightarrow \infty$, implies (5.2.7) for $B \neq 0$.

Let us now extend this argument to $B = 0$. There, the problem is that (5.2.15) may have several solutions, including the obvious solution 0 as well as the solutions $z^*(\beta) = \lim_{B \searrow 0} z^*(\beta, B)$, as well as $-z^*(\beta)$ by symmetry. Let us investigate this in more detail. Inspecting $z \mapsto \log \cosh(z\sqrt{\beta} + B) - z^2/2$, we see that for $B = z = 0$, it gives the trivial value 0. The question arises whether there is a solution that achieves a higher value. When $z \mapsto \log \cosh(z\sqrt{\beta}) - z^2/2$ is increasing in z for z small, the maximal solution cannot be 0. The derivative of $z \mapsto \log \cosh(z\sqrt{\beta}) - z^2/2$ at $z = 0$ equals 0, so this does not help. Hence, whether $z \mapsto \log \cosh(z\sqrt{\beta}) - z^2/2$ is larger than 0 for z small is determined by sign of the double derivative, which is $\beta - 1$. When $\beta > 1$, the two maximizers are $z^*(\beta) = \lim_{B \searrow 0} z^*(\beta, B)$ and $-z^*(\beta)$, where $z^*(\beta) > 0$. For $\beta \leq 1$, on the other hand, the maximal solution is $z^*(\beta) = 0$. In the latter case, however, the solution $z^*(\beta) = 0$ is again *unique*, so also $z^*(\beta) = \lim_{B \searrow 0} z^*(\beta, B) = 0$. This completes the argument \square

The above analysis made use of the Stratonovic-Hubbard transformation. However, the existence of the pressure in (5.2.6) can also be directly obtained:

EXERCISE 5.9 (Direct computation of Curie-Weiss pressure). *Prove the existence of the pressure in (5.2.6) directly, by using that there are precisely $\binom{n}{(n+m)/2}$ vectors of spins with total spin $m_n(\vec{\sigma}) = m/n$. Can you also verify that the final formulas for the pressure agree? [Hint: See Bovier and Kurkova [66] for inspiration if necessary.]*

Thermodynamic limits of magnetization, internal energy and susceptibility.

Following the proof of Theorem 5.2, we can also identify the thermodynamic limits of related quantities in the Curie-Weiss model:

THEOREM 5.3 (Thermodynamic limits Curie-Weiss model). *Fix $B > 0$. The Curie-Weiss model satisfies the following:*

(a) Magnetization: *Let $M_n(\beta, B) = \frac{1}{n} \sum_{i \in [n]} \langle \sigma_i \rangle_{\mu_n}$ be the magnetization per vertex. Then, its thermodynamic limit exists and is given by*

$$(5.2.17) \quad M(\beta, B) \equiv \lim_{n \rightarrow \infty} M_n(\beta, B) = \tanh(z^*(\beta, B)\sqrt{\beta} + B) = \frac{\partial}{\partial B} \varphi(\beta, B).$$

Consequently, $M(\beta, B)$ is the largest solution to

$$(5.2.18) \quad M(\beta, B) = \tanh(M(\beta, B)\beta + B),$$

as well as to $M(\beta, B) = \sqrt{\beta} \operatorname{argmax}[\log \cosh(m\beta + B) - z^2/2]$.

(b) Internal energy: *Let*

$$(5.2.19) \quad U_n(\beta, B) = -\frac{\partial}{\partial \beta} \psi_n(\beta, B) = -\frac{1}{2n^2} \sum_{(i,j) \in E} \langle \sigma_i \sigma_j \rangle_{\mu_n}$$

be the internal energy per vertex. Then, its thermodynamic limit exists and is given by

$$(5.2.20) \quad U(\beta, B) \equiv \lim_{n \rightarrow \infty} U_n(\beta, B) = \frac{1}{2} \tanh(z^*(\beta, B)\sqrt{\beta} + B)^2 = -\frac{\partial}{\partial \beta} \varphi(\beta, B).$$

(c) Susceptibility: *Let*

$$(5.2.21) \quad \chi_n(\beta, B) = \frac{1}{n} \sum_{(i,j) \in E_n} (\langle \sigma_i \sigma_j \rangle_{\mu_n} - \langle \sigma_i \rangle_{\mu_n} \langle \sigma_j \rangle_{\mu_n}) = \frac{\partial M_n}{\partial B}(\beta, B)$$

be the susceptibility. Then, its thermodynamic limit exists and is given by

$$(5.2.22) \quad \chi(\beta, B) \equiv \lim_{n \rightarrow \infty} \chi_n(\beta, B) = \frac{\partial^2}{\partial B^2} \varphi(\beta, B) = \chi(\beta, B).$$

Proof of Theorem 5.3. Denote the expected value of the magnetization by

$$(5.2.23) \quad \langle M_n \rangle_{\mu_n} = \frac{1}{Z_n(\beta, B)} \sum_{\vec{\sigma} \in \{-1, +1\}^n} m_n(\vec{\sigma}) \exp \{ n\beta m_n(\vec{\sigma})^2/2 + Bnm_n(\vec{\sigma}) \}.$$

Following the steps in the proof of Theorem 5.2, we can write

$$\begin{aligned}
(5.2.24) \quad & \sum_{\vec{\sigma} \in \{-1, +1\}^n} m_n(\vec{\sigma}) \exp \left\{ n\beta m_n(\vec{\sigma})^2 / 2 + Bn m_n(\vec{\sigma}) \right\} \\
&= \sum_{\vec{\sigma} \in \{-1, +1\}^n} \mathbb{E} \left[m_n(\vec{\sigma}) \exp \left\{ (Z\sqrt{n\beta} + Bn) m_n(\vec{\sigma}) \right\} \right] \\
&= 2^n \mathbb{E} \left[\sinh (Z\sqrt{\beta/n} + B) \cosh (Z\sqrt{\beta/n} + B)^{n-1} \right] \\
&= 2^n \mathbb{E} \left[\tanh (Z\sqrt{\beta/n} + B) \cosh (Z\sqrt{\beta/n} + B)^n \right].
\end{aligned}$$

This gives rise to an identical integral as for $Z_n(\beta, B)$ as in (5.2.12), apart from the single factor $\tanh(z\sqrt{\beta} + B)$. This factor does not change by which regime in z the integral is dominated, and merely replaces the limit by a factor $\tanh(z^*(\beta, B)\sqrt{\beta} + B)$, as required. To see how (5.2.18) arises, note by (5.2.15) and (5.2.17) that $M(\beta, B) = z^*(\beta, B)/\sqrt{\beta}$. Substituting this into (5.2.15) gives the claim.

In a similar way,

$$(5.2.25) \quad \langle U_n \rangle_{\mu_n} = \frac{1}{2Z_n(\beta, B)} \sum_{\vec{\sigma} \in \{-1, +1\}^n} m_n(\vec{\sigma})^2 \exp \left\{ n\beta m_n(\vec{\sigma})^2 / 2 + Bn m_n(\vec{\sigma}) \right\},$$

and

$$\begin{aligned}
& \sum_{\vec{\sigma} \in \{-1, +1\}^n} m_n(\vec{\sigma})^2 \exp \left\{ n\beta m_n(\vec{\sigma})^2 / 2 + Bn m_n(\vec{\sigma}) \right\} \\
&= \sum_{\vec{\sigma} \in \{-1, +1\}^n} \mathbb{E} \left[m_n(\vec{\sigma})^2 \exp \left\{ (Z\sqrt{n\beta} + Bn) m_n(\vec{\sigma}) \right\} \right] \\
&= \frac{2^n}{n^2} \mathbb{E} \left[\left(n(n-1) \sinh (Z\sqrt{\beta/n} + B)^2 + n \right) \cosh (Z\sqrt{\beta/n} + B)^{n-2} \right] \\
(5.2.26) \quad &= 2^n \mathbb{E} \left[\left(\tanh (Z\sqrt{\beta/n} + B)^2 + \frac{1}{n} \right) \cosh (Z\sqrt{\beta/n} + B)^n \right].
\end{aligned}$$

We omit further details.

The computation for the susceptibility is substantially harder for the Curie-Weiss model, since (5.2.21) contains a double sum containing $O(n^2)$ terms, whereas the prefactor is $1/n$. Using (5.2.24) and (5.2.26), we see that

$$(5.2.27) \quad \chi_n(\beta, B) = n \text{Var}_n \left(\tanh (Z\sqrt{\beta/n} + B) \right) + 1,$$

where Z_1, Z_2 are two i.i.d. standard normals and where Var_n denotes the variance under the measure \mathbb{E}_n given by

$$(5.2.28) \quad \mathbb{E}_n[f(Z)] = \frac{\mathbb{E} \left[f(Z) \cosh (Z\sqrt{\beta/n} + B)^n \right]}{\mathbb{E} \left[\cosh (Z\sqrt{\beta/n} + B)^n \right]}.$$

EXERCISE 5.10 (Proof of (5.2.27)). *Prove that (5.2.27) holds.*

Rather than pursuing this route, we instead follow another one that will also yield a central limit theorem for the total spin at no added cost. For this, we note that

$$(5.2.29) \quad \chi_n(\beta, B) = \frac{1}{n} \text{Var}_{\mu_n}(S_n) = \text{Var}_{\mu_n} \left(\frac{S_n - nM_n(\beta, B)}{\sqrt{n}} \right), \quad \text{where} \quad S_n = \sum_{v \in [n]} \sigma_v,$$

so that S_n denotes the total spin. Rather than studying the variance of $\bar{S}_n = (S_n - nM_n(\beta, B))/\sqrt{n}$, we study its moment generating function. Convergence of the moment generating function certainly implies convergence of the variance, as the argument will also show. Let us start by setting up the notation. Define

$$(5.2.30) \quad c_n(t) = \psi_n(\beta, B + t) - \psi_n(\beta, B).$$

Note that $c_n(t) \rightarrow \varphi(\beta, B + t)$ by Theorem 5.2. Then

$$(5.2.31) \quad \frac{1}{n} \log \left\langle e^{tS_n} \right\rangle_{\mu_n} = c_n(t).$$

We conclude that, with $t_n = t/\sqrt{n}$,

$$(5.2.32) \quad \log \left\langle e^{t\bar{S}_n} \right\rangle_{\mu_n} = n \left[c_n(t_n) - t_n c'_n(0) \right].$$

This smells like a Taylor expansion, but there are some technical issues. First, we would like to replace $c_n(t)$ by its limit $c(t) = \varphi(\beta, B + t) - \varphi(\beta, B)$. Doing so would tell us that, also using that $c_n(0) = c(0) = 0$,

$$(5.2.33) \quad \log \left\langle e^{t\bar{S}_n} \right\rangle_{\mu_n} \approx n \left[c(t_n) - t_n c'(0) \right] \approx \frac{t^2}{2} \frac{\partial^2}{\partial B^2} \varphi(\beta, B) = \frac{t^2}{2} \chi(\beta, B),$$

by the definition in (5.2.21). However, we are not allowed to simply replace $c_n(t)$ by its limit so easily. In order to prove that, we use some analytical properties. We note that

$$(5.2.34) \quad c'_n(t) = \frac{d}{dt} \psi_n(\beta, B + t) = M_n(\beta, B + t).$$

We know that $c'_n(t) \rightarrow c'(t) = M(\beta, B + t) = \frac{d}{dt} \varphi(\beta, B + t)$ by part (a) of this theorem (which has already been proved). It turns out that $t \mapsto c'_n(t)$ is also *concave*, which is a consequence of an important correlation inequality that is described just below (see the GHS inequality in Lemma 5.5). Now, for a concave function that converges pointwise, also its derivative converges pointwise, so that $c''_n(t) \rightarrow \frac{\partial^2}{\partial B^2} \varphi(\beta, B + t)$. This can be extended to show that also for $t_n \rightarrow t$, $c''_n(t_n) \rightarrow \frac{\partial^2}{\partial B^2} \varphi(\beta, B + t)$. This is explained in more detail in Section 5.6, see in particular Lemma 5.35. A Taylor expansion yields that

$$(5.2.35) \quad \log \left\langle e^{t\bar{S}_n} \right\rangle_{\mu_n} = n \left[c_n(t_n) - t_n c'_n(0) \right] = \frac{t^2}{2} c''_n(t_n^*),$$

for some t_n^* in between 0 and $t_n = t/\sqrt{n}$. In particular, $t_n^* \rightarrow 0$, so that by the above arguments and $c_n(0) = 0$,

$$(5.2.36) \quad \log \left\langle e^{t\bar{S}_n} \right\rangle_{\mu_n} \rightarrow \frac{t^2}{2} \chi(\beta, B).$$

To obtain the convergence of $\chi_n(\beta, B)$, we merely note that $\chi_n(\beta, B) = c''_n(0)$. \square

The above argument actually shows more, namely, also that S_n satisfies a central limit theorem:

EXERCISE 5.11 (Central limit for Curie-Weiss total spin). *Prove that*

$$(5.2.37) \quad \frac{S_n - nM_n(\beta, B)}{\sqrt{n}} \xrightarrow{d} Z,$$

where Z has a normal distribution with mean 0 and variance $\chi(\beta, B)$.

Trick of the trade: Correlation inequalities. Recall below (5.1.12) that it is yet unclear that the value of β_c is well defined. We now discuss some extremely useful properties of the Ising model on general graphs that go under the name of *correlation inequalities*. The first such correlation inequality is the *Griffiths, Kelly, Sherman (GKS) inequality*, which gives various monotonicity properties in the parameters (β, B) . Here we assume that we are in the setting that each vertex $v \in V$ has its own magnetic field B_v :

LEMMA 5.4 (GKS inequality). *Consider two Ising measures μ and μ' on graphs $G = (V, E)$ and $G' = (V, E')$, with inverse temperatures β and β' and external fields \underline{B} and \underline{B}' , respectively. If $E \subseteq E'$, $\beta \leq \beta'$ and $0 \leq B_v \leq B'_v$ for all $v \in V$, then, for any $U \subseteq V$,*

$$(5.2.38) \quad 0 \leq \langle \prod_{u \in U} \sigma_u \rangle_\mu \leq \langle \prod_{u \in U} \sigma_u \rangle_{\mu'}.$$

In the following exercises, we highlight the importance of the GKS inequality in Lemma 5.4, firstly by showing that β_c in (5.1.12) is well defined, and secondly by showing that for $B > 0$, the *maximal* solution to (5.2.15), which is the unique positive solution:

EXERCISE 5.12 (Well-definedness of spontaneous magnetization). *Assume that the thermodynamic limit of the magnetization in (5.1.10) exists for all $B > 0$. Use the GKS inequality in Lemma 5.4 to conclude that $B \mapsto M(\beta, B)$ is non-decreasing, and conclude that $M(\beta, 0^+)$ is well-defined. Show further that for every $B \geq 0$, $\beta \mapsto M(\beta, B)$ is non-decreasing and conclude that β_c in (5.1.12) is well defined.*

EXERCISE 5.13 (Well-definedness of critical value). *Show that for every $B \geq 0$, $\beta \mapsto M(\beta, B)$ is non-decreasing and conclude that β_c in (5.1.12) is well defined.*

EXERCISE 5.14 (Uniqueness of maximizer (cont.)). *For $B > 0$, (5.2.15) can have several solutions by Exercise 5.7. Use the GKS inequality in Lemma 5.4 to deduce that the unique positive solution is the one we look for in (5.2.14).*

A weaker version of this correlation inequality was first proved by Griffiths [142] and later generalized by Kelly and Sherman [193]. The second result on ferromagnetic Ising models is an inequality by Griffiths, Hurst and Sherman [143] which shows the concavity of the magnetization in the external (positive) magnetic fields. We have already seen that this is an extremely powerful result, as it gives a central limit theorem for the total spin almost for free (recall Exercise 5.11):

LEMMA 5.5 (GHS inequality). *Let $\beta \geq 0$ and $B_v \geq 0$ for all $v \in V$. Denote by*

$$(5.2.39) \quad m_j(\underline{B}) = \mu(\{\sigma : \sigma_j = +1\}) - \mu(\{\sigma : \sigma_j = -1\})$$

the magnetization of vertex j when the external fields at the vertices are \underline{B} . Then, for any three vertices $j, k, l \in V$,

$$(5.2.40) \quad \frac{\partial^2}{\partial B_k \partial B_l} m_j(\underline{B}) \leq 0.$$

Lemma 5.5 in particular allows us to prove that the magnetization $B \mapsto M_n(\beta, B)$ is concave, so that also $t \mapsto c'_n(t)$ in (5.2.34) is concave:

EXERCISE 5.15 (Concavity of magnetization in external field). *Use the GHS inequality in Lemma 5.5 to prove that $B \mapsto M_n(\beta, B)$ is concave.*

The strength of the correlation inequalities in Lemmas 5.4 and 5.5 is that they are valid for all graphs. They will prove to be extremely useful to conclude the thermodynamic limits in various settings, when explicit computations as for the Curie-Weiss model in Theorem 5.3 are impossible.

Critical value of the Curie-Weiss model. The proof of Theorem 5.2 already hints at the special role of $\beta = 1$. The following theorem shows that $\beta_c = 1$ is the critical value of the Curie-Weiss model:

THEOREM 5.6 (Critical value of Curie-Weiss model). *For the Curie-Weiss model, the critical value equals $\beta_c = 1$. More precisely, the thermodynamic limit of the magnetization $M(\beta, B)$, which is the largest solution to $M(\beta, B) = \beta \tanh(M(\beta, B)\beta + B)$, satisfies that $M(\beta, 0+) > 0$ for $\beta > 1$, while $M(\beta, 0+) = 0$ for $\beta \leq 1$.*

PROOF. Important results needed in the proof of Theorem 5.6 were already derived in the proof of the case that $B = 0$ in Theorem 5.2. Indeed, we proved there that $z^*(\beta) = \lim_{B \searrow 0} z^*(\beta, B) > 0$ for $\beta > 1$, while the maximal solution is $z^*(\beta) = \lim_{B \searrow 0} z^*(\beta, B) = 0$ for $\beta \leq 1$. Since $M(\beta, B) = z^*(\beta, B)/\sqrt{\beta}$, this proves the claim. \square

Critical exponents of the Curie-Weiss model. Having identified the critical value of the Curie-Weiss model, we next analyse the critical exponents:

THEOREM 5.7 (Critical exponents Curie-Weiss model). *For the Curie-Weiss model, the critical exponents β , δ and γ defined in Definition 5.1 exist and satisfy $\beta = \frac{1}{2}$, $\delta = 3$, and $\gamma = 1$.*

PROOF. We start with the analysis of β and δ , for which we need to take $\beta \geq \beta_c = 1$ and $B \geq 0$. For β , we then need to take $B \searrow 0$ and $\beta > 1$, while for δ , we need to take $\beta = \beta_c = 1$ and $B > 0$ small. Recall from (5.2.18) that $M(\beta, B)$ is the largest solution to $M(\beta, B) = \tanh(M(\beta, B)\beta + B)$. It can be expected that $M(\beta, B)$ vanishes when $(\beta, B) \approx (\beta_c, 0)$, but this requires a proof. We leave this proof to the reader, and refer ahead to a similar computation for the Ising model on random graphs where this property is being proved rigorously:

EXERCISE 5.16 (Phase transition is continuous). *Show that $\lim_{(\beta, B) \searrow (\beta_c, 0)} M(\beta, B) = 0$, so that the phase transition of the Curie-Weiss model is continuous. Use the GKS inequality in Lemma 5.4 to argue that $\lim_{(\beta, B) \searrow (\beta_c, 0)} M(\beta, B)$ is a monotone limit and thus exists. Denote it by c . Argue by contradiction by assuming that $c > 0$ and use that $\tanh(c) < c$ to reach a contradiction. [Hint: You can also compare to the related result in Lemma 5.23 for the Ising model on random graphs.]*

By Exercise 5.16, we know that $M(\beta, B)$ is small when $(\beta, B) \approx (\beta_c, 0)$. A Taylor expansion gives, for small x ,

$$(5.2.41) \quad \tanh(x) = x - \frac{x^3}{6} + o(x^3).$$

Thus,

$$(5.2.42) \quad M(\beta, B) = \tanh(M(\beta, B)\beta + B) = (M(\beta, B)\beta + B) - \frac{1}{6}(M(\beta, B)\beta + B)^3(1 + o(1)).$$

We start with β , for which we let $B \searrow 0$ for $\beta > 1$ to arrive at

$$(5.2.43) \quad M(\beta, 0^+) = M(\beta, 0^+)\beta - \frac{1}{6}(M(\beta, 0^+)\beta)^3(1 + o(1)).$$

Dividing through by $M(\beta, 0^+)$ (which is allowed since $M(\beta, 0^+) > 0$ for $\beta > \beta_c = 1$) leads to

$$(5.2.44) \quad M(\beta, 0^+)^2(1 + o(1)) = 6(\beta - 1) = 6(\beta - \beta_c),$$

so that

$$(5.2.45) \quad M(\beta, 0^+) = \sqrt{6(\beta - \beta_c)}(1 + o(1)).$$

This establishes that $\beta = \frac{1}{2}$.

For δ , we let $B > 0$ and take $\beta \searrow \beta_c = 1$ to arrive at

$$(5.2.46) \quad M(1, B) = (M(1, B) + B) - \frac{1}{6}(M(1, B) + B)^3(1 + o(1)).$$

Subtracting $M(1, B)$ from both sides leads to

$$(5.2.47) \quad M(1, B)^3(1 + o(1)) = 6B,$$

so that

$$(5.2.48) \quad M(1, B) = \sqrt[3]{6B}(1 + o(1)).$$

Since this exponent is written as $1/\delta$, this establishes that $\delta = 3$. We omit the proof of $\gamma = 1$. \square

Non-classical limit theorems for the total spin. The critical exponents in Theorem 5.7 arise when first taking the thermodynamic limit of large graphs, followed by the limit of the parameters (β, B) to their critical value $(\beta_c, 0) = (1, 0)$. We now investigate what happens when $(\beta, B) = (\beta_c, 0)$. We already know that then the magnetization vanishes (Theorem 5.7 gives much stronger results than that), so when considering the total spin $S_n = \sum_{v \in [n]} \sigma_v$, we can expect that $S_n/n \xrightarrow{\mathbb{P}} 0$. The following theorem describes rather precisely *how* this occurs:

THEOREM 5.8 (Non-classical limit theorem at criticality). *Fix $\beta = \beta_c$ and $B = 0$ in the Curie-Weiss model. There exists a random variable X such that*

$$(5.2.49) \quad \frac{S_n}{n^{3/4}} \xrightarrow{d} X, \quad \text{as } n \rightarrow \infty,$$

where the convergence is w.r.t. the measure μ_n at inverse temperature $\beta_c = 1$ and external field $B = 0$. The random variable X has a density proportional to $\exp(-x^4/12)$.

PROOF. We consider the moment generating function of the total spin,

$$(5.2.50) \quad \phi_n(r) = \left\langle e^{rn^{-3/4}S_n} \right\rangle_{\mu_n}.$$

It suffices to show that this moment generating function converges to that of X . For this, we use (5.2.11) to rewrite

$$(5.2.51) \quad \phi_n(r) = \frac{\mathbb{E} \left[\cosh \left(Z/\sqrt{n} + rn^{-3/4} \right)^n \right]}{\mathbb{E} \left[\cosh \left(Z/\sqrt{n} \right)^n \right]} = \frac{H_n(r)}{H_n(0)},$$

where

$$(5.2.52) \quad H_n(r) = \mathbb{E} \left[\cosh \left(Z/\sqrt{n} + rn^{-3/4} \right)^n \right].$$

Let us now study $H_n(r)$ more precisely. Writing out the expectation over Z explicitly, we obtain

$$(5.2.53) \quad H_n(r) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-z^2/2} \cosh \left(z/\sqrt{n} + rn^{-3/4} \right)^n dz.$$

A substitution $y = z/\sqrt{n}$ leads to

$$(5.2.54) \quad H_n(r) = \frac{\sqrt{n}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp \left\{ n[\log \cosh(y + rn^{-3/4}) - y^2/2] \right\} dy.$$

A shift of the integrand over $rn^{-3/4}$ leads to

$$(5.2.55) \quad \begin{aligned} H_n(r) &= \frac{\sqrt{n}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp \left\{ n[\log \cosh(y) - (y - rn^{-3/4})^2/2] \right\} dy \\ &= \frac{\sqrt{n}}{\sqrt{2\pi}} e^{-r^2 n^{-1/2}} \int_{-\infty}^{\infty} e^{n^{1/4} r y} \exp \left\{ n[\log \cosh(y) - y^2/2] \right\} dy. \end{aligned}$$

The term $e^{-r^2 n^{-1/2}} = 1 + o(1)$ can be omitted. We have again arrived at an integral that needs a Laplace method. However, the integral is special, since $\log \cosh(y) - y^2/2 \leq 0$, with the unique minimum attained at $y = 0$, while a Taylor expansion leads to

$$(5.2.56) \quad \log \cosh(y) - y^2/2 = -\frac{1}{12}y^4 + O(y^6).$$

This gives rise to

$$(5.2.57) \quad H_n(r) = (1 + o(1)) \frac{\sqrt{n}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{n^{1/4} r y} e^{-ny^4/12} dx.$$

The final substitution $x = yn^{1/4}$ then yields

$$(5.2.58) \quad H_n(r) = (1 + o(1)) \frac{n^{3/4}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{rx} e^{-x^4/12} dx.$$

We conclude that

$$(5.2.59) \quad \phi_n(r) \rightarrow \frac{\int_{-\infty}^{\infty} e^{rx} e^{-x^4/12} dx}{\int_{-\infty}^{\infty} e^{-x^4/12} dx}.$$

This proves the claim, as the right hand side is equal to the moment generating function of a random variable X with density proportional to $e^{-x^4/12}$. \square

Discussion and Ising models on random graphs. The beauty of the Curie-Weiss model is that it is analytically tractable, in the sense that virtually all computations can be performed explicitly. The aim of the remainder of this chapter is to extend such analyses to *random graphs*, where such explicit computations can not be done so cleanly. However, some of the main tools, such as the GKS and GHS inequalities in Lemmas 5.4-5.5, remain valid in this setting. The main tool that replaces the exact computations is *local weak convergence* (recall Section 2.2). The GKS and GHS inequalities in Lemmas 5.4-5.5 will show that the thermodynamic limit of the pressure can be seen as a continuous functional, and thus converges. This allows us to compute the thermodynamic limits, the

critical value, as well as some critical exponents. The non-classical central limit theorems as in Theorem 5.8 do not follow from such local weak convergence arguments, and so far can only be performed in annealed settings, where the annealed measure looks like an inhomogeneous version of the Curie-Weiss model.

A central property of the random graphs that we consider is their *inhomogeneity*. Just like for percolation on random graphs and its relation to the Erdős-Rényi random graph, we will see that the behavior of the Ising model is like that for the Curie-Weiss model when the inhomogeneity is not too pronounced, while it changes dramatically when it is too pronounced. Remarkably, the decisive criterion for the Ising model is the finiteness of the *fourth* moment of the degrees, rather than the third moment as for percolation. Let us start by analysing the thermodynamic limit on random graphs.

5.3. Thermodynamic limits on the configuration model

In this section we study the behavior of the Ising model on the configuration model. We assume that the degree distribution has a *strongly finite mean*, with which we mean that the limiting degree D satisfies that there exists $\tau > 2$ and c_D such that

$$(5.3.1) \quad 1 - F_D(x) \leq c_D x^{-(\tau-1)}.$$

EXERCISE 5.17 (Strongly finite mean). *Let D satisfy (5.3.1) for some $\tau > 2$. Show that $\mathbb{E}[D^p] < \infty$ for any $p < \tau - 1$.*

In this section, we follow the work with Dommers and Giardinà [109], which, in turn, was inspired by Dembo and Montanari [92]. We first investigate the thermodynamic limit of the pressure:

THEOREM 5.9 (Thermodynamic limit of the pressure). *Consider $\text{CM}_n(\mathbf{d})$ where the degree distribution satisfies Conditions 1.6(a)-(b), and has a strongly finite mean as in (5.3.1). Then, for all $0 \leq \beta < \infty$ and all $B \in \mathbb{R}$, the thermodynamic limit of the pressure exists and is deterministic:*

$$(5.3.2) \quad \psi_n(\beta, B) \xrightarrow{\mathbb{P}} \varphi(\beta, B),$$

The thermodynamic limit of the pressure satisfies, for $B < 0$, $\varphi(\beta, B) = \varphi(\beta, -B)$, and $\varphi(\beta, 0) = \lim_{B \downarrow 0} \varphi(\beta, B)$ for $B = 0$. For $B > 0$, it equals

$$(5.3.3) \quad \varphi(\beta, B) = \frac{\mathbb{E}[D]}{2} \log \cosh(\beta) - \frac{\mathbb{E}[D]}{2} \mathbb{E}[\log(1 + \tanh(\beta) \tanh(h_1) \tanh(h_2))] \\ + \mathbb{E} \left[\log \left(e^B \prod_{i=1}^D \{1 + \tanh(\beta) \tanh(h_i)\} + e^{-B} \prod_{i=1}^D \{1 - \tanh(\beta) \tanh(h_i)\} \right) \right],$$

where

- (i) $(h_i)_{i \geq 1}$ are i.i.d. copies of the fixed point $h^* = h^*(\beta, B)$ of the distributional recursion

$$(5.3.4) \quad h^{(t+1)} \stackrel{d}{=} B + \sum_{i=1}^{\Delta_t} \text{atanh}(\tanh(\beta) \tanh(h_i^{(t)})),$$

where $h^{(0)} \equiv B$, $(\Delta_t)_{t \geq 1}$, are i.i.d. random variables with distribution $D^* - 1$ and $(h_i^{(t)})_{i \geq 1}$ are i.i.d. copies of $h^{(t)}$ independent of Δ_t ;

(ii) D and $(h_i)_{i \geq 1}$ are independent.

The quantity $\varphi(\beta, B)$ can be seen as the infinite volume pressure of the Ising model on the random Bethe lattice, where every vertex has degree distribution D . This connection will be made more precise below. Baxter [30] studies the Ising model on the regular Bethe lattice, which is a special case arising for random regular graphs:

EXERCISE 5.18 (Pressure on the random 2-regular graph). *Show that*

$$(5.3.5) \quad \varphi(\beta, B) = \beta + \log \left(\cosh(B) + \sqrt{\sinh(B)^2 + e^{-4\beta}} \right).$$

for the random 2-regular graphs, where $d_i = 2$ for all $i \in [n]$.

EXERCISE 5.19 (Pressure on the random regular graph). *Can you solve the recursion (5.3.4) for random regular graphs, where $d_i = r$ for some $r \geq 3$ and every $i \in [n]$, and this way conclude what the thermodynamic limit of the pressure is in this case?*

One of the main difficulties in proving Theorem 5.9 will be to deal with the *inhomogeneity* in the random graph. This inhomogeneity gives rise to the random variables D and $(h_i)_{i \geq 1}$, as well as in the recursion relation in (5.3.4). The fact that the limiting pressure is *deterministic* can be interpreted as a sign that the randomness washes out in the large graph limit. More technically, we will see that this is closely related to the local weak convergence *in probability* as proved in Theorem 2.11.

Various thermodynamic quantities can be computed by taking the proper derivative of the function $\varphi(\beta, B)$, as we have already seen in detail for the Curie-Weiss model in Theorem 5.3:

THEOREM 5.10 (Thermodynamic quantities). *Consider $\text{CM}_n(\mathbf{d})$ where the degree distribution satisfies Conditions 1.6(a)-(b) hold, and has a strongly finite mean as in (5.3.1). Then, for all $\beta \geq 0$ and $B \neq 0$, each of the following statements holds:*

(a) Magnetization: *Let $M_n(\beta, B) = \frac{1}{n} \sum_{i \in [n]} \langle \sigma_i \rangle_{\mu_n}$ be the magnetization per vertex. Then, its thermodynamic limit exists and is given by*

$$(5.3.6) \quad M_n(\beta, B) \xrightarrow{\mathbb{P}} M(\beta, B) = \frac{\partial}{\partial B} \varphi(\beta, B).$$

(b) Internal energy: *Let $U_n(\beta, B) = -\frac{1}{n} \sum_{(i,j) \in E} \langle \sigma_i \sigma_j \rangle_{\mu_n}$ be the internal energy per vertex. Then, its thermodynamic limit exists and is given by*

$$(5.3.7) \quad U_n(\beta, B) \xrightarrow{\mathbb{P}} U(\beta, B) = -\frac{\partial}{\partial \beta} \varphi(\beta, B).$$

(c) Susceptibility: *Let $\chi_n(\beta, B) = \frac{1}{n} \sum_{(i,j) \in E_n} (\langle \sigma_i \sigma_j \rangle_{\mu_n} - \langle \sigma_i \rangle_{\mu_n} \langle \sigma_j \rangle_{\mu_n}) = \frac{\partial M_n}{\partial B}(\beta, B)$ be the susceptibility. Then, its thermodynamic limit exists and is given by*

$$(5.3.8) \quad \chi_n(\beta, B) \xrightarrow{\mathbb{P}} \chi(\beta, B) = \frac{\partial^2}{\partial B^2} \varphi(\beta, B).$$

EXERCISE 5.20 (Thermodynamic limits on the random 2-regular graph). *Compute the thermodynamic limits in Theorem 5.10 for the random 2-regular graph, for which $d_i = 2$ for every $i \in [n]$.*

EXERCISE 5.21 (Thermodynamic limits on the random d -regular graph for $d \geq 3$). Compute the thermodynamic limits in Theorem 5.10 for the random regular graph, for which $d_i = d$ for some $d \geq 3$ for every $i \in [n]$.

Another physical quantity studied in the physics literature is the *specific heat*,

$$(5.3.9) \quad C_n(\beta, B) \equiv -\beta^2 \frac{\partial U_n}{\partial \beta}.$$

Unfortunately, we were not able to prove that this converges to $\beta^2 \frac{\partial^2}{\partial \beta^2} \varphi(\beta, B)$, because we do not have convexity or concavity of the internal energy in β . We expect, however, that this limit also converges to the appropriate limit.

Taking the derivatives of the form of the pressure in (5.3.3) in Theorem 5.10 we can also give explicit expressions for the magnetization and internal energy which have a physical interpretation:

COROLLARY 5.11 (Explicit expressions for thermodynamic quantities). Consider $\text{CM}_n(\mathbf{d})$ where the degree distribution satisfies Conditions 1.6(a)-(b), and has a strongly finite mean as in (5.3.1). Then, for all $\beta \geq 0$ and $B \in \mathbb{R}$, each of the following statements holds a.s.:

(a) **Magnetization:** Let ν_{D+1} be the random Ising measure on a tree with $D+1$ vertices (one root and D leaves), defined by

$$(5.3.10) \quad \nu_{D+1}(\sigma) = \frac{1}{Z_{D+1}(\beta, h^*)} \exp \left\{ \beta \sum_{i=1}^D \sigma_0 \sigma_i + B \sigma_0 + \sum_{i=1}^D h_i \sigma_i \right\},$$

where $(h_i)_{i \geq 1}$ are i.i.d. copies of h^* , independent of D . Then, the thermodynamic limit of the magnetization per vertex is given by

$$(5.3.11) \quad M(\beta, B) = \mathbb{E} [\langle \sigma_0 \rangle_{\nu_{D+1}}],$$

where the expectation is taken over D and $(h_i)_{i \geq 1}$. More explicitly,

$$(5.3.12) \quad M(\beta, B) = \mathbb{E} \left[\tanh \left(B + \sum_{i=1}^D \text{atanh}(\tanh(\beta) \tanh(h_i)) \right) \right].$$

(b) **Internal energy:** Let ν'_2 be the random Ising measure on one edge, defined by

$$(5.3.13) \quad \nu'_2(\sigma) = \frac{1}{Z_2(\beta, h_1, h_2)} \exp \{ \beta \sigma_1 \sigma_2 + h_1 \sigma_1 + h_2 \sigma_2 \},$$

where h_1 and h_2 are i.i.d. copies of h^* . Then the thermodynamic limit of the internal energy per vertex is given by

$$(5.3.14) \quad U(\beta, B) = -\frac{\mathbb{E}[D]}{2} \mathbb{E} [\langle \sigma_1 \sigma_2 \rangle_{\nu'_2}],$$

where the expectation is taken over h_1 and h_2 . More explicitly,

$$(5.3.15) \quad U(\beta, B) = -\frac{\mathbb{E}[D]}{2} \mathbb{E} \left[\frac{\tanh(\beta) + \tanh(h_1) \tanh(h_2)}{1 + \tanh(\beta) \tanh(h_1) \tanh(h_2)} \right].$$

Note that the magnetization and internal energy are *local* observables, i.e., they are *spin* or *edge* variables averaged out over the graph. This is not true for the susceptibility, which is an average over pairs of spins, and hence it is much harder to give an explicit expression for this quantity.

5.3.1. Overview of the proof of Theorems 5.9. In this section, we give an overview of the proof of Theorem 5.9.

We will first analyze the case where $B > 0$ and deal with $B \leq 0$ later. We know that the configuration model is locally homogeneously tree-like (recall Theorem 2.11). The analysis for the Curie-Weiss model is based on an explicit characterization of the pressure per particle as an integral as in (5.2.12). For the Ising model on general graphs, such a nice rewrite is not possible, and the proof follows instead by taking the derivative w.r.t. β and proving convergence of the arising derivative. Let us sketch this argument. By the fundamental theorem of calculus,

$$(5.3.16) \quad \lim_{n \rightarrow \infty} \psi_n(\beta, B) = \lim_{n \rightarrow \infty} \left[\psi_n(0, B) + \int_0^\beta \frac{\partial}{\partial \beta'} \psi_n(\beta', B) d\beta' \right].$$

For all $n \geq 1$, we can compute

$$(5.3.17) \quad \psi_n(0, B) = \log(2 \cosh(B)) = \varphi(0, B),$$

so this is also true for $n \rightarrow \infty$. Further,

$$(5.3.18) \quad \frac{\partial}{\partial \beta} \psi_n(\beta, B) = \frac{1}{n} \sum_{(i,j) \in E_n} \langle \sigma_i \sigma_j \rangle_{\mu_n} = \frac{\mathbb{E}[D_n]}{2} \mathbb{E} \left[\langle \sigma_{I_n} \sigma_{J_n} \rangle_{\mu_n} \right],$$

where (I_n, J_n) is a *uniformly chosen* edge from the set of edges E_n and the expectation is w.r.t. the law of (I_n, J_n) . For an edge (i, j) , let $B_{(i,j)}(t)$ be the graph distance ball of radius t of the vertices i, j and $\partial B_{(i,j)}(t)$ its boundary. By the GKS inequality in Lemma 5.4, we can bound, for every $t \geq 0$,

$$(5.3.19) \quad \langle \sigma_{I_n} \sigma_{J_n} \rangle_{B_{(I_n, J_n)}(t)}^f \leq \langle \sigma_{I_n} \sigma_{J_n} \rangle_{\mu_n} \leq \langle \sigma_{I_n} \sigma_{J_n} \rangle_{B_{(I_n, J_n)}(t)}^+,$$

where $\langle \sigma_i \sigma_j \rangle_{B_{(i,j)}(t)}^{+/-f}$ is the correlation in the Ising model on $B_{(i,j)}(t)$ with $+/-$ free boundary conditions on $\partial B_{(i,j)}(t)$. By local weak convergence in probability, $B_{(I_n, J_n)}(t)$ converges in distribution to two independent copies of a branching process with offspring distribution Δ that are joined by a single edge at the root up to generation t . We denote this branching process by $\text{BP}_{(I,J)}(t)$. This implies that the upper and lower bounds in (5.3.19) converge to the respective quantities on a branching process tree, i.e.,

$$(5.3.20) \quad \begin{aligned} \mathbb{E}_n \left[\langle \sigma_{I_n} \sigma_{J_n} \rangle_{B_{(I_n, J_n)}(t)}^f \right] &\xrightarrow{\mathbb{P}} \mathbb{E} \left[\langle \sigma_I \sigma_J \rangle_{\text{BP}_{(I,J)}(t)}^f \right], \\ \mathbb{E}_n \left[\langle \sigma_{I_n} \sigma_{J_n} \rangle_{B_{(I_n, J_n)}(t)}^+ \right] &\xrightarrow{\mathbb{P}} \mathbb{E} \left[\langle \sigma_I \sigma_J \rangle_{\text{BP}_{(I,J)}(t)}^+ \right]. \end{aligned}$$

Here, the expectation \mathbb{E}_n on the left-hand sides of (5.3.20) are with respect to the *uniform edge* only, while the expectation \mathbb{E} on the right hand sides is with respect to the random tree. In particular, the left hand sides are *random variables*, as they still depend on the random graph, while the right hand sides are deterministic.

We have been a little sloppy here, as the local weak convergence result in Theorem 2.11 does not necessarily hold for the *two* trees that are attached to a random edge. We leave this extension of Theorem 2.11 as an exercise:

EXERCISE 5.22 (Local weak convergence for uniform edges). *Extend the local weak convergence in probability in Theorem 2.11 to local weak convergence in probability of the neighborhood of a uniformly chosen edge. It helps to realize that drawing a random edge is equivalent to drawing a vertex in a size-biased manner, and then taking a random edge incident to it.*

We next investigate Ising models on trees, the key example that we have in mind being a branching process tree. The computation of the Ising measure on a tree can be simplified to the computation of Ising measures on subtrees, by a *pruning argument*, as will prove to be extremely useful for the analysis of the Ising model on the configuration model:

LEMMA 5.12 (Pruning trees [92, Lemma 4.1]). *For U a subtree of a finite tree \mathcal{T} , let ∂U be the subset of vertices of U that connect to a vertex in $W \equiv \mathcal{T} \setminus U$. Denote by $\langle \sigma_u \rangle_{\mu_{W,u}}$ the magnetization of vertex $u \in \partial U$ of the Ising model on $W \cup \{u\}$. Then, the marginal Ising measure on U , $\mu_U^\mathcal{T}$, is the same as the Ising measure on U with magnetic fields*

$$(5.3.21) \quad B'_u = \begin{cases} \operatorname{atanh}(\langle \sigma_u \rangle_{\mu_{W,u}}), & u \in \partial U, \\ B_u, & u \in U \setminus \partial U. \end{cases}$$

The proof of this lemma follows from a direct application of the Boltzmann distribution given in (5.1.1), see also Dembo and Montanari [92, Lemma 4.1]. We leave a special case as an exercise:

EXERCISE 5.23 (Pruning trees). *Prove Lemma 5.12 in the case where $\mathcal{T} \setminus U$ consists of a single vertex.*

By Lemma 5.12, we can compute the expectation of the Ising model on a branching process tree from the leaves backwards. By the fact that the offspring distribution has law Δ , this gives rise to the recursion in (5.3.4). It turns out that (5.3.4) has a *unique* distributional solution when $B > 0$:

PROPOSITION 5.13 (Tree recursion). *Let $B > 0$ and let $(\Delta_t)_{t \geq 1}$ be i.i.d. according to some distribution. Consider the sequence of random variables $(h^{(t)})_{t \geq 0}$ defined by $h^{(0)} \equiv B$ and, for $t \geq 0$, by (5.3.4). Then, the distributions of $h^{(t)}$ are stochastically monotone and $h^{(t)} \xrightarrow{d} h^*$, where h^* is the unique fixed point h^* of the recursion (5.3.4) that is supported on $[0, \infty)$. Consequently,*

$$(5.3.22) \quad \frac{\partial}{\partial \beta} \psi_n(\beta, B) \xrightarrow{\mathbb{P}} \frac{\mathbb{E}[D]}{2} \mathbb{E}[\langle \sigma_1 \sigma_2 \rangle_{\nu'_2}],$$

where ν'_2 is defined in (5.3.13).

Note that Proposition 5.13 in fact proves (5.3.13) in Corollary 5.11. While Proposition 5.13 is crucial to show that the internal energy converges, it is not so convenient to use inside the intergral in (5.3.16), as it is unclear how to integrate the right hand side of

(5.3.22). In the proposition below, we identify the quantity in (5.3.22) as $\frac{\partial}{\partial\beta}\varphi(\beta, B)$, which is obviously more convenient:

PROPOSITION 5.14 (Identification of the limit of the internal energy). *Consider $\text{CM}_n(\mathbf{d})$ where the degree distribution satisfies Conditions 1.6(a)-(b), and has a strongly finite mean as in (5.3.1). Let $\beta > 0$. Then,*

$$(5.3.23) \quad \frac{\mathbb{E}[D]}{2} \mathbb{E} [\langle \sigma_1 \sigma_2 \rangle_{\nu'_2}] = \frac{\partial}{\partial\beta} \varphi(\beta, B),$$

where $\varphi(\beta, B)$ is given in (5.3.3).

Now we are ready to complete the proof of Theorem 5.9:

Proof of Theorem 5.9. By the fundamental theorem of calculus,

$$(5.3.24) \quad \begin{aligned} \psi_n(\beta, B) &= \left[\psi_n(0, B) + \int_0^\beta \frac{\partial}{\partial\beta'} \psi_n(\beta', B) d\beta' \right] \\ &= \left[\psi_n(0, B) + \int_0^\varepsilon \frac{\partial}{\partial\beta'} \psi_n(\beta', B) d\beta' + \int_\varepsilon^\beta \frac{\partial}{\partial\beta'} \psi_n(\beta', B) d\beta' \right], \end{aligned}$$

for any $0 < \varepsilon < \beta$. For all $n \geq 1$,

$$(5.3.25) \quad \psi_n(0, B) = \log(2 \cosh(B)) = \varphi(0, B),$$

so this is also true for $n \rightarrow \infty$.

By the fact that the number of edges in $\text{CM}_n(\mathbf{d})$ is $\ell_n/2 = O(n)$,

$$(5.3.26) \quad \left| \frac{\partial}{\partial\beta} \psi_n(\beta, B) \right| = \left| \frac{1}{n} \sum_{(i,j) \in E_n} \langle \sigma_i \sigma_j \rangle_{\mu_n} \right| \leq \frac{|E_n|}{n} \leq c,$$

for some constant c . Thus, uniformly in n ,

$$(5.3.27) \quad \left| \int_0^\varepsilon \frac{\partial}{\partial\beta'} \psi_n(\beta', B) d\beta' \right| \leq c\varepsilon.$$

Using the boundedness of the derivative for $\beta' \in [\varepsilon, \beta]$, together with Propositions 5.13–5.14 and bounded convergence,

$$(5.3.28) \quad \int_\varepsilon^\beta \frac{\partial}{\partial\beta'} \psi_n(\beta', B) d\beta' \xrightarrow{\mathbb{P}} \int_\varepsilon^\beta \frac{\partial}{\partial\beta'} \varphi(\beta', B) d\beta' = \varphi(\beta, B) - \varphi(\varepsilon, B),$$

again by the fundamental theorem of calculus.

Observing that $0 \leq \tanh(h^*) \leq 1$, one can show that, by dominated convergence, $\varphi(\beta, B)$ is right-continuous in $\beta = 0$. Thus, letting $\varepsilon \downarrow 0$,

$$(5.3.29) \quad \begin{aligned} \psi_n(\beta, B) &\xrightarrow{\mathbb{P}} \lim_{\varepsilon \downarrow 0} \left[\varphi(0, B) + \int_0^\varepsilon \frac{\partial}{\partial\beta'} \varphi(\beta', B) d\beta' + \int_\varepsilon^\beta \frac{\partial}{\partial\beta'} \varphi(\beta', B) d\beta' \right] \\ &= \varphi(0, B) + \lim_{\varepsilon \downarrow 0} (\varphi(\beta, B) - \varphi(\varepsilon, B)) = \varphi(\beta, B), \end{aligned}$$

which completes the proof for $B > 0$.

The Ising model with $B < 0$ is equivalent to the case $B > 0$, because one can multiply all spin variables $(\sigma_i)_{i \in [n]}$ and B by -1 without changing Boltzmann distribution (5.1.1). Furthermore, note that,

$$(5.3.30) \quad \left| \frac{\partial}{\partial B} \psi_n(\beta, B) \right| = \left| \frac{1}{n} \sum_{i \in [n]} \langle \sigma_i \rangle_{\mu_n} \right| \leq 1,$$

so that $B \mapsto \psi_n(\beta, B)$ is uniformly Lipschitz continuous with Lipschitz constant one. Therefore,

$$(5.3.31) \quad \psi_n(\beta, 0) = \lim_{B \downarrow 0} \psi_n(\beta, B) \xrightarrow{\mathbb{P}} \lim_{B \downarrow 0} \varphi(\beta, B).$$

□

The remainder of this section is organized as follows. We study the tree recursion of (5.3.4) in Section 5.3.2, and conclude Proposition 5.13 from it. We prove Proposition 5.14 in Section 5.3.3. In Section 5.3.4, we study the thermodynamic quantities and prove Corollary 5.11.

5.3.2. Tree recursion: proof of Propositions 5.13. To prove Proposition 5.13, we will first study the Ising model on a tree with ℓ generations, $\mathcal{T}(\ell)$, with either $+$ or free boundary conditions, where the Ising models on the tree $\mathcal{T}(\ell)$ with $+$ /free boundary conditions are defined by the Boltzmann distributions

$$(5.3.32) \quad \mu^{\ell,+}(\sigma) = \frac{1}{Z^{\ell,+}(\beta, \underline{B})} \exp \left\{ \beta \sum_{(i,j) \in \mathcal{T}(\ell)} \sigma_i \sigma_j + \sum_{i \in \mathcal{T}(\ell)} B_i \sigma_i \right\} \mathbb{1}_{\{\sigma_i = +1, \text{ for all } i \in \partial \mathcal{T}(\ell)\}},$$

and

$$(5.3.33) \quad \mu^{\ell,f}(\sigma) = \frac{1}{Z^{\ell,f}(\beta, \underline{B})} \exp \left\{ \beta \sum_{(i,j) \in \mathcal{T}(\ell)} \sigma_i \sigma_j + \sum_{i \in \mathcal{T}(\ell)} B_i \sigma_i \right\},$$

respectively, where $Z^{\ell,+/f}$ are the proper normalization factors and $\partial \mathcal{T}(\ell)$ denotes the vertices in the ℓ th generation of $\mathcal{T}(\ell)$.

The proof is organized as follows. In the next lemma we show that the effect of these boundary conditions vanishes when $\ell \rightarrow \infty$. We then show that the recursion (5.3.4) has a fixed point and use a coupling with the root magnetization in trees and Lemma 5.15 to show that this fixed point does not depend on the initial distribution $h^{(0)}$, thus showing that (5.3.4) has a *unique* fixed point.

We start by studying the effect of the boundary conditions:

LEMMA 5.15 (Vanishing effect of boundary conditions). *Let $m^{\ell,+/f}(\underline{B})$ denote the root magnetization given $\mathcal{T}(\ell)$ with external field per vertex $B_i \geq B_{\min} > 0$ when the tree has $+$ /free boundary conditions. Let $0 \leq \beta \leq \beta_{\max} < \infty$. Then, there exists an $K = K(\beta_{\max}, B_{\min}) < \infty$ such that, a.s.,*

$$(5.3.34) \quad m^{\ell,+}(\underline{B}) - m^{\ell,f}(\underline{B}) \leq \frac{K}{\ell},$$

for all $\ell \geq 1$.

Note that Lemma 5.15 is extremely general. For example, it also applies to trees arising from multitype branching processes, or even inhomogeneous trees.

PROOF. The lemma clearly holds for $\beta = 0$, so we assume that $\beta > 0$ in the remainder of the proof.

Recall that we can label the vertices of the tree $\mathcal{T}(\ell)$ by the root \emptyset and $w = w_1 \cdots w_k$ for some $k \in [\ell]$. Denote by $m^\ell(\underline{B}, \underline{H})$ the root magnetization given $\mathcal{T}(\ell)$ with free boundary conditions, when the external field on the vertices $w \in \partial\mathcal{T}(\ell)$ is $B_w + H_w$ and B_w on all other vertices $w \in \mathcal{T}(\ell - 1)$. We use Lemma 5.12 on the (finite) tree $\mathcal{T}(\ell)$. Thus, for $1 \leq k \leq \ell$,

$$(5.3.35) \quad m^{k,+}(\underline{B}) \equiv m^k(\underline{B}, \infty) = m^{k-1}(\underline{B}, (\beta d_w)_{w \in \partial\mathcal{T}(k-1)}),$$

where d_w is the forward degree of $w \in \partial\mathcal{T}(k-1)$. By the GKS inequality in Lemma 5.4,

$$(5.3.36) \quad m^{k-1}(\underline{B}, (\beta d_w)_{w \in \partial\mathcal{T}(k-1)}) \leq m^{k-1}(\underline{B}, \infty).$$

Since the magnetic field at all vertices in $\partial\mathcal{T}(k)$ is at least B_{\min} we can write, using Lemma 5.12 and the GKS inequality in Lemma 5.4, that

$$(5.3.37) \quad m^{k,f}(\underline{B}) \equiv m^k(\underline{B}, 0) \geq m^{k-1}(\underline{B}, \xi \cdot (d_w)_{w \in \partial\mathcal{T}(k-1)}),$$

where

$$(5.3.38) \quad \xi = \xi(\beta, B_{\min}) = \operatorname{atanh}(\tanh(\beta) \tanh(B_{\min})).$$

This inequality holds with equality when $B_w = B_{\min}$ for all $w \in \partial\mathcal{T}(k)$. Using the GKS inequality in Lemma 5.4 again,

$$(5.3.39) \quad m^{k-1}(\underline{B}, \xi(d_w)_{w \in \partial\mathcal{T}(k-1)}) \geq m^{k-1}(\underline{B}, 0).$$

Note that $0 \leq \xi(\beta, B_{\min}) \leq \beta$. The function $H \mapsto m^k(\underline{B}, H \cdot (d_w)_{w \in \partial\mathcal{T}(k-1)})$ is concave in H because of the GHS inequality in Lemma 5.5:

EXERCISE 5.24 (GHS inequality and concavity). *Prove that the GHS inequality in Lemma 5.5 implies that $H \mapsto m^k(\underline{B}, H \cdot (d_w)_{w \in \partial\mathcal{T}(k-1)})$ is concave.*

By the concavity of $H \mapsto m^k(\underline{B}, H \cdot (d_w)_{w \in \partial\mathcal{T}(k-1)})$,

$$(5.3.40) \quad \begin{aligned} & m^{k-1}(\underline{B}, \beta(d_w)_{w \in \partial\mathcal{T}(k-1)}) - m^{k-1}(\underline{B}, 0) \\ & \leq K (m^{k-1}(\underline{B}, \xi \cdot (d_w)_{w \in \partial\mathcal{T}(k-1)}) - m^{k-1}(\underline{B}, 0)), \end{aligned}$$

where

$$(5.3.41) \quad K = K(\beta_{\max}, B_{\min}) = \sup_{0 < \beta \leq \beta_{\max}} \frac{\beta}{\xi(\beta, B_{\min})} < \infty.$$

Thus, we can rewrite $m^{k,+}(\underline{B})$ using (5.3.35) and bound $m^{k,f}(\underline{B})$ using (5.3.37) and (5.3.39), to obtain

$$(5.3.42) \quad m^{k,+}(\underline{B}) - m^{k,f}(\underline{B}) \leq m^{k-1}(\underline{B}, \beta(d_w)_{w \in \partial\mathcal{T}(k-1)}) - m^{k-1}(\underline{B}, 0).$$

By (5.3.40),

$$(5.3.43) \quad \begin{aligned} m^{k,+}(\underline{B}) - m^{k,f}(\underline{B}) & \leq K (m^{k-1}(\underline{B}, \xi \cdot (d_w)_{w \in \partial\mathcal{T}(k-1)}) - m^{k-1}(\underline{B}, 0)) \\ & \leq K (m^k(\underline{B}, 0) - m^{k-1}(\underline{B}, 0)), \end{aligned}$$

where we have used (5.3.37) in the last inequality.

By (5.3.35) and (5.3.36), $m^{k,+}(\underline{B}) = m^{k,+}(\underline{B}, 0)$ is non-increasing in k and, by (5.3.37) and (5.3.39), $m^{k,f}(\underline{B}) = m^{k,f}(\underline{B}, 0)$ is non-decreasing in k . Thus, by summing the inequality in (5.3.43) over k gives

$$(5.3.44) \quad \begin{aligned} \ell (m^{\ell,+}(\underline{B}) - m^{\ell,f}(\underline{B})) &\leq \sum_{k=1}^{\ell} (m^{k,+}(\underline{B}) - m^{k,f}(\underline{B})) \leq K \sum_{k=1}^{\ell} (m^k(\underline{B}, 0) - m^{k-1}(\underline{B}, 0)) \\ &= K (m^{\ell}(\underline{B}, 0) - m^0(\underline{B}, 0)) \leq K, \end{aligned}$$

since $0 \leq m^{\ell/0}(\underline{B}, 0) \leq 1$. □

We are now ready to prove Proposition 5.13:

PROOF OF PROPOSITION 5.13. We start by proving the uniqueness of the solution of the fixed point of (5.3.4). Condition on the branching process tree $\text{BP}_{(I,J)} = \mathcal{T}$. Then

$$(5.3.45) \quad h^{(t)} \equiv \text{atanh}(m^{t,f}(B))$$

satisfies the recursive distribution (5.3.4) because of Lemma 5.12. Since, by the GKS inequality in Lemma 5.4, $m^{t,f}(B)$, and hence also $h^{(t)}$, are monotonically increasing in t , we have that $B = h^{(0)} \leq h^{(t)} \leq B + D_0 < \infty$ for all $t \geq 0$, where D_0 is the degree of the root. So, $h^{(t)}$ converges to some limit \underline{h} . Since this holds a.s. for any branching process $\text{BP}_{(I,J)}$, the distribution of \underline{h} also exists and one can show that this limit is a fixed point of (5.3.4) (see Dembo and Montanari [92, Proof of Lemma 2.2]). In a similar way, $h^{(t,+)} \equiv \text{atanh}(m^{t,+}(B))$ satisfies (5.3.4) when starting with $h^{(0,+)} = \infty$. Then, $h^{(t,+)}$ is monotonically decreasing and, for $t \geq 1$, $B \leq h^{(t)} \leq B + D_0 < \infty$, so $h^{(t,+)}$ also converges to some limit \bar{h} .

Let h be a fixed point of (5.3.4), condition on this h and let $h^{(0,*)} = h$. Then $h^{(t,*)}$ converges as above to a limit h^* say, when applying (5.3.4). Note that $h^{(0)} \leq h^{(0,*)} \leq h^{(0,+)}$. Coupling so as to have the same degrees $(\Delta_t)_{t \geq 1}$ while applying the recursion (5.3.4), this order is preserved by the GKS inequality, so that $h^{(t)} \leq h^{(t,*)} \leq h^{(t,+)}$ for all $t \geq 0$. By Lemma 5.15,

$$(5.3.46) \quad |\tanh(h^{(t)}) - \tanh(h^{(t,+)})| = |m^{t,f}(B) - m^{t,+}(B)| \rightarrow 0, \quad \text{for } t \rightarrow \infty.$$

Since the above holds a.s. for any tree \mathcal{T} and any realization of h^* , the distributions of \underline{h}, \bar{h} and h^* are equal, and, since h is a fixed point of (5.3.4), are all equal in distribution to h . This proves the uniqueness statement in Proposition 5.13.

We continue with (5.3.22). Recall (5.3.18)–(5.3.20). By Lemma 5.12 and the uniqueness of the fixed point of (5.3.4) (as we have already proved),

$$(5.3.47) \quad \lim_{t \rightarrow \infty} \mathbb{E} \left[\langle \sigma_I \sigma_J \rangle_{\text{BP}_{(I,J)}(t)}^{+/f} \right] = \mathbb{E} \left[\langle \sigma_1 \sigma_2 \rangle_{\nu'_2} \right],$$

thus also proving (5.3.22). This completes the proof of Proposition 5.13. □

5.3.3. Identification of the internal energy: proof of Proposition 5.14. For $\beta > 0$, we will show that the partial derivative with respect to β of $\psi_n(\beta, B)$ converges to the partial derivative with respect to β of $\varphi(\beta, B)$. For this, we need that we can in fact ignore the dependence of h^* on β when computing the latter derivative as we shall show first:

PROPOSITION 5.16 (Dependence of φ on (β, B) via h^*). *Assume that D has strongly finite mean for some $\tau \in (2, 3)$. Fix $B_1, B_2 > 0$ and $0 < \beta_1, \beta_2 < \infty$. Let h_1^* and h_2^* be the fixed points of (5.3.4) for (β_1, B_1) and (β_2, B_2) , respectively. Let $\varphi_{h^*}(\beta, B)$ be defined as in (5.3.3) with $(h_i)_{i \geq 1}$ replaced by i.i.d. copies of the specified h^* . Then,*

(a) *For $B_1 = B_2$, there exists a $\lambda_1 < \infty$ such that*

$$(5.3.48) \quad |\varphi_{h_1^*}(\beta_1, B_1) - \varphi_{h_2^*}(\beta_1, B_1)| \leq \lambda_1 |\beta_1 - \beta_2|^{\tau-1}.$$

(b) *For $\beta_1 = \beta_2$, there exists a $\lambda_2 < \infty$ such that*

$$(5.3.49) \quad |\varphi_{h_1^*}(\beta_1, B_1) - \varphi_{h_2^*}(\beta_1, B_1)| \leq \lambda_2 |B_1 - B_2|^{\tau-1}.$$

We need part (b) of the proposition above later in the proof of Corollary 5.11. Before proving Proposition 5.16, we first use it to complete the proof of Proposition 5.14:

Tree analysis: proof of Proposition 5.14. This proof requires a bit of analysis. Let us start by defining some notation. Let $(X_i)_{i \geq 1}$ be i.i.d. copies of $\tanh(h^*)$, also independent of D . Let $\hat{\beta} = \tanh(\beta)$ and, for $\ell \geq 2$, define

$$(5.3.50) \quad H_\ell(x_1, \dots, x_\ell) = \log \left\{ e^B \prod_{i=1}^{\ell} (1 + \hat{\beta}x_i) + e^{-B} \prod_{i=1}^{\ell} (1 - \hat{\beta}x_i) \right\} - \frac{1}{\ell-1} \sum_{1 \leq i < j \leq \ell} \log(1 + \hat{\beta}x_i x_j),$$

and

$$(5.3.51) \quad H_1(x_1, x_2) = \frac{1}{2} \left(\log(e^B(1 + \hat{\beta}x_1) + e^{-B}(1 - \hat{\beta}x_1)) + \log(e^B(1 + \hat{\beta}x_2) + e^{-B}(1 - \hat{\beta}x_2)) - \log(1 + \hat{\beta}x_1 x_2) \right).$$

Then, inspection of (5.3.3) reveals that

$$(5.3.52) \quad \varphi(\beta, B) = \frac{\mathbb{E}[D]}{2} \log \cosh(\beta) + \mathbb{E}[H_D(X_1, \dots, X_{\max\{2, D\}})],$$

as you may verify in the following exercise:

EXERCISE 5.25 (Limiting pressure per particle). *Verify that $\varphi(\beta, B)$ in (5.3.3) can be rewritten as (5.3.52).*

Differentiating (5.3.50) gives, for $\ell \geq 2$,

$$(5.3.53) \quad \frac{\partial}{\partial x_1} H_\ell(x_1, \dots, x_\ell) = \psi(x_1, g_\ell(x_2, \dots, x_\ell)) - \frac{1}{\ell-1} \sum_{j=2}^{\ell} \psi(x_1, x_j),$$

where $\psi(x, y) = \hat{\beta}y/(1 + \hat{\beta}xy)$ and

$$(5.3.54) \quad g_\ell(x_2, \dots, x_\ell) = \tanh \left(B + \sum_{j=2}^{\ell} \operatorname{atanh}(\hat{\beta}x_j) \right),$$

while differentiating (5.3.51) gives

$$(5.3.55) \quad \frac{\partial}{\partial x_1} H_1(x_1, x_2) = \frac{1}{2} [\psi(x_1, g_1) - \psi(x_1, x_2)],$$

where we extend (5.3.54) to $\ell = 1$ by letting $g_1 = \tanh(B)$:

EXERCISE 5.26 (Partial derivatives). *Verify that (5.3.53) and (5.3.55) indeed hold. [Hint: This is a quite nasty computation. You are advised to check the computations around (5.3.108) if you do not manage here.]*

Using that $\ell\mathbb{P}(D = \ell) = \mathbb{E}[D]\mathbb{P}(\Delta = \ell - 1)$, for all x ,

$$(5.3.56) \quad \begin{aligned} \mathbb{E}[D\psi(x, g_D(X_2, \dots, X_D))] &= \mathbb{E}[D]\mathbb{E}[\psi(x, g_{\Delta+1}(X_2, \dots, X_{\Delta+1}))] \\ &= \mathbb{E}[D]\mathbb{E}[\psi(x, X_2)], \end{aligned}$$

because $g_{\Delta+1}(X_2, \dots, X_{\Delta+1})$ is a fixed point of (5.3.4), so that $g_{\Delta+1}(X_2, \dots, X_{\Delta+1}) \stackrel{d}{=} X_2$ and is independent of X_1 . We conclude that, for all x ,

$$(5.3.57) \quad \begin{aligned} \mathbb{E}\left[\max\{2, D\} \frac{\partial}{\partial x_1} H_D(x, X_2, \dots, X_{\max\{2, D\}})\right] \\ = \sum_{\ell \geq 2} \ell p_\ell \mathbb{E}\left[\psi(x, g_\ell(X_2, \dots, X_\ell)) - \frac{1}{\ell-1} \sum_{j=2}^{\ell} \psi(x, X_j)\right] + p_1 \mathbb{E}[\psi(x, g_1) - \psi(x, X_2)] \\ = \sum_{\ell \geq 2} \ell p_\ell \mathbb{E}\left[\psi(x, g_\ell(X_2, \dots, X_\ell)) - \psi(x, X_2)\right] + p_1 \mathbb{E}[\psi(x, g_1) - \psi(x, X_2)] = 0. \end{aligned}$$

We further compute that

$$(5.3.58) \quad \frac{\partial H_\ell}{\partial \beta}(x_1, \dots, x_\ell) = (1 - \widehat{\beta}^2) \sum_{i=1}^{\ell} x_i \psi(x_i, g_\ell(\vec{x}_\ell^i)) - \frac{(1 - \widehat{\beta}^2)}{\ell - 1} \sum_{1 \leq i < j \leq \ell} \frac{x_i x_j}{1 + \widehat{\beta} x_i x_j},$$

where \vec{x}_ℓ^i is the vector of length $\ell - 1$ where x_i is removed, and

$$(5.3.59) \quad \frac{\partial H_1}{\partial \beta}(x_1, x_2) = \frac{1}{2}(1 - \widehat{\beta}^2) \left(x_1 \psi(x_1, g_1) + x_2 \psi(x_2, g_1) - \frac{x_1 x_2}{1 + \widehat{\beta} x_1 x_2} \right).$$

From Proposition 5.16 it follows that we can assume that β is fixed in h^* when differentiating $\varphi(\beta, B)$ with respect to β . Thus, taking the derivative of (5.3.52), using that each X_i gives one contribution that are equal by symmetry, and using (5.3.56), we obtain that

$$(5.3.60) \quad \begin{aligned} \frac{\partial}{\partial \beta} \varphi(\beta, B) &= \frac{\mathbb{E}[D]}{2} \widehat{\beta} + \mathbb{E}\left[\max\{2, D\} \frac{\partial H_D}{\partial x_1}(X_1, \dots, X_{\max\{2, D\}})\right] + \mathbb{E}\left[\frac{\partial H_D}{\partial \beta}(X_1, \dots, X_{\max\{2, D\}})\right] \\ &= \frac{\mathbb{E}[D]}{2} \widehat{\beta} + \mathbb{E}\left[\frac{\partial H_D}{\partial \beta}(X_1, \dots, X_{\max\{2, D\}})\right] \\ &= \frac{\mathbb{E}[D]}{2} \widehat{\beta} - (1 - \widehat{\beta}^2) \frac{\mathbb{E}[D]}{2} \mathbb{E}[\psi(X_1, X_2)] + (1 - \widehat{\beta}^2) \mathbb{E}[DX_1 \psi(X_1, g_\ell(X_2, \dots, X_\ell))]. \end{aligned}$$

Again by (5.3.57),

$$(5.3.61) \quad \mathbb{E}[DX_1 \psi(X_1, g_\ell(X_2, \dots, X_\ell))] = \mathbb{E}[D]\mathbb{E}[X_1 \psi(X_1, X_2)],$$

so that we arrive at

$$\begin{aligned}
\frac{\partial}{\partial \beta} \varphi(\beta, B) &= \frac{\mathbb{E}[D]}{2} \widehat{\beta} + \frac{\mathbb{E}[D]}{2} (1 - \widehat{\beta}^2) \mathbb{E}[X_1 \psi(X_1, X_2)] \\
&= \frac{\mathbb{E}[D]}{2} \mathbb{E} \left[\widehat{\beta} (1 + \widehat{\beta} X_1 X_2) + (1 - \widehat{\beta}^2) \frac{X_1 X_2}{1 + \widehat{\beta} X_1 X_2} \right] \\
(5.3.62) \quad &= \frac{\mathbb{E}[D]}{2} \mathbb{E} \left[\frac{\widehat{\beta} + X_1 X_2}{1 + \widehat{\beta} X_1 X_2} \right].
\end{aligned}$$

Since, with h_1, h_2 i.i.d. copies of h^* ,

$$\begin{aligned}
\mathbb{E} \left[\frac{\widehat{\beta} + X_1 X_2}{1 + \widehat{\beta} X_1 X_2} \right] &= \mathbb{E} \left[\frac{\tanh(\beta) + \tanh(h_1) \tanh(h_2)}{1 + \tanh(\beta) \tanh(h_1) \tanh(h_2)} \right] \\
(5.3.63) \quad &= \mathbb{E} \left[\frac{e^{\beta+h_1+h_2} - e^{-\beta-h_1+h_2} - e^{-\beta+h_1-h_2} + e^{\beta-h_1-h_2}}{e^{\beta+h_1+h_2} + e^{-\beta-h_1+h_2} + e^{-\beta+h_1-h_2} + e^{\beta-h_1-h_2}} \right] = \mathbb{E} [\langle \sigma_1 \sigma_2 \rangle_{\nu'_2}],
\end{aligned}$$

where the measure ν'_2 is defined in (5.3.13), we have proved the proposition. \square

We will now prove Proposition 5.16 by first bounding the dependence of φ on h^* in Lemma 5.17 and subsequently bounding the dependence of h^* on β and B in Lemmas 5.18 and 5.19 respectively.

LEMMA 5.17 (Dependence of φ on h^*). *Assume that D has strongly finite mean for some $\tau \in (2, 3)$. Fix $B_1, B_2 > 0$ and $0 < \beta_1, \beta_2 < \infty$. Let h_1^* and h_2^* be the fixed points of (5.3.4) for (β_1, B_1) and (β_2, B_2) , respectively. Let $\varphi_{h^*}(\beta, B)$ be defined as in (5.3.3) with $(h_i)_{i \geq 1}$ replaced by i.i.d. copies of the specified h^* . Then, for some $\lambda < \infty$,*

$$(5.3.64) \quad |\varphi_{h_1^*}(\beta_1, B_1) - \varphi_{h_2^*}(\beta_1, B_1)| \leq \lambda \|\tanh(h_1^*) - \tanh(h_2^*)\|_{\text{MK}}^{\tau-1},$$

where $\|X - Y\|_{\text{MK}}$ denotes the Monge-Kantorovich-Wasserstein distance between random variables X and Y , i.e., $\|X - Y\|_{\text{MK}}$ is the infimum of $\mathbb{E}[|\hat{X} - \hat{Y}|]$ over all couplings (\hat{X}, \hat{Y}) of X and Y .

PROOF. Let X_i and Y_i be i.i.d. copies of $X = \tanh(h_1^*)$ and $Y = \tanh(h_2^*)$ respectively and also independent of D . When $\|X - Y\|_{\text{MK}} = 0$ or $\|X - Y\|_{\text{MK}} = \infty$, the statement in the lemma clearly holds. Thus, without loss of generality, we fix $\gamma > 1$ and assume that (X_i, Y_i) are i.i.d. pairs, independent of D , that are coupled in such a way that $\mathbb{E}|X_i - Y_i| \leq \gamma \|X - Y\|_{\text{MK}} < \infty$.

Recall H_ℓ in (5.3.50). Then,

$$\begin{aligned}
(5.3.65) \quad \varphi_{h_1^*}(\beta_1, B_1) &= \frac{\mathbb{E}[D]}{2} \log \cosh(\beta) + \mathbb{E}[H_D(X_1, \dots, X_{\max\{2, D\}})], \\
\varphi_{h_2^*}(\beta_1, B_1) &= \frac{\mathbb{E}[D]}{2} \log \cosh(\beta) + \mathbb{E}[H_D(Y_1, \dots, Y_{\max\{2, D\}})].
\end{aligned}$$

In the remainder of the proof we will assume that H_1 is defined as in (5.3.50). The proof, however, also works for H_1 as defined in (5.3.51).

We will split the absolute difference between $\varphi_{h_1^*}(\beta_1, B_1)$ and $\varphi_{h_2^*}(\beta_1, B_1)$ into two parts depending on whether D is small or large, i.e., for some constant $\theta > 0$ to be chosen later

on, we split

$$(5.3.66) \quad \left| \mathbb{E}[H_D(Y_1, \dots, Y_D) - H_D(X_1, \dots, X_D)] \right| \leq \left| \mathbb{E}[(H_D(Y_1, \dots, Y_D) - H_D(X_1, \dots, X_D)) \mathbb{1}_{\{D \leq \theta\}}] \right| + \left| \mathbb{E}[(H_D(Y_1, \dots, Y_D) - H_D(X_1, \dots, X_D)) \mathbb{1}_{\{D > \theta\}}] \right|.$$

Note that

$$(5.3.67) \quad \begin{aligned} H_\ell(Y_1, \dots, Y_\ell) - H_\ell(X_1, \dots, X_\ell) &= \int_0^1 \frac{d}{ds} H_\ell(sY_1 + (1-s)X_1, \dots, sY_\ell + (1-s)X_\ell) \Big|_{s=t} dt \\ &= \int_0^1 \sum_{i=1}^{\ell} (Y_i - X_i) \frac{\partial H_\ell}{\partial x_i}(tY_1 + (1-t)X_1, \dots, tY_\ell + (1-t)X_\ell) dt \\ &= \sum_{i=1}^{\ell} (Y_i - X_i) \int_0^1 \frac{\partial H_\ell}{\partial x_i}(tY_1 + (1-t)X_1, \dots, tY_\ell + (1-t)X_\ell) dt. \end{aligned}$$

As can be observed by (5.3.53)–(5.3.55), $\frac{\partial H_\ell}{\partial x_i}$ is uniformly bounded, so that

$$(5.3.68) \quad \left| H_\ell(Y_1, \dots, Y_\ell) - H_\ell(X_1, \dots, X_\ell) \right| \leq \lambda_1 \sum_{i=1}^{\ell} |Y_i - X_i|,$$

where λ_1 is allowed to change from line to line. Hence,

$$(5.3.69) \quad \left| \mathbb{E}[(H_D(Y_1, \dots, Y_D) - H_D(X_1, \dots, X_D)) \mathbb{1}_{\{D > \theta\}}] \right| \leq \mathbb{E} \left[\sum_{i=1}^D |Y_i - X_i| c_1 \mathbb{1}_{\{D > \theta\}} \right] \leq \lambda_1 \|X - Y\|_{\text{MK}} \mathbb{E}[D \mathbb{1}_{\{D > \theta\}}].$$

We use Lemma 2.38 with $a = 1$ to compute

$$(5.3.70) \quad \left| \mathbb{E}[(H_D(Y_1, \dots, Y_D) - H_D(X_1, \dots, X_D)) \mathbb{1}_{\{D > \theta\}}] \right| \leq \lambda_1 \|X - Y\|_{\text{MK}} \theta^{-(\tau-2)}.$$

By the fundamental theorem of calculus, we can also write

$$(5.3.71) \quad H_\ell(Y_1, \dots, Y_\ell) - H_\ell(X_1, \dots, X_\ell) = \sum_{i=1}^{\ell} \Delta_i H_\ell + \sum_{i \neq j}^{\ell} (Y_i - X_i)(Y_j - X_j) f_{ij}^{(\ell)},$$

with

$$(5.3.72) \quad \Delta_i H_\ell = (Y_i - X_i) \int_0^1 \frac{\partial H_\ell}{\partial x_i}(X_1, \dots, tY_i + (1-t)X_i, \dots, X_\ell) dt,$$

and

$$(5.3.73) \quad f_{ij}^{(\ell)} = \int_0^1 \int_0^t \frac{\partial^2 H_\ell}{\partial x_i \partial x_j}(sY_1 + (1-s)X_1, \dots, sY_i + (1-s)X_i, \dots, sY_\ell + (1-s)X_\ell) ds dt.$$

Therefore,

$$\begin{aligned}
& \left| \mathbb{E}[(H_D(Y_1, \dots, Y_D) - H_D(X_1, \dots, X_D)) \mathbb{1}_{\{D \leq \theta\}}] \right| \\
& \leq \left| \mathbb{E} \left[\sum_{i=1}^D \Delta_i H_D \mathbb{1}_{\{D \leq \theta\}} \right] \right| \\
(5.3.74) \quad & + \left| \mathbb{E} \left[\sum_{i \neq j}^D (Y_i - X_i)(Y_j - X_j) f_{ij}^{(D)} \mathbb{1}_{\{D \leq \theta\}} \right] \right|.
\end{aligned}$$

Since $\frac{\partial^2 H_\ell}{\partial x_i \partial x_j}$ is also uniformly bounded (this is proved by Dembo and Montanari in [92, Corollary 6.3], but can also be directly observed by differentiating (5.3.53) and (5.3.55) once more), we obtain

$$\begin{aligned}
& \left| \mathbb{E} \left[\sum_{i \neq j}^D (Y_i - X_i)(Y_j - X_j) f_{ij}^{(L)} \mathbb{1}_{\{D \leq \theta\}} \right] \right| \leq \lambda_2 \mathbb{E} \left[\sum_{i \neq j}^D |Y_i - X_i| |Y_j - X_j| \mathbb{1}_{\{D \leq \theta\}} \right] \\
(5.3.75) \quad & \leq \lambda_2 \|X - Y\|_{\text{MK}}^2 \mathbb{E}[D^2 \mathbb{1}_{\{D \leq \theta\}}] \\
& \leq \lambda_2 \|X - Y\|_{\text{MK}}^2 \theta^{-(\tau-3)},
\end{aligned}$$

where λ_2 is allowed to change from line to line, and we use Lemma 2.39 with $a = 2$. We split

$$(5.3.76) \quad \left| \mathbb{E} \left[\sum_{i=1}^D \Delta_i H_D \mathbb{1}_{\{D \leq \theta\}} \right] \right| \leq \left| \mathbb{E} \left[\sum_{i=1}^D \Delta_i H_D \right] \right| + \left| \mathbb{E} \left[\sum_{i=1}^D \Delta_i H_D \mathbb{1}_{\{D > \theta\}} \right] \right|.$$

By symmetry of the functions H_ℓ with respect to their arguments, for i.i.d. (X_i, Y_i) independent of D ,

$$\begin{aligned}
& \mathbb{E} \left[\sum_{i=1}^D \Delta_i H_D \right] = \mathbb{E} [D \Delta_1 H_D] \\
(5.3.77) \quad & = \mathbb{E} \left[D(Y_1 - X_1) \int_0^1 \frac{\partial H_D}{\partial x_1}(tY_1 + (1-t)X_1, X_2, \dots, X_D) dt \right].
\end{aligned}$$

By (5.3.57),

$$(5.3.78) \quad \mathbb{E} \left[D \frac{\partial H_D}{\partial x_1}(x, X_2, \dots, X_{\max\{2, D\}}) \right] = 0, \quad \text{for all } x \in [-1, 1].$$

Since $\frac{\partial H_D}{\partial x_1}$ is uniformly bounded, $L \frac{\partial H_D}{\partial x_1}$ is integrable, so that, by Fubini's theorem and (5.3.78),

$$(5.3.79) \quad \mathbb{E} \left[\sum_{i=1}^D \Delta_i H_D \right] = \mathbb{E} \left[(Y_1 - X_1) \int_0^1 \mathbb{E} \left[D \frac{\partial H_D}{\partial x_1}(tY_1 + (1-t)X_1, X_2, \dots, X_D) \middle| X_1, Y_1 \right] dt \right] = 0.$$

Furthermore, by (5.3.72) and the uniform boundedness of $\frac{\partial H_\ell}{\partial x_i}$, as well as Lemma 2.39 with $a = 1$,

$$(5.3.80) \quad \left| \mathbb{E} \left[\sum_{i=1}^D \Delta_i H_D \mathbb{1}_{\{D > \theta\}} \right] \right| \leq \mathbb{E} \left[\sum_{i=1}^D |Y_i - X_i| c_1 \mathbb{1}_{\{D > \theta\}} \right] \leq \lambda_1 \|X - Y\|_{\text{MK}} \theta^{-(\tau-2)}.$$

Therefore,

$$(5.3.81) \quad \left| \mathbb{E} \left[(H_D(Y_1, \dots, Y_D) - H_D(X_1, \dots, X_D)) \mathbb{1}_{\{D \leq \theta\}} \right] \right| \\ \leq \lambda_1 \|X - Y\|_{\text{MK}} \theta^{-(\tau-2)} + \lambda_2 \|X - Y\|_{\text{MK}}^2 \theta^{-(\tau-3)}.$$

Combining (5.3.70) and (5.3.81) and letting $\theta = \|X - Y\|_{\text{MK}}^{-1}$ yields the desired result. \square

In the following two lemmas, we investigate the dependence of $h^* = h^*(\beta, B)$ on β and B , respectively. We start with its dependence on β :

LEMMA 5.18 (Dependence of h^* on β). *Fix $B > 0$ and $0 < \beta_1, \beta_2 \leq \beta_{\max}$. Let $h_{\beta_1}^*$ and $h_{\beta_2}^*$, where we made the dependence of h^* on β explicit, be the fixed points of (5.3.4) for (β_1, B) and (β_2, B) , respectively. Then, there exists a $\lambda < \infty$ such that*

$$(5.3.82) \quad \|\tanh(h_{\beta_1}^*) - \tanh(h_{\beta_2}^*)\|_{\text{MK}} \leq \lambda |\beta_1 - \beta_2|.$$

PROOF. For a given branching process tree BP we can, as in the proof of Proposition 5.13, couple $\tanh(h_{\beta}^*)$ to the root magnetizations $m_{\beta}^{\ell, f/+}(B)$ such that, for all $\beta \geq 0$ and $\ell \geq 0$,

$$(5.3.83) \quad m_{\beta}^{\ell, f}(B) \leq \tanh(h_{\beta}^*) \leq m_{\beta}^{\ell, +}(B),$$

where we have made the dependence of $m^{\ell, f/+}$ on β explicit. Without loss of generality, we may assume that $0 < \beta_1 \leq \beta_2 \leq \beta_{\max}$. Then, by the GKS inequality in Lemma 5.4,

$$(5.3.84) \quad |\tanh(h_{\beta_2}^*) - \tanh(h_{\beta_1}^*)| \leq m_{\beta_2}^{\ell, +}(B) - m_{\beta_1}^{\ell, f}(B) \\ = m_{\beta_2}^{\ell, +}(B) - m_{\beta_2}^{\ell, f}(B) + m_{\beta_2}^{\ell, f}(B) - m_{\beta_1}^{\ell, f}(B).$$

By Lemma 5.15, a.s.,

$$(5.3.85) \quad m_{\beta_2}^{\ell, +}(B) - m_{\beta_2}^{\ell, f}(B) \leq \frac{K}{\ell},$$

for some $K < \infty$. Since $m_{\beta}^{\ell, f}(B)$ is non-decreasing in β by the GKS inequality in Lemma 5.4,

$$(5.3.86) \quad m_{\beta_2}^{\ell, f}(B) - m_{\beta_1}^{\ell, f}(B) \leq (\beta_2 - \beta_1) \sup_{\beta_1 \leq \beta \leq \beta_{\max}} \frac{\partial m^{\ell, f}}{\partial \beta}.$$

Letting $\ell \rightarrow \infty$, it thus suffices to show that $\partial m^{\ell, f} / \partial \beta$ is, a.s., bounded uniformly in ℓ and $0 < \beta_1 \leq \beta \leq \beta_{\max}$.

We can write

$$(5.3.87) \quad \frac{\partial}{\partial \beta} m^{\ell, f}(\beta, B) = \sum_{(i, j) \in T(\ell)} \left(\langle \sigma_{\emptyset} \sigma_i \sigma_j \rangle_{\mu} - \langle \sigma_{\emptyset} \rangle_{\mu} \langle \sigma_i \sigma_j \rangle_{\mu} \right).$$

If i is on the path from the root \emptyset to j , then the spins σ_{\emptyset} and σ_j are conditionally independent given σ_i . It is not hard to see that in this case

$$(5.3.88) \quad \langle \sigma_{\emptyset} \sigma_i \sigma_j \rangle_{\mu} - \langle \sigma_{\emptyset} \rangle_{\mu} \langle \sigma_i \sigma_j \rangle_{\mu} = \gamma \langle \sigma_{\emptyset} \sigma_i \rangle_{\mu},$$

where γ is the arithmetic mean of the conditional expected value of σ_j for $\sigma_i = -1$ and the conditional expected value of σ_j for $\sigma_i = 1$, so that $|\gamma| \leq 1$. Further, $\langle \sigma_{\emptyset} \sigma_i \rangle_{\mu}$ is

non-negative by the GKS-inequality (Lemma 5.4). Therefore,

$$(5.3.89) \quad \frac{\partial}{\partial \beta} m^{\ell, f}(\beta, B) \leq \sum_{k=0}^{\ell-1} V_{k, \ell},$$

with

$$(5.3.90) \quad V_{k, \ell} = \sum_{w \in \partial \mathcal{T}(k)} d_w \frac{\partial}{\partial B_w} m^\ell(\underline{B}, 0) \Big|_{\underline{B}=B}.$$

By Lemma 5.12 and the GHS inequality in Lemma 5.5,

$$(5.3.91) \quad \begin{aligned} \frac{\partial}{\partial B_i} m^\ell(\underline{B}, 0) &= \frac{\partial}{\partial B_i} m^{\ell-1}(\underline{B}, \underline{H}) \\ &\leq \frac{\partial}{\partial B_i} m^{\ell-1}(\underline{B}, 0), \end{aligned}$$

for some field \underline{H} , so that $V_{k, \ell}$ is non-increasing in ℓ . We may assume that $B_w \geq B_{\min}$ for all $w \in \text{BP}(\ell)$ for some B_{\min} . Thus, also using Lemma 5.12,

$$(5.3.92) \quad \begin{aligned} V_{k, \ell} \leq V_{k, k+1} &= \sum_{w \in \partial \text{BP}(k)} d_w \frac{\partial}{\partial B_w} m^{k+1}(\underline{B}, 0) \Big|_{\underline{B}=B} \leq \sum_{w \in \partial \text{BP}(k)} d_w \frac{\partial}{\partial B_w} m^k(\underline{B}, \xi \cdot (d_w)_{w \in \partial \text{BP}(k)}) \Big|_{\underline{B}=B} \\ &= \frac{\partial}{\partial H} m^k(B, H \cdot (d_w)_{w \in \partial \text{BP}(k)}) \Big|_{H=\xi(\beta, B_{\min})}, \end{aligned}$$

where $\xi = \xi(\beta, B_{\min})$ is defined in (5.3.38). By the GHS inequality in Lemma 5.5, this derivative is non-increasing in H , so that, by Lemma 5.12, the above is at most

$$(5.3.93) \quad \frac{1}{\xi} [m^k(B, \xi \cdot (d_w)_{w \in \partial \text{BP}(k)}) - m^k(B, 0)] \leq \frac{1}{\xi} [m^{k+1}(B, 0) - m^k(B, 0)].$$

Therefore,

$$(5.3.94) \quad \frac{\partial}{\partial \beta} m^{\ell, f}(\beta, B) \leq \sum_{k=0}^{\ell-1} V_{k, \ell} \leq \frac{1}{\xi} \sum_{k=0}^{\ell-1} [m^{k+1}(B, 0) - m^k(B, 0)] \leq \frac{1}{\xi} < \infty,$$

for $0 < \beta_1 \leq \beta \leq \beta_{\max}$. □

We continue to study the dependence of $h^* = h^*(\beta, B)$ on B :

LEMMA 5.19 (Dependence of h^* on B). *Fix $\beta \geq 0$ and $B_1, B_2 \geq B_{\min} > 0$. Let $h_{B_1}^*$ and $h_{B_2}^*$, where we have made the dependence of h^* on B explicit, be the fixed points of (5.3.4) for (β, B_1) and (β, B_2) , respectively. Then, there exists a $\lambda < \infty$ such that*

$$(5.3.95) \quad \|\tanh(h_{B_1}^*) - \tanh(h_{B_2}^*)\|_{\text{MK}} \leq \lambda |B_1 - B_2|.$$

PROOF. This lemma can be proved along the same lines as Lemma 5.18. Indeed, for a given tree BP , we can couple $\tanh(h_B^*)$ to the root magnetizations $m^{\ell, f/+}(B)$ such that, for all $B > 0$ and $\ell \geq 0$,

$$(5.3.96) \quad m^{\ell, f}(B) \leq \tanh(h_B^*) \leq m^{\ell, +}(B).$$

Without loss of generality, we assume that $0 < B_{\min} \leq B_1 \leq B_2$. Then, by the GKS inequality (Lemma 5.4),

$$(5.3.97) \quad \begin{aligned} |\tanh(h_{B_2}^*) - \tanh(h_{B_1}^*)| &\leq m^{\ell,+}(B_2) - m^{\ell,f}(B_1) \\ &= m^{\ell,+}(B_2) - m^{\ell,f}(B_2) + m^{\ell,f}(B_2) - m^{\ell,f}(B_1). \end{aligned}$$

By Lemma 5.15, a.s.,

$$(5.3.98) \quad m^{\ell,+}(B_2) - m^{\ell,f}(B_2) \leq \frac{K}{\ell},$$

for some $K < \infty$. Since $m^{\ell,f}(B)$ is non-decreasing in B by the GKS inequality (Lemma 5.4),

$$(5.3.99) \quad m^{\ell,f}(B_2) - m^{\ell,f}(B_1) \leq (B_2 - B_1) \sup_{B \geq B_{\min} > 0} \frac{\partial m^{\ell,f}}{\partial B}.$$

Letting $\ell \rightarrow \infty$, it thus suffices to show that $\partial m^{\ell,f}/\partial B$ is bounded uniformly in ℓ and $B \geq B_{\min} > 0$. This follows from the GHS inequality in Lemma 5.5, since

$$(5.3.100) \quad \sup_{B \geq B_{\min} > 0} \frac{\partial m^{\ell,f}}{\partial B} \leq \left. \frac{\partial m^{\ell,f}}{\partial B} \right|_{B=B_{\min}} \leq \frac{2}{B_{\min}} [m^{\ell,f}(B_{\min}) - m^{\ell,f}(B_{\min}/2)] \leq \frac{2}{B_{\min}} < \infty.$$

□

Proof of Proposition 5.16. We start with Proposition 5.16(a), for which we combine Lemma 5.17 with Lemma 5.18. For Proposition 5.16(b), we combine Lemma 5.17 with Lemma 5.19 instead. □

5.3.4. Thermodynamic quantities: proofs of Theorem 5.10 and Corollary 5.11.

To prove the statements in Theorem 5.10 we need to show that we can interchange the limit of $n \rightarrow \infty$ and the derivatives of the finite volume pressure. We can do this using the monotonicity properties of the Ising model and the following lemma:

LEMMA 5.20 (Interchanging limits and derivatives). *Let $(f_n(x))_{n \geq 1}$ be a sequence of functions that are twice differentiable in x . Assume that*

- (i) $\lim_{n \rightarrow \infty} f_n(x) = f(x)$ for some function $x \mapsto f(x)$ which is differentiable in x ;
- (ii) $\frac{d}{dx} f_n(x)$ is monotone in $[x - h, x + h]$ for all $n \geq 1$ and some $h > 0$.

Then,

$$(5.3.101) \quad \lim_{n \rightarrow \infty} \frac{d}{dx} f_n(x) = \frac{d}{dx} f(x).$$

PROOF. First, suppose that $\frac{d^2}{dy^2} f_n(y) \geq 0$ for all $y \in [x - h, x + h]$, all $n \geq 1$ and some $h > 0$. Then, for $h > 0$ sufficiently small and all $n \geq 1$,

$$(5.3.102) \quad \frac{f_n(x - h) - f_n(x)}{-h} \leq \frac{d}{dx} f_n(x) \leq \frac{f_n(x + h) - f_n(x)}{h},$$

and taking $n \rightarrow \infty$ we get, by assumption (i), that

$$(5.3.103) \quad \frac{f(x - h) - f(x)}{-h} \leq \liminf_{n \rightarrow \infty} \frac{d}{dx} f_n(x) \leq \limsup_{n \rightarrow \infty} \frac{d}{dx} f_n(x) \leq \frac{f(x + h) - f(x)}{h}.$$

Taking $h \downarrow 0$ now proves the result. The proof for $\frac{d^2}{dx^2} f_n(x) \leq 0$ is similar. □

Lemma 5.20 is quite convenient to deal with the convergence of derivatives of functions of which we know their convexity and concavity. In our case, these convexity/concavity properties follow from the GKS inequality in Lemma 5.4 and the GHS inequality in Lemma 5.5. Let us now provide the details to the proof of Theorem 5.10:

PROOF OF THEOREM 5.10. We apply Lemma 5.20 with f_n equal to $B \mapsto \psi_n(\beta, B)$. Then,

$$(5.3.104) \quad M_n(\beta, B) = \frac{1}{n} \sum_{i \in [n]} \langle \sigma_i \rangle_{\mu_n} = \frac{\partial}{\partial B} \psi_n(\beta, B),$$

and $\psi_n(\beta, B) \xrightarrow{\mathbb{P}} \varphi(\beta, B)$ by Theorem 5.9 and $B \mapsto M_n(\beta, B)$ is non-decreasing by the GKS inequality in Lemma 5.4. Therefore,

$$(5.3.105) \quad M_n(\beta, B) = \frac{\partial}{\partial B} \psi_n(\beta, B) \xrightarrow{\mathbb{P}} \frac{\partial}{\partial B} \varphi(\beta, B),$$

which proves part (a). Part (b) follows immediately from Propositions 5.13–5.14. Part (c) is proved using Lemma 5.20 by combining part (a) of this theorem with the fact that $B \mapsto \frac{\partial}{\partial B} M_n(\beta, B)$ is non-increasing by the GHS inequality in Lemma 5.5. \square

We can now prove each of the statements in Corollary 5.11 by taking the proper derivative of $\varphi(\beta, B)$:

PROOF OF COROLLARY 5.11. It follows from Theorem 5.10 (a) that the magnetization per vertex is given by

$$(5.3.106) \quad M(\beta, B) = \frac{\partial}{\partial B} \varphi(\beta, B).$$

Similar to the proof of Proposition 5.14, we can ignore the dependence of h^* on B when differentiating $\varphi(\beta, B)$ with respect to B by Proposition 5.16. Therefore, with $\hat{\beta} = \tanh(\beta)$,

$$\frac{\partial}{\partial B} \varphi(\beta, B) = \frac{\partial}{\partial B} \mathbb{E} \left[\log \left(e^B \prod_{i=1}^D \{1 + \tanh(\beta) \tanh(h_i)\} + e^{-B} \prod_{i=1}^D \{1 - \tanh(\beta) \tanh(h_i)\} \right) \right]$$

$$(5.3.107) \quad = \mathbb{E} \left[\frac{e^B \prod_{i=1}^D (1 + \hat{\beta} \tanh(h_i)) - e^{-B} \prod_{i=1}^D (1 - \hat{\beta} \tanh(h_i))}{e^B \prod_{i=1}^D (1 + \hat{\beta} \tanh(h_i)) + e^{-B} \prod_{i=1}^D (1 - \hat{\beta} \tanh(h_i))} \right]$$

$$= \mathbb{E} \left[\frac{e^B \prod_{i=1}^D \left(\frac{1 + \hat{\beta} \tanh(h_i)}{1 - \hat{\beta} \tanh(h_i)} \right)^{1/2} - e^{-B} \prod_{i=1}^D \left(\frac{1 - \hat{\beta} \tanh(h_i)}{1 + \hat{\beta} \tanh(h_i)} \right)^{1/2}}{e^B \prod_{i=1}^D \left(\frac{1 + \hat{\beta} \tanh(h_i)}{1 - \hat{\beta} \tanh(h_i)} \right)^{1/2} + e^{-B} \prod_{i=1}^D \left(\frac{1 - \hat{\beta} \tanh(h_i)}{1 + \hat{\beta} \tanh(h_i)} \right)^{1/2}} \right],$$

where $(h_i)_{i \geq 1}$'s are i.i.d. copies of h^* independent of D . Using that $\operatorname{atanh}(x) = \frac{1}{2} \log \left(\frac{1+x}{1-x} \right)$ the above simplifies to

$$(5.3.108) \quad \mathbb{E} \left[\frac{e^B \prod_{i=1}^D e^{\operatorname{atanh}(\hat{\beta} \tanh(h_i))} - e^{-B} \prod_{i=1}^D e^{-\operatorname{atanh}(\hat{\beta} \tanh(h_i))}}{e^B \prod_{i=1}^D e^{\operatorname{atanh}(\hat{\beta} \tanh(h_i))} + e^{-B} \prod_{i=1}^D e^{-\operatorname{atanh}(\hat{\beta} \tanh(h_i))}} \right] \\ = \mathbb{E} \left[\tanh \left(B + \sum_{i=1}^D \operatorname{atanh}(\hat{\beta} \tanh(h_i)) \right) \right].$$

By Lemma 5.12, this indeed equals $\mathbb{E}[\langle \sigma_0 \rangle_{\nu_{L+1}}]$, where ν_{L+1} is given in (5.3.10), which proves part (a). Part (b) immediately follows from Propositions 5.13–5.14. \square

5.4. Critical temperature on the Ising model

In this section we compute the critical temperature for $\text{CM}_n(\mathbf{d})$:

THEOREM 5.21 (Critical temperature). *Consider $\text{CM}_n(\mathbf{d})$ where the degree distribution satisfies Conditions 1.6(a)–(b). Then, a.s., the critical temperature β_c equals*

$$(5.4.1) \quad \beta_c = \text{atanh}(1/\nu).$$

Note that if $\nu \searrow 1$ then $\beta_c \rightarrow \infty$ which is to be expected as there is no phase transition for $\nu \leq 1$ at any positive temperature. The other extreme is when $\nu = \infty$, which is the case, e.g., if the degree distribution obeys a power law with exponent $\tau \in (2, 3]$. In that case $\beta_c = 0$ and hence the spontaneous magnetization is positive for *any* finite temperature. This is similar to the instantaneous percolation phase for $\text{CM}_n(\mathbf{d})$ when $\tau \in (2, 3)$. The following exercise investigates the critical value for random regular graphs, for which $d_v = d$ for all $v \in [n]$:

EXERCISE 5.27 (Ising critical value for random regular graphs). *Show that $\beta_c = \infty$ for the random two-regular graph, while $\beta_c = \frac{1}{2} \log(d/(d-2))$ for the random d -regular graph with $d \geq 3$.*

Before starting with the proof, we state and prove a convenient lemma, about bounds on the function $\xi(x) = \text{atanh}(\widehat{\beta} \tanh x)$, where we recall the abbreviation $\widehat{\beta} = \tanh(\beta)$:

LEMMA 5.22 (Properties of $x \mapsto \xi(x)$). *For all $x, \beta \geq 0$,*

$$(5.4.2) \quad \widehat{\beta}x - \frac{\widehat{\beta}}{3(1-\widehat{\beta}^2)}x^3 \leq \xi(x) \leq \widehat{\beta}x.$$

The upper bound holds with strict inequality if $x, \beta > 0$.

PROOF. By Taylor's theorem,

$$(5.4.3) \quad \xi(x) = \xi(0) + \xi'(0)x + \xi''(\zeta)\frac{x^2}{2},$$

for some $\zeta \in (0, x)$. It is easily verified that $\xi(0) = 0$,

$$(5.4.4) \quad \xi'(0) = \left. \frac{\widehat{\beta}(1 - \tanh^2 x)}{1 - \widehat{\beta}^2 \tanh^2 x} \right|_{x=0} = \widehat{\beta},$$

and

$$(5.4.5) \quad \xi''(\zeta) = -\frac{2\widehat{\beta}(1-\widehat{\beta}^2)(\tanh \zeta)(1-\tanh^2 \zeta)}{(1-\widehat{\beta}^2 \tanh^2 \zeta)^2} \leq 0,$$

thus proving the upper bound. If $x, \beta > 0$ then also $\zeta > 0$ and hence the above holds with strict inequality.

For the lower bound, note that $\xi''(0) = 0$ and

$$(5.4.6) \quad \begin{aligned} \xi'''(\zeta) &= -\frac{2\widehat{\beta}(1-\widehat{\beta}^2)(1-\tanh^2\zeta)}{(1-\widehat{\beta}^2\tanh^2\zeta)^3} \left(1-3(1-\widehat{\beta}^2)\tanh^2\zeta-\widehat{\beta}^2\tanh^4\zeta\right) \\ &\geq -\frac{2\widehat{\beta}(1-\widehat{\beta}^2)(1-\tanh^2\zeta)}{(1-\widehat{\beta}^2)^2(1-\tanh^2\zeta)} = -\frac{2\widehat{\beta}}{1-\widehat{\beta}^2}. \end{aligned}$$

Thus, for all $\zeta \in (0, x)$,

$$(5.4.7) \quad \xi(x) = \xi(0) + \xi'(0)x + \xi''(0)\frac{x^2}{2} + \xi'''(\zeta)\frac{x^3}{3!} \geq \widehat{\beta}x - \frac{2\widehat{\beta}}{1-\widehat{\beta}^2}\frac{x^3}{3!}.$$

□

PROOF OF THEOREM 5.21. Let $\beta^* = \operatorname{atanh}(1/\nu)$. We first show that if $\beta < \beta^*$, then

$$(5.4.8) \quad \lim_{B \searrow 0} M(\beta, B) = 0,$$

which implies that $\beta_c \geq \beta^*$. Note that this only arises when $\nu < \infty$. Later, we show that if $\lim_{B \searrow 0} M(\beta, B) = 0$ then $\beta \leq \beta^*$, implying that $\beta_c \leq \beta^*$.

Proof of $\beta_c \geq \beta^*$. Suppose that $\beta < \beta^*$. Then, by the fact that $\tanh x \leq x$ and Wald's identity,

$$(5.4.9) \quad M(\beta, B) = \mathbb{E}\left[\tanh\left(B + \sum_{i=1}^D \xi(h_i)\right)\right] \leq B + \mathbb{E}[D]\mathbb{E}[\xi(h)].$$

We use the upper bound in Lemma 5.22 to get

$$(5.4.10) \quad \mathbb{E}[\xi(h)] = \mathbb{E}[\operatorname{atanh}(\widehat{\beta} \tanh h)] \leq \widehat{\beta}\mathbb{E}[h] = \widehat{\beta}(B + \nu\mathbb{E}[\xi(h)]).$$

Further, note that

$$(5.4.11) \quad \mathbb{E}[\xi(h)] = \mathbb{E}[\operatorname{atanh}(\widehat{\beta} \tanh h)] \leq \beta,$$

because $\tanh h \leq 1$. Applying inequality (5.4.10) ℓ times to (5.4.9) and subsequently using inequality (5.4.11) once gives

$$(5.4.12) \quad M(\beta, B) \leq B + B\widehat{\beta}\mathbb{E}[D]\frac{1-(\widehat{\beta}\nu)^\ell}{1-\widehat{\beta}\nu} + \beta\mathbb{E}[D](\widehat{\beta}\nu)^\ell.$$

Hence,

$$(5.4.13) \quad \begin{aligned} M(\beta, B) &\leq \limsup_{\ell \rightarrow \infty} \left(B + B\widehat{\beta}\mathbb{E}[D]\frac{1-(\widehat{\beta}\nu)^\ell}{1-\widehat{\beta}\nu} + \beta\mathbb{E}[D](\widehat{\beta}\nu)^\ell\right) \\ &= B \left(1 + \widehat{\beta}\mathbb{E}[D]\frac{1}{1-\widehat{\beta}\nu}\right), \end{aligned}$$

because $\widehat{\beta} < \widehat{\beta}^* = 1/\nu$. Therefore,

$$(5.4.14) \quad \lim_{B \searrow 0} M(\beta, B) \leq \lim_{B \searrow 0} B \left(1 + \widehat{\beta}\mathbb{E}[D]\frac{1}{1-\widehat{\beta}\nu}\right) = 0.$$

This proves the lower bound on β_c .

Proof of $\beta_c \leq \beta^*$. We adapt Lyons' proof in [211] for the critical temperature of deterministic trees to the random tree to show that $\beta_c \leq \beta^*$. Take $\beta > \beta^*$, and assume that $\lim_{B \searrow 0} M(\beta, B) = 0$. We aim to arrive at a contradiction.

Note that Theorem 5.10(a) shows that the magnetization $M(\beta, B)$ is equal to the expectation over the random branching process tree \mathbf{BP} of the root magnetization. Hence, denoting the root of the tree \mathbf{BP} by \emptyset , we get $M(\beta, B) = \mathbb{E}[\langle \sigma_\emptyset \rangle]$. It follows from our assumption on $M(\beta, B)$ that, a.s., $\lim_{B \searrow 0} \langle \sigma_\emptyset \rangle = 0$, since the latter limit exists by the GKS inequalities in Lemma 5.4.

We therefore condition on the tree $T = \mathbf{BP}$. Define, for $v \in T$,

$$(5.4.15) \quad h(v) = \langle \sigma_v \rangle \quad \text{and} \quad h^{\ell,+}(v) = \langle \sigma_v \rangle^{\ell,+},$$

and let $|v|$ denote the graph distance from \emptyset to v . Furthermore, we say that $w \leftarrow v$ if $\{w, v\}$ is an edge in T and $|w| = |v| + 1$. By the Pruning Lemma 5.12, for $|v| < \ell$,

$$(5.4.16) \quad h^{\ell,+}(v) = B + \sum_{w \leftarrow v} \xi(h^{\ell,+}(w)).$$

Since this recursion has a unique solution by Proposition 5.13, we have $h(\emptyset) = \lim_{\ell \rightarrow \infty} h^{\ell,+}(\emptyset)$. Therefore, if we suppose that $\lim_{B \searrow 0} \langle \sigma_\emptyset \rangle = 0$, then also $\lim_{B \searrow 0} h(\emptyset) = 0$ and then it thus also holds that $\lim_{B \searrow 0} \lim_{\ell \rightarrow \infty} h^{\ell,+}(\emptyset) = 0$. Because of (5.4.16), we must then have, for all $v \in T$,

$$(5.4.17) \quad \lim_{B \searrow 0} \lim_{\ell \rightarrow \infty} h^{\ell,+}(v) = 0.$$

Now, fix $0 < \beta_0 < \beta$ and choose ℓ large enough and B small enough such that, for some $\varepsilon = \varepsilon(\beta_0, \beta) > 0$ that we choose later,

$$(5.4.18) \quad h^{\ell,+}(v) \leq \varepsilon,$$

for all $v \in T$ with $|v| = 1$. Note that $h^{\ell,+}(v) = \infty > \varepsilon$ for $v \in T$ with $|v| = \ell$.

We now follow Lyons in [211], and say that Π is a *cutset* if Π is a finite subset of $T \setminus \{\emptyset\}$ and every path from \emptyset to infinity intersects Π at exactly one vertex $v \in \Pi$. We write $v \leq \Pi$ if every infinite path from v intersects Π and write $\sigma < \Pi$ if $\sigma \leq \Pi$ and $\sigma \notin \Pi$. Then, since $h^{\ell,+}(v) \leq \varepsilon$ for $v \in \mathcal{T}$ with $|v| = 1$ and $h^{\ell,+}(v) = \infty > \varepsilon$ for $v \in T$ with $|v| = \ell$, there is a unique cutset Π_ℓ , such that $h^{\ell,+}(v) \leq \varepsilon$ for all $v \leq \Pi_\ell$, and for all $v \in \Pi_\ell$ there is at least one $w \leftarrow v$ such that $h^{\ell,+}(w) > \varepsilon$.

It follows from the lower bound in Lemma 5.22 that, for $v < \Pi_\ell$,

$$(5.4.19) \quad h^{\ell,+}(v) = B + \sum_{w \leftarrow v} \xi(h^{\ell,+}(w)) \geq \sum_{w \leftarrow v} \widehat{\beta} h^{\ell,+}(w) - \frac{\widehat{\beta} h^{\ell,+}(w)^3}{3(1 - \widehat{\beta}^2)} \geq \sum_{w \leftarrow v} \widehat{\beta} h^{\ell,+}(w) \left(1 - \frac{\varepsilon^2}{3(1 - \widehat{\beta}^2)}\right),$$

while, for $v \in \Pi_\ell$,

$$(5.4.20) \quad h^{\ell,+}(v) = B + \sum_{w \leftarrow v} \xi(h^{\ell,+}(w)) > \xi(\varepsilon).$$

If we now choose $\varepsilon > 0$ such that

$$(5.4.21) \quad \widehat{\beta} \left(1 - \frac{\varepsilon^2}{3(1 - \widehat{\beta}^2)^2}\right) = \widehat{\beta}_0,$$

which is possible because $\beta_0 < \beta$, then, iterating (5.4.19) in each direction until Π_ℓ and then using (5.4.20),

$$(5.4.22) \quad h^{\ell,+}(\phi) \geq \sum_{v \in \Pi_\ell} \widehat{\beta}_0^{|v|} \xi(\varepsilon).$$

Since $\xi(\varepsilon) > 0$ and $\lim_{B \searrow 0} \lim_{\ell \rightarrow \infty} h^{\ell,+}(\phi) = 0$,

$$(5.4.23) \quad \inf_{\Pi} \sum_{v \in \Pi} \widehat{\beta}_0^{|v|} = 0.$$

Lyons [212, Proposition 6.4] then shows that $\widehat{\beta}_0 \leq 1/\nu$. This holds for all $\beta_0 < \beta$, so

$$(5.4.24) \quad \beta \leq \operatorname{atanh}(1/\nu) = \beta^*.$$

We have arrived at a contradiction, as we took $\beta > \beta^*$. This proves the upper bound on β_c , thus concluding the proof. \square

We next show that the phase transition at this critical temperature is *continuous*. This will be highly useful when studying the *critical exponents* for the Ising model on the configuration model, which we do in the next section:

LEMMA 5.23 (Continuous phase transition). *Let $((\beta_n, B_n))_{n \geq 1}$ be a sequence with β_n and B_n non-increasing, $\beta_n \geq \beta_c$ and $B_n > 0$, and $\beta_n \searrow \beta_c$ and $B_n \searrow 0$ as $n \rightarrow \infty$. Then,*

$$(5.4.25) \quad \lim_{n \rightarrow \infty} \mathbb{E}[\xi(\beta_n, B_n)] = 0.$$

In particular,

$$(5.4.26) \quad \lim_{B \searrow 0} \mathbb{E}[\xi(h(\beta_c, B))] = 0, \quad \text{and} \quad \lim_{\beta \searrow \beta_c} \mathbb{E}[\xi(h(\beta, 0^+))] = 0.$$

PROOF. For all sequences $((\beta_n, B_n))_{n \geq 1}$ satisfying the assumptions stated in the lemma, $\mathbb{E}[\xi(\beta_n, B_n)]$ is non-increasing in n and it is also non-negative so that the limit as $n \rightarrow \infty$ exists. By the concavity of $h \mapsto \xi(h)$ and Jensen's inequality,

$$(5.4.27) \quad 0 \leq c \equiv \lim_{n \rightarrow \infty} \mathbb{E}[\xi(\beta_n, B_n)] \leq \lim_{n \rightarrow \infty} \xi(B_n + \nu \mathbb{E}[\xi(h(\beta_n, B_n))]) = \xi(\nu c).$$

Since $\xi(x) < \widehat{\beta}_c x$ for $x > 0$ by Lemma 5.22 and using $\widehat{\beta}_c = 1/\nu$, we obtain

$$(5.4.28) \quad \xi(\nu c) < \widehat{\beta}_c \nu c = c,$$

leading to a contradiction when $c > 0$. \square

EXERCISE 5.28 (Continuity of the magnetization). *Recall the formula for the magnetization $M(\beta, B)$ in (5.3.12) in Corollary 5.11. Use Lemma 5.23 to prove that $\lim_{n \rightarrow \infty} M(\beta_n, B_n) = 0$ when β_n and B_n non-increasing, $\beta_n \geq \beta_c$ and $B_n > 0$, and $\beta_n \searrow \beta_c$ and $B_n \searrow 0$ as $n \rightarrow \infty$.*

5.5. Ising critical exponents on locally-tree like random graphs

We now investigate the critical behavior of the Ising model on the configuration model with power-law degree sequences. Near the critical temperature the behavior of the Ising model can be described by critical exponents, recall Definition 5.1. The values of these critical exponents for different values of τ are stated in the following theorem:

THEOREM 5.24 (Critical exponents). *Consider $\text{CM}_n(\mathbf{d})$ where the degree distribution satisfies Conditions 1.6(a)-(b), and has a strongly finite mean as in (5.3.1). Assume that $\mathbb{E}[D^4] < \infty$, or that there exist a $\tau \in (3, 5]$ and constants c_D, C_D such that, for all $x \geq 1$,*

$$(5.5.1) \quad c_D x^{-(\tau-1)} \leq 1 - F_D(x) \leq C_D x^{-(\tau-1)}.$$

Then, the critical exponents β, δ and γ defined in Definition 5.1 exist and satisfy

	$\tau \in (3, 5)$	$\mathbb{E}[D^4] < \infty$
β	$1/(\tau - 3)$	$1/2$
δ	$\tau - 2$	3
γ	1	1

The exponent γ' defined in Definition 5.1 satisfies $\gamma' \geq 1$.

For the boundary case $\tau = 5$ there are logarithmic corrections for $\beta = 1/2$ and $\delta = 3$, but not for $\gamma = 1$ and for the lower bound $\gamma' \geq 1$. Indeed, (5.1.16) holds with $\gamma = 1$ and the lower bound in (5.1.17) holds with $\gamma' = 1$, while

$$(5.5.2) \quad M(\beta, 0^+) \asymp \left(\frac{\beta - \beta_c}{\log 1/(\beta - \beta_c)} \right)^{1/2} \quad \text{for } \beta \searrow \beta_c, \quad M(\beta_c, B) \asymp \left(\frac{B}{\log(1/B)} \right)^{1/3} \quad \text{for } B \searrow 0.$$

From the previous theorem we can also derive the joint scaling of the magnetization as $(\beta, B) \searrow (\beta_c, 0)$:

COROLLARY 5.25 (Joint scaling in B and $(\beta - \beta_c)$). *Under the conditions of Theorem 5.24 with $\tau \neq 5$,*

$$(5.5.3) \quad M(\beta, B) = \Theta((\beta - \beta_c)^\beta + B^{1/\delta}),$$

where $f(\beta, B) = \Theta(g(\beta, B))$ means that there exist constants $c_1, C_1 > 0$ such that $c_1 g(\beta, B) \leq f(\beta, B) \leq C_1 g(\beta, B)$ for all $B \in (0, \varepsilon)$ and $\beta \in (\beta_c, \beta_c + \varepsilon)$ with ε small enough.

For $\tau = 5$,

$$(5.5.4) \quad M(\beta, B) = \Theta\left(\left(\frac{\beta - \beta_c}{\log 1/(\beta - \beta_c)} \right)^{1/2} + \left(\frac{B}{\log(1/B)} \right)^{1/3} \right).$$

Organization of the proof of Theorem 5.24. The remainder of this section is organized as follows. We start with some preliminary computations. The proof that the exponents stated in Theorem 5.24 are indeed the correct values of β, δ and γ is given in Section 5.5.3.

Preliminaries: formula for the magnetization. The starting point of our analysis is the formula (5.3.12) for the magnetization in Corollary 5.11, which we restate here for convenience as

$$(5.5.5) \quad M(\beta, B) = \mathbb{E} \left[\tanh \left(B + \sum_{i=1}^D \xi(h_i) \right) \right],$$

where $(h_i)_{i \geq 1}$ are i.i.d. copies of the fixed point of the distributional recursion (5.3.4). Here we present a more intuitive proof based on local weak convergence and correlation inequalities:

ALTERNATIVE PROOF OF (5.3.12). Let \emptyset be a vertex picked uniformly at random from $[n]$ and \mathbb{E}_n be the corresponding expectation. Then,

$$(5.5.6) \quad M_n(\beta, B) = \frac{1}{n} \sum_{i=1}^n \langle \sigma_i \rangle_{\mu_n} = \mathbb{E}_n[\langle \sigma_{\emptyset} \rangle_{\mu_n}].$$

Denote by $\langle \cdot \rangle_{\mu_n}^{\ell, +/f}$ the expectations with respect to the Ising measure with $+/\text{free}$ boundary conditions on vertices at graph distance ℓ from \emptyset . Note that $\langle \sigma_{\emptyset} \rangle_{\mu_n}^{\ell, +/f}$ only depends on the spins of vertices in $B_{\emptyset}(\ell)$. By the GKS inequality in Lemma 5.4,

$$(5.5.7) \quad \langle \sigma_{\emptyset} \rangle_{\mu_n}^{\ell, f} \leq \langle \sigma_{\emptyset} \rangle_{\mu_n} \leq \langle \sigma_{\emptyset} \rangle_{\mu_n}^{\ell, +}.$$

Taking the limit $n \rightarrow \infty$ and using the local weak convergence result in Theorem 2.11, the ball $B_{\emptyset}(\ell)$ has the same distribution as the branching process tree $\text{BP}(\ell)$, because of the locally tree-like nature of the graph sequence. With $\langle \cdot \rangle^{\ell, +/f}$ denoting the expectations with respect to the Ising measure with $+/\text{free}$ boundary conditions on $\text{BP}(\ell)$, this means that

$$(5.5.8) \quad \lim_{n \rightarrow \infty} \langle \sigma_{\emptyset} \rangle_{\mu_n}^{\ell, +/f} = \langle \sigma_{\emptyset} \rangle^{\ell, +/f}.$$

Conditioned on the tree BP , we can prune the tree using Lemma 5.12, to obtain that

$$(5.5.9) \quad \langle \sigma_{\emptyset} \rangle^{\ell, f} = \tanh \left(B + \sum_{i=1}^D \xi(h_i^{(\ell-1)}) \right).$$

Similarly,

$$(5.5.10) \quad \langle \sigma_{\emptyset} \rangle^{\ell, +} = \tanh \left(B + \sum_{i=1}^D \xi(h_i'^{(\ell-1)}) \right),$$

where $h_i'^{(t+1)}$ also satisfies (5.3.4), but has initial value $h'^{(0)} = \infty$. Since this recursion has a unique fixed point by Proposition 5.13, we prove (5.3.12) by taking the limit $\ell \rightarrow \infty$ and taking the expectation over the branching process tree BP . \square

5.5.1. Bounds on moments of $\xi(h)$. In this section, we analyze moments of $\xi(h)$ that will be crucial in our analysis of the critical exponents of the magnetization. Throughout Section 5.5.1 we assume that B is sufficiently close to zero and $\beta_c < \beta < \beta_c + \varepsilon$ for ε sufficiently small, so that we can apply Lemma 5.23 to make sure $\mathbb{E}[\xi(h)]$ is sufficiently small.

Preliminaries: error term notation. We start by introducing some notation to be able to deal appropriately with error terms in our Taylor expansions. In the following, we write $c_i, C_i, i \geq 1$ for constants that (i) only depend on β and moments of D (they do not depend on B); (ii) that satisfy

$$(5.5.11) \quad 0 < \liminf_{\beta \searrow \beta_c} c_i(\beta) \leq \limsup_{\beta \searrow \beta_c} c_i(\beta) < \infty,$$

and the same holds for the C_i . The index i is just a label for the constants. For reader convenience we try to use the i th label to denote a constant appearing in a bound involving the i th moment of $\xi(h)$. However this is not always possible and therefore in general nothing should be sought from the labeling of the constants. What we consistently do is to use C_i for constants appearing in upper bounds, while c_i appears in lower bounds.

Furthermore, we write e_i , for $i \geq 1$, (again the labeling is arbitrary) for error functions that only depend on $\beta, B, \mathbb{E}[\xi(h)]$ and moments of Δ , and satisfy for all $\beta \in (\beta_c, \beta_c + \epsilon)$

$$(5.5.12) \quad \limsup_{B \searrow 0} e_i(\beta, B) < \infty \quad \text{and} \quad \lim_{B \searrow 0} e_i(\beta_c, B) = 0.$$

The expressions of c_i, C_i, e_i are given in the proofs of the results of this section and conditions (5.5.11) (5.5.12) are verified explicitly. Finally, we write $\nu_k = \mathbb{E}[D(D-1) \cdots (D-k+1)]$ for the k th factorial moment of D , so that $\nu_2/\nu_1 = \nu$.

Bounds on higher moments of $\xi(h)$. In this section, we derive appropriate bounds on the moments of $\xi(h)$ that we will rely on later. We start by bounding the second moment of $\xi(h)$:

LEMMA 5.26 (Bounds on second moment of $\xi(h)$). *Let $\beta \geq \beta_c$ and $B > 0$. Then,*

$$(5.5.13) \quad \mathbb{E}[\xi(h)^2] \leq \begin{cases} C_2 \mathbb{E}[\xi(h)]^2 + B e_2 & \text{when } \mathbb{E}[D^3] < \infty, \\ C_2 \mathbb{E}[\xi(h)]^2 \log(1/\mathbb{E}[\xi(h)]) + B e_2 & \text{when } \tau = 4, \\ C_2 \mathbb{E}[\xi(h)]^{\tau-2} + B e_2 & \text{when } \tau \in (3, 4). \end{cases}$$

PROOF. We first treat the case $\mathbb{E}[D^3] < \infty$. We use Lemma 5.22 and the recursion in (5.3.4) to obtain

$$(5.5.14) \quad \begin{aligned} \mathbb{E}[\xi(h)^2] &\leq \widehat{\beta}^2 \mathbb{E}[h^2] = \widehat{\beta}^2 \mathbb{E}\left[\left(B + \sum_{i=1}^{\Delta} \xi(h_i)\right)^2\right] \\ &= \widehat{\beta}^2 \left(B^2 + 2B\nu \mathbb{E}[\xi(h)] + \nu_2 \mathbb{E}[\xi(h)]^2 + \nu \mathbb{E}[\xi(h)^2]\right). \end{aligned}$$

Since $1 - \widehat{\beta}^2 \nu > 0$, because β is sufficiently close to β_c and $\widehat{\beta}_c = 1/\nu < 1$, the lemma holds with

$$(5.5.15) \quad C_2 = \frac{\widehat{\beta}^2 \nu_2}{1 - \widehat{\beta}^2 \nu}, \quad \text{and} \quad e_2 = \frac{B \widehat{\beta}^2 + 2 \widehat{\beta}^2 \nu \mathbb{E}[\xi(h)]}{1 - \widehat{\beta}^2 \nu}.$$

It is not hard to see that (5.5.11) holds. For e_2 the first property of (5.5.12) can also easily be seen. The second property in (5.5.12) follows from Lemma 5.23.

If $\tau \leq 4$, then $\mathbb{E}[D^3] = \infty$ and the above does not work. To analyze this case, we apply the recursion (5.3.4) and split the expectation over Δ in small and large degrees:

$$(5.5.16) \quad \mathbb{E}[\xi(h)^2] = \mathbb{E}\left[\xi\left(B + \sum_{i=1}^{\Delta} \xi(h_i)\right)^2 \mathbb{1}_{\{\Delta \leq \ell\}}\right] + \mathbb{E}\left[\xi\left(B + \sum_{i=1}^{\Delta} \xi(h_i)\right)^2 \mathbb{1}_{\{\Delta > \ell\}}\right].$$

We use Lemma 5.22 to bound the first term as follows:

$$(5.5.17) \quad \begin{aligned} \mathbb{E}\left[\xi\left(B + \sum_{i=1}^{\Delta} \xi(h_i)\right)^2 \mathbb{1}_{\{\Delta \leq \ell\}}\right] &\leq \widehat{\beta}^2 \mathbb{E}\left[\left(B + \sum_{i=1}^{\Delta} \xi(h_i)\right)^2 \mathbb{1}_{\{\Delta \leq \ell\}}\right] \\ &\leq \widehat{\beta}^2 \left(B^2 + 2B\nu \mathbb{E}[\xi(h)] + \mathbb{E}[\Delta^2 \mathbb{1}_{\{\Delta \leq \ell\}}] \mathbb{E}[\xi(h)]^2 + \nu \mathbb{E}[\xi(h)^2]\right), \end{aligned}$$

where in the second inequality we used the independence of the $\xi(h_i)$ and Wald's identity, and bounded $\mathbb{E}[\Delta \mathbb{1}_{\{\Delta \leq \ell\}}] \leq \nu$. By Lemma 2.39 with $a = 3$, for $\tau \in (3, 4)$,

$$(5.5.18) \quad \mathbb{E}[\Delta^2 \mathbb{1}_{\{\Delta \leq \ell\}}] \leq C_{2,\tau} \ell^{4-\tau},$$

while, for $\tau = 4$,

$$(5.5.19) \quad \mathbb{E}[\Delta^2 \mathbb{1}_{\{\Delta \leq \ell\}}] \leq C_{2,4} \log \ell.$$

To bound the second sum in (5.5.16), note that $\xi(h) \leq \beta$. Hence,

$$(5.5.20) \quad \mathbb{E}\left[\xi\left(B + \sum_{i=1}^{\Delta} \xi(h_i)\right)^2 \mathbb{1}_{\{\Delta > \ell\}}\right] \leq \beta^2 \mathbb{E}[\mathbb{1}_{\{\Delta > \ell\}}] \leq C_{0,\tau} \beta^2 \ell^{2-\tau}.$$

The optimal bound (up to a constant) can be achieved by choosing ℓ such that $\ell^{4-\tau} \mathbb{E}[\xi(h)]^2$ and $\ell^{2-\tau}$ are of the same order of magnitude. Hence, we choose $\ell = 1/\mathbb{E}[\xi(h)]$. This choice of the truncation value ℓ turns out to be the relevant choice when making the distinction between small and large values of Δ also in further computations and hence will also be used (up to a constant) there.

Combining the two upper bounds gives the desired result with

$$(5.5.21) \quad C_2 = \frac{1}{1 - \widehat{\beta}^2 \nu} \left(C_{2,\tau} \widehat{\beta}^2 + C_{0,\tau} \beta^2 \right),$$

where, for $\tau = 4$, we have also used that $\mathbb{E}[\xi(h)]^2 \leq \mathbb{E}[\xi(h)]^2 \log(1/\mathbb{E}[\xi(h)])$, and

$$(5.5.22) \quad e_2 = \frac{B \widehat{\beta}^2 + 2 \widehat{\beta}^2 \nu \mathbb{E}[\xi(h)]}{1 - \widehat{\beta}^2 \nu}.$$

□

We continue with an upper bound on the third moment of $\xi(h)$, whose proof we do not provide as it is quite similar to that of Lemma 5.26:

LEMMA 5.27 (Bounds on third moment of $\xi(h)$). *Let $\beta \geq \beta_c$ and $B > 0$. Then,*

$$(5.5.23) \quad \mathbb{E}[\xi(h)^3] \leq \begin{cases} C_3 \mathbb{E}[\xi(h)]^3 + B e_3 & \text{when } \mathbb{E}[D^4] < \infty, \\ C_3 \mathbb{E}[\xi(h)]^3 \log(1/\mathbb{E}[\xi(h)]) + B e_3 & \text{when } \tau = 5, \\ C_3 \mathbb{E}[\xi(h)]^{\tau-2} + B e_3 & \text{when } \tau \in (3, 5). \end{cases}$$

5.5.2. Bounds on first moment of $\xi(h)$. In this section, we derive sharp bounds on the first moment of $\xi(h)$, using the bounds on the higher moments as derived in the previous section. The main upper bound is as follows:

PROPOSITION 5.28 (Upper bound on first moment of $\xi(h)$). *Let $\beta \geq \beta_c$ and $B > 0$. Then, there exists a $C_1 > 0$ such that*

$$(5.5.24) \quad \mathbb{E}[\xi(h)] \leq \widehat{\beta} B + \widehat{\beta} \nu \mathbb{E}[\xi(h)] - C_1 \mathbb{E}[\xi(h)]^\delta,$$

where δ takes the values as stated in Theorem 5.24. For $\tau = 5$,

$$(5.5.25) \quad \mathbb{E}[\xi(h)] \leq \widehat{\beta} B + \widehat{\beta} \nu \mathbb{E}[\xi(h)] - C_1 \mathbb{E}[\xi(h)]^3 \log(1/\mathbb{E}[\xi(h)]).$$

PROOF. We first use recursion (5.3.4) and rewrite it as

$$(5.5.26) \quad \mathbb{E}[\xi(h)] = \mathbb{E}\left[\xi\left(B + \sum_{i=1}^{\Delta} \xi(h_i)\right)\right] = \widehat{\beta} B + \widehat{\beta} \nu \mathbb{E}[\xi(h)] + T_1 + T_2,$$

where

$$(5.5.27) \quad T_1 = \mathbb{E} \left[\xi \left(B + \Delta \mathbb{E}[\xi(h)] \right) - \widehat{\beta} \left(B + \Delta \mathbb{E}[\xi(h)] \right) \right],$$

and

$$(5.5.28) \quad T_2 = \mathbb{E} \left[\xi \left(B + \sum_{i=1}^{\Delta} \xi(h_i) \right) - \xi \left(B + \Delta \mathbb{E}[\xi(h)] \right) \right].$$

Here, T_1 can be seen as the error of a first order Taylor series approximation of $\xi(B + \Delta \mathbb{E}[\xi(h)])$ around 0, whereas T_2 is the error made by replacing $\xi(h_i)$ by its expected value in the sum. By Lemma 5.22, $T_1 \leq 0$ and by concavity of $x \mapsto \xi(x)$ and Jensen's inequality $T_2 \leq 0$. We bound T_1 separately for the cases where $\mathbb{E}[D^4] < \infty$, $\tau \in (3, 5)$ and $\tau = 5$. Since these bounds on T_1 are already of the correct order to prove the proposition it suffices to use $T_2 \leq 0$ and we do not bound T_2 more sharply.

Bound on T_1 when $\mathbb{E}[D^4] < \infty$. To bound T_1 for $\mathbb{E}[D^4] < \infty$ we use that $\xi''(0) = 0$, so that it follows from Taylor's theorem that, a.s.,

$$(5.5.29) \quad \xi(B + \Delta \mathbb{E}[\xi(h)]) - \widehat{\beta}(B + \Delta \mathbb{E}[\xi(h)]) = \frac{\xi'''(\zeta)}{6} (B + \Delta \mathbb{E}[\xi(h)])^3,$$

for some $\zeta \in (0, B + \Delta \mathbb{E}[\xi(h)])$. Note that

$$(5.5.30) \quad \xi'''(\zeta) = -\frac{2\widehat{\beta}(1 - \widehat{\beta}^2)(1 - \tanh^2 \zeta)}{(1 - \widehat{\beta}^2 \tanh^2 \zeta)^3} \left(1 - 3(1 - \widehat{\beta}^2) \tanh^2 \zeta - \widehat{\beta}^2 \tanh^4 \zeta \right),$$

which we would like to bound from above by a negative constant. By Lemma 5.22, a.s.,

$$(5.5.31) \quad \xi(B + \Delta \mathbb{E}[\xi(h)]) - \widehat{\beta}(B + \Delta \mathbb{E}[\xi(h)]) \leq 0,$$

so that we are allowed to assume that $B + \Delta \mathbb{E}[\xi(h)]$ is sufficiently small. Indeed, from the above equation it follows that, for any constant a ,

$$(5.5.32) \quad T_1 \leq \mathbb{E} \left[\left(\xi(B + \Delta \mathbb{E}[\xi(h)]) - \widehat{\beta}(B + \Delta \mathbb{E}[\xi(h)]) \right) \mathbb{1}_{\{B + \Delta \mathbb{E}[\xi(h)] \leq a\}} \right].$$

If, for convenience, we choose $a = a \tanh \frac{1}{2}$, then

$$(5.5.33) \quad \xi'''(\zeta) \leq -\frac{3}{8} \widehat{\beta}(1 - \widehat{\beta}^2),$$

and hence,

$$(5.5.34) \quad \begin{aligned} T_1 &\leq -\frac{1}{16} \widehat{\beta}(1 - \widehat{\beta}^2) \mathbb{E} \left[(B + \Delta \mathbb{E}[\xi(h)])^3 \mathbb{1}_{\{B + \Delta \mathbb{E}[\xi(h)] \leq a \tanh \frac{1}{2}\}} \right] \\ &\leq -\frac{1}{16} \widehat{\beta}(1 - \widehat{\beta}^2) \mathbb{E} [\Delta^3 \mathbb{1}_{\{\Delta \mathbb{E}[\xi(h)] \leq a \tanh \frac{1}{2} - B\}}] \mathbb{E} [\xi(h)]^3. \end{aligned}$$

Note that $B + \Delta \mathbb{E}[\xi(h)]$ converges to zero for both limits of interest. Thus $\mathbb{E}[\Delta^3 \mathbb{1}_{\{\Delta \mathbb{E}[\xi(h)] \leq a \tanh \frac{1}{2} - B\}}]$ tends to infinity in the case $\mathbb{E}[D^4] = \infty$ and hence this bound is not useful. We provide better bounds in the next paragraph for this case.

Bound on T_1 when $\tau \in (3, 5]$. For $\tau \in (3, 5]$, we make the expectation over Δ explicit:

$$(5.5.35) \quad T_1 = \sum_{k=0}^{\infty} \mathbb{P}(\Delta = k) \left(\xi(B + k\mathbb{E}[\xi(h)]) - \widehat{\beta}(B + k\mathbb{E}[\xi(h)]) \right),$$

where it should be noted that all terms in this sum are negative because of Lemma 5.22. Define $f(k) = \xi(B + k\mathbb{E}[\xi(h)]) - \widehat{\beta}(B + k\mathbb{E}[\xi(h)])$ and note that $f(k)$ is non-increasing. We use (2.6.3) to rewrite

$$(5.5.36) \quad \begin{aligned} T_1 &= \sum_{k=0}^{\infty} f(k) \mathbb{P}(\Delta = k) = f(0) + \sum_{k \geq 1} [f(k) - f(k-1)] \mathbb{P}(\Delta \geq k) \\ &\leq f(0) + c_D \sum_{k \geq 1} [f(k) - f(k-1)] k^{-(\tau-2)} \\ &= f(0) + c_D \sum_{k \geq 1} [f(k) - f(k-1)] \sum_{\ell \geq k} (\ell^{-(\tau-2)} - (\ell+1)^{-(\tau-2)}) \end{aligned}$$

Then, we can again interchange the summation order as we did to obtain (2.6.3) to rewrite this as

$$(5.5.37) \quad \begin{aligned} T_1 &\leq f(0) + c_D \sum_{\ell \geq 1} \sum_{k=1}^{\ell} [f(k) - f(k-1)] (\ell^{-(\tau-2)} - (\ell+1)^{-(\tau-2)}) \\ &= f(0)(1 - c_D) + c_D \sum_{\ell \geq 1} f(\ell) (\ell^{-(\tau-2)} - (\ell+1)^{-(\tau-2)}). \end{aligned}$$

Using the convexity of $\ell^{-(\tau-2)}$ this can be bounded as

$$(5.5.38) \quad T_1 \leq f(0)(1 - c_D) + (\tau - 2)c_D \sum_{\ell \geq 1} f(\ell) (\ell + 1)^{-(\tau-1)}.$$

Since we can assume that $c_D \leq 1$, $f(0)(1 - c_D) \leq 0$. Hence,

$$(5.5.39) \quad \begin{aligned} T_1 &\leq (\tau - 2)c_D (\mathbb{E}[\xi(h)])^{\tau-1} \sum_{k=0}^{\infty} ((k+1)\mathbb{E}[\xi(h)])^{-(\tau-1)} \left(\xi(B + k\mathbb{E}[\xi(h)]) - \widehat{\beta}(B + k\mathbb{E}[\xi(h)]) \right) \\ &\leq (\tau - 2)c_D (\mathbb{E}[\xi(h)])^{\tau-1} \sum_{k=a/\mathbb{E}[\xi(h)]}^{b/\mathbb{E}[\xi(h)]} (k\mathbb{E}[\xi(h)])^{-(\tau-1)} \left(\xi(B + k\mathbb{E}[\xi(h)]) - \widehat{\beta}(B + k\mathbb{E}[\xi(h)]) \right), \end{aligned}$$

where we choose a and b such that $0 < a < b < \infty$. We use dominated convergence on the above sum. The summands are uniformly bounded, and $\mathbb{E}[\xi(h)] \rightarrow 0$ for both limits of interest. Further, when $k\mathbb{E}[\xi(h)] = y$, the summand converges pointwise to $y^{-(\tau-1)} \left(\xi(B + y) - \widehat{\beta}(B + y) \right)$. Hence, we can write the above sum as

$$(5.5.40) \quad \mathbb{E}[\xi(h)]^{-1} \left(\int_a^b y^{-(\tau-1)} \left(\xi(B + y) - \widehat{\beta}(B + y) \right) dy + o(1) \right),$$

where $o(1)$ is a function tending to zero for both limits of interest [176, 216 A]. The integrand is uniformly bounded for $y \in [a, b]$ and hence we can bound the integral from

above by a (negative) constant $-I$ for B sufficiently small and β sufficiently close to β_c . Hence,

$$(5.5.41) \quad \mathbb{E}[\xi(h)] \leq \widehat{\beta}B + \widehat{\beta}\nu\mathbb{E}[\xi(h)] - (\tau - 1)c_D I \mathbb{E}[\xi(h)]^{\tau-2}.$$

Logarithmic corrections in the bound for $\tau = 5$. We complete the proof by identifying the logarithmic correction for $\tau = 5$. Since the random variable in the expectation in T_1 is non-positive, we can bound

$$(5.5.42) \quad T_1 \leq \mathbb{E} \left[\xi(B + \Delta\mathbb{E}[\xi(h)]) - \widehat{\beta}(B + \Delta\mathbb{E}[\xi(h)]) \mathbb{1}_{\{\Delta \leq \varepsilon/\mathbb{E}[\xi(h)]\}} \right].$$

Taylor expansion $h \mapsto \xi(h)$ to third order, using that $\xi(0) = \xi''(0) = 0$, while the linear term cancels, leads to

$$(5.5.43) \quad T_1 \leq \mathbb{E} \left[\frac{\xi'''(\zeta)}{6} (B + \Delta\mathbb{E}[\xi(h)])^3 \mathbb{1}_{\{\Delta \leq \varepsilon/\mathbb{E}[\xi(h)]\}} \right],$$

for some $\zeta \in (0, \Delta\mathbb{E}[\xi(h)])$. On the event that $\Delta \leq \varepsilon/\mathbb{E}[\xi(h)]$, we thus have that $\zeta \in (0, \varepsilon)$, and $\xi'''(\zeta) \leq -c_\varepsilon \equiv \sup_{x \in (0, \varepsilon)} \xi'''(x) < 0$ when ε is sufficiently small. Thus, recalling the notation $(m)_k = m(m-1) \cdots (m-k+1)$ for the k th factorial of m ,

$$(5.5.44) \quad \begin{aligned} T_1 &\leq -\frac{c_\varepsilon}{6} \mathbb{E} \left[(B + \Delta\mathbb{E}[\xi(h)])^3 \mathbb{1}_{\{\Delta \leq \varepsilon/\mathbb{E}[\xi(h)]\}} \right] \\ &\leq -\frac{c_\varepsilon}{6} \mathbb{E}[\xi(h)]^3 \mathbb{E} \left[(\Delta)_3 \mathbb{1}_{\{\Delta \leq \varepsilon/\mathbb{E}[\xi(h)]\}} \right]. \end{aligned}$$

When $\tau = 5$, by Lemma 2.39 with $a = 4$, $\mathbb{E}[(\Delta)_3 \mathbb{1}_{\{\Delta \leq \ell\}}] \geq c_{3,5} \log \ell$, which completes the proof. \square

After having completed the upper bound on the first moment of $\xi(h)$ in Proposition 5.28, we continue with a lower bound:

PROPOSITION 5.29 (Lower bound on first moment of $\xi(h)$). *Let $\beta \geq \beta_c$ and $B > 0$. Then, there exists a constant $C_2 > 0$ such that*

$$(5.5.45) \quad \mathbb{E}[\xi(h)] \geq \widehat{\beta}B + \widehat{\beta}\nu\mathbb{E}[\xi(h)] - c_1\mathbb{E}[\xi(h)]^\delta - Be_1,$$

where

$$(5.5.46) \quad \delta = \begin{cases} 3 & \text{when } \mathbb{E}[D^4] < \infty, \\ \tau - 2 & \text{when } \tau \in (3, 5). \end{cases}$$

For $\tau = 5$,

$$(5.5.47) \quad \mathbb{E}[\xi(h)] \geq \widehat{\beta}B + \widehat{\beta}\nu\mathbb{E}[\xi(h)] - c_1\mathbb{E}[\xi(h)]^3 \log(1/\mathbb{E}[\xi(h)]) - Be_1.$$

PROOF. We again use the split in (5.5.26) and we bound T_1 and T_2 . Since the proof is quite similar to that of Proposition 5.28, we omit further details. \square

5.5.3. Determining the critical exponents β, δ and γ . We now have all the tools ready to derive the critical exponents for the configuration model. As for the Curie-Weiss model in the proof of Theorem 5.7, the proof will involve a Taylor expansion in the fixed-point equation (5.3.4) that determines the pressure. Since this recursion is quite a bit more involved than that for the Curie-Weiss model in (5.2.18), we have to work substantially harder. We start by analyzing β and δ .

Critical exponents β and δ . It remains to show that the bounds on $\mathbb{E}[\xi(h)]$ give us the desired result:

THEOREM 5.30 (Values of β and δ). *The critical exponents β and δ equal the values as stated in Theorem 5.24 when $\mathbb{E}[D^4] < \infty$ and $\tau \in (3, 5)$. Furthermore, for $\tau = 5$, (5.5.2) holds.*

PROOF. We prove the upper and the lower bounds separately, starting with the upper bound.

The upper bounds on the magnetization. We start by bounding the magnetization from above:

$$(5.5.48) \quad M(\beta, B) = \mathbb{E} \left[\tanh \left(B + \sum_{i=1}^D \xi(h_i) \right) \right] \leq B + \mathbb{E}[D] \mathbb{E}[\xi(h)].$$

We first perform the analysis for β . Taking the limit $B \searrow 0$ in (5.5.24) in Proposition 5.28 yields

$$(5.5.49) \quad \mathbb{E}[\xi(h_0)] \leq \widehat{\beta\nu} \mathbb{E}[\xi(h_0)] - C_1 \mathbb{E}[\xi(h_0)]^\delta,$$

where $h_0 = h(\beta, 0^+)$. For $\beta > \beta_c$, by definition, $\mathbb{E}[\xi(h_0)] > 0$ and thus we can divide through by $\mathbb{E}[\xi(h_0)]$ to obtain

$$(5.5.50) \quad \mathbb{E}[\xi(h_0)]^{\delta-1} \leq \frac{\widehat{\beta\nu} - 1}{C_1}.$$

By Taylor's theorem,

$$(5.5.51) \quad \widehat{\beta\nu} - 1 \leq \nu(1 - \widehat{\beta_c^2})(\beta - \beta_c).$$

Hence,

$$(5.5.52) \quad \mathbb{E}[\xi(h_0)] \leq \left(\frac{\nu(1 - \widehat{\beta_c^2})}{C_1} \right)^{1/(\delta-1)} (\beta - \beta_c)^{1/(\delta-1)}.$$

Using that $\beta = 1/(\delta - 1)$,

$$(5.5.53) \quad M(\beta, 0^+) \leq \mathbb{E}[D] \left(\frac{\nu(1 - \widehat{\beta_c^2})}{C_1} \right)^\beta (\beta - \beta_c)^\beta,$$

from which it easily follows that

$$(5.5.54) \quad \limsup_{\beta \searrow \beta_c} \frac{M(\beta, 0^+)}{(\beta - \beta_c)^\beta} < \infty.$$

We complete the analysis for β by analyzing $\tau = 5$. Since (5.5.24) also applies to $\tau = 5$, (5.5.54) holds as well. We now improve upon this using (5.5.25) in Proposition 5.28, which yields in a similar way as above that

$$(5.5.55) \quad \mathbb{E}[\xi(h_0)]^2 \leq \frac{\widehat{\beta\nu} - 1}{C_1 \log(1/\mathbb{E}[\xi(h_0)])}.$$

Since $x \mapsto 1/\log(1/x)$ is increasing on $(0, 1)$ and $\mathbb{E}[\xi(h_0)] \leq C(\beta - \beta_c)^{1/2}$ for some $C > 0$, we immediately obtain that

$$(5.5.56) \quad \mathbb{E}[\xi(h_0)]^2 \leq \frac{\widehat{\beta}\nu - 1}{C_1 \log(1/\mathbb{E}[\xi(h_0)])} \leq \frac{\widehat{\beta}\nu - 1}{C_1 \log(1/[C(\beta - \beta_c)^{1/2}])}.$$

Taking the limit of $\beta \searrow \beta_c$ as above then completes the proof.

We continue with the analysis for δ . Setting $\beta = \beta_c$ in (5.5.24) and rewriting gives

$$(5.5.57) \quad \mathbb{E}[\xi(h_c)] \leq \left(\frac{\widehat{\beta}_c}{C_1}\right)^{1/\delta} B^{1/\delta},$$

with $h_c = h(\beta_c, B)$. Hence,

$$(5.5.58) \quad M(\beta_c, B) \leq B + \mathbb{E}[D] \left(\frac{\widehat{\beta}_c}{C_1}\right)^{1/\delta} B^{1/\delta},$$

so that, using $1/\delta < 1$,

$$(5.5.59) \quad \limsup_{B \searrow 0} \frac{M(\beta_c, B)}{B^{1/\delta}} < \infty.$$

The analysis for δ for $\tau = 5$ can be performed in an identical way as for β .

The lower bounds on the magnetization. For the lower bound on the magnetization we use that

$$(5.5.60) \quad \frac{d^2}{dx^2} \tanh x = -2 \tanh x (1 - \tanh^2 x) \geq -2,$$

so that

$$(5.5.61) \quad \tanh x \geq x - x^2.$$

Hence,

$$(5.5.62) \quad \begin{aligned} M(\beta, B) &\geq B + \mathbb{E}[D]\mathbb{E}[\xi(h)] - \mathbb{E}\left[\left(B + \sum_{i=1}^D \xi(h_i)\right)^2\right] \\ &\geq B + \mathbb{E}[D]\mathbb{E}[\xi(h)] - Be_6 - \mathbb{E}[D(D-1)]\mathbb{E}[\xi(h)]^2 - \mathbb{E}[D]C_2\mathbb{E}[\xi(h)]^{2\wedge(\tau-2)} \\ &= B + (\mathbb{E}[D] - e_7)\mathbb{E}[\xi(h)] - Be_6, \end{aligned}$$

with $a \wedge b$ denoting the minimum of a and b , where e_7 satisfies (5.5.12) both also $\liminf_{\beta \searrow \beta_c} e_7(\beta, 0^+) = 0$, because $\mathbb{E}[\xi(h)]$ converges to zero for both limits of interest.

We again first perform the analysis for β and $\tau \neq 5$. We get from (5.5.45) in Proposition 5.29 that

$$(5.5.63) \quad \mathbb{E}[\xi(h_0)] \geq \left(\frac{\widehat{\beta}\nu - 1}{c_1}\right)^{1/(\delta-1)} \geq \left(\frac{\nu(1 - \widehat{\beta}^2)}{c_1}\right)^\beta (\beta - \beta_c)^\beta,$$

where the last inequality holds because, by Taylor's theorem,

$$(5.5.64) \quad \widehat{\beta}\nu - 1 \geq \nu(1 - \widehat{\beta}^2)(\beta - \beta_c).$$

Hence,

$$(5.5.65) \quad M(\beta, 0^+) \geq (\mathbb{E}[D] - e_7) \left(\frac{\nu(1 - \widehat{\beta}^2)}{c_1} \right)^\beta (\beta - \beta_c)^\beta,$$

so that

$$(5.5.66) \quad \liminf_{\beta \searrow \beta_c} \frac{M(\beta, 0^+)}{(\beta - \beta_c)^\beta} \geq \mathbb{E}[D] \left(\frac{\nu(1 - \widehat{\beta}^2)}{c_1} \right)^\beta > 0.$$

For $\tau = 5$, we note that (5.5.47) as well as the fact that $\log 1/x \leq A_\varepsilon x^{-\varepsilon}$ for all $x \in (0, 1)$ and some $A_\varepsilon > 0$, yields that

$$(5.5.67) \quad \mathbb{E}[\xi(h_0)] \geq \left(\frac{\widehat{\beta}\nu - 1}{A_\varepsilon c_1} \right)^{1/(2+\varepsilon)} \geq \left(\frac{\nu(1 - \widehat{\beta}^2)}{A_\varepsilon c_1} \right)^{1/(2+\varepsilon)} (\beta - \beta_c)^{1/(2+\varepsilon)}.$$

Then again using (5.5.47) yields, for some constant $c > 0$,

$$(5.5.68) \quad \mathbb{E}[\xi(h_0)] \geq \left(\frac{\widehat{\beta}\nu - 1}{c_1 \log(1/\mathbb{E}[\xi(h_0)])} \right)^{1/2} \geq c \left(\frac{\beta - \beta_c}{\log(1/(\beta - \beta_c))} \right)^{1/2},$$

once more since $x \mapsto 1/(\log(1/x))$ is increasing.

We continue with the analysis for δ . Again, setting $\beta = \beta_c$ in (5.5.45), we get

$$(5.5.69) \quad \mathbb{E}[\xi(h_c)] \geq \left(\frac{\widehat{\beta}_c - e_1}{c_1} \right)^{1/\delta} B^{1/\delta},$$

from which it follows that

$$(5.5.70) \quad M(\beta_c, B) \geq (\mathbb{E}[D] - e_7) \left(\frac{\widehat{\beta}_c - e_1}{c_1} \right)^{1/\delta} B^{1/\delta} - B e_6,$$

and hence,

$$(5.5.71) \quad \liminf_{B \searrow 0} \frac{M(\beta_c, B)}{B^{1/\delta}} \geq \mathbb{E}[D] \left(\frac{\widehat{\beta}_c}{c_1} \right)^{1/\delta} > 0,$$

as required. The extension to $\tau = 5$ can be dealt with in an identical way as in (5.5.67)–(5.5.68). This proves the theorem. \square

We leave the proof of the joint scaling of the magnetization as $(\beta, B) \searrow (\beta_c, 0)$ as an exercise:

EXERCISE 5.29 (Joint critical exponents magnetization). *Prove the joint scaling of the magnetization in Corollary 5.25.*

We next study the susceptibility, and identify γ :

The critical exponent γ . For the susceptibility in the *subcritical* phase, i.e., in the high-temperature region $\beta < \beta_c$, we can not only identify the critical exponent γ , but we can also identify the constant:

THEOREM 5.31 (Critical exponent γ). *For $\mathbb{E}[\Delta] < \infty$ and $\beta < \beta_c$,*

$$(5.5.72) \quad \chi(\beta, 0^+) = 1 + \frac{\mathbb{E}[D]\widehat{\beta}}{1 - \nu\widehat{\beta}}.$$

In particular,

$$(5.5.73) \quad \lim_{\beta \nearrow \beta_c} \chi(\beta, 0^+) (\beta_c - \beta) = \frac{\mathbb{E}[D]\widehat{\beta}_c^2}{1 - \widehat{\beta}_c^2},$$

and $\gamma = 1$.

PROOF. The proof is divided into three steps. We first reduce the susceptibility on the random graph to the one on the branching process tree. Secondly, we rewrite the susceptibility on the tree using transfer matrix techniques. Finally, we use this rewrite (which applies to *all* β and $B > 0$) to prove that $\gamma = 1$.

Reduction to the random tree. Let \emptyset denote a vertex selected uniformly at random from $[n]$ and let \mathbb{E}_\emptyset denote its expectation. Then we can write the susceptibility as

$$(5.5.74) \quad \chi_n \equiv \frac{1}{n} \sum_{i,j=1}^n \left(\langle \sigma_i \sigma_j \rangle_{\mu_n} - \langle \sigma_i \rangle_{\mu_n} \langle \sigma_j \rangle_{\mu_n} \right) = \mathbb{E}_\emptyset \left[\sum_{j=1}^n \left(\langle \sigma_\emptyset \sigma_j \rangle_{\mu_n} - \langle \sigma_\emptyset \rangle_{\mu_n} \langle \sigma_j \rangle_{\mu_n} \right) \right].$$

Note that

$$(5.5.75) \quad \langle \sigma_i \sigma_j \rangle_{\mu_n} - \langle \sigma_i \rangle_{\mu_n} \langle \sigma_j \rangle_{\mu_n} = \frac{\partial \langle \sigma_i \rangle_{\mu_n}}{\partial B_j},$$

which is, by the GHS inequality in Lemma 5.5, decreasing in external fields at all other vertices $k \in [n]$. Denote by $\langle \cdot \rangle^{t,+/f}$ the Ising model with $+/\text{free}$ boundary conditions, respectively, at all vertices at graph distance t from \emptyset . Then, for all $t \geq 1$,

$$(5.5.76) \quad \chi_n \geq \mathbb{E}_\emptyset \left[\sum_{j=1}^n \left(\langle \sigma_\emptyset \sigma_j \rangle_{\mu_n}^{t,+} - \langle \sigma_\emptyset \rangle_{\mu_n}^{t,+} \langle \sigma_j \rangle_{\mu_n}^{t,+} \right) \right].$$

By introducing boundary conditions, only vertices in the ball $B_\emptyset(t)$ contribute to the sum. Hence, by taking the limit $n \rightarrow \infty$ and using that the graph is locally tree-like,

$$(5.5.77) \quad \chi \geq \mathbb{E} \left[\sum_{j \in T_t} \left(\langle \sigma_\emptyset \sigma_j \rangle^{t,+} - \langle \sigma_\emptyset \rangle^{t,+} \langle \sigma_j \rangle^{t,+} \right) \right],$$

where the expectation now is over the first t generations of a branching process tree $T_t \sim \text{BP}_t$ with root \emptyset .

For an upper bound on χ_n we use a trick similar to one used in the proof of [92, Corollary 4.5]: Let $B'_j = B$ if $j \in B_t(\emptyset)$ and $B'_j = B + H$ if $j \notin B_t(\emptyset)$ for some $H > -B$. Denote by $\langle \cdot \rangle_H$ the associated Ising expectation. Then, because of (5.5.75),

$$(5.5.78) \quad \mathbb{E}_\emptyset \left[\sum_{j \notin B_t(\emptyset)} \left(\langle \sigma_\emptyset \sigma_j \rangle - \langle \sigma_\emptyset \rangle \langle \sigma_j \rangle \right) \right] = \mathbb{E}_\emptyset \left[\frac{\partial}{\partial H} \langle \sigma_\emptyset \rangle_H \Big|_{H=0} \right],$$

By the GHS inequality (Lemma 5.5), $\langle \sigma_\emptyset \rangle_H$ is a concave function of H and hence,

$$(5.5.79) \quad \mathbb{E}_\emptyset \left[\left. \frac{\partial}{\partial H} \langle \sigma_\emptyset \rangle_H \right|_{H=0} \right] \leq \mathbb{E}_\emptyset \left[\frac{2}{B} (\langle \sigma_\emptyset \rangle_{H=0} - \langle \sigma_\emptyset \rangle_{H=-B/2}) \right].$$

Using the GKS inequality (Lemma 5.4), this can be bounded from above by

$$(5.5.80) \quad \mathbb{E}_\emptyset \left[\frac{2}{B} (\langle \sigma_\emptyset \rangle_{H=0}^{t,+} - \langle \sigma_\emptyset \rangle_{H=-B/2}^{t,f}) \right] = \mathbb{E}_\emptyset \left[\frac{2}{B} (\langle \sigma_\emptyset \rangle^{t,+} - \langle \sigma_\emptyset \rangle^{t,f}) \right],$$

where the equality holds because the terms depend only on the system in the ball $B_t(\emptyset)$ and hence not on H . By letting $n \rightarrow \infty$, by the local weak convergence to a branching process in Theorem 2.11, this is equal to

$$(5.5.81) \quad \frac{2}{B} \mathbb{E} [(\langle \sigma_\emptyset \rangle^{t,+} - \langle \sigma_\emptyset \rangle^{t,f})],$$

where the expectation and the Ising model now is over the unimodular Galton-Watson tree \mathbf{BP} with root offspring distribution $(p_k)_{k \geq 1}$ with root \emptyset . Let \mathbf{BP}_t denote its for t generations. By Lemma 5.15, we know that this expectation can be bounded from above by M/t for some constant $M = M(\beta, B) < \infty$. Hence, if $t \rightarrow \infty$,

$$(5.5.82) \quad \lim_{t \rightarrow \infty} \mathbb{E} \left[\sum_{j \in \mathbf{BP}_t} (\langle \sigma_\emptyset \sigma_j \rangle^{t,+} - \langle \sigma_\emptyset \rangle^{t,+} \langle \sigma_j \rangle^{t,+}) \right] \leq \chi \leq \lim_{t \rightarrow \infty} \mathbb{E} \left[\sum_{j \in \mathbf{BP}_t} (\langle \sigma_\emptyset \sigma_j \rangle^{t,f} - \langle \sigma_\emptyset \rangle^{t,f} \langle \sigma_j \rangle^{t,f}) \right].$$

Rewrite of the susceptibility on trees. It remains to study the susceptibility on trees. For this, condition on the branching process tree \mathbf{BP} . Then, for some vertex j at height $\ell \leq t$ in the tree, denote the vertices on the unique path from \emptyset to j by $\emptyset = v_0, v_1, \dots, v_\ell = j$ and let, for $0 \leq i \leq \ell$, $S_{\leq i} = (\sigma_{v_0}, \dots, \sigma_{v_i})$. We first compute the expected value of a spin σ_{v_i} on this path, conditioned on the spin values $S_{\leq i-1}$. Note that under this conditioning the expected spin value only depends on the spin value $\sigma_{v_{i-1}}$ and the effective field $h_{v_i} = h_{v_i}^{t,+/f}$ obtained by pruning the tree at vertex v_i , i.e., by removing all edges at vertex v_i going away from the root and replacing the external magnetic field at vertex v_i by h_{v_i} which can be exactly computed using Lemma 5.12. Hence,

$$(5.5.83) \quad \langle \sigma_{v_i} | S_{\leq i-1} \rangle^{t,+/f} = \frac{e^{\beta \sigma_{v_{i-1}} + h_{v_i}} - e^{-\beta \sigma_{v_{i-1}} - h_{v_i}}}{e^{\beta \sigma_{v_{i-1}} + h_{v_i}} + e^{-\beta \sigma_{v_{i-1}} - h_{v_i}}}.$$

We can write the indicators $\mathbb{1}_{\{\sigma_{v_{i-1}} = \pm 1\}} = \frac{1}{2}(1 \pm \sigma_{v_{i-1}})$, so that the above equals

$$(5.5.84) \quad \begin{aligned} & \frac{1}{2}(1 + \sigma_{v_{i-1}}) \frac{e^{\beta + h_{v_i}} - e^{-\beta - h_{v_i}}}{e^{\beta + h_{v_i}} + e^{-\beta - h_{v_i}}} + \frac{1}{2}(1 - \sigma_{v_{i-1}}) \frac{e^{-\beta + h_{v_i}} - e^{\beta - h_{v_i}}}{e^{-\beta + h_{v_i}} + e^{\beta - h_{v_i}}} \\ &= \sigma_{v_{i-1}} \frac{1}{2} \left(\frac{e^{\beta + h_{v_i}} - e^{-\beta - h_{v_i}}}{e^{\beta + h_{v_i}} + e^{-\beta - h_{v_i}}} - \frac{e^{-\beta + h_{v_i}} - e^{\beta - h_{v_i}}}{e^{-\beta + h_{v_i}} + e^{\beta - h_{v_i}}} \right) + \frac{1}{2} \left(\frac{e^{\beta + h_{v_i}} - e^{-\beta - h_{v_i}}}{e^{\beta + h_{v_i}} + e^{-\beta - h_{v_i}}} + \frac{e^{-\beta + h_{v_i}} - e^{\beta - h_{v_i}}}{e^{-\beta + h_{v_i}} + e^{\beta - h_{v_i}}} \right). \end{aligned}$$

By pairwise combining the terms over a common denominator the above equals

$$(5.5.85) \quad \sigma_{v_{i-1}} \frac{1}{2} \frac{(e^{\beta+h_{v_i}} - e^{-\beta-h_{v_i}})(e^{-\beta+h_{v_i}} + e^{\beta-h_{v_i}}) - (e^{-\beta+h_{v_i}} - e^{\beta-h_{v_i}})(e^{\beta+h_{v_i}} + e^{-\beta-h_{v_i}})}{(e^{\beta+h_{v_i}} + e^{-\beta-h_{v_i}})(e^{-\beta+h_{v_i}} + e^{\beta-h_{v_i}})} + \frac{1}{2} \frac{(e^{\beta+h_{v_i}} - e^{-\beta-h_{v_i}})(e^{-\beta+h_{v_i}} + e^{\beta-h_{v_i}}) + (e^{-\beta+h_{v_i}} - e^{\beta-h_{v_i}})(e^{\beta+h_{v_i}} + e^{-\beta-h_{v_i}})}{(e^{\beta+h_{v_i}} + e^{-\beta-h_{v_i}})(e^{-\beta+h_{v_i}} + e^{\beta-h_{v_i}})}.$$

By expanding all products, this equals, after cancellations,

$$(5.5.86) \quad \sigma_{v_{i-1}} \frac{e^{2\beta} + e^{-2\beta}}{e^{2\beta} + e^{-2\beta} + e^{2h_{v_i}} + e^{-2h_{v_i}}} + \frac{e^{2h_{v_i}} + e^{-2h_{v_i}}}{e^{2\beta} + e^{-2\beta} + e^{2h_{v_i}} + e^{-2h_{v_i}}} = \sigma_{v_{i-1}} \frac{\sinh(2\beta)}{\cosh(2\beta) + \cosh(2h_{v_i})} + \frac{\sinh(2h_{v_i})}{\cosh(2\beta) + \cosh(2h_{v_i})}.$$

This gives

$$(5.5.87) \quad \langle \sigma_{v_\ell} \rangle^{t,+/f} = \langle \langle \sigma_{v_\ell} | S_{\leq \ell-1} \rangle^{t,+/f} \rangle^{t,+/f} = \langle \sigma_{v_{\ell-1}} \rangle^{t,+/f} \frac{\sinh(2\beta)}{\cosh(2\beta) + \cosh(2h_{v_\ell})} + \frac{\sinh(2h_{v_\ell})}{\cosh(2\beta) + \cosh(2h_{v_\ell})}.$$

Applying this recursively, we get

$$(5.5.88) \quad \langle \sigma_{v_\ell} \rangle^{t,+/f} = \langle \sigma_{v_0} \rangle^{t,+/f} \prod_{i=1}^{\ell} \frac{\sinh(2\beta)}{\cosh(2\beta) + \cosh(2h_{v_i})} + \sum_{i=1}^{\ell} \left(\frac{\sinh(2h_{v_i})}{\cosh(2\beta) + \cosh(2h_{v_i})} \prod_{k=i+1}^{\ell} \frac{\sinh(2\beta)}{\cosh(2\beta) + \cosh(2h_{v_k})} \right).$$

Similarly,

$$(5.5.89) \quad \langle \sigma_{v_0} \sigma_{v_\ell} \rangle^{t,+/f} = \left\langle \sigma_{v_0} \left(\sigma_{v_0} \prod_{i=1}^{\ell} \frac{\sinh(2\beta)}{\cosh(2\beta) + \cosh(2h_{v_i})} + \sum_{i=1}^{\ell} \left(\frac{\sinh(2h_{v_i})}{\cosh(2\beta) + \cosh(2h_{v_i})} \prod_{k=i+1}^{\ell} \frac{\sinh(2\beta)}{\cosh(2\beta) + \cosh(2h_{v_k})} \right) \right) \right\rangle^{t,+/f} = \prod_{i=1}^{\ell} \frac{\sinh(2\beta)}{\cosh(2\beta) + \cosh(2h_{v_i})} + \langle \sigma_{v_0} \rangle^{t,+/f} \sum_{i=1}^{\ell} \left(\frac{\sinh(2h_{v_i})}{\cosh(2\beta) + \cosh(2h_{v_i})} \prod_{k=i+1}^{\ell} \frac{\sinh(2\beta)}{\cosh(2\beta) + \cosh(2h_{v_k})} \right).$$

Combining the above yields

$$(5.5.90) \quad \langle \sigma_{v_0} \sigma_{v_\ell} \rangle^{t,+/f} - \langle \sigma_{v_0} \rangle^{t,+/f} \langle \sigma_{v_\ell} \rangle^{t,+/f} = \left(1 - (\langle \sigma_{v_0} \rangle^{t,+/f})^2 \right) \prod_{i=1}^{\ell} \frac{\sinh(2\beta)}{\cosh(2\beta) + \cosh(2h_{v_i})}.$$

By taking the limit $t \rightarrow \infty$, we obtain

$$(5.5.91) \quad \langle \sigma_{v_0} \sigma_{v_\ell} \rangle - \langle \sigma_{v_0} \rangle \langle \sigma_{v_\ell} \rangle = \left(1 - (\langle \sigma_{v_0} \rangle)^2 \right) \prod_{i=1}^{\ell} \frac{\sinh(2\beta)}{\cosh(2\beta) + \cosh(2h_{v_i})},$$

where now the expectation is with respect to the Ising model on the infinite tree \mathbf{BP} . Note that the above formula is true almost surely in \mathbf{BP} .

We take the expectation over \mathbf{BP} to obtain

$$(5.5.92) \quad \chi = \mathbb{E} \left[\sum_{j \in \mathbf{BP}} (1 - \langle \sigma_{v_0} \rangle^2) \prod_{i=1}^{|j|} \frac{\sinh(2\beta)}{\cosh(2\beta) + \cosh(2h_{v_i})} \right].$$

Finally, we can rewrite

$$(5.5.93) \quad \frac{\sinh(2\beta)}{\cosh(2\beta) + \cosh(2h_{v_i})} = \frac{2 \sinh(\beta) \cosh(\beta)}{2 \cosh(\beta)^2 - 1 + \cosh(2h_{v_i})} = \frac{\widehat{\beta}}{1 + \frac{\cosh(2h_{v_i}) - 1}{2 \cosh(\beta)^2}},$$

so that

$$(5.5.94) \quad \chi(\beta, B) = \mathbb{E} \left[(1 - \langle \sigma_{v_0} \rangle^2) \sum_{w \in \mathbf{BP}} \widehat{\beta}^{|w|} \prod_{i=1}^{|w|} \left(1 + \frac{\cosh(2h_{v_i}) - 1}{2 \cosh(\beta)^2} \right)^{-1} \right].$$

The rewrite in (5.5.94) is valid for all β and $B > 0$, and provides the starting point for all our results on the susceptibility.

Identification of the susceptibility for $\beta < \beta_c$. In order to prove (5.5.72), we perform a branching process computation. We take the limit $B \searrow 0$, for $\beta < \beta_c$, and apply dominated convergence. First of all, all fields h_i converge to zero by the definition of β_c , so we have pointwise convergence. Secondly, $1 + \frac{\cosh(2h_{v_i}) - 1}{2 \cosh(\beta)^2} \geq 1$, so that the random variable in the expectation is bounded from above by $\sum_{w \in \mathbf{BP}} \widehat{\beta}^{|w|}$, which has finite expectation as we show below. Thus, by dominated convergence, the above converges to

$$(5.5.95) \quad \lim_{B \searrow 0} \chi(\beta, B) = \mathbb{E} \left[\sum_{w \in \mathbf{BP}} \widehat{\beta}^{|w|} \right].$$

Denote by Z_ℓ the number of vertices at distance ℓ from the root. Then,

$$(5.5.96) \quad \mathbb{E} \left[\sum_{w \in \mathbf{BP}} \widehat{\beta}^{|w|} \right] = \mathbb{E} \left[\sum_{\ell=0}^{\infty} Z_\ell \widehat{\beta}^\ell \right] = \sum_{\ell=0}^{\infty} \mathbb{E}[Z_\ell] \widehat{\beta}^\ell,$$

because $Z_\ell \geq 0$, a.s. Note that $Z_\ell / (\mathbb{E}[D] \nu^{\ell-1})$ is a martingale, because the offspring of the root has expectation $\mathbb{E}[D]$ and all other vertices have expected offspring ν . Hence,

$$(5.5.97) \quad \lim_{B \searrow 0} \chi(\beta, B) = \sum_{\ell=0}^{\infty} \mathbb{E}[Z_\ell] \widehat{\beta}^\ell = 1 + \sum_{\ell=1}^{\infty} \mathbb{E}[D] \nu^{\ell-1} \widehat{\beta}^\ell = 1 + \frac{\mathbb{E}[D] \widehat{\beta}}{1 - \widehat{\beta} \nu}.$$

This proves (5.5.72). We continue to prove (5.5.73), which follows by using (5.5.51) and (5.5.64):

$$(5.5.98) \quad 1 + \frac{\mathbb{E}[D] \widehat{\beta}}{\nu(1 - \widehat{\beta}^2)} (\beta_c - \beta)^{-1} \leq 1 + \frac{\mathbb{E}[D] \widehat{\beta}}{1 - \widehat{\beta} \nu} \leq 1 + \frac{\mathbb{E}[D] \widehat{\beta}}{\nu(1 - \widehat{\beta}^2)} (\beta_c - \beta)^{-1}.$$

This completes the proof. \square

We close this section stating an open problem about γ' . In [110], we proved a one-sided bound, that $\gamma' \geq 1$. We believe that $\gamma' = 1$, but do not know how to prove the upper bound:

OPEN PROBLEM 5.1 (Identification of γ'). *Show that $\gamma' = 1$.*

5.6. Central limit theorem for the non-critical total spin

In this section, we prove a central limit theorem for the total spin, which is valid throughout the uniqueness regime. We follow the work with Giardinà, Giberti and Prioriello [137]. For the Ising model, the spins are quite dependent, which might destroy the central limit theorem as observed for independent random variables. However, in the uniqueness regime, it turns out that correlations between spins of far away vertices decay fast, thus creating an ‘effective’ short range interaction for which the central limit theorem remains valid.

In order to state the central limit theorem for the total spin, we define the *uniqueness regime* as the set of parameters (β, B) satisfying

$$(5.6.1) \quad \mathcal{U}^{\text{qu}} := \{(\beta, B) : \beta \geq 0, B \neq 0 \text{ or } 0 < \beta < \beta_c^{\text{qu}}, B = 0\}.$$

Then, the central limit theorem for the total spin reads as follows:

THEOREM 5.32 (Central limit theorem for the total spin). *Consider $\text{CM}_n(\mathbf{d})$ where the degree distribution satisfies Conditions 1.6(a)-(b), and has a strongly finite mean as in (5.3.1). For all $(\beta, B) \in \mathcal{U}^{\text{qu}}$,*

$$(5.6.2) \quad \frac{S_n - nM_n}{\sqrt{n}} \xrightarrow{d} Z \quad \text{as } n \rightarrow \infty,$$

where $\chi = \chi(\beta, B)$ is the susceptibility defined in (5.3.8) and $Z \sim \mathcal{N}(0, \chi)$ denotes a centered Gaussian random variable with variance χ .

We note that the variance of the total spin in the central limit theorem is given by another important physical thermodynamic quantity, namely, the susceptibility. This is a general feature, as also outlined for the Curie-Weiss model in Theorem 5.3. Our proof will again highlight why this occurs more clearly. Theorem 5.32, together with the fact that the susceptibility $\chi(\beta, B)$ can be arbitrarily large by Theorem 5.21. Should the spins all be independent, we would have that $\text{Var}(S_n) = n[1 - M(\beta, B)^2]$:

EXERCISE 5.30 (Variance spin variables). *Prove that $\lim_{n \rightarrow \infty} \text{Var}(\sigma_i) = [1 - M(\beta, B)^2]$, and conclude that $\text{Var}(S_n) = n[1 - M(\beta, B)^2](1 + o(1))$ when the spins would all be independent.*

We conclude that, particularly close to the critical point $(\beta_c, 0)$, the variance of the total spin for the Ising model on the configuration model is much larger than it would be for a sum of i.i.d. random variables. This is due to the strong interactions between spins, an effect that becomes more pronounced when (β, B) approaches $(\beta_c, 0)$.

We follow the proof strategy for the convergence of the susceptibility in Theorem 5.3(c). This proof is remarkably simple and general. We consider the so-called *cumulant generating functions* of S_n , given by

$$(5.6.3) \quad c_n(t) = \frac{1}{n} \log \mu_{G_n} [\exp(tS_n)].$$

Taking the second derivative of (5.6.3) and evaluating it in zero gives $\text{Var}_{\mu_{G_n}}(S_n)/n$:

EXERCISE 5.31 (Moments and cumulant generating function). *Prove that $c'_n(0) = M_n(\beta, B)$ and $c''_n(0) = \text{Var}_{\mu_{G_n}}(S_n)/n$.*

The crucial argument in the proof of Theorem 5.32 will be to show the existence of these variances in the limit $n \rightarrow \infty$. This in turn will be a consequence of the existence of the limit for the sequence $(c_n(t))_{n \geq 1}$ and the sequence of its derivatives. The existence of the limit $c(t) := \lim_{n \rightarrow \infty} c_n(t)$ is a simple consequence of the existence of the pressure. Let us now give some more details.

We give the proof for $B \geq 0$ only, the case $B < 0$ is handled similarly. The strategy of the proof is to show that the moment generating function of

$$(5.6.4) \quad \bar{S}_n = \frac{S_n - nM_n(\beta, B)}{\sqrt{n}}.$$

converges to the moment generating function of a Gaussian random variable with variance $\chi(\beta, B)$ given in (5.3.8), i.e., for all $t \in [0, \alpha)$ and some $\alpha > 0$,

$$(5.6.5) \quad \lim_{n \rightarrow \infty} \mu_{G_n}(\exp(t\bar{S}_n)) = \exp(\chi(\beta, B)t^2/2).$$

This can be done by expressing $\mu_{G_n}(\exp(t\bar{S}_n))$ in terms of the second derivative of the cumulant generating function $c_n(t)$ defined in (5.6.3). A simple computation shows that

$$(5.6.6) \quad c'_n(t) = \frac{1}{n} \frac{\mu_{G_n}(S_n \exp(tS_n))}{\mu_{G_n}(\exp(tS_n))} = \mu_{G_n(\beta, B+t)}\left(\frac{S_n}{n}\right),$$

where, in order to stress the dependence on the magnetic field, we used the symbol $\mu_{G_n(\beta, B+t)}(\cdot)$ to denote the μ_{G_n} -average in the presence of the field $B + t$. Note that $c'_n(0) = \frac{1}{n} \mu_{G_n}(S_n) = M_n(\beta, B)$ by Exercise 5.31. Further, again by Exercise 5.31,

$$(5.6.7) \quad c''_n(t) = \text{Var}_{\mu_{G_n(\beta, B+t)}}\left(\frac{S_n}{n}\right).$$

By (5.3.8) in Theorem 5.10, as $n \rightarrow \infty$,

$$(5.6.8) \quad c''_n(0) = \chi_n(\beta, B) \rightarrow \chi(\beta, B).$$

Let us take $t > 0$ and set $t_n = t/\sqrt{n}$. Using the fact that $c_n(0) = 0$ and applying the Taylor's theorem with Lagrange remainder, we obtain

$$\begin{aligned} \log \mu_{G_n}(\exp(t\bar{S}_n)) &= \log \mu_{G_n}\left(\exp\left(\frac{tS_n - tnM_n(\beta, B)}{\sqrt{n}}\right)\right) \\ &= \log[\mu_{G_n}(\exp(t_n S_n))] - t\sqrt{n}M_n(\beta, B) \\ &= n[c_n(t_n) - t_n c'_n(0)] = n \frac{t_n^2}{2} c''_n(t_n^*) = \frac{t^2}{2} c''_n(t_n^*), \end{aligned}$$

for some $t_n^* \in [0, t_n]$. In order to control the limiting behavior of $c''_n(t_n^*)$, we exploit the following property of $c_n(t)$:

PROPOSITION 5.33 (Uniform convergence of second derivatives). *Under the conditions of Theorem 5.32 and for $B \geq 0$, $\lim_{n \rightarrow \infty} c''_n(t_n) = \chi(\beta, B)$ for all $t \geq -B$.*

Proof of Theorem 5.32. Proposition 5.33 immediately implies that

$$(5.6.9) \quad \lim_{n \rightarrow \infty} \log \mu_{G_n} \left(e^{t\bar{S}_n} \right) = \frac{1}{2} \chi(\beta, B) t^2,$$

which proves Theorem 5.32. \square .

The remainder of this section is devoted to the proof of Proposition 5.33. It relies on the concavity of the functions $c'_n(t)$ as stated in the following lemma:

LEMMA 5.34 (Concavity of magnetization). *For $B \geq 0$, $c'_n(t)$ is concave on $[0, \infty)$.*

PROOF. For $B \geq 0$, the concavity of $c'_n(t)$ on $[0, \infty)$ can be obtained by observing that this function is the magnetization per particle w.r.t. μ_{G_n} in presence of the magnetic field $B + t$, see (5.6.6), and then by applying the GHS inequality. Indeed, for $t \geq -B$, the GHS inequality (Lemma 5.5) implies that

$$(5.6.10) \quad \frac{\partial^2}{\partial t^2} c'_n(t) = \frac{1}{n} \sum_{i \in [n]} \frac{\partial^2}{\partial B^2} \mu_{G_n(\beta, B+t)}(\sigma_i) \leq 0.$$

We conclude that $c'_n(t)$ is concave on $[-B, \infty)$ for $B \geq 0$. \square

The proof of Proposition 5.33 requires the following lemma that is proven by Ellis in [121, Lemma V.7.5]:

LEMMA 5.35. *Let $(f_n)_{n \geq 1}$ be a sequence of convex functions on an open interval A of \mathbb{R} such that $f(t) = \lim_{n \rightarrow \infty} f_n(t)$ exists for every $t \in A$. Let $(t_n)_{n \geq 1}$ be a sequence in A which converges to a point $t_0 \in A$. If $f'_n(t_n)$ exists for every $n \geq 1$ and $f'(t_0)$ exists, then $\lim_{n \rightarrow \infty} f'_n(t_n)$ exists and equals $f'(t_0)$.*

PROOF OF PROPOSITION 5.33. In the following proof we use subscripts to denote the dependence on β and B of the functions $c_n(t)$ and $c(t)$. We first consider $B > 0$. According to Lemma 5.34, each function $-c'_{n,\beta,B}(t)$ is convex on the interval $[-B, \infty)$, which contains the origin in its interior since $B > 0$. We first prove that $-c'_{\beta,B}(t) = \lim_{n \rightarrow \infty} (-c'_{n,\beta,B}(t))$ exists. By Hölder's inequality, $t \mapsto c_{n,\beta,B}(t)$ is a convex function on \mathbb{R} and by the existence of the pressure in Theorem 5.9, as $n \rightarrow \infty$,

$$(5.6.11) \quad c_{n,\beta,B}(t) \longrightarrow c_{\beta,B}(t) = \psi(\beta, B+t) - \psi(\beta, B).$$

For $B > 0$, the derivative $\frac{\partial}{\partial B} \psi(\beta, B_0)$ exists for all B_0 in some neighborhood of B . Hence there exists $\alpha > 0$ such that $c'(t) = \frac{\partial}{\partial B} \psi(\beta, B+t)$ exists for all $-\alpha < t < \alpha$. Now Lemma 5.35 implies that, as $n \rightarrow \infty$,

$$(5.6.12) \quad c'_{n,\beta,B}(t) \longrightarrow c'_{\beta,B}(t) \quad \text{for all } -\alpha < t < \alpha.$$

By (5.3.8) in Theorem 5.10, $\frac{\partial^2}{\partial B^2} \psi(\beta, B) = \chi(\beta, B)$, so that

$$(5.6.13) \quad c''_{\beta,B}(0) = \frac{\partial^2}{\partial B^2} \psi(\beta, B) = \chi(\beta, B).$$

This completes the proof of Proposition 5.33 for $B > 0$, because Lemma 5.35 implies that for any $0 \leq t_n < \alpha$ with $t_n \rightarrow 0$, as $n \rightarrow \infty$,

$$(5.6.14) \quad -c''_{n,\beta,B}(t_n) \longrightarrow -c''_{\beta,B}(0) = -\chi(\beta, B).$$

The proof of Proposition 5.33 for $B > 0$ made use of the fact that all the functions $-c'_{n,\beta,B}(t)$ are convex on an interval that contains the origin in its interior. But for $B = 0$, $-c'_{n,\beta,B}(t)$ is convex on $[0, \infty)$ and by symmetry is concave on $(-\infty, 0]$. Hence the above proof must be modified. We fix $0 < \beta < \beta_c^{\text{qu}}$ and assume that $\chi(\beta, 0)$ is finite. Since for $0 < \beta < \beta_c^{\text{qu}}$, $\frac{\partial}{\partial B}\psi(\beta, B)$ exists for all B real, $c_{\beta,0}(t) = \psi(\beta, t) - \psi(\beta, 0)$ is differentiable for all t real. We define new functions

$$(5.6.15) \quad h_n(t) = \begin{cases} -c'_{n,\beta,0}(t) & \text{for } t \geq 0, \\ -c''_{n,\beta,0}(0) \cdot t & \text{for } t < 0, \end{cases} \quad h(t) = \begin{cases} -c'_{\beta,0}(t) & \text{for } t \geq 0, \\ -\chi(\beta, 0) \cdot t & \text{for } t < 0. \end{cases}$$

Then h_n is continuous at 0 ($c'_{n,\beta,0}(0) = M(\beta, 0) = 0$) and is convex on \mathbb{R} . By (5.6.8), $c''_{n,\beta,B}(0) \rightarrow \chi(\beta, B)$, and so $h_n(t) \rightarrow h(t)$ for all t real. Since $\chi(\beta, 0)$ is assumed to be finite, $h'(0) = -c''_{\beta,0}(0) = -\chi(\beta, 0)$. Lemma 5.35 implies that, for any $t_n \geq 0$ with $t_n \rightarrow 0$, as $n \rightarrow \infty$,

$$(5.6.16) \quad h'_n(t_n) = -c''_{n,\beta,0}(t_n) \rightarrow h'(0) = -\chi(\beta, 0).$$

This proves Proposition 5.33 for $B = 0$ in the uniqueness regime. \square

5.7. Non-classical limit theorems for the critical total spin

In this section, we investigate non-classical limit theorems for the total spin at the critical point $(\beta, B) = (\beta_c, 0)$. While central limit theorems for the total spin as in Theorem 5.32 are quite robust, and hold in great generality, the behavior at the critical point depends sensitively on the precise setting. On random graphs, we will see that the critical behavior is sensitive to the level of inhomogeneity in the model, as described by the limiting degree distribution. We have already seen this effect in the critical exponents in Theorem 5.24, now we will investigate the critical behavior of the total spin. Unfortunately, such results cannot be derived easily, and we will go to another setting where such results can be obtained. We follow Dommers et al. [108] and Giardinà et al. [138].

This section is organised as follows. In Section 5.7.1, we investigate the annealed Ising model on rank-1 inhomogeneous random graphs. In Section 5.7.1, we consider the configuration model, where results are more difficult and restricted to strong conditions on the degree distribution. One setting that can be handled completely is the setting of random regular graphs of arbitrary degree.

5.7.1. Annealed Ising model on rank-1 inhomogeneous random graphs. In this section, we investigate the annealed Ising model on rank-1 inhomogeneous random graphs. In these models, due to the fact that edges are present *independently*, the annealed partition function can be computed explicitly. We start by explaining that the *quenched* rank-1 inhomogeneous random graph also fits within the settings described so far.

Extension to rank-1 inhomogeneous random graphs. As we explain in more detail in Section 5.8 below, the results in Sections 5.3, 5.5 and 5.6 apply more generally than only for the configuration model. Indeed, they hold when the random graph is homogeneously locally tree-like with a limiting distribution that has a finite $(1 + \varepsilon)$ moment. See Section 5.8 for more details. In particular, they apply to the rank-1 setting. Here, we focus on the generalized random graph $\text{GRG}_n(\mathbf{w})$, where the weight distribution satisfies

Condition 1.1(a)-(b). Note that the fact that the results in Sections 5.3, 5.5 and 5.6 apply here can also be deduced from the relation between $\text{CM}_n(\mathbf{d})$ and $\text{GRG}_n(\mathbf{w})$ in this case (recall Theorem 2.15).

In this case, it turns out that the quenched critical value is given by

$$(5.7.1) \quad \beta_c^{\text{qu}} = \text{atanh}(1/\nu),$$

where ν is defined in (2.3.4). Indeed, ν plays the same role for $\text{GRG}_n(\mathbf{w})$ as ν does for $\text{CM}_n(\mathbf{d})$. The major advantage of inhomogeneous random graphs is that the edge statuses are *independent*, while at the same time inhomogeneity is quite pronounced. Unfortunately, when studying the Ising model for a fixed random graph, i.e., in the *quenched* setting, we cannot profit from this independence. We next introduce the *annealed* measure of the Ising model on a random graph, where this independence can be used rather effectively.

Annealed measure and thermodynamic quantities. The *annealed measure*, as discussed in more detail in Section 5.1, arises by taking expectations on both sides of the ratio in (5.1.1), leading to

$$(5.7.2) \quad \tilde{\mu}_n(\vec{\sigma}) = \frac{\mathbb{E}_n \left(\exp \left[\beta \sum_{(i,j) \in E_n} \sigma_i \sigma_j + B \sum_{i \in [n]} \sigma_i \right] \right)}{\mathbb{E}_n [Z_n(\beta, B)],}$$

where $Z_n(\beta, B)$ is the partition function defined in (5.1.2). It will often be convenient to write $\tilde{Z}_n(\beta, B) = \mathbb{E}_n [Z_n(\beta, B)]$. There is an obvious relation between the quenched partition function and the annealed one:

EXERCISE 5.32 (Relation quenched and annealed pressures). *Prove that*

$$(5.7.3) \quad \tilde{\psi}_n(\beta, B) = \frac{1}{n} \log(\mathbb{E}[Z_n(\beta, B)]) \leq \mathbb{E}[\tilde{\psi}_n(\beta, B)] = \frac{1}{n} \mathbb{E}[\log Z_n(\beta, B)].$$

We can interpret the annealed measure as corresponding to the stationary distribution corresponding to the Glauber dynamics on the set of spins is defined as a Markov chain $\vec{\sigma}(t)$ on $\{-1, +1\}^n$ where we choose $j \in [n]$ uniformly at random, and let $\vec{\sigma}(t+1)$ be $\vec{\sigma}^j(t)$, but now with probability

$$(5.7.4) \quad \max \left\{ \frac{\mathbb{E}_n [e^{-H(\vec{\sigma}^j(t))}]}{\mathbb{E}_n [e^{-H(\vec{\sigma}(t))}]}, 1 \right\},$$

where the Hamiltonian H is defined in (5.1.24). Thus, instead of observing the *graph* at each time, the *average* over the random graphs is being taken.

The thermodynamic quantities with respect to the annealed measure can be defined in a similar way as in (5.1.9)–(5.1.10). For example, the *annealed pressure* is given by

$$(5.7.5) \quad \tilde{\psi}_n(\beta, B) = \frac{1}{n} \log(\mathbb{E}_n [Z_n(\beta, B)]) = \frac{1}{n} \log(\tilde{Z}_n(\beta, B)).$$

while the *annealed magnetization* is given by

$$(5.7.6) \quad \tilde{M}_n(\beta, B) = \tilde{\mu}_n \left(\frac{S_n}{n} \right),$$

and the *annealed susceptibility* equals

$$(5.7.7) \quad \tilde{\chi}_n(\beta, B) := \frac{\partial}{\partial B} \tilde{M}_n(\beta, B) = \text{Var}_{\tilde{\mu}_n} \left(\frac{S_n}{\sqrt{n}} \right).$$

It is not hard to see that the thermodynamic limits of the above objects exist. Also, there exists an *annealed critical point* denoted by β_c^{an} . In this section, we will write β_c^{qu} for the *quenched* critical point that we have studied so far, so as to distinguish it from the annealed critical point β_c^{an} .

EXERCISE 5.33 (Relation quenched and annealed pressures (cont.)). *Prove that when the thermodynamic limits of the pressure per particle $\tilde{\varphi}(\beta, B)$ and $\varphi(\beta, B)$ in the annealed and quenched settings both agree, that then $\tilde{\varphi}(\beta, B) \geq \varphi(\beta, B)$.*

We briefly explain the thermodynamic limits in the annealed setting, as they will be relevant for our analysis of the critical behavior. The statement is that for all $0 \leq \beta < \infty$ and for all $B \in \mathbb{R}$, the annealed pressure exists in the thermodynamic limit $n \rightarrow \infty$ and is given by

$$(5.7.8) \quad \tilde{\psi}(\beta, B) := \lim_{n \rightarrow \infty} \tilde{\psi}_n(\beta, B),$$

and its value is

$$(5.7.9) \quad \tilde{\psi}(\beta, B) = \log 2 + \alpha(\beta) + \sup_z \left[\mathbb{E} \left[\log \cosh \left(\sqrt{\frac{\sinh(\beta)}{\mathbb{E}[W]}} Wz + B \right) \right] - \frac{z^2}{2} \right],$$

where $z^* = z^*(\beta, B)$ is the solution of the fixed-point equation

$$(5.7.10) \quad z = \mathbb{E} \left[\tanh \left(\sqrt{\frac{\sinh(\beta)}{\mathbb{E}[W]}} Wz + B \right) \sqrt{\frac{\sinh(\beta)}{\mathbb{E}[W]}} W \right],$$

with W the limiting random variable defined in Condition 1.1(a)-(c). Further, for all (β, B) , the *magnetization per vertex* exists in the limit $n \rightarrow \infty$, i.e.,

$$(5.7.11) \quad \tilde{M}(\beta, B) := \lim_{n \rightarrow \infty} \tilde{M}_n(\beta, B).$$

For $B \neq 0$ the limit value $\tilde{M}(\beta, B)$ equals $\tilde{M}(\beta, B) = \frac{\partial}{\partial B} \tilde{\psi}(\beta, B)$ and is given by

$$\tilde{M}(\beta, B) = \mathbb{E} \left[\tanh \left(\sqrt{\frac{\sinh(\beta)}{\mathbb{E}[W]}} Wz^* + B \right) \right],$$

From this, it is not hard to verify that the *annealed critical inverse temperature* is

$$(5.7.12) \quad \beta_c^{\text{an}} = \text{asinh}(1/\nu),$$

where we recall that ν is defined in (2.3.4). In particular, if $\nu > 1$, then $\beta_c^{\text{an}} < \beta_c^{\text{qu}}$. Thus, the annealed measure has a *different* critical point compared to the quenched measure, showing that the annealing has a rather dramatic effect.

The main result in this section concerns the scaling limit of the total spin at criticality. We aim to verify when the inhomogeneity in the vertex weights destroys the limit of $S_n/n^{3/4}$ as derived for the Curie-Weiss model in Theorem 5.8. The next theorem provides

the correct scaling and the limit distribution of S_n at criticality. For $\text{GRG}_n(\mathbf{w})$, we define the inverse temperature sequence

$$(5.7.13) \quad \beta_n = \text{asinh}(1/\nu_n),$$

where

$$(5.7.14) \quad \nu_n = \frac{\mathbb{E}[W_n^2]}{\mathbb{E}[W_n]},$$

so that $\beta_{c,n} \rightarrow \beta_c$ for $n \rightarrow \infty$. The main result in this section is the following:

THEOREM 5.36 (Non-classical limit theorem at criticality). *Consider the $\text{GRG}_n(\mathbf{w})$ where the weights \mathbf{w} satisfy Conditions 1.1(a)-(c), and let δ have the respective values stated in Theorem 5.24. Assume further that for $\tau \in (3, 5)$, the weights $\mathbf{w} = (w_i)_{i \in [n]}$ are chosen as in (1.3.16) with, for $w \geq c_w$,*

$$(5.7.15) \quad 1 - F_w(w) = (c_w/w)^{(\tau-1)}.$$

Then, there exists a random variable X such that

$$(5.7.16) \quad \frac{S_n}{n^{\delta/(\delta+1)}} \xrightarrow{d} X, \quad \text{as } n \rightarrow \infty,$$

where the convergence is w.r.t. the annealed measure $\tilde{\mu}_n$ at inverse temperature $\beta_n = \text{asinh}(1/\nu_n)$ and external field $B = 0$. The random variable X has a density proportional to $\exp(-f(x))$ with

$$(5.7.17) \quad f(x) = \begin{cases} \frac{1}{12} \frac{\mathbb{E}[W^4]}{\mathbb{E}[W]^4} x^4 & \text{when } \mathbb{E}[W^4] < \infty, \\ \sum_{i \geq 1} \left(\frac{1}{2} \left(\frac{\tau-2}{\tau-1} x i^{-1/(\tau-1)} \right)^2 - \log \cosh \left(\frac{\tau-2}{\tau-1} x i^{-1/(\tau-1)} \right) \right) & \text{when } \tau \in (3, 5). \end{cases}$$

We will see that in both the case where the fourth moment is finite as well as when it is infinite,

$$(5.7.18) \quad \lim_{x \rightarrow \infty} \frac{f(x)}{x^{1+\delta}} = C,$$

with

$$(5.7.19) \quad C = \begin{cases} \frac{1}{12} \frac{\mathbb{E}[W^4]}{\mathbb{E}[W]^4} & \text{when } \mathbb{E}[W^4] < \infty, \\ \left(\frac{\tau-2}{\tau-1} \right)^{\tau-1} \int_0^\infty \left(\frac{1}{2} y^{-2/(\tau-1)} - \log \cosh y^{-1/(\tau-1)} \right) dy & \text{when } \tau \in (3, 5). \end{cases}$$

This result extends the non-classical limit theorem for the Curie-Weiss model in Theorem 5.8 to the annealed $\text{GRG}_n(\mathbf{w})$. Theorem 5.36 again highlights the importance of the finiteness of the fourth moment of the weights, which in turn holds when the fourth moment of the degrees is finite. This has been a recurring feature within this chapter.

Organisation of the proof of the non-classical limit theorem. In the remainder of this section, we derive Theorem 5.36. The proof is divided into several steps. We start by cashing in on the independence of the edge statuses in order to explicitly compute the annealed partition function. After that, we will interpret the annealed partition function as the partition function corresponding to a certain *inhomogeneous* Curie-Weiss model, which will be the work horse of this section. In this inhomogeneous Curie-Weiss model, the underlying geometry is still the complete graph, but the coupling constants are different

for every pair of vertices. Due to the approximate product structure of these coupling constants, we can adapt the Curie-Weiss arguments to this inhomogeneous setting.

Annealed partition function. We start by analyzing the average of the partition function for $\text{GRG}_n(\mathbf{w})$. By remembering that the edge statuses $(I_{ij})_{1 \leq i < j \leq n}$, with I_{ij} the indicator that the edge between vertices i and j is present, are independent Bernoulli random variables with parameter $p_{ij} = w_i w_j / (\ell_n + w_i w_j)$, we can compute

$$\begin{aligned} \mathbb{E}_n(Z_n(\beta, B)) &= \mathbb{E}_n\left(\sum_{\vec{\sigma} \in \{-1, +1\}^n} \exp\left[\beta \sum_{i < j} I_{ij} \sigma_i \sigma_j + B \sum_{i \in [n]} \sigma_i\right]\right) \\ &= \sum_{\vec{\sigma} \in \{-1, +1\}^n} e^{B \sum_{i \in [n]} \sigma_i} \mathbb{E}_n\left(e^{\beta \sum_{i < j} I_{ij} \sigma_i \sigma_j}\right) \\ &= \sum_{\vec{\sigma} \in \{-1, +1\}^n} e^{B \sum_{i \in [n]} \sigma_i} \prod_{i < j} \mathbb{E}_n\left(e^{\beta I_{ij} \sigma_i \sigma_j}\right) \\ &= \sum_{\vec{\sigma} \in \{-1, +1\}^n} e^{B \sum_{i \in [n]} \sigma_i} \prod_{i < j} (e^{\beta \sigma_i \sigma_j} p_{ij} + (1 - p_{ij})). \end{aligned}$$

We rewrite

$$(5.7.20) \quad e^{\beta \sigma_i \sigma_j} p_{ij} + (1 - p_{ij}) = C_{ij} e^{\beta_{ij} \sigma_i \sigma_j},$$

where β_{ij} and C_{ij} are chosen such that

$$(5.7.21) \quad e^{-\beta} p_{ij} + (1 - p_{ij}) = C_{ij} e^{-\beta_{ij}}, \quad \text{and} \quad e^{\beta} p_{ij} + (1 - p_{ij}) = C_{ij} e^{\beta_{ij}}.$$

Now, by adding and dividing the two equations of the system above, we get

$$(5.7.22) \quad C_{ij} \cosh(\beta_{ij}) = p_{ij} \cosh(\beta) + (1 - p_{ij}), \quad \beta_{ij} = \frac{1}{2} \log \frac{e^{\beta} p_{ij} + (1 - p_{ij})}{e^{-\beta} p_{ij} + (1 - p_{ij})}.$$

Then, using the symmetry $\beta_{ij} = \beta_{ji}$ we arrive at

$$\begin{aligned} \mathbb{E}_n(Z_n(\beta, B)) &= \left(\prod_{i < j} C_{ij}\right) \sum_{\vec{\sigma} \in \{-1, +1\}^n} e^{B \sum_{i \in [n]} \sigma_i} e^{\sum_{i < j} \beta_{ij} \sigma_i \sigma_j} \\ (5.7.23) \quad &= G(\beta) \sum_{\vec{\sigma} \in \{-1, +1\}^n} e^{B \sum_{i \in [n]} \sigma_i} e^{\frac{1}{2} \sum_{i, j \in [n]} \beta_{ij} \sigma_i \sigma_j}, \end{aligned}$$

where $G(\beta) = \prod_{i < j} C_{ij} \prod_{i \in [n]} e^{-\beta_{ii}/2}$ and we write $p_{ii} = w_i^2 / (\ell_n + w_i^2)$. This is the starting point of our analysis. We can recognize the right hand side as the partition function of an inhomogeneous Ising model on the complete graph, where the coupling constant between vertices i and j is equal to β_{ij} . In the next step, we analyze this partition function in detail.

Towards an inhomogeneous Curie-Weiss model. We continue by showing that β_{ij} is close to factorizing into a contribution due to i and to j . For this, by a Taylor expansion of $x \mapsto \log(1 + x)$,

$$\begin{aligned} \beta_{ij} &= \frac{1}{2} \log(1 + p_{ij}(e^{\beta} - 1)) - \frac{1}{2} \log(1 + p_{ij}(e^{-\beta} - 1)) \\ &= \frac{1}{2} p_{ij}(e^{\beta} - 1) - \frac{1}{2} p_{ij}(e^{-\beta} - 1) + O(p_{ij}^2) = \sinh(\beta) p_{ij} + O(p_{ij}^2). \end{aligned}$$

Then,

$$(5.7.24) \quad \mathbb{E}_n(Z_n(\beta, B)) = G(\beta) \sum_{\vec{\sigma} \in \{-1, +1\}^n} e^{B \sum_{i \in [n]} \sigma_i} e^{\frac{1}{2} \sinh(\beta) \sum_{i, j \in [n]} p_{ij} \sigma_i \sigma_j} + O(\sum_{i, j \in [n]} p_{ij}^2 \sigma_i \sigma_j).$$

To control the error in the exponent, we use $p_{ij} \leq w_i w_j / \ell_n$ and the assumptions in Condition 1.1(a)-(c), to obtain

$$\left| \sum_{i, j \in [n]} p_{ij}^2 \sigma_i \sigma_j \right| \leq \sum_{i, j \in [n]} \left(\frac{w_i w_j}{\ell_n} \right)^2 = \left(\frac{\sum_{i \in [n]} w_i^2}{\ell_n} \right)^2 = o(n).$$

Then,

$$\begin{aligned} \mathbb{E}_n(Z_n(\beta, B)) &= G(\beta) e^{o(n)} \sum_{\vec{\sigma} \in \{-1, +1\}^n} e^{B \sum_{i \in [n]} \sigma_i} e^{\frac{1}{2} \sinh(\beta) \sum_{i, j \in [n]} \frac{w_i w_j}{\ell_n} \sigma_i \sigma_j} \\ &= G(\beta) e^{o(n)} \sum_{\vec{\sigma} \in \{-1, +1\}^n} e^{B \sum_{i \in [n]} \sigma_i} e^{\frac{1}{2} \frac{\sinh(\beta)}{\ell_n} (\sum_{i \in [n]} w_i \sigma_i)^2}. \end{aligned}$$

When $w_i \equiv w$ for all i , so that $\text{GRG}_n(\mathbf{w})$ is the Erdős-Rényi random graph, we retrieve the Curie-Weiss model at inverse temperature $\beta' = \sinh(\beta)w$. In our inhomogeneous setting, we obtain an inhomogeneous Curie-Weiss model that we will analyze next.

Analysis of the inhomogeneous Curie-Weiss model. We use the Hubbard-Stratonovich identity, i.e., we write $e^{t^2/2} = \mathbb{E}[e^{tZ}]$, with Z standard Gaussian. Then, we find

$$\begin{aligned} \mathbb{E}_n(Z_n(\beta, B)) &= G(\beta) e^{o(n)} \sum_{\vec{\sigma} \in \{-1, +1\}^n} e^{B \sum_{i \in [n]} \sigma_i} \mathbb{E} \left[e^{\sqrt{\frac{\sinh(\beta)}{\ell_n}} (\sum_{i \in [n]} w_i \sigma_i) Z} \right] \\ &= G(\beta) e^{o(n)} 2^n \mathbb{E} \left[\prod_{i=1}^n \cosh \left(\sqrt{\frac{\sinh(\beta)}{\ell_n}} w_i Z + B \right) \right] \\ &= G(\beta) e^{o(n)} 2^n \mathbb{E} \left[\exp \left\{ \sum_{i=1}^n \log \cosh \left(\sqrt{\frac{\sinh(\beta)}{\ell_n}} w_i Z + B \right) \right\} \right]. \end{aligned}$$

We rewrite the sum in the exponential, using the fact that $W_n = w_{I_n}$, where we recall that I_n is a uniform vertex in $[n]$, to obtain

$$\begin{aligned} \mathbb{E}_n(Z_n(\beta, B)) &= G(\beta) e^{o(n)} 2^n \mathbb{E} \left[\exp \left\{ n \mathbb{E} \left[\log \cosh \left(\sqrt{\frac{\sinh(\beta)}{n \mathbb{E}[W_n]}} W_n Z + B \right) \middle| Z \right] \right\} \right] \\ (5.7.25) \quad &= G(\beta) e^{o(n)} 2^n \mathbb{E} \left[e^{n F_n \left(\frac{Z}{\sqrt{n}} \right)} \right], \end{aligned}$$

where

$$(5.7.26) \quad F_n(z) = \mathbb{E} \left[\log \cosh \left(\sqrt{\frac{\sinh(\beta)}{\mathbb{E}[W_n]}} W_n z + B \right) \right],$$

and the expectation in (5.7.26) is only with respect to the random variable W_n . The rewrite in (5.7.25) is highly similar as the one for the Curie-Weiss model in (5.2.11), and we will use similar techniques to estimate it.

We continue by analyzing $F_n(z)$. We claim that, uniformly for $|z| \leq a$ and any $a < \infty$,

$$(5.7.27) \quad \sup_{|z| \leq a} |F_n(z) - F(z)| = o(1),$$

where

$$F(z) = \mathbb{E} \left[\log \cosh \left(\sqrt{\frac{\sinh(\beta)}{\mathbb{E}[W]}} Wz + B \right) \right].$$

To see (5.7.27), we note that $F_n(z) \rightarrow F(z)$ for every z fixed by Condition 1.1(a)-(c), and the fact that $\log \cosh(x) \leq |x|$. Further,

$$|F'_n(z)| = \frac{\sinh(\beta)}{\mathbb{E}[W_n]} \mathbb{E} \left[\tanh \left(\sqrt{\frac{\sinh(\beta)}{\mathbb{E}[W_n]}} W_n z + B \right) W_n \right] \leq \sinh(\beta),$$

since $|\tanh(x)| \leq 1$ for all x , so that $|F'_n(z)|$ is uniformly bounded in n and z . Therefore, $(F_n)_{n \geq 1}$ forms a uniformly equicontinuous family of functions, so that (5.7.27) follows from Arzelà-Ascoli. Since $F_n(z) \leq \sinh(\beta)|z|$, it further follows that, for $a > 4 \sinh(\beta)$,

$$\begin{aligned} \mathbb{E} \left[e^{nF_n\left(\frac{z}{\sqrt{n}}\right)} \mathbb{1}_{\{|Z| > a\sqrt{n}\}} \right] &\leq \mathbb{E} \left[e^{\sqrt{n} \sinh(\beta) |Z|} \mathbb{1}_{\{|Z| > a\sqrt{n}\}} \right] \\ &= 2 \mathbb{E} \left[e^{\sqrt{n} \sinh(\beta) Z} \mathbb{1}_{\{Z > a\sqrt{n}\}} \right] \\ &= \frac{2}{\sqrt{2\pi}} \int_{a\sqrt{n}}^{\infty} e^{\sqrt{n} \sinh(\beta) z} e^{-z^2/2} dz \\ &\leq e^{a \sinh(\beta) n - a^2 n/2} \int_0^{\infty} e^{\sqrt{n}(\sinh(\beta) - a)x} dx \leq e^{-a^2 n/4}, \end{aligned}$$

which, for a sufficiently large, is negligible compared to $\mathbb{E} \left[e^{nF_n(Z/\sqrt{n})} \mathbb{1}_{\{|Z| \leq a\sqrt{n}\}} \right]$. We conclude that

$$(5.7.28) \quad \mathbb{E}_n(Z_n(\beta, B)) = G(\beta) e^{o(n)} 2^n \mathbb{E} \left[e^{nF\left(\frac{Z}{\sqrt{n}}\right)} \right].$$

A large deviation analysis. The expectation in (5.7.28) is an expectation of an exponential functional, to which we apply large deviation machinery. The Gaussian variable Z/\sqrt{n} satisfies a large deviation principle with rate function $I(z) = z^2/2$ and speed n , because $Z/\sqrt{n} \stackrel{d}{=} \frac{1}{n}(Z_1 + \dots + Z_n)$, where $(Z_i)_{i \in [n]}$ are i.i.d. standard Gaussian variables.¹ Using Varadhan's Lemma and the fact that $z \mapsto F(z)$ is continuous, we calculate the thermodynamic limit of the pressure as

$$(5.7.29) \quad \begin{aligned} \lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{E}_n(Z_n(\beta, B)) &= \log 2 + \lim_{n \rightarrow \infty} \frac{1}{n} \log G(\beta) + \sup_z [F(z) - I(z)] \\ &= \log 2 + \alpha(\beta) \\ &\quad + \sup_z \left[\mathbb{E} \left[\log \cosh \left(\sqrt{\frac{\sinh(\beta)}{\mathbb{E}[W]}} Wz + B \right) \right] - \frac{z^2}{2} \right], \end{aligned}$$

¹Alternatively, we can also write the expectation on the right hand side of (5.7.28) as an explicit integral with respect to the density of Z .

where $\alpha(\beta) = \lim_{n \rightarrow \infty} \frac{1}{n} \log G(\beta)$. The equation that defines the supremum is

$$(5.7.30) \quad z^* = z^*(\beta, B) = \mathbb{E} \left[\tanh \left(\sqrt{\frac{\sinh(\beta)}{\mathbb{E}[W]}} W z^* + B \right) \sqrt{\frac{\sinh(\beta)}{\mathbb{E}[W]}} W \right].$$

This equation is an inhomogeneous version of the fixed point equation for the Curie-Weiss model (recall (5.2.15)). As a result, many properties can be deduced from this point onwards, such as the critical temperature, the thermodynamic limits of the magnetization, internal energy and susceptibility, as well as the critical exponents. Interestingly, even though the *critical value* has changed due to the annealing (recall the annealed critical value in (5.7.12) and compare it to the q unquenched one in (5.7.1)), it turns out that the *critical exponents* remain the same. This is a strong form of *universality*.

We leave the derivation of the critical value as an exercise:

EXERCISE 5.34 (Annealed critical value). *Prove that $\beta_c^{\text{an}} = \text{asinh}(1/\nu)$ using (5.7.29) and (5.7.30).*

Non-classical limit theorems at criticality: proof of Theorem 5.36. We now prove the non-classical limit theorem in Theorem 5.36. For this, we follow the strategy of the proof for the Curie-Weiss model in Theorem 5.8. It suffices to prove that, for any real number r ,

$$(5.7.31) \quad \lim_{n \rightarrow \infty} \mathbb{E}_{\tilde{\mu}_n} \left[e^{r S_n / n^{\delta/(\delta+1)}} \right] = \frac{\int_{-\infty}^{\infty} \exp(rz - f(z)) dz}{\int_{-\infty}^{\infty} \exp(-f(z)) dz}.$$

As observed above, the measure $\tilde{\mu}_n$ is approximately equal to the inhomogeneous Curie-Weiss measure

$$(5.7.32) \quad \begin{aligned} \tilde{\mu}_n^{(\text{CW})}(g) &= \frac{1}{\tilde{Z}_n^{(\text{CW})}} \sum_{\vec{\sigma} \in \{-1, +1\}^n} g(\vec{\sigma}) e^{\frac{1}{2} \sinh \beta \sum_{i,j \in [n]} \frac{w_i w_j}{\ell_n} \sigma_i \sigma_j} \\ &= \frac{1}{\tilde{Z}_n^{(\text{CW})}} \sum_{\vec{\sigma} \in \{-1, +1\}^n} g(\vec{\sigma}) e^{\frac{1}{2} \frac{\sinh \beta}{\ell_n} (\sum_{i \in [n]} w_i \sigma_i)^2}, \end{aligned}$$

where $g(\vec{\sigma})$ is any bounded function defined in $\{-1, +1\}^n$ and $\tilde{Z}_n^{(\text{CW})}$ is the associated normalization factor, i.e.,

$$(5.7.33) \quad \tilde{Z}_n^{(\text{CW})} = \sum_{\vec{\sigma} \in \{-1, +1\}^n} e^{\frac{1}{2} \frac{\sinh \beta}{\ell_n} (\sum_{i \in [n]} w_i \sigma_i)^2}.$$

We first prove the theorem for this measure $\tilde{\mu}_n^{(\text{CW})}$, which we call a *rank-1 inhomogeneous Curie-Weiss model*, with β replaced with $\sinh(\beta)$. Below, we write $\mathbb{E}_{\tilde{\mu}_n^{(\text{CW})}}$ for the expectation with respect to $\tilde{\mu}_n^{(\text{CW})}$.

Remaining steps in the proof of Theorem 5.36 for the rank-1 inhomogeneous Curie-Weiss model. For this, we use the Hubbard-Stratonovich identity to rewrite $\mathbb{E}_{\tilde{\mu}_n^{(\text{CW})}} [e^{r S_n / n^\lambda}]$ as a fraction of two integrals of an exponential function in Lemma 5.37. We next split the analysis into the cases $\mathbb{E}[W^4] < \infty$ and $\tau \in (3, 5)$. For both these cases we analyze the exponents in the integrals and use Taylor expansions to show that they converge in Lemmas 5.38 and 5.40, respectively. We then use dominated convergence to show that the integrals also converge in Lemmas 5.39 and 5.41, respectively. The tail

behavior of $f(x)$ for $\tau \in (3, 5)$ is analyzed in Lemma 5.42. Combining these results we conclude the proof of Theorem 5.36.

Rewrite of the moment generating function. To ease the notation we first rescale S_n by n^λ and later set $\lambda = \delta/(\delta + 1)$. We rewrite $\mathbb{E}_{\tilde{\mu}_n}^{(\text{CW})} [e^{rS_n/n^\lambda}]$ in the following lemma:

LEMMA 5.37 (Moment generating function of S_n/n^λ). *For $B = 0$,*

$$(5.7.34) \quad \mathbb{E}_{\tilde{\mu}_n}^{(\text{CW})} [e^{rS_n/n^\lambda}] = \frac{\int_{-\infty}^{\infty} e^{-nG_n(z;r)} dz}{\int_{-\infty}^{\infty} e^{-nG_n(z;0)} dz},$$

where

$$(5.7.35) \quad G_n(z; r) = \frac{1}{2}z^2 - \mathbb{E} \left[\log \cosh \left(\alpha_n(\beta) W_n z + \frac{r}{n^\lambda} \right) \right], \quad \text{with} \quad \alpha_n(\beta) = \sqrt{\frac{\sinh \beta}{\mathbb{E}[W_n]}}.$$

PROOF. We use the Hubbard-Stratonovich identity, i.e., we write $e^{t^2/2} = \mathbb{E} [e^{tZ}]$, with Z standard Gaussian, to obtain

$$(5.7.36) \quad \begin{aligned} \tilde{Z}_n^{(\text{CW})} \mathbb{E}_{\tilde{\mu}_n}^{(\text{CW})} [e^{rS_n/n^\lambda}] &= \sum_{\vec{\sigma} \in \{-1, +1\}^n} e^{\frac{r}{n^\lambda} \sum_{i \in [n]} \sigma_i} e^{\frac{1}{2} \frac{\sinh \beta}{\ell_n} (\sum_{i \in [n]} w_i \sigma_i)^2} \\ &= \sum_{\vec{\sigma} \in \{-1, +1\}^n} e^{\frac{r}{n^\lambda} \sum_{i \in [n]} \sigma_i} \mathbb{E} \left[e^{\sqrt{\frac{\sinh \beta}{\ell_n}} (\sum_{i \in [n]} w_i \sigma_i) Z} \right] \\ &= 2^n \mathbb{E} \left[\prod_{i \in [n]} \cosh \left(\sqrt{\frac{\sinh \beta}{\ell_n}} w_i Z + \frac{r}{n^\lambda} \right) \right] \\ &= 2^n \mathbb{E} \left[e^{\sum_{i \in [n]} \log \cosh \left(\sqrt{\frac{\sinh \beta}{\ell_n}} w_i Z + \frac{r}{n^\lambda} \right)} \right]. \end{aligned}$$

We rewrite the sum in the exponential, using the fact that $W_n = w_{U_n}$, where U_n is a uniformly chosen vertex in $[n]$, as

$$(5.7.37) \quad \begin{aligned} \tilde{Z}_n^{(\text{CW})} \mathbb{E}_{\tilde{\mu}_n}^{(\text{CW})} [e^{rS_n/n^\lambda}] &= 2^n \mathbb{E} \left[\exp \left\{ n \mathbb{E} \left[\log \cosh \left(\sqrt{\frac{\sinh \beta}{n \mathbb{E}[W_n]}} W_n Z + \frac{r}{n^\lambda} \right) \mid Z \right] \right\} \right] \\ &= \frac{2^n}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-z^2/2} \exp \left\{ n \mathbb{E} \left[\log \cosh \left(\alpha_n(\beta) W_n \frac{z}{\sqrt{n}} + \frac{r}{n^\lambda} \right) \right] \right\} dz. \end{aligned}$$

By substituting z/\sqrt{n} for z , we get

$$(5.7.38) \quad \begin{aligned} \tilde{Z}_n^{(\text{CW})} \mathbb{E}_{\tilde{\mu}_n}^{(\text{CW})} [e^{rS_n/n^\lambda}] &= 2^n \sqrt{\frac{n}{2\pi}} \int_{-\infty}^{\infty} e^{-nz^2/2} \exp \left\{ n \mathbb{E} \left[\log \cosh \left(\alpha_n(\beta) W_n z + \frac{r}{n^\lambda} \right) \right] \right\} dz \\ &= 2^n \sqrt{\frac{n}{2\pi}} \int_{-\infty}^{\infty} e^{-nG_n(z;r)} dz. \end{aligned}$$

In a similar way we can rewrite

$$(5.7.39) \quad \tilde{Z}_n^{(\text{CW})} = 2^n \sqrt{\frac{n}{2\pi}} \int_{-\infty}^{\infty} e^{-nG_n(z;0)} dz,$$

so that the lemma follows. \square

Convergence for $\mathbb{E}[W^4] < \infty$. We next analyze the asymptotics of the function $z \mapsto G_n(z; r)$:

LEMMA 5.38 (Asymptotics of G_n for $\mathbb{E}[W^4] < \infty$). *For $\beta = \beta_n = \operatorname{asinh}(1/\nu_n)$, $B = 0$ and $\mathbb{E}[W^4] < \infty$,*

$$(5.7.40) \quad \lim_{n \rightarrow \infty} nG_n(z/n^{1/4}; r) = -zr \sqrt{\frac{\mathbb{E}[W]}{\nu}} + \frac{1}{12} \frac{\mathbb{E}[W^4]}{\mathbb{E}[W^2]^2} z^4.$$

PROOF. Taylor expanding $\log \cosh(x)$ around $x = 0$ gives that

$$(5.7.41) \quad \log \cosh(x) = \frac{x^2}{2} - \frac{1}{12} x^4 + O(x^6).$$

We want to use this to analyze $nG_n(z/n^{1/4}; r)$ and hence need to analyze the second, fourth and sixth moment of $\sqrt{\frac{\sinh(\beta_n)}{\mathbb{E}[W_n]}} W_n \frac{z}{n^{1/4}} + \frac{r}{n^\lambda}$.

The second moment equals, using that $\lambda = \delta/(\delta + 1) = 3/4$,

$$(5.7.42) \quad \begin{aligned} \mathbb{E} \left[\left(\alpha_n(\beta_n) W_n \frac{z}{n^{1/4}} + \frac{r}{n^\lambda} \right)^2 \right] &= \sinh(\beta_n) \nu_n \frac{z^2}{\sqrt{n}} + 2\sqrt{\sinh(\beta_n) \mathbb{E}[W_n]} \frac{zr}{n} + \frac{r^2}{n^{6/4}} \\ &= \frac{z^2}{\sqrt{n}} + 2\frac{zr}{n} \sqrt{\frac{\mathbb{E}[W_n]}{\nu_n}} + o(1/n), \end{aligned}$$

where we have used that $\sinh(\beta_n) = 1/\nu_n$ in the second equality.

For the fourth moment we use that by assumption the first four moments of W_n are $O(1)$. Hence, for all r ,

$$(5.7.43) \quad \begin{aligned} \mathbb{E} \left[\left(\alpha_n(\beta_n) W_n \frac{z}{n^{1/4}} + \frac{r}{n^\lambda} \right)^4 \right] &= \frac{\sinh^2 \beta_{c,n} \mathbb{E}[W_n^4] z^4}{\mathbb{E}[W_n]^2 n} + O\left(\frac{1}{n^{3/4+\lambda}} + \frac{1}{n^{2/4+2\lambda}} + \frac{1}{n^{1/4+3\lambda}} + \frac{1}{n^{4\lambda}} \right) \\ &= \frac{\mathbb{E}[W_n^4] z^4}{\mathbb{E}[W_n^2]^2 n} + o\left(\frac{1}{n}\right). \end{aligned}$$

For the sixth moment, we have to be a bit more careful since $\mathbb{E}[W^6]$ is potentially infinite. We can, however, use that

$$(5.7.44) \quad \mathbb{E}[W_n^6] = \frac{1}{n} \sum_{i=1}^n w_i^6 \leq \left(\max_{i=1}^n w_i \right)^2 \frac{1}{n} \sum_{i=1}^n w_i^4 = \left(\max_i w_i \right)^2 \mathbb{E}[W_n^4].$$

Note that $\max_{i \in [n]} w_i = o(n^{1/4})$ when $W_n \xrightarrow{d} W$ and $\mathbb{E}[W_n^4] \rightarrow \mathbb{E}[W^4] < \infty$. Hence,

$$(5.7.45) \quad \mathbb{E} \left[\left(\alpha_n(\beta_n) W_n \frac{z}{n^{1/4}} \right)^6 \right] = \frac{\sinh^3(\beta_n)}{\mathbb{E}[W_n]^3} \mathbb{E}[W_n^6] \frac{z^6}{n^{6/4}} = \frac{o(n^{1/2}) \mathbb{E}[W_n^4] z^6}{\mathbb{E}[W_n^2]^3 n^{6/4}} = o\left(\frac{1}{n}\right).$$

In a similar way, it can be shown that

$$(5.7.46) \quad \mathbb{E} \left[\left(\alpha_n(\beta_n) W_n \frac{z}{n^{1/4}} + \frac{r}{n^\lambda} \right)^6 \right] = o\left(\frac{1}{n}\right).$$

Putting everything together and using that the first four moments of W_n converge by assumption, we finally arrive at

$$(5.7.47) \quad \begin{aligned} \lim_{n \rightarrow \infty} nG_n(z/n^{1/4}; r) &= \lim_{n \rightarrow \infty} \left(\frac{\sqrt{n}}{2} z^2 - n \mathbb{E} \left[\log \cosh \left(\alpha_n(\beta_{c,n}) W_n \frac{z}{n^{1/4}} + \frac{r}{n^\lambda} \right) \right] \right) \\ &= -zr \sqrt{\frac{\mathbb{E}[W]}{\nu}} + \frac{1}{12} \frac{\mathbb{E}[W^4]}{\mathbb{E}[W^2]^2} z^4. \end{aligned}$$

□

From Lemma 5.38 it also follows that the integral converges as we show next:

LEMMA 5.39 (Convergence of the integral for $\mathbb{E}[W^4] < \infty$). For $\beta = \beta_n = \operatorname{asinh}(1/\nu_n)$, $B = 0$ and $\mathbb{E}[W^4] < \infty$,

$$(5.7.48) \quad \lim_{n \rightarrow \infty} \int_{-\infty}^{\infty} e^{-nG_n(z/n^{1/4}; r)} dz = \int_{-\infty}^{\infty} e^{zr \frac{\mathbb{E}[W]}{\sqrt{\mathbb{E}[W^2]}} - \frac{1}{12} \frac{\mathbb{E}[W^4]}{\mathbb{E}[W^2]^2} z^4} dz.$$

PROOF. We prove this lemma using dominated convergence. Hence, we need to find a lower bound on $nG_n(z/n^{1/4}; r)$. We first rewrite this function by using that

$$(5.7.49) \quad \mathbb{E} \left[\frac{1}{2} \alpha_n(\beta_n)^2 W_n^2 \left(\frac{z}{n^{1/4}} \right)^2 \right] = \frac{1}{2} \left(\frac{z}{n^{1/4}} \right)^2.$$

Hence,

$$(5.7.50) \quad \begin{aligned} G_n(z/n^{1/4}; r) &= \mathbb{E} \left[\frac{1}{2} \alpha_n(\beta_n)^2 W_n^2 \left(\frac{z}{n^{1/4}} \right)^2 - \log \cosh \left(\alpha_n(\beta_n) W_n \frac{z}{n^{1/4}} + \frac{r}{n^\lambda} \right) \right] \\ &= \mathbb{E} \left[\frac{1}{2} \left(\alpha_n(\beta_n) W_n \frac{z}{n^{1/4}} \right)^2 - \log \cosh \left(\alpha_n(\beta_n) W_n \frac{z}{n^{1/4}} \right) \right] \\ &\quad - \mathbb{E} \left[\log \cosh \left(\alpha_n(\beta_n) W_n \frac{z}{n^{1/4}} + \frac{r}{n^\lambda} \right) - \log \cosh \left(\alpha_n(\beta_n) W_n \frac{z}{n^{1/4}} \right) \right]. \end{aligned}$$

Since

$$(5.7.51) \quad \frac{d^2}{dx^2} \left(\frac{1}{2} x^2 - \log \cosh x \right) = 1 - (1 - \tanh^2(x)) = \tanh^2(x) \geq 0,$$

the function $\frac{1}{2}x^2 - \log \cosh x$ is convex and we can use Jensen's inequality to bound

$$(5.7.52) \quad \begin{aligned} &\mathbb{E} \left[\frac{1}{2} \left(\alpha_n(\beta_{c,n}) W_n \frac{z}{n^{1/4}} \right)^2 - \log \cosh \left(\alpha_n(\beta_{c,n}) W_n \frac{z}{n^{1/4}} \right) \right] \\ &\geq \frac{1}{2} \left(\alpha_n(\beta_{c,n}) \mathbb{E}[W_n] \frac{z}{n^{1/4}} \right)^2 - \log \cosh \left(\alpha_n(\beta_{c,n}) \mathbb{E}[W_n] \frac{z}{n^{1/4}} \right) \\ &= \frac{1}{2} \left(\sqrt{\frac{\mathbb{E}[W_n]}{\nu_n}} \frac{z}{n^{1/4}} \right)^2 - \log \cosh \left(\sqrt{\frac{\mathbb{E}[W_n]}{\nu_n}} \frac{z}{n^{1/4}} \right). \end{aligned}$$

As observed by Ellis in the proof of [121, Theorem V.9.5], there exist positive constants A and ε so that

$$(5.7.53) \quad \frac{1}{2} x^2 - \log \cosh x \geq d(x) := \begin{cases} \varepsilon x^4, & \text{for } |x| \leq A, \\ \varepsilon x^2, & \text{for } |x| > A. \end{cases}$$

To bound the second term in (5.7.50), we can use the Taylor expansion

$$(5.7.54) \quad \log \cosh(a+x) = \log \cosh(a) + \tanh(\xi)x,$$

for some $\xi \in (a, a + x)$, and that $|\tanh(\xi)| \leq |\xi| \leq |a| + |x|$ to obtain

$$\begin{aligned}
 & \mathbb{E} \left[\log \cosh \left(\alpha_n(\beta_{c,n}) W_n \frac{z}{n^{1/4}} + \frac{r}{n^\lambda} \right) - \log \cosh \left(\alpha_n(\beta_{c,n}) W_n \frac{z}{n^{1/4}} \right) \right] \\
 & \leq \mathbb{E} \left[\left| \log \cosh \left(\alpha_n(\beta_{c,n}) W_n \frac{z}{n^{1/4}} + \frac{r}{n^\lambda} \right) - \log \cosh \left(\alpha_n(\beta_{c,n}) W_n \frac{z}{n^{1/4}} \right) \right| \right] \\
 & \leq \mathbb{E} \left[\left(\left| \alpha_n(\beta_{c,n}) W_n \frac{z}{n^{1/4}} \right| + \frac{|r|}{n^\lambda} \right) \frac{|r|}{n^\lambda} \right] = \alpha_n(\beta_{c,n}) \mathbb{E}[W_n] \frac{|zr|}{n^{1/4+\lambda}} + \frac{r^2}{n^{2\lambda}}. \\
 (5.7.55) \quad & = \sqrt{\frac{\mathbb{E}[W_n]}{\nu_n}} \frac{|zr|}{n} + \frac{r^2}{n^{3/2}}.
 \end{aligned}$$

Hence,

$$(5.7.56) \quad e^{-nG_n(z/n^{1/4}; r)} \leq \exp \left\{ \sqrt{\frac{\mathbb{E}[W_n]}{\nu_n}} |zr| + \frac{r^2}{n^{1/2}} - nd \left(\sqrt{\frac{\mathbb{E}[W_n]}{\nu_n}} \frac{z}{n^{1/4}} \right) \right\},$$

which can be used as a dominating function. Hence, we need to prove that the integral of this function over $z \in \mathbb{R}$ is uniformly bounded, which we leave as an exercise:

EXERCISE 5.35 (Using dominated convergence). *Use the dominated convergence theorem to prove that the integral in (5.7.47) converges.*

Together with the pointwise convergence proved in Lemma 5.38, Exercise 5.35 proves Lemma 5.39. □

Convergence for $\tau \in (3, 5)$. We next analyze $G_n(z; r)$ for $\tau \in (3, 5)$. Recall that we assume that the weights $\mathbf{w} = (w_i)_{i \in [n]}$ are chosen as in (1.3.16).

LEMMA 5.40 (Asymptotics of G_n for $\tau \in (3, 5)$). *Assume that the weights $\mathbf{w} = (w_i)_{i \in [n]}$ are chosen as in (1.3.16), for some F having finite second moment. For $\beta = \beta_n = \operatorname{asinh}(1/\nu_n)$, $B = 0$ and $\tau \in (3, 5)$,*

$$(5.7.57) \quad \lim_{n \rightarrow \infty} nG_n(z/n^{1/(\tau-1)}; r) = -zr \sqrt{\frac{\mathbb{E}[W]}{\nu}} + f \left(\sqrt{\frac{\mathbb{E}[W]}{\nu}} z \right),$$

where $f(z)$ is defined in (5.7.17).

PROOF. Define the function

$$(5.7.58) \quad g(w, z) = \frac{1}{2} (\alpha_n(\beta_n) wz)^2 - \log \cosh (\alpha_n(\beta_n) wz),$$

so that we can rewrite, in a similar way as in (5.7.50),

$$\begin{aligned}
 (5.7.59) \quad nG_n(z/n^{1/(\tau-1)}; r) &= n\mathbb{E}[g(W_n, z/n^{1/(\tau-1)})] \\
 &\quad - n\mathbb{E} \left[\log \cosh \left(\frac{1}{\sqrt{\mathbb{E}[W_n^2]}} W_n \frac{z}{n^{1/(\tau-1)}} + \frac{r}{n^\lambda} \right) - \log \cosh \left(\frac{1}{\sqrt{\mathbb{E}[W_n^2]}} W_n \frac{z}{n^{1/(\tau-1)}} \right) \right].
 \end{aligned}$$

By the definition of W_n , we can rewrite

$$(5.7.60) \quad \mathbb{E}[g(W_n, z/n^{1/(\tau-1)})] = \frac{1}{n} \sum_{i=1}^n g(w_i, z/n^{1/(\tau-1)}).$$

With the deterministic choice of the weights as in (1.3.16), and assuming that (5.7.15) holds so that $w_i = c_w(n/i)^{1/(\tau-1)}$ for some constant c_w ,

$$(5.7.61) \quad \begin{aligned} g(w_i, z/n^{1/(\tau-1)}) &= \frac{1}{2} \left(\alpha_n(\beta_n) \frac{w_i z}{n^{1/(\tau-1)}} \right)^2 - \log \cosh \left(\alpha_n(\beta_n) \frac{w_i z}{n^{1/(\tau-1)}} \right) \\ &= \frac{1}{2} \left(\frac{1}{\sqrt{\mathbb{E}[W_n^2]}} \frac{c_w z}{i^{1/(\tau-1)}} \right)^2 - \log \cosh \left(\frac{1}{\sqrt{\mathbb{E}[W_n^2]}} \frac{c_w z}{i^{1/(\tau-1)}} \right). \end{aligned}$$

From this it clearly follows that, for all $i \geq 1$,

$$(5.7.62) \quad \lim_{n \rightarrow \infty} g(w_i, z/n^{1/(\tau-1)}) = \frac{1}{2} \left(\frac{1}{\sqrt{\mathbb{E}[W^2]}} \frac{c_w z}{i^{1/(\tau-1)}} \right)^2 - \log \cosh \left(\frac{1}{\sqrt{\mathbb{E}[W^2]}} \frac{c_w z}{i^{1/(\tau-1)}} \right).$$

It remains to show that also the sum converges, which we do using dominated convergence. For this, we use a Taylor expansion of $\log \cosh(x)$ about $x = 0$ up to the fourth order

$$(5.7.63) \quad \log \cosh(x) = \frac{x^2}{2} + (-2 + 2 \tanh^2(\xi) + 6 \tanh^2(\xi)(1 - \tanh^2(\xi))) \frac{x^4}{4!} \geq \frac{x^2}{2} - \frac{x^4}{12},$$

for some $\xi \in (0, x)$. Hence,

$$(5.7.64) \quad g(w_i, z/n^{1/(\tau-1)}) \leq \frac{1}{12} \left(\frac{1}{\sqrt{\mathbb{E}[W_n^2]}} \frac{c_w z}{i^{1/(\tau-1)}} \right)^4 = \frac{1}{12} \frac{1}{\mathbb{E}[W^2]^2 + o(1)} \frac{(c_w z)^4}{i^{4/(\tau-1)}}.$$

Since $\tau \in (3, 5)$, it holds that $4/(\tau - 1) > 1$, so that

$$(5.7.65) \quad \lim_{n \rightarrow \infty} \sum_{i=1}^n \frac{1}{12} \frac{1}{\mathbb{E}[W^2]^2 + o(1)} \frac{(c_w z)^4}{i^{4/(\tau-1)}} < \infty.$$

We conclude that

$$(5.7.66) \quad \begin{aligned} \lim_{n \rightarrow \infty} \sum_{i=1}^n g(w_i, z/n^{1/(\tau-1)}) &= \sum_{i=1}^{\infty} \left(\frac{1}{2} \left(\frac{1}{\sqrt{\mathbb{E}[W^2]}} \frac{c_w z}{i^{1/(\tau-1)}} \right)^2 - \log \cosh \left(\frac{1}{\sqrt{\mathbb{E}[W^2]}} \frac{c_w z}{i^{1/(\tau-1)}} \right) \right) \\ &= f \left(\sqrt{\frac{\mathbb{E}[W]}{\nu}} z \right), \end{aligned}$$

where in the last equality we have used that $\mathbb{E}[W_n] \rightarrow \mathbb{E}[W] = c_w(\tau - 1)/(\tau - 2)$.

To analyze the second term in (5.7.59), we can use the Taylor expansions

$$(5.7.67) \quad \begin{aligned} \log \cosh(a + x) &= \log \cosh(a) + \tanh(a)x + (1 - \tanh^2(\xi))x^2 \\ &= \log \cosh(a) + (a - \tanh \zeta(1 - \tanh^2 \zeta)a^2)x + (1 - \tanh^2(\xi))x^2, \end{aligned}$$

for some $\xi \in (a, a + x)$ and $\zeta \in (0, a)$. This gives

$$\begin{aligned}
 & n\mathbb{E}\left[\log \cosh\left(\frac{1}{\sqrt{\mathbb{E}[W_n^2]}}W_n\frac{z}{n^{1/(\tau-1)}} + \frac{r}{n^\lambda}\right) - \log \cosh\left(\frac{1}{\sqrt{\mathbb{E}[W_n^2]}}W_n\frac{z}{n^{1/(\tau-1)}}\right)\right] \\
 &= n\sqrt{\frac{\mathbb{E}[W_n]}{\nu_n}}\frac{z}{n^{1/(\tau-1)}}\frac{r}{n^\lambda} - n\mathbb{E}\left[\tanh \zeta(1 - \tanh^2 \zeta)W_n^2\right]\frac{1}{\mathbb{E}[W_n^2]}\frac{z^2}{n^{2/(\tau-1)}}\frac{r}{n^\lambda} \\
 &\quad + n\mathbb{E}\left[(1 - \tanh^2(\xi))\right]\frac{r^2}{n^{2\lambda}} \\
 (5.7.68) \quad &= \sqrt{\frac{\mathbb{E}[W_n]}{\nu_n}}zr + o(1),
 \end{aligned}$$

where the last equality follows from $\lambda = (\tau - 2)/(\tau - 1)$ and $\tau \in (3, 5)$. □

Again it follows that also the integral converges:

LEMMA 5.41 (Convergence of the integral for $\tau \in (3, 5)$). *For $\beta = \beta_{c,n}$, $B = 0$ and $\tau \in (3, 5)$,*

$$(5.7.69) \quad \lim_{n \rightarrow \infty} \int_{-\infty}^{\infty} e^{-nG_n(z/n^{1/(\tau-1)};r)} dz = \int_{-\infty}^{\infty} e^{zr\sqrt{\frac{\mathbb{E}[W]}{\nu}} - f\left(\sqrt{\frac{\mathbb{E}[W]}{\nu}}z\right)} dz.$$

PROOF. We again start from the rewrite of G_n in (5.7.50). As before, (5.7.70)

$$n\mathbb{E}[g(W_n, z/n^{1/(\tau-1)})] = \sum_{i=1}^n \left[\frac{1}{2} \left(\frac{1}{\sqrt{\mathbb{E}[W_n^2]}} \frac{c_W z}{i^{1/(\tau-1)}} \right)^2 - \log \cosh \left(\frac{1}{\sqrt{\mathbb{E}[W_n^2]}} \frac{c_W z}{i^{1/(\tau-1)}} \right) \right],$$

where it is easy to see that the summands are positive and decreasing in i . Hence, (5.7.71)

$$n\mathbb{E}[g(W_n, z/n^{1/(\tau-1)})] \geq \int_1^n \left[\frac{1}{2} \left(\frac{1}{\sqrt{\mathbb{E}[W_n^2]}} \frac{c_W z}{y^{1/(\tau-1)}} \right)^2 - \log \cosh \left(\frac{1}{\sqrt{\mathbb{E}[W_n^2]}} \frac{c_W z}{y^{1/(\tau-1)}} \right) \right] dy.$$

We aim use (5.7.53), for which we can split the integral in the region where $|\frac{1}{\sqrt{\mathbb{E}[W_n^2]}} \frac{c_W z}{y^{1/(\tau-1)}}|$ is bigger or smaller than A . We omit further details. Together with the pointwise convergence in the previous lemma, this proves this lemma for $r = 0$. For $r \neq 0$, the proof can be adapted as for the case $\mathbb{E}[W^4] < \infty$. □

We next analyze the large x behavior of $f(x)$ arising in the density of the limiting random variable:

LEMMA 5.42 (Asymptotics of f for $\tau \in (3, 5)$). *For $\tau \in (3, 5)$,*

$$(5.7.72) \quad \lim_{x \rightarrow \infty} \frac{f(x)}{x^{\tau-1}} = \left(\frac{\tau - 2}{\tau - 1} \right)^{\tau-1} \int_0^\infty \left(\frac{1}{2y^{2/(\tau-1)}} - \log \cosh \frac{1}{y^{1/(\tau-1)}} \right) dy < \infty.$$

PROOF. We first prove that the integral is finite. For this, define

$$(5.7.73) \quad h(y) = \frac{1}{2}y^2 - \log \cosh y,$$

so that $h(y) \geq 0$. Then,

$$(5.7.74) \quad \int_0^\infty \left(\frac{1}{2y^{2/(\tau-1)}} - \log \cosh \frac{1}{y^{1/(\tau-1)}} \right) dy = \int_0^\infty h \left(\frac{1}{y^{1/(\tau-1)}} \right) dy.$$

Since $\log \cosh y \geq 0$, we have $h(y) \leq \frac{1}{2}y^2$, and hence

$$(5.7.75) \quad h \left(\frac{1}{y^{1/(\tau-1)}} \right) \leq \frac{1}{2y^{2/(\tau-1)}}.$$

This is integrable for $y \rightarrow 0$, because $2/(\tau - 1) < 1$ for $\tau \in (3, 5)$.

Using (5.7.63), for y large,

$$(5.7.76) \quad h \left(\frac{1}{y^{1/(\tau-1)}} \right) \leq \frac{1}{12} \frac{1}{y^{4/(\tau-1)}}.$$

This is integrable for $y \rightarrow \infty$, because $4/(\tau - 1) > 1$ for $\tau \in (3, 5)$.

To prove that $f(x)/x^{\tau-1}$ converges to the integral as $x \rightarrow \infty$ we rewrite, with $a = (\tau - 2)/(\tau - 1)$,

$$(5.7.77) \quad \begin{aligned} \frac{f(x)}{x^{\tau-1}} &= \frac{1}{x^{\tau-1}} \sum_{i=1}^\infty h \left(a \frac{x}{i^{1/(\tau-1)}} \right) = a^{\tau-1} \frac{1}{(ax)^{\tau-1}} \sum_{i=1}^\infty h \left(\left(\frac{1}{i/(ax)^{\tau-1}} \right)^{1/(\tau-1)} \right) \\ &= a^{\tau-1} \int_0^\infty h \left(\frac{1}{y^{1/(\tau-1)}} \right) dy (1 + o(1)). \end{aligned} \quad \square$$

Proof of Theorem 5.36. We now complete the proof of Theorem 5.36, following the argument in the proof of Theorem 5.8. We can do a change of variables so that

$$(5.7.78) \quad \int_{-\infty}^\infty e^{-nG_n(z;r)} dz = n^{1/(\delta+1)} \int_{-\infty}^\infty e^{-nG_n(z/n^{1/(\delta+1)};r)} dz.$$

Hence, using Lemma 5.37

$$(5.7.79) \quad \tilde{\mu}_n \left(\exp \left(r \frac{S_n}{n^{\delta/(\delta+1)}} \right) \right) = \frac{\int_{-\infty}^\infty e^{-nG_n(z/n^{1/(\delta+1)};r)} dz}{\int_{-\infty}^\infty e^{-nG_n(z/n^{1/(\delta+1)};0)} dz}.$$

It follows from Lemma 5.39 for $\mathbb{E}[W^4] < \infty$ and from Lemma 5.41 for $\tau \in (3, 5)$ that

$$(5.7.80) \quad \lim_{n \rightarrow \infty} \tilde{\mu}_n \left(\exp \left(r \frac{S_n}{n^{\delta/(\delta+1)}} \right) \right) = \frac{\int_{-\infty}^\infty e^{zr\sqrt{\frac{\mathbb{E}[W]}{\nu}} - f\left(\sqrt{\frac{\mathbb{E}[W]}{\nu}}z\right)} dz}{\int_{-\infty}^\infty e^{-f\left(\sqrt{\frac{\mathbb{E}[W]}{\nu}}z\right)} dz} = \frac{\int_{-\infty}^\infty e^{xr - f(x)} dx}{\int_{-\infty}^\infty e^{-f(x)} dx},$$

where we made the change of variables $x = \sqrt{\frac{\mathbb{E}[W]}{\nu}}z$ in both integrals to obtain the last equality.

As mentioned, this is sufficient to prove the convergence in distribution of $S_n/n^{\delta/(\delta+1)}$ to the random variable X (see the argument in the proof of Theorem 5.8, or Ellis [121, Theorem A.8.7(a)]).

For the case $\mathbb{E}[W^4] < \infty$,

$$(5.7.81) \quad \lim_{x \rightarrow \infty} \frac{f(x)}{x^{1+\delta}} = \lim_{x \rightarrow \infty} \frac{\frac{1}{12} \frac{\mathbb{E}[W^4]}{\mathbb{E}[W]^4} x^4}{x^4} = \frac{1}{12} \frac{\mathbb{E}[W^4]}{\mathbb{E}[W]^4}.$$

For $\tau \in (3, 5)$, the proof that $\lim_{x \rightarrow \infty} \frac{f(x)}{x^{1+\delta}} = C$ is given in Lemma 5.42.

The above proves the claim for the rank-1 inhomogeneous Curie-Weiss model. The extension to the annealed $\text{GRG}_n(\mathbf{w})$ is quite standard from the rather explicit formula for the partition function in (5.7.23). The formal result that is needed is stated below:

LEMMA 5.43. For $\mathbb{E}[W^4] < \infty$ and $\tau \in (3, 5)$,

$$(5.7.82) \quad \lim_{n \rightarrow \infty} \mathbb{E}_{\tilde{\mu}_n}^{(\text{CW})} \left[e^{rS_n/n^\lambda} \right] - \mathbb{E}_{\tilde{\mu}_n} \left[e^{rS_n/n^\lambda} \right] = 0.$$

We leave the details to the reader. \square

Further discussion on non-classical limit theorems. It would be of great interest to extend the non-classical limit theorem for *annealed* generalized random graphs in Theorem 5.36 to quenched settings:

OPEN PROBLEM 5.2 (Non-classical limit theorems in quenched settings). *Extend Theorem 5.36 to the quenched setting of $\text{CM}_n(\mathbf{d})$ or $\text{GRG}_n(\mathbf{w})$.*

Unfortunately, we do not know how to proceed for this, as we lack an integral identity as in (5.7.38), which was the essential ingredient to the proof. In the following section, we discuss another method that can be applied to the random regular graph, and possibly could be extended also to other settings.

5.7.2. Annealed Ising model on random regular graphs. In this section, we discuss some recent results of Can [73, 74] on the behavior of the annealed Ising model on the configuration model. The main results specialize to the random regular graph, where explicit computations are possible. However, Can [73, 74] also proves some results for the configuration model $\text{CM}_n(\mathbf{D})$ with i.i.d. degrees where the degree distribution has exponential tails. We explain those at the end of this section. We start by discussing thermodynamic limits of the annealed Ising model as in Can [73], followed by non-classical limit theorems at the critical point as in Can [74].

Thermodynamic limits of annealed Ising model on d -regular random graphs. The first main result of Can [73] concerns the existence of the thermodynamic limits of the Ising model on random regular graphs:

THEOREM 5.44 (Thermodynamic quantities annealed d -regular configuration model). *Consider the random d -regular graph with $d \geq 2$ fixed. The annealed thermodynamic limits of the pressure per particle $\tilde{\psi}_n(\beta, B)$, the magnetization $\tilde{M}_n(\beta, B)$, the internal energy $\tilde{U}_n(\beta, B)$ and the susceptibility $\tilde{\chi}_n(\beta, B)$ exist and are equal to those of the quenched setting in Theorem 5.10.²*

Can [73] only proves Theorem 5.44 in the (multigraph) configuration model setting. It is not hard to conclude Theorem 5.44 from this, by conditioning on simplicity:

EXERCISE 5.36 (From configuration model to random d -regular graphs). *Show that when Theorem 5.44 holds for the d -regular configuration model, then it also holds for the random d -regular graph.*

²Beware that Can [73] uses $\tilde{\psi}_n(\beta, B)$ for the *quenched* pressure, and $\psi_n(\beta, B)$ for the annealed, rather than we do it here.

So, let us consider the d -regular configuration model from now on. Can [73] makes clever use of the symmetry that is present in the model in the annealed setting, and which replaces the explicit expectation over the edge statuses as performed for $\text{GRG}_n(\mathbf{w})$ in the previous section. For every set $A \subset [n]$, denote the (random) number of edges between A and $A^c = [n] \setminus A$ by $e(A, A^c)$. Write $\ell_A = \sum_{a \in A} d_a$ for the total degree of the subset A . We will study the *distribution* of $e(A, A^c)$ in the annealed setting, so that we can take expectations over it. The symmetry that is present is most easily described in terms of the random matching between the half-edges in the configuration model, as we explain next.

Using symmetry for the annealed partition function. Denote the random matching of size ℓ_n used to construct $\text{CM}_n(\mathbf{d})$ by $G_{\ell_n, 1}$. Fix $m \geq 1$, and let us investigate $G_{m, 1}$. For any $k \in [m]$, let

$$(5.7.83) \quad X(k, m) = \#\{\text{edges between } [k] \text{ and } [m] \setminus [k]\}.$$

The symmetry is summarized in the fact that, for every $A \subseteq [n]$,

$$(5.7.84) \quad e(A, A^c) \stackrel{d}{=} X(\ell_A, \ell_n).$$

This allows us to give the following nice and simple form for the annealed partition function in the following lemma:

LEMMA 5.45 (Annealed partition function). *Consider the random d -regular graph with $d \geq 2$ fixed. For every B, β ,*

$$(5.7.85) \quad \mathbb{E}[Z_n(\beta, B)] = e^{(\beta d/2 - B)n} \sum_{j=0}^n \binom{n}{j} e^{2Bj} g_\beta(dj, dn),$$

where

$$(5.7.86) \quad g_\beta(k, m) = \mathbb{E}[e^{-2\beta X(k, m)}].$$

PROOF. Having explained the available symmetry in the model, now it is time to apply it. Let $A_+ = \{a \in [n] : \sigma_a = +1\}$ denote the set of vertices that have positive spin. Note that

$$(5.7.87) \quad \sum_{v \in [n]} \sigma_v = 2|A_+| - n, \quad \sum_{(i, j) \in E_n} \sigma_i \sigma_j = |E_n| - 2e(A_+, A_+^c).$$

Here, since we are dealing with the multigraph setting, we have to be careful in the definition of $\sum_{(i, j) \in E_n} \sigma_i \sigma_j$, which we interpret as $\sum_{1 \leq i < j \leq n} x_{ij} \sigma_i \sigma_j$, where x_{ij} is the number of edges between $i, j \in [n]$ (and thus equals the number of self-loops when $i = j$).

We can thus rewrite the Hamiltonian $H_n(\beta, B)$ appearing in the exponent in (5.1.1) as

$$(5.7.88) \quad \begin{aligned} H_n(\beta, B) &= -\beta \sum_{(i, j) \in E_n} \sigma_i \sigma_j - \sum_{i \in [n]} B_i \sigma_i \\ &= (Bn - \beta|E_n|) + 2\beta e(A_+, A_+^c) - 2B|A_+|. \end{aligned}$$

Thus, also using that $|E_n| = \ell_n/2$, and encoding the spins by their set of plus spins $A = A_+$, we arrive at

$$(5.7.89) \quad Z_n(\beta, B) = e^{(\beta\ell_n/2 - Bn)} \sum_{A \subseteq [n]} e^{-2\beta e(A, A^c) + 2B|A|}.$$

Taking expectations and using the fact that any set of size j gives the same contribution leads to

$$(5.7.90) \quad \begin{aligned} \mathbb{E}[Z_n(\beta, B)] &= e^{(\beta\ell_n/2 - Bn)} \sum_{j=0}^n e^{+2Bj} \mathbb{E}[e^{-2\beta e([j], [n] \setminus [j])}] \\ &= e^{(\beta\ell_n/2 - Bn)} \sum_{j=0}^n \binom{n}{j} e^{2Bj} g_\beta(dj, dn), \end{aligned}$$

thus concluding the proof. \square

Lemma 5.45 yields a beautiful symmetry for the annealed Ising model, where all the action is in the function $g_\beta(k, m)$. When this function can be determined explicitly, then also the annealed partition function can be computed explicitly. In the exercise below, you are requested to extend Lemma 5.45 to general degrees:

EXERCISE 5.37 (Extension annealed partition function to general degrees). *Consider $\text{CM}_n(\mathbf{d})$ with general degrees. Extend (5.7.91) to the statement that*

$$(5.7.91) \quad \mathbb{E}[Z_n(\beta, B)] = e^{(\beta\mathbb{E}[D_n]/2 - B)n} \sum_{A \subseteq [n]} e^{2B|A|} g_\beta(\ell_A, \ell_n).$$

The thermodynamic limit of the annealed partition function. By Lemma 5.45, we immediately conclude that

$$(5.7.92) \quad \begin{aligned} \tilde{\psi}_n(\beta, B) &= \frac{1}{n} \log \mathbb{E}[Z_n(\beta, B)] \\ &= (\beta\mathbb{E}[D_n]/2 - B) + \frac{1}{n} \max_{j=0}^n \log \binom{n}{j} + \log g_\beta(dj, dn) + o(1). \end{aligned}$$

Thus, the remainder all revolves around the study of $\frac{1}{n} \log g_\beta(dj, dn)$.

Fix $t \in [0, 1]$ and define

$$(5.7.93) \quad L_{n,\beta}(t) = \frac{1}{n} \log \binom{n}{tn} + \frac{1}{n} \log g_\beta(dtn, dn),$$

so that

$$(5.7.94) \quad \tilde{\psi}_n(\beta, B) = (\beta\mathbb{E}[D_n]/2 - B) + \max_{t_n \in [n]/n} [2Bt_n + L_{n,\beta}(t_n)] + o(1).$$

The first two terms in (5.7.93) are relatively simple. Indeed, by Stirling's formula,

$$(5.7.95) \quad \frac{1}{n} \log \binom{n}{tn} = t \log(1/t) + (1-t) \log(1/(1-t)) + o(1),$$

where the error term is uniform for $t \in [0, 1]$. The asymptotics of $\frac{1}{n} \log g_\beta(dtn, dn)$ is more difficult, and is contained in the following proposition:

PROPOSITION 5.46 (Moment generating function number of edges out of set). As $m \rightarrow \infty$, uniformly in $t \in [0, 1]$,

$$(5.7.96) \quad \frac{1}{m} \log g_\beta(tm, m) = F_\beta(t) + o(1),$$

where $F_\beta(t)$ satisfies that $F_\beta(t) = F_\beta(1 - t)$ and, for $t \in [0, \frac{1}{2}]$,

$$(5.7.97) \quad F_\beta(t) = \int_0^t \log f(s) ds, \quad \text{where} \quad f_\beta(s) = \frac{e^{-2\beta}(1-2s) + \sqrt{1 + (e^{-4\beta} - 1)(1-2t)^2}}{2(1-s)}.$$

Further, there exists $b > 0$ such that, uniformly in $k \in [m]$,

$$(5.7.98) \quad \left| \log g_\beta(k+1, m) - \log g_\beta(k, m) - \log f_\beta(k/m) \right| \leq \frac{b}{m}.$$

SKETCH OF PROOF OF PROPOSITION 5.46. The symmetry $g_\beta(k, m) = g_\beta(m - k, m)$ shows that it is enough to consider $k \in [0, m/2]$. Obviously, (5.7.96) follows from (5.7.98), so we will focus on (5.7.98). We do not give all the details, particularly not the more technical and analytical ones, but we do give the main outline. Recall that $g_\beta(k, m) = \mathbb{E}[e^{-2\beta X(k, m)}]$. We analyze $g_\beta(k, m)$ by using a nice recursion relation, that follows from combining two relations between $k \mapsto g_\beta(k, m)$ and $k \mapsto g_\beta(k, m - 2)$. For this, we note that pairing one of the first $[k]$ the half-edges, we create an edge between $[k]$ and its complement with probability $(m - k)/(m - 1)$, and none otherwise. In the first case, the distribution of $X(k, m)$ is equal to $X(k - 2, m - 2)$, the the second, it is equal to $1 + X(k - 1, m - 2)$. We conclude that

$$(5.7.99) \quad g_\beta(k, m) = \frac{k-1}{m-1} g_\beta(k-2, m-2) + e^{-2\beta} \frac{m-k}{m-1} g_\beta(k-1, m-2).$$

On the other hand, when pairing one of the $[m] \setminus [k]$ last half-edges, we create an edge between $[k]$ and its complement with probability $k/(m - 1)$ and none otherwise, so that now

$$(5.7.100) \quad g_\beta(k, m) = \frac{m-k-1}{m-1} g_\beta(k, m-2) + e^{-2\beta} \frac{k}{m-1} g_\beta(k-1, m-2).$$

Equating these two formulas, and solving for $g_\beta(k, m - 2)$ yields

$$(5.7.101) \quad g_\beta(k, m - 2) = \frac{k-1}{m-k-1} g_\beta(k-2, m-2) + e^{-2\beta} \frac{m-2k}{m-k-1} g_\beta(k-1, m-2).$$

Replacing $m - 2$ by m gives the nice recurrence relation

$$(5.7.102) \quad g_\beta(k, m) = \frac{k-1}{m-k+1} g_\beta(k-2, m) + e^{-2\beta} \frac{m-2k+2}{m-k+1} g_\beta(k-1, m).$$

We now denote

$$(5.7.103) \quad h_\beta(k, m) = \frac{g_\beta(k, m)}{g_\beta(k-1, m)},$$

so that

$$(5.7.104) \quad \log g_\beta(k, m) = \sum_{i=1}^k \log h_\beta(i, m).$$

Thus, to prove (5.7.98), we need to show that $|\log h_\beta(k, m) - \log f_\beta(k/m)| \leq b/m$. For this, we note that (5.7.102) is equivalent to (5.7.105)

$$h_\beta(k, m) = \frac{k-1}{(m-k+1)h_\beta(k-1, m)} + c_\beta \frac{m-2k+2}{m-k+1}, \quad \text{where } c_\beta = e^{-2\beta}.$$

Further, since $g_\beta(0, m) = 1$, $g_\beta(1, m) = e^{-2\beta}$, this recurrence relation starts with $h_\beta(1, m) = e^{-2\beta} = c_\beta$. Equation (5.7.105) is a highly non-trivial non-linear recurrence relation that can be interpreted in terms of continued fractions. However, due to its explicit nature, detailed computations are possible. Let us start by explaining the intuition of the relation between (5.7.105) and the form of f_β in (5.7.97). Assuming that $h_\beta(k, m) \approx f_\beta(k/m)$ for some nice function f_β , and then approximating $h_\beta(k-1, m) \approx f_\beta(k/m)$ as well, we see that $f_\beta(t)$ satisfies

$$(5.7.106) \quad f(t) = \frac{c_\beta(1-2t)}{1-t} + \frac{t}{(1-t)f(t)},$$

the solution of which is given by $f(t)$ in (5.7.97). This explains (5.7.98). In order to make this precise, Can [73] proves some crucial estimates on $f_\beta(t)$, such as the statements that $f_\beta(t) \in [c_\beta, 1]$ for all $t \in [0, \frac{1}{2}]$, that $f'_\beta \in [\varepsilon, 1/\varepsilon]$ for all $t \in (0, \frac{1}{2})$. The main technical statement is the fact that there exists $b > 0$ such that

$$(5.7.107) \quad |h_\beta(k, m) - f_\beta((k-1)/m)| \leq \frac{b}{m}.$$

This is proved by showing that, for all m large enough and k such that $m/2 - k$ is sufficiently large (both bounds being independent of m), it holds that

$$(5.7.108) \quad f_\beta((k-1)/m) \leq h_\beta(k, m) \leq f_\beta(k/m).$$

This is proved by induction on k . Let us perform the lower bound, so as to highlight the idea. We note that

$$(5.7.109) \quad f_\beta(k/m) = \frac{c_\beta(m-2k)}{m-k} + \frac{k}{(m-k)f_\beta(k/m)},$$

while

$$(5.7.110) \quad h_\beta(k+1, m) = \frac{c_\beta(m-2k)}{m-k} + \frac{k}{(m-k)h_\beta(k, m)}.$$

Using the recurrence relation on $h_\beta(k, m) \leq f_\beta(k/m)$, we obtain that

$$(5.7.111) \quad h_\beta(k+1, m) \geq \frac{c_\beta(m-2k)}{m-k} + \frac{k}{(m-k)f_\beta(k/m)} = f_\beta(k/m),$$

thus advancing this part of the induction hypothesis. The upper bound is slightly more involved, and we omit this here. \square

We are now ready to complete the proof of the convergence of the pressure per particle in Theorem 5.44:

CONVERGENCE OF PRESSURE PER PARTICLE IN THEOREM 5.44. By (5.7.96),

$$(5.7.112) \quad \frac{1}{n} \log g_\beta(tdn, dn) = dF_\beta(t) + o(1),$$

which, by (5.7.92), proves that

$$(5.7.113) \quad \tilde{\psi}_n(\beta, B) = \beta \mathbb{E}[D]/2 - B + \max_{t \in [n]/n} [2Bt + L_\beta(t)] + o(1),$$

where we write

$$(5.7.114) \quad L_\beta(t) = t \log(1/t) + (1-t) \log(1/(1-t)) + d \log F_\beta(t).$$

Since $t \mapsto 2Bt + L_\beta(t)$ is continuous (even differentiable), it follows that the maximum over $t \in [n]/n$ can be replaced by the maximum over $t \in [0, 1]$, at the expense of a small error. This proves the claim of convergence. In analyzing many other quantities, the maximization problem $\max_{t \in [0, 1]} [2Bt + L_\beta(t)]$ plays a crucial role. Denote the maximizer by $t_{\beta, B}^*$. Then, for $B > 0$, by the symmetry $L_\beta(t) = L_\beta(1-t)$, we see that $t_{\beta, B}^* > \frac{1}{2}$.

The convergence of the magnetization and susceptibility requires some more calculus. For example, for the magnetization, we see that

$$(5.7.115) \quad \widetilde{M}_n(\beta, B) = \frac{1}{\mathbb{E}[Z_n(\beta, B)]} e^{(\beta d/2 - B)n} \sum_{j=0}^n \binom{n}{j} e^{2Bj} \left(\frac{2j}{n} - 1 \right) g_\beta(dj, dn),$$

suggesting that the thermodynamic limit of the magnetization indeed exists and equals $\widetilde{M}(\beta, B) = 2t_{\beta, B}^* - 1$, where $t_{\beta, B}^*$ is the maximizer of $t \mapsto 2Bt + L_\beta(t)$ in $[0, 1]$. This can be made precise by an appropriate Laplace method on the arising sum, and Taylor expanding around the maximizer. In turn, this suggests that $\tilde{\chi}(\beta, B) = 2\partial t_{\beta, B}^*/\partial B$.

Similarly,

$$(5.7.116) \quad -\tilde{U}_n(\beta, B) = \frac{1}{\mathbb{E}[Z_n(\beta, B)]} e^{(\beta d/2 - B)n} \sum_{j=0}^n \binom{n}{j} e^{2Bj} \frac{\partial}{\partial \beta} g_\beta(dj, dn),$$

where

$$(5.7.117) \quad -\frac{\partial}{\partial \beta} g_\beta(k, m) = \mathbb{E}[2X(k, m)e^{-2\beta X(k, m)}].$$

We refrain from giving more details. □

We now know that the pressure per particle converges, but now yet what the limit is. Theorem 5.44 claims that it is equal to the quenched pressure, as given in Theorem 5.10. This computation takes about 4 pages of technical computations, and will be omitted here.

Annealed Ising model on configuration models with i.i.d. degrees. Can [73] extend his results to annealed Ising models with arbitrary i.i.d. degrees, for which he assumes that the moment generating function exists. This is close to necessary, as the following exercise shows:

EXERCISE 5.38 (Annealed Ising model with i.i.d. degrees needs exponential moments). *Show that $\mathbb{E}[Z_n(\beta, B)] = \infty$ if and only if $\mathbb{E}[e^{\beta D/2}] = \infty$, where we emphasize that we take the expectation over random graph as well as its random degrees.*

However, when fixing the degrees in the configuration model, this may not be necessary, as already hinted at in Exercise 5.37:

OPEN PROBLEM 5.3 (Non-classical limit theorems for general degrees). *Extend Theorem 5.44 to the annealed configuration models $\text{CM}_n(\mathbf{d})$ where the degrees satisfy Conditions 1.6(a)-(b). When are the quenched and annealed pressures the same? When are the critical values the same?*

We expect the quenched and annealed pressures to be in general different, except in the random d -regular setting as studied above. Giardinà, Giberti and Prioriello [138] give an example where this is proved, in the case of configuration models with degrees 1 and 2. Thus, the regular graph setting can be expected to be special rather than typical. There, however, the critical values are both infinite. We are still lacking an example where the annealed and quenched critical values are different for the configuration model.

Critical exponents for annealed Ising model on random d -regular graphs.

We continue by discussing the critical behavior of the annealed Ising model on random d -regular graphs, starting with its critical exponents:

THEOREM 5.47 (Critical exponents annealed Ising model on random d -regular graphs). *For the annealed Ising model on the random d -regular graph with $d \geq 2$, the critical exponents $\beta, \delta, \gamma, \gamma', \alpha$ and α' defined in Definition 5.1 exist and satisfy $\beta = \frac{1}{2}, \delta = 3, \gamma = \gamma' = 1$ and $\alpha = \alpha' = 0$.*

Note that since the critical exponents correspond to the thermodynamic limits, and the thermodynamic limits of the annealed and quenched Ising model on random d -regular graphs agree by Theorem 5.44, also $\gamma' = 1$ and $\alpha = \alpha' = 0$ for the quenched Ising model on the random d -regular graph with $d \geq 2$.

The proof of Theorem 5.47 follows by Taylor expanding the equality $\widetilde{M}(\beta, B) = 2t_{\beta, B}^* - 1$ (recall (5.7.115)), where $t_{\beta, B}^*$ maximizes $t \mapsto 2Bt + t \log(1/t) + (1-t) \log(1/(1-t)) + F_\beta(t)$ in $[0, 1]$. We remark that Can [74] defines the critical exponents $\alpha, \alpha' = 0$ directly as those corresponding to $\partial^2 \partial \beta^2 \varphi(\beta, 0^+)$, but he does not show that the thermodynamic limit of the specific heat exists. Yet, it is very interesting that he manages to identify these critical exponents, which so far has remained illusive. We omit further details.

Non-classical limits for annealed Ising model on random d -regular graphs.

We next extend to limit theorems for the total spin of the annealed Ising model on the random d -regular graph at the critical value. The main result by Can [74] is as follows:

THEOREM 5.48 (Non-classical limit theorem at criticality). *Consider the random d -regular graph. Then, there exists a random variable X such that*

$$(5.7.118) \quad \frac{S_n}{n^{3/4}} \xrightarrow{d} X, \quad \text{as } n \rightarrow \infty,$$

where the convergence is w.r.t. the annealed measure $\tilde{\mu}_n$ at inverse temperature $\beta_c = \frac{1}{2} \log(d/(d-2))$ and external field $B = 0$. The random variable X has a density proportional to $e^{-(d-1)(d-2)x^4/(12d^2)}$.

SKETCH OF PROOF OF THEOREM 5.48. Fix $\beta = \beta_c$ and $B = 0$. Recall (5.7.90) and (5.7.93). Performing a careful analysis of the sum, this leads to

$$\begin{aligned} \mathbb{E}_{\tilde{\mu}_n} \left[e^{rS_n/n^{3/4}} \right] &= \frac{\sum_{j=0}^n e^{r(2j-n)/n^{3/4}} e^{L_{n,\beta}(j/n)}}{\sum_{j=0}^n e^{L_{n,\beta}(j/n)}} (1 + o(1)) \\ (5.7.119) \qquad \qquad \qquad &= \frac{\sum_{j=0}^n e^{r(2j-n)/n^{3/4}} e^{L_\beta(j/n)}}{\sum_{j=0}^n e^{L_\beta(j/n)}} (1 + o(1)), \end{aligned}$$

where we recall from (5.7.114) that $L_\beta(t) = t \log(1/t) + (1-t) \log(1/(1-t)) + d \log F_\beta(t)$.

Recall that $t_{\beta,B}^*$ denotes the maximizer of $t \mapsto 2Bt + L_\beta(t)$. When $\beta = \beta_c$ and $B = 0$, Can first shows that $t_{\beta_c,0}^* = \frac{1}{2}$, and then that, denoting the i th derivative of L_β by $L_\beta^{(i)}$,

$$(5.7.120) \qquad \qquad \qquad L_{\beta_c}^{(1)}\left(\frac{1}{2}\right) = L_{\beta_c}^{(2)}\left(\frac{1}{2}\right) = L_{\beta_c}^{(3)}\left(\frac{1}{2}\right) = 0,$$

while

$$(5.7.121) \qquad \qquad \qquad L_\beta^{(4)}\left(\frac{1}{2}\right) = -\frac{4(d-1)(d-2)}{3d^2}.$$

The fact that $L_{\beta_c}^{(1)}\left(\frac{1}{2}\right) = L_{\beta_c}^{(3)}\left(\frac{1}{2}\right) = 0$ follows by symmetry for $B = 0$ ($L_\beta^{(1)}\left(\frac{1}{2}\right) = 0$ also since $t_{\beta_c,0}^* = \frac{1}{2}$ is the maximizer of $L_{\beta_c}(t)$). The fact that $L_{\beta_c}^{(2)}\left(\frac{1}{2}\right) = 0$ is the crucial fact, and is due to the fact that $\beta = \beta_c$. Let us do this computation to show how the special role of $\beta_c = \frac{1}{2} \log(d/(d-2))$ enters.

Remco: Update from here!

□

5.8. Further results on Ising models on random graphs

Extension to homogeneously locally tree-like random graphs. Dembo and Montanari [92] proved their results in the finite-variance degree case in more generality than just for the configuration model (even though the configuration model was one of their prime examples), and also the results in [109] extend to this setting. Let us comment on this here. The results in Sections 5.3, 5.5 and 5.6 apply more generally than only for the configuration model. Indeed, they hold when the random graph is homogeneously locally tree-like with a limiting distribution that has a finite $(1 + \varepsilon)$ moment. This, for example, applies to rank-1 inhomogeneous random graphs. We next explain this setting. Recall the definition of local-weak convergence towards a homogeneous unimodular Galton-Watson tree in Theorem 2.11, as well as in Theorem 2.14. We further say that the graph sequence $(G_n)_{n \geq 1}$ is *uniformly sparse* when, a.s.,

$$(5.8.1) \qquad \qquad \qquad \lim_{\ell \rightarrow \infty} \limsup_{n \rightarrow \infty} \frac{1}{n} \sum_{i \in [n]} d_i \mathbb{1}_{\{d_i \geq \ell\}} = 0.$$

Note that uniform sparsity follows if $\frac{1}{n} \sum_{i \in [n]} d_i \rightarrow \mathbb{E}[D]$ a.s., by the weak convergence of the degree of a uniform vertex. Then, the results in Sections 5.3, 5.5 and 5.6 hold whenever the graph sequence $(G_n)_{n \geq 1}$ converges locally-weakly towards a homogeneous unimodular Galton-Watson, is uniformly sparse, and has a strongly-finite mean.

The antiferromagnetic Ising model. The Ising model studied in this chapter is called *ferromagnetic*, as neighboring spins tend to wish to align and take the same value. There is also substantial interest in the *antiferromagnetic*. In the social context, individuals wish to align with those they look up to, which created the ferromagnetic interaction. They may also wish to set themselves apart from some people (think about adolescents and their parents). In this case, the coupling constant would not be non-negative, but rather negative. There is some work on the antiferromagnetic Ising model on random graphs.

Contucci, Dommers, Giardinà and Starr [84] use interpolation arguments on a suitable variational principle to prove the existence of a phase transition at a positive critical temperature for the antiferromagnetic Ising model on the Erdős-Rényi random graph. They also provide upper and lower bounds on the critical temperature critical value. Salez [236] proves the existence of the thermodynamic limit of the pressure using the fact that *concave* graph parameters converge in the local weak convergence topology, and showing that the pressure is indeed concave. This is surprising, since for the antiferromagnetic Ising model for sufficiently sufficiently negative β , it is not difficult to show that the free energy per spin on random d -regular graphs is asymptotically lower than on random bipartite d -regular graphs. Thus, local weak convergence is generally not a sufficient condition for existence of the pressure per particle.

Metastability for the Ising model on random graphs. Recently, there has also been extensive interest in the *dynamical* properties of the Ising model, in particular its mixing times and its metastable behavior. For a general background on metastability, we refer to the recent book by Bovier and den Hollander [65]. Metastability refers to the setting where it takes a system that evolves dynamically an enormous amount of time to reach its stationary state. In this case, there is a so-called *metastable state*, which is highly favorable, yet not optimal, and it takes the system a long time to leave from this local optimum. In the Ising model, this should correspond to settings where the temperature is low (so that β is high, and there is order in the system), we take a small magnetic field $B > 0$ while starting close to the all minus configuration. While it seems obvious that it will take the system an enormous amount of time to depart from such a system, so far the results are considerably weaker than that, as this is a very difficult problem. Dommers [107] investigates metastability properties for the random d -regular graph with d large. He takes n large, and investigates the metastability of the all minus configuration as the temperature tends to zero (i.e., $\beta \rightarrow \infty$).

A very exciting (but probably hard) problem is to extend the metastability results to the setting of β large (but fixed) and $n \rightarrow \infty$:

OPEN PROBLEM 5.4 (Metastability for small temperatures in large-graph limit). *Prove metastability results on the configuration model when β is large but fixed, in the limit as $n \rightarrow \infty$.*

The Potts model on random graphs. There are many more interesting statistical mechanics models that one might consider. See Liggett [205] and Contucci and Giardinà [86] for examples. Here we discuss some work on the Potts model, which poses quite a few novel challenges. See also Grimmett [145] for the relation between FK-percolation and Potts models.

In the Potts model, rather than having only two possible states, there are q of them. Thus, the setting of $q = 2$ reduces to the Ising model. Surprisingly, changing the value of q to values larger than 2 has a rather dramatic effect. The setting on locally tree-like random graphs is not understood, but there is a highly interesting work on the random d -regular graph by Dembo, Montanari, Sly and Sun [93]. Let us explain how the difficulty of the Potts model for $q > 2$ may be understood. Recall the local weak limit argument applying to the Ising model around (5.3.20). The fact that both limits are equal (see Proposition 5.13) is crucial in the proof of the convergence of the pressure per particle. Later, such arguments have also been used successfully for the magnetization (see also the alternative proof of (5.3.12) in the proof of Theorem 5.24). This makes crucial use of a monotonicity argument. In a similar way, the local weak limit of Potts measures on G_n can be sandwiched between Potts model with free boundary conditions for the lower bound, and the so-called maximally 1-biased automorphism-invariant Gibbs measure for the upper bound. For the Ising model for which $q = 2$, these measures luckily coincide for all $\beta \geq 0, B > 0$. For $q > 2$, unfortunately, these measures disagree in certain regimes of the parameter space (β, B) . This makes it unclear which is in fact the relevant measure in this setting.

Statistical physics folklore predicts that the fixed point with the highest free energy density on the tree should be selected. However, as Dembo, Montanari, Sly and Sun [93] point out, in the physics literature this is justified only via analogy with other models, without providing arguments that apply to locally tree-like graphs. The result by Dembo, Montanari, Sly and Sun [93] is the first rigorous verification of this variational principle in a non-trivial example for locally tree-like graphs.

Dembo, Montanari, Sly and Sun [93] also provide a probabilistic interpretation for this variational principle for random d -regular graphs, in relation to the *annealed* setting (see also the discussion in Section 5.7). They show that $\tilde{\psi}_n(\beta, B) = \frac{1}{n} \log(\mathbb{E}[Z_n(\beta, B)])$ and $\mathbb{E}[\tilde{\psi}_n(\beta, B)]$ have the same limits for d even (see [93, Theorem 1], the easier lower bound is proved in [94] and applies to all d). Obviously, upper bounds $\tilde{\psi}_n(\beta, B) \leq \mathbb{E}[\tilde{\psi}_n(\beta, B)] = \frac{1}{n} \mathbb{E}[\log Z_n(\beta, B)]$, recall Exercises 5.32 and 5.33. In fact, the computation of $\mathbb{E}[Z_n(\beta, B)]$ can be understood to correspond in a very precise manner to the folklore prescription of maximizing Bethe free energy over all its fixed points.

A natural (yet very hard) problem would be to generalize the results for the Potts model to general configuration models:

OPEN PROBLEM 5.5 (Potts models on the configuration model).
Extend the results by Dembo, Montanari, Sly and Sun [93] to Potts models on general configuration models.

One can expect this to be an extremely difficult problem as even the setting on Galton-Watson trees is not understood. We refer to Dembo, Montanari, Sly and Sun [93] for an extensive discussion.

Remarkably, the *antiferromagnetic* setting of the Potts model turns out to be highly similar to that of the Ising model, in the sense that the number of spin values matters a lot less there. The results of Contucci, Dommers, Giardinà and Starr [84] and Salez [236] discussed above also apply to the Potts setting.

5.9. Notes and discussion of Ising models on random graphs

Notes on Section 5.1. We refer to Niss [224, 225] for a historical account of the Ising model. Its inception is certainly interesting. The model was proposed to Ising by his PhD supervisor Lenz. Ising wrote his thesis, mainly containing the analysis of the one-dimensional case, where there is no phase transition. By analogy, he then conjectured there not to be a phase transition in any dimension, which was later disproved by Onsager for \mathbb{Z}^2 . The use of the general statistical mechanics models on random graphs is also motivated by applications in statistics, combinatorial optimization and counting problems. See the extensive overview by Dembo and Montanari [91] for these connections.

Notes on Section 5.2. For more background on the Curie-Weiss model, see Kochmański, Paszkiewicz and Wolski [197], who also use the Hubbard-Stratonovitz transformation. We also found Bovier and Kurkova [66], who rather rely on explicit counting of the number of collections of spins with given total spin, useful. Theorem 5.8 is proved by Ellis in [121, Theorem V.9.5].

Notes on Section 5.3. I have removed a few typos in the proof of Proposition 5.14 that were present in [109].

A key idea to analyze the Ising model on random graphs is to use the fact that expectations of local quantities coincide with the corresponding values for the Ising model on suitable random trees [92]. Statistical mechanics models on deterministic trees have been studied extensively in the literature (see for instance Baxter [30], Lyons [211], and its relation to “broadcasting on trees” in work of Evens, Kenyon, Peres and Schonmann [129] and Mezard and Montanari [214]). The analysis on random trees is more recent and has been triggered by the study of models on random graphs. Dembo and Montanari [91] give a general introduction to the various aspects of Ising models on random graphs that makes for a highly recommendable read.

The proof of the existence of the thermodynamic limits in Theorems 5.9 and 5.10 follows the line of argument used first by Dembo and Montanari in [92]. We have used some improvements in work with Dommers and Giardinà [109] in order to prove Propositions 5.13, 5.16 and 5.14 in the infinite-variance case. To prove Proposition 5.13, the proof of Dembo and Montanari is adapted by using the actual forward degrees, instead of using Jensen’s inequality to replace them by expected forward degrees, which are potentially infinite. This also makes a separate analysis of vertices that have zero offspring superfluous, which considerably simplifies the analysis.

The proof of Proposition 5.16(a) is somewhat more elaborate in [109] compared to [92], because of the necessary distinction between the cases where D in (5.3.3) is small or large. The techniques used remain similar. By, again, taking into account the actual degrees more precisely, the analysis is simplified however: we, for example, do not rely on the exponential decay of the correlations. Part (b) of this proposition is new and can be proved with similar techniques. The proof of Proposition 5.14 is a minor adaptation of that in [92].

An overview of results by physicists can for example be found in Dorogovtsev, Goltsev and Mendes [113]. De Sanctis and Guerra studied this model on Erdős-Rényi random graphs in the high and zero temperature regime [89]. Dembo and Montanari [92] study a ferromagnetic Ising model on locally tree-like graphs, where they assume that the degree distribution of the graph has finite variance. The Ising model on the d -regular graph

where there is no external magnetic field is studied in more detail by Montanari, Mossel and Sly [217].

Notes on Section 5.5. Theorem 5.24 confirms the physics predictions in Dorogovtsev, Goltsev and Mendes [112] and Leone, Vázquez, Vespignani and Zecchina [203]. For $\tau \leq 3$, one has $\nu = \infty$ and hence $\beta_c = 0$ by Theorem 5.21, so that the critical behavior coincides with the infinite temperature limit. Since in this case there is no phase transition at finite temperature, we do not study the critical behavior here. For $\tau = 5$, Dorogovtsev, Goltsev and Mendes [112], also compute the logarithmic correction for $\beta = 1/2$ in (5.5.2), but not that of $\delta = 3$.

Theorem 5.24 only gives a lower bound on the critical exponent γ' . It is predicted that $\gamma' = 1$ for all $\tau > 3$. The proof with Dommers and Giardinà in [110] shows that $\chi(\beta, 0^+) \geq c(\beta - \beta_c)^{-1}$ for some constant $c > 0$, which, if the critical exponent γ' exists, implies that it must satisfy $\gamma' \geq 1$. Unfortunately, we cannot prove that the critical exponent γ' exists, see the discussion in [110] for more details on this issue. Further, [110] also proves related results for the Ising model on branching process trees with power-law offspring distributions. We refrain from discussing these results in more detail.

There are also predictions for other critical exponents in the physics literature. For instance the critical exponent α' for the specific heat in the low-temperature phase satisfies $\alpha' = 0$ when $\mathbb{E}[\Delta^3] < \infty$ and $\alpha' = (\tau - 5)/(\tau - 3)$ in the power-law case with $\tau \in (3, 5)$ (see [112, 203]). The critical exponent α' for the specific heat is beyond our current methods, partly since we are not able to relate the specific heat on a random graph to that on the random Bethe tree.

Notes on Section 5.6. The proof of the central limit theorem in Theorem follows the strategy of Ellis [121], who proved a central limit theorem for the total spin of the Ising model in a large box in \mathbb{Z}^d in the uniqueness regime. Related central limit theorems were proved in another work with Giardinà, Giberti and Prioriello [138], which focussed on annealed random graphs. Examples were $\text{CM}_n(\mathbf{d})$ where all degrees are 1 or 2, and the generalized random graph. The latter will also be discussed in Section 5.7, and we refrain from going into more details here.

Notes on Section 5.7. The rank-1 inhomogeneous Curie-Weiss model was introduced and studied in work with Dommers et al. [108]. Even though it seems quite natural to us, it seems that it had not been defined before. The fact that annealed and quenched critical exponents agree in this setting is a strong form of universality. This is particularly remarkable since the critical values *are* different.

The fact that the thermodynamic limit of the pressure per particle on the d -regular graph, proved by Can [73], is the same in the annealed and quenched settings has already been observed by Dembo, Montanari, Sly and Sun [93] (see also the remark below [93, Theorem 1], where it is mentioned that the Ising result also holds for odd degree), who apply it even to the Potts model for even degree. For $d = 2$, it was also observed in [138].

Can [73] proves some further interesting results that we summarize here. Can proves concentration of the total spin around $nM(\beta, 0^+)$ in [73, Proposition 1.4], also for $\beta > \beta_c$ which is outside of the uniqueness regime.

Further, in [74, Theorem 1.2], Can also proves the existence of all critical exponents that have been identified so far in Section 5.5, including γ' , α and α .

Notes on Section 5.8.

Related stochastic processes on random graphs

6.1. Competition processes

In this section, we discuss the behavior of competition processes on random graphs. Many versions of such processes can be imagined, and we discuss several of them. We start by discussing competition processes that spread like a rumor on the graph, but where vertices, once invaded, keep their type forever. There is a close relation to information diffusion on random graphs, as the two processes are equal up to the moment where the invaded regions of different types touch each other.

One can imagine such competition processes as a caricature model for viral marketing of expensive products, say laptops. When I have decided which laptop to buy, I will stick to that choice for a while. Thus, at least for the time being, my choice is fixed and I am not susceptible to competitors.

Let us now formally define the model. Fix a graph $G = (V, E)$, finite or infinite. Start invasion processes from two starting points. Let each of the types have edge-weights $(Y_e^{(1)})_{e \in E}$ and $(Y_e^{(2)})_{e \in E}$ associated to the edges. Each vertex performs first-passage percolation, but each time a vertex is found by each of the types, it only accepts it when the vertex has not been found yet by the other type. This means that at each time t , the vertex sets of vertices of each of the two types, denoted by $V_t^{(1)}$ and $V_t^{(2)}$, are disjoint and connected sets. Of course, other settings where there are more starting points, more types, etc., are all feasible and of interest, but here we will stick to this simplest setting. The basic question is how many vertices each of the types finally reach. We will predominantly deal with *connected* graphs, and the dynamics will be such that each vertex will eventually be found by one of the two types. It is possible that the traversal times $(Y_e^{(1)})_{e \in E}$ and $(Y_e^{(2)})_{e \in E}$ are *different* for the different types. The easiest setting is that $Y_e^{(1)}$ has the same distribution as $Y_e^{(2)}/\lambda$, where without loss of generality we may assume that $\lambda \geq 1$. Here λ could be related to the attractiveness of each of the different species. When $\lambda = 1$, then the products are equally attractive, while if $\lambda > 1$, the type 1 product spreads more quickly on average, making type 1 more attractive. We will call λ the speed of type 1 compared to type 2.

The main question of interest is who wins the majority of the vertices, and whether (and if so how) this depends on the speed in the system. When the graph G is the hypercubic lattice and the $(Y_e^{(1)})_{e \in E}$ and $(Y_e^{(2)})_{e \in E}$ each are i.i.d. sequences of exponential random variables, possibly with different means, this question has attracted substantial attention. The first-passage percolation setting which this resembles is called the *Richardson model*. Since we are on an infinite graph, we can wonder whether there might be *infinite co-existence*, meaning that each of the different types wins an infinite number of vertices. One would guess that this is only possible when $\lambda = 1$ and the types are equally attractive. The Richardson model has also been extended to a two-type version that describes a competition between two infection types; see Häggström and Pemantle [149]. Infinite

co-existence then refers to the event that both infection types occupy infinite parts of the lattice, and it is conjectured that this has positive probability if and only if the infections have the same intensity. The if-direction was proved by Häggström and Pemantle [149] for $d = 2$, and for general d independently Garet and Marchand [136] and Hoffman [158]. The only-if-direction remains unproved. Häggström and Pemantle [150] give convincing partial results.

The aim of this section is to investigate the setting on random graphs. Since random graphs are finite, co-existence needs to be reformulated. We will say that *finite co-existence* occurs when each of the two types wins a positive proportion of the graph with positive probability. When this is not the case, then we speak of a *winner-take-it-all* scenario. In the latter case, the main question of interest is how many vertices the losing type wins. In formulas, let $N_1^{(n)}$ and $N_2^{(n)}$ denote the number of vertices that the types 1 and 2 win in the graph $G_n = (V_n, E_n)$ of size n . Finite co-existence corresponds to the setting where there exists an $\varepsilon > 0$ such that, with positive probability, $N_1^{(n)} \geq \varepsilon n$ and $N_2^{(n)} \geq \varepsilon n$. In the winner-takes-it-all scenario, $\max\{N_1^{(n)}, N_2^{(n)}\}/n \xrightarrow{\mathbb{P}} 1$ and we are mainly interested in the behavior of $\min\{N_1^{(n)}, N_2^{(n)}\}$.

Organisation of this section. This section is organised as follows. In Section 6.1.1, we will fix ideas by studying competition processes on the complete graph. In Section 6.1.2, we study the setting of competition on random regular graphs, and formulate some conjectures for competition processes on general configuration models with finite-variance degrees. In Section 6.1.3, we study competition on configuration models with infinite-variance degrees, where surprisingly rich phenomena occur. We close this section in Section 6.1.4 by studying a competition process on the preferential attachment model, where the dynamics is incorporated into the competition process.

6.1.1. Fixing ideas: Competition on the complete graph. Let us start by analyzing the complete graph. We assume that the edge weights are exponential. More precisely, we let $Y_e^{(1)} \sim \text{Exp}(\lambda)$ have an exponential distribution with mean $1/\lambda$, while $Y_e^{(2)} \sim \text{Exp}(1)$. All edge weights are independent.

Equal speeds. Let us first describe the setting where each of the types has equal speed, as this turns out (somewhat surprisingly) to be the easier case. The main result is as follows:

THEOREM 6.1 (Co-existence for equal-speed competition on complete graph). *Consider competition on the complete graph, starting from two distinct sources. Let $N_1^{(n)}, N_2^{(n)}$ denote the number of vertices of type 1 and 2 in the limiting configuration. Then,*

$$(6.1.1) \quad \frac{N_1^{(n)}}{n} \xrightarrow{d} U,$$

where U has a standard uniform distribution.

Theorem 6.1 implies that for the equal speeds case, the competition process gives rise to co-existence. Even, co-existence occurs almost surely. We will give two separate proofs of Theorem 6.1. The first relies on Pólya urn schemes, as discussed in Theorem 2.17, and is highly specific to the setting of regular graphs (in particular, the complete graph). The second proof relies on growth of first-passage percolation smallest-weight graphs and their

connection to continuous-time branching processes, which is a tool that has the potential of being more robust as we will explain in more detail below.

PROOF OF THEOREM 6.1 USING PÓLYA URN SCHEMES. By the memoryless property of the exponential distribution, together with the fact that the speeds are equal for the two processes, we see that at each stage k of the competition, given the number of vertices $V_k^{(i)}$ of types 1 and 2, respectively, the next vertex will be of type 1 with conditional probability $V_k^{(1)}/(V_k^{(1)} + V_k^{(2)})$. We conclude that $(V_k^{(1)}, V_k^{(2)})_{k \geq 1}$ is exactly equal to a Pólya urn schemes with, in the notation of Theorem 2.17, $a_r = a_b = 0$ and $r_0 = b_0 = 1$. Thus, Theorem 2.17 proves the claim, when we note that the limiting Beta(1,1) random variable is nothing but a standard uniform random variable. \square

EXERCISE 6.1 (Starting from multiple vertices). *Suppose that we start our competition process from r_1 vertices of type 1 and r_2 vertices of type 2. Show that*

$$(6.1.2) \quad \frac{N_1^{(n)}}{n} \xrightarrow{d} U,$$

and determine the distribution of U .

PROOF OF THEOREM 6.1 USING CONTINUOUS-TIME BRANCHING PROCESSES. This proof is a little more technical, but also more robust. We do not give all details. We investigate the growth of the smallest-weight graphs from vertices 1 and 2 at time t/n . The rescaling of time by $1/n$ is equivalent to rescaling the parameters of the exponential edge-weights by a factor n . We grow both smallest-weight graphs up to time $t_n = \frac{1-\varepsilon}{2} \log n$, ignoring the competition effect for the time being. Then, their sizes $Z_{t_n}^{(i)}$, for $i \in \{1, 2\}$, for $i \in \{1, 2\}$ will be close to

$$(6.1.3) \quad E_i e^{t_n} = E_i n^{(1-\varepsilon)/2}.$$

Recall the analysis in Section 3.2, as well as the description of the Yule process in Section 3.4.1. Moreover, since $n^{(1-\varepsilon)/2} \ll \sqrt{n}$, these sets are with high probability disjoint:

EXERCISE 6.2 (Disjointness of smallest-weight graphs). *Perform first-passage percolation on the complete graph K_n from two sources 1 and 2. Assume that the respective sizes are n_1, n_2 . Prove that the two sets are with high probability disjoint when $n_1 n_2 = o(n)$.*

Now, we wish to analyse how the number of vertices conquered by types 1 and 2 behaves. For this, we condition on the sizes $Z_{t_n}^{(i)}$, for $i \in \{1, 2\}$, and grow smallest-weight graphs from *all other* vertices, until the moment that they connect to one of the vertices that has already been conquered by type 1 or type 2. By symmetry, we may assume that this other vertex is vertex 3. We color a vertex type 1 when it first connects to one of the $Z_{t_n}^{(1)}$ type 1 vertices, and type 2 when it first connects to one of the $Z_{t_n}^{(2)}$ type 2 vertices. Let $M_n^{(i)}$ be equal to the number of vertices conquered by type i in this way, so that $N_n^{(i)} = M_n^{(i)} + Z_{t_n}^{(i)}$. We conclude that

$$(6.1.4) \quad \frac{N_n^{(i)}}{n} = \frac{M_n^{(i)}}{n} + o_{\mathbb{P}}(1),$$

so it suffices to study $M_n^{(i)}$. Now, conditionally on $(Z_{t_n}^{(1)}, Z_{t_n}^{(2)})$, each time k a novel vertex is found, it has probability $Z_{t_n}^{(1)}/(n - k)$, since the next vertex is chosen uniformly from

all vertices that are not yet in the smallest-weight graph of vertex 3. This suggests that the probability that vertex 3 becomes type 1 is equal to

$$(6.1.5) \quad \frac{Z_{t_n}^{(1)}}{Z_{t_n}^{(1)} + Z_{t_n}^{(2)}} \approx \frac{E_1}{E_1 + E_2}.$$

Assuming that $M_n^{(i)}$ concentrates (which can be shown by a conditional variance computation), we arrive at the claim that

$$(6.1.6) \quad \frac{M_n^{(i)}}{n} \xrightarrow{\mathbb{P}} \frac{E_1}{E_1 + E_2}.$$

Now, $E_1/(E_1 + E_2)$ has a uniform distribution:

EXERCISE 6.3 (Uniform random variable in terms of exponentials). *Let E_1, E_2 be two independent standard exponential random variables. Prove that $E_1/(E_1 + E_2)$ has a standard uniform distribution.*

□

EXERCISE 6.4 (Starting from multiple vertices (cont.)). *Suppose that we start our competition process from r_1 vertices of type 1 and r_2 vertices of type 2. Show that*

$$(6.1.7) \quad \frac{M_n^{(1)}}{n} \xrightarrow{d} \frac{E_1^{(1)} + \cdots + E_{r_1}^{(1)}}{E_1^{(1)} + \cdots + E_{r_1}^{(1)} + E_1^{(2)} + \cdots + E_{r_2}^{(2)}},$$

where all exponential random variables are independent. This gives an alternative description of the random variable appearing in Exercise 6.1.

Unequal speeds. We next investigate what happens when the speeds are different. Assume that the rate at which type 1 occupies its neighbors is λ , while that of type 2 is 1. Our investigation will be at the heuristic level, there is no literature on this problem yet. We grow the smallest-weight graph of type 1 to size $\varepsilon_n n$, where $\varepsilon_n = o(1)$ is some sequence to be chosen later on. Throughout the argument, we will frequently use the analysis of Section 3.2, where it was shown that the number of vertices found by first-passage percolation starting from a vertex v in time nt is close to

$$(6.1.8) \quad E_v e^t,$$

recall in particular (3.2.6), as well as the relation to Yule processes in Section 3.4.1.

Using (6.1.8) and recalling that the occupation process from of type 1 runs at rate $\lambda > 1$ rather than 1, we obtain that the time it takes the smallest-weight graph of type 1 to grow to size $\varepsilon_n n T_n^{(1)}$ is such that

$$(6.1.9) \quad n[T_n^{(1)} - (1/\lambda) \log(\varepsilon_n n)] \xrightarrow{d} \log(1/E_1).$$

At this moment, type 2 has only occupied $M_n^{(2)}$ vertices, where $M_n^{(2)} e^{-T_n^{(1)}} \xrightarrow{d} E_2$. We conclude that

$$(6.1.10) \quad M_2^{(n)} \approx E_2 e^{T_n^{(1)}} \approx (\varepsilon_n n)^{1/\lambda} \frac{E_2}{E_1^{1/\lambda}}.$$

You are asked to make these statements precise in the following exercise:

EXERCISE 6.5 (Initial number of vertices occupied by losing type). *Make the convergence statement in (6.1.10) precise as a limit in distribution.*

We arrive at the statement that each of the two types has occupied a certain number of vertices. Type 1, the winner, is in the lead and has already occupied $\varepsilon_n n$ vertices, while type 2, the loser, is stuck at a number of occupied vertices equal to $M_2^{(n)}$. Denote by $N_2^{(n)}$ the total number of vertices occupied by the losing type 2. We aim to do a conditional second moment method on $N_2^{(n)}$, conditionally on $\mathcal{F}_{T_n^{(1)}}$, the σ -algebra of all that has happened up to time $T_n^{(1)}$. Then, we have that

$$(6.1.11) \quad N_2^{(n)} = M_2^{(n)} + \sum_{v \in [n] \setminus [M_2^{(n)} + \varepsilon_n n]} \mathbb{1}_{\{v \text{ type 2}\}},$$

where, for simplicity, we label all the vertices occupied by types 1 and 2 combined by $[M_2^{(n)} + \varepsilon_n n]$. We aim to show that

$$(6.1.12) \quad \frac{N_2^{(n)}}{n^{1/\lambda}} \xrightarrow{\mathbb{P}} \Gamma(1 + 1/\lambda) \Gamma(1 - 1/\lambda) \frac{E_2}{E_1^{1/\lambda}}.$$

Here we note that the constant $\Gamma(1 + 1/\lambda) \Gamma(1 - 1/\lambda)$ is finite precisely when $\lambda > 1$, which is encouraging. We will only consider the conditional first moment of $N_2^{(n)}$, the variance being similar. Fix $v \in [n] \setminus [M_2^{(n)} + \varepsilon_n n]$. We know that $M_2^{(n)} n^{-1/\lambda} \xrightarrow{\mathbb{P}} 0$ by Exercise 6.5, so we can ignore this contribution. Thus,

$$(6.1.13) \quad n^{-1/\lambda} \mathbb{E}[N_2^{(n)} \mid \mathcal{F}_{T_n^{(1)}}] = n^{-1/\lambda} \mathbb{P}(v \text{ type 2} \mid \mathcal{F}_{T_n^{(1)}}) + o_{\mathbb{P}}(1).$$

To investigate the arising probability, we let the occupation processes of types 1 and 2 evolve independently starting from vertex v . We first only consider the occupation process of type 1 from vertex v , until the moment that a first vertex occupied by type 1 is found. The size of the tree at that time is then roughly $n/(n\varepsilon_n) = 1/\varepsilon_n$, since each time a new vertex is occupied, with probability ε_n it will be one of the $n\varepsilon_n$ first vertices. More precisely, the size is close to a Geometric random variable with success probability ε_n , so that the size $I_n^{(v)}$ satisfies that $\varepsilon_n I_n^{(v)} \xrightarrow{d} E_v^{(1,1)}$. The time it takes for this to happen is thus close to

$$(6.1.14) \quad T_{I_n^{(v)}}^{(v)} \approx \frac{1}{n\lambda} \left[\log(E_v^{(1,1)}/\varepsilon_n) + \log(1/E_v^{(1,2)}) \right].$$

At this moment, the number of vertices occupied by type 2 is approximately

$$(6.1.15) \quad K_2^{(n)} = E_v^{(2)} e^{T_{I_n^{(v)}}^{(v)}} \approx E_v^{(2)} \left(\frac{E_v^{(1,1)}}{E_v^{(1,2)}} \right) \varepsilon_n^{-1/\lambda}.$$

Now, for type 2 to occupy vertex v , the backwards flow from vertex v of the type 2 occupation process must reach one of the initial type 2 vertices *before* the backwards flow from vertex v of the type 1 occupation process reaches one of the initial type 1 vertices. Informally, this means that one of the $K_2^{(n)}$ vertices must be occupied by type 2. This has probability

$$(6.1.16) \quad K_2^{(n)} \frac{M_2^{(n)}}{n} \approx E_v^{(2)} \left(\frac{E_v^{(1,1)}}{E_v^{(1,2)}} \right)^{1/\lambda} \varepsilon_n^{-1/\lambda} (\varepsilon_n n)^{1/\lambda} \frac{E_2}{E_1^{1/\lambda}} = \frac{E_2}{E_1^{1/\lambda}} n^{1-1/\lambda} E_v^{(2)} \left(\frac{E_v^{(1,1)}}{E_v^{(1,2)}} \right)^{1/\lambda}.$$

This suggests that

$$\begin{aligned}
 \mathbb{P}(v \text{ type 2} \mid \mathcal{F}_{T_n^{(1)}}) &\approx \frac{E_2}{E_1^{1/\lambda}} n^{1-1/\lambda} \mathbb{E} \left[\left(\frac{E_v^{(1,1)}}{E_v^{(1,2)}} \right)^{1/\lambda} \right] \\
 (6.1.17) \qquad \qquad \qquad &= \Gamma(1 + 1/\lambda) \Gamma(1 - 1/\lambda) \frac{E_2}{E_1^{1/\lambda}}.
 \end{aligned}$$

Of course, the above analysis is far from a proof. There are several missing steps. First of all, we have only looked at the conditional expectation of $N_2^{(n)}$ given $\mathcal{F}_{T_n^{(1)}}$, rather than also treating the conditional variance, so as to show that conditionally on $\mathcal{F}_{T_n^{(1)}}$, the random variable $N_2^{(n)}/n$ converges in probability. Further, we have considered the occupation processes as being independent, which they of course clearly are not. In fact, we can run the processes independently from two vertices, but then we should process the competitive nature of the processes correctly, by killing occupied vertices (and all their offspring!) upon realizing that they have been occupied by the other type first. In the unequal speed context, however, this should not make much of a difference, as the competition process of type 1 moves much faster than that of type 2. It would be of interest to make the above argument precise.

Yet, it is pretty illustrative.

Remco: Update from here!

6.1.2. Competition on random regular graphs. Here we discuss [20].

6.1.3. Competition on scale-free configuration models. In the case where the configuration model has infinite-variance degrees, first-passage percolation has several universality classes (recall Section 3.6), corresponding to the explosive and conservative settings, respectively. Thus, one would also expect competition on such graphs to have rather different possible behavior. In this section, we discuss the available results in this setting. We first discuss the setting of *exponential* edge-weights, which corresponds to the explosive setting, and after this discuss *deterministic* edge-weights, which corresponds to a conservative setting. Our results show when co-existence can occur, as well as how many vertices are found by the losing type in the winner-takes-all scenarios. We also speculate whether co-existence holds in more general settings.

Competition on scale-free configuration models with exponential edge-weights. Here we discuss competition on scale-free configuration models with exponential edge-weights. Recall Theorem 3.31, where it was shown that the weight of the smallest weight path between two uniformly chosen individuals converges to the explosion time of two independent Bellman-Harris branching processes. Here we discuss the work with Deijfen [90] that investigates the competition problem. Again, we take $Y_e^{(1)} \sim \text{Exp}(\lambda)$ and $Y_e^{(2)} \sim \text{Exp}(1)$, where $\lambda \geq 1$ denotes the advantage of type 1 compared to type 2. The main result is as follows:

THEOREM 6.2 (The winner takes it all). *Fix $\lambda \geq 1$. Assume that there exist $\tau \in (2, 3)$ and $c_D \in (0, \infty)$ such that*

$$(6.1.18) \qquad [1 - F_D](x) = c_D x^{-(\tau-1)} (1 + o(1)).$$

- (a) The fraction $\bar{N}_1(n)$ of type 1 infected vertices converges in distribution to the indicator variable $\mathbb{1}_{\{Q_1 < \lambda Q_2\}}$ as $n \rightarrow \infty$, where Q_1 and Q_2 are the explosion times in Theorem 3.31.
- (b) The total number $N_{\text{los}}(n)$ of vertices occupied by the losing type converges in distribution to a proper discrete random variable N_{los} .

There are two surprises in Theorem 6.2. The first surprise is that both types have a positive probability of winning the largest part of the graph. Indeed, since Q_1, Q_2 has support on $(0, \infty)$, this is an easy exercise:

EXERCISE 6.6 (Both types can win). *Show that $\mathbb{P}(Q_1 < \lambda Q_2) \in (0, 1)$ for every $\lambda \in (0, \infty)$, so that each of the two types can win.*

This can be understood as follows. Due to the scale-free nature of $\text{CM}_n(\mathbf{D})$ when $\tau \in (2, 3)$, the hubs are the most relevant: the type that gets there first wins almost the entire market. Since each of the types explodes in finite time, each can explode first and thus also reach the hubs first. This explains the ‘both-can-win’ principle in Theorem 6.2.

The second surprise is that the other type wins so little, it really is an unfair world! In fact, the limiting random variable N_{los} in part (b) has an explicit characterization involving the (almost surely finite) extinction time of a certain Markov process, but it is quite involved indeed. We explain its shape in some more detail below. The proof reveals that the limiting number of vertices that is captured by the losing type is equal to 1 with strictly positive probability, which is the smallest possible value. Thus, the ABBA lyrics

‘The winner takes it all. The loser’s standing small...’

could not be more appropriate.

Let us first give a short heuristic explanation. Whp, the initially infected vertex 1 and vertex 2 will not be located very close to each other in the graph and hence the infection types will initially evolve without interfering with each other. This means that the initial stages of the spread of each one of the infections can be approximated by a continuous-time branching process, which has infinite mean when the degree distribution has infinite variance (because of size biasing). These two processes will both explode in finite time, and the type that explodes *first* is random and asymptotically equal to 1 precisely when $Q_1 < \lambda Q_2$. Theorem 6.2 follows from the fact that the type with the smallest explosion time will get a lead that is impossible to catch up with for the other type. More specifically, the type that explodes first will whp occupy all vertices of high degree – often referred to as *hubs* – in the graph shortly after the time of explosion, while the other type occupies only a finite number of vertices. From the hubs the exploding type will then rapidly invade the rest of the graph before the other type makes any substantial progress at all.

We next give some more details of the proof. Let a_n be some sequence that tends to infinity sufficiently fast, we take $a_n = n^{(\tau-2)/(\tau-1)}$ as this is the time at which the largest hub is typically found. Let R_m denote the time at which both types together reach m vertices for the first time. Define $T_{Ra_n}^{(1,2)}$ to be the time at which this happens. You should have in mind that $T_{Ra_n}^{(1,2)} \approx \min\{\mu Q_1, Q_2\}$, where $\mu = 1/\lambda$, i.e., the time at which the first type explodes. At this moment, the other type has not yet exploded, and thus has

only seen *finitely many* vertices. Let us define some of the necessary notation. We write $\mathcal{L}_n \in \{1, 2\}$ for the losing type and $\mathcal{W}_n \in \{1, 2\}$ for the winning type, and $\mu^{(\mathcal{W})}$ and $\mu^{(\mathcal{L})}$ are the mean edge traversal times of the winning and losing types, respectively. Thus, $(\mu^{(\mathcal{W})}, \mu^{(\mathcal{L})}) = (\mu, 1)$ when type 1 wins, and $(\mu^{(\mathcal{W})}, \mu^{(\mathcal{L})}) = (1, \mu)$ otherwise.

The fact that the losing type has only occupied finitely many vertices is formalized in the next lemma:

LEMMA 6.3 (Status at explosion time). *Let $N_{\mathcal{L}_n}^* = \max\{m: T_{R_m}^{(\mathcal{L}_n)} \leq T_{R_{a_n}}^{(1,2)}\}$. Then, as $n \rightarrow \infty$,*

$$(T_{R_{a_n}}^{(1,2)}, N_{\mathcal{L}_n}^*) \xrightarrow{d} (Q^{(\mathcal{W})}, N_{\text{los}}^*),$$

where

$$(6.1.19) \quad N_{\text{los}}^* \stackrel{d}{=} \max \left\{ m: \mu^{(\mathcal{L})} \sum_{j=1}^m E_j^{(\mathcal{L})} / S_j^{(\mathcal{L})} \leq Q^{(\mathcal{W})} \right\}.$$

After the approximate explosion at time $T_{R_{a_n}}^{(1,2)}$, the losing type thus has seen a finite number of vertices. This means that it will find the next vertex after a positive time. However, the winning vertex has already exploded, and will thus invade the rest of the graph extremely fast. Indeed, at time $T_{R_{a_n}}^{(1,2)} + \varepsilon$, the winning type will have conquered a *positive proportion* of the vertices in the graph, and the vertices with higher degree are more likely to have been found by the winning type. Formalizing this statement requires some more notation, as we need to keep track of the number of remaining vertices of each degree, as well as the total number of ‘free’ half-edges that remain.

Write $\bar{N}_{\mathcal{W}_n}^{(t,k)}$ for the fraction of vertices that have degree k and that have been captured by type \mathcal{W}_n at time $T_{R_{a_n}}^{(1,2)} + t$, that is,

$$(6.1.20) \quad \bar{N}_{\mathcal{W}_n}^{(t,k)} = \#\{v: d_v = k \text{ and } v \text{ is infected by type } \mathcal{W}_n \text{ at time } T_{R_{a_n}}^{(1,2)} + t\} / n.$$

We further define $Q(k)$ to be the explosion time of a continuous time branching process started from k individuals, so that $Q_i \stackrel{d}{=} Q(D)$. More precisely,

$$(6.1.21) \quad Q(k) = \sum_{j=0}^{\infty} E_j / S_j(k), \quad \text{where} \quad S_j(k) = k + \sum_{i=1}^j (\Delta_i - 1),$$

Then, the dynamics of the winning vertex after time the explosion time $T_{R_{a_n}}^{(1,2)}$ is given by the following proposition:

PROPOSITION 6.4 (Fraction of fixed degree winning type vertices and edges at fixed time). *As $n \rightarrow \infty$,*

$$(6.1.22) \quad \bar{N}_{\mathcal{W}_n}^{(t,k)} \xrightarrow{\mathbb{P}} \mathbb{P}(\mu^{(\mathcal{W})} Q(k) \leq t) \mathbb{P}(D = k).$$

EXERCISE 6.7 (Epidemic curve). *Perform first-passage percolation from a single vertex. Let T_{a_n} denote the time at which a_n vertices have been occupied. Let $\bar{N}_n(t)$ denote the proportion of vertices occupied at time $T_{a_n} + t$. The function $\bar{N}(t)$ is sometimes called the epidemic curve. Use Proposition 6.4 to show that, for every $t > 0$ (and thus uniformly in $t > 0$),*

$$(6.1.23) \quad \bar{N}_n(t) \xrightarrow{\mathbb{P}} \mathbb{P}(Q \leq t).$$

We have already argued that the losing type has only occupied finitely many vertices at time $T_{Ran}^{(1,2)}$. We now argue that it only finds a finite number after time $T_{Ran}^{(1,2)}$ as well. By Proposition 6.4, the winning type find more and more vertices. In particular, it is more likely to find vertices of high degree, since $Q(k)$ becomes quite small when $k \rightarrow \infty$, as can be concluded from the following exercise:

EXERCISE 6.8 (Explosion occurs more quickly from high-degree vertices). *Show that $Q(k) = \min_{i=1}^k Q_i(1)$, where $(Q_i(1))_{i \in [k]}$ are i.i.d. copies of $Q(1)$. Conclude that $Q(k) \xrightarrow{\mathbb{P}} 0$ as $k \rightarrow \infty$.*

By Exercise 6.8, it can be concluded that at *any* time $t > 0$, the degree distribution of the remaining vertices has very thin tails. In particular, the exploration of the losing type has only vertices with small degrees at its exposal, so that it becomes Malthusian. As a result, it grows at most exponentially, rather than being explosive. Thus, it will find finitely many vertices at *any time*. Further, the occupation process of the winning type is progressive, so less and less ground is left for the losing type. This results in the property that it finds only finitely many vertices. The precise distribution of N_{los} is derived in [90], and it is quite involved, as it needs tracking precisely how the losing type occupies territory while being pushed back by the winning type. We omit further details.

Competition on scale-free configuration models with deterministic edge-weights. Here we discuss the results on competition on configuration models with infinite-variance degrees and *deterministic* edge-weights, as proved with Baroni and Komjáthy for unequal speeds in [27] and with Komjáthy for equal speeds in [169]. We start with unequal speeds, where the results are arguably the least surprising. We will again assume that there exist constants c_D, C_D and $\tau \in (2, 3)$ such that

$$(6.1.24) \quad \frac{c_D}{x^{\tau-1}} \leq [1 - F_D](x) = \mathbb{P}(D > x) \leq \frac{C_D}{x^{\tau-1}}.$$

Let us introduce some notation. Let $Z_k^{(1)}, Z_k^{(2)}$ denote the number of individuals in the k th generation of two independent copies of a Galton-Watson process described as follows: the size of the first generation has distribution F satisfying (6.1.24), and all the further generations have offspring distribution F^* . Then, for a fixed but small $\rho > 0$ let us define

$$(6.1.25) \quad Y_1^{(n)} := (\tau - 2)^{t(n^\rho)} \log(Z_{t(n^\rho)}^{(1)}), \quad Y_2^{(n)} := (\tau - 2)^{\lfloor t(n^\rho)/\lambda \rfloor} \log(Z_{\lfloor t(n^\rho)/\lambda \rfloor}^{(2)}),$$

where $t(n^\rho) = \inf_k \{Z_k^{(\tau)} \geq n^\rho\}$. Let us further introduce

$$(6.1.26) \quad Y_1 := \lim_{k \rightarrow \infty} (\tau - 2)^k \log(Z_k^{(1)}), \quad Y_2 := \lim_{k \rightarrow \infty} (\tau - 2)^k \log(Z_k^{(2)}).$$

We see that $(Y_1^{(n)}, Y_2^{(n)}) \xrightarrow{d} (Y_1, Y_2)$ from (6.1.26) as $n \rightarrow \infty$. In terms of this notation, we have the following scaling behavior:

THEOREM 6.5 (Winner-takes-it all phenomenon for unequal speeds). *Fix $\lambda > 1$. Then,*

$$(6.1.27) \quad \frac{N_1^{(n)}}{n} \xrightarrow{\mathbb{P}} 1.$$

Further, there exists $H_n(Y_1^{(n)}, Y_2^{(n)})$ such that, as $n \rightarrow \infty$,

$$(6.1.28) \quad \frac{\log N_2^{(n)}}{(\log n)^{\frac{2}{\lambda+1}} H_n(Y_1^{(n)}, Y_2^{(n)})} \xrightarrow{d} \left(\frac{Y_2^\lambda}{Y_1} \right)^{\frac{1}{\lambda+1}}.$$

Here $H_n(Y_1^{(n)}, Y_2^{(n)})$ is a deterministic, oscillating (non-convergent) function of $\tau, \lambda, n, Y_1^{(n)}, Y_2^{(n)}$ that is uniformly bounded from above and below by finite and positive constants.

Thus, we see that, as expected, the faster species wins. Since it takes each of the species time of the order $\log \log n$ to reach a large part of the graph, the type that has a higher speed has an enormous advantage. This means that the faster species will get to the hub first, and then starts invading all the vertices of high degrees. At a certain moment, the two species will meet and that describes the scaling for the losing type in (6.1.28). We will give some more precise intuition for these results below, after also treating the case with equal speeds.

We next study the setting of equal speeds. When the types have equal speeds, again the type that reaches the highest-degree vertices first will have an enormous competitive advantage. Depending on the values of Y_1, Y_2 , we can decide who will get there first, but it is possible that both types get there roughly at the same time. Then, the situation is rather delicate, and we need to analyse what happens at the moment that both types try to conquer the highest-degree vertices extremely precisely. Here, we have to make an additional assumption on the precise distribution of the limiting variables Y_1, Y_2 , as their precise distributional properties are highly relevant. The following assumption requires these random variables to have an *absolutely continuous density* on some interval with zero on its boundary:

ASSUMPTION 6.6 (Absolute continuity branching process limit in infinite mean case). *Consider Z_k , the size of generation k in a branching process with offspring distribution F^* , and let $Y := \lim_{k \rightarrow \infty} (\tau - 2)^k \log Z_k$. We assume that the distribution function of Y is strictly increasing and absolutely continuous on some interval $(0, K)$, $K \in \mathbb{R}^+ \cup \{\infty\}$.*

When studying random graphs, one often arrives at settings where the branching process theory is not entirely developed, and this is an example. There is some literature on this problem though, particularly by Seneta [237, 238], who gives necessary conditions for Y to have a density.

Remco: Continue from here!

THEOREM 6.7 (Possible coexistence for equal speeds). *Consider the fixed speed competition on the configuration model started from two uniformly chosen vertices U_1, U_2 with independent tie-breaking rule with probability $p_{\text{tie}} \in (0, 1)$. The following results hold:*

- (1) *Let U_1, U_2 be such that $Y_{\min}^{(n)}/Y_{\max}^{(n)} \leq \tau - 2$. Then,*

$$(6.1.29) \quad \frac{N_{\text{win}}^{(n)}}{n} \xrightarrow{\mathbb{P}} 1,$$

while and $N_{\text{lose}}^{(n)} = n^{H_n(Y_{\max}^{(n)}, Y_{\min}^{(n)}) (1+o(1))}$ where $H_n(x, y) \in (0, 1)$ a function sequence that does not converge pointwise.

- (2) *Let U_1, U_2 be such that $Y_{\min}^{(n)}/Y_{\max}^{(n)} > \tau - 2$. Under Assumption 6.6, there exists a monotone increasing function $q_{p_{\text{tie}}} : (\tau - 2, 1] \rightarrow [0, 1/2]$, with $\lim_{x \downarrow \tau - 2} q_{p_{\text{tie}}}(x) = 0$ such that \mathbb{P}_Y -whp*

$$(6.1.30) \quad q_{p_{\text{tie}}}(Y_{\min}^{(n)}/Y_{\max}^{(n)}) < \frac{N_1^{(n)}}{n}, \frac{N_2^{(n)}}{n} < 1 - q_{p_{\text{tie}}}(Y_{\min}^{(n)}/Y_{\max}^{(n)}),$$

i.e., there is co-existence. The function $q_{p_{\text{tie}}}$ depends on the tie-breaking parameter $p_{\text{tie}} \in (0, 1)$.

The asymptotic probability of co-existence converges to $p_{\text{co}} = \mathbb{P}(Y_{\min}/Y_{\max} > \tau - 2)$.

A bootstrap-like coloring game. The way to establish the coexistence in Theorems 6.7 is to consider a third vertex w chosen uniformly in $[n] \setminus Z_{t(n\rho')}^{(1)} \cup Z_{t(n\rho')}^{(2)}$ and investigate the probability that this vertex is colored red and blue, respectively. As is the recurring theme in this text, the local neighborhood of such a vertex is well-approximated by a unimodular Galton-Watson tree, where the first generation has distribution F and the offspring in every consecutive generations is from F^* . The tree-approximation is only valid up to the moment that the total size of the tree is not too large or equivalently, the maximal degree in the tree does not exceed some value $Q = Q_n$. It is thus a stopped tree in the sense that its growth is stopped at the generation - say generation κ - where this condition is violated first. The problem that one needs to resolve is to understand how the two colors *enter* this random tree from its leaves and how they spread towards the root w .

From an analysis regarding how the types occupy the core of the graph (the high-degree vertices), it turns out that precisely when $Y_{\min}/Y_{\max} > \tau - 2$, the two colors reach the last generation $\Delta B_{\kappa}(w)$ of the tree *simultaneously*. At that moment, one of the types, which we will call the *winner type*, reaches *many more vertices* in $\Delta B_{\kappa}(w)$, namely, all vertices with degree larger than Q , while the losing type reaches all vertices with degree larger than Q^{γ} for some $\gamma > 1$. Thus, at that moment, among vertices in $\Delta B_{\kappa}(w)$, with degree at least Q^{γ} , the *tie-breaking rule* applies and each vertex is receives type 1 and 2 independently with probability p_{tie} and $1 - p_{\text{tie}}$, respectively. On the other hand, the vertices with degree in the interval $[Q, Q^{\gamma})$ are all occupied by the winning type. The rest of the vertices in $\Delta B_{\kappa}(w)$ is not yet assigned a type.

From this point on a race towards the root between the vertices of the different types starts, and the root w is receives its final type κ time steps later. The tie-breaking rule governs the coloring of the types towards the root: when a vertex in $\Delta B_i(w)$ has at least one child of each types in $\Delta B_{i+1}(w)$, then its type is decided by an i.i.d. coin flip. The question that then arises is as follows: What is the probability that the root is receives the loser/winner type, respectively, in this coloring scheme? Its answer provides the asymptotic proportion of vertices the loser/winner type occupies, respectively.

We find this problem interesting in its own right. It resembles bootstrap percolation (as discussed in Section 6.3), which explains the name of this paragraph. Here we make the above heuristics more formal and define the problem on Galton-Watson trees. By renaming the types if necessary, we assume that in the colouring scheme below the winning type is 1 and the losing type is 2.

Let us consider a unimodular Galton-Watson BP with root offspring distribution D . We write D_x^* for the degree of vertex x in this BP. Denote the set of vertices in generation i by Z_i . Fix $Q \geq 1$ and define the stopping time $\kappa = \kappa(Q)$ as

$$(6.1.31) \quad \kappa := \inf\{j: \max_{x \in Z_j} D_x^* \geq Q\}.$$

We fix $\gamma \in (1, 1/(\tau - 2))$, $p \in (0, 1)$ and partially decide the types of the vertices in Z_{κ} depending on their number of offspring (in generation $\kappa + 1$), *independently of each other* as follows

Starting Rule: a vertex $x \in Z_\kappa$ gets colour

- (i) red when $D_x^* \in [Q, Q^\gamma)$,
- (ii) blue with probability p , red with prob. $1 - p$, when $D_x^* \geq Q^\gamma$,
- (iii) neutral when $D_x^* < Q$.

Bootstrap rule: If a vertex $x \in Z_i$ has children in Z_{i+1} that

- (i) are all neutral, then vertex x gets colour neutral,
- (ii) are all either neutral or red, then it gets colour red,
- (iii) are all either neutral or blue, then it gets colour blue,
- (iv) have both colours red and blue, then it gets colour blue with probability p , otherwise red.

Note that the definition of κ ensures that there is at least one coloured vertex in Z_κ . Thus, the root of the BP will eventually be painted either red or blue.

PROPOSITION 6.8 (The branching process coloring game). *Fix $\gamma \in [1, 1/(\tau - 2))$, $p \in (0, 1)$ and consider the colouring scheme of a unimodular BP with root offspring distribution D . Assume further that Assumption 6.6 holds for F_D . Then there exist constants $0 < c_{p,\gamma} \leq C_{p,\gamma} < 1$ such that*

$$(6.1.32) \quad c_{p,\gamma} \leq \liminf_{Q \rightarrow \infty} \mathbb{P}(\text{root is occupied by 2}) \leq \limsup_{Q \rightarrow \infty} \mathbb{P}(\text{root is occupied by 2}) \leq C_{p,\gamma}.$$

Note that Proposition 6.8 is non-trivial since the proportion of blue vertices among all coloured vertices in the last generation tends to zero as $Q \rightarrow \infty$, and further, the generation where the process is stopped also tends to infinity as $Q \rightarrow \infty$. As a result of these two effects, a smaller and smaller proportion of blue vertices have to ‘make their way’ down to the root that is further and further away. Heuristically speaking, the rule that a vertex flips a coin that *does not depend on the number of its red and blue children* saves the blue colour: this effect ‘exaggerates’ the proportion of blue vertices as the generation number decreases towards the root.

To gain a more precise result on the proportion of vertices in the graph in Theorem 6.7, one has to improve upon the estimate for probability that the root is painted blue in this colouring scheme. In particular, the dependence of $\mathbb{P}(\text{root is blue})$ on the parameter γ directly translates to the dependence of $N_{\text{los}}^{(n)}/n$ on the ratio $q = Y_{\min}/Y_{\max} \in (\tau - 2, 1)$. A more precise result on $\mathbb{P}(\text{root is occupied by type 2})$ can only be achieved by knowing more about the distribution function of the limiting variable $Y = \lim_{k \rightarrow \infty} (\tau - 2)^k \log Z_k$.

6.1.4. Competition on preferential attachment models. [21]

6.2. Contact process

Organisation of this section.

6.2.1. Fixing ideas: Contact process on the complete graph.

6.2.2. Contact proces on configuration models.

6.2.3. Contact proces on preferential attachment models.

6.3. Bootstrap percolation

Bibliography

- [1] R. Abraham, J.-F. Delmas, and P. Hoscheit. A note on the Gromov-Hausdorff-Prokhorov distance between (locally) compact metric measure spaces. *Electron. J. Probab.*, **18**:no. 14, 1–21, (2013).
- [2] L. Addario-Berry. Partition functions of discrete coalescents: from Cayley’s formula to Frieze’s $\zeta(3)$ limit theorem. In *XI Symposium on Probability and Stochastic Processes*, volume **68** of *Progress in Probability Volume*. Birkhauser, Basel, (2015).
- [3] L. Addario-Berry, S. Bhamidi, and S. Sen. A probabilistic approach to the leader problem in random graphs. Available at arXiv:1703.09908 [math.PR], Preprint (2017).
- [4] L. Addario-Berry, N. Broutin, and C. Goldschmidt. Critical random graphs: limiting constructions and distributional properties. *Electron. J. Probab.*, **15**(25):741–775, (2010).
- [5] L. Addario-Berry, N. Broutin, and C. Goldschmidt. The continuum limit of critical random graphs. *Probab. Theory Related Fields*, **152**(3-4):367–406, (2012).
- [6] L. Addario-Berry, N. Broutin, C. Goldschmidt, and G. Miermont. The scaling limit of the minimum spanning tree of the complete graph. Available at arXiv:1301.1664 [math.PR], Preprint (2013).
- [7] L. Addario-Berry, N. Broutin, and G. Lugosi. The longest minimum-weight path in a complete graph. *Combin. Probab. Comput.*, **19**(1):1–19, (2010).
- [8] L. Addario-Berry, N. Broutin, and B. Reed. The diameter of the minimum spanning tree of a complete graph. In *Fourth Colloquium on Mathematics and Computer Science Algorithms, Trees, Combinatorics and Probabilities*, Discrete Math. Theor. Comput. Sci. Proc., AG, pages 237–248. Assoc. Discrete Math. Theor. Comput. Sci., Nancy, (2006).
- [9] D. Aldous. Brownian excursions, critical random graphs and the multiplicative coalescent. *Ann. Probab.*, **25**(2):812–854, (1997).
- [10] D. Aldous and V. Limic. The entrance boundary of the multiplicative coalescent. *Electron. J. Probab.*, **3**:No. 3, 59 pp. (electronic), (1998).
- [11] D. Aldous and R. Lyons. Processes on unimodular random networks. *Electron. J. Probab.*, **12**:no. 54, 1454–1508, (2007).
- [12] D. Aldous and J. Steele. The objective method: probabilistic combinatorial optimization and local weak convergence. In *Probability on discrete structures*, volume **110** of *Encyclopaedia Math. Sci.*, pages 1–72. Springer, Berlin, (2004).
- [13] N. Alon and J. Spencer. *The probabilistic method*. Wiley-Interscience Series in Discrete Mathematics and Optimization. John Wiley & Sons, New York, second edition, (2000).
- [14] H. Amini, M. Draief, and M. Lelarge. Flooding in weighted sparse random graphs. *SIAM Journal on Discrete Mathematics*, **27**(1):1–26, (2013).

- [15] H. Amini and M. Lelarge. Upper deviations for split times of branching processes. *J. Appl. Probab.*, **49**(4):1134–1143, (2012).
- [16] H. Amini and M. Lelarge. The diameter of weighted random graphs. *Ann. Appl. Probab.*, **25**(3):1686–1727, (2015).
- [17] H. Amini and Y. Peres. Shortest-weight paths in random regular graphs. *SIAM J. Discrete Math.*, **28**(2):656–672, (2014).
- [18] O. Amini, L. Devroye, S. Griffiths, and N. Olver. On explosions in heavy-tailed branching random walks. *Ann. Probab.*, **41**(3B):1864–1899, (2013).
- [19] V. Anantharam and J. Salez. The densest subgraph problem in sparse random graphs. *Ann. Appl. Probab.*, **26**(1):305–327, (2016).
- [20] T. Antunovic, Y. Dekel, E. Mossell, and Y. Peres. Competiting first passage percolation on random regular graphs. Available at arXiv:1109.2575 [math.PR], Preprint (2011).
- [21] T. Antunović, E. Mossel, and M. Rácz. Coexistence in preferential attachment networks. *Combin. Probab. Comput.*, **25**(6):797–822, (2016).
- [22] R. Arratia, A. Barbour, and S. Tavaré. *Logarithmic combinatorial structures: a probabilistic approach*. European Mathematical Society, (2003).
- [23] K. Athreya and P. Ney. *Branching processes*. Springer-Verlag, New York, (1972). Die Grundlehren der mathematischen Wissenschaften, Band 196.
- [24] A. Auffinger, M. Damron, and J. Hanson. 50 years of first passage percolation. Available at arXiv:1511.03262 [math.PR], Preprint (2015).
- [25] A.-L. Barabási. *Linked: The New Science of Networks*. Perseus Publishing, Cambridge, Massachusetts, (2002).
- [26] A.-L. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, **286**(5439):509–512, (1999).
- [27] E. Baroni, R. v. d. Hofstad, and J. Komjáthy. Fixed speed competition on the configuration model with infinite variance degrees: unequal speeds. *Electron. J. Probab.*, **20**:Paper No. 116, 48, (2015).
- [28] E. Baroni, R. v. d. Hofstad, and J. Komjáthy. Tight fluctuations of weight-distances in random graphs with infinite-variance degrees. *Preprint arXiv:1609.07269 [math.PR]*, (2016).
- [29] E. Baroni, R. v. d. Hofstad, and J. Komjáthy. Nonuniversality of weighted random graphs with infinite variance degree. *J. Appl. Probab.*, **54**(1):146–164, (2017).
- [30] R. Baxter. *Exactly solved models in statistical mechanics*. Academic Press Inc. [Harcourt Brace Jovanovich Publishers], London, (1982).
- [31] E. Bender and E. Canfield. The asymptotic number of labelled graphs with given degree sequences. *Journal of Combinatorial Theory (A)*, **24**:296–307, (1978).
- [32] I. Benjamini, R. Lyons, and O. Schramm. Unimodular random trees. *Ergodic Theory Dynam. Systems*, **35**(2):359–373, (2015).
- [33] I. Benjamini and O. Schramm. Recurrence of distributional limits of finite planar graphs. *Electron. J. Probab.*, **6**:no. 23, 13 pp. (electronic), (2001).
- [34] N. Berger, C. Borgs, J. Chayes, and A. Saberi. Asymptotic behavior and distributional limits of preferential attachment graphs. *Ann. Probab.*, **42**(1):1–40, (2014).
- [35] S. Bhamidi, N. Broutin, S. Sen, and X. Wang. Scaling limits of random graph models at criticality: Universality and the basin of attraction of the Erdős-Rényi random graph. Available at arXiv:1411.3417 [math.PR], Preprint (2014).

- [36] S. Bhamidi, A. Budhiraja, and X. Wang. The augmented multiplicative coalescent and critical dynamic random graph models. Available at arXiv:1212.5493 [math.PR], Preprint (2012).
- [37] S. Bhamidi, S. Dhara, R. van der Hofstad, and S. Sen. Universality for critical heavy-tailed network models: Gromov-Hausdorff-Prohorov convergence. In preparation, (2017+).
- [38] S. Bhamidi, S. Dhara, R. van der Hofstad, and S. Sen. Universality for critical heavy-tailed network models: metric structure of maximal components. Available at arXiv:1703.07145 [math.PR], Preprint (2017).
- [39] S. Bhamidi and R. v. d. Hofstad. Diameter of the stochastic mean-field model of distance. To appear in *Combinatorics, Probability and Computing*, Preprint (2013).
- [40] S. Bhamidi, R. v. d. Hofstad, and G. Hooghiemstra. Extreme value theory, Poisson-Dirichlet distributions, and first passage percolation on random networks. *Adv. in Appl. Probab.*, **42**(3):706–738, (2010).
- [41] S. Bhamidi, R. v. d. Hofstad, and G. Hooghiemstra. First passage percolation on random graphs with finite mean degrees. *Ann. Appl. Probab.*, **20**(5):1907–1965, (2010).
- [42] S. Bhamidi, R. v. d. Hofstad, and G. Hooghiemstra. First passage percolation on the Erdős-Rényi random graph. *Combin. Probab. Comput.*, **20**(5):683–707, (2011).
- [43] S. Bhamidi, R. v. d. Hofstad, and J. v. Leeuwaarden. Scaling limits for critical inhomogeneous random graphs with finite third moments. *Electronic Journal of Probability*, **15**:1682–1702, (2010).
- [44] S. Bhamidi, R. v. d. Hofstad, and J. v. Leeuwaarden. Novel scaling limits for critical inhomogeneous random graphs. *Ann. Probab.*, **40**(6):2299–2361, (2012).
- [45] S. Bhamidi, R. v. d. Hofstad, and S. Sen. The multiplicative coalescent, inhomogeneous continuum random trees, and new universality classes for critical random graphs. To appear in *Probab. Theory Applic.* Available at arXiv:1508.04645 [math.PR], Preprint (2015).
- [46] S. Bhamidi and S. Sen. Geometry of the vacant set left by random walk on random graphs, wright’s constants, and critical random graphs with prescribed degrees. Available at arXiv: 1608.07153 [math.PR], Preprint (2016).
- [47] S. Bhamidi and R. van der Hofstad. Weak disorder asymptotics in the stochastic mean-field model of distance. *Ann. Appl. Probab.*, **22**(1):29–69, (2012).
- [48] B. Bollobás. A probabilistic proof of an asymptotic formula for the number of labelled regular graphs. *European J. Combin.*, **1**(4):311–316, (1980).
- [49] B. Bollobás. The evolution of random graphs. *Trans. Amer. Math. Soc.*, **286**(1):257–274, (1984).
- [50] B. Bollobás. The evolution of sparse graphs. In *Graph theory and combinatorics (Cambridge, 1983)*, pages 35–57. Academic Press, London, (1984).
- [51] B. Bollobás. *Random graphs*. Academic Press, Inc. [Harcourt Brace Jovanovich, Publishers], London, (1985).
- [52] B. Bollobás. *Random graphs*, volume **73** of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, second edition, (2001).
- [53] B. Bollobás, C. Borgs, J. Chayes, and O. Riordan. Directed scale-free graphs. In *Proceedings of the Fourteenth Annual ACM-SIAM Symposium on Discrete Algorithms (Baltimore, MD, 2003)*, pages 132–139, New York, (2003). ACM.

- [54] B. Bollobás, D. Gamarnik, O. Riordan, and B. Sudakov. On the value of a random minimum weight Steiner tree. *Combinatorica*, **24**(2):187–207, (2004).
- [55] B. Bollobás, S. Janson, and O. Riordan. The phase transition in the uniformly grown random graph has infinite order. *Random Structures Algorithms*, **26**(1-2):1–36, (2005).
- [56] B. Bollobás, S. Janson, and O. Riordan. The phase transition in inhomogeneous random graphs. *Random Structures Algorithms*, **31**(1):3–122, (2007).
- [57] B. Bollobás and O. Riordan. The diameter of a scale-free random graph. *Combinatorica*, **24**(1):5–34, (2004).
- [58] B. Bollobás and O. Riordan. *Percolation*. Cambridge University Press, New York, (2006).
- [59] B. Bollobás, O. Riordan, J. Spencer, and G. Tusnády. The degree sequence of a scale-free random graph process. *Random Structures Algorithms*, **18**(3):279–290, (2001).
- [60] C. Borgs, J. Chayes, R. v. d. Hofstad, G. Slade, and J. Spencer. Random subgraphs of finite graphs. I. The scaling window under the triangle condition. *Random Structures Algorithms*, **27**(2):137–184, (2005).
- [61] C. Borgs, J. Chayes, R. v. d. Hofstad, G. Slade, and J. Spencer. Random subgraphs of finite graphs. II. The lace expansion and the triangle condition. *Ann. Probab.*, **33**(5):1886–1944, (2005).
- [62] C. Borgs, J. Chayes, R. v. d. Hofstad, G. Slade, and J. Spencer. Random subgraphs of finite graphs. III. The phase transition for the n -cube. *Combinatorica*, **26**(4):395–410, (2006).
- [63] S. Bornholdt and F. Wagner. Stability of money: phase transitions in an ising economy. *Physica A: Statistical Mechanics and its Applications*, **316**(1):453–468, 2002.
- [64] A. Bovier. *Statistical mechanics of disordered systems*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, (2006). A mathematical perspective.
- [65] A. Bovier and F. den Hollander. *Metastability*, volume **351** of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer, Cham, (2015). A potential-theoretic approach.
- [66] A. Bovier and I. Kurkova. A short course on mean field spin glasses. In *Spin glasses: statics and dynamics*, volume **62** of *Progr. Probab.*, pages 3–44. Birkhäuser Verlag, Basel, (2009).
- [67] L. Braunstein, S. Buldyrev, R. Cohen, S. Havlin, and H. Stanley. Optimal paths in disordered complex networks. *Physical review letters*, **91**(16):168701, (2003).
- [68] T. Britton, M. Deijfen, and A. Martin-Löf. Generating simple random graphs with prescribed degree distribution. *J. Stat. Phys.*, **124**(6):1377–1397, (2006).
- [69] N. Broutin and J.-F. Marckert. A new encoding of coalescent processes: applications to the additive and multiplicative cases. *Probab. Theory Related Fields*, **166**(1-2):515–552, (2016).
- [70] W. Bühler. Generations and degree of relationship in supercritical Markov branching processes. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, **18**:141–152, (1971).
- [71] D. Burago, Y. Burago, and S. Ivanov. *A course in metric geometry*, volume **33** of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, (2001).

- [72] D. Callaway, J. Hopcroft, J. M. Kleinberg, M. E. J. Newman, and S. Strogatz. Are randomly grown graphs really random? *Physical Review E*, **64**:041902, (2001).
- [73] V. Can. Annealed limit theorems for the Ising model on random regular graphs. Available at arXiv: 1701.08639 [math.PR], Preprint (2017).
- [74] V. Can. Critical behavior of the annealed Ising model on random regular graphs. Available at arXiv: 1701.08628 [math.PR], Preprint (2017).
- [75] F. Caravenna, A. Garavaglia, and R. v. d. Hofstad. Diameter in ultra-small scale-free random graphs. Preprint arXiv:1605.02714 [math.PR], (2016).
- [76] J. Černý and A. Teixeira. *From random walk trajectories to random interlacements*. Rio de Janeiro: Sociedade Brasileira de Matemática, (2012).
- [77] J. Černý and A. Teixeira. Critical window for the vacant set left by random walk on random regular graphs. *Random Structures Algorithms*, **43**(3):313–337, (2013).
- [78] J. Černý, A. Teixeira, and D. Windisch. Giant vacant component left by a random walk in a random d -regular graph. *Ann. Inst. Henri Poincaré Probab. Stat.*, **47**(4):929–968, (2011).
- [79] F. Chung and L. Lu. The average distances in random graphs with given expected degrees. *Proc. Natl. Acad. Sci. USA*, **99**(25):15879–15882 (electronic), (2002).
- [80] F. Chung and L. Lu. Connected components in random graphs with given expected degree sequences. *Ann. Comb.*, **6**(2):125–145, (2002).
- [81] F. Chung and L. Lu. The average distance in a random graph with given expected degrees. *Internet Math.*, **1**(1):91–113, (2003).
- [82] F. Chung and L. Lu. *Complex graphs and networks*, volume **107** of *CBMS Regional Conference Series in Mathematics*. Published for the Conference Board of the Mathematical Sciences, Washington, DC, (2006).
- [83] F. Chung and L. Lu. The volume of the giant component of a random graph with given expected degrees. *SIAM J. Discrete Math.*, **20**:395–411, (2006).
- [84] P. Contucci, S. Dommers, C. Giardinà, and S. Starr. Antiferromagnetic Potts model on the Erdős-Rényi random graph. *Comm. Math. Phys.*, **323**(2):517–554, (2013).
- [85] P. Contucci, I. Gallo, and G. Menconi. Phase transitions in social sciences: two-population mean field theory. *International Journal of Modern Physics B*, **22**(14):2199–2212, 2008.
- [86] P. Contucci and C. Giardinà. *Perspectives on spin glasses*. Cambridge University Press, Cambridge, (2013).
- [87] C. Cooper and A. Frieze. Component structure of the vacant set induced by a random walk on a random graph. *Random Structures Algorithms*, **42**(2):135–158, (2013).
- [88] P. L. Davies. The simple branching process: a note on convergence when the mean is infinite. *J. Appl. Probab.*, **15**(3):466–480, (1978).
- [89] L. De Sanctis and F. Guerra. Mean field dilute ferromagnet: high temperature and zero temperature behavior. *J. Stat. Phys.*, **132**(5):759–785, (2008).
- [90] M. Deijfen and R. v. d. Hofstad. The winner takes it all. (2016).
- [91] A. Dembo and A. Montanari. Gibbs measures and phase transitions on sparse random graphs. *Braz. J. Probab. Stat.*, **24**(2):137–211, (2010).
- [92] A. Dembo and A. Montanari. Ising models on locally tree-like graphs. *The Annals of Applied Probability*, **20**(2):565–592, (2010).

- [93] A. Dembo, A. Montanari, A. Sly, and N. Sun. The replica symmetric solution for Potts models on d -regular graphs. *Comm. Math. Phys.*, **327**(2):551–575, (2014).
- [94] A. Dembo, A. Montanari, and N. Sun. Factor models on locally tree-like graphs. *Ann. Probab.*, **41**(6):4162–4213, (2013).
- [95] A. Dembo and A.-S. Sznitman. On the disconnection of a discrete cylinder by a random walk. *Probab. Theory Related Fields*, **136**(2):321–340, (2006).
- [96] S. Dereich, C. Mönch, and P. Mörters. Typical distances in ultrasmall random networks. *Adv. in Appl. Probab.*, **44**(2):583–601, (2012).
- [97] S. Dereich, C. Mönch, and P. Mörters. Distances in scale free networks at criticality. Available at arXiv:1604.00779 [math.PR], Preprint (2016).
- [98] S. Dereich and P. Mörters. Random networks with sublinear preferential attachment: Degree evolutions. *Electronic Journal of Probability*, **14**:122–1267, (2009).
- [99] S. Dereich and P. Mörters. Random networks with concave preferential attachment rule. *Jahresber. Dtsch. Math.-Ver.*, **113**(1):21–40, (2011).
- [100] S. Dereich and P. Mörters. Random networks with sublinear preferential attachment: the giant component. *Ann. Probab.*, **41**(1):329–384, (2013).
- [101] L. Devroye. Branching processes in the analysis of the heights of trees. *Acta Inform.*, **24**(3):277–298, (1987).
- [102] L. Devroye and N. Fraiman. Connectivity of inhomogeneous random graphs. *Random Structures Algorithms*, **45**(3):408–420, (2014).
- [103] S. Dhara, R. v. d. Hofstad, J. Leeuwaarden, and S. Sen. Critical window for the configuration model: finite third moment degrees. Available at arXiv:1605.02868 [math.PR], Preprint (2016).
- [104] S. Dhara, R. v. d. Hofstad, J. Leeuwaarden, and S. Sen. Heavy-tailed configuration models at criticality. Available at arXiv:1612.00650 [math.PR], Preprint (2016).
- [105] J. Ding, J. H. Kim, E. Lubetzky, and Y. Peres. Diameters in supercritical random graphs via first passage percolation. *Combin. Probab. Comput.*, **19**(5-6):729–751, (2010).
- [106] J. Ding, J. H. Kim, E. Lubetzky, and Y. Peres. Anatomy of a young giant component in the random graph. *Random Structures Algorithms*, **39**(2):139–178, (2011).
- [107] S. Dommers. Metastability of the Ising model on random regular graphs at zero temperature. *Probab. Theory Related Fields*, **167**(1-2):305–324, (2017).
- [108] S. Dommers, C. Giardinà, C. Giberti, R. v. d. Hofstad, and M. Prioriello. Ising critical behavior of inhomogeneous Curie-Weiss models and annealed random graphs. *Comm. Math. Phys.*, **348**(1):221–263, (2016).
- [109] S. Dommers, C. Giardinà, and R. v. d. Hofstad. Ising models on power-law random graphs. *J. Stat. Phys.*, **141**(4):638–660, (2010).
- [110] S. Dommers, C. Giardinà, and R. v. d. Hofstad. Ising critical exponents on random trees and graphs. *Comm. Math. Phys.*, **328**(1):355–395, (2014).
- [111] S. Dommers, R. v. d. Hofstad, and G. Hooghiemstra. Diameters in preferential attachment graphs. *Journ. Stat. Phys.*, **139**:72–107, (2010).
- [112] S. Dorogovtsev, A. Goltsev, and J. Mendes. Pseudofractal scale-free web. *Phys. Rev. E.*, **65**:066122, (2002).
- [113] S. Dorogovtsev, A. Goltsev, and J. Mendes. Critical phenomena in complex networks. *Reviews of Modern Physics*, **80**(4):1275–1335, (2008).

- [114] M. Draief and L. Massoulié. *Epidemics and rumours in complex networks*, volume **369** of *London Mathematical Society Lecture Note Series*. Cambridge University Press, Cambridge, (2010).
- [115] R. Durrett. *Lecture notes on particle systems and percolation*. The Wadsworth & Brooks/Cole Statistics/Probability Series. Wadsworth & Brooks/Cole Advanced Books & Software, Pacific Grove, CA, (1988).
- [116] R. Durrett. Rigorous result for the CHKNS random graph model. In *Discrete random walks (Paris, 2003)*, Discrete Math. Theor. Comput. Sci. Proc., AC, pages 95–104. Assoc. Discrete Math. Theor. Comput. Sci., Nancy, (2003).
- [117] M. Eckhoff, J. Goodman, R. v. d. Hofstad, and F. Nardi. Short paths for first passage percolation on the complete graph. *J. Stat. Phys.*, **151**(6):1056–1088, (2013).
- [118] M. Eckhoff, J. Goodman, R. v. d. Hofstad, and F. Nardi. Long paths in first passage percolation on the complete graph I. Local PWIT dynamics. Available at arXiv:1512.06152 [math.PR], Preprint (2015).
- [119] M. Eckhoff, J. Goodman, R. v. d. Hofstad, and F. Nardi. Long paths in first passage percolation on the complete graph II. Global branching dynamics. Available at arXiv:1512.06145 [math.PR], Preprint (2015).
- [120] M. Eckhoff and P. Mörters. Vulnerability of robust preferential attachment networks. *Electron. J. Probab.*, **19**:no. 57, 47, (2014).
- [121] R. Ellis. *Entropy, large deviations, and statistical mechanics*, volume **271** of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, New York, (1985).
- [122] P. Erdős and A. Rényi. On random graphs. I. *Publ. Math. Debrecen*, **6**:290–297, (1959).
- [123] P. Erdős and A. Rényi. On the evolution of random graphs. *Magyar Tud. Akad. Mat. Kutató Int. Közl.*, **5**:17–61, (1960).
- [124] P. Erdős and A. Rényi. On the evolution of random graphs. *Bull. Inst. Internat. Statist.*, **38**:343–347, (1961).
- [125] P. Erdős and A. Rényi. On the strength of connectedness of a random graph. *Acta Math. Acad. Sci. Hungar.*, **12**:261–267, (1961).
- [126] H. v. d. Esker, R. v. d. Hofstad, and G. Hooghiemstra. Universality for the distance in finite variance random graphs. *J. Stat. Phys.*, **133**(1):169–202, (2008).
- [127] H. v. d. Esker, R. v. d. Hofstad, G. Hooghiemstra, and D. Znamenski. Distances in random graphs with infinite mean degrees. *Extremes*, **8**:111–140, (2006).
- [128] S. N. Ethier and T. G. Kurtz. *Markov processes*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons Inc., New York, (1986). Characterization and convergence.
- [129] W. Evans, C. Kenyon, Y. Peres, and L. Schulman. Broadcasting on trees and the Ising model. *Ann. Appl. Probab.*, **10**(2):410–433, (2000).
- [130] C. Faloutsos, P. Faloutsos, and M. Faloutsos. On power-law relationships of the internet topology. *Computer Communications Rev.*, **29**:251–262, (1999).
- [131] L. Federico and R. v. d. Hofstad. Critical window for connectivity in the configuration model. Available at arXiv:1603.03254 [math.PR], Preprint (2016).
- [132] D. Fernholz and V. Ramachandran. The diameter of sparse random graphs. *Random Structures Algorithms*, **31**(4):482–516, (2007).
- [133] N. Fountoulakis. Percolation on sparse random graphs with given degree sequence. *Internet Math.*, **4**(4):329–356, (2007).

- [134] D. Fraiman, P. Balenzuela, J. Foss, and D. Chialvo. Ising-like dynamics in large-scale functional brain networks. *Physical Review E*, **79**(6):061922, (2009).
- [135] A. M. Frieze. On the value of a random minimum spanning tree problem. *Discrete Appl. Math.*, **10**(1):47–56, (1985).
- [136] O. Garet and R. Marchand. Coexistence in two-type first-passage percolation models. *Ann. Appl. Probab.*, **15**(1A):298–330, (2005).
- [137] C. Giardinà, C. Giberti, R. van der Hofstad, and M. Prioriello. Quenched central limit theorems for the Ising model on random graphs. *J. Stat. Phys.*, **160**(6):1623–1657, (2015).
- [138] C. Giardinà, C. Giberti, R. van der Hofstad, and M. L. Prioriello. Annealed central limit theorems for the Ising model on random graphs. *ALEA Lat. Am. J. Probab. Math. Stat.*, **13**(1):121–161, (2016).
- [139] E. N. Gilbert. Random graphs. *Ann. Math. Statist.*, **30**:1141–1144, (1959).
- [140] A. Greven, P. Pfaffelhuber, and A. Winter. Convergence in distribution of random metric measure spaces (Λ -coalescent measure trees). *Probab. Theory Related Fields*, **145**(1-2):285–322, (2009).
- [141] D. R. Grey. Explosiveness of age-dependent branching processes. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, **28**:129–137, (1973/74).
- [142] R. Griffiths. Correlations in Ising ferromagnets. II. External magnetic fields. *Journal of Mathematical Physics*, **8**:484, (1967).
- [143] R. B. Griffiths, C. A. Hurst, and S. Sherman. Concavity of magnetization of an Ising ferromagnet in a positive external field. *J. Mathematical Phys.*, **11**:790–795, (1970).
- [144] G. Grimmett. *Percolation*. Springer, Berlin, 2nd edition, (1999).
- [145] G. Grimmett. *The random-cluster model*. Berlin: Springer, 2006.
- [146] G. Grimmett and D. Stirzaker. *Probability and random processes*. Oxford University Press, New York, third edition, (2001).
- [147] M. Gromov. *Metric structures for Riemannian and non-Riemannian spaces*. Modern Birkhäuser Classics. Birkhäuser Boston, Inc., Boston, MA, english edition, (2007). Based on the 1981 French original. With appendices by M. Katz, P. Pansu and S. Semmes, Translated from the French by Sean Michael Bates.
- [148] L. Gulikers. Explosiveness of age-dependent branching processes with contagious and incubation periods. Preprint (2015).
- [149] O. Häggström and R. Pemantle. First passage percolation and a model for competing spatial growth. *J. Appl. Probab.*, **35**(3):683–692, (1998).
- [150] O. Häggström and R. Pemantle. Absence of mutual unbounded growth for almost all parameter values in the two-type Richardson model. *Stochastic Process. Appl.*, **90**(2):207–222, (2000).
- [151] P. Halmos. *Measure Theory*. D. Van Nostrand Company, Inc., New York, N. Y., (1950).
- [152] J. M. Hammersley. First-passage percolation. *J. Roy. Statist. Soc. Ser. B*, **28**:491–496, (1966).
- [153] T. Harris. *The theory of branching processes*. Die Grundlehren der Mathematischen Wissenschaften, Bd. 119. Springer-Verlag, Berlin, (1963).
- [154] S. Havlin, L. Braunstein, S. Buldyrev, R. Cohen, T. Kalisky, S. Sreenivasan, and E. Stanley. Optimal path in random networks with disorder: A mini review. *Physica A: Statistical Mechanics and its Applications*, **346**(1-2):82–92, (2005).

- [155] M. Heydenreich and R. v. d. Hofstad. Random graph asymptotics on high-dimensional tori. *Comm. Math. Phys.*, **270**(2):335–358, (2007).
- [156] M. Heydenreich and R. v. d. Hofstad. Random graph asymptotics on high-dimensional tori II: volume, diameter and mixing time. *Probab. Theory Related Fields*, **149**(3-4):397–415, (2011).
- [157] M. Heydenreich and R. v. d. Hofstad. Progress in high-dimensional percolation and random graphs. To appear in the *CRM Short Courses Series* with Springer, (2017).
- [158] C. Hoffman. Coexistence for Richardson type competing spatial growth models. *Ann. Appl. Probab.*, **15**(1B):739–747, (2005).
- [159] R. v. d. Hofstad. Critical behavior in inhomogeneous random graphs. *Random Structures and Algorithms*, **42**(4):480–508, (2013).
- [160] R. v. d. Hofstad. *Random graphs and complex networks. Volume 1.* Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, (2017).
- [161] R. v. d. Hofstad. *Random graphs and complex networks. Volume 2.* (2018+). In preparation, see <http://www.win.tue.nl/~rhofstad/NotesRGCNII.pdf>.
- [162] R. v. d. Hofstad, G. Hooghiemstra, and P. Van Mieghem. Distances in random graphs with finite variance degrees. *Random Structures Algorithms*, **27**(1):76–123, (2005).
- [163] R. v. d. Hofstad, G. Hooghiemstra, and P. Van Mieghem. Size and weight of shortest path trees with exponential link weights. *Combin. Probab. Comput.*, **15**(6):903–926, (2006).
- [164] R. v. d. Hofstad, G. Hooghiemstra, and P. Van Mieghem. The weight of the shortest path tree. *Random Structures Algorithms*, **30**(3):359–379, (2007).
- [165] R. v. d. Hofstad, G. Hooghiemstra, and D. Znamenski. Distances in random graphs with finite mean and infinite variance degrees. *Electron. J. Probab.*, **12**(25):703–766 (electronic), (2007).
- [166] R. v. d. Hofstad, S. Janson, and M. Luczak. Component structure of the configuration model: barely supercritical case. Available at arXiv:1611.05728 [math.PR], Preprint (20126).
- [167] R. v. d. Hofstad, W. Kager, and T. Müller. A local limit theorem for the critical random graph. *Electron. Commun. Probab.*, **14**:122–131, (2009).
- [168] R. v. d. Hofstad and M. Keane. An elementary proof of the hitting time theorem. *Amer. Math. Monthly*, **115**(8):753–756, (2008).
- [169] R. v. d. Hofstad and J. Komjáthy. Fixed speed competition on the configuration model with infinite variance degrees: equal speeds. Available at arXiv:1503.09046 [math.PR], Preprint (2015).
- [170] R. v. d. Hofstad and J. Komjáthy. When is a scale-free graph ultra-small? Available at arXiv:1611.03639 [math.PR], Preprint (2016).
- [171] R. v. d. Hofstad, J. v. Leeuwaarden, and C. Stegehuis. Hierarchical configuration model. Available at arXiv:1512.08397 [math.PR], Preprint (2015).
- [172] R. v. d. Hofstad and A. Nachmias. Unlacing hypercube percolation: a survey. *Metrika*, **77**(1):23–50, 2014.
- [173] R. v. d. Hofstad and A. Nachmias. Hypercube percolation. To appear in *Journ. Europ. Math. Soc.*, Preprint (2012).
- [174] R. v. d. Hofstad and J. Spencer. Counting connected graphs asymptotically. *European J. Combin.*, **27**(8):1294–1320, (2006).

- [175] C. Howard. Models of first-passage percolation. In *Probability on discrete structures*, volume **110** of *Encyclopaedia Math. Sci.*, pages 125–173. Springer, Berlin, (2004).
- [176] K. Itô, editor. *Encyclopedic dictionary of mathematics. Vol. I–IV*. MIT Press, Cambridge, MA, second edition, (1987). Translated from the Japanese.
- [177] P. Jagers. *Branching processes with biological applications*. Wiley-Interscience [John Wiley & Sons], London, (1975). Wiley Series in Probability and Mathematical Statistics—Applied Probability and Statistics.
- [178] P. Jagers and O. Nerman. The growth and composition of branching populations. *Adv. in Appl. Probab.*, **16**(2):221–259, (1984).
- [179] S. Janson. One, two and three times $\log n/n$ for paths in a complete graph with random weights. *Combin. Probab. Comput.*, **8**:347–361, (1999).
- [180] S. Janson. The largest component in a subcritical random graph with a power law degree distribution. *Ann. Appl. Probab.*, **18**(4):1651–1668, (2008).
- [181] S. Janson. On percolation in random graphs with given vertex degrees. *Electron. J. Probab.*, **14**:no. 5, 87–118, (2009).
- [182] S. Janson. Asymptotic equivalence and contiguity of some random graphs. *Random Structures Algorithms*, **36**(1):26–45, (2010).
- [183] S. Janson. Susceptibility of random graphs with given vertex degrees. *J. Comb.*, **1**(3-4):357–387, (2010).
- [184] S. Janson, D. Knuth, T. Łuczak, and B. Pittel. The birth of the giant component. *Random Structures Algorithms*, **4**(3):231–358, (1993). With an introduction by the editors.
- [185] S. Janson and M. Łuczak. A new approach to the giant component problem. *Random Structures Algorithms*, **34**(2):197–216, (2009).
- [186] S. Janson, T. Łuczak, and A. Ruciński. *Random graphs*. Wiley-Interscience Series in Discrete Mathematics and Optimization. Wiley-Interscience, New York, (2000).
- [187] S. Janson and J. Spencer. A point process describing the component sizes in the critical window of the random graph evolution. *Combin. Probab. Comput.*, **16**(4):631–658, (2007).
- [188] S. Janson and L. Warnke. On the critical probability in percolation. Available at arXiv:1611.08549 [math.PR], Preprint (2016).
- [189] A. Joseph. The component sizes of a critical random graph with given degree sequence. *Ann. Appl. Probab.*, **24**(6):2560–2594, (2014).
- [190] S. Kalikow and B. Weiss. When are random graphs connected? *Israel J. Math.*, **62**(3):257–268, (1988).
- [191] O. Kallenberg. *Random measures*. Akademie-Verlag, Berlin, fourth edition, (1986).
- [192] I. Karatzas and S. E. Shreve. *Brownian motion and stochastic calculus*, volume 113 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, (1991).
- [193] D. Kelly and S. Sherman. General Griffiths’ inequalities on correlations in Ising ferromagnets. *Journal of Mathematical Physics*, **9**:466, (1968).
- [194] H. Kesten. *Percolation theory for mathematicians*, volume **2** of *Progress in Probability and Statistics*. Birkhäuser Boston, Mass., (1982).
- [195] H. Kesten. Aspects of first passage percolation. In *École d’été de probabilités de Saint-Flour, XIV—1984*, volume **1180** of *Lecture Notes in Math.*, pages 125–264. Springer, Berlin, (1986).

- [196] H. Kesten and B. P. Stigum. Limit theorems for decomposable multi-dimensional Galton-Watson processes. *J. Math. Anal. Appl.*, **17**:309–338, (1967).
- [197] M. Kochmański, T. Paszkiewicz, and S. Wolski. Curie–Weiss magnet – a simple model of phase transition. *European Journal of Physics*, **34**(6):1555, (2013).
- [198] G. Kohring. Ising models of social impact: the role of cumulative advantage. *Journal de Physique I*, **6**(2):301–308, (1996).
- [199] I. Kolossváry and J. Komjáthy. First passage percolation on inhomogeneous random graphs. *Adv. in Appl. Probab.*, **47**(2):589–610, (2015).
- [200] J. Komjáthy. Explosive Crump-Mode-Jagers branching processes. Available at arXiv:1602.01657 [math.PR], Preprint (2016).
- [201] J. Komjáthy and V. Vadon. First passage percolation on the Newman-Watts small world model. *J. Stat. Phys.*, **162**(4):959–993, (2016).
- [202] D. Krioukov, M. Kitsak, R. Sinkovits, D. Rideout, D. Meyer, and M. Boguñá. Network cosmology. *Scientific reports*, **2**, (2012).
- [203] M. Leone, A. Vázquez, A. Vespignani, and R. Zecchina. Ferromagnetic ordering in graphs with arbitrary degree distribution. *The European Physical Journal B-Condensed Matter and Complex Systems*, **28**(2):191–197, (2002).
- [204] J. Leskovec, K. Lang, A. Dasgupta, and M. Mahoney. Community structure in large networks: Natural cluster sizes and the absence of large well-defined clusters. *Internet Mathematics*, **6**(1):29–123, 2009.
- [205] T. M. Liggett. Stochastic models of interacting systems. *Ann. Probab.*, **25**(1):1–29, (1997).
- [206] T. M. Liggett. Interacting particle systems—an introduction. In *School and Conference on Probability Theory*, ICTP Lect. Notes, XVII, pages 1–56 (electronic). Abdus Salam Int. Cent. Theoret. Phys., Trieste, (2004).
- [207] R. Lipster and A. Shiriyayev. *Theory of Martingales*. Springer, Dordrecht, (1989).
- [208] T. Łuczak. Component behavior near the critical point of the random graph process. *Random Structures Algorithms*, **1**(3):287–310, (1990).
- [209] T. Łuczak. On the number of sparse connected graphs. *Random Structures Algorithms*, **1**(2):171–173, (1990).
- [210] T. Łuczak, B. Pittel, and J. Wierman. The structure of a random graph at the point of the phase transition. *Trans. Amer. Math. Soc.*, **341**(2):721–748, (1994).
- [211] R. Lyons. The Ising model and percolation on trees and tree-like graphs. *Comm. Math. Phys.*, **125**(2):337–353, (1989).
- [212] R. Lyons. Random walks and percolation on trees. *Ann. Probab.*, **18**(3):931–958, (1990).
- [213] R. Lyons, R. Pemantle, and Y. Peres. Conceptual proofs of $L \log L$ criteria for mean behavior of branching processes. *Ann. Probab.*, **23**(3):1125–1138, (1995).
- [214] M. Mézard and A. Montanari. Reconstruction on trees and spin glass transition. *J. Stat. Phys.*, **124**(6):1317–1350, (2006).
- [215] M. Molloy and B. Reed. A critical point for random graphs with a given degree sequence. *Random Structures Algorithms*, **6**(2-3):161–179, (1995).
- [216] M. Molloy and B. Reed. The size of the giant component of a random graph with a given degree sequence. *Combin. Probab. Comput.*, **7**(3):295–305, (1998).
- [217] A. Montanari, E. Mossel, and A. Sly. The weak limit of Ising models on locally tree-like graphs. *Probab. Theory Related Fields*, **152**(1-2):31–51, (2012).

- [218] J. W. Moon. The distance between nodes in recursive trees. In *Combinatorics (Proc. British Combinatorial Conf., Univ. Coll. Wales, Aberystwyth, 1973)*, pages 125–132. London Math. Soc. Lecture Note Ser., No. 13. Cambridge Univ. Press, London, (1974).
- [219] A. Nachmias and Y. Peres. Critical random graphs: diameter and mixing time. *Ann. Probab.*, **36**(4):1267–1286, (2008).
- [220] A. Nachmias and Y. Peres. Critical percolation on random regular graphs. *Random Structures Algorithms*, **36**(2):111–148, (2010).
- [221] A. Nachmias and Y. Peres. The critical random graph, with martingales. *Israel J. Math.*, **176**:29–41, (2010).
- [222] M. E. J. Newman. *Networks: an introduction*. Oxford University Press, (2010).
- [223] M. E. J. Newman, D. J. Watts, and A.-L. Barabási. *The Structure and Dynamics of Networks*. Princeton Studies in Complexity. Princeton University Press, (2006).
- [224] M. Niss. History of the Lenz-Ising model 1920–1950: from ferromagnetic to cooperative phenomena. *Arch. Hist. Exact Sci.*, **59**(3):267–318, (2005).
- [225] M. Niss. History of the Lenz-Ising model 1950–1965: from irrelevance to relevance. *Arch. Hist. Exact Sci.*, **63**(3):243–287, (2009).
- [226] I. Norros and H. Reittu. On a conditionally Poissonian graph process. *Adv. in Appl. Probab.*, **38**(1):59–75, (2006).
- [227] R. Pastor-Satorras, C. Castellano, P. Van Mieghem, and A. Vespignani. Epidemic processes in complex networks. *Rev. Modern Phys.*, **87**(3):925–979, (2015).
- [228] B. Pittel. On tree census and the giant component in sparse random graphs. *Random Structures Algorithms*, **1**(3):311–342, (1990).
- [229] B. Pittel. Note on the heights of random recursive trees and random m -ary search trees. *Random Structures Algorithms*, **5**(2):337–347, (1994).
- [230] B. Pittel. On the largest component of the random graph at a nearcritical stage. *J. Combin. Theory Ser. B*, **82**(2):237–269, (2001).
- [231] A. Rényi. On connected graphs. I. *Magyar Tud. Akad. Mat. Kutató Int. Közl.*, **4**:385–388, (1959).
- [232] S. Resnick. *Extreme values, regular variation and point processes*. Springer Series in Operations Research and Financial Engineering. Springer, New York, (2008). Reprint of the 1987 original.
- [233] O. Riordan. The small giant component in scale-free random graphs. *Combin. Probab. Comput.*, **14**(5-6):897–938, (2005).
- [234] O. Riordan. The phase transition in the configuration model. *Combinatorics, Probability and Computing*, **21**:1–35, (2012).
- [235] L. Rogers and D. Williams. *Diffusions, Markov processes, and martingales. Vol. 1*. Cambridge Mathematical Library. Cambridge University Press, Cambridge, (2000). Foundations, Reprint of the second (1994) edition.
- [236] J. Salez. The interpolation method for random graphs with prescribed degrees. *Combin. Probab. Comput.*, **25**(3):436–447, (2016).
- [237] E. Seneta. The simple branching process with infinite mean. I. *J. Appl. Probability*, **10**:206–212, (1973).
- [238] E. Seneta. Regularly varying functions in the theory of simple branching processes. *Advances in Appl. Probability*, **6**:408–420, (1974).
- [239] Y. Shavitt and E. Shir. Dimes - let the internet measure itself. *ACM SIGCOMM Computer Communication Review*, **35**(5):74, (2005).

- [240] R. Smythe and H. Mahmoud. A survey of recursive trees. *Theory Probab. Math. Statist.*, (51):1–27, (1996).
- [241] R. T. Smythe and J. C. Wierman. *First-passage percolation on the square lattice*, volume **671** of *Lecture Notes in Mathematics*. Springer, Berlin, (1978).
- [242] R. Solomonoff and A. Rapoport. Connectivity of random nets. *Bull. Math. Biophys.*, **13**:107–117, (1951).
- [243] J. Spencer. Enumerating graphs and Brownian motion. *Comm. Pure Appl. Math.*, **50**(3):291–294, (1997).
- [244] F. Spitzer. *Principles of Random Walk*. Springer, New York, 2nd edition, (1976).
- [245] S. Sreenivasan, T. Kalisky, L. Braunstein, S. Buldyrev, S. Havlin, and H. Stanley. Effect of disorder strength on optimal paths in complex networks. *Physical Review E*, **70**(4):46133, (2004).
- [246] C. Stegehuis, R. v. d. Hofstad, and J. v. Leeuwaarden. Epidemic spreading on complex networks with community structures. *Scientific Reports*, **6**:29748, (2016).
- [247] C. Stegehuis, R. v. d. Hofstad, and J. v. Leeuwaarden. Power-law relations in random networks with communities. *Physical Review E*, **94**:012302, (2016).
- [248] D. J. Watts. *Small worlds. The dynamics of networks between order and randomness*. Princeton Studies in Complexity. Princeton University Press, Princeton, NJ, (1999).
- [249] D. J. Watts. *Six degrees. The science of a connected age*. W. W. Norton & Co. Inc., New York, (2003).
- [250] W. Whitt. Some useful functions for functional limit theorems. *Math. Oper. Res.*, **5**(1):67–85, (1980).
- [251] W. Whitt. *Stochastic-process limits*. Springer Series in Operations Research. Springer-Verlag, New York, (2002). An introduction to stochastic-process limits and their application to queues.
- [252] W. Whitt. Proofs of the martingale FCLT. *Probab. Surv.*, **4**:268–302, (2007).
- [253] W. Willinger, D. Alderson, and J. Doyle. Mathematics and the internet: A source of enormous confusion and great potential. *Notices of the American Mathematical Society*, **56**(5):586–599, (2009).
- [254] E. M. Wright. The number of connected sparsely edged graphs. *J. Graph Theory*, **1**(4):317–330, (1977).
- [255] E. M. Wright. The number of connected sparsely edged graphs. III. Asymptotic results. *J. Graph Theory*, **4**(4):393–407, (1980).

Course Outline

Saint Flour Summer School in Probability

The summer school course will follow the following schedule:

Lecture 1: Random graphs as models for real-world networks: In this lecture, we present an overview to real-world networks and random graph models for them, as described in Chapter 1. We discuss real-world networks in Section 1.1 and random graph models for them in Section 1.2. We then continue to define some of the most highly studied random graph models in Section 1.3, the Erdős-Rényi random graph, inhomogeneous random graphs, the configuration model and preferential attachment models. We focus on their degree structure here, that determines the inhomogeneity in these models. After this, we formulate the main aim of this lecture series, stochastic processes on random graphs as models for network functionality, in Section 1.5.

Recommended exercises: Exercises 1.1, 1.2, 1.7, 1.8, 1.10, 1.11.

Lecture 2: Topology of random graphs: In this lecture based on Chapter 2, we discuss properties of random graphs that we will rely on in what follows. We show that these random graph models all are *locally tree-like*, which is an extremely useful property. Rank-1 inhomogeneous random graphs and configuration models even locally weakly converge to *homogeneous* branching processes. We continue to discuss connectivity of random graphs as well as their graph distances. We state when there is a giant component, as well as the connectivity transition.

Recommended exercises: Exercises 2.1, 2.2, 2.7, 2.12, 2.34.

Lecture 3: Rumor spread on random graphs with finite-variance degrees:

We treat first-passage percolation, which is a highly simplified model for the rumor spread on random graphs, as discussed in Chapter 3. We discuss the motivation for first-passage percolation in Section 3.1. In this lecture, we then focus on the finite-variance setting of the configuration model, in which the variance of the degree distribution remains uniformly bounded. We briefly describe continuous-time branching processes (continuous-time branching process) in Section 3.4. continuous-time branching processes are the key tool in this chapter. We then come to the key result in this chapter, which is Theorem 3.21 in Section 3.5. This result shows that first-passage percolation on configuration models with finite-variance degrees has a large amount of universality in terms of weighted graph distances between uniform vertices as well as the number of edges in the smallest-weight path.

Recommended exercises: Read through Section 3.2 on first-passage percolation on the complete graph as inspiration, and make some of the exercises. Exercises 3.18, 3.19, 3.20.

Lecture 4: Rumor spread on scale-free random graphs: Here we take some time to complete the discussion of Theorem 3.21. After this, we continue with random graph settings where the degree distribution has infinite variance. Theorem 3.28 shows that the continuous-time branching process approximation of

first-passage percolation on the graph can then explode in finite time. Theorem 3.30 gives a precise condition for explosion to occur. In the explosive setting, we show that the graph distance between two uniformly chosen vertices remains tight, and converges in distribution to the sum of two i.i.d. copies of the explosion time of the corresponding continuous-time branching process in Theorem 3.31, as discussed in Section 3.6. In the conservative setting, we show that the weight-distance between two uniform vertices tends to infinity in probability instead. We study one conservative example, where all the edge weights exceed 1, in more detail in Theorem 3.33.

Recommended exercises: Exercises 3.24, 3.28, 3.29, 3.31, 3.37.

Lecture 5: Percolation on random graphs: Critical value: We discuss the motivation of percolation on finite graphs in Section 4.1. We then continue with percolation on the configuration model in Section 4.3, using Janson's construction. We identify the asymptotic critical percolation threshold in Theorem 4.5. We make a brief start with Section 4.4.

Recommended exercises: Exercises 4.1. Read through Section 4.2 and make some of the exercises. Exercises 4.11.

Lecture 6: Percolation on random graphs: Scaling limits: We discuss scaling limits of the cluster sizes in critical percolation on configuration models. We start by discussing the case of finite third-moment degrees in Theorem 4.9, where the scaling limit (apart from rescaling factors) is the same as for the Erdős-Rényi random graph. We give a very brief sketch of the proof by discussing the exploration process of clusters as well as its scaling limit, which in fact is close to proving the scaling limit result of the cluster sizes. We continue in Section 4.5 by showing that for heavy-tailed degrees with finite second moment, but infinite third moments, the scaling limit is quite different (see Theorem 4.33). Again we sketch the proof, by discussing the exploration process of clusters as well as its scaling limit in Theorem 4.35.

Recommended exercises: Exercises 4.30, 4.31, 4.32, 4.33.

Lecture 7: Ising models on random graphs: We discuss the motivation of Ising models, as well as its definition on general (finite) graphs in Section 5.1. We then focus on the existence of thermodynamic limits for the Ising model on the configuration model in Section 5.3, when the graph size tends to infinity. We derive the thermodynamic limit of the pressure in Theorem 5.9, as well as those of the magnetization, internal energy and susceptibility in Theorem 5.10. We present part of the proof of Theorem 5.9, focussing on the uniqueness of the solution on the branching process tree in Proposition 5.13, and a local weak convergence argument to lift this to the random graph. The thermodynamic limit has a phase transition, and we describe the critical point in Theorem 5.21 in Section 5.4. Further, we discuss the critical exponents corresponding with it in Theorem 5.24 in Section 5.5 without going into proofs.

Recommended exercises: Read through the Ising model on the complete graph or Curie-Weiss model in Section 5.2, and try out some of the exercises there. Exercises 5.19, 5.21, 5.23, 5.28.

Lecture 8: Competition on random graphs: This lecture focuses on competition on random graphs as discussed in Section 6.1. We focus on Sections 6.1.2 and 6.1.3.

Recommended exercises: Exercises 6.6, 6.7, 6.8.