Random Graphs and Complex Networks. Vol. II

Remco van der Hofstad

Department of Mathematics and Computer Science
Eindhoven University of Technology
P.O. Box 513
5600 MB Eindhoven, The Netherlands
rhofstad@win.tue.nl

July 11, 2014

Preface

These lecture notes are intended to be used for master courses, where the students have a limited prior knowledge of special topics in probability. Therefore, we have included many of the preliminaries, such as convergence of random variables, probabilistic bounds, coupling, martingales and branching processes. These notes are aimed to be self-contained, and to give the readers an insight in the history of the field of random graphs.

The field of random graphs was started in 1959-1960 by Erdős and Rényi, see [50, 51, 52, 53]. At first, the study of random graphs was used to prove deterministic properties of graphs. For example, if we can show that a random graph has with a positive probability a certain property, then a graph must exist with this property. The method of proving deterministic statements using probabilistic arguments is called the *probabilistic method*, and goes back a long way. See among others the preface of a standard work in random graphs by Bollobás [22], or the work devoted to it *The Probabilistic Method* [6]. Erdős was one of the first to use this method, see e.g., [49], where it was shown that the Ramsey number $R(k)$ is at least $2^{k/2}$. The Ramsey number $R(k)$ is the value n for which any graph of size at least n is such that either itself or its complement contains a complete graph of size at least k. Erdős in [49] shows that for $n \leq 2^{k/2}$, the fraction of graphs for which the graph or its complement contains a complete graph of size k is bounded by $1/2$, so that there must be graphs of size $n \leq 2^{k/2}$ for which the graph nor its complement contains a complete graph of size k.

The initial work by Erdős and Rényi on random graphs has incited a great amount of work on the field. See the standard references on the subject [22] and [79] for the state of the art. Particularly [51] is a highly impressive paper. In it, a rather complete picture is given of the various phase transitions that occur on the Erdős-Rényi random graph. An interesting quote appearing in [51, Page 2-3] is the following:

> "*It seems to us worthwhile to consider besides graphs also more complex structures from the same point of view, i.e. to investigate the laws governing their evolution in a similar spirit. This may be interesting not only from a purely mathematical point of view. In fact, the evolution of graphs can be seen as a rather simplified model of the evolution of certain communication nets...*"

This was an excellent prediction indeed! Later, interest in random graphs of a different nature arose. Due to the increase of computer power, it has become possible to study so-called *real networks*. Many of these networks turned out to share similar properties, such as the fact that they are *small worlds*, and are *scale-free*, which means that they have degrees obeying *power laws*. The Erdős-Rényi random graph does not obey these properties, and, therefore, new graph models were invented. In fact, already in [51], Erdős and Rényi remark that

> *"Of course, if one aims at describing such a real situation, one should replace the hypothesis of equiprobability of all connection by some more realistic hypothesis."*

See [87] and [4] for two reviews of real networks and their properties to see what 'more realistic' could mean. These other models are also partly covered in the classical works [22] and [79], but up to today, there is no text treating random graphs and random graph models for complex networks in a relatively elementary way. See [46] for the most recent book on random graph, and, particularly, dynamical processes living on them. Durrett covers part of the material in this book, and much more, but the intended audience is rather different. The goal of these notes is to provide a source for a 'Random graphs' course at master level.

We treat both results for the Erdős-Rényi random graph, as well as for random graph models for complex networks. The aim is to give the simplest possible proofs for classical results, such as the phase transition for the largest connected component in the Erdős-Rényi random graph. Some proofs are more technical and difficult, and the sections containing these proofs will be indicated with a star *. These sections can be omitted without losing the logic in the results. We also give many exercises that help the reader to obtain a deeper understanding of the material by working at their solutions.

These notes would not have been possible without the help of many people. I thank Gerard Hooghiemstra for the encouragement to write these notes, and for using them at Delft University of Technology almost simultaneously while I used these notes at Eindhoven University of Technology in the spring of 2006 and again in the fall of 2008. Together with Piet Van Mieghem, we entered the world of random graphs in 2001, and I have tremendously enjoyed exploring this field together with you, as well as with Henri van den Esker, Dmitri Znamenski, Mia Deijfen and Shankar Bhamidi. I particularly wish to thank Gerard for many useful comments on these notes, solutions to exercises and suggestions for improvements of the presentation of particularly Chapters 2–5.

I thank Christian Borgs, Jennifer Chayes, Gordon Slade and Joel Spencer for joint work on random graphs which are like the Erdős-Rényi random graph, but do have geometry. This work has deepened my understanding of the basic properties of random graphs, and many of the proofs presented here have been inspired by our work in [26, 27, 28]. Special thanks go to Gordon Slade, who has introduced me to the world of percolation, which is a close neighbor of random graphs (see also [59]). It is peculiar to see that two communities work on two so related topics with quite different methods and even terminology, while it has taken such a long time to build bridges between the subjects.

Further I wish to thank Shankar Bhamidi, Finbar Bogerd, Kota Chisaki, Mia Deijfen, Michel Dekking, Henri van den Esker, Jesse Goodman, Markus Heydenreich, Yusuke Ide, Martin van Jole, Willemien Kets, Juliá Komjáthy, John Lapeyre, Norio Konno, Abbas Mehrabian, Mislav Mišković, Mirko Moscatelli, Jan Nagel, Sidharthan Nair, Alex Olssen, Mariana Olvera-Cravioto, Helena Peña, Nathan Ross, Karoly Simon, Xiaotin Yu for remarks, corrections of typos and ideas that have improved the

content and presentation of these notes substantially. Wouter Kager has entirely read the February 2007 version of the notes, giving many ideas for improvements of the arguments and of the methodology. Artëm Sapozhnikov, Maren Eckhoff and Gerard Hooghiemstra read and commented the October 2011 version.

I especially wish to thank Dennis Timmers, Eefje van den Dungen and Joop van de Pol, who, as, my student assistants, have been a great help in the development of these notes, in making figures, providing solutions to the exercises, typing up parts of proof from the literature, checking the proofs and keeping the references up to date. Maren Eckhoff and Gerard Hooghiemstra also provided many solutions to the exercises, for which I am grateful!

I have tried to give as many references to the literature as possible. However, the number of papers on random graphs is currently exploding. In MathSciNet, see

`http://www.ams.org/mathscinet/`,

there are, on December 21, 2006, a total of 1,428 papers that contain the phrase 'random graphs' in the review text, on September 29, 2008, this number increased to 1614, and to 2346 on April 9, 2013. These are merely the papers on the topic in the math community. What is special about random graph theory is that it is extremely multidisciplinary, and many papers are currently written in economics, biology, theoretical physics and computer science, using random graph models. For example, in Scopus (see `http://www.scopus.com/scopus/home.url`), again on December 21, 2006, there are 5,403 papers that contain the phrase 'random graph' in the title, abstract or keywords, on September 29, 2008, this has increased to 7,928 and on April 9, 2013 to 13,987. It can be expected that these numbers will increase even faster in the coming period, rendering it impossible to review most of the literature.

Contents

Chapter 1
Introduction to real-world networks and random graphs revisited

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

In Chapters 2–6, we shall return to the models discussed in [I, Chapters 6–8], and focus on their critical behavior as well as on the distances in these random graph models. Interestingly, a large chunk of the non-rigorous physics literature suggests that the behavior in various *different* random graph models can be described by only a *few* essential parameters. The key parameter of each of these models in the power-law degree exponent, and the physics literature predicts that the behavior in random graph models with similar degree sequences is similar. This is an example of the notion of *universality*, a notion which is central in statistical physics. Despite its importance, there are only few example of universality that can be rigorously proved. In Chapters 2–6, we investigate the level of universality present in random graph models.

We will often refer to Volume I. When we do, we write [I, Theorem 2.15] to mean that we refer to Theorem 2.15 in [65].

Chapter 2
General inhomogeneous random graphs

Abstract

In this chapter, we introduce the general setting of inhomogeneous random graphs. The inhomogeneous random graph is a generalization of the Erdős-Rényi random graph $\mathrm{ER}_n(p)$ as well as the inhomogeneous random graphs studied in [I, Chapter 6]. We start by motivating its choice, which is inspired by [I, Example 6.1]. The main results in this section concern the degree structure of such inhomogeneous random graphs, multitype branching process approximations to neighborhoods and the phase transition in these random graphs.

2.1 Motivation

Example 2.1 (Population of two types (Cont.))**.** Suppose that we have a complex network in which there are n_1 vertices of type 1 and n_2 of type 2. Type 1 individuals have on average m_1 neighbors, type 2 individuals m_2, where $m_1 \neq m_2$. Further, suppose that the probability that a type 1 individual is a friend of a type 2 individual is quite different from the probability that a type 1 individual is a friend of a type 1 individual.

In the model proposed in [I, Example 6.3], the probability that a type i individual is a friend of a type j individual (where $i, j, \in \{1, 2\}$) is equal to $m_i m_j/(\ell_n + m_i m_j)$, where $\ell_n = n_1 m_1 + n_2 m_2$. Approximating this probability by $m_i m_j/\ell_n$, we see that the probability that a type 1 individual is friend of a type 2 individual is quite related to the probability that a type 1 individual is friend of a type 1 individual. Indeed, take two type 1 and two type 2 individuals. Then, the probability that the type 1 individuals are friends and the type 2 individuals are friends is almost the same as the probability that first type 1 individual is friend with the first type 2 individual, and the second type 1 individual is friend of the second type 2 individual. This relation seems quite artificial, and is in many practical situations unwanted. The problem originates in the *product structure* of the edge probabilities in the generalized random graph.

Organization of this chapter In Section 3.6, we state some recent related results, and in Section 2.7, we close this chapter with notes and discussion.

2.2 Definition of the model

We assume that our individuals have types which are in a certain type space $\mathcal{S}$. When there are individuals of just 2 types, as in Example 2.1, then it suffices to take $\mathcal{S} = \{1, 2\}$. However, the model allows for rather general sets of types of the individuals, both finite as well as (countably or uncountably) infinite. An example of an uncountably infinite type space could be types related to the *ages* of the individuals in the population. We therefore also need to know how many individuals there are of a given type. This is described in terms of a *measure* μ_n, where, for $A \subseteq \mathcal{S}$, $\mu_n(A)$ denotes the proportion of individuals having a type in A.

In our general model, instead of vertex weights, the edge probabilities are moderated by a *kernel* $\kappa\colon \mathcal{S}^2 \to [0, \infty)$. The probability that two individuals of types x_1 and x_2 are friends is approximately $\kappa(x_1, x_2)/n$. Since there are many choices for κ, we arrive at a rather flexible model, where individuals have types and connection probabilities are related to the types of the individuals involved.

We start by making the above definitions formal, by defining what our ground space is and what a kernel is:

Definition 2.2 (Kernel). *(i) A ground space is a pair $(\mathcal{S}, \mu)$, where $\mathcal{S}$ is a separable metric space and μ is a Borel probability measure on $\mathcal{S}$.*

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

$$\mu_n(A) = \#\{i\colon x_i \in A\}/n \to \mu(A) \tag{2.2.1}$$

for every μ-continuity set $A \subseteq \mathcal{S}$. The convergence in (2.2.1) is denoted by $\mu_n \xrightarrow{p} \mu$.

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

Before defining the precise random graph model, we state the necessary conditions on our kernels. We write $E(G)$ for the number of edges in a graph G. Note that

$$\mathbb{E}[E(\mathrm{IRG}_n(\boldsymbol{p}))] = \sum_{i<j} p_{ij}, \tag{2.2.2}$$

so that our model has bounded degree in expectation precisely when $\frac{1}{n}\sum_{i<j} p_{ij}$ remains bounded. In our applications, we wish that the average degree per vertex in fact *converges*. This explains the main conditions we pose on the kernel κ

Definition 2.3 (Graphical and irreducible kernels). *(i) A kernel κ is* graphical *if the following conditions hold:*

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

(b)

$$\iint_{\mathcal{S}^2} \kappa(x,y)\mu(dx)\mu(dy) < \infty; \tag{2.2.3}$$

(c)

$$\frac{1}{n}\mathbb{E}[E(\mathrm{IRG}_n(\boldsymbol{p}(\kappa)))] \to \frac{1}{2}\iint_{\mathcal{S}^2} \kappa(x,y)\mu(dx)\mu(dy). \tag{2.2.4}$$

Similarly, a sequence (κ_n) *of kernels is called* graphical with limit κ *when*

$$y_n \to y \quad \text{and} \quad z_n \to z \quad \text{imply that} \quad \kappa_n(y_n,z_n) \to \kappa(y,z), \tag{2.2.5}$$

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

$$\frac{1}{n}\mathbb{E}[E(\mathrm{IRG}_n(\boldsymbol{p}(\kappa_n)))] \to \frac{1}{2}\iint_{\mathcal{S}^2} \kappa(x,y)\mu(dx)\mu(dy). \tag{2.2.6}$$

(ii) A kernel κ *is called* reducible *if*

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

otherwise κ *is irreducible.*

We now discuss the above definitions. The assumptions in (2.2.3), (2.2.4), (2.2.6) imply that the expected number of edges is proportional to n, and that the proportionality constant is precisely equal to $\iint_{\mathcal{S}^2} \kappa(x,y)\mu(dx)\mu(dy)$. Thus, in the terminology of Chapter 1, $\mathrm{IRG}_n(\boldsymbol{p}(\kappa))$ is *sparse*.

Roughly speaking, κ is reducible if the vertex set of $\mathrm{IRG}_n(\boldsymbol{p}(\kappa))$ can be split in two parts so that the probability of an edge from one part to the other is zero, and irreducible otherwise. When κ is reducible, then the random graph splits into two independent random graphs on the two disjoint subsets A and $\mathcal{S} \setminus A$. Therefore, we could have equally started with each of them separately, explaining why the notion of irreducibility is quite natural.

In many cases, we shall take $\mathcal{S} = [0,1]$, $x_i = i/n$ and μ the Lebesgue-measure on $[0,1]$. Then, clearly, (2.2.1) is satisfied. In fact, in [76], it is shown that we can always restrict to $\mathcal{S} = [0,1]$ by suitably adapting the other choices of our model. However, for notational purposes, it is more convenient to work with general $\mathcal{S}$. For example, where $\mathcal{S} = \{1\}$ is just a single type, the model reduces to the Erdős-Rényi random graph, and in the setting where $\mathcal{S} = [0,1]$, this is slightly more cumbersome:

Exercise 2.1 (Erdős-Rényi random graph)**.** *Show that when* $\mathcal{S} = [0,1]$ *and* $p_{ij} = \kappa(i/n, j/n)/n$ *with* $\kappa\colon [0,1]^2 \to [0,\infty)$ *being continuous, then the model is the Erdős-Rényi random graph with edge probability* λ/n *precisely when* $\kappa(x,y) = \lambda$*. Is this also true when* $\kappa\colon [0,1]^2 \to [0,\infty)$ *is not continuous?*

Now we come to the definition of our random graph. Given a kernel κ, for $n \in \mathbb{N}$, we let $\mathrm{IRG}_n(\boldsymbol{p}(\kappa))$ be the random graph on $[n]$, each possible edge ij, $i,j \in [n]$, is present with probability

$$p_{ij}(\kappa) = p_{ij} = \frac{1}{n}[\kappa(x_i,x_j) \wedge n], \tag{2.2.7}$$

and the events that different edges are present are independent. Similarly, $\mathrm{IRG}_n(\boldsymbol{p}(\kappa_n))$ is defined with κ_n replacing κ in (2.2.7).

For $\mathrm{CL}_n(\boldsymbol{w})$ with $\boldsymbol{w} = (w_i)_{i\in[n]}$ as in (6.1.14), we take $\mathcal{S} = [0,1], x_i = i/n$ and

$$\kappa_n(x,y) = [1-F]^{-1}(x)[1-F]^{-1}(y)n/\ell_n. \tag{2.2.8}$$

For $\mathrm{CL}_n(\boldsymbol{w})$ with $\boldsymbol{w} = (w_i)_{i\in[n]}$ satisfying Condition 6.4, instead, we take $\mathcal{S} = [0,1]$, $x_i = i/n$ and

$$\kappa_n(i/n, j/n) = w_i w_j/\mathbb{E}[W_n]. \tag{2.2.9}$$

We next study the Chung-Lu random graph:

Exercise 2.2 (The Chung-Lu model)**.** *Prove that when κ is given by*

$$\kappa(x,y) = [1-F]^{-1}(x)[1-F]^{-1}(y)/\mathbb{E}[W], \tag{2.2.10}$$

then κ is graphical precisely when $\mathbb{E}[W] < \infty$, where W has distribution function F. Further, κ is always irreducible.

Exercise 2.3 (The Chung-Lu model repeated)**.** *Let $\tilde{w}_i = [1-F]^{-1}(i/n)\sqrt{n\mathbb{E}[W]/\ell_n}$ and $w_i = [1-F]^{-1}(i/n)$ as in [I, (6.1.14)]. Then $\mathrm{CL}_n(\tilde{\boldsymbol{w}})$ and $\mathrm{CL}_n(\boldsymbol{w})$ are asymptotically equivalent whenever $(\frac{\mathbb{E}[W_n]}{\ell_n} - 1)^2 = o(n)$.*

In [24], also the choices

$$p_{ij}^{(\mathrm{NR})}(\kappa_n) = 1 - \mathrm{e}^{-\kappa_n(x_i,x_j)/n}, \tag{2.2.11}$$

or

$$p_{ij}^{(\mathrm{GRG})}(\kappa_n) = p_{ij} = \frac{\kappa(x_i,x_j)}{n+\kappa(x_i,x_j)} \tag{2.2.12}$$

are investigated. All results in [24] remain valid for the choices in (2.2.11) and (2.2.12). When

$$\sum_{i,j\in[n]} \kappa_n(x_i,x_j)^3 = o(n^{3/2}), \tag{2.2.13}$$

this follows immediately from [I, Theorem 6.18]:

Exercise 2.4 (Asymptotic equivalence for general IRGs)**.** *Prove that the random graphs $\mathrm{IRG}_n(\boldsymbol{p})$ with p_{ij} as in (2.2.7) is asymptotically equivalent to $\mathrm{IRG}_n(\boldsymbol{p})$ with $p_{ij} = p_{ij}^{(\mathrm{NR})}(\kappa_n)$ and to $\mathrm{IRG}_n(\boldsymbol{p})$ with $p_{ij} = p_{ij}^{(\mathrm{GRG})}(\kappa_n)$ when (2.2.13) holds.*

In the next section, we discuss some examples of inhomogeneous random graphs.

2.2.1 Examples of inhomogeneous random graphs

The Erdős-Rényi random graph. If $\mathcal{S}$ is general and $\kappa(x,y) = \lambda$ for every $x, y \in \mathcal{S}$, then the edge probabilities p_{ij} given by (2.2.7) are all equal to λ/n (for $n > \lambda$). Then $\mathrm{IRG}_n(\boldsymbol{p}(\kappa)) = \mathrm{ER}_n(\lambda/n)$. The simplest choice here is to take $\mathcal{S} = \{1\}$.

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

Exercise 2.5 (Definitions 2.2-2.3 for homogeneous bipartite graph)**.** *Prove that Definitions 2.2-2.3 hold for the homogeneous bipartite graph.*

The finite-type case. Fix $r \geq 2$ and suppose we have a graph with r different types of vertices. Let $\mathcal{S} = \{1, \dots, r\}$. Let n_i denote the number of vertices of type i, and let $\mu_n(i) = n_i/n$. Let $\mathrm{IRG}_n(\boldsymbol{p}(\kappa))$ be the random graph where two vertices of types i and j, respectively, joined by an edge with probability $n^{-1}\kappa(i, j)$ (for $n \geq \max \kappa$). Then κ is equivalent to an $r \times r$ matrix, and the random graph $\mathrm{IRG}_n(\boldsymbol{p}(\kappa))$ has vertices of r different types (or colors). The finite-types case has been studied by Söderberg [105, 106, 107, 108]. We conclude that our general IRG covers the cases of a finite (or even countably infinite) number of types.

Exercise 2.6 (Homogeneous bipartite graph)**.** *Prove that the homogeneous bipartite random graph is a special case of the finite-types case.*

Exercise 2.7 (Irreducibility for the finite-types case)**.** *Prove that, in the finite-type case, irreducibility follows when there exists an m such that the m^{th} power of the matrix $(\kappa(i, j))_{i,j\in[r]}$ contains no zeros.*

Exercise 2.8 (Graphical limit in the finite-types case)**.** *Prove that, in the finite-type case,* (2.2.1) *holds precisely when*

$$\lim_{n\to\infty} n_i/n = p_i. \tag{2.2.14}$$

2.2.2 Degree sequence of IRG

We next turn to the degrees of the vertices of $\mathrm{IRG}_n(\boldsymbol{p}(\kappa_n))$. As we shall see, the degree of a vertex of a given type x is asymptotically Poisson with a mean

$$\lambda(x) = \int_{\mathcal{S}} \kappa(x, y)\mu(dy) \tag{2.2.15}$$

that depends on x. This leads to a mixed Poisson distribution for the degree D of a (uniformly chosen) random vertex of $\mathrm{IRG}_n(\boldsymbol{p}(\kappa))$. We recall that N_k denotes the number of vertices of $\mathrm{IRG}_n(\boldsymbol{p}(\kappa))$ with degree k.

Theorem 2.4 (The degree sequence of $\mathrm{IRG}_n(\boldsymbol{p}(\kappa))$)**.** *Let (κ_n) be a graphical sequence of kernels with limit κ. For any fixed $k \geq 0$,*

$$N_k/n \xrightarrow{\mathbb{P}} \int_{\mathcal{S}} \frac{\lambda(x)^k}{k!} \mathrm{e}^{-\lambda(x)} \mu(dx) \tag{2.2.16}$$

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

$$\lambda(x) = \int_{\mathcal{S}} \kappa(x,y)\mu(dy). \tag{2.2.17}$$

Equivalently,

$$N_k/n \xrightarrow{\mathbb{P}} \mathbb{P}(\Xi = k), \tag{2.2.18}$$

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

$$\mathbb{P}(W_\lambda \le x) = \int_0^x \lambda(y)\mu(dy). \tag{2.2.19}$$

In the remainder of this section, we shall give a proof of Theorem 2.4. We start by proving Theorem 2.4 for the finite-types case, which is substantially easier. After this, we give a proof in the general case, for which we shall need to prove results on approximations of sequences of graphical kernels.

Proof of Theorem 2.4 in the finite-types case. Consider first the finite-type case. Take a vertex v of type i, let D_v be its degree, and let $D_{v,j}$ be the number of edges from v to vertices of type $j \in [r]$. Then, clearly, $D_v = \sum_j D_{v,j}$.

Recall that, in the finite-types case, the edge probability between vertices of types i and j is denoted by $(\kappa(i,j) \wedge n)/n$. Assume that $n \ge \max \kappa$. The random variables $(D_{v,j})_{j\in[r]}$ are independent, and $D_{v,j} \sim \mathsf{Bin}(n_j - \delta_{ij}, \kappa(i,j)/n) \xrightarrow{d} \mathsf{Poi}(\mu_j\kappa(i,j))$, where n_j are the number of vertices with type j and $\mu_j = \lim_{n\to\infty} n_j/n$. Hence,

$$D_v \xrightarrow{d} \mathsf{Poi}\left(\sum_j \mu_j\kappa(i,j)\right) = \mathsf{Poi}(\lambda(i)), \tag{2.2.20}$$

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

$$\mathbb{P}(D_v = k) \to \mathbb{P}(\mathsf{Poi}(\lambda(i)) = k) = \frac{\lambda(i)^k}{k!}\mathrm{e}^{-\lambda(i)}. \tag{2.2.21}$$

Let $N_{k,i}$ be the number of vertices in $G^{\mathcal{V}}(n,k)$ of type i with degree k. Then, for fixed $n_1, \dots, n_r$,

$$\frac{1}{n}\mathbb{E}[N_{k,i}] = \frac{1}{n}n_i\mathbb{P}(D_v = k) \to \mu_i\mathbb{P}(\mathsf{Poi}(\lambda(i)) = k). \tag{2.2.22}$$

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

$$\frac{1}{n}N_{k,i} \xrightarrow{\mathbb{P}} \mathbb{P}(\mathsf{Poi}(\lambda(i)) = k)\mu_i = \mathbb{P}(\Xi = k), \tag{2.2.23}$$

and thus, summing over i,

$$\frac{1}{n}N_k = \sum_i \frac{1}{n}N_{k,i} \xrightarrow{\mathbb{P}} \sum_i \mathbb{P}(\mathsf{Poi}(\lambda(i)) = k)\mu_i = \mathbb{P}(\Xi = k). \tag{2.2.24}$$

This proves Theorem 2.4 in the regular finitary case. □

In order to prove Theorem 2.4 in the general case, we shall be approximating a sequence of graphical kernels (κ_n) by appropriate regular finite kernels:

Approximations of sequences of graphical kernels. Recall that $\mathcal{S}$ is a separable metric space, and that μ is a Borel measure on $\mathcal{S}$ with $0 < \mu(\mathcal{S}) < \infty$. We usually assume that $\mu(\mathcal{S}) = 1$; in this section, this makes no difference. Here the metric and topological structure of $\mathcal{S}$ will be important.

Given a sequence of finite partitions $\mathcal{P}_m = \{A_{m1}, \dots, A_{mM_m}\}$, $m \ge 1$, of $\mathcal{S}$ and an $x \in \mathcal{S}$, we define $i_m(x)$ by requiring that

$$x \in A_{m,i_m(x)}. \tag{2.2.25}$$

As usual, for $A \subset \mathcal{S}$ we write $\operatorname{diam}(A)$ for $\sup\{|x-y| \colon x, y \in A\}$. By taking $\mathcal{P}_m$ as the dyadic partition into intervals of length 2^{-m} in $\mathcal{S}$, we easily see the following:

Lemma 2.5 (Approximating partition)**.** *There exists a sequence of finite partitions $\mathcal{P}_m = \{A_{m1}, \dots, A_{mM_m}\}$, $m \ge 1$, of $\mathcal{S}$ such that*

(i) each A_{mi} is measurable and $\mu(\partial A_{mi}) = 0$;

(ii) for each m, $\mathcal{P}_{m+1}$ refines $\mathcal{P}_m$, i.e., each A_{mi} is a union $\bigcup_{j \in J_{mi}} A_{m+1,j}$ for some set J_{mi};

(iii) for a.e. $x \in \mathcal{S}$, $\operatorname{diam}(A_{m,i_m(x)}) \to 0$ as $m \to \infty$, where $i_m(x)$ is defined by (2.2.25)

Recall that a kernel κ is a symmetric measurable function on $\mathcal{S} \times \mathcal{S}$. Fixing a sequence of partitions with the properties described in Lemma 2.5, we can define sequences of lower and upper approximations to κ by

$$\kappa_m^-(x,y) = \inf\{\kappa(x',y') \colon x' \in A_{m,i_m(x)}, y' \in A_{m,i_m(y)}\}, \tag{2.2.26}$$
$$\kappa_m^+(x,y) = \sup\{\kappa(x',y') \colon x' \in A_{m,i_m(x)}, y' \in A_{m,i_m(y)}\}. \tag{2.2.27}$$

We thus replace κ by its infimum or supremum on each $A_{mi} \times A_{mj}$. As κ_m^+ might be $+\infty$, we shall use it only for bounded κ.

By Lemma 2.5(ii),

$$\kappa_m^- \le \kappa_{m+1}^- \quad \text{and} \quad \kappa_m^+ \ge \kappa_{m+1}^+. \tag{2.2.28}$$

Furthermore, if κ is continuous a.e. then, by Lemma 2.5(iii),

$$\kappa_m^-(x,y) \to \kappa(x,y) \quad \text{and} \quad \kappa_m^+(x,y) \to \kappa(x,y) \quad \text{for a.e. } (x,y) \in \mathcal{S}^2. \tag{2.2.29}$$

Since $k_m^- \le \kappa$, we can obviously construct our random graph so that $\mathrm{IRG}_n(\kappa_m^-) \subseteq \mathrm{IRG}_n(\kappa)$, and in the sequel we shall assume this. Similarly, we shall assume that $\mathrm{IRG}_n(\kappa_m^+) \supseteq \mathrm{IRG}_n(\kappa)$ when κ is bounded.

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

$$\kappa_m^-(x,y) := \inf\{(\kappa \wedge \kappa_n)(x',y') \colon x' \in A_{m,i_m(x)}, y' \in A_{m,i_m(y)}, n \ge m\}, \tag{2.2.30}$$
$$\kappa_m^+(x,y) := \sup\{(\kappa \vee \kappa_n)(x',y') \colon x' \in A_{m,i_m(x)}, y' \in A_{m,i_m(y)}, n \ge m\}. \tag{2.2.31}$$

By Lemma 2.5, we have $\kappa_m^- \le \kappa_{m+1}^-$, and from Lemma 2.5(iii) and 2.3(ii) we see that

$$\kappa_m^-(x,y) \nearrow \kappa(x,y) \quad \text{as } m \to \infty, \text{ for a.e. } (x,y) \in \mathcal{S}^2. \tag{2.2.32}$$

Moreover, when $n \geq m$, we have

$$\kappa_n \geq \kappa_m^-, \tag{2.2.33}$$

and we may assume that $\mathrm{IRG}_n(\kappa_m^-) \subseteq \mathrm{IRG}_n(\kappa_n)$. By the convergence of the sequence of kernels (κ_n), we further obtain that also the number of edges converges:

Now we are ready to complete the proof of Theorem 2.4 for general sequences of graphical kernels (κ_n). Define κ_m^- by (2.2.30). Let $\varepsilon > 0$ be given. From (2.2.6) and monotone convergence, there is an m such that

$$\iint_{\mathcal{S}^2} \kappa_m^-(x,y)\mu(dx)\mu(dy) > \iint_{\mathcal{S}^2} \kappa(x,y)\mu(dx)\mu(dy) - \varepsilon. \tag{2.2.34}$$

For $n \geq m$, we have $\kappa_m^- \leq \kappa_n$ by (2.2.33), so we may assume that $\mathrm{IRG}_n(\boldsymbol{p}(\kappa_m^-)) \subseteq \mathrm{IRG}_n(\boldsymbol{p}(\kappa_n))$. Then, by (2.2.6) and (2.2.34),

$$\begin{aligned} &\frac{1}{n} E(\mathrm{IRG}_n(\boldsymbol{p}(\kappa_n)) \setminus \mathrm{IRG}_n(\boldsymbol{p}(\kappa_m^-))) \\ &= \frac{1}{n} E(G(n,\kappa_n)) - \frac{1}{n} E(\mathrm{IRG}_n(\boldsymbol{p}(\kappa_m^-))) \\ &\xrightarrow{\mathbb{P}} \frac{1}{2} \iint_{\mathcal{S}^2} \kappa(x,y)\mu(dx)\mu(dy) - \frac{1}{2} \iint_{\mathcal{S}^2} \kappa_m^-(x,y)\mu(dx)\mu(dy) < \frac{\varepsilon}{2}, \end{aligned} \tag{2.2.35}$$

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

$$|N_k^{(m)} - N_k| < 2\varepsilon n \tag{2.2.36}$$

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

$$|N_k^{(m)}/n - \mathbb{P}(\Xi^{(m)} = k)| < \varepsilon. \tag{2.2.37}$$

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

$$\begin{aligned} \mathbb{P}(\Xi \neq \Xi^{(m)}) &= \mathbb{P}(\Xi - \Xi^{(m)} \geq 1) \\ &\leq \mathbb{E}[\Xi - \Xi^{(m)}] = \iint_{\mathcal{S}^2} \kappa(x,y)\mu(dx)\mu(dy) - \iint_{\mathcal{S}^2} \kappa_m^-(x,y)\mu(dx)\mu(dy) < \varepsilon. \end{aligned} \tag{2.2.38}$$

Combining (2.2.36), (2.2.37) and (2.2.38), we see that $|N_k/n - \mathbb{P}(\Xi = k))| < 4\varepsilon$ **whp**. □

Let Λ be the random variable $\lambda(U)$, where U is a random variable on $\mathcal{S}$ having distribution μ. Then we can also describe the mixed Poisson distribution of Ξ as $\mathsf{Poi}(\Lambda)$. Under mild conditions, the tail probabilities $\mathbb{P}(\Xi > t)$ and $\mathbb{P}(\Lambda > t)$ are similar for large t. We state this for the case of power-law tails; the result generalizes to regularly varying tails. As above, let D be the degree of a random vertex in $\mathrm{IRG}_n(\boldsymbol{p}(\kappa_n))$. Let $N_{\geq k}$ be the number of vertices with degree at least k.

Corollary 2.6 (Power-law tails for the degree sequence)**.** *Let (κ_n) be a graphical sequence of kernels with limit κ. Suppose that $\mathbb{P}(\Lambda > t) = \mu\{x\colon \lambda(x) > t\} \sim at^{-(\tau-1)}$ as $t \to \infty$, for some $a > 0$ and $\tau > 2$. Then*

$$N_{\geq k}/n \xrightarrow{\mathbb{P}} \mathbb{P}(\Xi \geq k) \sim ak^{-(\tau-1)}, \tag{2.2.39}$$

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

Proof. It suffices to show that $\mathbb{P}(\Xi \geq k) \sim ak^{-(\tau-1)}$; the remaining conclusions then follow from Theorem 2.4. For any $\varepsilon > 0$, $\mathbb{P}(\mathsf{Poi}(\Lambda) > t|\Lambda > (1+\varepsilon)t) \to 1$ and $\mathbb{P}(\mathsf{Poi}(\Lambda) > t|\Lambda < (1-\varepsilon)t) = o(t^{-(\tau-1)})$ as $t \to \infty$. It follows that $\mathbb{P}(\Xi > t) = \mathbb{P}(\mathsf{Poi}(\Lambda) > t) \sim at^{-(\tau-1)}$ as $t \to \infty$. □

This result shows that the general IRG does include natural cases with power-law degree distributions. Naturally, we have already observed in [I, Theorem 6.7] that this is the case for the $\mathrm{GRG}_n(\boldsymbol{w})$ when the weights sequence $\boldsymbol{w}$ is chosen appropriately.

2.2.3 Multitype branching processes

In order to study further properties of $\mathrm{IRG}_n(\boldsymbol{p}(\kappa))$, we need to understand the neighborhood structure of vertices. For simplicity, let us restrict ourselves to the finite types case. Then, for a vertex of type i, the number of neighbors of type j is close to Poisson distributed with approximate mean $\kappa(i,j)$. Even when we assume independence of the neighborhood structures of different vertices, we still do not arrive at a classical branching process as discussed in [I, Chapter 3]. Instead, we can describe the neighborhood structure with a branching process in which we keep track of the type of each of the vertices. For general κ and μ, we can even have a continuum of types. Such branching processes are called *multitype branching processes*. See e.g. [12, Chapter V] or [62, Chapter III] for more background on multitype branching processes. In this section, we shall only discuss the basics and we shall quickly go to the special case of multitype branching processes where every offspring has a Poisson distribution.

Multitype branching processes with finitely many types. Multitype branching process can be analyzed using linear algebra. In order to do so, we first introduce some notation. We first assume that we are in the finite types case, and denote the number of types by r. We let $\boldsymbol{j} = (j_1, \ldots, j_r) \in \mathbb{N}_0^r$ be a vector of non-negative integers, and denote by $p_{\boldsymbol{j}}^{(i)}$ the probability that an individual of type i gives rise to an offspring $\boldsymbol{j}$, i.e., j_1 children of type 1, j_2 children of type 2, etc. The offsprings of the different individuals are all mutually independent. Denote by $Z_{n,j}^{(i)}$ the number of individuals of type j in generation n when starting from a single particle of type i and $\mathbf{Z}_n^{(i)} = (Z_{n,1}^{(i)}, \ldots, Z_{n,j}^{(i)})$. We shall be interested in the survival or extinction of multitype branching processes, and in the growth of the generation sizes. In the multitype case, we are naturally lead to a matrix setup. We now discuss the survival versus

extinction of multitype branching processes. We denote the survival probability of the multitype branching process when starting from a single individual of type i by

$$q^{(i)} = \mathbb{P}(\mathbf{Z}_n^{(i)} \neq \mathbf{0} \text{ for all } n), \tag{2.2.40}$$

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

Multitype branching processes and generating functions. We write $\boldsymbol{p}(\boldsymbol{j}) = (p_{\boldsymbol{j}}^{(1)}, \ldots, p_{\boldsymbol{j}}^{(r)})$ and we let

$$G^{(i)}(\boldsymbol{s}) = \sum_{\boldsymbol{j}} p_{\boldsymbol{j}}^{(i)} \prod_{a=1}^{r} s_a^{j_a} \tag{2.2.41}$$

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

$$G_n^{(i)}(\boldsymbol{s}) = \mathbb{E}[\prod_{a=1}^{r} s_a^{Z_{n,a}^{(i)}}], \tag{2.2.42}$$

and $\mathbf{G}_n(\boldsymbol{s}) = (G_n^{(1)}(\boldsymbol{s}), \ldots, G_n^{(r)}(\boldsymbol{s}))$. Then, we have that $\mathbf{G}_{n+1}(\boldsymbol{s}) = \mathbf{G}_n(\mathbf{G}(\boldsymbol{s})) = \mathbf{G}(\mathbf{G}_n(\boldsymbol{s}))$ and $\boldsymbol{q} = \mathbf{1} - \lim_{n\to\infty} \mathbf{G}_n(\mathbf{0})$. Naturally, the extinction probability depends sensitively on the type of the ancestor of the branching process. On the other hand, under reasonable assumptions, the positivity of the survival probability is independent of the initial type. A necessary and sufficient condition for this property is that, with positive probability, a particle of type i arises as a descendent of of a particle of type j for each type i and j.

We note that when $\mathbf{G}(\boldsymbol{s}) = \mathbf{M}\boldsymbol{s}$ for some matrix $\mathbf{M}$, then each individual in the Markov chain has precisely one offspring, and we call this case *singular*. When each particle has precisely one offspring, the multitype branching process is equivalent to a Markov chain, and the process a.s. survives. Thus, in this case, there is no survival vs. extinction phase transition. We shall assume throughout the remainder that the multitype branching process is non-singular.

Survival vs. extinction and mean offspring. We continue to describe the survival versus extinction of multitype branching processes in terms of the mean offspring. Let κ_{ij} denote the expected offspring of type j of a single individual of type i, and let $\mathbf{T}_\kappa = \{\kappa_{ij}\}_{i,j=1}^r$ be the matrix of offsprings. We shall assume that there exists an l such that the matrix $\mathbf{M}^l$ has only strictly positive entries. This is sometimes called *irreducibility*, as it implies that the Markov chain of the number of individuals of the various types is an irreducible Markov chain. By the Perron-Frobenius theorem, the matrix $\mathbf{M}$ has a unique largest eigenvalue $\|\mathbf{T}_\kappa\|$ with non-negative left-eigenvector $\mathbf{x}_\kappa$, and the eigenvalue $\|\mathbf{T}_\kappa\|$ can be computed as

$$\|\mathbf{T}_\kappa\| = \sup_{\mathbf{x}:\, \|\mathbf{x}\|_2 \leq 1} \|\mathbf{T}_\kappa \mathbf{x}\|_2, \quad \text{where} \quad \|\mathbf{x}\|_2 = \sqrt{\sum_{i=1}^{r} x_i^2}. \tag{2.2.43}$$

We note that

$$\mathbb{E}[\mathbf{Z}_{n+1}^{(i)}|\mathbf{Z}_n^{(i)} = \mathbf{z}] = \mathbf{T}_\kappa \mathbf{z}, \tag{2.2.44}$$

so that

$$\mathbb{E}[\mathbf{Z}_{n+1}^{(i)}] = \mathbf{T}_\kappa^n \mathbf{e}^{(i)}. \tag{2.2.45}$$

where $\mathbf{T}_\kappa^n$ denotes the n-fold application of the matrix $\mathbf{T}_\kappa$, and $\mathbf{e}^{(i)}$ is the vector which has on the i^{th} position a 1, and further only zeroes. The identifications in (2.2.44) and (2.2.45) have several important consequences concerning the phase transition of multitype branching processes, as we shall now discuss in more detail.

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

$$\mathbb{E}[\mathbf{Z}_{n+1}^{(i)}] \leq \|\mathbf{T}_\kappa\|^n \|\mathbf{e}^{(i)}\|_2, \tag{2.2.46}$$

which converges to 0 exponentially fast. Therefore, by the Markov inequality (Theorem 2.15), the multitype branching process dies out a.s. When $\|\mathbf{T}_\kappa\| > 1$, on the other hand, the sequence

$$M_n = \mathbf{x}_\kappa \mathbf{Z}_{n+1}^{(i)} \|\mathbf{T}_\kappa\|^{-n} \tag{2.2.47}$$

is a non-negative martingale, by (2.2.44) and the fact that $\mathbf{x}_\kappa$ is a left-eigenvector with eigenvalue $\|\mathbf{T}_\kappa\|$, since $\mathbf{x}_\kappa \mathbf{T}_\kappa = \|\mathbf{T}_\kappa\| \mathbf{x}_\kappa$. By the Martingale convergence theorem (Theorem 2.22), the martingale M_n converges a.s. When we further assume some further restrictions on M_n, for example that M_n has finite second moment, then we obtain that $M_n \xrightarrow{a.s.} M_\infty$ and $\mathbb{E}[M_n] \to \mathbb{E}[M_\infty]$. More precisely, there is a multitype analog of the Kesten-Stigum Theorem (Theorem 3.10). Since $\mathbb{E}[M_n] = \mathbb{E}[M_0] = \mathbf{x}_\kappa \mathbf{e}^{(i)} > 0$, we thus have that $\mathbf{Z}_{n+1}^{(i)}$ grows exponentially with a strictly positive probability, which implies that the survival probability is positive. Theorem 3.1 can be adapted to show that $\mathbf{Z}_{n+1}^{(i)} \xrightarrow{\mathbb{P}} \mathbf{0}$ when $\|\mathbf{T}_\kappa\| = 1$. See e.g. [62, Sections II.6-II.7]. We conclude that, for non-singular and irreducible multitype branching processes, we have that $\boldsymbol{q} > \mathbf{0}$ precisely when $\|\mathbf{T}_\kappa\| > 1$.

Poisson multitype branching processes. We call a multitype branching processes *Poisson* when all the number of children of each type are independent Poisson random variables. Thus, $\mathbf{Z}^{(i)} = (Z_{1,1}^{(i)}, \ldots, Z_{1,r}^{(i)})$ is a vector of independent Poisson random variables with means $(\kappa_{1,i}, \ldots, \kappa_{r,i})$. As we shall see later, Poisson multitype branching processes arise naturally when exploring a component of $\mathrm{IRG}_n(\boldsymbol{p}(\kappa))$ starting at a vertex of type x. This is directly analogous to the use of the single-type Poisson branching process in the analysis of the Erdős-Rényi graph $\mathrm{ER}_n(\lambda/n)$ as discussed in detail in Chapters 4 and 5.

For Poisson multitype branching processes, we obtain that

$$G^{(i)}(\boldsymbol{s}) = \mathbb{E}[\prod_{a=1}^{r} s_a^{Z_{1,a}^{(i)}}] = \mathrm{e}^{\sum_{a=1}^r \kappa_{a,i}(s_a - 1)} = \mathrm{e}^{(\mathbf{T}_\kappa(\boldsymbol{s}-\mathbf{1}))_i}. \tag{2.2.48}$$

Thus, the vector of survival probabilities $\boldsymbol{q}$ satisfies

$$\boldsymbol{q} = \mathbf{1} - \mathrm{e}^{-\mathbf{T}_\kappa \boldsymbol{q}}. \tag{2.2.49}$$

This leads us to the investigation of eigenfunctions of non-linear operators of the form $f \mapsto 1 - \mathrm{e}^{-\mathbf{T}_\kappa f}$. We now extend the above setting of finite-type Poisson multitype branching processes to the infinite type case.

Poisson multitype branching processes with infinitely many types. Let κ be a kernel. We define the Poisson multitype branching processes with kernel κ as follows. Each individual of type $x \in \mathcal{S}$ is replaced in the next generation by a set of individuals distributed as a Poisson process on $\mathcal{S}$ with intensity $\kappa(x,y)\mu(dy)$. Thus, the number of children with types in a subset $A \subseteq \mathcal{S}$ has a Poisson distribution with mean $\int_A \kappa(x,y)\mu(dy)$, and these numbers are independent for disjoint sets A and for different particles; see e.g., Kallenberg [82].

Let $\zeta_\kappa(x)$ be the survival probability of the Poisson multitype branching process with kernel κ, starting from an ancestor of type $x \in \mathcal{S}$. Set

$$\zeta_\kappa = \int_{\mathcal{S}} \zeta_\kappa(x)\mu(dx). \tag{2.2.50}$$

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

$$(\mathbf{T}_\kappa f)(x) = \int_{\mathcal{S}} \kappa(x,y) f(y)\mu(dy), \tag{2.2.51}$$

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

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

As we shall see, the analysis of multitype branching processes with a possibly uncountable number of types is a bit more functional analytic. Similarly to the finite-type case in (2.2.43), we define

$$\|\mathbf{T}_\kappa\| = \sup\Bigl\{\|\mathbf{T}_\kappa f\|_2 \colon f \geq 0, \|f\|_2 \leq 1\Bigr\} \leq \infty. \tag{2.2.52}$$

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

$$\|\mathbf{T}_\kappa\| \leq \|\mathbf{T}_\kappa\|_{\mathrm{HS}} = \|\kappa\|_{L^2(\mathcal{S}\times\mathcal{S})} = \biggl(\iint_{\mathcal{S}^2} \kappa(x,y)^2\mu(dx)\mu(dy)\biggr)^{1/2}. \tag{2.2.53}$$

We also define the non-linear operator Φ_κ by

$$\bigl(\Phi_\kappa f\bigr)(x) = 1 - \mathrm{e}^{-(\mathbf{T}_\kappa f)(x)}, \qquad x \in \mathcal{S}, \tag{2.2.54}$$

for $f \geq 0$. Note that for such f we have $0 \leq \mathbf{T}_\kappa f \leq \infty$, and thus $0 \leq \Phi_\kappa f \leq 1$. We shall characterize the survival probability $\zeta_\kappa(x)$, and thus ζ_κ, in terms of the non-linear

operator Φ_κ, showing essentially that the function $x \mapsto \zeta_\kappa(x)$ is the maximal fixed point of the non-linear operator Φ_κ (recall (2.2.49)). Again, the survival probability satisfies that $\zeta_\kappa > 0$ precisely when $\|\mathbf{T}_\kappa\| > 1$, recall the finite-types case discussed in detail above.

We call a multitype branching process *supercritical* when $\|\mathbf{T}_\kappa\| > 1$, *critical* when $\|\mathbf{T}_\kappa\| < 1$, and *subcritical* when $\|\mathbf{T}_\kappa\| < 1$. Then, the above discussion can be summarized by saying that a multitype branching process survives with positive probability precisely when it is supercritical.

2.3 The phase transition for inhomogeneous random graphs

In this section, we discuss the phase transition in $\mathrm{IRG}_n(\boldsymbol{p}(\kappa))$. The main result shows that there is a giant component when the associated multitype branching process is supercritical, while otherwise there is not:

Theorem 2.7 (Giant component of IRG)**.** *Let (κ_n) be a sequence of irreducible graphical kernels with limit κ, and let $\mathcal{C}_{\max}$ denote the largest connected component of $\mathrm{IRG}_n(\boldsymbol{p}(\kappa_n))$. Then,*

$$|\mathcal{C}_{\max}|/n \xrightarrow{\mathbb{P}} \zeta_\kappa. \tag{2.3.1}$$

In all cases $\zeta_\kappa < 1$, while $\zeta_\kappa > 0$ precisely when $\|\mathbf{T}_\kappa\| > 1$.

Theorem 2.7 is a generalization of the law of large numbers for the largest connected component in Theorem 4.8 for $\mathrm{ER}_n(\lambda/n)$:

Exercise 2.9 (LLN for $\mathcal{C}_{\max}$ for $\mathrm{ER}_n(\lambda/n)$)**.** *Prove that, for the Erdős-Rényi random graph, Theorem 2.7 implies that $|\mathcal{C}_{\max}|/n \xrightarrow{\mathbb{P}} \zeta_\lambda$, where ζ_λ is the survival probability of a Poisson branching process with mean λ offspring.*

We close this section by discussing a few examples of Theorem 2.7.

The bipartite random graph. We let n be even and take $\mathcal{S} = \{1, 2\}$ and

$$\kappa_n(x, y) = \kappa(x, y) = \lambda \mathbb{1}_{\{x \neq y\}}/2. \tag{2.3.2}$$

Thus, for $i < j$, the edge probabilities p_{ij} given by (2.2.7) are equal to $\lambda/(2n)$ (for $2n > \lambda$) when $i \in [n/2]$ and $j \in [n] \setminus [n/2]$.

In this case, $\|\mathbf{T}_\kappa\| = \lambda$ with corresponding eigenfunction $f(x) = 1$ for all $x \in \mathcal{S}$. Thus, Theorem 2.7 proves that there is a phase transition at $\lambda = 2$. Furthermore, the function $\zeta_\lambda(x)$ reduces to the single value $\zeta_{\lambda/2}$, which is the survival probability of a Poisson branching process with mean offspring $\lambda/2$. This is not surprising, since the degree of each vertex is $\mathsf{Bin}(n/2, \lambda/n)$, so the bipartite random graph of size n is quite closely related the Erdős-Rényi random graph of size $n/2$.

The finite-type case. The bipartite random graph can also be viewed as a random graph with two types of vertices (i.e., the vertices $[n/2]$ and $[n] \setminus [n/2]$). We now generalize the results to the finite-type case, in which we have seen that κ_n is equivalent to an $r \times r$-matrix $(\kappa_n(i,j))_{i,j\in[r]}$, where r denotes the number of types. In this case, $\mathrm{IRG}_n(\boldsymbol{p}(\kappa))$ has vertices of r different types (or colors), say n_i vertices of type i, with two vertices of type i and j joined by an edge with probability $n^{-1}\kappa_n(i,j)$ (for $n \geq \max \kappa_n$). This case has been studied by Söderberg [105, 106, 108, 107], who noted Theorem 2.7 in this case.

Exercise 2.10 (Phase transition for $r = 2$)**.** *Compute ζ_κ in the case of two types, and give necessary and sufficient conditions for $\zeta_\kappa > 0$.*

Exercise 2.11 (The size of small components in the finite-types case)**.** *Prove that, in the finite-types case, when (κ_n) converges, then $\sup_{x,y,n} \kappa_n(x,y) < \infty$ holds, so that the results of Theorem 2.9 apply in the sub- and supercritical cases.*

The random graph with prescribed expected degrees. We next consider the Chung-Lu model or expected degree random graph, where κ_n is given by (2.2.9), i.e., $\kappa_n(i/n, j/n) = w_i w_j/\mathbb{E}[W_n]$.

We first assume that [I, Condition 6.4(a)-(c)] holds, so that in particular $\mathbb{E}[W^2] < \infty$, where W has distribution function F. A particular instance of this case is the choice $w_i = [1-F]^{-1}(i/n)$ in [I, (6.1.14)]. In this case, the sequence (κ_n) converges to κ, where the limit κ is given by (recall (2.2.10))

$$\kappa(x,y) = \psi(x)\psi(y)/\mathbb{E}[W], \tag{2.3.3}$$

where $\psi(x) = [1-F]^{-1}(x)$ Then, we note that for each $f \geq 0$ with $\|f\|_2 = 1$,

$$(\mathbf{T}_\kappa f)(x) = \psi(x)\frac{\int_{\mathcal{S}} \psi(x)f(x)\mu(dx)}{\int_{\mathcal{S}} \psi(x)\mu(dx)}, \tag{2.3.4}$$

so that $\|\mathbf{T}_\kappa f\|_2 = \int_{\mathcal{S}} \psi(x)f(x)\mu(dx)/\int_{\mathcal{S}} \psi(x)\mu(dx)$, which is maximal when $f(x) = \psi(x)/\|\psi\|_2$. We conclude that $\|\mathbf{T}_\kappa\| = \|\psi\|_2^2/\int_{\mathcal{S}} \psi(x)\mu(dx) = \mathbb{E}[W^2]/\mathbb{E}[W]$. Thus,

$$\|\mathbf{T}_\kappa\| = \mathbb{E}[W^2]/\mathbb{E}[W], \tag{2.3.5}$$

and we recover the results in [36, 38] in the case where $\mathbb{E}[W^2] < \infty$. In the case where $\mathbb{E}[W^2] = \infty$, on the other hand, we see that $\|\mathbf{T}_\kappa f\|_2 = \infty$ for every f with $\|f\|_2 = 1$ such that $\int_{\mathcal{S}} \psi(x)f(x)\mu(dx) \neq 0$, so that $\|\mathbf{T}_\kappa\| = \infty$, so that $\mathrm{CL}_n(\boldsymbol{w})$ is always supercritical in this regime.

Exercise 2.12 (The size of small components for $\mathrm{CL}_n(\boldsymbol{w})$)**.** *Prove that, for $\mathrm{CL}_n(\boldsymbol{w})$ with weights given by [I, (6.1.14)] and with $1 < \nu < \infty$, the second largest cluster has size $|\mathcal{C}_{(2)}| = O_{\mathbb{P}}(\log n)$ when W has bounded support or is a.s. bounded below by $\varepsilon > 0$, while if $\nu < 1$, $|\mathcal{C}_{\max}| = O(\log n)$ when W has bounded support. Here W is a random variable with distribution function F.*

2.4 Small-world effect in inhomogeneous random graphs

In this section, we consider the distances between vertices of $\mathrm{IRG}_n(\boldsymbol{p}(\kappa_n))$ where, as usual, (κ_n) is a graphical sequence of kernels with limit κ.

Let us write $\mathrm{dist}_G(i,j)$ for the graph distance between the vertices $i,j \in [n]$ in a graph G, where the graph distance is the minimum number of edges in the graph G that form a path from i to j, and, by convention, we let $\mathrm{dist}_G(i,j) = \infty$ when i,j are in different connected components.

We define the *typical graph distance* to be H_n, where $H_n = \mathrm{dist}_G(V_1, V_2)$ is the graph distance between two vertices V_1, V_2 which are chosen uniformly at random from $[n]$.

We start by discussing logarithmic asymptotics of the typical graph distance in the case where $\nu = \|\mathbf{T}_\kappa\| \in (1,\infty)$. When $\|\mathbf{T}_\kappa\| = \infty$, then our results also prove that $H_n = o_{\mathbb{P}}(\log n)$, but they do not tell us much about their exact asymptotics.

Logarithmic asymptotics of typical graph distance in $\mathrm{IRG}_n(\boldsymbol{p}(\kappa_n))$. The main result on typical graph distances in $\mathrm{IRG}_n(\boldsymbol{p}(\kappa_n))$ is as follows:

Theorem 2.8 (Typical distances in $\mathrm{IRG}_n(\boldsymbol{p}(\kappa_n))$)**.** *Let (κ_n) be graphical sequence of kernels with limit κ, and with $\nu = \|\mathbf{T}_\kappa\| \in (1,\infty)$. Let $\varepsilon > 0$ be fixed. Then, for $\mathrm{IRG}_n(\boldsymbol{p}(\kappa_n))$,*

(i) If $\sup_{x,y,n} \kappa_n(x,y) < \infty$, then

$$\mathbb{P}(H_n \le (1-\varepsilon)\log_\nu n) = o(1). \tag{2.4.1}$$

(ii) If κ is irreducible, then

$$\mathbb{P}(H_n \le (1+\varepsilon)\log_\nu n) = \zeta_\kappa^2 + o(1). \tag{2.4.2}$$

In the terminology of Section 1.2, Theorem 2.8(ii) implies that $\mathrm{IRG}_n(\boldsymbol{p}(\kappa))$ is a small world when κ is irreducible and $\nu = \|\mathbf{T}_\kappa\| < \infty$. Theorem 2.8(i) shows that the graph distances are of order $\Theta(\log n)$ when $\sup_{x,y,n} \kappa_n(x,y) < \infty$, so that $\mathrm{IRG}_n(\boldsymbol{p}(\kappa_n))$ is not an ultra-small world. The intuition behind Theorem 2.8 is that, by (2.2.45) and (2.2.47), a Poisson multitype branching process with kernel κ has neighborhoods that grow exponentially, i.e., the number of vertices at distance k grows like $\|\mathbf{T}_\kappa\|^k$. Thus, if we examine the distance between two vertices V_1 and V_2 chosen uniformly at random from $[n]$, then we need to explore the neighborhood of vertex V_1 up to the moment that it 'catches' vertex V_2. In this case, the neighborhood must be of size $\sim n$, so that we need that $\|\mathbf{T}_\kappa\|^k = \nu^k \sim n$, i.e., $k = k_n \sim \log_\nu n$. However, proving of such a fact is quite tricky, since there are far fewer possible further vertices to explore when the neighborhood has size $\sim n$. The proof overcomes this fact by exploring from the two vertices V_1 and V_2 simultaneously up to the first moment that these neighborhoods share a common vertex. At this moment, we have found the shortest path.

Theorem 2.8(i) is closely related to Theorem 3.5 in the next chapter, and we refer to the discussion there.

2.5 Related results for inhomogeneous random graphs

In this section, we discuss some related results for inhomogeneous random graphs. While we give intuition about their proofs, we shall not include them in full detail.

The largest subcritical cluster. For the classical random graph $\mathrm{ER}_n(\lambda/n)$, it is well-known that in the subcritical case for which $\lambda < 1$, the stronger bound $|\mathcal{C}_{\max}| = \Theta(\log n)$ holds (see [I, Theorems 4.4–4.5]), and that in the supercritical case for which $\lambda > 1$, $|\mathcal{C}_{(2)}| = \Theta(\log n)$. These bounds do not always hold in the general framework we are considering here, but if we add some conditions, then we can improve the estimates in Theorem 2.7 for the subcritical case to $O(\log n)$:

Theorem 2.9 (Subcritical phase and duality principle of the IRG)**.** *Consider the inhomogeneous random graph* $\mathrm{IRG}_n(\boldsymbol{p}(\kappa_n))$*, where* (κ_n) *is a graphical sequence of kernels with limit* κ*. Then,*

(*i*) *if* κ *is subcritical and* $\sup_{x,y,n} \kappa_n(x,y) < \infty$*, then* $|\mathcal{C}_{\max}| = O_{\mathbb{P}}(\log n)$.

(*ii*) *if* κ *is supercritical,* κ *is irreducible, and either* $\inf_{x,y,n} \kappa_n(x,y) > 0$ *or* $\sup_{x,y,n} \kappa_n(x,y) < \infty$*, then* $|\mathcal{C}_{(2)}| = O_{\mathbb{P}}(\log n)$.

When $\lim_{n\to\infty} \sup_{x,y} \kappa_n(x,y) = \infty$, the largest subcritical clusters can have rather different behavior, as we now show for the rank-1 case. Note that, by Theorem 2.7 as well as the fact that $\|\mathbf{T}_\kappa\| = \nu = \mathbb{E}[W^2]/\mathbb{E}[W]$, a rank-1 model can only be subcritical when $\mathbb{E}[W^2] < \infty$, i.e., in the case of finite variance degrees. However, when W has a power-law tail, i.e., when $\mathbb{P}(W \geq w) \sim w^{-(\tau-1)}$, then the highest weight can be much larger than $\log n$. When this is the case, then also the largest subcritical cluster is much larger than $\log n$, as proved in the following theorem:

Theorem 2.10 (Subcritical phase for rank-1 IRG)**.** *Let* $\boldsymbol{w}$ *satisfy [I, Condition 6.4(a)-(c)] with* $\nu = \mathbb{E}[W^2]/\mathbb{E}[W] < 1$*, and, further, that there exist* $\tau > 3$ *and* $c_2 > 0$ *such that*

$$[1 - F_n](x) \leq c_2 x^{-(\tau-1)}. \tag{2.5.1}$$

Then, for $\mathrm{NR}_n(\boldsymbol{w})$ *with* $\Delta = \max_{j\in[n]} w_j$*,*

$$|\mathcal{C}_{\max}| = \frac{\Delta}{1-\nu} + o_{\mathbb{P}}(n^{1/(\tau-1)}). \tag{2.5.2}$$

Theorem 3.25 is most interesting in the case where the limiting distribution function F in Condition 6.4 has a power-law tail. For example, for $\boldsymbol{w}$ as in [I, (6.1.14)], let F satisfy

$$[1 - F](x) = cx^{-(\tau-1)}(1 + o(1)). \tag{2.5.3}$$

Then, $\Delta = w_1 = [1-F]^{-1}(1/n) = (cn)^{1/(\tau-1)}(1 + o(1))$. Therefore,

$$|\mathcal{C}_{\max}| = (cn)^{1/(\tau-1)}/(1-\nu) + o(n^{1/(\tau-1)}. \tag{2.5.4}$$

Thus, the largest connected component is much larger than for $\mathrm{ER}_n(\lambda/n)$ with $\lambda < 1$.

Theorem 3.25 can be intuitively understood as follows. The connected component of a typical vertex is close to a branching process, so that it is with high probability bounded since the expected value of its cluster will be close to $1/(1-\nu)$. Thus, the best way to obtain a large connected component is to start with a vertex with high weight w_i, and let all of its roughly w_i children be independent branching processes. Therefore, in expectation, each of these children is connected to another $1/(1-\nu)$ different vertices, leading to a cluster size of roughly $w_i/(1-\nu)$. This is clearly largest when $w_i = \max_{j\in[n]} w_j = \Delta$, leading to an intuitive explanation of Theorem 3.25.

Theorems 2.9 and 3.25 raise the question what the precise conditions for $|\mathcal{C}_{\max}|$ to be of order $\log n$ are. Intuitively, when $\Delta \gg \log n$, then $|\mathcal{C}_{\max}| = \Delta/(1-\nu)(1+o_{\mathbb{P}}(1))$, whereas if $\Delta = \Theta(\log n)$, then $|\mathcal{C}_{\max}| = \Theta_{\mathbb{P}}(\log n)$ as well. In [110], it was proved that $|\mathcal{C}_{\max}|/\log n$ converges in probability to a finite constant when $\nu < 1$ and the weights are i.i.d. with distribution function F with $\mathbb{E}[\mathrm{e}^{\alpha W}] < \infty$ for some $\alpha > 0$, i.e., *exponential tails* are sufficient.

The diameter in inhomogeneous random graphs. We shall also be interested in the *diameter* of $\mathrm{IRG}_n(\boldsymbol{p}(\kappa_n))$, which is defined to be the maximal finite graph distance between any pair of vertices, i.e., the diameter $\mathrm{diam}(G)$ of the graph G equals

$$\mathrm{diam}(G) = \max_{u,v:\, \mathrm{dist}_G(u,v)<\infty} \mathrm{dist}_G(u,v). \tag{2.5.5}$$

We shall see that for $\mathrm{IRG}_n(\boldsymbol{p}(\kappa))$, the diameter tends to be much larger than the typical graph distances, which is due to long thin lines which are distributed as subcritical $\mathrm{IRG}_n(\boldsymbol{p}(\kappa))$ with a subcritical κ by a duality principle for $\mathrm{IRG}_n(\boldsymbol{p}(\kappa))$. Before we state the results, we introduce the notion of the *dual kernel*:

Definition 2.11 (Dual kernel for $\mathrm{IRG}_n(\boldsymbol{p}(\kappa))$)**.** *Let (κ_n) be a sequence of supercritical kernels with limit κ. The dual kernel is the kernel $\widehat{\kappa}$ defined by $\widehat{\kappa}(x,y) = \kappa(x,y)$, with* reference measure $d\widehat{\mu}(x) = (1-\zeta_\kappa(x))\mu(dx)$.

The dual kernel shall describe the graph after the removal of the giant component. Here, the reference measure $\widehat{\mu}$ measures the size of the graph. In this case, a vertex x is in the giant component with probability $1-\zeta_\kappa(x)$, in which case it must be removed. Thus, $\widehat{\mu}$ describes the proportion of vertices of the various types which are outside the giant component. As before, we define the operator $\mathbf{T}_{\widehat{\kappa}}$ by the equality

$$(\mathbf{T}_{\widehat{\kappa}} f)(x) = \int_{\mathcal{S}} \widehat{\kappa}(x,y) f(y) d\widehat{\mu}(y) = \int_{\mathcal{S}} \kappa(x,y) f(y)[1-\zeta_\kappa(x)]\mu(dy), \tag{2.5.6}$$

and we write $\|\mathbf{T}_{\widehat{\kappa}}\|$ for

$$\|\mathbf{T}_{\widehat{\kappa}}\| = \sup\big\{\|\mathbf{T}_{\widehat{\kappa}} f\|_2 \colon f \ge 0, \|f\|_{\widehat{\mu},2} = 1\big\}, \tag{2.5.7}$$

where

$$\|f\|^2_{\widehat{\mu},2} = \int_{\mathcal{S}} f^2(x)\widehat{\mu}(dx). \tag{2.5.8}$$

Theorem 2.12 (The diameter of $\mathrm{IRG}_n(\boldsymbol{p}(\kappa))$ in the finite-types case)**.** *Let (κ_n) be a sequence of kernels with limit κ, which has finitely many types. If $0 < \|\mathbf{T}_\kappa\| < 1$, then*

$$\frac{\mathrm{diam}(\mathrm{IRG}_n(\boldsymbol{p}(\kappa_n)))}{\log n} \xrightarrow{\mathbb{P}} \frac{1}{\log \|\mathbf{T}_\kappa\|^{-1}} \tag{2.5.9}$$

as $n \to \infty$. If $\|\mathbf{T}_\kappa\| > 1$ and κ irreducible, then

$$\frac{\mathrm{diam}(\mathrm{IRG}_n(\boldsymbol{p}(\kappa_n)))}{\log n} \xrightarrow{\mathbb{P}} \frac{2}{\log \|\mathbf{T}_{\widehat{\kappa}}\|^{-1}} + \frac{1}{\log \|\mathbf{T}_\kappa\|}, \tag{2.5.10}$$

where $\widehat{\kappa}$ is the dual kernel to κ.

If we compare Theorem 2.12 to Theorem 3.28, we see that the diameter has the same scaling as the typical graph distance, but that the limit in probability of $\mathrm{diam}(\mathrm{IRG}_n(\boldsymbol{p}(\kappa)))/\log n$ is strictly larger than the one for $H_n/\log n$ conditioned on $H_n < \infty$. This effect is particularly noticeable in the case when $\tau \in (2,3)$, where $H_n/\log\log n$, conditionally on $H_n < \infty$, converges in probability to a finite limit, while $\mathrm{diam}(\mathrm{IRG}_n(\boldsymbol{p}(\kappa)))/\log n$ converges to a non-zero limit. This can be explained by noticing that the diameter in $\mathrm{IRG}_n(\boldsymbol{p}(\kappa))$ arises due to very this lines of length of order $\log n$. Since these this lines involve only very few vertices, they will not contribute to H_n, but they do to $\mathrm{diam}(\mathrm{IRG}_n(\boldsymbol{p}(\kappa)))$. This is another argument why we prefer to work with typical graph distances than with the diameter.

2.6 Related models

Clustered inhomogeneous random graphs. General inhomogeneous random graphs have rather low clustering. Indeed, assuming that $\kappa_n(x,y \le n$, we can compute that the expected number of triangles in an $\mathrm{IRG}_n(\boldsymbol{p}(\kappa))$ is equal to

$$\mathbb{E}[\#\text{ triangles in } \mathrm{IRG}_n(\boldsymbol{p}(\kappa))] = \frac{1}{n^3}\sum_{i,j,k\in[n]} \kappa_n(x_i,x_j)\kappa_n(x_j,x_k)\kappa_n(x_k,x_i). \tag{2.6.1}$$

Under relatively weak conditions on the kernel κ_n, it follows that

$$\mathbb{E}[\#\text{ triangles in } \mathrm{IRG}_n(\boldsymbol{p}(\kappa))] \to \int_{\mathcal{S}^3} \kappa(x_1,x_2)\kappa(x_2,x_3)\kappa(x_3,x_1)\mu(dx_1)\mu(dx_2)\mu(dx_3). \tag{2.6.2}$$

Therefore, the clustering coefficient converges to zero as $1/n$. In many real-world networks, particularly in social networks, the clustering coefficient is strictly positive. In this section, we discuss a model similar to the inhomogeneous random graph $\mathrm{IRG}_n(\boldsymbol{p}(\kappa))$ that incorporates clustering.

The idea behind this model is that instead of only adding *edges* independently, we can also add other graphs on r vertices in an independent way. For example, we could study a graph where each pair of vertices is independently connected with probability λ/n, as for $\mathrm{ER}_n(\lambda/n)$, but also each collection of triples forms a triangle with probability μ/n^2, independently for all triplets and independently of the status

of the edges. Here the exponent n^{-2} is chosen to as to make the expected number of triangles containing a vertex bounded.

Exercise 2.13 (Clustering in model with edges and triangles)**.** *Show that the clustering coefficient in the model where each pair of vertices is independently connected with probability λ/n, as for $\mathrm{ER}_n(\lambda/n)$ and each triples forms a triangle with probability μ/n^2, independently for all triplets and independently of the status of the edges, converges to $\mu/(\mu+\lambda^2)$.*

In social networks, also complete graphs of size four, five, etc., are present more often than in usual random graph. Therefore, we also wish to add those independently. In order to formulate this general version of the model, we start by introducing some notation.

Let $\mathcal{F}$ consist of one representative of each isomorphism class of finite connected graphs, chosen so that if $F \in \mathcal{F}$ has r vertices then $V(F) = [r] = \{1, 2, \dots, r\}$. Simple examples of such F are the complete graphs on r vertices, but also other examples are possible. Recall that $\mathcal{S}$ denotes the type space. Given $F \in \mathcal{F}$ with r vertices, let κ_F be a measurable function from $\mathcal{S}^r$ to $[0, \infty)$. The function κ_F is called the *kernel corresponding to* F. A sequence $\widetilde{\kappa} = (\kappa_F)_{F\in\mathcal{F}}$ is a *kernel family*.

Let $\widetilde{\kappa}$ be a kernel family and n an integer. We define a random graph $\mathrm{IRG}_n(\widetilde{\kappa})$ with vertex set $[n] = \{1, \dots, n\}$. First let $x_1, x_2, \dots, x_n \in \mathcal{S}$ be i.i.d. (independent and identically distributed) with the distribution μ. Given $\mathbf{x} = (x_1, \dots, x_n)$, construct $\mathrm{IRG}_n(\widetilde{\kappa})$ as follows, starting with the empty graph. For each r and each $F \in \mathcal{F}$ with $|F| = r$, and for every r-tuple of distinct vertices $(v_1, \dots, v_r) \in [n]^r$, add a copy of F on the vertices $v_1, \dots, v_r$ (with vertex i of F mapped to v_i) with probability

$$p = p(v_1, \dots, v_r; F) = \frac{\kappa_F(x_{v_1}, \dots, x_{v_r})}{n^{r-1}}, \tag{2.6.3}$$

all these choices being independent. Here, if $p > 1$, by convention we simply add a copy with probability 1. We shall often call the added copies of the various F that together form $\mathrm{IRG}_n(\widetilde{\kappa})$ *atoms* as they may be viewed as indivisible building blocks. Sometimes we refer to them as *small graphs*, although there is in general no bound on their sizes.

The reason for dividing by n^{r-1} in (2.6.3) is that we wish to consider sparse graphs; indeed, our main interest is the case when $\mathrm{IRG}_n(\widetilde{\kappa})$ has $O(n)$ edges. As it turns out, we can be slightly more general; however, when κ_F is integrable (which we shall always assume), the expected number of added copies of each graph F is $O(n)$. Note that all incompletely specified integrals are with respect to the appropriate r-fold product measure μ^r on $\mathcal{S}^r$.

There are several plausible choices for the normalization in (2.6.3). The one we have chosen means that if $\kappa_F = c$ is constant, then (asymptotically) there are on average cn copies of F in total, and each vertex is on average in rc copies of F. An alternative is to divide the expression in (2.6.3) by r; then (asymptotically) each vertex would on average be in c copies of F. Another alternative, natural when adding cliques only but less so in the general case, would be to divide by $r!$; this

is equivalent to considering unordered sets of r vertices instead of ordered r-tuples. When there is only one kernel, corresponding to adding edges, this would correspond to the normalization used in [24], and in particular to that of the classical model $\mathrm{ER}_n(\lambda/n)$; the normalization we use here differs from this by a factor of 2.

In the special case where all κ_F are zero apart from κ_{K_2}, the kernel corresponding to an edge, we recover (essentially) a special case of the model of [24]; we call this the *edge-only* case, since we add only edges, not larger graphs. We write κ_2 for κ_{K_2}. Note that in the edge-only case, given $\mathbf{x}$, two vertices i and j are joined with probability

$$\frac{\kappa_2(x_i,x_j)+\kappa_2(x_j,x_j)}{n}+O\left(\frac{(\kappa_2(x_i,x_j)+\kappa_2(x_j,x_i))^2}{n^2}\right). \tag{2.6.4}$$

The correction term will never matter, so we may as well replace κ_2 by its symmetrized version. In fact, we shall always assume that κ_F is invariant under permutations of the vertices of the graph F.

For any kernel family $\widetilde{\kappa}$, let κ_e be the corresponding *edge kernel*, defined by

$$\kappa_e(x,y)=\sum_F \sum_{ij\in E(F)} \int_{\mathcal{S}^{V(F)\setminus\{i,j\}}} \kappa_F(x_1,\dots,x_{i-1},x,x_{i+1},\dots,x_{j-1},y,x_{j+1},\dots,x_{|F|}), \tag{2.6.5}$$

where the second sum runs over all $2E(F)$ ordered pairs (i,j) with $ij \in \mathit{Edges}(F)$, and we integrate over all variables apart from x and y. Note that the sum need not always converge; since every term is positive this causes no problems: we simply allow $\kappa_e(x,y)=\infty$ for some x,y. Given x_i and x_j, the probability that i and j are joined in $G(n,\widetilde{\kappa})$ is at most $\kappa_e(x_i,x_j)/n+O(1/n^2)$. In other words, κ_e captures the edge probabilities in $G(n,\widetilde{\kappa})$, but not the correlations.

Before proceeding to deeper properties, let us note that the expected number of added copies of F is $(1+O(n^{-1}))n\int_{\mathcal{S}^{|F|}}\kappa_F$. Unsurprisingly, the actual number turns out to be concentrated about this mean. Let

$$\xi(\widetilde{\kappa})=\sum_{F\in\mathcal{F}} E(F)\int_{\mathcal{S}^{|F|}}\kappa_F=\frac{1}{2}\int_{\mathcal{S}^2}\kappa_e\le\infty \tag{2.6.6}$$

be the *asymptotic edge density* of $\widetilde{\kappa}$. Since every copy of F contributes $E(F)$ edges, the following theorem is almost obvious, provided we can ignore overlapping edges.

Theorem 2.13 (The edge density in $\mathrm{IRG}_n(\widetilde{\kappa})$)**.** *As $n\to\infty$, $\mathbb{E}[E(\mathrm{IRG}_n(\widetilde{\kappa}))/n]\to\xi(\widetilde{\kappa})\le\infty$ Moreover, $E(\mathrm{IRG}_n(\widetilde{\kappa}))/n$ converges in probability to the asymptotic edge density $\xi(\widetilde{\kappa})$. In other words, if $\xi(\widetilde{\kappa})<\infty$ then $E(\mathrm{IRG}_n(\widetilde{\kappa}))=\xi(\widetilde{\kappa})n+o_p(n)$, and if $\xi(\widetilde{\kappa})=\infty$ then $E(\mathrm{IRG}_n(\widetilde{\kappa}))>Cn$ for every constant C, whp.*

The main focus is the emergence of the giant component. By the *component structure* of a graph G, we mean the set of vertex sets of its components, i.e., the structure encoding only which vertices are in the same component, not the internal structure of the components themselves. When studying the component structure of $\mathrm{IRG}_n(\widetilde{\kappa})$, the model can be simplified somewhat. Recalling that the atoms $F\in\mathcal{F}$ are connected by definition, when we add an atom F to a graph G, the effect on the

component structure is simply to unite all components of G that meet the vertex set of F, so only the vertex set of F matters, not its graph structure. We say that $\widetilde{\kappa}$ is a *clique kernel family* if the only non-zero kernels are those corresponding to complete graphs; the corresponding random graph model $\mathrm{IRG}_n(\widetilde{\kappa})$ is a *clique model*. For questions corresponding component structure, it suffices to study clique models. For clique kernels we write κ_r for κ_{K_r}; as above, we always assume that κ_r is symmetric, here meaning invariant under all permutations of the coordinates of $\mathcal{S}^r$. Given a general kernel family $\widetilde{\kappa}$, the corresponding (symmetrized) clique kernel family is given by $\underline{\widetilde{\kappa}} = (\kappa_r)_{r\geq 2}$ with

$$\kappa_r(x_1,\dots,x_r) = \sum_{F\in\mathcal{F}:\,|F|=r} \frac{1}{r!} \sum_{\pi\in\mathfrak{S}_r} \kappa_F(x_{\pi(1)},\dots,x_{\pi(r)}), \tag{2.6.7}$$

where $\mathfrak{S}_r$ denotes the symmetric group of permutations of $[r]$. (This is consistent with our notation $\kappa_2 = \kappa_{K_2}$.) When considering the size (meaning number of vertices) of the giant component in $\mathrm{IRG}_n(\widetilde{\kappa})$, we may always replace $\underline{\widetilde{\kappa}}$ by the corresponding clique kernel family.

It is often convenient to think of a clique model as a random *hypergraph*, with the cliques as the hyperedges; for this reason we call a clique kernel family a *hyperkernel*. Note that each unordered set of r vertices corresponds to $r!$ r-tuples, so the probability that we add a K_r on a given set of r vertices is $r!\kappa_r(x_{v_1},\dots,x_{v_r})/n^{r-1}$. (More precisely, this is the expected number of K_rs added with this vertex set.)

In our analysis we also consider the linear operator T_{κ_e} defined by

$$T_{\kappa_e}(f)(x) = \int_{\mathcal{S}} \kappa_e(x,y) f(y) d\mu(y), \tag{2.6.8}$$

where κ_e is defined by (2.6.5). We need to impose some sort of integrability condition on our kernel family:

Definition 2.14. *(i) A kernel family $\widetilde{\kappa} = (\kappa_F)_{F\in\mathcal{F}}$ is integrable if*

$$\int \widetilde{\kappa} = \sum_{F\in\mathcal{F}} |F| \int_{\mathcal{S}^{|F|}} \kappa_F < \infty. \tag{2.6.9}$$

This means that the expected number of atoms containing a given vertex is bounded.

(ii) A kernel family $\widetilde{\kappa} = (\kappa_F)_{F\in\mathcal{F}}$ is edge integrable if

$$\sum_{F\in\mathcal{F}} E(F) \int_{\mathcal{S}^{|F|}} \kappa_F < \infty; \tag{2.6.10}$$

equivalently, $\xi(\kappa) < \infty$ or $\int_{\mathcal{S}^2} \kappa_e < \infty$. This means that the expected number of edges in $G(n,\widetilde{\kappa})$ is $O(n)$, see Theorem 2.13, and thus the expected degree of a given vertex is bounded.

Note that a hyperkernel (κ_r) is integrable if and only if $\sum_{r\geq 2} r \int_{\mathcal{S}^r} \kappa_r < \infty$, and edge integrable

The main results concerning the phase transition on $\mathrm{IRG}_n(\widetilde{\kappa})$ is that if $\widetilde{\kappa}$ is an integrable kernel family satisfying a certain extra assumption, then the normalized size of the giant component in $\mathrm{IRG}_n(\widetilde{\kappa})$ is simply $\zeta(\widetilde{\kappa}) + o_p(1)$. The extra assumption is an irreducibility assumption similar to Definition 2.3(ii) that essentially guarantees that the graph does not split into two pieces: we say that a symmetric kernel $\kappa_e : \mathcal{S}^2 \to [0,\infty)$ is *reducible* if

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

otherwise κ_e is irreducible. Thus, κ_e is irreducible if

$$A \subseteq \mathcal{S} \text{ and } \kappa_e = 0 \text{ a.e. on } A \times (\mathcal{S} \setminus A) \text{ implies } \mu(A) = 0 \text{ or } \mu(\mathcal{S} \setminus A) = 0.$$

We are now ready to formulate the main result in this section involving the phase transition in $\mathrm{IRG}_n(\widetilde{\kappa})$. Recall that $|\mathcal{C}_{\max}|$ denotes the number of vertices in the largest connected component of the graph under consideration, and $|\mathcal{C}_{(2)}|$ the number of vertices in its second largest component.

Theorem 2.15 (The phase transition on clustered inhomogeneous random graphs)**.** *Let $\widetilde{\kappa}' = (\kappa'_F)_{F\in\mathcal{F}}$ be an irreducible, integrable kernel family, and let $\widetilde{\kappa} = (\kappa_r)_{r\geq 2}$ be the corresponding hyperkernel, given by (2.6.7). Then, there exists a $\zeta(\widetilde{\kappa}) \in [0,1)$ such that*

$$|\mathcal{C}_{\max}| = \zeta(\widetilde{\kappa})n + o_p(n), \tag{2.6.11}$$

and $|\mathcal{C}_{(2)}| = o_p(n)$.

Theorem 2.15 is proved by showing that (in the hyperkernel case) the branching process that captures the 'local structure' of $\mathrm{IRG}_n(\widetilde{\kappa})$. For Theorem 2.15 to be useful we would like to know something about $\zeta(\widetilde{\kappa})$, which can be calculated from $\zeta_{\widetilde{\kappa}}$, which is in turn the largest solution to the functional equation:

$$f = 1 - \mathrm{e}^{-S_{\widetilde{\kappa}}(f)(x)}. \tag{2.6.12}$$

The question when $\zeta(\widetilde{\kappa}) > 0$ is settled in the following theorem:

Theorem 2.16 (Condition for existence giant component)**.** *Let $\widetilde{\kappa}$ be an integrable hyperkernel. Then, $\zeta(\widetilde{\kappa}) > 0$ if and only if $||T_{\kappa_e}|| > 1$. Furthermore, if $\widetilde{\kappa}$ is irreducible and $||T_{\kappa_e}|| > 1$, then $\zeta_{\widetilde{\kappa}}(x)$ is the unique non-zero solution to the functional equation (2.6.12), and $\zeta_{\widetilde{\kappa}}(x) > 0$ holds for a.e. x.*

In general, $||T_{\kappa_e}||$ may be rather hard to calculate. When we suppose that each κ_r is constant, however, this can be done. Indeed, say that $\kappa_r = c_r$. Then $\kappa_e(x,y) = \sum_r r(r-1)c_r = 2\xi(\kappa)$ for all x and y, so

$$||T_{\kappa_e}|| = 2\xi(\kappa). \tag{2.6.13}$$

This is perhaps surprising: it tells us that for such uniform hyperkernels, the critical point where a giant component emerges is determined only by the total number of

edges added; it does not matter what size cliques they lie in, even though, for example, the third edge in every triangle is 'wasted'. This turns out not to be true for arbitrary kernel families, where, rather each atom needs to be replaced by a clique.

2.7 Notes and discussion

Notes on Section 3.5. The path counting techniques in Proposition 3.17 are novel. Related proofs for the upper bound on H_n when $\nu < \infty$ often rely on branching process comparisons up to a generation $m = m_n \to \infty$.

Notes on Section 2.2. Theorem 2.4 is a special case of [24, Theorem 3.13]. Theorem 2.7 is a special case of [24, Theorem 3.1]. Earlier versions for random graphs with given expected degrees or Chung-Lu model appeared in [34, 38] (see also the monograph [37]).

The seminal paper [24] studies inhomogeneous random graph in an even more general setting, where the number of vertices in the graph need not be equal to n. In this case, the vertex space is called a *generalized vertex space*. We simplify the discussion here by assuming that the number of vertices is always equal to n. An example where the extension to a random number of vertices is crucially used is in [112], which studies an interpolation between percolation and $\mathrm{ER}_n(p)$.

In [24], there are various other results concerning the giant component of $\mathrm{IRG}_n(\boldsymbol{p}(\kappa))$. For example, [24, Theorem 3.9] proves that the giant component of $\mathrm{IRG}_n(\boldsymbol{p}(\kappa))$ is stable in the sense that its size does not change much if we add or delete a few edges. Note that the edges added or deleted do not have to be random or independent of the existing graph, rather, they can be chosen by a adversary after inspecting the whole of $\mathrm{IRG}_n(\boldsymbol{p}(\kappa))$. More precisely, [24, Theorem 3.9] shows that, for small enough $\delta > 0$, the giant component of $\mathrm{IRG}_n(\boldsymbol{p}(\kappa))$ in the supercritical regime does change by more than εn vertices if we remove any collection of δn edges.

Theorem 2.8 is a simplified version of [24, Theorem 3.14]. A first version of Theorem 2.8 was proved in [33, 35] for the expected degree random graph, in the case of *admissible* deterministic weights. We refer to [35, p. 94] for the definition of admissible degree sequences.

Theorem 3.4 for the expected degree random graph or Chung-Lu model is first proved in [33, 35], in the case of deterministic weights $w_i = c \cdot (i/n)^{-1/(\tau-1)}$, having average degree strictly greater than 1 and maximum degree m satisfying $\log m \gg \log n / \log\log n$. These restrictions were lifted in [46, Theorem 4.5.2]. Indeed, the bound on the average distance is not necessary, since, for $\tau \in (2,3)$, $\nu = \infty$ and therefore the IRG is *always* supercritical. An upper bound as in Theorem 3.4 for the Norros-Reittu model with i.i.d. weights is proved in [93].

Chapter 3
Small-world phenomena in rank-1 inhomogeneous random graphs

Abstract

In this chapter, we investigate the connectivity structure in rank-1 inhomogeneous random graphs. We focus on the phase transition in such random graphs, and on their small-world properties.

3.1 The phase transition in generalized random graphs

In this section, we investigate the size of the largest connected component in generalized random graphs.

Example 3.1 (Population of two types (Cont.))**.** Recall [I, Example 6.1], where we assumed that two distinct types of vertices are present. The first type has on average m_1 neighbors, the second type m_2, where $m_1 \neq m_2$. We have modeled this with a $\mathrm{GRG}_n(\boldsymbol{w})$ where n_1 vertices have weight m_1 and n_2 vertices have weight m_2, and write $n = n_1 + n_2$. Assume that $n_1/n \to p$. Under what conditions does $\mathrm{GRG}_n(\boldsymbol{w})$ have a giant component? When we compare with [I, Theorem 4.8], we may suspect that a giant component exists whenever a uniform vertex has on average more than 1 neighbor. The latter is true when

$$pm_1 + (1-p)m_2 > 1. \tag{3.1.1}$$

In this section, we examine whether this intuition is correct.

Exercise 3.1 (Average degree in two populations)**.** Show that the average degree is close to $pm_1 + (1-p)m_2$ in the setting of Example 3.1.

In order to state the main result in this section, we recall the limiting degree distribution in the generalized random graph given in (6.3.2)

$$p_k = \mathbb{E}\big[\mathrm{e}^{-W}\frac{W^k}{k!}\big], \qquad k \geq 0. \tag{3.1.2}$$

Then, Theorem 6.10 shows that the degree sequence in $\mathrm{GRG}_n(\boldsymbol{w})$ is close to $(p_k)_{k\geq 0}$. The main result is as follows:

Theorem 3.2 (Phase transition in generalized random graphs)**.** *Suppose that [I, Condition 6.4(a)-(b)] hold and consider the random graph* $\mathrm{NR}_n(\boldsymbol{w})$*, letting* $n \to \infty$*. Let* $\mathcal{C}_{\max}$ *and* $\mathcal{C}_{(2)}$ *be the largest and second largest components of* $\mathrm{NR}_n(\boldsymbol{w})$*.*

(a) If $\nu = \mathbb{E}[W^2]/\mathbb{E}[W] > 1$, *then there exist* $\xi \in (0,1), \zeta \in (0,1)$ *such that*

$$\begin{aligned} |\mathcal{C}_{\max}|/n &\xrightarrow{\mathbb{P}} \zeta, \\ v_k(\mathcal{C}_{\max})/n &\xrightarrow{\mathbb{P}} p_k(1-\xi^k), \text{ for every } k \geq 0, \\ |E(\mathcal{C}_{\max})|/n &\xrightarrow{\mathbb{P}} \tfrac{1}{2}\mathbb{E}[W](1-\xi^2). \end{aligned}$$

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

(b) If $\nu = \mathbb{E}[W^2]/\mathbb{E}[W] \leq 1$, *then* $|\mathcal{C}_{\max}|/n \xrightarrow{\mathbb{P}} 0$ *and* $|E(\mathcal{C}_{\max})|/n \xrightarrow{\mathbb{P}} 0$.

The above results apply to $\mathrm{GRG}_n(\boldsymbol{w})$ *and* $\mathrm{CL}_n(\boldsymbol{w})$ *under the same conditions.*

The proof of Theorem 3.2 is deferred to Section 4.1.2 in Chapter 4, where a similar result is proved for the configuration model. By the strong relation between the configuration model and the generalized random graph, this result can be seen to imply Theorem 3.2.

Exercise 3.2 (The phase transition for two populations)**.** *Show that the* $\zeta > 0$ *precisely when* $[pm_1^2 + (1-p)m_2^2]/[pm_1 + (1-p)m_2] > 1$ *in the setting of Example 3.1. Find an example of* p, m_1, m_2 *where the average degree is less than one, yet there exists a giant component.*

Reformulation in terms of branching processes. We start by reformulating the results in Theorem 3.2 in terms of *branching processes*. We can interpret ξ as the extinction probability of a branching process, and ζ as the survival probability of a related two-stage branching process.

We start by introducing two-stage branching processes with a mixed Poisson offspring. We define the branching process $(\mathcal{Z}_l)_{l\geq 0}$ as starting from $\mathcal{Z}_0 = 1$, where in the first generation the offspring distribution is equal to $(p_k)_{k\geq 0}$ given in (3.1.2), whereas in the second and further generations the offspring is chosen in accordance to

$$g_k = \frac{(k+1)p_{k+1}}{\mathbb{E}[W]} = \frac{1}{\mathbb{E}[W]}\mathbb{E}\big[\mathrm{e}^{-W}\frac{W^{k+1}}{k!}\big], \qquad k \geq 0, \tag{3.1.3}$$

where W is the mixing distribution. The expected number of offspring in the second and further generations is given by

$$\sum_{k=1}^{\infty} k g_k = \frac{1}{\mathbb{E}[W]}\mathbb{E}\big[\sum_{k=1}^{\infty} k(k+1)\mathrm{e}^{-W}\frac{W^{k+1}}{(k+1)!}\big] = \mathbb{E}[W^2]/\mathbb{E}[W] = \nu. \tag{3.1.4}$$

In particular, a branching process with mixed Poisson offspring distribution is supercritical when $\mathbb{E}[W^2] > \mathbb{E}[W]$. In Section 3.4.1, we relate the neighborhoods of $\mathrm{NR}_n(\boldsymbol{W})$ to an ordinary branching process having asymptotic expected offspring equal to $\nu = \mathbb{E}[W^2]/\mathbb{E}[W]$. Therefore, $\nu > 1$ is equivalent to the branching process being supercritical.

We now heuristically explain this relation to branching processes by describing the exploration of a vertex U chosen uniformly from the vertex set $[n]$. By Theorem 6.10,

the probability that its degree is k converges to p_k, for every $k \geq 1$. This explains the offspring of the root of our branching process approximation.

To describe the offspring of the direct neighbors of the root, we need to examine the forward degree of a uniform neighbor of the root. Here, by forward degree, we mean the number of vertices unequal to the root to which a neighbor of the root is connected. Intuitively, this forward degree is not much different from the degree minus one of a vertex contained in a *uniform edge*. We apply Theorem 6.6, which states that the total number of edges in $\mathrm{GRG}_n(\boldsymbol{w})$ and $\mathrm{NR}_n(\boldsymbol{w})$ is well approximated by its mean ℓ_n. We also use that the probability that an edge is occupied is close to $w_i w_j/\ell_n$, so that the probability $g_k^{(n)}$ that the forward degree equals k is close to

$$\begin{aligned} g_k^{(n)} &\approx \frac{1}{\ell_n} \sum_{i,j\in[n]} \mathbb{P}(ij \text{ occupied}, D_j = k+1) \\ &= \frac{1}{\ell_n} \sum_{i,j\in[n]} \frac{w_i w_j}{\ell_n} \mathbb{P}(X_j = k) \\ &= \frac{1}{\ell_n} \sum_{j\in[n]} w_j \mathbb{P}(X_j = k), \end{aligned} \tag{3.1.5}$$

where $X_j = \sum_{s\in[n]:\, s\neq i,j} I_{sj}$, where (I_{sj}) are independent Bernoulli variables with $\mathbb{P}(I_{sj} = 1) = w_s w_j/\ell_n$. By Theorem 2.10, X_j is close to a Poisson random variable with mean $\sum_{s\in[n]} w_s w_j/\ell_n = w_j$, so that

$$g_k^{(n)} \approx \frac{1}{\ell_n} \sum_{j\in[n]} w_j \mathbb{P}(\mathsf{Poi}(w_j) = k) = \mathbb{P}(\mathsf{Poi}(W_n^*) = k), \tag{3.1.6}$$

where W_n^* is the size-biased distribution of W_n. Here, for a non-negative random variable X with $\mathbb{E}[X] > 0$, we let X^* denote its *size-biased distribution* given by

$$\mathbb{P}(X^* \leq x) = \frac{\mathbb{E}[X \mathbb{1}_{\{X\leq x\}}]}{\mathbb{E}[X]}. \tag{3.1.7}$$

When [I, Condition 6.4(a)-(b)] holds, we have that $W_n^* \xrightarrow{d} W^*$, which is the size-biased distribution of W. We arrive at

$$g_k^{(n)} \approx \mathbb{P}(\mathsf{Poi}(W^*) = k) = \frac{1}{\mathbb{E}[W]} \mathbb{E}\big[\mathrm{e}^{-W} \frac{W^{k+1}}{k!}\big] = g_k, \tag{3.1.8}$$

which explains (3.1.3). It turns out that the steps in the above heuristic explanation can be made precise. See Section 3.4.1, where we relate the neighborhood of a uniform vertex in $\mathrm{NR}_n(\boldsymbol{w})$ to a mixed Poisson random variable.

The above heuristically argues that the number of vertices unequal to the root connected to any neighbor of the root has asymptotic law $(g_k)_{k\geq 0}$. However, every time we find a vertex in the cluster of the root, the number of *available* vertices decreases, in a similar way as the depletion-of-points effect in the exploration of clusters for the Erdős-Rényi random graph $\mathrm{ER}_n(\lambda/n)$. Since the number of vertices

is originally n, we can grow the cluster of the root for a long time before we note this effect.

As a result, the degrees of different vertices in the exploration process is close to being i.i.d., leading to a branching process approximation. In terms of this branching process, we can interpret ζ in Theorem 3.2 as the survival probability of the above two-stage branching process, so that ζ satisfies

$$\zeta = \sum_{k=1}^{\infty} p_k(1-\xi^k), \tag{3.1.9}$$

where ξ is the extinction probability of the branching process with offspring distribution $(g_k)_{k\geq 0}$. Clearly, $\xi = 1$ precisely when

$$\nu = \sum_{k\geq 0} k g_k \leq 1, \tag{3.1.10}$$

which explains the condition on ν in Theorem 3.2(a). Further, by Theorem 6.10, there are approximately np_k vertices with degree k. Assuming approximate independence of each of these k neighbors, each of them survives with probability $1-\xi$, so that the probability that at least one survives equals $1-\xi^k$. When one of the neighbors survive, the vertex itself will be part of the giant component, which explains why $v_k(\mathcal{C}_{\max})/n \xrightarrow{\mathbb{P}} p_k(1-\xi^k)$. Finally, an edge consists of two half-edges, and an edge is part of the giant component precisely when one of the vertices incident to it is, which occurs with probability $1-\xi^2$. There are in total approximately $\ell_n/2 \approx n\mathbb{E}[W]/2$ edges, which explains why $|E(\mathcal{C}_{\max})|/n \xrightarrow{\mathbb{P}} \frac{1}{2}\mathbb{E}[W](1-\xi^2)$. Therefore, all results in Theorem 3.2 have a simple explanation in terms of the branching process approximation of the connected component of a uniform vertex in $[n]$ for $\mathrm{NR}_n(\boldsymbol{w})$.

Exercise 3.3 (Degree sequence of giant component)**.** *Show that the proportion of vertices of the giant component $\mathcal{C}_{\max}$ having degree k is close to $p_k(1-\xi^k)/\zeta$.*

Exercise 3.4 (Degree sequence of complement of giant component)**.** *Show that when $\xi < 1$, the proportion of vertices outside the giant component $\mathcal{C}_{\max}$ having degree k is close to $p_k\xi^k/(1-\zeta)$. Conclude that the degree sequence of the complement of the giant component never satisfies a power law. Can you give an intuitive explanation for this?*

We close this section by discussing the consequences of the phase transition for the attack vulnerability of $\mathrm{CL}_n(\boldsymbol{w})$:

Attack vulnerability of $\mathrm{CL}_n(\boldsymbol{w})$. Suppose an adversary attacks a network by removing some of its vertices. A *clever* adversary would remove the vertices in a clever way, this is often referred to as a *deliberate* attack. On the other hand, the vertices might also be exposed to random failures, which is often referred to as a *random* attack. The results as stated above do not specifically apply to these settings, but do have intuitive consequences. We model a deliberate attack as the removal of a proportion of the vertices with highest weights, whereas a random attack is modeled

by random removal of the vertices with a given probability. One of the aims is to quantify the effect of such attacks, and in particular the difference in random and deliberate attacks. We shall denote the proportion of removed vertices by p. We shall always assume that $\nu > 1$, so that a giant component exists, and we investigate under what conditions on p and the graph $\mathrm{CL}_n(\boldsymbol{w})$, the giant component remains to exist.

We start by addressing the case of random attack for the $\mathrm{CL}_n(\boldsymbol{w})$ model under [I, Condition 6.4(a)-(c)], where $\mathbb{E}[W^2] < \infty$. One of the difficulties of the above set-up is that we remove *vertices* rather than *edges*, so that the resulting graph is no longer an IRG. In percolation jargon, we deal with site percolation rather than with edge percolation. We start by relating the obtained graph to an IRG.

Note that when we *explore* a cluster of a vertex after an attack, then the vertex may not have been affected by the attack, which has probability p. After this, in the exploration, we shall always inspect an edge between a vertex which is unaffected by the attack and a vertex of which we do not yet know whether it has been attacked or not. As a result, for random attacks, the probability that it is affected is precisely equal to p. Therefore, it is similar to the random graph where p_{ij} is replaced with $(1-p) \times p_{ij}$. For a branching process, this identification is exact, and we have that $\zeta_{\kappa,p} = (1-p)\zeta_{(1-p)\kappa}$, where $\zeta_{\kappa,p}$ denotes the survival probability of the branching process where each individual is killed with probability p independently of all other randomness. For $\mathrm{CL}_n(\boldsymbol{w})$, this equality is only asymptotic. In the case where $\mathbb{E}[W^2] < \infty$, so that $\nu < \infty$, this means that there exists a critical value $p_c = 1 - 1/\nu$, such that if $p < p_c$, the $\mathrm{CL}_n(\boldsymbol{w})$ where vertices are removed with probability p, the giant component persists, while if $p > p_c$, then the giant component is destroyed. Thus, when $\mathbb{E}[W^2] < \infty$, the $\mathrm{CL}_n(\boldsymbol{w})$ is *sensitive to random attacks*. When $\mathbb{E}[W^2] = \infty$, on the other hand, $\nu = \infty$, so that the giant component persists for *every* $p \in [0,1)$, and the graph is called *robust to random attacks*. Here we must note that the *size* of the giant component does decrease, since $\zeta_{\kappa,p} < p\zeta_\kappa$!

For a deliberate attack, we remove the proportion p of vertices with highest weight. This means that $\boldsymbol{w}$ is replaced with $\boldsymbol{w}(p)$, which is equal to $w_i(p) = w_i \mathbb{1}_{\{i > np\}}$, and we denote the resulting edge probabilities by

$$p_{ij}(p) = \max\{1, w_i(p)w_j(p)/\ell_n\}. \tag{3.1.11}$$

In this case, the resulting graph on $[n] \setminus [np]$ is again a Chung-Lu model, for which ν is replaced with $\nu(p)$ given by

$$\nu(p) = \mathbb{E}[[1-F]^{-1}(U)^2 \mathbb{1}_{\{U>p\}}]/\mathbb{E}[W], \tag{3.1.12}$$

where U is uniform on $[0,1]$. Now, for any distribution function F, $\mathbb{E}[[1-F]^{-1}(U)^2 \mathbb{1}_{\{U>p\}}] < \infty$, so that, for p sufficiently close to 1, $\nu(p) < 1$. Thus, the $\mathrm{CL}_n(\boldsymbol{w})$ model is always sensitive to deliberate attacks.

Exercise 3.5 (Finiteness of $\nu(p)$)**.** *Prove that $\nu(p) < \infty$ for every $p \in (0,1]$ and any distribution function F.*

Exercise 3.6 (Connectivity of uniformly chosen vertices)**.** *Suppose we draw two vertices uniformly at random from $[n]$. Prove that Theorem 3.2 implies that the probability that the vertices are connected converges to ζ^2.*

3.2 The small-world phenomenon in generalized random graphs

In this section, we discuss typical distances in $\mathrm{NR}_n(\boldsymbol{w})$. We define H_n to be the graph distance between two vertices chosen uniformly at random from $[n]$, 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 $H_n = +\infty$. By Theorem 3.2, $\mathbb{P}(H - n = +\infty) \to 1 - \zeta^2 > 0$, since $\zeta < 1$ (see Exercise 3.6). In particular, when $\zeta = 0$, which is equivalent to $\nu \leq 1$, $\mathbb{P}(H - n = +\infty) \to 1$. Therefore, in our main results, we shall condition on $H_n < \infty$.

Distances in inhomogeneous random graphs with finite variance weights. We start by investigating the behavior of H_n for $\mathrm{NR}_n(\boldsymbol{w})$ in the case where the weights have finite variance:

Theorem 3.3 (Typical distances in $\mathrm{NR}_n(\boldsymbol{w})$ for finite-variance weights)**.** *In the Norros-Reittu model* $\mathrm{NR}_n(\boldsymbol{w})$*, where the weights* $\boldsymbol{w} = (w_i)_{i\in[n]}$ *satisfy [I, Condition 6.4(a)-(c)] and where* $\nu > 1$*, conditionally on* $H_n < \infty$*,*

$$H_n/\log n \xrightarrow{\mathbb{P}} 1/\log \nu. \tag{3.2.1}$$

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

We give a complete proof of Theorem 3.3 in Sections 3.3-3.5 below. The intuition behind Theorem 3.3 is as follows. In Section 3.1, we have argued that the neighborhood of a uniform vertex in $\mathrm{NR}_n(\boldsymbol{w})$ is well-approximated by a two-stage branching process, where the second and all later generations have offspring distribution $(g_k)_{k\geq 0}$ in (3.1.3). When $\nu = \sum_{k\geq 0} k g_k < \infty$, then the number of vertices at distance k is close to $M\nu^k$, where M is the martingale limit of $\mathcal{Z}_k/\nu^k$. To know what H_n 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 $\mathcal{Z}_k \approx n$, which suggests that the relevant k is such that $\nu^k \approx n$, so that $k \approx \log_\nu n$.

While the above heuristic is quite convincing, the argument is fatally flawed. Indeed, as argued in Section 3.1, the neighborhoods of a uniform vertex are well-approximated by a branching process as long as the number of vertices found is much smaller than n. When the number of vertices found becomes of order n, the depletion-of-points effect has already started to kick in. Therefore, the above approach is doomed to fail. Our proof instead, is divided in a lower and and upper bound on the typical distance H_n. For the proof of the lower bound in Section 3.3.1, we show that the *expected* number of paths of k edges between two uniform vertices is approximately ν^k/ℓ_n, so that such a path whpdoes not exist when $k \leq (1-\varepsilon)\log_\nu n$. For the proof of the upper bound in Section 3.5, we use a second moment method to show that, conditionally on the two uniformly chosen vertices being in the giant component, whpthere exists a path of $(1+\varepsilon)\log_\nu n$ edges.

Exercise 3.7 (Typical distances in $\mathrm{ER}_n(\lambda/n)$)**.** *Prove that* $H_n/\log n \xrightarrow{\mathbb{P}} 1/\log\lambda$ *in* $\mathrm{ER}_n(\lambda/n)$*.*

Theorem 3.3 leaves open what happens when $\nu = \infty$. We can use Theorem 3.3 to show that $H_n = o_{\mathbb{P}}(\log n)$:

Exercise 3.8 (Typical distances when $\nu = \infty$). *Prove that* $H_n/\log n \xrightarrow{\mathbb{P}} 0$ *when* $\nu = \infty$.

Distances in inhomogeneous random graphs with infinite variance weights. In this section, we study typical distances in the Norros-Reittu random graph $\mathrm{NR}_n(\boldsymbol{w})$, 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 $H_n = o_{\mathbb{P}}(\log n)$ (recall Exercise 3.8).

Many of our arguments also apply to the generalized random graph $\mathrm{GRG}_n(\boldsymbol{w})$ and the Chung-Lu model $\mathrm{CL}_n(\boldsymbol{w})$. In this section, we discuss the setting where the weights $\boldsymbol{w}$ are heavy tailed. Recall that $F_n(x)$ denotes the proportion of vertices i for which $w_i \le x$. Then, we assume that there exists a $\tau \in (2,3)$ such that for all $\delta > 0$, there exists $c_1 = c_1(\delta)$ and $c_2 = c_2(\delta)$ such that, uniformly in n,

$$c_1 x^{-(\tau-1+\delta)} \le [1 - F_n](x) \le c_2 x^{-(\tau-1-\delta)}, \tag{3.2.2}$$

where the upper bound is expected to hold for every $x \ge 1$, while the lower bound is only required to hold for $1 \le x \le n^{\alpha}$ for some $\alpha > 1/2$.

The assumption in (3.2.2) 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 (3.2.2) cannot be valid for *all* x, since $F_n(x) > 0$ implies that $F_n(x) \ge 1/n$, so that the lower and upper bound in (3.2.2) 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 $x \le n^{\alpha}$ for *some* $\alpha \in (1/2, 1)$.

We now give simpler conditions for (3.2.2) in special cases:

Exercise 3.9 (Power-law tails in key example of deterministic weights). *Let* $\boldsymbol{w}$ *be defined as in* (6.1.14)*, and assume that* F *satisfies*

$$1 - F(x) = x^{-(\tau-1)} L(x), \tag{3.2.3}$$

where the exponent satisfies $\tau \in (2,3)$*, and where* $x \mapsto L(x)$ *is slowly varying. Prove that* (3.2.2) *holds.*

Exercise 3.10 (Power-law tails for i.i.d. weights). *For i.i.d. weights* $\boldsymbol{w} = (w_i)_{i \in [n]}$ *with distribution* F *satisfying that* (3.2.3) *with* $\tau \in (2,3)$*, and where* $x \mapsto L(x)$ *is slowly varying. Prove that* (3.2.2) *holds with probability converging to 1.*

Our main result is as follows:

Theorem 3.4 (Typical distances in $\mathrm{NR}_n(\boldsymbol{w})$ for $\tau \in (2,3)$). *Fix the Norros-Reittu model* $\mathrm{NR}_n(\boldsymbol{w})$*, where the weights* $\boldsymbol{w} = (w_i)_{i \in [n]}$ *satisfy [I, Condition 6.4(a)-(b)] and* (3.2.2)*. Then, conditionally on* $H_n < \infty$*,*

$$\frac{H_n}{\log\log n} \xrightarrow{\mathbb{P}} \frac{2}{|\log(\tau-2)|}. \tag{3.2.4}$$

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

Theorem 3.4 implies that $\mathrm{NR}_n(\boldsymbol{w})$ with $\boldsymbol{w}$ as in (6.1.14), for $\tau \in (2,3)$, is an ultra-small world when (3.3.25) is satisfied.

The main tool to study distances in $\mathrm{NR}_n(\boldsymbol{w})$ is a comparison to branching processes, which is particularly pretty for $\mathrm{NR}_n(\boldsymbol{w})$. In the next two sections, we prove Theorems 3.3–3.4. When $\tau > 3$, then the branching process approximation has *finite mean*, and we can make use of the martingale limit results of the number of individuals in generation k as $k \to \infty$. When $\tau \in (2,3)$, on the other hand, the branching process has infinite mean. In this case, the number of individuals in generation k, conditionally on survival, grows *superexponentially*, which explains why distances grow doubly logarithmically. See Section 5.4, where this is explained in more detail in the context of the configuration model. The superexponential growth implies that a path between two vertices typically passes through vertices with growing weights as we move away from the two vertices. Thus, starting from the first vertex $U_1 \in [n]$, the path connecting U_1 to U_2 uses vertices that first grow until the midpoint of the path is reached, and then decrease again to reach U_2. This can be understood by noting that the probability that vertex with weight w is not connected to any vertex with weight larger than $y > w$ in $\mathrm{NR}_n(\boldsymbol{w})$ is

$$\mathrm{e}^{-\sum_{i:\, w_i > y} w w_i/\ell_n} = \mathrm{e}^{-w[1-F_n^*](y)}, \tag{3.2.5}$$

where $F_n^*(y) = \sum_{i:\, w_i \le y} w_i/\ell_n$ is the distribution function of W_n^* introduced in (3.1.6). When (3.2.2) holds, it follows that $[1-F_n^*](y)$ is close to $y^{-(\tau-2)}$, the size-biasing increasing the power by one. Therefore, the probability that vertex with weight w is not connected to any vertex with weight larger than $y > w$ in $\mathrm{NR}_n(\boldsymbol{w})$ is approximately $\mathrm{e}^{-wy^{-(\tau-2)}}$. Take w large, then this probability is small when $y \gg w^{1/(\tau-2)}$. Thus, a vertex of weight w is whp connected to a vertex of weight $w^{1/(\tau-2)}$. Since $1/(\tau-2) > 1$ when $\tau \in (2,3)$, we obtain that vertices with large weights w are whp connected to vertices with weight at least $w^{1/(\tau-2)}$.

The proof of Theorems 3.3–3.4 are organized as follows. In Section 3.3, we prove the lower bounds on the typical distance in $\mathrm{NR}_n(\boldsymbol{w})$, both when $\tau > 3$ and when $\tau \in (2,3)$. In Section 3.4, we describe a stochastic domination of the neighborhoods in $\mathrm{NR}_n(\boldsymbol{w})$ in terms of a marked Poisson branching process and prove the $\log\log n$ upper bound for $\tau \in (2,3)$. In Section 3.5, we investigate the number of paths between sets of vertices in $\mathrm{NR}_n(\boldsymbol{w})$, and use this to prove the $\log n$ upper bound when $\tau > 3$. In each of our proofs, we formulate the precise results as separate theorems, and prove them under conditions that are slightly weaker than those in Theorems 3.3–3.4.

3.3 Lower bounds on typical distances

In this section, we prove lower bounds on typical graph distances. In Section 3.3.1, we prove the lower bound in Theorem 2.8(i) in the setting of Theorem 3.3.

3.3.1 Logarithmic lower bound graph distance in finite variance case

In this section, we prove a logarithmic lower bound on the graph distance in $\mathrm{NR}_n(\boldsymbol{w})$. The main result is as follows:

Theorem 3.5 (Logarithmic lower bound graph distances $\mathrm{NR}_n(\boldsymbol{w})$)**.** *Assume that*

$$\limsup_{n\to\infty} \nu_n = \nu, \tag{3.3.1}$$

where $\nu \in (1,\infty)$ *and*

$$\nu_n = \mathbb{E}[W_n^2]/\mathbb{E}[W_n] = \sum_{i\in[n]} w_i^2 / \sum_{i\in[n]} w_i. \tag{3.3.2}$$

Then, for any $\varepsilon > 0$*,*

$$\mathbb{P}(H_n \le (1-\varepsilon)\log_\nu n) = o(1). \tag{3.3.3}$$

The same results hold for $\mathrm{CL}_n(\boldsymbol{w})$ *and* $\mathrm{GRG}_n(\boldsymbol{w})$ *under the same conditions.*

Proof. We recall that H_n is the graph distance in $\mathrm{NR}_n(\boldsymbol{w})$ between two vertices drawn uniformly at random from $[n]$. The idea behind the proof of Theorem 3.5 is that it is quite unlikely that a path exists that is much shorter than $\log_\nu n$ edges. In order to show this, we use a first moment bound and show that the *expected* number of occupied paths connecting the two vertices chosen uniformly at random from $[n]$ having length at most k is $o(1)$. We now fill in the details.

We denote the graph distance between $i,j\in[n]$ in $\mathrm{NR}_n(\boldsymbol{w})$ by $\mathrm{dist}_{\mathrm{NR}_n(\boldsymbol{w})}(i,j)$ and abbreviate $k_n = \lceil (1-\varepsilon)\log_\nu n\rceil$. Then, conditioning on the uniform vertices chosen and Boole's inequality gives

$$\begin{aligned}\mathbb{P}(H_n \le k_n) &= \frac{1}{n^2}\sum_{i,j\in[n]} \mathbb{P}(\mathrm{dist}_{\mathrm{NR}_n(\boldsymbol{w})}(i,j) \le k_n)\\ &= \frac{1}{n^2}\sum_{i,j\in[n]}\sum_{k=0}^{k_n} \mathbb{P}(\mathrm{dist}_{\mathrm{NR}_n(\boldsymbol{w})}(i,j) = k).\end{aligned} \tag{3.3.4}$$

We make use of *path counting techniques.* A path $\pi = (\pi_0,\dots,\pi_k)$ of length k between vertices i and j is a sequence of vertices connecting $\pi_0 = i$ to $\pi_k = j$. We call a path π *self-avoiding* when it visits every vertex at most once, i.e., $\pi_i \neq \pi_j$ for every $i \neq j$. Let $\mathcal{P}_k(i,j)$ denote the set of k-step self-avoiding paths between vertices i and j. See Figure 3.3.1 for an example of a 12-step self-avoiding path between i and j.

When $\mathrm{dist}_{\mathrm{NR}_n(\boldsymbol{w})}(i,j) = k$, there must be path of length k such that all edges (π_l,π_{l+1}) are occupied in $\mathrm{NR}_n(\boldsymbol{w})$, for $l = 0,\dots,k-1$. The probability in $\mathrm{NR}_n(\boldsymbol{w})$ that the edge (π_l,π_{l+1}) is occupied is equal to

$$1 - \mathrm{e}^{-w_{\pi_l}w_{\pi_{l+1}}/\ell_n} \le w_{\pi_l}w_{\pi_{l+1}}/\ell_n. \tag{3.3.5}$$

For $\mathrm{CL}_n(\boldsymbol{w})$ and $\mathrm{GRG}_n(\boldsymbol{w})$, an identical upper bound holds.

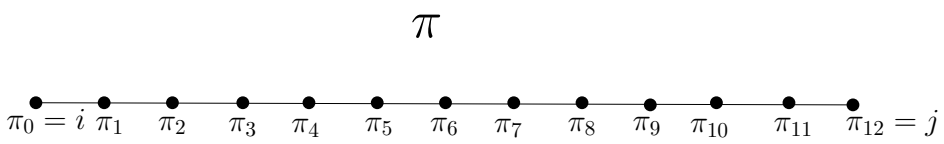

Figure 3.1: A 12-step self-avoiding path connecting vertices i and j.

We say that π is *occupied* when all edges in π are occupied in $\mathrm{NR}_n(\boldsymbol{w})$. Then, by Boole's inequality,

$$\mathbb{P}(\mathrm{dist}_{\mathrm{NR}_n(\boldsymbol{w})}(i,j)=k) \le \mathbb{P}(\exists \pi \in \mathcal{P}_k(i,j)\colon \pi \text{ occupied}) \le \sum_{\pi\in\mathcal{P}_k(i,j)} \mathbb{P}(\pi \text{ occupied}). \tag{3.3.6}$$

For any path $\pi \in \mathcal{P}_k(i,j)$,

$$\begin{aligned}\mathbb{P}(\pi \text{ occupied}) &= \prod_{s=0}^{k-1} \mathbb{P}((\pi_l,\pi_{l+1}) \text{ occupied}) \le \prod_{l=0}^{k-1} w_{\pi_l} w_{\pi_{l+1}}/\ell_n \\ &= \frac{w_{\pi_0} w_{\pi_k}}{\ell_n} \prod_{l=1}^{k} w_{\pi_l}^2/\ell_n = \frac{w_i w_j}{\ell_n} \prod_{l=1}^{k} w_{\pi_l}^2/\ell_n.\end{aligned} \tag{3.3.7}$$

Therefore,

$$\begin{aligned}\mathbb{P}(\mathrm{dist}_{\mathrm{NR}_n(\boldsymbol{w})}(i,j)=k) &\le \frac{w_i w_j}{\ell_n} \sum_{\pi\in\mathcal{P}_k(i,j)} \prod_{l=1}^{k} \frac{w_{\pi_l}^2}{\ell_n} \\ &= \frac{w_i w_j}{\ell_n} \prod_{l=1}^{k} \Big(\sum_{\pi_l\in[n]} \frac{w_{\pi_l}^2}{\ell_n} \Big) = \frac{w_i w_j}{\ell_n} \nu_n^k,\end{aligned} \tag{3.3.8}$$

where ν_n is defined in (3.3.2), so that

$$\mathbb{P}(H_n \le k_n) \le \frac{1}{n^2} \sum_{i,j\in[n]} \sum_{k=0}^{k_n} \frac{w_i w_j}{\ell_n} \nu_n^k = \frac{\ell_n}{n^2} \sum_{k=0}^{k_n} \nu_n^k = \frac{\ell_n}{n^2} \frac{\nu_n^{k_n+1}-1}{\nu_n - 1}. \tag{3.3.9}$$

By (3.3.1), $\limsup_{n\to\infty} \nu_n = \nu \in (1,\infty)$, so that, for n large enough, $\nu_n \ge (\nu-\delta) > 1$, while $\ell_n/n = \mathbb{E}[W_n] \to \mathbb{E}[W] < \infty$. Thus, since $\nu \mapsto (\nu^{k+1}-1)/(\nu-1)$ is increasing for every integer $k \ge 0$,

$$\mathbb{P}(H_n \le k_n) \le O((\nu-\delta)^{k_n}/n) = o(1), \tag{3.3.10}$$

when $\delta = \delta(\varepsilon) > 0$ is chosen such that $(1-\varepsilon)/\log(\nu-\delta) < 1$, and since $k_n = \lceil (1-\varepsilon)\log_\nu n \rceil$. □

The condition (3.3.1) is slightly weaker than [I, Condition 6.4], which is assumed in Theorem 3.3:

Exercise 3.11 (Conditions (3.3.1) and [I, Condition 6.4]). *Show that when there is precisely one vertex with weight $w_1 = \sqrt{n}$, whereas $w_i = \lambda > 1$, then (3.3.1) holds, but [I, Condition 6.4] does not. Argue that the upper bound derived in Theorem 3.5 is not sharp, since the vertex 1 can occur at most once in a self-avoiding path.*

Exercise 3.12 (Lower bound on fluctuations). *Adapt the proof of Theorem 3.5 to show that for every ε, we can find a constant $K = K(\varepsilon) > 0$ such that*

$$\mathbb{P}\big(H_n \le \frac{\log n}{\log \nu_n} - K\big) \le \varepsilon. \tag{3.3.11}$$

Conclude that if $\log \nu_n = \log \nu + o(1/\log n)$, then the same statement holds with $\log_\nu n$ replacing $\frac{\log n}{\log \nu_n}$.

We close this section by extending the above result to settings where ν_n is not necessarily bounded:

Corollary 3.6 (Lower bound graph distances $\mathrm{NR}_n(\boldsymbol{w})$ for $\tau = 3$). *Let ν_n be given in (3.3.2). Then, for any $\varepsilon > 0$,*

$$\mathbb{P}\Big(H_n \le (1-\varepsilon)\log_{\nu_n} n\Big) = o(1). \tag{3.3.12}$$

The same results hold for $\mathrm{CL}_n(\boldsymbol{w})$ and $\mathrm{GRG}_n(\boldsymbol{w})$ under the same conditions.

The proof of Corollary 3.6 is left as an exercise:

Exercise 3.13 (Proof Corollary 3.6). *Adapt the proof to Theorem 3.5 to prove Corollary 3.6.*

Exercise 3.14 (Lower bound on typical distances for $\tau = 3$). *Let $w_i = c\sqrt{(n/i)}$, so that $\tau = 3$. Prove that $\nu_n/\log n \to c$. Use Corollary 3.6 to obtain that for any $\varepsilon > 0$,*

$$\mathbb{P}\Big(H_n \le (1-\varepsilon)\frac{\log n}{\log\log n}\Big) = o(1). \tag{3.3.13}$$

Exercise 3.15 (Lower bound on typical distances for $\tau \in (2,3)$). *Let $w_i = c/i^{1/(\tau-1)}$ with $\tau \in (2,3)$. Prove that there exists a constant $c' > 0$ such that $\nu_n \ge c' n^{(3-\tau)/(\tau-1)}$. Show that Corollary 3.6 implies that $H_n \ge (\tau-1)/(\tau-3)$ in this case. How useful is this bound?*

Discussion of the proof of Theorem 2.8(i). Theorem 2.8(i) is closely related to Theorem 3.5. The proof of Theorem 3.5 can be extended to $\mathrm{IRG}_n(\boldsymbol{p}(\kappa_n))$ for certain (κ_n) by noting that

$$\mathbb{P}(\mathrm{dist}_{\mathrm{IRG}_n(\boldsymbol{p}(\kappa))}(i,j) = k) \le \sum_{i_1,\ldots,i_{k-1}\in[n]} \prod_{l=0}^{k-1} \frac{\kappa_n(x_{i_l}, x_{i_{l+1}})}{n}, \tag{3.3.14}$$

where $i_0 = i, i_k = j$, so that

$$\mathbb{P}(H_n = k) \le \frac{1}{n^k} \sum_{i_0,i_1,\ldots,i_{k-1},i_k \in [n]} \prod_{l=0}^{k-1} \kappa_n(x_{i_l}, x_{i_{l+1}}). \tag{3.3.15}$$

If the above $(k+1)$-dimensional discrete integrals could be replaced by the continuous integral, then we would arrive at

$$\frac{1}{n} \int_{\mathcal{S}} \cdots \int_{\mathcal{S}} \prod_{l=0}^{k} \kappa(x_l, x_{l+1}) \prod_{i=0}^{k} \mu(dx_i) = \frac{1}{n} \|\mathbf{T}_\kappa^{k+1} \mathbf{1}\|_1, \tag{3.3.16}$$

which is bounded from above by $\frac{1}{n}\|\mathbf{T}_\kappa\|^{k+1}$. Repeating the bound in (3.3.10) would then prove that, when $\nu = \|\mathbf{T}_\kappa\| > 1$,

$$\mathbb{P}(H_n \le (1-\varepsilon) \log_\nu n) = o(1). \tag{3.3.17}$$

However, in the general case, it is not so easy to replace the $(k+1)$-fold discrete sum in (3.3.15) by a $(k+1)$-fold integral. In the exercises below, we show how we can carry through the above argument:

Exercise 3.16 (Logarithmic lower bound graph distances $\mathrm{IRG}_n(\boldsymbol{p}(\kappa_n))$ of finite–types)**.** *Suppose that $\mathrm{IRG}_n(\boldsymbol{p}(\kappa_n))$ is of finite type, and assume that $n_i/n \to p_i$, where n_i is the number of vertices of type i. Suppose further that the probability that a vertex of type i is directly connected to one of type j is equal to $[\kappa_{ij} \wedge n]/n$ for some matrix $(\kappa_{ij})_{i,j\in[r]}$ with largest eigenvalue $\nu = \|\mathbf{T}_\kappa\| > 1$. Prove that*

$$\mathbb{P}(H_n \le (1-\varepsilon) \log_\nu n) = o(1). \tag{3.3.18}$$

Exercise 3.17 (Extension to the setting of Theorem 2.8)**.** *Assume that the conditions in Theorem 2.8 hold. Recall the bound in (2.2.31), which bounds κ_n from above by κ_m^+, which is of finite-type. Then, use the fact that $\|\mathbf{T}_{\kappa_m^+}\| \downarrow \|\mathbf{T}_\kappa\| = \nu > 1$ to conclude that $\mathbb{P}(H_n \le (1-\varepsilon)\log_\nu n) = o(1)$ holds under the conditions of Theorem 2.8.*

Note that

$$|\{\{v,w\}\colon \mathrm{dist}_G(v,w) < \infty\}| = \sum_i \binom{|\mathcal{C}_{(i)}|}{2}, \tag{3.3.19}$$

where $\{v,w\}$ denotes an unordered pair of distinct vertices of G and where $\mathcal{C}_{(i)}$ denote the connected components of G arranged in decreasing order, so that $\mathcal{C}_{(1)} = \mathcal{C}_{\max}$. Thus, by Theorem 2.7, if κ is irreducible, then

$$\mathbb{P}(H_n < \infty) = \zeta_\kappa^2 + o(1). \tag{3.3.20}$$

Thus, Theorem 2.8(i–ii) can be reformulated by saying that if $\sup_{x,y,n} \kappa_n(x,y) < \infty$ and if $\nu = \|\mathbf{T}_\kappa\| > 1$, then

$$\mathbb{P}\Big(1-\varepsilon \le \frac{H_n}{\log_\nu n} \le 1+\varepsilon \Big| H_n < \infty\Big) \to 1. \tag{3.3.21}$$

As a result, conditionally on $H_n < \infty$, we have that $H_n/\log n \xrightarrow{\mathbb{P}} 1/\log \nu$:

Exercise 3.18 (Convergence in probability of typical distance in $\mathrm{IRG}_n(\boldsymbol{p}(\kappa_n))$). *Suppose that the graphical sequence of kernels (κ_n) satisfies $\sup_{x,y,n}\kappa_n(x,y)<\infty$, where the limit κ is irreducible and $\nu=\|\mathbf{T}_\kappa\|>1$. Prove that Theorem 2.7 together with Theorem 2.8(i–ii) imply that, conditionally on $H_n<\infty$,*

$$H_n/\log n \xrightarrow{\mathbb{P}} 1/\log\nu. \tag{3.3.22}$$

Theorem 2.8 leaves open the case when $\|\mathbf{T}_\kappa\|=\infty$, which, for example for $\mathrm{CL}_n(\boldsymbol{w})$, is the case when F has infinite second moment. [24, Theorem 3.14(iv)] states that when $\|\mathbf{T}_\kappa\|=\infty$, the typical graph distance is smaller than $\log n$. More precisely, [24, Theorem 3.14(iv)] states that if κ is irreducible and $\|\mathbf{T}_\kappa\|=\infty$, then there is a function $f(n)=o(\log n)$ such that

$$\mathbb{P}(H_n\le f(n))=\zeta_\kappa^2+o(1). \tag{3.3.23}$$

Exercise 3.19 (Convergence in probability of typical distance in $\mathrm{IRG}_n(\boldsymbol{p}(\kappa_n))$). *Suppose that the graphical sequence of kernels (κ_n) converges to κ, where κ is irreducible and $\|\mathbf{T}_\kappa\|=\infty$. Prove that Theorem 2.7 together with Theorem 2.8(iii) imply, conditionally on $H_n<\infty$,*

$$H_n/\log n \xrightarrow{\mathbb{P}} 0. \tag{3.3.24}$$

3.3.2 A $\log\log$ lower bound on typical distances in the infinite variance case

In this section, we prove a $\log\log$-lower bound on the typical distances of $\mathrm{NR}_n(\boldsymbol{w})$ for $\tau\in(2,3)$. The main result we prove is the following theorem:

Theorem 3.7 (Loglog lower bound on typical distances in $\mathrm{NR}_n(\boldsymbol{w})$). *Suppose that the weights $\boldsymbol{w}=(w_i)_{i\in[n]}$ satisfy [I, Condition 6.4(a)] and that there exists a $\tau\in(2,3)$ and c_2 such that, for all $x\ge 1$,*

$$[1-F_n](x)\le c_2 x^{-(\tau-1)}, \tag{3.3.25}$$

Then, for every $\varepsilon>0$,

$$\mathbb{P}\Big(H_n\le(1-\varepsilon)\frac{2\log\log n}{|\log(\tau-2)|}\Big)=o(1). \tag{3.3.26}$$

We follow the proof of Theorem 3.5 as closely as possible. The problem with that proof is that, under the condition in (3.3.25), ν_n is too large. Indeed, Exercise 3.15 shows that the lower bound obtained in Corollary 3.6 is a constant, which is not very useful. What fails is that there are too many vertices with too high weight. However, it is quite unlikely that a vertex with a high weight is chosen. Indeed, as argued in (3.2.5), when starting from a vertex with weight w, say, the probability that it is directly connected to a vertex having weight a_n is at most

$$\sum_{j:\,w_j\ge y}\frac{ww_j}{\ell_n}=w[1-F_n^*](y), \tag{3.3.27}$$

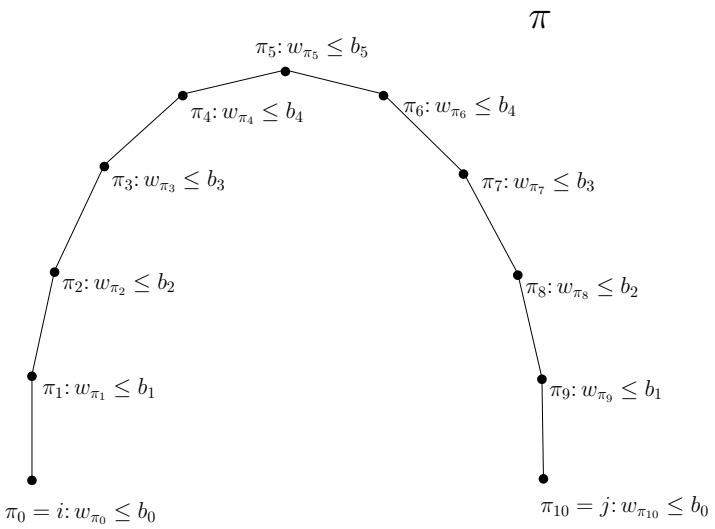

Figure 3.2: A 10-step good path connecting i and j and the upper bounds on the weight of its vertices. The height of a vertex is high for vertices with large weights.

which is small when y is too large. On the other hand, the main contribution to ν_n comes from vertices having maximal weight of the order $n^{1/(\tau-1)}$. This problem is resolved by a suitable truncation argument on the weights of the vertices in the occupied paths, which effectively removes these high-weight vertices. Therefore, instead of obtaining $\nu_n = \sum_{s\in[n]} w_s^2/\ell_n$, we obtain a partial sum of this restricted to vertices having a relatively small weight. Effectively, this means that we split the space of all paths into *good paths*, i.e., paths that avoid vertices with too large weight, and *bad paths*, which are paths that jump to vertices with too high weight.

We now present the details for this argument. We again start from

$$\mathbb{P}(H_n \le k_n) = \frac{1}{n^2} \sum_{i,j\in[n]} \mathbb{P}(\mathrm{dist}_{\mathrm{NR}_n(\boldsymbol{w})}(i,j) \le k_n). \tag{3.3.28}$$

When $\mathrm{dist}_{\mathrm{NR}_n(\boldsymbol{w})}(i,j) \le k_n$, there exists an occupied path $\pi \in \mathcal{P}_k(i,j)$ for some $k \le k_n$.

We fix an *increasing* sequence of numbers $(b_l)_{l=0}^{\infty}$ that serve as truncation values for the weights of vertices along our occupied path. We determine the precise values of $(b_l)_{l=0}^{\infty}$ below. We say that a path $\pi \in \mathcal{P}_k(i,j)$ is *good* when $w_{\pi_l} \le b_l \wedge b_{k-l}$ for every $l = 0,\dots,k$, and bad otherwise. The condition $w_{\pi_l} \le b_l \wedge b_{k-l}$ for every $l = 0,\dots,k$ is equivalent to the statement that $w_{\pi_l} \le b_l$ for $l \le \lceil k/2 \rceil$, while $w_{\pi_l} \le b_{k-l}$ for $\lceil k/2 \rceil < l \le k$. Thus, b_l provides an upper bound on the weight of the l^{th} vertex and the $(k-l)^{\text{th}}$ vertex of the occupied path, ensuring that the weights occurring in the occupied path can not be too large. See Figure 3.3.2 for a description of a good path and the bounds on the weight of its vertices.

Let $\mathcal{GP}_k(i,j)$ be the set of good paths in $\mathcal{P}_k(i,j)$. Let

$$\mathcal{E}_k(i,j) = \{\exists \pi \in \mathcal{GP}_k(i,j) \colon \pi \text{ occupied}\} \tag{3.3.29}$$

denote the event that there exists a good path of length k.

When $\mathrm{dist}_{\mathrm{NR}_n(\boldsymbol{w})}(i,j) \le k_n$, but there does not exist a $k \le k_n$ and a good occupied path $\pi \in \mathcal{GP}_k(i,j)$, then either there exists an $l \le \lceil k/2 \rceil$ such that $w_{\pi_s} \le b_s$ for every $s < l$, while $w_{\pi_l} > b_l$, or there exists an $l \le \lceil k/2 \rceil$ such that $w_{\pi_{k-s}} \le b_{k-s}$ for every $s < l$, while $w_{\pi_{k-l}} > b_{k-l}$. Let $\mathcal{P}_k(i) = \cup_{l \in [n]} \mathcal{P}_k(i,l)$ denote the set of all paths of length k from i, and let

$$\mathcal{BP}_k(i) = \{\pi \in \mathcal{P}_k(i) \colon w_{\pi_l} > b_l, w_{\pi_s} \le b_s \forall s < l\} \tag{3.3.30}$$

denote the set of *bad paths* of length k, i.e., those $\pi \in \mathcal{P}_k(i)$ that are not in $\mathcal{GP}_k(i, \pi_k)$. Let $\mathcal{F}_l(i)$ be the event that there exists a bad path of length l starting from i, i.e.,

$$\mathcal{F}_l(i) = \{\exists \pi \in \mathcal{BP}_l(i) \colon \pi \text{ occupied}\}. \tag{3.3.31}$$

Then, since $\mathrm{dist}_{\mathrm{NR}_n(\boldsymbol{w})}(i,j) \le k_n$ implies that there either is a good path or a bad path,

$$\{\mathrm{dist}_{\mathrm{NR}_n(\boldsymbol{w})}(i,j) \le k_n\} \subseteq \bigcup_{k \le k_n} \Big(\mathcal{F}_k(i) \cup \mathcal{F}_k(j) \cup \mathcal{E}_k(i,j) \Big), \tag{3.3.32}$$

so that, by Boole's inequality,

$$\mathbb{P}(\mathrm{dist}_{\mathrm{NR}_n(\boldsymbol{w})}(i,j) \le k_n) \le \sum_{k=0}^{k_n} \big[\mathbb{P}(\mathcal{F}_k(i)) + \mathbb{P}(\mathcal{F}_k(j)) + \mathbb{P}(\mathcal{E}_k(i,j)) \big]. \tag{3.3.33}$$

In order to estimate the probabilities $\mathbb{P}(\mathcal{F}_k(i))$ and $\mathbb{P}(\mathcal{E}_k(i,j))$, we introduce some notation. For $b \ge 0$, let

$$\nu_n(b) = \frac{1}{\ell_n} \sum_{i \in [n]} w_i^2 \mathbb{1}_{\{w_i \le b\}}, \tag{3.3.34}$$

be the restriction of ν_n to vertices with weights at most b, and let

$$F_n^*(x) = \frac{1}{\ell_n} \sum_{i \in [n]} w_i \mathbb{1}_{\{w_i \le x\}} \tag{3.3.35}$$

be the distribution function of W_n^*, the size-biased version of W_n. The following lemma gives bounds on $\mathbb{P}(\mathcal{F}_k(i))$ and $\mathbb{P}(\mathcal{E}_k(i,j))$:

Lemma 3.8 (Truncated path probabilities)**.** *For every $k \ge 1$, $(b_l)_{l \ge 0}$ with $b_l \ge 0$ and $l \mapsto b_l$ non-decreasing,*

$$\mathbb{P}(\mathcal{F}_k(i)) \le w_i [1 - F_n^*](b_k) \prod_{l=1}^{k-1} \nu_n(b_l), \tag{3.3.36}$$

and

$$\mathbb{P}(\mathcal{E}_k(i,j)) \le \frac{w_i w_j}{\ell_n} \prod_{l=1}^{k-1} \nu_n(b_l \wedge b_{k-l}). \tag{3.3.37}$$

When $b_l = \infty$ for each l, the bound in (3.3.37) equals that obtained in (3.3.8).

Proof. We start by proving (3.3.36). By Boole's inequality,

$$\mathbb{P}(\mathcal{F}_k(i)) = \mathbb{P}(\exists \pi \in \mathcal{BP}_l(i)\colon \pi \text{ occupied}) \le \sum_{\pi \in \mathcal{BP}_l(i)} \mathbb{P}(\pi \text{ occupied}). \tag{3.3.38}$$

By (3.3.7), (3.3.34) and (3.3.35),

$$\begin{aligned}\mathbb{P}(\mathcal{F}_k(i)) &\le \sum_{\pi \in \mathcal{BP}_l(i)} \frac{w_i w_{\pi_k}}{\ell_n} \prod_{l=1}^{k-1} w_{\pi_l}^2/\ell_n \\ &\le w_i \sum_{\pi_k\colon w_{\pi_k} \ge b_k} \frac{w_{\pi_k}}{\ell_n} \times \prod_{l=1}^{k-1} \sum_{\pi_l\colon w_{\pi_l} \le b_l} w_{\pi_l}^2/\ell_n \\ &= w_i [1 - F_n^*](b_k) \prod_{l=1}^{k} \nu_n(b_l).\end{aligned} \tag{3.3.39}$$

The proof of (3.3.37) is similar. Indeed, by (3.3.7),

$$\mathbb{P}(\mathcal{E}_k(i,j)) \le \sum_{\pi \in \mathcal{GP}_k(i,j)} \frac{w_i w_j}{\ell_n} \prod_{l=1}^{k-1} w_{\pi_l}^2/\ell_n \le \frac{w_i w_j}{\ell_n} \prod_{l=1}^{k-1} \nu(b_l \wedge b_{k-l}). \tag{3.3.40}$$

Now follow the steps in the proof of (3.3.36). □

Exercise 3.20 (Distance between fixed vertices)**.** *Show that* (3.3.33) *and Lemma 3.8 imply that for all* $a, b \in [n]$ *with* $a \neq b$,

$$\begin{aligned}\mathbb{P}(\mathrm{dist}_{\mathrm{NR}_n(\boldsymbol{w})}(a,b) \le k_n) \le \frac{w_a w_b}{\ell_n} \sum_{k=1}^{k_n} \prod_{l=1}^{k-1} \nu_n(b_l \wedge b_{k-l}) \\ + (w_a + w_b) \sum_{k=1}^{k^*} [1 - F_n^*](b_k) \prod_{l=1}^{k} \nu_n(b_l).\end{aligned} \tag{3.3.41}$$

We continue by proving upper bounds on $[1 - F_n^*](x)$ and $\nu_n(b)$:

Lemma 3.9 (Bounds on sums)**.** *Suppose that the weights* $\boldsymbol{w} = (w_i)_{i \in [n]}$ *satisfy [I, Condition 6.4(a)] and that there exist* $\tau \in (2,3)$ *and* c_2 *such that, for all* $x \ge 1$,

$$[1 - F_n](x) \le c_2 x^{-(\tau-1)}. \tag{3.3.42}$$

Then, there exists a constant $c_2^* > 0$ *such that, for all* $x \ge 1$,

$$[1 - F_n^*](x) \le c_2^* x^{-(\tau-2)}, \tag{3.3.43}$$

and there exists a $c_\nu > 0$ *such that for all* $b \ge 1$,

$$\nu_n(b) \le c_\nu b^{3-\tau}. \tag{3.3.44}$$

Proof. For (3.3.44), we bound

$$\nu_n(b) = \mathbb{E}[W_n^* \mathbb{1}_{\{W_n^* \le b\}}] \le = \int_0^b [1 - F_n^*](x)dx \le c_2^* \int_0^b x^{-(\tau-2)}dx = \frac{c_2^*}{3-\tau} b^{3-\tau}, \tag{3.3.45}$$

which implies (3.3.44) for $c_\nu = c_2^*/(3-\tau)$. Write out

$$[1 - F_n^*](x) = \frac{1}{\ell_n} \sum_{i \in [n]} w_i \mathbb{1}_{\{w_i > x\}} = \frac{\mathbb{E}[W_n \mathbb{1}_{\{W_n > x\}}]}{\mathbb{E}[W_n]}. \tag{3.3.46}$$

Now use the fact that, for any non-negative random variable X,

$$\mathbb{E}[X] = \int_0^\infty \mathbb{P}(X > x)dx. \tag{3.3.47}$$

Applying this to $X = W_n \mathbb{1}_{\{W_n > x\}}$ yields

$$[1 - F_n^*](x) = \frac{1}{\mathbb{E}[W_n]} \int_x^\infty [1 - F_n(y)]dy. \tag{3.3.48}$$

By (3.3.42),

$$[1 - F_n^*](x) \le \frac{1}{\mathbb{E}[W_n]} \int_x^\infty c_2 y^{-(\tau-1)}dy = \frac{c_2}{(\tau-2)\mathbb{E}[W_n]} x^{-(\tau-2)} \le c_2^* x^{-(\tau-2)}, \tag{3.3.49}$$

when $c_2^* = c_2/[(\tau-2)(\mathbb{E}[W] - \varepsilon)$ and n is sufficiently large. □

With Lemmas 3.8 and 3.9 at hand, we are ready to choose $(b_l)_{l \ge 0}$ and to complete the proof of Theorem 3.7:

Proof of Theorem 3.7. Take $k_n = 2(1-\varepsilon)\log\log n/|\log(\tau-2)|$. By (3.3.28) and (3.3.32),

$$\mathbb{P}(H_n \le k_n) \le \frac{1}{n} + \sum_{k=1}^{k_n} \Big[\frac{2}{n} \sum_{i \in [n]} \mathbb{P}(\mathcal{F}_k(i)) + \frac{1}{n^2} \sum_{i,j \in [n]:\, i \ne j} \mathbb{P}(\mathcal{E}_k(i,j))\Big],$$

where the contribution $1/n$ is due to $i = j$ for which $\mathrm{dist}_{\mathrm{NR}_n(\boldsymbol{w})}(i,i) = 0$. We use Lemmas 3.8 and 3.9 to provide bounds on $\mathbb{P}(\mathcal{F}_k(i))$, $\mathbb{P}(\mathcal{F}_k(j))$ and $\mathbb{P}(\mathcal{E}_k(i,j))$. These bounds are quite similar.

We first describe how we choose the truncation values $(b_l)_{l=0}^\infty$ so that $[1-F_n^*](b_k)$ is so small that $\mathbb{P}(\mathcal{F}_k(i))$ is small, and, for this choice of $(b_l)_{l=0}^\infty$, we show that $\mathbb{P}(\mathcal{E}_k(i,j))$ is small. Intuitively, this means that it is quite unlikely that i or j is connected to a vertex at distance k with too high weight, i.e., having weight at least b_k. At the same time, it is also unlikely that there is a path $\pi \in \mathcal{P}_k(i,j)$ whose weights are all small, i.e., for which $w_{\pi_k} \le b_k$ for every $k \le k_n$, because k_n is to small.

By Lemma 3.8, we wish to choose b_k so that $\mathbb{P}(\mathcal{F}_k(i)) = [1-F_n^*](b_k)\prod_{l=0}^{k-1}\nu_n(b_l)$ is small. Below (3.2.5), it is argued that $b_k \approx b_{k-1}^{1/(\tau-2)}$. In order to make this probability

small, we will take b_k somewhat larger. We now present the details. We take $\delta \in (0, \tau-2)$ sufficiently small and let

$$a = 1/(\tau - 2 - \delta) > 1. \tag{3.3.50}$$

Take $b_0 = \mathrm{e}^A$ for some constant $A \geq 0$ sufficiently large and define $(b_l)_{l\geq 0}$ recursively by

$$b_l = b_{l-1}^a, \qquad \text{so that} \qquad b_l = b_0^{a^l} = \mathrm{e}^{A(\tau-2-\delta)^{-l}}. \tag{3.3.51}$$

We start from (3.3.33). By Lemma 3.8, we obtain an upper bound on $\mathbb{P}(\mathcal{F}_k(i))$ in terms of factors $\nu_n(b_l)$ and $[1-F_n^*](b_k)$, which are bounded in Lemma 3.9. We start by applying the bound on $\nu_n(b_l)$ to obtain

$$\begin{aligned}\prod_{l=1}^{k-1} \nu_n(b_l) &\leq \prod_{l=1}^{k-1} c_\nu b_l^{\tau-3} = c_\nu^k \mathrm{e}^{K(3-\tau)\sum_{l=1}^{k-1} a^l} \\ &\leq c_\nu^{k-1} \mathrm{e}^{K(3-\tau)a^k/(a-1)} = c_2^* w_i c_\nu^{k-1} b_k^{(3-\tau)/(a-1)}.\end{aligned} \tag{3.3.52}$$

Combining (3.3.52) with the bound on $[1-F_n^*](b_k)$ in Lemma 3.9 yields

$$\mathbb{P}(\mathcal{F}_k(i)) \leq c_2^* w_i c_\nu^k b_k^{-(\tau-2)+(3-\tau)/(a-1)}. \tag{3.3.53}$$

Since $3-\tau+\delta < 1$ when $\tau \in (2,3)$ and $\delta \in (0, \tau-2)$,

$$\begin{aligned}(\tau-2) - (3-\tau)/(a-1) &= (\tau-2) - (3-\tau)(\tau-2-\delta)/(3-\tau+\delta) \\ &= \delta/(3-\tau+\delta) > \delta,\end{aligned} \tag{3.3.54}$$

so that

$$\mathbb{P}(\mathcal{F}_k(i)) \leq c_2^* w_i c_\nu^k b_k^{-\delta}. \tag{3.3.55}$$

As a result, for each $\delta > 0$

$$\frac{1}{n}\sum_{i\in[n]}\sum_{k=0}^{k_n} \mathbb{P}(\mathcal{F}_k(i)) \leq c_2^* \frac{1}{n}\sum_{i\in[n]} w_i \mathbb{1}_{\{w_i>K\}} + \frac{1}{n}\sum_{i\in[n]} c_2^* w_i \sum_{k\geq 1} c_\nu^k b_k^{-\delta} \leq \varepsilon, \tag{3.3.56}$$

when we take $A = A(\delta, \varepsilon)$ sufficiently large.

Similarly, since $b_l \geq 1$, by (3.3.52),

$$\mathbb{P}(\mathcal{E}_k(i,j)) \leq \frac{w_i w_j}{\ell_n} \prod_{l=1}^{k-1} \nu_n(b_l \wedge b_{k-l}) \leq \frac{w_i w_j}{\ell_n} c_\nu^{k-1} b_{\lceil k/2\rceil}^{2(3-\tau)/(a-1)}, \tag{3.3.57}$$

so that, using further that $l \mapsto b_l$ is increasing,

$$\begin{aligned}\sum_{k=1}^{k_n} \frac{1}{n^2} \sum_{i,j\in[n]} \mathbb{P}(\mathcal{E}_k(i,j)) &\leq \frac{1}{n^2}\sum_{k=1}^{k_n}\sum_{i,j\in[n]} \frac{w_i w_j}{\ell_n} c_\nu^{k-1} b_{\lceil k/2\rceil}^{2(3-\tau)/(a-1)} \\ &\leq \frac{\ell_n}{n^2} k_n c_\nu^{k_n-1} b_{\lceil k_n/2\rceil}^{2(3-\tau)/(a-1)}.\end{aligned} \tag{3.3.58}$$

Recall that $k \le k_n = 2(1-\varepsilon)\log\log n/|\log(\tau-2)|$. Take $\delta = \delta(\varepsilon) > 0$ so small that $(\tau-2-\delta)^{-(k_n+1)/2} \le (\log n)^{1-\varepsilon/4}$. Then,

$$b_{\lceil k_n/2\rceil} \le \mathrm{e}^{A(\tau-2-\delta)^{-(k_n+1)/2}} \le \mathrm{e}^{A(\log n)^{1-\varepsilon/4}}, \tag{3.3.59}$$

and we conclude that

$$\sum_{k=1}^{k_n} \frac{1}{n^2} \sum_{i,j\in[n]} \mathbb{P}(\mathcal{E}_k(i,j)) \le \frac{\ell_n}{n^2} k_n c_\nu^{k_n} \exp\left(2A(3-\tau)(\log n)^{1-\varepsilon/4}\right) = o(1), \tag{3.3.60}$$

since $k_n = O(\log\log n)$ and $\ell_n/n^2 = \Theta(1/n)$. This completes the proof of Theorem 3.7. □

Exercise 3.21 (Lower bound on fluctuations*). *Adapt the proof of Theorem 3.7 to show that for every ε, we can find a constant $K = K(\varepsilon) > 0$ such that*

$$\mathbb{P}\big(H_n \le \frac{2\log\log n}{|\log(\tau-2)|} - K\big) \le \varepsilon. \tag{3.3.61}$$

Hint: choose $b_k = Lb_{k-1}^{1/(\tau-2)}$, where the constant $L >$ is chosen sufficiently large.

3.4 Branching process comparisons and the log log upper bound

In this section, we prove the log log upper bound on typical graph distances in the infinite variance case. In Section 3.4.1, we start by comparing the neighborhoods of vertices in $\mathrm{NR}_n(\boldsymbol{w})$ to branching processes, a technique that is crucial in the derivation of all our lower bounds. In Section 3.4.2, we use this comparison to prove the log log upper bound on typical distances when $\tau \in (2,3)$.

3.4.1 Comparison to branching processes

In this section, we describe a beautiful comparison of the neighborhoods of a uniformly chosen vertex in inhomogeneous random graphs, such as the generalized random graph, the Chung-Lu model and the Norros-Reittu model, and a marked branching process. This comparison is particularly pretty when considering the Norros-Reittu model, where there is an explicit stochastic domination result of these neighborhoods are bounded by a so-called *two-stage* branching process with a mixed Poisson offspring distribution. [1]

Stochastic domination of clusters by a branching process. We shall dominate the cluster of a vertex in the Norros-Reittu model by the total progeny of a two-stage branching processes with mixed Poisson offspring. This domination is such that we also control the difference, and makes the heuristic argument below Theorem 3.2 precise.

[1] In [54], the two-stage branching process is called a delayed branching process.

We now describe the cluster exploration of a uniformly chosen vertex $U \in [n]$. Define the *mark distribution* to be the random variable M with distribution

$$\mathbb{P}(M = m) = w_m/\ell_n, \qquad m \in [n]. \tag{3.4.1}$$

Let $(X_w)_w$ be a collection of independent random variables, where

(a) the number of children of the root $X_\varnothing$ has a mixed Poisson distribution with random parameter $w_{M_\varnothing}$, where $M_\varnothing$ is uniformly chosen in $[n]$;

(b) X_w has a mixed Poisson distribution with random parameter w_{M_w}, where $(M_w)_{w\neq\varnothing}$ are i.i.d. random marks with distribution (3.4.1) independently of $M_\varnothing$.

We call $(X_w, M_w)_w$ a *marked mixed-Poisson branching process* (MMPBP).

Clearly, $w_U = w_{M_\varnothing}$ has distribution W_n defined in (6.1.17), while the distribution of w_{M_w} for each w with $|w| \geq 1$ is i.i.d. with distribution w_M given by

$$\mathbb{P}(w_M \leq x) = \sum_{m=1}^{n} \mathbb{1}_{\{w_m \leq x\}} \mathbb{P}(M = m) = \frac{1}{\ell_n} \sum_{m=1}^{n} w_m \mathbb{1}_{\{w_m \leq x\}} = \mathbb{P}(W_n^* \leq x) = F_n^*(x), \tag{3.4.2}$$

where W_n^* is the *size-biased distribution* of W_n and F_n^* is defined in (3.3.35).

When we are only interested in *numbers of individuals*, then we obtain a two-stage branching process since the random variables $(X_w)_w$ are independent, and the random variables $(X_w)_{w\neq\varnothing}$ are i.i.d. However, in the sequel, we make explicit use of the marks $(M_w)_{w\neq\varnothing}$, as the complete information $(X_w, M_w)_w$ gives us a way to retrieve the cluster of the vertex $M_\varnothing$, something that would not be possible on the basis of $(X_w)_w$ only.

In order to define the cluster exploration in $\mathrm{NR}_n(\boldsymbol{w})$, we introduce a *thinning* that guarantees that we only inspect a vertex once. We think of M_w as being the vertex label in $\mathrm{NR}_n(\boldsymbol{w})$ of the tree vertex w, and $X_w = \mathsf{Poi}(w_{M_w})$ as its *potential number of children*. These potential children effectively become children when their marks correspond to vertices in $\mathrm{NR}_n(\boldsymbol{w})$ that have not yet appeared. The thinning ensures this. To describe the thinning, we set $\varnothing$ unthinned, and, for w with $w \neq \varnothing$, we thin w when either (i) one of the tree vertices on the (unique) path between the root $\varnothing$ and w has been thinned, or (ii) when $M_w = M_{w'}$ for some unthinned vertex $w' < w$. We now make the connection between the thinned marked mixed Poisson branching process and the cluster exploration precise:

Proposition 3.10 (Clusters as thinned marked branching processes)**.** *The cluster of a uniformly chosen vertex $\mathcal{C}(U)$ is equal in distribution to $\{M_w\colon w \text{ unthinned}\}$, the marks of unthinned vertices encountered in the marked mixed Poisson branching process up to the end of the exploration. Similarly, the set of vertices at graph distance k from U has the same distribution as*

$$\Big(\{M_w\colon w \text{ unthinned}, |w| = k\}\Big)_{k\geq 0}. \tag{3.4.3}$$

Proof. We prove the two statements simultaneously. By construction, the distribution of U is the same as that of $M_\varnothing$, the mark of the root of the marked mixed Poisson branching process. We continue by proving that the direct neighbors of the root $\varnothing$ agree in both constructions. In $\mathrm{NR}_n(\boldsymbol{w})$, the direct neighbors are equal to $\{j \in [n] \setminus \{l\} \colon I_{lj} = 1\}$, where $(I_{lj})_{j\in[n]\setminus\{l\}}$ are independent $\mathsf{Be}(p_{lj})$ random variables with $p_{lj} = 1 - \mathrm{e}^{-w_l w_k/\ell_n}$.

We now prove that the same is true for the marked mixed Poisson branching process. Conditionally on $M_\varnothing = l$, the root has a $\mathsf{Poi}(w_l)$ number of children, where these $\mathsf{Poi}(w_l)$ offspring receive i.i.d. marks. We make use of the following fundamental property of the Poisson distribution:

Lemma 3.11 (A Poisson number of multinomial trials)**.** *Let X have a Poisson distribution with parameter λ. Perform X multinomial trials, where the i^{th} outcome appears with probability p_i for some probabilities $(p_i)_{i=1}^k$. Let $(X_i)_{i=1}^k$, where X_i denotes the total number of outcomes i. Then $(X_i)_{i=1}^k$ is a sequence of* independent *Poisson random variables with parameters $(\lambda p_i)_{i=1}^k$.*

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

$$\begin{aligned}\mathbb{P}((X_i)_{i=1}^k = (x_i)_{i=1}^k) &= \mathbb{P}(X = x)\mathbb{P}((X_i)_{i=1}^k = (x_i)_{i=1}^k \mid X = x) \\ &= \mathrm{e}^{-\lambda}\frac{\lambda^x}{x!}\binom{x}{x_1, x_2, \dots, x_k} p_1^{x_1}\cdots p_k^{x_k} = \prod_{i=1}^k \mathrm{e}^{-\lambda x_i}\frac{\lambda^{x_i}}{(x_i)!}.\end{aligned} \tag{3.4.4}$$

□

By Lemma 3.11, the random vector $(X_{\varnothing,j})_{j\in[n]}$, where $X_{\varnothing,j}$ is the number of offspring of the root that receive mark j, is a vector of *independent* Poisson random variables with parameters $w_l w_j/\ell_n$. Due to the thinning, a mark occurs precisely when $X_{\varnothing,j} \geq 1$. Therefore, the mark j occurs, independently for all $j \in [n]$, with probability $1 - \mathrm{e}^{-w_l w_j/\ell_n} = p_{jk}^{(\mathrm{NR})}$. This proves that the set of marks of children of the root in the MMPBD has the same distribution as the set of neighbors of the chosen vertex in $\mathrm{NR}_n(\boldsymbol{w})$.

Next, we look at the number of new elements of $\mathcal{C}(U)$ neighboring the vertex which has received word w. First, condition on $M_w = l$, and assume that w is not thinned. Conditionally on $M_w = l$, the number of children of w in the MMPBP has distribution $\mathsf{Poi}(w_l)$. Each of these $\mathsf{Poi}(w_l)$ children receives an i.i.d. mark. Let $X_{w,j}$ denote the number of children of w that receive mark j.

By Lemma 3.11, $(X_{w,j})_{j\in[n]}$ is again a vector of independent Poisson random variables with parameters $w_l w_j/\ell_n$. Due to the thinning, a mark appears within the offspring of individual w precisely when $X_{w,j} \geq 1$, and these events are independent. In particular, for each j that has not appeared as the mark of an unthinned vertex, the probability that it occurs equals $1 - \mathrm{e}^{-w_j w_k/\ell_n} = p_{jk}^{(\mathrm{NR})}$, as required. □

The law of the branching process. For given weights $(w_i)_{i\in[n]}$, we now describe the distribution of the marked mixed Poisson branching process (MMPBP). Since the

marks are mutually *independent*, the marked Poisson process is a branching process if we ignore the information about the marks. The offspring distribution $p^{(n)}$ of Z_1, i.e., the first generation of $(Z_l)_{l\geq 0}$, is given by

$$p_k^{(n)} = \mathbb{P}\big(\mathsf{Poi}(w_V) = k\big) = \sum_{i\in[n]} \mathbb{P}\big(\mathsf{Poi}(w_i) = k \big| V = i\big)\mathbb{P}(V = i) = \frac{1}{n}\sum_{i\in[n]} \mathrm{e}^{-w_i}\frac{w_i^k}{k!}, \tag{3.4.5}$$

for $k \geq 0$, so that $p^{(n)}$ is a mixed Poisson distribution with mixing distribution W_n, i.e.,

$$p_k^{(n)} = \mathbb{E}\big[\mathrm{e}^{-W_n}\frac{W_n^k}{k!}\big]. \tag{3.4.6}$$

Recall that individuals in the second and further generations have a random mark distributed as an independent copy of M given by (3.4.1). Hence, if we denote the offspring distribution of the second and further generations by $g^{(n)}$, then we obtain, for all $k \geq 0$,

$$\begin{aligned} g_k^{(n)} &= \mathbb{P}\big(\mathsf{Poi}(w_M) = k\big) = \sum_{i\in[n]} \mathbb{P}\big(\mathsf{Poi}(w_i) = k \big| M = i\big)\mathbb{P}(M = i) \\ &= \frac{1}{\ell_n}\sum_{i\in[n]} \mathrm{e}^{-w_i}\frac{w_i^{k+1}}{k!}. \end{aligned} \tag{3.4.7}$$

We recognize the offspring distribution in (3.4.7) as a mixed Poisson distribution, where the mixing random variable now has distribution W_n^*:

$$g_k^{(n)} = \mathbb{E}\big[\mathrm{e}^{-W_n^*}\frac{(W_n^*)^k}{k!}\big]. \tag{3.4.8}$$

Together, (3.4.5) and (3.4.7) identify the distribution $(Z_l)_{l\geq 0}$ as a two-stage branching process, where (1) the first generation has a mixed Poisson distribution with mixing random variable w_V, where $V \in [n]$ is chosen uniformly at random, and where (2) the offspring distribution of the second and further generations has a mixed Poisson distribution with mixing random variable w_M, where $M \in [n]$ has distribution given by (3.4.1). This yields a stochastic upper bound on the neighborhoods of a uniformly chosen $V \in [n]$ in the Norros-Reittu model $\mathrm{NR}_n(\boldsymbol{w})$. In the case where $\boldsymbol{w}$ is *constant*, the above gives an interesting direct upper bound of $|\mathcal{C}(U)|$ in terms of a Poisson branching process:

Exercise 3.22 (Erdős-Rényi random graph)**.** *Prove that* $\mathrm{NR}_n(\boldsymbol{w}) = \mathrm{ER}_n(\lambda/n)$ *when* $\boldsymbol{w}$ *is constant with* $w_i = -n\log(1-\lambda/n)$.

Exercise 3.23 (Erdős-Rényi random graph (Cont.))**.** *Show that Exercise 3.22 together with Proposition 3.10 imply that* $|\mathcal{C}(U)| \preceq T^*$, *where* T^* *is the total progeny of a Poisson branching process with mean* $-n\log(1-\lambda/n)$ *offspring.*

The limiting two-stage branching process. In the previous section, we have described the neighborhood of a uniform in terms of a (thinned) two-stage branching process. This leads to the definitions of the probability mass functions $p^{(n)}$ and $g^{(n)}$ in (3.4.5) and (3.4.7). We now study what happens when $n \to \infty$.

Recall the two-stage branching processes with a mixed Poisson offspring introduced in Section 3.1. We proceed by proving that (3.1.2) and (3.1.3) are the limits of (3.4.5) and (3.4.7):

Lemma 3.12 (Weak convergence of MMPBP)**.** *Let the weights $\boldsymbol{w} = (w_i)_{i\in[n]}$ satisfy [I, Condition 6.4(a)]. Then, for every $k \geq 0$,*

$$\lim_{n\to\infty} p_k^{(n)} = p_k. \tag{3.4.9}$$

When, instead, the weights $\boldsymbol{w} = (w_i)_{i\in[n]}$ satisfy [I, Condition 6.4(a)-(b)] with $\mathbb{E}[W] < \infty$, where W has distribution function F, then, for all $k \geq 0$,

$$\lim_{n\to\infty} g_k^{(n)} = g_k. \tag{3.4.10}$$

Proof. By [I, Condition 6.4(a)], $W_n \xrightarrow{d} W$, where W has distribution function F. By (3.4.6), and the fact that, for each $k \geq 0$, the function $w \mapsto \mathrm{e}^{-w}w^k/k!$ is a bounded continuous function, we obtain from the continuous mapping theorem that $p_k^{(n)} \to p_k$.

For (3.4.10), we note that

$$g_k^{(n)} = \frac{(k+1)p_{k+1}^{(n)}}{\mathbb{E}[W_n]}. \tag{3.4.11}$$

By [I, Condition 6.4(b)], $\mathbb{E}[W_n] \to \mathbb{E}[W]$. By (3.4.9), also $p_{k+1}^{(n)} \to p_{k+1}$, so that, indeed, $g_k^{(n)} \to \frac{(k+1)p_{k+1}}{\mathbb{E}[W]} = g_k$. □

Branching process notation. We now define the generation sizes of our branching process. We define $Z_0 = 1$, and, for $m \geq 1$,

$$Z_m = \#\{w\colon |w| = m\}. \tag{3.4.12}$$

Since each individual in generation m has precisely one parent in generation $m-1$, we can alternatively write

$$Z_m = \sum_{w\colon |w|=m-1} X_w. \tag{3.4.13}$$

We see that $(Z_m)_{m\geq 0}$ has the same distribution as the generation sizes of a so-called *two-stage mixed Poisson branching process*, in which the root has $Z_1 \sim \mathsf{Poi}(w_U)$ children, where U is chosen uniformly in $[n]$, and all other individuals have offspring distribution given by $\mathsf{Poi}(w_M)$. Throughout the remainder of this chapter, we use $(Z_l)_{l\geq 0}$ denote the n-dependent two-stage marked mixed-Poisson branching process

(MMPBP) with offspring distributions $p^{(n)}$ in (3.4.6) for the root, and offspring distributions $g^{(n)}$ in (3.4.7) for all later generations. We also let $(\mathcal{Z}_l)_{l\geq 0}$ denote the limiting two-stage MMPBP with offspring distributions p in (3.1.2) for the root, and offspring distributions g in (3.1.3) for all later generations. We write $(\widetilde{Z}_m)_{m\geq 0}$ for the *thinned* MMPBP by $\widetilde{Z}_0 = 1$ and

$$\widetilde{Z}_m = \sum_{w\colon |w|=m} (1 - \mathbb{1}_{\{w \text{ thinned}\}}). \tag{3.4.14}$$

Thus, $\widetilde{Z}_m$ denotes the number of unthinned words w of generation $|w| = m$. Then, obviously, by the coupling between the two processes and the fact that the thinning only removes vertices,

$$\widetilde{Z}_m \leq Z_m \qquad \text{a.s.} \tag{3.4.15}$$

We are interested in *typical distances*, which is the graph distance between the vertices V_1 and V_2, where $V_1, V_2 \in [n]$ are chosen uniformly at random. Therefore, we sometimes use a superscript (1) or (2), and write, e.g., $\widetilde{Z}_m^{(1)}$, when we consider the constructions from vertex V_1 and V_2, respectively. A straightforward adaptation of Proposition 3.10 implies that, on the event that $H_n \geq 2m$,

$$\widetilde{Z}_m^{(1)} \leq Z_m^{(1)}, \quad \text{and} \quad \widetilde{Z}_m^{(2)} \leq Z_m^{(2)} \qquad \text{a.s.} \tag{3.4.16}$$

Indeed, when $H_n \geq 2m$, and conditionally on $\mathcal{N}_m(V_1)$, in the graph exploration from V_2, we must avoid the vertices in $\mathcal{N}_m(V_1)$, as well as thin the vertices in $(\mathcal{N}_l(V_2))_{l=0}^m$. This explains (3.4.16).

Corollary 3.13 (Weak convergence of neighborhoods of two vertices)**.** *Let the weights* $\boldsymbol{w} = (w_i)_{i\in[n]}$ *satisfy [I, Condition 6.4(a)-(b)]. Then, for each* $m \geq 1$,

$$\mathbb{P}(H_n \leq 2m) = o(1), \tag{3.4.17}$$

and

$$(\widetilde{Z}_m^{(1)}, \widetilde{Z}_m^{(2)}) \xrightarrow{d} (\mathcal{Z}_m^{(1)}, \mathcal{Z}_m^{(2)}), \tag{3.4.18}$$

where $(\mathcal{Z}_l^{(1)}, \mathcal{Z}_l^{(2)})_{l\geq 0}$ *are two independent limiting two-stage marked mixed-Poisson branching processes.*

Proof. We start by proving that $\widetilde{Z}_m^{(1)} \xrightarrow{d} \mathcal{Z}_m^{(1)}$. By Lemma 3.12, $Z_m^{(1)} \xrightarrow{d} \mathcal{Z}_m^{(1)}$, and, by (3.4.16), $\widetilde{Z}_m^{(1)} \leq Z_m^{(1)}$. Further, by Proposition 3.10, $\widetilde{Z}_m^{(1)} = Z_m^{(1)}$ unless one of the vertices in $Z_m^{(1)}$ has been thinned. Since $Z_m^{(1)} \xrightarrow{d} \mathcal{Z}_m^{(1)}$, which is a bounded random variable, whp we draw at most $Z_m^{(1)} \leq a_n$ marks for any $a_n \to \infty$. The probability that within these a_n draws, the mark i appears *twice*, thus causing the second individual to be thinned, is at most

$$\left(\frac{w_i}{\ell_n}\right)^2 \binom{a_n}{2} \leq \frac{(a_n w_i)^2}{\ell_n^2}, \tag{3.4.19}$$

since there at most $\binom{a_n}{2}$ pairs of times that mark i can be drawn. By Boole's inequality, the probability that an individual from the first a_n individuals is thinned is therefore at most

$$\frac{a_n^2}{\ell_n^2}\sum_{i\in[n]} w_i^2. \tag{3.4.20}$$

By [I, Exercise 6.3], [I, Condition 6.4(a) and (b)] imply that $\max_{i\in[n]} w_i = o(n)$, so that the probability that an individual from the first a_n individuals is thinned is at most $o(1)a_n^2$. Choosing $a_n \to \infty$ sufficiently slowly proves that $\mathbb{P}(\widetilde{Z}_m^{(1)} \neq Z_m^{(1)}) = o(1)$, completing the proof that $\widetilde{Z}_m^{(1)} \xrightarrow{d} \mathcal{Z}_m^{(1)}$.

Further, let $\mathcal{N}_{\le m}(U_1)$ denote the vertices in $\mathrm{NR}_n(\boldsymbol{w})$ that are at distance at most m from U_1. Similarly as in the proof of $\widetilde{Z}_m^{(1)} \xrightarrow{d} \mathcal{Z}_m^{(1)}$, we can also show that $\mathcal{N}_{\le m}(U_1)$ converges in distribution to the set of marks in the MMPBP up to generation m.

We continue show that $H_n > 2m$ occurs whp. We know that $\widetilde{Z}_{2m}^{(1)} \xrightarrow{d} \mathcal{Z}_{2m}^{(1)}$, which is a finite random variable. Now, $H_n \le 2m$ occurs precisely when U_2 is one of the $\mathcal{Z}_{2m}^{(1)}$ individuals, which, conditionally on $\mathcal{Z}_{2m}^{(1)}$, has probability $\mathcal{Z}_{2m}^{(1)}/n = o_{\mathbb{P}}(1)$. Therefore, $\mathbb{P}(H_n \le 2m) = o(1)$, as required. We condition on $\mathcal{N}_{\le m}(U_1)$ and on $H_n > 2m$. Then, the cluster of U_2 up to graph distance m is again an inhomogeneous random graph on $[n]\setminus\mathcal{N}_{\le m}(U_1)$ with edge probabilities given by $p_{ij} = 1 - \mathrm{e}^{-w_i w_j/\ell_n}$. Since $|\mathcal{N}_{\le m}(U_1)| = \widetilde{Z}_m^{(1)} \xrightarrow{d} \mathcal{Z}_m^{(2)}$, which is a finite random variable, it follows that, conditionally on $\mathcal{N}_{\le m}(U_1)$, the number of vertices in $[n]\setminus\mathcal{N}_{\le m}(U_1)$ at distance m from U_2 converges in distribution to $\mathcal{Z}_m^{(2)}$. Since this is true *conditionally on* $\mathcal{N}_{\le m}(U_1)$, and the limit is independent of $\mathcal{N}_{\le m}(U_1)$, also $\mathcal{Z}_m^{(1)}$ and $\mathcal{Z}_m^{(2)}$) are independent. This completes the proof. □

Exercise 3.24 (The diameter tends to infinity)**.** *Let the weights satisfy [I, Condition 6.4(a)-(b)]. Use (3.4.17) to show that the diameter of* $\mathrm{NR}_n(\boldsymbol{w})$ *tends to infinity in probability.*

3.4.2 A log log upper bound for $\tau \in (2,3)$

In this section, we prove a $\log\log n$ upper bound on the typical distance H_n using the comparison to branching processes obtained in the previous section. Throughout this section, we assume that there exist $\tau \in (2,3)$, $\alpha > 1/2$ and c_1 such that, uniformly in n and $x \le n^\alpha$,

$$[1 - F_n](x) \ge c_1 x^{-(\tau-1)}. \tag{3.4.21}$$

The bound in (3.4.21) corresponds to the lower bound in (3.2.2). The main result in this section is the following theorem:

Theorem 3.14 (A log log upper bound on typical distance for $\tau \in (2,3)$)**.** *Suppose that empirical distribution function F_n of the weights $\boldsymbol{w} = (w_i)_{i\in[n]}$ satisfies [I, Condition 6.4(a)-(b)] and (3.4.21). Then, for every $\varepsilon > 0$,*

$$\lim_{n\to\infty} \mathbb{P}\Big(H_n \le \frac{2(1+\varepsilon)\log\log n}{|\log(\tau-2)|} \mid H_n < \infty\Big) = 1. \tag{3.4.22}$$

Proof. The proof Theorem 3.14 is organized as follows. We start by showing that the giant-weight vertices, i.e., the vertices with extremely high weight, are all connected to one another. Thus, the giant-weight vertices form a complete graph. This is often referred to as a *clique* in the random graph community. In the second step, we show that connections from a vertex to the set of giant weight vertices occur at distance at most $\frac{(1+\varepsilon)\log\log n}{|\log(\tau-2)|}$. The latter is only true when the vertex is in the giant component, a fact we need to carefully into account. In the final step, we complete the proof of Theorem 3.14. We now start by defining the set of giant-weight vertices.

The giant-weight vertices form a clique. Recall the definition of $\alpha > 1/2$ in (3.4.21). Let

$$\mathrm{Giant}_n = \{i\colon w_i \geq n^\alpha\} \tag{3.4.23}$$

denote the set of vertices with giant weights. Let $A \subseteq [n]$. We say that A forms a *clique* when the edges a_1a_2 are occupied for all $a_1, a_2 \in A$. We continue by proving that, **whp**, Giant_n forms a clique:

Lemma 3.15 (High-weight vertices form clique)**.** *Under the conditions of Theorem 3.14,*

$$\mathbb{P}(\mathrm{Giant}_n \textit{ does not form clique}) \leq n^2 \mathrm{e}^{-n^{2\alpha}/\ell_n} \tag{3.4.24}$$

Proof. Let $g_1, g_2 \in \mathrm{Giant}_n$, so that $w_{g_1}, w_{g_2} \geq n^\alpha$. There are at most $|\mathrm{Giant}_n|^2 \leq n^2$ pairs of vertices in Giant_n, so that

$$\mathbb{P}(\mathrm{Giant}_n \text{ does not form clique}) \leq n^2 \max_{g_1,g_2\in\mathrm{Giant}_n} \mathbb{P}(g_1g_2 \text{ vacant}). \tag{3.4.25}$$

The edge g_1g_2 is vacant with probability

$$\mathbb{P}(g_1g_2 \text{ vacant}) = \mathrm{e}^{-w_{g_1}w_{g_2}/\ell_n} \leq \mathrm{e}^{-n^{2\alpha}/\ell_n}, \tag{3.4.26}$$

since $w_g \geq n^\alpha$ for every $g \in \mathrm{Giant}_n$. Multiplying out gives the result. □

Connections to Giant_n occur at $\log\log n$ distances. We next show that vertices that survive up to distance m have a high probability of connecting to Giant_n using a path of at most $(1+\varepsilon)\frac{\log\log n}{|\log(\tau-2)|}$ edges:

Proposition 3.16 (Connecting to Giant_n)**.** *Let $i \in [n]$ be such that $w_i > 1$. Under the conditions of Theorem 3.14, there exist $c, c_1^* > 0$ and $\eta > 0$ such that*

$$\mathbb{P}\Big(\mathrm{dist}_{\mathrm{NR}_n(\boldsymbol{w})}(i, \mathrm{Giant}_n) \geq (1+\varepsilon)\frac{\log\log n}{|\log(\tau-2)|}\Big) \leq c\mathrm{e}^{-c_1^* w_i^\eta}. \tag{3.4.27}$$

Consequently, with $\mathcal{N}_m(i)$ denoting the vertices at graph distance m from i, $\mathcal{N}_{\leq m}(i)$ the vertices at graph distance at most m from i, and $\mathcal{W}_m(i) = \sum_{k\in\mathcal{N}_m(i)} w_k$ denoting the weight of vertices in $\mathcal{N}_m(i)$,

$$\mathbb{P}\Big(\mathrm{dist}_{\mathrm{NR}_n(\boldsymbol{w})}(\mathcal{N}_m(i), \mathrm{Giant}_n) \geq (1+\varepsilon)\frac{\log\log n}{|\log(\tau-2)|} \mid \mathcal{N}_{\leq m}(i)\Big) \leq c\mathrm{e}^{-c_1^*\mathcal{W}_m(i)^\eta}. \tag{3.4.28}$$

Proof. We start by proving (3.4.27). The bound in (3.4.27) is trivial unless w_i is large. We let $x_0 = i$, and define, recursively,

$$x_\ell = \max\{j \in [n] \colon x_{\ell-1} j \text{ occupied}\}. \tag{3.4.29}$$

Thus, x_ℓ is maximal weight neighbor of $x_{\ell-1}$. We stop the above recursion when $w_{x_\ell} \geq n^\alpha$, since then $x_\ell \in \mathrm{Giant}_n$. Recall the heuristic below (3.2.5), which shows that a vertex with weight w is whp connected to a vertex with weight $w^{1/(\tau-2)}$. We now make this precise.

We take $a = 1/(\tau-2+\delta)$, where we choose $\delta > 0$ so small that $a > 1$. By (3.3.35),

$$\mathbb{P}(w_{x_{\ell+1}} < w_{x_\ell}^a \mid (x_s)_{s\leq \ell}) = \mathrm{e}^{-w_{x_\ell} \sum_{l\colon w_l \geq w_{x_\ell}^a} w_l/\ell_n} = \mathrm{e}^{-w_{x_\ell}[1-F_n^*](w_{x_\ell}^a)}. \tag{3.4.30}$$

We split the argument depending on whether $w_{x_\ell}^a \leq n^\alpha$ or not. Firstly, when $w_{x_\ell}^a \leq n^\alpha$, by (3.4.21) and uniformly for $x \leq n^\alpha$,

$$[1-F_n^*](x) \geq \frac{xn}{\ell_n}[1-F_n](x) \geq c_1^* x^{-(\tau-2)}, \tag{3.4.31}$$

where, for n large enough, we can take $c_1^* = c_1/(2\mathbb{E}[W])$. Therefore,

$$\mathbb{P}(w_{x_{\ell+1}} < w_{x_\ell}^a \mid (x_s)_{s\leq \ell}) \leq \mathrm{e}^{-c_1^* w_{x_\ell}^{1-(\tau-2)a}} \leq \mathrm{e}^{-c_1^* w_{x_\ell}^\delta}, \tag{3.4.32}$$

since $a = 1/(\tau-2+\delta) > 1$ so that $1-(\tau-2)a = a\delta > \delta$.

Secondly, when $w_{x_\ell}^a > n^\alpha$, but $w_{x_\ell} < n^\alpha$, we can use (3.4.31) for $x = n^\alpha$ to obtain

$$\mathbb{P}(w_{x_{\ell+1}} < n^\alpha \mid (x_s)_{s\leq \ell}) \leq \mathrm{e}^{-c_1^* w_{x_\ell} n^{\alpha(\tau-2)}} \leq \mathrm{e}^{-c_1^* n^{\alpha[1-(\tau-2)]/a}} \leq \mathrm{e}^{-c_1^* n^{\alpha\delta/a}}. \tag{3.4.33}$$

Therefore, in both cases, and with $\eta = \alpha\delta/a$,

$$\mathbb{P}(w_{x_{\ell+1}} < (n^\alpha \wedge w_{x_\ell}^a) \mid (x_s)_{s\leq \ell}) \mathrm{e}^{-c_1^* w_{x_\ell}^\eta}. \tag{3.4.34}$$

As a result, when x_ℓ is such that w_{x_ℓ} is quite large, whp, $w_{x_{\ell+1}} \geq w_{x_\ell}$. This produces, whp, a short path to Giant_n. We now investigate the properties of this path.

Let the recursion stop at some integer time k. The key observation is that when this occurs, we must have that $w_{x_{\ell+1}} > w_{x_\ell}^a$ for each $\ell \leq k-1$ where k is such that $w_{x_{k-1}} \in [n^{\alpha/a}, n^\alpha]$, and at the same time $w_{x_k} \geq n^\alpha$. Then, the following holds:

(1) $w_{x_\ell} \geq w_{x_0}^{a^\ell} = w_i^{a^\ell}$ for every $\ell \leq k-1$,

(2) $\mathrm{dist}_{\mathrm{NR}_n(\boldsymbol{w})}(i, \mathrm{Giant}_n) \leq k$.

By (1), $w_{x_{k-1}} \geq w_i^{a^{k-1}}$, and $w_{x_{k-1}} \in [n^{\alpha/a}, n^\alpha]$. Therefore, $w_i^{a^{k-1}} \leq n^\alpha$, which, in turn, implies that

$$a^{k-1} \leq \alpha \log n, \qquad \text{or} \qquad k-1 \leq (\log\log n + \log\alpha)(\log a). \tag{3.4.35}$$

Let $k_n = (1+\varepsilon)\frac{\log\log n}{|\log(\tau-2)|}$. By (1) and (2), when $\mathrm{dist}_{\mathrm{NR}_n(\boldsymbol{w})}(i, \mathrm{Giant}_n) > k_n$ occurs, then there must exist an $\ell \le k_n$ such that $w_{x_{\ell+1}} \le n^\alpha \wedge w_{x_\ell}^a$. We conclude that

$$\begin{aligned}\mathbb{P}\Big(\mathrm{dist}_{\mathrm{NR}_n(\boldsymbol{w})}(i, \mathrm{Giant}_n) \ge k_n\Big) &\le \sum_{\ell=0}^{k_n} \mathbb{P}(w_{x_{\ell+1}} \le w_{x_\ell}^a) \\ &\le \sum_{\ell=0}^{k_n} \mathbb{E}[\mathbb{P}(w_{x_{\ell+1}} \le w_{x_\ell}^a \mid (x_s)_{s\le\ell})] \\ &\le \sum_{\ell=0}^{k_n} \mathbb{E}[\mathrm{e}^{-c_1^* w_{x_\ell}^\eta}] \le \sum_{\ell=0}^{k_n} \mathrm{e}^{-c_1^* w_i^{\delta a^\ell}} \le c\mathrm{e}^{-c_1^* w_i^\eta}.\end{aligned} \tag{3.4.36}$$

The proof of (3.4.28) is similar, by conditioning on $\mathcal{N}_m(i)$ and by noting that we can interpret $\mathcal{N}_m(i)$ as a single vertex having weight $\mathcal{W}_m(i) = \sum_{k\in\mathcal{N}_m(i)} w_k$. □

Completion of the proof of Theorem 3.14. To prove the upper bound in Theorem 3.14, for $\varepsilon \in (0,1)$, we take

$$k_n = (1+\varepsilon)\frac{\log\log n}{|\log(\tau-2)|}, \tag{3.4.37}$$

so that it suffices to show, for every $\varepsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}(H_n \le 2k_n \mid H_n < \infty) = 1. \tag{3.4.38}$$

Since

$$\mathbb{P}(H_n \le 2k_n \mid H_n < \infty) = \frac{\mathbb{P}(H_n \le 2k_n)}{\mathbb{P}(H_n < \infty)}, \tag{3.4.39}$$

this follows from the bounds

$$\liminf_{n\to\infty} \mathbb{P}(H_n < \infty) \le \zeta^2, \tag{3.4.40}$$

$$\limsup_{n\to\infty} \mathbb{P}(H_n \le 2k_n) \ge \zeta^2, \tag{3.4.41}$$

with $\zeta > 0$ the survival probability of the underlying branching process approximation to the neighborhood shells of $\mathrm{NR}_n(\boldsymbol{w})$. For (3.4.40), we split, for some $m \ge 1$,

$$\mathbb{P}(H_n < \infty) \le \mathbb{P}(H_n \le 2m) + \mathbb{P}(\widetilde{Z}_m^{(1)} > 0, \widetilde{Z}_m^{(2)} > 0, H_n > 2m). \tag{3.4.42}$$

By (3.4.17) in Corollary 3.13, $\mathbb{P}(H_n \le 2m) = o(1)$, and, by (3.4.18) in Corollary 3.13,

$$\lim_{n\to\infty} \mathbb{P}(\widetilde{Z}_m^{(1)} > 0, \widetilde{Z}_m^{(2)} > 0, H_n > 2m) = \mathbb{P}(\mathcal{Z}_m > 0)^2, \tag{3.4.43}$$

which converges to ζ^2 when $m \to \infty$. This proves (3.4.40).

To prove (3.4.41), we fix $m \ge 1$ and write

$$\begin{aligned}\mathbb{P}(2m < H_n \le 2k_n) &\ge \mathbb{P}\big(\mathrm{dist}_{\mathrm{NR}_n(\boldsymbol{w})}(V_i, \mathrm{Giant}_n) \le k_n, i = 1,2, H_n > 2m\big) \\ &\ge \mathbb{P}(\widetilde{Z}_m^{(1)} > 0, \widetilde{Z}_m^{(2)} > 0, H_n > 2m) \\ &\quad - 2\mathbb{P}\big(\mathrm{dist}_{\mathrm{NR}_n(\boldsymbol{w})}(V_1, \mathrm{Giant}_n) < k_n, \widetilde{Z}_m^{(1)} > 0\big).\end{aligned} \tag{3.4.44}$$

By (3.4.43), the first term converges to ζ_m^2, which in turn converges to ζ^2 when $m \to \infty$.

For the second term, we condition on $\mathcal{N}_{\le m}(V_1), \mathcal{N}_{\le m}(V_2)$, and use that $\widetilde{Z}_m^{(1)}$ is measurable w.r.t. $\mathcal{N}_{\le m}(V_1)$ to obtain

$$\begin{aligned}&\mathbb{P}\big(\mathrm{dist}_{\mathrm{NR}_n(\boldsymbol{w})}(V_1, \mathrm{Giant}_n) < k_n, \widetilde{Z}_m^{(1)} > 0\big) \\ &\quad = \mathbb{E}\Big[\mathbb{P}\big(\mathrm{dist}_{\mathrm{NR}_n(\boldsymbol{w})}(V_1, \mathrm{Giant}_n) > k_n \mid \mathcal{N}_{\le m}(V_1)\big)\mathbb{1}_{\{\widetilde{Z}_m^{(1)}>0\}}\Big].\end{aligned} \tag{3.4.45}$$

By Proposition 3.16,

$$\mathbb{P}\big(\mathrm{dist}_{\mathrm{NR}_n(\boldsymbol{w})}(V_1, \mathrm{Giant}_n) > k_n \mid \mathcal{N}_{\le m}(V_1)\big) \le c\mathrm{e}^{-c_1^* \mathcal{W}_m(V_1)^\eta}. \tag{3.4.46}$$

By Lemma 3.12 and the fact that $(w_k)_{k\in\mathcal{N}_m(V_1)}$ are i.i.d. copies of the random variable W_n, conditionally on being distinct and different from $(w_l)_{l\in\mathcal{N}_{\le m-1}(V_1)}$,

$$\mathcal{W}_m(V_1) \xrightarrow{d} \sum_{i=1}^{\mathcal{Z}_m^{(1)}} W(i), \tag{3.4.47}$$

where $(W(i))_{i\ge 1}$ are i.i.d. random variables with distribution function F. Therefore,

$$\mathcal{W}_m(V_1) \xrightarrow{\mathbb{P}} \infty \tag{3.4.48}$$

when first $n \to \infty$ followed by $m \to \infty$, and we use that $\mathcal{Z}_m^{(1)} \xrightarrow{\mathbb{P}} \infty$ since $\mathcal{Z}_m^{(1)} > 0$. As a result,

$$\mathbb{P}\big(\mathrm{dist}_{\mathrm{NR}_n(\boldsymbol{w})}(V_1, \mathrm{Giant}_n) > k_n \mid \mathcal{N}_{\le m}(V_1)\big)\mathbb{1}_{\{\widetilde{Z}_m^{(1)}>0\}} \xrightarrow{\mathbb{P}} 0, \tag{3.4.49}$$

which by Lebesgues Dominated Convergence Theorem (Theorem A.17) implies that

$$\mathbb{E}\Big[\mathrm{e}^{-c_1^* \mathcal{W}_m(V_1)^\eta}\mathbb{1}_{\{\widetilde{Z}_m^{(1)}>0\}}\Big] \to 0, \tag{3.4.50}$$

when first $n \to \infty$ followed by $m \to \infty$. This proves (3.4.41), and thus completes the proof of the upper bound in Theorem 3.4. □

3.5 Path counting and the log upper bound for $\tau > 3$

3.5.1 Path counting techniques

In this section, we study path counting techniques in the context of inhomogeneous random graphs (IRGs). We generalize the setting somewhat, and consider an IRG on the vertices $\mathcal{I}$ with edge probabilities $p_{ij} = u_i u_j$, for some weights $(u_i)_{i\in\mathcal{I}}$. We obtain $\mathrm{CL}_n(\boldsymbol{w})$ by taking $u_i = w_i/\sqrt{\ell_n}$ and $\mathcal{I} = [n]$. Since the $\mathrm{NR}_n(\boldsymbol{w})$ random graph is closely related to $\mathrm{CL}_n(\boldsymbol{w})$, this suffices for our purposes.

For $a, b \in \mathcal{I}$ and $k \ge 1$, let

$$N_k(a, b) = \#\{\pi \in \mathcal{P}_k(a, b) \colon \pi \text{ occupied}\} \tag{3.5.1}$$

denote the number of paths of length k between the vertices a and b. Let

$$n_k(a,b) = \mathbb{E}[N_k(a,b)] \tag{3.5.2}$$

denote the expected number of occupied paths of length k connecting a and b. Define

$$\bar{n}_k(a,b) = u_a u_b \Big(\sum_{i \in \mathcal{I} \setminus \{a,b\}} u_i^2 \Big)^{k-1}, \qquad \underline{n}_k(a,b) = u_a u_b \Big(\sum_{i \in \mathcal{I}_{a,b,k}} u_i^2 \Big)^{k-1}, \tag{3.5.3}$$

where $\mathcal{I}_{a,b,k}$ is the subset of $\mathcal{I}$ in which a and b, as well as the $k+2$ indexes with highest weights have been removed. In Section 3.3, we have implicitly proved an upper bound on $\mathbb{E}[N_k(a,b)]$ of the form

$$n_k(a,b) \le \bar{n}_k(a,b). \tag{3.5.4}$$

Exercise 3.25 (Upper bound on the expected number of paths)**.** *Prove* (3.5.4) *for an inhomogeneous random graph with vertex set* $\mathcal{I}$ *and with edge probabilities* $p_{ij} = u_i u_j$ *for every* $i, j \in \mathcal{I}$.

Let

$$\nu_{\mathcal{I}} = \sum_{i \in \mathcal{I}} u_i^2, \qquad \gamma_{\mathcal{I}} = \sum_{i \in \mathcal{I}} u_i^3 \tag{3.5.5}$$

denote the sums of squares and third powers of $(u_i)_{i \in \mathcal{I}}$, respectively. Our aim is to show that whp paths of length k exist between the vertices a and b. We do this by applying a second moment method on $N_k(a,b)$, for which we need a lower bound on $\mathbb{E}[N_k(a,b)]$ and an upper bound on $\mathrm{Var}(N_k(a,b))$, which are such that $\mathrm{Var}(N_k(a,b)) = o(\mathbb{E}[N_k(a,b)]^2)$ (recall Theorem 2.16. We prove lower bounds on $\mathbb{E}[N_k(a,b)]$ and upper bounds on $\mathrm{Var}(N_k(a,b))$ in the following proposition:

Proposition 3.17 (Variance of numbers of paths)**.** *For any* $k \ge 1$, $a, b \in \mathcal{I}$ *and* $(u_i)_{i \in \mathcal{I}}$,

$$\mathbb{E}[N_k(a,b)] \ge \underline{n}_k(a,b), \tag{3.5.6}$$

while, assuming that $\nu_{\mathcal{I}} > 1$,

$$\mathrm{Var}(N_k(a,b)) \le n_k(a,b) + \bar{n}_k(a,b)^2 \Big(\frac{\gamma_{\mathcal{I}} \nu_{\mathcal{I}}^2}{\nu_{\mathcal{I}} - 1} \Big(\frac{1}{u_a} + \frac{1}{u_b} \Big) + \frac{\gamma_{\mathcal{I}}^2 \nu_{\mathcal{I}}}{u_a u_b (\nu_{\mathcal{I}} - 1)^2} + e_k \Big), \tag{3.5.7}$$

where

$$e_k = \Big(1 + \frac{\gamma_{\mathcal{I}}}{u_a \nu_{\mathcal{I}}}\Big)\Big(1 + \frac{\gamma_{\mathcal{I}}}{u_b \nu_{\mathcal{I}}}\Big) \frac{\nu_{\mathcal{I}}}{\nu_{\mathcal{I}} - 1} \big(\mathrm{e}^{2k^3 \gamma_{\mathcal{I}}^2 / \nu_{\mathcal{I}}^3} - 1 \big).$$

We apply Proposition 3.17 in cases where $\mathbb{E}[N_k(a,b)] = n_k(a,b) \to \infty$, by taking $\mathcal{I}$ is a large subset of $[n]$ and $u_i = w_i / \sqrt{\ell_n}$. In this case, $\nu_{\mathcal{I}} \approx \nu_n \approx \nu > 1$. In our applications of Proposition 3.17, the ratio $\bar{n}_k(a,b)/\underline{n}_k(a,b)$ will be bounded, and $k^3 \gamma_{\mathcal{I}}^2 / \nu_{\mathcal{I}}^3 = o(1)$, so that the last term is an error term. The starting and end vertices

$a, b \in \mathcal{I}$ will correspond to a union of vertices in $[n]$ of quite large size. As a result, $\gamma_{\mathcal{I}}/u_a$ and $\gamma_{\mathcal{I}}/u_a$ are typically small, so that

$$\frac{\mathrm{Var}(N_k(a,b))}{\mathbb{E}[N_k(a,b)]^2} \approx \frac{\gamma_{\mathcal{I}}\nu_{\mathcal{I}}^2}{\nu_{\mathcal{I}}-1}\Big(\frac{1}{u_a}+\frac{1}{u_b}\Big) + \frac{\gamma_{\mathcal{I}}^2\nu_{\mathcal{I}}}{u_a u_b(\nu_{\mathcal{I}}-1)^2} \tag{3.5.8}$$

is small. The choice of a, b and $\mathcal{I}$ is quite delicate, which explains why we formulate Proposition 3.17 in such generality.

Proof. We note that $N_k(a,b)$ is a sum of indicators

$$N_k(a,b) = \sum_{\pi\in\mathcal{P}_k(a,b)} \mathbb{1}_{\{\pi \text{ occupied}\}}. \tag{3.5.9}$$

As a result,

$$\begin{aligned}\mathbb{E}[N_k(a,b)] &= \sum_{\pi\in\mathcal{P}_k(a,b)} \mathbb{P}(\pi \text{ occupied}) = \sum_{\pi\in\mathcal{P}_k(a,b)} \prod_{l=0}^{k} u_{\pi_l}u_{\pi_{l+1}} \\ &= u_{\pi_0}u_{\pi_k} \sum_{\pi\in\mathcal{P}_k(a,b)} \prod_{l=1}^{k-1} u_{\pi_l}^2.\end{aligned} \tag{3.5.10}$$

For $\pi \in \mathcal{P}_k(a,b)$, $\pi_0 = a, \pi_k = b$. Further,

$$\sum_{\pi\in\mathcal{P}_k(a,b)} \prod_{l=1}^{k-1} u_{\pi_l}^2 = \sum_{i_1,\ldots,i_{k-1}\in\mathcal{I}\setminus\{a,b\}}^{*} \prod_{l=1}^{k-1} u_{\pi_l}^2, \tag{3.5.11}$$

where we recall that $\sum^*_{i_1,\ldots,i_r\in\mathcal{I}}$ denotes a sum over *distinct* indices. Each sum over i_j yields a factor that is at least $\sum_{i\in\mathcal{I}_{a,b,k}} u_i^2$, which proves (3.5.6).

To compute $\mathrm{Var}(N_k(a,b))$, we again start from (3.5.9), which yields

$$\mathrm{Var}(N_k(a,b)) = \sum_{\pi,\rho\in\mathcal{P}_k(a,b)} \big[\mathbb{P}(\pi,\rho \text{ occupied}) - \mathbb{P}(\pi \text{ occupied})\mathbb{P}(\rho \text{ occupied})\big]. \tag{3.5.12}$$

For π, ρ, we denote by $\pi \cap \rho$ the edges the paths π and ρ have in common. The occupation statuses of π and ρ are independent precisely when $\pi \cap \rho = \varnothing$, so that

$$\mathrm{Var}(N_k(a,b)) \le \sum_{\substack{\pi,\,\rho\,\in\,\mathcal{P}_k(a,b)\\ \pi\cap\rho\neq\varnothing}} \mathbb{P}(\pi,\rho \text{ occupied}). \tag{3.5.13}$$

Define $\rho \setminus \pi$ to be the edges in ρ that are *not* part of π, so that

$$\begin{aligned}\mathbb{P}(\pi,\rho \text{ occupied}) &= \mathbb{P}(\pi \text{ occupied})\mathbb{P}(\rho \text{ occupied} \mid \pi \text{ occupied}) \\ &= \prod_{l=0}^{k} u_{\pi_l}u_{\pi_{l+1}} \prod_{e\in\rho\setminus\pi} u_{\bar{e}}u_{\underline{e}},\end{aligned} \tag{3.5.14}$$

where, for an edge $e = \{x, y\}$, we write $\bar{e} = x, \underline{e} = y$. When $\pi = \rho$, then

$$\mathbb{P}(\pi, \rho \text{ occupied}) = \mathbb{P}(\pi \text{ occupied}), \tag{3.5.15}$$

and this contributes $n_k(a, b)$ to $\mathrm{Var}(N_k(a, b))$. From now on, we consider $\pi \neq \rho$.

The probability in (3.5.14) needs to be summed over all possible pairs of paths (π, ρ) with $\pi \neq \rho$ that share at least one edge. In order to do this effectively, we start by introducing some notation.

Let $l = |\pi \cap \rho|$ denote the number of edges in $\pi \cap \rho$, so that $l \geq 1$ precisely when $\pi \cap \rho \neq \varnothing$. Since $\pi \neq \rho$, $l \leq k - 1$. When $\pi \neq \rho$, we have that $l \leq k - 1$, and since π and ρ are self-avoiding paths between a and b, l cannot be equal to $k - 1$, so that we consider $l \leq k - 2$ from now on. Let $k - l = |\rho \setminus \pi|$ be the number of edges in ρ that are not part of π.

Let m denote the number of connected subpaths in $\rho \setminus \pi$, so that $m \geq 1$ whenever $\pi \neq \rho$. Since $\pi_0 = \rho_0 = a$ and $\pi_k = \rho_k = b$, these subpaths start and end in vertices along the path π. We can view these subpaths as excursions of the path ρ from the walk π. By construction, between two excursions, there is at least one edge that π and ρ have in common.

Fix m. We define $\mathrm{Shape}(\pi, \rho)$, the *shape* of the pair (π, ρ), by

$$\mathrm{Shape}(\pi, \rho) = (\vec{x}_{m+1}, \vec{s}_m, \vec{t}_m, \vec{o}_{m+1}, \vec{r}_{m+1}), \tag{3.5.16}$$

where

(1) $\vec{x}_{m+1} \in \mathbb{N}_0^{m+1}$ and $x_j \geq 0$ is the length of the subpath in $\rho \cap \pi$ in between the $(j-1)^{\text{st}}$ and j^{th} subpath of $\pi \setminus \rho$. Here x_1 is the number of common edges in the subpath of $\rho \cap \pi$ that contains a, while x_{m+1} is the number of common edges in the subpath of $\rho \cap \pi$ that contains b, so that $x_1 \geq 0$ and $x_{m+1} \geq 0$. For $j \in \{2, \ldots, m\}$, $x_j \geq 1$;

(2) $\vec{s}_m \in \mathbb{N}^m$ and $s_j \geq 1$ is the number of edges in the j^{th} subpath of $\pi \setminus \rho$;

(3) $\vec{t}_m \in \mathbb{N}^m$ and $t_j \geq 1$ is the number of edges in the j^{th} subpath of $\rho \setminus \pi$;

(4) $\vec{o}_{m+1} \in [m+1]^{m+1}$ and o_j is the order of the j^{th} common subpath in $\rho \cap \pi$ of the path π in ρ, i.e., $o_2 = 5$ means that the second subpath that π has in common with ρ is the 5^{th} subpath that ρ has in common with π. Note that $o_1 = 1$ and $o_{m+1} = m + 1$, since π and ρ start and end in a and b, respectively;

(5) $\vec{r}_{m+1} \in \{0, 1\}^{m+1}$ is such that r_j describes the direction in which the j^{th} common subpath in $\rho \cap \pi$ of the path π is traversed by ρ, with $r_j = 1$ when the direction is the same for π and ρ and 0 otherwise. Thus, $r_1 = r_{m+1} = 1$.

The information in $\mathrm{Shape}(\pi, \rho)$ is precisely that necessary to piece together the topology of the two paths, except for the information of the *vertices* involved in π and ρ. See Figure 3.5.1 for an example of a pair of paths (π, ρ) and its corresponding shape.

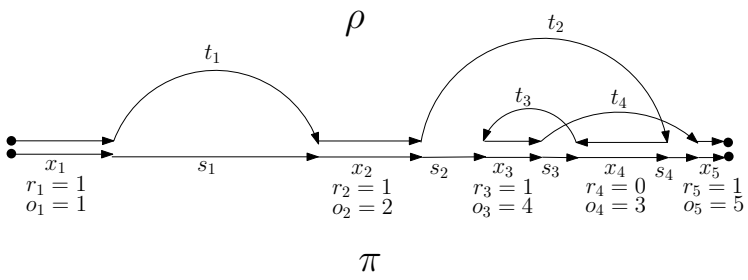

Figure 3.3: An example of a pair of paths (π, ρ) and its corresponding shape.

With $l = |\pi \cap \rho|$, we have

$$\sum_{j=1}^{m+1} x_j = l, \qquad \sum_{j=1}^{m} s_j = \sum_{j=1}^{m} t_j = k - l. \tag{3.5.17}$$

Let $\text{Shape}_{m,l}$ denote the set of shapes corresponding to pairs of paths (π, ρ) with m excursions and l common edges, so that (3.5.17) hold. Then,

$$\text{Var}(N_k(a,b)) \le \bar{n}_k(a,b) + \sum_{l=1}^{k-2} \sum_{m=1}^{k-l} \sum_{\sigma \in \text{Shape}_{m,l}} \sum_{\substack{\pi,\rho \in \mathcal{P}_k(a,b) \\ \text{Shape}(\pi,\rho) = \sigma}} \mathbb{P}(\pi, \rho \text{ occupied}). \tag{3.5.18}$$

When $\text{Shape}(\pi, \rho) = \sigma$ for some $\sigma \in \text{Shape}_{m,l}$, and since π and ρ both start and end in a and b, the union of paths $\pi \cup \rho$ visits $k + 1 + l - m$ distinct vertices. The vertex a is in $1 + \delta_{x_1,0}$ edges, and b in $1 + \delta_{x_{m+1},0}$ edges. Of the other $k - 1$ vertices in π, precisely $2m - \delta_{x_1,0} - \delta_{x_{m+1},0}$ are part of of three edges, and $k - 1 - 2m + \delta_{x_1,0} + \delta_{x_{m+1},0}$ are part of in two edges. The remaining $k - l - m$ vertices in ρ that are not part of π are part of precisely 2 edges. By construction, the $k + 1$ vertices of both π and ρ are disjoint, but the remaining $k - l - m$ vertices in ρ may intersect those of π. Therefore, denoting $a_1 = \delta_{x_1,0}, a_{m+1} = \delta_{x_{m+1},0}$,

$$\mathbb{P}(\pi, \rho \text{ occupied}) = u_a^{1+a_1} u_b^{1+a_{m+1}} \prod_{s=1}^{2m-a_1-a_{m+1}} u_{v_s}^3 \prod_{t=2m-a_1-a_{m+1}+1}^{2(k-1)-l-m} u_{v_t}^2, \tag{3.5.19}$$

where $\{(v_1, \ldots, v_{k+1+l-m})\} \in \mathcal{I}^{k-1+l-m}$.

For a fixed $\sigma \in \text{Shape}_{m,l}$ now bound the sum over $\pi, \rho \in \mathcal{P}_k(a,b)$ such that $\text{Shape}(\pi, \rho) = \sigma$ from above by summing (3.5.19) over *all* $\{(v_1, \ldots, v_{k-1+l-m})\} \in \mathcal{I}^{k-1+l-m}$, to obtain for any $\sigma \in \text{Shape}_{m,l}$,

$$\begin{aligned} &\sum_{\substack{\pi,\rho \in \mathcal{P}_k(a,b) \\ \text{Shape}(\pi,\rho) = \sigma}} \mathbb{P}(\pi, \rho \text{ occupied}) \qquad (3.5.20) \\ &\le u_a u_b \gamma_{\mathcal{I}}^{2m} \nu_{\mathcal{I}}^{2k-1-3m-l} \Big(\frac{u_a \nu_{\mathcal{I}}}{\gamma_{\mathcal{I}}}\Big)^{\delta_{x_1,0}} \Big(\frac{u_b \nu_{\mathcal{I}}}{\gamma_{\mathcal{I}}}\Big)^{\delta_{x_{m+1},0}} \\ &= \bar{n}_k(a,b)^2 \gamma_{\mathcal{I}}^{2(m+1)} \nu_{\mathcal{I}}^{-3(m-1)-l} \Big(\frac{\gamma_{\mathcal{I}}}{u_a \nu_{\mathcal{I}}}\Big)^{1-\delta_{x_1,0}} \Big(\frac{\gamma_{\mathcal{I}}}{u_b \nu_{\mathcal{I}}}\Big)^{1-\delta_{x_{m+1},0}}. \end{aligned}$$

Therefore, we arrive at

$$\mathrm{Var}(N_k(a,b)) \leq n_k(a,b) + \bar{n}_k(a,b)^2 \sum_{l=1}^{k-2} \sum_{m=1}^{k-l} \gamma_{\mathcal{I}}^{2(m-1)} \nu_{\mathcal{I}}^{-3(m-1)-l} \times \sum_{\sigma \in \mathrm{Shape}_{m,l}} \Big(\frac{\gamma_{\mathcal{I}}}{u_a \nu_{\mathcal{I}}}\Big)^{1-\delta_{x_1,0}} \Big(\frac{\gamma_{\mathcal{I}}}{u_b \nu_{\mathcal{I}}}\Big)^{1-\delta_{x_{m+1},0}}. \tag{3.5.21}$$

We continue to compute the number of shapes in the following lemma:

Lemma 3.18 (The number of shapes)**.** *Fix $m \geq 1$ and $l \leq k-2$. For $m = 1$, the number of shapes in* $\mathrm{Shape}_{m,l}$ *fixed $a_1 = \delta_{x_1,0}, a_{m+1} = \delta_{x_{m+1},0}$ equals l when $a_1 = a_{m+1} = 0$, 1 when $a_1 + a_{m+1} = 1$ and 0 when $a_1 = a_{m+1} = 1$. For $m \geq 2$, the number of shapes in* $\mathrm{Shape}_{m,l}$ *fixed $a_1 = \delta_{x_1,0}, a_{m+1} = \delta_{x_{m+1},0}$ is bounded by*

$$2^{m-1}(m-1)! \binom{k-l-1}{m-1}^2 \binom{l}{m-a_1-a_{m+1}}. \tag{3.5.22}$$

Proof. Since $r_1 = r_{m+1} = 1$, there are 2^{m-1} directions in which the common parts can be traversed. Since there are m distinct parts, there are $m+1$ common parts. The first part contains vertex a, the last part contains vertex b. Thus, there are $(m-1)!$ orders $\vec{o}_{m+1}$ of the common parts when we have fixed the directions the paths can be traversed.

In counting the number of $\vec{x}_{m+1}, \vec{s}_m, \vec{t}_m$, we repeatedly use the fact that there are $\binom{a-1}{b-1}$ possible sequences $(y_1, \ldots, y_b) \in \mathbb{N}_0^b$ such that $\sum_{j=1}^b y_j = a$. This can be seen by representing a as a sequence of a ones, separated by $a-1$ zeros. We draw b zeros, which we can do in $\binom{a-1}{b-1}$ possible ways. Then, we note that a sequence $(y_1, \ldots, y_b) \in \mathbb{N}_0^b$ such that $\sum_{j=1}^b y_j = a$ can be obtained uniquely by letting y_i be the number of ones in between the $(i-1)^{\text{st}}$ and i^{th} chosen zero. Similarly, there are $\binom{a+b-1}{b-1}$ possible sequences $(y_1, \ldots, y_b) \in \mathbb{N}^b$ such that $\sum_{j=1}^b y_j = a$, since we can apply the previous equality to $(y_1+1, \ldots, y_b+1) \in \mathbb{N}^b$.

Using the above, we continue to count the number of shapes. The number of $(s_1, \ldots, s_m) \in \mathbb{N}^m$ such that $s_j \geq 1$ and $\sum_{j=1}^m s_j = k-l$ equals

$$\binom{k-l-1}{m-1}. \tag{3.5.23}$$

The same applies to $(t_1, \ldots, t_m) \in \mathbb{N}^m$ such that $t_j \geq 1$ and $\sum_{j=1}^m t_j = l$. In counting the number of possible $\vec{x}_{m+1}$ such that $\sum_{j=1}^{m+1} x_j = l$, we need to count their numbers separately for $x_1 = 0$ and $x_1 \geq 1$, and for $x_{m+1} = 0$ and $x_{m+1} \geq 1$. When $m = 1$, the number is zero when $x_1 = x_2 = 0$, since $x_1 = x_2 = 0$ implies that the paths share no edges. Denote $a_1 = \delta_{x_1,0}, a_{m+1} = \delta_{x_{m+1},0}$, and suppose that $m - a_1 - a_{m+1} \geq 0$. Then, there are

$$\binom{l}{m-a_1-a_{m+1}} \tag{3.5.24}$$

possible choice of $\vec{x}_{m+1}$ with fixed $a_1 = \delta_{x_1,0}, a_{m+1} = \delta_{x_{m+1},0}$. The claim follows by multiplying these bounds on the number of choices for $\vec{r}_{m+1}$, $\vec{o}_{m+1}$, $\vec{s}_m$, $\vec{t}_m$ and $\vec{x}_{m+1}$. □

We complete the proof of Proposition 3.17. By (3.5.21) and applying Lemma 3.18, it suffices to sum

$$2^{m-1}(m-1)!\binom{k-l-1}{m-1}^2\binom{l}{m-a_1-a_{m+1}} \times \Big(\frac{2\gamma_{\mathcal{I}}^2}{\nu_{\mathcal{I}}^3}\Big)^{m-1}\nu_{\mathcal{I}}^{-l}\Big(\frac{\gamma_{\mathcal{I}}}{u_a\nu_{\mathcal{I}}}\Big)^{1-a_1}\Big(\frac{\gamma_{\mathcal{I}}}{u_b\nu_{\mathcal{I}}}\Big)^{1-a_{m+1}} \tag{3.5.25}$$

over $l \in [k-2]$, $m \in [k-l]$ and $a_1, a_{m+1} \in \{0,1\}$, where, by convention, $\binom{l}{-1} = 0$.

We start with $m = 1$, for which we obtain that the sum of (3.5.25) over the other variables equals

$$\begin{aligned}&\gamma_{\mathcal{I}}\Big(\frac{1}{u_a}+\frac{1}{u_b}\Big)\sum_{l=1}^{\infty}\nu_{\mathcal{I}}^{-(l-1)} + \frac{\gamma_{\mathcal{I}}^2}{u_au_b\nu_{\mathcal{I}}}\sum_{l=1}^{\infty}l\nu_{\mathcal{I}}^{-(l-1)}\\ &\quad= \frac{\gamma_{\mathcal{I}}\nu_{\mathcal{I}}^2}{\nu_{\mathcal{I}}-1}\Big(\frac{1}{u_a}+\frac{1}{u_b}\Big) + \frac{\gamma_{\mathcal{I}}^2\nu_{\mathcal{I}}}{u_au_b(\nu_{\mathcal{I}}-1)^2},\end{aligned} \tag{3.5.26}$$

where we use that, for $a \in [0,1)$,

$$\sum_{l=0}^{\infty}a^{-l} = a/(1-a), \qquad \sum_{l=0}^{\infty}la^{-(l-1)} = a^2/(1-a)^2. \tag{3.5.27}$$

The terms in (3.5.26) are the first two terms appearing on the right-hand side of (3.5.7).

This leaves us to bound the contribution when $m \geq 2$. We continue by bounding

$$\binom{k-l-1}{m-1}(m-1)! = \frac{1}{(m-1)!}\Big(\frac{(k-l-1)!}{(k-l-m)!}\Big)^2 \leq \frac{k^{2(m-1)}}{(m-1)!}, \tag{3.5.28}$$

and, using that $\binom{a}{b} \leq a^b/b!$ and $l \leq k$,

$$\binom{l}{m-a_1-a_{m+1}} \leq \frac{l^{m-a_1-a_{m+1}}}{(m-a_1-a_{m+1})!} \leq k^m. \tag{3.5.29}$$

Therefore, the number of shapes in $\mathrm{Shape}_{m,l}$ is, for each $l \geq 1$ and $m \geq 2$, bounded by

$$2^{m-1}\frac{k^{2(m-1)}}{(m-1)!}k^m = k\frac{(2k^3)^{m-1}}{(m-1)!}. \tag{3.5.30}$$

Since the above is independent of l, we can start by summing (3.5.25) over $l \geq 1$, and over $a_1, a_{m+1} \in \{0,1\}$ to obtain a bound of the form

$$\begin{aligned}&k\Big(1+\frac{\gamma_{\mathcal{I}}}{u_a\nu_{\mathcal{I}}}\Big)\Big(1+\frac{\gamma_{\mathcal{I}}}{u_b\nu_{\mathcal{I}}}\Big)\frac{\nu_{\mathcal{I}}}{\nu_{\mathcal{I}}-1}\sum_{m\geq 2}\frac{(2k^3)^{m-1}}{(m-1)!}\Big(\frac{\gamma_{\mathcal{I}}^2}{\nu_{\mathcal{I}}^3}\Big)^{m-1}\\ &\quad= k\Big(1+\frac{\gamma_{\mathcal{I}}}{u_a\nu_{\mathcal{I}}}\Big)\Big(1+\frac{\gamma_{\mathcal{I}}}{u_b\nu_{\mathcal{I}}}\Big)\frac{\nu_{\mathcal{I}}}{\nu_{\mathcal{I}}-1}\big(\mathrm{e}^{2k^3\gamma_{\mathcal{I}}^2/\nu_{\mathcal{I}}^3}-1\big).\end{aligned} \tag{3.5.31}$$

The term in (3.5.31) is the last term appearing on the right-hand side of (3.5.7). Summing the bounds in (3.5.26) and (3.5.31) proves (3.5.7). □

Exercise 3.26 (Variance of two paths). *Prove that* $\mathrm{Var}(N_k(a,b)) \leq \mathbb{E}[N_k(a,b)]$ *for* $k = 2$.

Exercise 3.27 (Variance of three paths). *Compute* $\mathrm{Var}(N_3(a,b))$ *explicitly, and compare it to the bound in* (3.5.7).

Exercise 3.28 (Variance on paths for $\mathrm{ER}_n(\lambda/n)$). *Let* $A, B \subseteq [n]$, *and let* $N_k(A,B)$ *denote the number of paths of length* k *connecting* A *to* B *(where a path connecting* A *and* B *avoids* A *and* B *except for the starting and end point). Show that for* $k \leq K \log n$,

$$\mathbb{E}[N_k(A,B)] = \lambda^k |A||B|\big(1 - \frac{|A|+|B|}{n}\big)^k (1+o(1)). \tag{3.5.32}$$

Use Proposition 3.17 to bound the variance of $N_k(A,B)$, *and prove that*

$$N_k(A,B)/\mathbb{E}[N_k(A,B)] \xrightarrow{\mathbb{P}} 1 \tag{3.5.33}$$

when $|A|, |B| \to \infty$ *with* $|A| + |B| = o(n/k)$.

3.5.2 Logarithmic upper bound on typical distances in finite variance case

In this section, we prove that two uniformly chosen vertices that are conditioned to be connected are with high probability within distance $(1+\varepsilon)\log_\nu n$, as formulated in the following theorem:

Theorem 3.19 (Logarithmic upper bound graph distances $\mathrm{NR}_n(\boldsymbol{w})$). *Assume that [I, Condition 6.4(a)-(c)] hold, where* $\nu = \mathbb{E}[W^2]/\mathbb{E}[W] \in (1,\infty)$. *Then, for any* $\varepsilon > 0$,

$$\mathbb{P}(H_n \leq (1+\varepsilon)\log_\nu n \mid H_n < \infty) = 1 + o(1). \tag{3.5.34}$$

The same results hold for $\mathrm{CL}_n(\boldsymbol{w})$ *and* $\mathrm{GRG}_n(\boldsymbol{w})$ *under the same conditions.*

Organization of the proof of Theorem 3.19. We prove Theorem 3.19 by combining the *branching process comparison* to a *second moment method* using Proposition 3.17 on the number of paths of a given length. More precisely, we fix $m \geq 1$ large, and recall that $\mathcal{N}_{\leq m}(U_1)$ and $\mathcal{N}_{\leq m}(U_2)$ denote the vertices at distance at most m from V_1 and V_2 respectively, and let $\mathcal{N}_m(U_1)$ and $\mathcal{N}_m(U_2)$ denote the vertices at distance precisely equal to m. We condition on $\mathcal{N}_{\leq m}(U_1)$ and $\mathcal{N}_{\leq m}(U_2)$ such that $\mathcal{N}_m(U_1) \neq \varnothing$ and $\mathcal{N}_m(U_2) \neq \varnothing$. By Corollary 3.13, the probabilities of the latter event is close to ζ_m^2, where $\zeta_m = \mathbb{P}(\mathcal{Z}_m^{(1)} > 0)$ is the probability that the branching process survives to generation m. Then, $\zeta_m \to \zeta$ when $m \to \infty$, and, conditionally on $\mathcal{Z}_m^{(1)} > 0$, $\mathcal{Z}_m^{(1)} \geq M$ whp, for any M and as $m \to \infty$. This explains the branching process approximation. We take $u_i = w_i/\sqrt{\ell_n}$,

$$a = \mathcal{N}_m(U_1), \qquad b = \mathcal{N}_m(U_2), \tag{3.5.35}$$

so that

$$u_a = \frac{1}{\sqrt{\ell_n}} \sum_{i \in \mathcal{N}_m(U_1)} w_i = \mathcal{W}_m(U_1)/\sqrt{\ell_n}, \qquad u_b = \frac{1}{\sqrt{\ell_n}} \sum_{i \in \mathcal{N}_m(U_2)} w_i = \mathcal{W}_m(U_2)/\sqrt{\ell_n}. \tag{3.5.36}$$

We formalize the above ideas in the following lemma:

Lemma 3.20 (Branching process approximation). *As $n \to \infty$,*

$$(\mathcal{W}_m(U_1), \mathcal{W}_m(U_2)) \xrightarrow{d} \Big(\sum_{j=1}^{\mathcal{Z}_m^{(1)}} W^{(1)}(j), \sum_{j=1}^{\mathcal{Z}_m^{(2)}} W^{(2)}(j) \Big), \tag{3.5.37}$$

where $(\mathcal{Z}_m^{(1)}, \mathcal{Z}_m^{(2)})$ are the generation sizes of two independent branching processes, and $(W^{(1)}(j))_{j \geq 1}$ and $(W^{(2)}(j))_{j \geq 1}$ are two independent sequences of i.i.d. random variables with distribution F.

Proof. By Corollary 3.13, $\widetilde{Z}_m^{(1)} = |\mathcal{N}_m(U_1)|$ and $\widetilde{Z}_m^{(2)} = |\mathcal{N}_m(U_2)|$ jointly converge in distribution to $(\mathcal{Z}_m^{(1)}, \mathcal{Z}_m^{(2)})$, which are independent generation sizes of MMPBPs. Each of the individuals in $\mathcal{N}_m(U_1)$ and $\mathcal{N}_m(U_2)$ receives a mark M_i, and its weight is w_{M_i}. By Proposition 3.10, these marks are i.i.d. random variables conditioned to be distinct (if they were not distinct, then the corresponding individuals would have been thinned). Since, whp, the proof of Corollary 3.13 shows that no vertex is thinned, whp we can ignore the thinning. Then, $\mathcal{W}_m(U_1) = \sum_{j=1}^{\widetilde{Z}_m^{(1)}} W_n(j)$, where $(W_n(j))_{j \geq 1}$ are i.i.d. copies of W_n. By [I, Condition 6.4(a)], $W_n \xrightarrow{d} W$, so that $\mathcal{W}_m(U_1) \xrightarrow{d} \sum_{j=1}^{\mathcal{Z}_m^{(1)}} W^{(1)}(j)$. The joint convergence follows in a similar fashion. □

Second moment method and path counting. Fix $k = k_n = (1+\varepsilon)\log_\nu n - 2m$. We next present the details of the second moment method that shows that whp, on the event that $\mathcal{N}_m(U_1) \neq \varnothing$ and $\mathcal{N}_m(U_2) \neq \varnothing$, there exist a path of length $k_n - 2m$ connecting $\mathcal{N}_m(U_1)$ and $\mathcal{N}_m(U_2)$. This ensures that, on the event that $\mathcal{N}_m(U_1) \neq \varnothing$ and $\mathcal{N}_m(U_2) \neq \varnothing$, the event $H_n \leq k_n - 2m$ occurs whp. For this, we take $u_i = w_i/\sqrt{\ell_n}$. We fix $K \geq 1$ sufficiently large and take

$$\mathcal{I} = \{i \in [n] \colon w_i \leq K\} \setminus (\mathcal{N}_{\leq m}(U_1) \cup \mathcal{N}_{\leq m}(U_2)). \tag{3.5.38}$$

We investigate the constants appearing in Proposition 3.17 in the following lemma:

Lemma 3.21 (Parameters in path counting). *Conditionally on $\mathcal{N}_{\leq m}(U_1)$ and $\mathcal{N}_{\leq m}(U_2)$, and with $a = \mathcal{N}_m(U_1), b = \mathcal{N}_m(U_2)$, for $k = (1+\varepsilon)\log_\nu n$,*

$$\lim_{n \to \infty} n_k(a,b) = \infty, \qquad \lim_{n \to \infty} \frac{\bar{n}_k(a,b)}{\underline{n}_k(a,b)} = 1, \tag{3.5.39}$$

and, as $n \to \infty$,

$$\frac{\mathrm{Var}(N_k(a,b))}{\mathbb{E}[N_k(a,b)]^2} \leq \frac{K\nu^2}{\nu - 1}\Big(\frac{1}{\sqrt{\ell_n} u_a} + \frac{1}{\sqrt{\ell_n} u_b}\Big) + \frac{K^2\nu^2}{(\nu-1)\ell_n u_a u_b} + o_{\mathbb{P}}(1). \tag{3.5.40}$$

Proof. By (3.5.3),

$$\underline{n}_k(a,b) = u_a u_b \nu_{\mathcal{I}_{a,b}}^{k-1}, \tag{3.5.41}$$

and

$$\frac{\bar{n}_k(a,b)}{\underline{n}_k(a,b)} = \big(\nu_{\mathcal{I}_{a,b}}/\nu_{\mathcal{I}_{a,b,k}}\big)^{k-1}. \tag{3.5.42}$$

We start by investigating $\nu_{\mathcal{I}}$. Denote

$$\nu(K) = \frac{\mathbb{E}[W^2 \mathbb{1}_{\{W\le K\}}]}{\mathbb{E}[W]}. \tag{3.5.43}$$

Then, by (3.5.38) and the fact that $\mathcal{N}_{\le m}(U_1)$ and $\mathcal{N}_{\le m}(U_2)$ contain a finite number of vertices,

$$\lim_{n\to\infty} \nu_{\mathcal{I}} = \nu(K). \tag{3.5.44}$$

The same applies to $\nu_{\mathcal{I}_{a,b}}$ and $\nu_{\mathcal{I}_{a,b,k}}$. Then, with $K > 0$ chosen so large that $\nu(K) \ge \nu - \varepsilon/2$ and with $k = (1+\varepsilon)\log_\nu n$,

$$\nu_{\mathcal{I}_{a,b}}^{k-1} = \mathcal{W}_m(U_1)\mathcal{W}_m(U_2)\frac{n}{\ell_n} n^{(1+\varepsilon)\log \nu_{\mathcal{I}_{a,b}}/\log\nu - 1} \to \infty, \tag{3.5.45}$$

where K and n are so large that $(1+\varepsilon)\nu_{\mathcal{I}}/\nu > 1$. This proves the first property in (3.5.39).

To prove the second property in (3.5.39), we note that the set $\mathcal{I}_{a,b,k}$ is obtained from $\mathcal{I}_{a,b}$ by removing the k vertices with highest weight. Since $w_i \le K$ for all $i \in \mathcal{I}$ (recall (3.5.38)), $\nu_{\mathcal{I}_{a,b}} \le \nu_{\mathcal{I}_{a,b,k}} + kK/\ell_n$. Since $k \le A\log n$, we therefore arrive at

$$\frac{\bar{n}_k(a,b)}{\underline{n}_k(a,b)} \le \big(1 + kK/(\ell_n \nu_{\mathcal{I}_{a,b,k}})\big)^{k-1} = \mathrm{e}^{k^2 K/(\ell_n \nu_{\mathcal{I}_{a,b,k}})} \to 1, \tag{3.5.46}$$

as required.

To prove (3.5.40), we rely on Proposition 3.17. We have already shown that $n_k(a,b) = \mathbb{E}[N_k(a,b)] \to \infty$, so that the first term on the right-hand side of (3.5.7) is $o(\mathbb{E}[N_k(a,b)]^2)$. Further, by (3.5.38),

$$\gamma_{\mathcal{I}} \le \nu_{\mathcal{I}}(\max_{i\in\mathcal{I}} u_i) \le \frac{\nu_{\mathcal{I}} K}{\sqrt{\ell_n}}, \tag{3.5.47}$$

so that, for $k \le A\log n$ with $A > 1$ fixed,

$$(1 + \frac{\gamma_{\mathcal{I}}}{u_a \nu_{\mathcal{I}}})(1 + \frac{\gamma_{\mathcal{I}}}{u_b \nu_{\mathcal{I}}}) k(\mathrm{e}^{2k^3\gamma_{\mathcal{I}}^2/\nu_{\mathcal{I}}^3} - 1) = o_{\mathbb{P}}(1). \tag{3.5.48}$$

Substituting these bounds into (3.5.40) and using (3.5.39) yields the claim. □

Completion of the proof of Theorem 3.19. Now we are are ready to complete the proof of Theorem 3.19. We must show that

$$\mathbb{P}(k_n < H_n < \infty) = o(1). \tag{3.5.49}$$

Indeed, then $\mathbb{P}(H_n > k_n \mid H_n < \infty) = o(1)$ since, $\mathbb{P}(H_n < \infty) \to \zeta^2 > 0$ by Theorem 3.2. We rewrite

$$\begin{aligned}\mathbb{P}(k_n < H_n < \infty) &= \mathbb{P}(k_n < H_n < \infty, \mathcal{N}_m(U_1) \neq \varnothing, \mathcal{N}_m(U_2) \neq \varnothing) \\ &\leq \mathbb{P}(N_{k_n-2m}(\mathcal{N}_m(U_1), \mathcal{N}_m(U_2)) = 0, \mathcal{N}_m(U_1) \neq \varnothing, \mathcal{N}_m(U_2) \neq \varnothing).\end{aligned} \tag{3.5.50}$$

Recall that $k = k_n = (1+\varepsilon)\log_\nu n$. By the Chebychev inequality (Theorem 2.16), and given $\mathcal{N}_{\leq m}(U_1), \mathcal{N}_{\leq m}(U_2)$, the conditional probability on $\{H_n > k_n\}$ is at most

$$\frac{\mathrm{Var}(N_{k-2m}(a,b))}{\mathbb{E}[N_{k-2m}(a,b)]^2} \leq \frac{K\nu^2}{\nu-1}\Big(\frac{1}{\sqrt{\ell_n}u_a} + \frac{1}{\sqrt{\ell_n}u_b}\Big) + \frac{K^2\nu^2}{(\nu-1)\ell_n u_a u_b} + o_{\mathbb{P}}(1). \tag{3.5.51}$$

When $\mathcal{N}_m(U_1) \neq \varnothing$ and $\mathcal{N}_m(U_2) \neq \varnothing$, by (3.5.36),

$$\frac{1}{\sqrt{\ell_n}u_a} + \frac{1}{\sqrt{\ell_n}u_b} \xrightarrow{\mathbb{P}} \Big(\sum_{j=1}^{\mathcal{Z}_m^{(1)}} W^{(1)}(j)\Big)^{-1} + \Big(\sum_{j=1}^{\mathcal{Z}_m^{(2)}} W^{(2)}(j)\Big)^{-1} \xrightarrow{\mathbb{P}} 0, \tag{3.5.52}$$

when $m \to \infty$. Therefore,

$$\mathbb{P}\Big(N_{k-2m}(a,b) = 0 \mid \mathcal{N}_m(U_1) \neq \varnothing, \mathcal{N}_m(U_2) \neq \varnothing\Big) \xrightarrow{\mathbb{P}} 0, \tag{3.5.53}$$

and, by Lebesgues Dominated Convergence Theorem (Theorem A.17),

$$\mathbb{P}(H_n > k_n, \mathcal{N}_m(U_1) \neq \varnothing, \mathcal{N}_m(U_2) \neq \varnothing) \to 0, \tag{3.5.54}$$

which completes the proof. □

We close this section by describing what happens when $\tau = 3$, and there are no slowly varying functions:

Distances for the critical case $\tau = 3$ When $\tau = 3$, w_i is approximately $c(n/i)^{1/2}$. It turns our that this changes the distances only by a doubly logarithmic factor:

Theorem 3.22 (Logarithmic upper bound graph distances $\mathrm{NR}_n(\boldsymbol{w})$)**.** *Assume that [I, Condition 6.4(a)-(b)] hold, and that there exists constants $c_2 > c_1 > 0$ and $\alpha > 0$ such that for all $x \leq n^\alpha$,*

$$[1 - F_n](x) \geq c_1/x^2, \tag{3.5.55}$$

and for all $x \geq 0$,

$$[1 - F_n](x) \leq c_2/x^2. \tag{3.5.56}$$

Then, conditionally on $H_n < \infty$,

$$\frac{H_n \log\log n}{\log n} \xrightarrow{\mathbb{P}} 1. \tag{3.5.57}$$

The same results hold for $\mathrm{CL}_n(\boldsymbol{w})$ and $\mathrm{GRG}_n(\boldsymbol{w})$ under the same conditions.

The lower bound in Theorem 3.22 is already stated in Corollary 3.6. The upper bound can be proved using the path counting techniques in Proposition 3.17 and adaptations of it. We now sketch the proof.

Let $\eta > 0$ and let

$$\alpha_n = \mathrm{e}^{\nu_n^{1-\eta}}. \tag{3.5.58}$$

Define the *core* of $\mathrm{NR}_n(\boldsymbol{w})$ to be

$$\mathrm{Core}_n = \{i \colon w_i \geq \alpha_n\}. \tag{3.5.59}$$

The proof of Theorem 3.22 follows from the following two propositions:

Proposition 3.23 (Typical distances in the core)**.** *Under the conditions of Theorem 3.22, let $V_1', V_2' \in \mathrm{Core}_n$ be chosen with probability proportional to their weight, i.e.,*

$$\mathbb{P}(V_i' = j) = \frac{w_j}{\sum_{v \in \mathrm{Core}_n} w_v}, \tag{3.5.60}$$

and let H_n' be the graph distance between V_1', V_2' in Core_n. Then, for any $\varepsilon > 0$, there exists an $\eta > 0$ such that

$$\mathbb{P}\big(H_n' \leq \frac{(1+\varepsilon)\log n}{\log\log n}\big) \to 1. \tag{3.5.61}$$

Proposition 3.24 (From the periphery to the core)**.** *Under the conditions of Theorem 3.22, let V_1, V_2 be two vertices chosen uniformly at random from $[n]$. Then, for any $\eta > 0$,*

$$\mathbb{P}(d_{\mathrm{NR}_n(\boldsymbol{w})}(V_1, \mathrm{Core}_n) \leq \nu_n^{1-\eta}, d_{\mathrm{NR}_n(\boldsymbol{w})}(V_2, \mathrm{Core}_n) \leq \nu_n^{1-\eta}) \to \zeta^2. \tag{3.5.62}$$

To see that Propositions 3.23–3.24 imply Theorem 3.22, we note that

$$\begin{aligned} H_n &= d_{\mathrm{NR}_n(\boldsymbol{w})}(V_1, V_2) \\ &\leq d_{\mathrm{NR}_n(\boldsymbol{w})}(V_1, \mathrm{Core}_n) + d_{\mathrm{NR}_n(\boldsymbol{w})}(V_2, \mathrm{Core}_n) + d_{\mathrm{NR}_n(\boldsymbol{w})}(V_1', V_2'), \end{aligned} \tag{3.5.63}$$

where $V_1', V_2' \in \mathrm{Core}_n$ are the vertices in Core_n found first in the breadth-first search. Then, by Proposition 3.10, $V_1', V_2' \in \mathrm{Core}_n$ are chosen with probability proportional to their weight. Therefore, when n is so large that $\nu_n^{1-\eta} \leq \varepsilon \log n/(4 \log\log n)$,

$$\begin{aligned} &\mathbb{P}(H_n \leq (1+\varepsilon)\frac{\log n}{\log\log n}) \\ &\quad \geq \mathbb{P}\big(d_{\mathrm{NR}_n(\boldsymbol{w})}(V_i, \mathrm{Core}_n) \leq \nu_n^{1-\eta}, i = 1,2, d_{\mathrm{NR}_n(\boldsymbol{w})}(V_1', V_2') \leq (1+\varepsilon/2)\frac{\log n}{\log\log n}\big). \end{aligned} \tag{3.5.64}$$

By Proposition 3.24, the probability of the first event converges to ζ^2, and by Proposition 3.24, the probability of the second event converges to 1. We conclude that

$$\mathbb{P}\big(H_n \leq (1+\varepsilon)\frac{\log n}{\log\log n}\big) \to \zeta^2.$$

Since also $\mathbb{P}(H_n < \infty) \to \zeta^2$, this completes the proof. □

The proofs of Propositions 3.23–3.24 follow from path counting techniques similar to the ones carried out above. We now sketch their proofs, starting with Proposition 3.23. For Proposition 3.23, we take

$$a = V_1', \qquad b = V_2', \qquad \mathcal{I} = \{w_i \colon w_i \in [K, \sqrt{\alpha_n}]\}. \tag{3.5.65}$$

The whole point is that there exists a constant $c > 0$ such that

$$\nu_{\mathcal{I}} \geq c \log \alpha_n = c\nu_n^{1-\eta}, \tag{3.5.66}$$

while $u_a \geq \alpha_n/\sqrt{\ell_n}$, $u_b \geq \alpha_n/\sqrt{\ell_n}$, so that

$$\mathbb{E}[N_k(a,b)] \approx \alpha_n^2 c^k \nu_n^{k(1-\eta)}/\ell_n \to \infty \tag{3.5.67}$$

for $k = \log n/((1-\eta)\log \nu_n) \leq (1+\varepsilon/2)\log n/\log \nu_n$ when η is such that $1/(1-\eta) \leq 1 + \varepsilon/2$. Further,

$$\gamma_{\mathcal{I}} \leq \sqrt{\alpha_n}/\sqrt{\ell_n}, \tag{3.5.68}$$

so that $\mathrm{Var}(N_k(a,b))/\mathbb{E}[N_k(a,b)]^2 \to 0$ by Proposition 3.17.

Exercise 3.29 (ν_n bound for $\tau = 3$)**.** *Prove that* (3.5.55) *and* (3.5.56) *imply that* $\nu_{\mathcal{I}} \geq c \log \alpha_n$ *by using*

$$\frac{1}{n}\sum_{i\in\mathcal{I}} w_i^2 = \mathbb{E}[W_n^2 \mathbb{1}_{\{W_n \in [K,\sqrt{\alpha_n}]\}}] = 2\int_K^{\sqrt{\alpha_n}} x[F_n(\sqrt{\alpha_n}) - F_n(x)]dx. \tag{3.5.69}$$

Exercise 3.30 (Expected number of paths within Core_n diverges)**.** *Prove that*

$$\mathbb{E}[N_k(a,b)] \to \infty$$

for $a = V_1'$, $b = V_2'$ *and* $k = \log n/((1-\eta)\log \nu_n)$.

Exercise 3.31 (Concentration of number of paths within Core_n)**.** *Prove that*

$$\mathrm{Var}(N_k(a,b))/\mathbb{E}[N_k(a,b)]^2 \to 0$$

for $a = V_1'$, $b = V_2'$ *and* $k = \log n/((1-\eta)\log \nu_n)$.

For Proposition 3.23, we again condition on $\mathcal{N}_m(U_1) \neq \varnothing, \mathcal{N}_m(U_2) \neq \varnothing$, the probability of which converges to ζ^2 when first $n \to \infty$ followed by $m \to \infty$. Then, we perform a second moment method on the number of paths between $\mathcal{N}_m(U_i)$ and Core_n. For this, we take $k = \nu_n^{1-\eta}$ and

$$a = \mathcal{N}_m(U_1), \qquad b = \mathrm{Core}_n, \qquad \mathcal{I} = \{i \colon w_i \leq K\} \setminus (\mathcal{N}_{\leq m}(U_1) \cup \mathcal{N}_{\leq m}(U_2)). \tag{3.5.70}$$

Then we follow the proof in (3.5.51)–(3.5.54) to show that

$$\mathbb{P}(d_{\mathrm{NR}_n(\boldsymbol{w})}(V_1, \mathrm{Core}_n) > \nu_n^{1-\eta}, \mathcal{N}_m(U_1) \neq \varnothing, \mathcal{N}_m(U_2) \neq \varnothing) \to 0, \tag{3.5.71}$$

as required. Note, for this, that, conditionally on $\mathcal{N}_{\le m}(U_1), \mathcal{N}_{\le m}(U_2)$

$$\mathbb{E}[N_k(a,b)] \approx \nu(K)^k \mathcal{W}_m(U_1) \frac{1}{\ell_n} \sum_{i \in \mathrm{Core}_n} w_i, \tag{3.5.72}$$

where $\nu(K) \to \infty$ as $K \to \infty$, and where, by (3.5.55),

$$\frac{1}{n} \sum_{i \in \mathrm{Core}_n} w_i \ge \alpha_n [1 - F_n](\alpha_n) \ge c/\alpha_n. \tag{3.5.73}$$

Therefore, $\mathbb{E}[N_k(a,b)] \to \infty$ as soon as $k \ge 2 \log \alpha_n / \log \nu(K)$, which is satisfied for K sufficiently large and $k = \nu_n^{1-\eta}$.

Exercise 3.32 (Completion proof Proposition 3.23). *Complete the proof of Proposition 3.23 by adapting the arguments in* (3.5.51)–(3.5.54).

3.6 Related results for rank-1 inhomogeneous random graphs

In this section, we discuss some related results for inhomogeneous random graphs. While we give intuition about their proofs, we shall not include them in full detail.

When $\lim_{n\to\infty} \sup_{x,y} \kappa_n(x,y) = \infty$, the largest subcritical clusters can have rather different behavior, as we now show for the rank-1 case. Note that, by Theorem 2.7 as well as the fact that $\|\mathbf{T}_\kappa\| = \nu = \mathbb{E}[W^2]/\mathbb{E}[W]$, a rank-1 model can only be subcritical when $\mathbb{E}[W^2] < \infty$, i.e., in the case of finite variance degrees. However, when W has a power-law tail, i.e., when $\mathbb{P}(W \ge w) \sim w^{-(\tau-1)}$, then the highest weight can be much larger than $\log n$. When this is the case, then also the largest subcritical cluster is much larger than $\log n$, as proved in the following theorem:

Theorem 3.25 (Subcritical phase for rank-1 IRG). *Let $\boldsymbol{w}$ satisfy [I, Condition 6.4(a)-(c)] with $\nu = \mathbb{E}[W^2]/\mathbb{E}[W] < 1$, and, further, that there exist $\tau > 3$ and $c_2 > 0$ such that*

$$[1 - F_n](x) \le c_2 x^{-(\tau-1)}. \tag{3.6.1}$$

Then, for $\mathrm{NR}_n(\boldsymbol{w})$ with $\Delta = \max_{j\in[n]} w_j$,

$$|\mathcal{C}_{\max}| = \frac{\Delta}{1-\nu} + o_{\mathbb{P}}(n^{1/(\tau-1)}). \tag{3.6.2}$$

Theorem 3.25 is most interesting in the case where the limiting distribution function F in [I, Condition 6.4] has a power-law tail. For example, for $\boldsymbol{w}$ as in (6.1.14), let F satisfy

$$[1 - F](x) = cx^{-(\tau-1)}(1 + o(1)). \tag{3.6.3}$$

Then, $\Delta = w_1 = [1-F]^{-1}(1/n) = (cn)^{1/(\tau-1)}(1+o(1))$. Therefore, $|\mathcal{C}_{\max}| = (cn)^{1/(\tau-1)}/(1-\nu) + o(n^{1/(\tau-1)})$. Thus, the largest connected component is much larger than for $\mathrm{ER}_n(\lambda/n)$ with $\lambda < 1$.

Theorem 3.25 can be intuitively understood as follows. The connected component of a typical vertex is close to a branching process, so that it is with high probability

bounded since the expected value of its cluster will be close to $1/(1-\nu)$. Thus, the best way to obtain a large connected component is to start with a vertex with high weight w_i, and let all of its roughly w_i children be independent branching processes. Therefore, in expectation, each of these children is connected to another $1/(1-\nu)$ different vertices, leading to a cluster size of roughly $w_i/(1-\nu)$. This is clearly largest when $w_i = \max_{j\in[n]} w_j = \Delta$, leading to an intuitive explanation of Theorem 3.25.

Theorems 2.9 and 3.25 raise the question what the precise conditions for $|\mathcal{C}_{\max}|$ to be of order $\log n$ are. Intuitively, when $\Delta \gg \log n$, then $|\mathcal{C}_{\max}| = \Delta/(1-\nu)(1+o_{\mathbb{P}}(1))$, whereas if $\Delta = \Theta(\log n)$, then $|\mathcal{C}_{\max}| = \Theta_{\mathbb{P}}(\log n)$ as well. In [110], it was proved that $|\mathcal{C}_{\max}|/\log n$ converges in probability to a finite constant when $\nu < 1$ and the weights are i.i.d. with distribution function F with $\mathbb{E}[\mathrm{e}^{\alpha W}] < \infty$ for some $\alpha > 0$, i.e., *exponential tails* are sufficient.

The critical behavior of rank-1 random graphs. We next discuss the effect of inhomogeneity on the size of the largest connected components in the *critical* case. As it turns out, the behavior is rather different depending on whether $\mathbb{E}[W^3] < \infty$ or not.

Theorem 3.26 (The critical behavior with finite third moments)**.** *Fix the Norros-Reittu random graph with weights $\boldsymbol{w}(t) = \boldsymbol{w}(1+tn^{(\tau-3)(\tau-1)})$. Assume that $\nu = 1$, that the weight sequence $\boldsymbol{w}$ satisfies [I, Condition 6.4(a)-(c)], and further assume that*

$$\mathbb{E}[W_n] = \mathbb{E}[W] + o(n^{-1/3}), \qquad \mathbb{E}[W_n^2] = \mathbb{E}[W^2] + o(n^{-1/3}), \qquad \mathbb{E}[W_n^3] = \mathbb{E}[W^3] + o(1) \tag{3.6.4}$$

Let $(|\mathcal{C}_{(i)}(t)|)_{i\geq 1}$ denote the clusters of $\mathrm{NR}_n(\boldsymbol{w}(t))$ with $\boldsymbol{w}(t) = (1+tn^{-1/3})\boldsymbol{w}$, ordered in size. Then, as $n\to\infty$, for all $t\in\mathbb{R}$,

$$\big(n^{-2/3}|\mathcal{C}_{(i)}(t)|\big)_{i\geq 1} \xrightarrow{d} \big(\gamma_i^*(t)\big)_{i\geq 1}, \tag{3.6.5}$$

in the product topology, for some limiting random variables $\big(\gamma_i^*(t)\big)_{i\geq 1}$.

The limiting random variables $\big(\gamma_i^*(t)\big)_{i\geq 1}$ are, apart from a multiplication by a constant and a time-rescaling, equal to those for $\mathrm{ER}_n(\lambda/n)$ in the scaling window (see Theorem 5.7).

When $\mathbb{E}[W^{3-\varepsilon}] = \infty$ for some $\varepsilon > 0$, it turns out that the scaling of the largest critical cluster is rather different:

Theorem 3.27 (Weak convergence of the ordered critical clusters for $\tau\in(3,4)$)**.** *Fix the Norros-Reittu random graph with weights $\boldsymbol{w}(t) = \boldsymbol{w}(1+tn^{(\tau-3)(\tau-1)})$ defined in (6.1.14). Assume that $\nu = 1$ and that there exists a $\tau\in(3,4)$ and $0 < c_F < \infty$ such that*

$$\lim_{x\to\infty} x^{\tau-1}[1-F(x)] = c_F. \tag{3.6.6}$$

Let $(|\mathcal{C}_{(i)}(t)|)_{i\geq 1}$ denote the clusters of $\mathrm{NR}_n(\boldsymbol{w}(t))$, ordered in size. Then, as $n\to\infty$, for all $t\in\mathbb{R}$,

$$\big(n^{-(\tau-2)/(\tau-1)}|\mathcal{C}_{(i)}(t)|\big)_{i\geq 1} \xrightarrow{d} (\gamma_i(t))_{i\geq 1}, \tag{3.6.7}$$

in the product topology, for some non-degenerate limit $(\gamma_i(t))_{i\geq 1}$.

In this chapter, we have already seen that distances depend sensitively on the finiteness of $\mathbb{E}[W^2]$. Now we see that the critical behavior is rather different when $\mathbb{E}[W^3] < \infty$ or $\mathbb{E}[W^3] = \infty$. Interestingly, in the power-law case as described in (3.6.6), the size of the largest clusters grows like $n^{(\tau-2)/(\tau-1)}$, which is much smaller than the $n^{2/3}$ scaling when $\mathbb{E}[W^3] < \infty$. The proof of Theorems 3.26 and 3.27 also reveals that the structure of large critical clusters is quite different. When $\mathbb{E}[W^3] < \infty$, then the vertex with largest weight is in the largest connected component with vanishing probability. Therefore, the largest connected component arises by many attempts to create a large cluster, and each trial has roughly the same probability. This can be formulated as *power to the masses.* In the other hand, for weights $\boldsymbol{w}$ as in (6.1.14) for which (3.6.6) holds, the vertices with largest weight are with probability bounded away from 0 and 1 in the largest cluster, while a vertex with small weight is in the largest cluster with vanishing probability. Thus, to find the largest clusters, it suffices to explore the clusters of the high-weight vertices: *power to the wealthy!*

Fluctuations of distances in the finite-variance case. We continue by studying the fluctuations of the typical graph distance when $\mathbb{E}[W^2] < \infty$. We shall impose a slightly stronger condition on the distribution function F of W, namely, that there exists a $\tau > 3$ and $c > 0$ such that

$$1 - F(x) \leq cx^{-(\tau-1)}. \tag{3.6.8}$$

Equation (3.6.8) implies that the degrees have finite variance:

Exercise 3.33 (Finite variance degrees when (3.6.8) holds)**.** *Prove that* (3.6.8) *implies that* $\mathbb{E}[W^2] < \infty$. *Use this to prove that the degrees have uniformly bounded variance when* (3.6.8) *holds.*

Theorem 3.28 (Limit law for the typical graph distance in $\mathrm{CL}_n(\boldsymbol{w})$)**.** *Assume that* (3.6.8) *is satisfied, and let* $\nu = \mathbb{E}[W^2]/\mathbb{E}[W] > 1$. *For* $k \geq 1$, *define* $a_k = \lfloor \log_\nu k \rfloor - \log_\nu k \in (-1, 0]$. *Then, for* $\mathrm{CL}_n(\boldsymbol{w})$ *with* $\boldsymbol{w}$ *as in* (6.1.14), *there exist random variables* $(R_a)_{a\in(-1,0]}$ *with* $\limsup_{K\to\infty} \sup_{a\in(-1,0]} \mathbb{P}(|R_a| < K) = 1$ *such that, as* $n \to \infty$,

$$\mathbb{P}\big(H_n - \lfloor \log_\nu n \rfloor = k \mid H_n < \infty\big) = \mathbb{P}(R_{a_n} = k) + o(1), \qquad k \in \mathbb{Z}. \tag{3.6.9}$$

While Theorem 2.8 implies that, conditionally on $H_n < \infty$, $H_n/\log n \xrightarrow{\mathbb{P}} 1/\log \nu$, Theorem 3.28 implies that the fluctuations of H_n around $\log_\nu n$ remain uniformly bounded in probability.

The random variables $(R_a)_{a\in(-1,0]}$ can be determined in terms of the limit law in a branching process approximation of the neighborhoods of $\mathrm{CL}_n(\boldsymbol{w})$, and depend sensitively on a, which implies that although $\{H_n - \lfloor \log_\nu n \rfloor\}_{n=2}^\infty$ is a *tight* sequence of random variables, it does *not* weakly converges.

Exercise 3.34 (Tightness of centered typical graph distances in $\mathrm{CL}_n(\boldsymbol{w})$)**.** *Prove that, under the conditions of Theorem 3.28, and conditionally on* $H_n < \infty$, *the sequence* $\{H_n - \lfloor \log_\nu n \rfloor\}_{n=2}^\infty$ *is tight.*

Exercise 3.35 (Non-convergence of centered typical graph distances in $\mathrm{CL}_n(\boldsymbol{w})$)**.** *Prove that, under the conditions of Theorem 3.28, and conditionally on $H_n < \infty$, the sequence $H_n - \lfloor \log_\nu n \rfloor$ does not weakly converge when the distribution of R_a depends continuously on a and when there are $a, b \in (-1, 0]$ such that the distribution of R_a is not equal to the one of R_b.*

Exercise 3.36 (Extension Theorem 3.28 to $\mathrm{GRG}_n(\boldsymbol{w})$ and $\mathrm{NR}_n(\boldsymbol{w})$)**.** *Use Theorem 6.18 to prove that Theorem 3.28 holds verbatim for $\mathrm{GRG}_n(\boldsymbol{w})$ and $\mathrm{NR}_n(\boldsymbol{w})$ when (3.6.8) holds.*

3.7 Notes and discussion

Notes on Section 3.1. Theorem 3.2 is taken from [78], where the giant component is investigated for the configuration model. We explain its proof in detail in Section 4.1, where we also prove how the result for the configuration model in Theorem 4.1 can be used to prove Theorem 3.2.

Theorem 3.3 has a long history, and many versions of it have been proven in the literature. We refer the reader to [33, 35] for the Chung-Lu model, and [54] for its extensions to the Norros-Reittu model and the generalized random graph. Theorem 3.4 has also been proved in many versions, both fully as well as in partial forms, see [93], [33, 35], as well as the recent paper [42].

Notes on Section 3.3. As far as we are aware, the proof of Theorem 3.5 is new in the present context. Similar arguments have been used often though to prove lower bounds on distances in various situations.

Notes on Section 3.4. Proposition 3.10 appears first as [93, Proposition 3.1], where the connection between $\mathrm{NR}_n(\boldsymbol{w})$ and Poisson branching processes were first exploited to prove versions of Theorem 3.4.

Notes on Section 3.5. The path counting techniques in Proposition 3.17 are novel. Related proofs for the upper bound on H_n when $\nu < \infty$ often rely on branching process comparisons up to a generation $m = m_n \to \infty$.

Theorem 2.8 is a simplified version of [24, Theorem 3.14]. A first version of Theorem 2.8 was proved in [33, 35] for the expected degree random graph, in the case of *admissible* deterministic weights. We refer to [35, p. 94] for the definition of admissible degree sequences.

Theorem 3.4 for the expected degree random graph or Chung-Lu model is first proved in [33, 35], in the case of deterministic weights $w_i = c \cdot (i/n)^{-1/(\tau-1)}$, having average degree strictly greater than 1 and maximum degree m satisfying $\log m \gg \log n / \log \log n$. These restrictions were lifted in [46, Theorem 4.5.2]. Indeed, the bound on the average distance is not necessary, since, for $\tau \in (2, 3)$, $\nu = \infty$ and therefore the IRG is *always* supercritical. An upper bound as in Theorem 3.4 for the Norros-Reittu model with i.i.d. weights is proved in [93].

Notes on Section 3.6. Theorem 3.25 is [75, Corollary 4.4]. Theorem 3.26 is proved in [19], a related version with a different proof can be found in [111]. Theorem 3.27 is proved in [20]. Theorem 3.28 is proved in [54], both in the case of i.i.d. degrees as well as for deterministic weights under a mild further condition on the distribution function. Theorem 2.12 is a special case of [24, Theorem 3.16]. Even in the special case of $\mathrm{ER}_n(\lambda/n)$, it is new, and it negatively answers a question of Chung and Lu [32]. Related results for the configuration model, which also imply results for the generalized random graph, can be found in [56].

Chapter 4
The phase transition in the configuration model

Abstract

In this chapter, we investigate the phase transition in the configuration model by investigating its largest connected component.

Organization of this chapter. This chapter is organized as follows. In Section 4.1, we study when a giant component exists for the configuration model. In Section 4.2, we study when the configuration model is connected. In Section 4.3, we state further results in the configuration model, and in Section 4.4 we discuss similar results in some related models.

4.1 Phase transition in the configuration model

In this section, we investigate the connected components in the configuration model. Alike for the Erdős-Rényi random graph, we shall identify when the configuration model with high probability has a giant component. Again, this condition has the interpretation that an underlying branching process describing the exploration of a cluster has a strictly positive survival probability.

We start by recalling some notation from [I, Chapter 7]. We investigate the configuration model $\mathrm{CM}_n(\boldsymbol{d})$, where in most cases, the degrees $\boldsymbol{d} = (d_i)_{i\in[n]}$ are assumed to satisfy [I, Condition 7.7(a)-(b)], and sometimes also [I, Condition 7.7(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 uniformly chosen from $[n]$. Equivalently,

$$\mathbb{P}(D_n = k) = n_k/n, \tag{4.1.1}$$

where n_k denotes the number of vertices of degree k. For a graph G, we write $v_k(G)$ for the number of vertices of degree k in G, and $|E(G)|$ for the number of edges. The main result concerning the size and structure of the largest connected components of $\mathrm{CM}_n(\boldsymbol{d})$ is the following:

Theorem 4.1 (Phase transition in $\mathrm{CM}_n(\boldsymbol{d})$)**.** *Suppose that [I, Condition 7.7(a)-(b)] hold and consider the random graph $\mathrm{CM}_n(\boldsymbol{d})$, letting $n \to \infty$. 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 $\mathrm{CM}_n(\boldsymbol{d})$.*

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

$$\begin{aligned} |\mathcal{C}_{\max}|/n &\xrightarrow{\mathbb{P}} \zeta, \\ v_k(\mathcal{C}_{\max})/n &\xrightarrow{\mathbb{P}} p_k(1-\xi^k) \quad \text{for every } k \geq 0, \\ |E(\mathcal{C}_{\max})|/n &\xrightarrow{\mathbb{P}} \frac{1}{2}\mathbb{E}[D](1-\xi^2). \end{aligned}$$

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

(b) If $\nu = \mathbb{E}[D(D-1)]/\mathbb{E}[D] \leq 1$, *then* $|\mathcal{C}_{\max}|/n \xrightarrow{\mathbb{P}} 0$ *and* $|E(\mathcal{C}_{\max})|/n \xrightarrow{\mathbb{P}} 0$.

Reformulation in terms of branching processes. We start by interpreting the results in Theorem 4.1 in terms of *branching processes*. As it turns out, we can interpret ξ as the extinction probability of a branching process, and ζ as the survival probability of a related branching process. Similarly to the cluster exploration for $\mathrm{NR}_n(\boldsymbol{w})$, we shall see that the root of the branching process plays a special role. Indeed, the offspring distribution at the root is equal to $(p_k)_{k\geq 0}$, where $p_k = \mathbb{P}(D = k)$ is the asymptotic degree distribution. The offspring distribution of the individuals in the first and later generations is given by

$$g_k = \frac{(k+1)p_{k+1}}{\mathbb{E}[D]}. \tag{4.1.2}$$

We now heuristically explain this relation to branching processes by intuitively describing the exploration of a vertex chosen uniformly from the vertex set $[n]$. By definition, the probability that its degree is k equals $\mathbb{P}(D_n = k)$, which, by [I, Condition 7.7(a)], converges to $p_k = \mathbb{P}(D = k)$, for every $k \geq 1$. This explains the offspring of the root of our branching process approximation.

To describe the offspring of the direct neighbors of the root, we examine the degree of the vertex to which the first half-edge incident to the root is paired. By the uniform matching of half-edges, the probability that a vertex of degree k is chosen is proportional to k. Ignoring the fact that the root and one half-edge have already been chosen (which does have an effect on the number of available or free half-edges), the degree of the vertex incident to the chosen half-edge equals k with probability equal to $kp_k^{(n)}/\mathbb{E}[D_n]$, where $p_k^{(n)} = n_k/n$ is the proportion of vertices with degree k, and

$$\mathbb{E}[D_n] = \frac{1}{n}\sum_{i\in[n]} d_i = \frac{1}{n}\sum_{i\in[n]}\sum_{k=0}^{\infty} k\mathbb{1}_{\{d_i=k\}} = \sum_{k=0}^{\infty} kp_k^{(n)} \tag{4.1.3}$$

is the average degree in $\mathrm{CM}_n(\boldsymbol{d})$. Thus, $(kp_k^{(n)}/\mathbb{E}[D_n])_{k\geq 0}$ is a probability mass function. However, one of the half-edges is used up to connect to the root, so that, for a vertex incident to the root to have k offspring, it needs to connect its half-edge to a vertex having degree $k+1$. Therefore, the probability that the offspring of any of the direct neighbors of the root is equal to k equals

$$g_k^{(n)} = \frac{(k+1)p_{k+1}^{(n)}}{\mathbb{E}[D_n]}. \tag{4.1.4}$$

Thus, $(g_k^{(n)})_{k\geq 0}$ can be interpreted as the *forward degree of vertices in the cluster exploration*. When [I, Condition 7.7(a)-(b)] hold, then also $g_k^{(n)} \to g_k$, where $(g_k)_{k\geq 0}$ is defined in (4.1.2). As a result, we often refer to $(g_k)_{k\geq 0}$ as the *asymptotic forward degree distribution*.

The above heuristically argues that the number of vertices unequal to the root connected to any direct neighbor of the root has asymptotic law $(g_k)_{k\geq 0}$. However, every time we pair two half-edges, the number of *free* or *available* half-edges decreases by 2. Similarly to the depletion of points effect in the exploration of clusters for the Erdős-Rényi random graph $\mathrm{ER}_n(\lambda/n)$, the configuration model $\mathrm{CM}_n(\boldsymbol{d})$ suffers from a *depletion of points and half-edges* effect. Thus, by iteratively connecting half-edges in a breadth-first way, the offspring distribution changes along the way, which gives potential trouble. Luckily, the number of available half-edges that we start with equals $\ell_n - 1$, which is very large when [I, Condition 7.7(a)-(b)] hold, since then $\ell_n/n = \mathbb{E}[D_n]/n \to \mathbb{E}[D] > 0$. Thus, we can pair many half-edges before we start noticing that their number decreases. As a result, the degrees of different vertices in the exploration process is close to being i.i.d., leading to a branching process approximation. In terms of this branching process, we can interpret ζ in Theorem 4.1 as the survival probability of the above two-stage branching process, so that ζ satisfies

$$\zeta = \sum_{k=1}^{\infty} p_k(1-\xi^k), \tag{4.1.5}$$

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

$$\xi = \sum_{k=0}^{\infty} g_k \xi^k. \tag{4.1.6}$$

Clearly, $\xi = 1$ precisely when

$$\nu = \sum_{k\geq 0} k g_k \leq 1. \tag{4.1.7}$$

By (4.1.2), we can rewrite

$$\nu = \frac{1}{\mathbb{E}[D]} \sum_{k\geq 0} k(k+1)p_{k+1} = \mathbb{E}[D(D-1)]/\mathbb{E}[D], \tag{4.1.8}$$

which explains the condition on ν in Theorem 4.1(a). Further, to understand the asymptotics of $v_k(\mathcal{C}_{\max})$, we note that there are $n_k = np_k^{(n)} \approx np_k$ vertices with degree k. Each of the k direct neighbors of a vertex of degree k survives with probability close to $1-\xi$, so that the probability that at least one of them survives is close to $1-\xi^k$. When one of the neighbors of the vertex of degree k survives, the vertex itself is part of the giant component, which explains why $v_k(\mathcal{C}_{\max})/n \xrightarrow{\mathbb{P}} p_k(1-\xi^k)$. Finally, an edge consists of two half-edges, and an edge is part of the giant component precisely when one of the vertices incident to it is, which occurs with probability $1-\xi^2$. There are in total $\ell_n/2 = n\mathbb{E}[D_n]/2 \approx n\mathbb{E}[D]/2$ edges, which explains why $|E(\mathcal{C}_{\max})|/n \xrightarrow{\mathbb{P}} \frac{1}{2}\mathbb{E}[D](1-\xi^2)$. Therefore, all results in Theorem 4.1 have a simple explanation in terms

of the branching process approximation of the connected component for $\mathrm{CM}_n(\boldsymbol{d})$ of a uniform vertex in $[n]$.

Reformulation in terms of generating functions. We next reformulate the results in terms of generating functions, which play a crucial role throughout our proof. Let

$$G_D(x) = \sum_{k=0}^{\infty} p_k x^k = \mathbb{E}[x^D] \tag{4.1.9}$$

be the probability generating function of the probability distribution $(p_k)_{k\geq 1}$. Recall that, for a non-negative random variable D, the random variable D^* denotes its size-biased distribution. Define further

$$G^\star_D(x) = \mathbb{E}[x^{D^*}] = \sum_{k=1}^{\infty} g_k x^k = G'_D(x)/G'_D(1), \tag{4.1.10}$$

$$H(x) = \mathbb{E}[D]x\big(x - G^\star_D(x)\big). \tag{4.1.11}$$

Note that $G^\star_D(1) = 1$, and thus $H(0) = H(1) = 0$. Note also that

$$H'(1) = 2\mathbb{E}[D] - \sum_k k^2 p_k = \mathbb{E}[2D - D^2] = -\mathbb{E}[D(D-2)] \tag{4.1.12}$$

For further properties of $x \mapsto H(x)$, see Lemma 4.6 below. We conclude that if $\mathbb{E}[D(D-2)] = \sum_k k(k-2)p_k > 0$ and if $g_1 > 0$, then there is a unique $\xi \in (0,1)$ such that $H(\xi) = 0$, or equivalently $G^\star_D(\xi) = \xi$, so that indeed ξ is the extinction probability of the branching process with offspring distribution $(g_k)_{k\geq 0}$. When $g_1 = 0$, instead, $\xi = 0$ is the unique solution in $[0,1)$ of $H(\xi) = 0$. The functions $x \mapsto H(x)$ and $x \mapsto G^\star_D(x)$ play a crucial role in our analysis of the problem.

We prove Theorem 4.1 in Section 4.1.2 below. We now remark upon the result.

The condition $\mathbb{P}(D = 2) = p_2 < 1$. Because isolated vertices do not matter, without loss of generality, we may assume that $p_0 = 0$. The case $p_2 = 1$, for which $\nu = 1$ is quite exceptional. In this case, $H(x) = 0$ for all x. We give three examples showing that then quite different behaviors are possible.

Our first example is when $d_i = 2$ for all $i \in [n]$, so we are studying a random 2-regular graph. In this case, the components are cycles and the distribution of cycle lengths in $\mathrm{CM}_n(\boldsymbol{d})$ is given by the Ewen's sampling formula $\mathrm{ESF}(1/2)$, see e.g. [11]. This implies that $|\mathcal{C}_{\max}|/n$ converges in distribution to a non-degenerate distribution on $[0,1]$ and not to any constant [11, Lemma 5.7]. Moreover, the same is true for $|\mathcal{C}_{(2)}|/n$ (and for $|\mathcal{C}_{(3)}|/n$,...), so in this case there are several large components. To intuitively see this result, we note that in the exploration of a cluster we start with one vertex with two half-edges. When pairing a half-edge, it connects to a vertex that again has two half-edges. Therefore, the number of half-edges to be paired is always equal to 2, up to the moment when the cycle is closed, and the cluster is completed. When there are $m = \alpha n$ free half-edges left, the probability of closing up

the cycle equals $1/m = 1/(\alpha n)$, and, thus, the time this takes is of order n. A slight extension of this reasoning shows that the time it takes to close a cycle is nT_n, where T_n converges to a limiting non-degenerate random variable:

Exercise 4.1 (Cluster size of vertex 1 in a 2-regular graph)**.** *Let $n_2 = n$, and let $\mathcal{C}(1)$ denote the cluster size of vertex 1. Show that*

$$|\mathcal{C}(1)|/n \xrightarrow{d} T, \tag{4.1.13}$$

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 $n_1 \to \infty$ be such that $n_1/n \to 0$, and $n_2 = n - n_1$. In this case, components can either be cycles, or strings of vertices with degree 2 terminated with two vertices with degree 1. When $n_1 \to \infty$, it is more likely to terminate a long string of vertices of degree 2 by a vertex of degree 1 than by closing the cycle, as for the latter we need to pair to a *unique* half-edge, while for the former, we have n_1 choices. Therefore, it is easy to s ee that this implies that $|\mathcal{C}_{\max}| = o_{\mathbb{P}}(n)$:

Exercise 4.2 (Cluster size in a 2-regular graph with some degree-1 vertices)**.** *Let $n_1 \to \infty$ with $n_1/n \to 0$, and $n_2 = n - n_1$. Let $\mathcal{C}(1)$ denote the cluster size of vertex 1. Show that*

$$|\mathcal{C}(1)|/n \xrightarrow{\mathbb{P}} 0. \tag{4.1.14}$$

Our third example with $p_2 = 1$ is obtained by instead adding a small number of vertices of degree 4 (i.e., $n_4 \to \infty$ such that $n_4/n \to 0$, and $n_2 = n - n_4$.) We can regard each vertex of degree 4 as two vertices of degree 2 that have been identified. Therefore, to obtain $\mathrm{CM}_n(\boldsymbol{d})$ 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 4.3 (Cluster size in a 2-regular graph with some degree-4 vertices)**.** *Let $n_4 \to \infty$ with $n_4/n \to 0$, and $n_2 = n - n_4$. Let $\mathcal{C}(1)$ denote the cluster size of vertex 1. Show that*

$$|\mathcal{C}(1)|/n \xrightarrow{\mathbb{P}} 1. \tag{4.1.15}$$

We conclude that the case where $p_2 = \mathbb{P}(D = 2) = 1$ is quite sensitive to the precise properties of the degree structure that are not captured by the limiting distribution $(p_k)_{k\geq 1}$ only. In the sequel, we shall ignore the case where $p_2 = 1$.

Reduction to the case where $\mathbb{P}(D = 1) = p_1 > 0$. In our proof, it is convenient to assume that $p_1 = \mathbb{P}(D = 1) > 0$. The extinction probability $\xi = 0$ and the survival probability $\zeta = 1$ when $p_1 = 0$, which causes technical difficulties in the proof. We now explain how we can reduce the case where $p_1 = 0$ to the case where $p_1 > 0$.

Let $d_{\min} = \min\{k\colon p_k > 0\}$ be the minimum of the support of the asymptotic degree distribution D. Fix $\varepsilon > 0$, and assume that $\varepsilon < p_k$. Consider the configuration model with $\tilde{n} = n + 2d_{\min}\varepsilon n$, and degree sequence $\tilde{\boldsymbol{d}} = (\tilde{d}_i)_{i\in[n]}$ with $\tilde{n}_k = n_k$ for all $k > d_{\min}$, $\tilde{n}_{d_{\min}} = n_{d_{\min}} - \varepsilon n$, $\tilde{n}_1 = 2d_{\min}\varepsilon n$. This configuration model can be obtained from $\mathrm{CM}_n(\boldsymbol{d})$ by replacing εn vertices of degree $d_{\min}$ by $d_{\min}$ vertices having degree 1, as if we have 'forgotten' that these vertices are actually equal.

Clearly, $\mathrm{CM}_n(\boldsymbol{d})$ can be retrieved by identifying εn collections of $d_{\min}$ vertices of degree 1 to a single vertex of degree $d_{\min}$. When $\boldsymbol{d}$ satisfies [I, Condition 7.7(a)-(b)], then so does $\tilde{\boldsymbol{d}}$ with limiting degree distribution $\tilde{p}_1 = 2d_{\min}\varepsilon/$, $\tilde{p}_{d_{\min}} = (p_{d_{\min}} - \varepsilon)/(1 + 2d_{\min}\varepsilon)$, $\tilde{p}_k = p_k/(1 + 2d_{\min}\varepsilon)$ for all $k > d_{\min}$. The above procedure clearly makes $|\mathcal{C}_{\max}|$ smaller. Further, with ζ_ε denoting the limit of $|\mathcal{C}_{\max}|/\tilde{n}$ for $\tilde{\boldsymbol{d}}$, we have that $\zeta_\varepsilon \to 1$ as $\varepsilon \downarrow 0$. As a result, Theorem 4.1 for $\zeta = 1, \xi = 0$ follows from Theorem 4.1 with $p_1 > 0$, for which $\zeta < 1$ and $\xi > 0$. In the remainder of the proof, we therefore without loss of generality assume that $\xi > 0$ and $\zeta < 1$.

Organization of the proof of Theorem 4.1. Theorem 4.1 is proved using a clever *randomization scheme* to explore the connected components one by one. This construction is explained terms of a simple continuous-time algorithm in Section 4.1.1 below. The algorithm describes the number of vertices of given degrees that have been found, as well as the total number of unpaired half-edges, at time $t > 0$. It is proved that, when $n \to \infty$, these quantities all converge in probability to *deterministic functions* described in terms of the functions $x \mapsto H(x)$ and $x \mapsto G^\star_D(x)$ above. In particular, the number of unpaired half-edges is given in terms of $x \mapsto H(x)$, so that the first zero of this function gives the size of the giant component. In Section 4.1.2, the algorithm is analyzed by showing that when $\zeta > 0$, after a short initial period of exploring small clusters, the giant component is found, and the exploration explores it completely, after which no large component is left. When $\zeta = 0$, instead, only small clusters are found. A crucial aspect in the proof resides in how to deal with the depletion of points and half-edges effect.

4.1.1 Finding the largest component

The components of an arbitrary finite graph or multigraph can be found by the following standard procedure. Pick an arbitrary vertex v and determine the component of v as follows: include all the neighbors of v in an arbitrary order; then add in the neighbors of the neighbors, and so on, until no more vertices can be added. The vertices included until this moment form the component of v. If there are still vertices left in the graph, then pick any such vertex w, and repeat the above to determine the second component (the component of vertex w). Carry on in this manner until all the components have been found.

The same result can be more conveniently obtained in the following way. Regard each edge as consisting of two half-edges, each half-edge having one endpoint. We will label the vertices as sleeping or awake (= used) and the half-edges as sleeping, active or dead; the sleeping and active half-edges are also called living. We start with all vertices and half-edges sleeping. Pick a vertex and label its half-edges as active.

Then take any active half-edge, say x and find its partner y in the graph; label these two half-edges as dead. Further, if the endpoint of y is sleeping, label it as awake and all other half-edges of the vertex incident to y as active. Repeat as long as there are active half-edge. When there is no active half-edge left, we have obtained the first component. Then start again with another vertex until all components are found.

We apply this algorithm to $\mathrm{CM}_n(\boldsymbol{d})$ with a given degree sequence, revealing its edges during the process. We thus initially only observe the vertex degrees and the half-edges, but not how they are joined to form edges. Hence, each time we need a partner of an half-edge, it is uniformly distributed over all other living half-edges, with the understanding that the dead half-edges are the ones that are already paired into edges. It is here that we are using the specific structure of the configuration model, which simplifies the analysis substantially.

We make the random choices of finding a partner to the edges by associating i.i.d. random maximal lifetimes τ_x to the half-edge x, where τ_x has an Exp(1) distribution. We interpret these lifetimes as *clocks*, and changes in our exploration process only occur when a clock of a half-edge rings. In other words, each half-edge dies spontaneously at rate 1 (unless killed earlier). Each time we need to find the partner of a half-edge x, we then wait until the next living half-edge $\neq x$ dies and take that one. This process in continuous-time can be formulated as an *algorithm*, constructing $\mathrm{CM}_n(\boldsymbol{d})$ and exploring its components simultaneously, as follows. Recall that we start with all vertices and half-edges sleeping. The exploration is then formalized in the following three steps:

Step 1 When there is no active half-edge (as in the beginning), select a sleeping vertex and declare it awake and all its half-edges active. For definiteness, we choose the vertex by choosing a half-edge uniformly at random among all sleeping half-edges. When there is no sleeping half-edge left, the process stops; the remaining sleeping vertices are all isolated and we have explored all other components.

Step 2 Pick an active half-edge (which one does not matter) and kill it, i.e., change its status to dead.

Step 3 Wait until the next half-edge dies (spontaneously, as a result of its clock ringing). This half-edge is joined to the one killed in the previous step Step 2 to form an edge of the graph. When the vertex incident to it is sleeping, we change this vertex to awake and all other half-edges incident to it to active. Repeat from Step 1.

The above randomized algorithm is such that components are created between the successive times Step 1 is performed, where we say that Step 1 is performed when there is no active half-edge and, as a result, a new vertex is chosen.

The vertices in the component created during one of these intervals between the successive times Step 1 is performed are the vertices that are awakened during the interval. Note also that a component is completed and Step 1 is performed exactly when the number of active half-edges is 0 and a half-edge dies at a vertex where all other half-edges (if any) are dead. In the next section, we investigate the behavior

of the key characteristics of the algorithm, such as the number of sleeping half-edges and the number of sleeping vertices of a given degree.

4.1.2 Analysis of the algorithm for $\mathrm{CM}_n(\boldsymbol{d})$

We start by introducing the key characteristics of the exploration algorithm. Let $S(t)$ and $A(t)$ be the numbers of sleeping and active half-edges, respectively, at time t, and let

$$L(t) = S(t) + A(t) \tag{4.1.16}$$

be the number of living half-edges. For definiteness, we define these random functions to be right-continuous.

Let us first look at $L(t)$. We start with ℓ_n half-edges, all sleeping and thus living, but we immediately perform Step 1 and Step 2 and kill one of them. Thus, $L(0) = \ell_n - 1$. In the sequel, as soon as a living half-edge dies, we perform Step 3 and then (instantly) either Step 2 or both Step 1 and Step 2. Since Step 1 does not change the number of living half-edges while Step 2 and Step 3 each decrease it by 1, the total result is that $L(t)$ is decreased by 2 each time one of the living half-edges dies, except when the last living one dies and the process terminates. Because of this simple dynamics of $t \mapsto L(t)$, we can give sharp asymptotics of $L(t)$ when $n \to \infty$:

Proposition 4.2 (The number of sleeping half-edges)**.** *As $n \to \infty$, for any $t_0 \geq 0$ fixed,*

$$\sup_{0 \leq t \leq t_0} |n^{-1} L(t) - \mathbb{E}[D_n] \mathrm{e}^{-2t}| \xrightarrow{\mathbb{P}} 0. \tag{4.1.17}$$

Proof. The process $t \mapsto L(t)$ satisfies $L(0) = \ell_n - 1$, and it degreases by 2 at rate $L(t)$. As a result, it is closely related to a *death process*. We study such processes in the following lemma:

Lemma 4.3 (Asymptotics of death processes)**.** *Let $d, \gamma > 0$ be given and let $(N^{(x)}(t))_{t \geq 0}$ be a Markov process such that $N^{(x)}(t) = x$ a.s., and the dynamics of $t \mapsto (N^{(x)}(t))_{t \geq 0}$ is such that when it is in position y, then it jumps down by d at rate γy. In other words, the waiting time until the next event is $\mathrm{Exp}(1/\gamma y)$ and each jump is of size d downwards. Then, for every $t_0 \geq 0$,*

$$\mathbb{E}\Big[\sup_{t \leq t_0} \big| N^{(x)}(t) - \mathrm{e}^{-\gamma d t} x \big|^2 \Big] \leq 8d(\mathrm{e}^{\gamma d t_0} - 1)x + 8d^2. \tag{4.1.18}$$

Proof. The proof follows by distinguishing several cases. First assume that $d = 1$ and that x is an integer. In this case, the process is a standard pure death process taking the values $x, x-1, x-2, \ldots, 0$, describing the number of particles alive when the particles die independently at rate $\gamma > 0$. As is well-known, and easily seen by regarding $N^{(x)}(t)$ as the sum of x independent copies of the process $N^{(1)}(t)$, the process $(\mathrm{e}^{\gamma t} N^{(x)}(t))_{t \geq 0}$, is a *martingale* starting in x. Furthermore, for every $t \geq 0$, the random variable $N^{(x)}(t)$ has a $\mathsf{Bin}(x, \mathrm{e}^{-\gamma t})$ distribution, since each of the x particles has a probability of dying before time t of $\mathrm{e}^{-\gamma t}$, and the different particles die independently.

Hence, by Doob's inequality,

$$\mathbb{E}\Big[\sup_{t\le t_0}\big|N^{(x)}(t)-\mathrm{e}^{-\gamma t}x\big|^2\Big] \le \mathbb{E}\Big[\sup_{t\le t_0}\big|\mathrm{e}^{\gamma t}N^{(x)}(t)-x\big|^2\Big] \le 4\mathbb{E}\Big[\big(\mathrm{e}^{\gamma t}N^{(x)}(t_0)-x\big)^2\Big]$$
$$= 4\mathrm{e}^{2\gamma t}\mathrm{Var}(N^{(x)}(t_0)) \le 4(\mathrm{e}^{\gamma t_0}-1)x. \tag{4.1.19}$$

This proves the claim for x being integer.

Next, still assume $d=1$, but let $x>0$ be arbitrary. We can couple the two processes $\big(N^{(x)}(t)\big)_{t\ge 0}$ and $\big(N^{(\lfloor x\rfloor)}(t)\big)_{t\ge 0}$ with different initial values such that whenever the smaller one jumps by 1, so does the other. This coupling keeps

$$|N^{(x)}(t)-N^{(\lfloor x\rfloor)}(t)|<1 \tag{4.1.20}$$

for all $t\ge 0$, and thus,

$$\sup_{t\le t_0}\big|N^{(\lfloor x\rfloor)}(t)-\mathrm{e}^{-\gamma t}\lfloor x\rfloor\big| \le \sup_{t\le t_0}\big|N^{(x)}(t)-\mathrm{e}^{-\gamma t}x\big|+2, \tag{4.1.21}$$

so that by (4.1.19), in turn,

$$\mathbb{E}\Big[\sup_{t\le t_0}\big|N^{(x)}(t)-\mathrm{e}^{-\gamma t}x\big|^2\Big] \le 8(\mathrm{e}^{\gamma t_0}-1)x+8. \tag{4.1.22}$$

Finally, for a general $d>0$, we observe that $N^{(x)}(t)/d$ is a process of the same type with the parameters (γ,d,x) replaced by $(\gamma d,1,x/d)$, and the general result follows from (4.1.22) and (4.1.19). □

The proof of Proposition 4.2 follows from Lemma 4.3 with $d=2, x=(\ell_n-1)=n\mathbb{E}[D_n]-1$ and $\gamma=1$. □

We continue by considering the sleeping half-edges $S(t)$. Let $V_k(t)$ be the number of sleeping vertices of degree k at time t, so that

$$S(t)=\sum_{k=1}^{\infty}kV_k(t). \tag{4.1.23}$$

Note that Step 2 does not affect sleeping half-edges, and that Step 3 implies that each sleeping vertex of degree k is eliminated (i.e., awakened) with intensity k, independently of all other vertices. There are also some sleeping vertices eliminated by Step 1, though, which complicates the dynamics of $t\mapsto V_k(t)$. It is here that the depletion of points and half-edges effect enters the analysis of the component structure of $\mathrm{CM}_n(\boldsymbol{d})$.

We first ignore the effect of Step 1 by letting $\widetilde{V}_k(t)$ be the number of vertices of degree k such that all its half-edges have maximal lifetimes $\tau_x>t$. Thus, none of its k half-edges would have died spontaneously up to time t, assuming they all escaped Step 1. It is reasonable to ignore the effect of Step 1 in the leading order, as we perform Step 1 until we hit the giant component, and then it takes a long time to find the entire giant component. When $\zeta>0$, the number of times we perform Step 1 until we find the giant component will be small, as each time we have a strictly

positive probability of choosing a vertex in the giant component. Thus, intuitively, we expect the difference between $V_k(t)$ and $\widetilde{V}_k(t)$ to be insignificant.

For a given half-edge, we call the half-edges incident to the same vertex its *brother half-edges*. Let further

$$\widetilde{S}(t) = \sum_k k\widetilde{V}_k(t) \tag{4.1.24}$$

denote the number of half-edges whose brother half-edges have escaped spontaneous death up to time t.

Recall the functions G_D, $G_D^\star$ from (4.1.9)–(4.1.10), and define

$$h(x) = x\mathbb{E}[D]G_D^\star(x). \tag{4.1.25}$$

Then, we can identify the asymptotics of $(\widetilde{V}_k(t))_{t\geq 0}$ in a similar way as in Proposition 4.2:

Lemma 4.4 (The number of sleeping vertices)**.** *Assume that [I, Condition 7.7(a)-(b)] hold. Then, as $n\to\infty$, for any $t_0\geq 0$ fixed*

$$\sup_{t\leq t_0} |n^{-1}\widetilde{V}_k(t) - p_k\mathrm{e}^{-kt}| \xrightarrow{\mathbb{P}} 0 \tag{4.1.26}$$

for every $k\geq 0$ and

$$\sup_{t\leq t_0} |n^{-1}\sum_{k=0}^{\infty}\widetilde{V}_k(t) - G_D(\mathrm{e}^{-t})| \xrightarrow{\mathbb{P}} 0, \tag{4.1.27}$$

$$\sup_{t\leq t_0} |n^{-1}\widetilde{S}(t) - h(\mathrm{e}^{-t})| \xrightarrow{\mathbb{P}} 0. \tag{4.1.28}$$

Proof. The statement (4.1.26) again follows from Lemma 4.3, now with $\gamma = k$, $x = n_k$ and $d = 1$. The case $k = 0$ is trivial, with $\widetilde{V}_0(t) = n_0$ for all t. We can replace $p_k^{(n)}$ by p_k by [I, Condition 7.7(a)].

By [I, Condition 7.7(b)], the sequence of random variables $(D_n)_{n\geq 1}$ is uniformly integrable, which means that for every $\varepsilon > 0$ there exists $K < \infty$ such that for all n, $\sum_{k>K} kn_k/n = \mathbb{E}[D_n|D_n > k] < \varepsilon$. We may further assume (or deduce from Fatou's inequality) that $\sum_{k>K} kp_k < \varepsilon$, and obtain by (4.1.26) that, **whp**,

$$\begin{aligned}
\sup_{t\leq t_0} |n^{-1}\widetilde{S}(t) - h(\mathrm{e}^{-t})| &= \sup_{t\leq t_0}\left|\sum_{k=1}^{\infty} k(n^{-1}\widetilde{V}_k(t) - p_k\mathrm{e}^{-kt})\right| \\
&\leq \sum_{k=1}^{K} k\sup_{t\leq t_0}|n^{-1}\widetilde{V}_k(t) - p_k\mathrm{e}^{-kt}| + \sum_{k>K} k\left(\frac{n_k}{n} + p_k\right) \\
&\leq \varepsilon + \varepsilon + \varepsilon,
\end{aligned}$$

proving (4.1.28). An almost identical argument yields (4.1.27). □

Remarkably, the difference between $S(t)$ and $\widetilde{S}(t)$ is easily estimated. The following result can be viewed as the key to why this approach works. Indeed, it gives a *uniform* upper bound on the difference due to the application of Step 1:

Lemma 4.5 (Effect of Step 1). *If* $\Delta := \max_{i\in[n]} d_i$ *is the maximum degree of* $\mathrm{CM}_n(\boldsymbol{d})$, *then*

$$0 \le \widetilde{S}(t) - S(t) < \sup_{0\le s\le t} (\widetilde{S}(s) - L(s)) + \Delta. \tag{4.1.29}$$

Proof. Clearly, $V_k(t) \le \widetilde{V}_k(t)$, and thus $S(t) \le \widetilde{S}(t)$. Furthermore, $\widetilde{S}(t) - S(t)$ increases only as a result of Step 1. Indeed, Step 1 acts to guarantee that $A(t) = L(t) - S(t) \ge 0$, and is only performed when $A(t) = 0$.

If Step 1 is performed at time t and a vertex of degree $j > 0$ is awakened, then Step 2 applies instantly and we have $A(t) = j - 1 < \Delta$, and consequently

$$\widetilde{S}(t) - S(t) = \widetilde{S}(t) - L(t) + A(t) < \widetilde{S}(t) - L(t) + \Delta. \tag{4.1.30}$$

Furthermore, $\widetilde{S}(t) - S(t)$ is never changed by Step 2 and either unchanged or decreased by Step 3. Hence, $\widetilde{S}(t) - S(t)$ does not increase until the next time Step 1 is performed. Consequently, for any time t, if s was the last time before (or equal to) t that Step 1 was performed, then $\widetilde{S}(t) - S(t) \le \widetilde{S}(s) - S(s)$, and the result follows by (4.1.30). □

Recall that $A(t) = L(t) - S(t)$ denotes the number of awakened vertices and let

$$\widetilde{A}(t) = L(t) - \widetilde{S}(t) = A(t) - (\widetilde{S}(t) - S(t)) \tag{4.1.31}$$

denote the number of awakened vertices ignoring the effect of Step 1. Thus, $\widetilde{A}(t) \le A(t)$ since $S(t) \le \widetilde{S}(t)$.

Then, by Lemmas 4.2 and 4.4 (and (4.1.11)), for any $t_0 \ge 0$,

$$\sup_{t\le t_0} |n^{-1}\widetilde{A}(t) - H(\mathrm{e}^{-t})| \xrightarrow{\mathbb{P}} 0. \tag{4.1.32}$$

Lemma 4.5 can be rewritten as

$$0 \le \widetilde{S}(t) - S(t) < -\inf_{s\le t} \widetilde{A}(s) + \Delta. \tag{4.1.33}$$

By (4.1.31) and (4.1.33),

$$\widetilde{A}(t) \le A(t) < \widetilde{A}(t) - \inf_{s\le t} \widetilde{A}(s) + \Delta, \tag{4.1.34}$$

which, perhaps, illuminates the relation between $A(t)$ and $\widetilde{A}(t)$. Now, the function $t \mapsto H(\mathrm{e}^{-t})$, which acts as the limit of $\widetilde{A}(t)$, is strictly positive in $(0, -\log\xi)$ and $H(1) = H(\xi) = 0$. Therefore, we expect $\widetilde{A}(t)$ to be positive for $t \in (0, -\log\xi)$, and, if so, $\inf_{s\le t} \widetilde{A}(s) = 0$. The idea is to continue our algorithm in Step 1-Step 3 until the giant component has been found, which implies that $A(t) > 0$ for the time of exploration of the giant component, and $A(t) = 0$ for the first time when we have completed the exploration of the giant component, which is $t = -\log\xi$. Thus, the term $\inf_{s\le t} \widetilde{A}(s)$ in (4.1.34) ought to be negligible. When [I, Condition 7.7(a)-(b)] hold, we further have that $\Delta = o(n)$, so that one can expect $\widetilde{A}(t)$ to be a good approximation of $A(t)$. The remainder of the proof makes this intuition precise. We start by summarizing the properties of $x \mapsto H(x)$ that we rely upon:

Lemma 4.6 (Properties of $x \mapsto H(x)$)**.** *Suppose that [I, Condition 7.7(a)-(b)] hold and let $H(x)$ be given by (4.1.11).*

(i) If $\nu = \mathbb{E}[D(D-1)]/\mathbb{E}[D] > 1$ and $p_1 > 0$, then there is a unique $\xi \in (0,1)$ such that $H(\xi) = 0$. Moreover, $H(x) < 0$ for all $x \in (0,\xi)$ and $H(x) > 0$ for all $x \in (\xi, 1)$.

(ii) If $\nu = \mathbb{E}[D(D-1)]/\mathbb{E}[D] \leq 1$, then $H(x) < 0$ for all $x \in (0,1)$.

Proof. As remarked earlier, $H(0) = H(1) = 0$ and $H'(1) = -\mathbb{E}[D(D-2)]$. Furthermore, if we define $\phi(x) := H(x)/x$, then $\phi(x) = \mathbb{E}[D](x - G^\star_D(x))$ is a concave function on $(0,1]$, and it is strictly concave unless $p_k = 0$ for all $k \geq 3$, in which case $H'(1) = -\mathbb{E}[D(D-2)] = p_1 > 0$. Indeed, $p_1 + p_2 = 1$ when $p_k = 0$ for all $k \geq 3$. Since we assume that $p_2 < 1$, we thus obtain that $p_1 > 0$ in this case.

In case (ii), we thus have that ϕ is concave and $\phi'(1) = H'(1) - H(1) \geq 0$, with either the concavity or the inequality strict, and thus $\phi'(x) > 0$ for all $x \in (0,1)$, whence $\phi(x) < \phi(1) = 0$ for $x \in (0,1)$.

In case (i), $H'(1) < 0$, and thus $H(x) > 0$ for x close to 1. Further, when $p_1 > 0$, $H'(0) = -h'(0) = -p_1 < 0$, and thus $H(x) \leq 0$ for x close to 0. Hence, there is at least one $\xi \in (0,1)$ with $H(\xi) = 0$, and since $H(x)/x$ is strictly concave and also $H(1) = 0$, there is at most one such ξ and the result follows. □

Now we are in the position to complete the proof of Theorem 4.1 in the following section.

4.1.3 Proof of Theorem 4.1

We start with the proof of Theorem 4.1(i). Let ξ be the zero of H given by Lemma 4.6(i) and let $\theta = -\log \xi$. Then, by Lemma 4.6, $H(\mathrm{e}^{-t}) > 0$ for $0 < t < \theta$, and thus $\inf_{t \leq \theta} H(\mathrm{e}^{-t}) = 0$. Consequently, (4.1.32) implies

$$n^{-1} \inf_{t \leq \theta} \widetilde{A}(t) = \inf_{t \leq \theta} n^{-1} \widetilde{A}(t) - \inf_{t \leq \theta} H(\mathrm{e}^{-t}) \xrightarrow{\mathbb{P}} 0 \tag{4.1.35}$$

Further, by [I, Condition 7.7(b)], $\Delta = o(n)$, and thus $n^{-1}\Delta \to 0$. Consequently, (4.1.33) and (4.1.35) yield

$$\sup_{t \leq \theta} n^{-1} |A(t) - \widetilde{A}(t)| = \sup_{t \leq \theta} n^{-1} |\widetilde{S}(t) - S(t)| \xrightarrow{\mathbb{P}} 0. \tag{4.1.36}$$

Thus, by (4.1.32),

$$\sup_{t \leq \theta} |n^{-1} A(t) - H(\mathrm{e}^{-t})| \xrightarrow{\mathbb{P}} 0. \tag{4.1.37}$$

Let $0 < \varepsilon < \theta/2$. Since $H(\mathrm{e}^{-t}) > 0$ on the compact interval $[\varepsilon, \theta - \varepsilon]$, (4.1.37) implies that $A(t)$ remains **whp** positive on $[\varepsilon, \theta - \varepsilon]$, and thus no new component is started during this interval.

On the other hand, again by Lemma 4.6(i), $H(\mathrm{e}^{-(\theta+\varepsilon)}) < 0$ and (4.1.32) implies that $n^{-1}\widetilde{A}(\theta+\varepsilon) \xrightarrow{\mathbb{P}} H(\mathrm{e}^{-(\theta+\varepsilon)})$, while $A(\theta+\varepsilon) \geq 0$. Thus, with $\delta = |H(\mathrm{e}^{-\theta-\varepsilon})|/2 > 0$, **whp**

$$\widetilde{S}(\theta+\varepsilon) - S(\theta+\varepsilon) = A(\theta+\varepsilon) - \widetilde{A}(\theta+\varepsilon) \geq -\widetilde{A}(\theta+\varepsilon) > n\delta, \tag{4.1.38}$$

while (4.1.36) yields that $\widetilde{S}(\theta) - S(\theta) < n\delta$ **whp**. Consequently, **whp** $\widetilde{S}(\theta+\varepsilon) - S(\theta+\varepsilon) > \widetilde{S}(\theta) - S(\theta)$, so **whp** Step 1 is performed between the times θ and $\theta+\varepsilon$.

Let T_1 be the last time Step 1 was performed before time $\theta/2$. Let T_2 be the next time Step 1 is performed (by convention, $T_2 = \infty$ if such a time does not exist). We have shown that for any $\varepsilon > 0$, and **whp** $0 \leq T_1 \leq \varepsilon$ and $\theta - \varepsilon \leq T_2 \leq \theta + \varepsilon$. In other words, $T_1 \xrightarrow{\mathbb{P}} 0$ and $T_2 \xrightarrow{\mathbb{P}} \theta$. We conclude that we have found one component that is explored between time $T_1 \xrightarrow{\mathbb{P}} 0$ and time $T_2 \xrightarrow{\mathbb{P}} \theta$. This is our candidate for the giant component, and we continue to study its properties, i.e., its size, its number of edges and its number of vertices of degree k. These properties are stated separately in the next lemma, so that we are able to reuse them later on:

Proposition 4.7 (Cluster properties)**.** *Let T_1^* and T_2^* be two random times when Step 1 is performed, with $T_1^* \leq T_2^*$, and assume that $T_1^* \xrightarrow{\mathbb{P}} t_1$ and $T_2^* \xrightarrow{\mathbb{P}} t_2$ where $0 \leq t_1 \leq t_2 \leq \theta < \infty$. If $\mathcal{C}^*$ is the union of all components explored between T_1^* and T_2^*, then*

$$v_k(\mathcal{C}^*)/n \xrightarrow{\mathbb{P}} p_k(\mathrm{e}^{-kt_1} - \mathrm{e}^{-kt_2}), \quad k \geq 0, \tag{4.1.39}$$

$$|\mathcal{C}^*|/n \xrightarrow{\mathbb{P}} G_D(\mathrm{e}^{-t_1}) - G_D(\mathrm{e}^{-t_2}), \tag{4.1.40}$$

$$|E(\mathcal{C}^*)|/n \xrightarrow{\mathbb{P}} \frac{1}{2}h(\mathrm{e}^{-t_1}) - \frac{1}{2}h(\mathrm{e}^{-t_2}). \tag{4.1.41}$$

In particular, if $t_1 = t_2$, then $|\mathcal{C}^*|/n \xrightarrow{\mathbb{P}} 0$ and $|E(\mathcal{C}^*)| \xrightarrow{\mathbb{P}} 0$.

We apply Proposition 4.7 to $T_1 = o_{\mathbb{P}}(1)$ and T_2, where $T_2 = \theta + o_{\mathbb{P}}(1)$. We can identify the values of the above constants:

Exercise 4.4 (Limiting constants)**.** *Prove that for $t_1 = 0$ and $t_2 = \theta$, $\mathrm{e}^{-kt_1} = 1, \mathrm{e}^{-kt_2} = \xi$, $G_D(\mathrm{e}^{-t_1}) = 1$, $G_D(\mathrm{e}^{-t_2}) = 1 - \zeta$, $h(\mathrm{e}^{-t_1}) = 2\mathbb{E}[D]$, $h(\mathrm{e}^{-t_2}) = 2\mathbb{E}[D]\xi^2$.*

By Proposition 4.7 and Exercise 4.4, Theorem 4.1(i) follows when we prove that the cluster found between times T_1 and T_2 is indeed the giant component. We now first prove Proposition 4.7, after which we show that cluster found between times T_1 and T_2 is indeed the giant component.

Proof. The set of vertices $\mathcal{C}^*$ contains all vertices awakened in the interval $[T_1^*, T_2^*)$ and no others, and thus

$$v_k(\mathcal{C}^*) = V_k(T_1^*-) - V_k(T_2^*-), \qquad k \geq 1. \tag{4.1.42}$$

Since $T_2^* \xrightarrow{\mathbb{P}} t_2 \leq \theta$ and H is continuous, we obtain that $\inf_{t \leq T_2^*} H(\mathrm{e}^{-t}) \xrightarrow{\mathbb{P}} \inf_{t \leq t_2} H(\mathrm{e}^{-t}) = 0$, where the latter equality follows since $H(1) = 0$. Now, (4.1.32)

and (4.1.33) imply, in analogy with (4.1.35) and (4.1.36), that $n^{-1}\inf_{t\le T_2^*}\widetilde{A}(t)\xrightarrow{\mathbb{P}}0$ and

$$\sup_{t\le T_2^*} n^{-1}|\widetilde{S}(t)-S(t)|\xrightarrow{\mathbb{P}}0. \tag{4.1.43}$$

Since $\widetilde{V}_j(t)\ge V_j(t)$ for every j and $t\ge 0$,

$$\widetilde{V}_k(t)-V_k(t)\le k^{-1}\sum_{j=1}^{\infty} j(\widetilde{V}_j(t)-V_j(t))=k^{-1}(\widetilde{S}(t)-S(t)),\quad k\ge 1. \tag{4.1.44}$$

Hence (4.1.43) implies, for every $k\ge 1$, $\sup_{t\le T_2^*}|\widetilde{V}_k(t)-V_k(t)|=o_{\mathbb{P}}(n)$. Consequently, using Lemma 4.4, for $j=1,2$,

$$V_k(T_j^*-)=\widetilde{V}_k(T_j^*-)+o_{\mathbb{P}}(n)=np_k\mathrm{e}^{-kT_j^*}+o_{\mathbb{P}}(n)=np_k\mathrm{e}^{-kt_j}+o_{\mathbb{P}}(n), \tag{4.1.45}$$

and (4.1.39) follows by (4.1.42). Similarly, using $\sum_{k=0}^{\infty}(\widetilde{V}_k(t)-V_k(t))\le\widetilde{S}(t)-S(t)$,

$$|\mathcal{C}^*|=\sum_{k=1}^{\infty}(V_k(T_1^*-)-V_k(T_2^*-))=\sum_{k=1}^{\infty}(\widetilde{V}_k(T_1^*-)-\widetilde{V}_k(T_2^*))+o_{\mathbb{P}}(n) \tag{4.1.46}$$

$$=nG_D(\mathrm{e}^{-T_1^*})-nG_D(\mathrm{e}^{-T_2^*})+o_{\mathbb{P}}(n), \tag{4.1.47}$$

and

$$2|E(\mathcal{C}^*)|=\sum_{k=1}^{\infty}k(V_k(T_1^*-)-V_k(T_2^*))=\sum_{k=1}^{\infty}k(\widetilde{V}_k(T_1^*-)-\widetilde{V}_k(T_2^*))+o_{\mathbb{P}}(n) \tag{4.1.48}$$

$$=nh(\mathrm{e}^{-T_1^*})-nh(\mathrm{e}^{-T_2^*})+o_{\mathbb{P}}(n), \tag{4.1.49}$$

and (4.1.40) and (4.1.41) follow from the convergence $T_i^*\xrightarrow{\mathbb{P}}t_i$ and the continuity of $t\mapsto G_D(\mathrm{e}^{-t})$ and $t\mapsto h(\mathrm{e}^{-t})$. □

Let $\mathcal{C}'_{\max}$ be the component created at time T_1 and explored until time T_2, where we recall that T_1 is the last time Step 1 was performed before time $\theta/2$ and let T_2 be the next time it is performed if this occurs and $T_2=\infty$ otherwise. Then, $T_1\xrightarrow{\mathbb{P}}0$ and $T_2\xrightarrow{\mathbb{P}}\theta$. The cluster $\mathcal{C}'_{\max}$ is our candidate for the giant component $\mathcal{C}_{\max}$, and we next prove that indeed it is, **whp**, the largest connected component.

By Proposition 4.7, with $t_1=0$ and $t_2=\theta$,

$$|v_k(\mathcal{C}'_{\max})|/n\xrightarrow{\mathbb{P}}p_k(1-\mathrm{e}^{-kt}), \tag{4.1.50}$$

$$|\mathcal{C}'_{\max}|/n\xrightarrow{\mathbb{P}}G_D(1)-G_D(\mathrm{e}^{-\theta})=1-G_D(\xi), \tag{4.1.51}$$

$$|E(\mathcal{C}'_{\max})|/n\xrightarrow{\mathbb{P}}\frac{1}{2}(h(1)-h(\mathrm{e}^{-\theta}))=\frac{1}{2}(h(1)-h(\xi))=\frac{\mathbb{E}[D]}{2}(1-\xi^2), \tag{4.1.52}$$

using Exercise 4.4. We have found one large component $\mathcal{C}'_{\max}$ with the claimed numbers of vertices and edges. It remains to show that **whp** there is no other large component.

For this, let $\eta > 0$, and apply Proposition 4.7 to $T_0 = 0$ and T_1. Then, since $T_1 \xrightarrow{\mathbb{P}} 0$, the total number of vertices and edges in *all* components found before $\mathcal{C}'_{\max}$, i.e., before time T_1, is $o_{\mathbb{P}}(n)$. Hence, recalling that $\ell_n = \Theta(n)$ by [I, Condition 7.7(b)],

$$\mathbb{P}(\text{a component } \mathcal{C} \text{ with } |E(\mathcal{C})| \geq \eta \ell_n \text{ is found before } \mathcal{C}'_{\max}) \to 0. \tag{4.1.53}$$

We conclude that **whp** no component containing at least $\eta\ell_n$ half-edges is found *before* $\mathcal{C}'_{\max}$ is found.

In order to study the probability of finding a component containing at least $\eta\ell_n$ edges *after* $\mathcal{C}'_{\max}$ is found, we start by letting T_3 be the first time after time T_2 that Step 1 is performed. Since $\widetilde{S}(t) - S(t)$ increases by at most $\Delta = o(n)$ each time Step 1 is performed, we obtain from (4.1.43) that

$$\sup_{t \leq T_3}(\widetilde{S}(t) - S(t)) \leq \sup_{t \leq T_2}(\widetilde{S}(t) - S(t)) + \Delta = o_{\mathbb{P}}(n). \tag{4.1.54}$$

Comparing this to (4.1.38), for every $\varepsilon > 0$ and **whp**, we have that $\theta + \varepsilon > T_3$. Since also $T_3 > T_2 \xrightarrow{\mathbb{P}} \theta$, it follows that $T_3 \xrightarrow{\mathbb{P}} \theta$. If $\mathcal{C}'$ is the component created between T_2 and T_3, then Proposition 4.7 applied to T_2 and T_3 yields $|\mathcal{C}'|/n \xrightarrow{\mathbb{P}} 0$ and $|E(\mathcal{C}')| \xrightarrow{\mathbb{P}} 0$.

On the other hand, if there would exist a component $\mathcal{C} \neq \mathcal{C}'_{\max}$ in $\mathrm{CM}_n(\boldsymbol{d})$ with at least $\eta\ell_n$ edges that has not been found before $\mathcal{C}'_{\max}$, then with probability at least η, the vertex chosen at random by Step 1 at time T_2 starting the component $\mathcal{C}'$ would belong to $\mathcal{C}$. When this occurs, we clearly have that $\mathcal{C} = \mathcal{C}'$. Consequently,

$$\mathbb{P}(\text{a component } \mathcal{C} \text{ with } |E(\mathcal{C})| \geq \eta \ell_n \text{ is found after } \mathcal{C}'_{\max}) \leq \eta^{-1}\mathbb{P}(|E(\mathcal{C}')| \geq \eta\ell_n) \to 0, \tag{4.1.55}$$

since $|E(\mathcal{C}')| \xrightarrow{\mathbb{P}} 0$.

Combining (4.1.53) and (4.1.55), we see that **whp** there is no component except $\mathcal{C}'_{\max}$ that has at least $\eta\ell_n$ edges. As a result, we must have that $\mathcal{C}'_{\max} = \mathcal{C}_{\max}$, where $\mathcal{C}_{\max}$ is the largest component. Further, again **whp**, $|E(\mathcal{C}_{(2)})| < \eta\ell_n$. Consequently, the results for $\mathcal{C}_{\max}$ follow from (4.1.50)-(4.1.52). We have further shown $|E(\mathcal{C}_{(2)}|/\ell_n \xrightarrow{\mathbb{P}} 0$, which implies $|E(\mathcal{C}_{(2)})|/n \xrightarrow{\mathbb{P}} 0$ and $|\mathcal{C}_{(2)}|/n \xrightarrow{\mathbb{P}} 0$ because $\ell_n = \Theta(n)$ and $|\mathcal{C}_{(2)}| \leq |E(\mathcal{C}_{(2)})| + 1$. This completes the proof of Theorem 4.1(i). □

The proof of Theorem 4.1(ii) is very similar to the last step in the proof for Theorem 4.1(i). Indeed, let $T_1 = 0$ and let T_2 be the next time Step 1 is performed, or $T_2 = \infty$ when this does not occur. Then,

$$\sup_{t \leq T_2} |A(t) - \widetilde{A}(t)| = \sup_{t \leq T_2} |\widetilde{S}(t) - S(t)| \leq 2\Delta = o(n). \tag{4.1.56}$$

For every $\varepsilon > 0$, $n^{-1}\widetilde{A}(\varepsilon) \xrightarrow{\mathbb{P}} H(\mathrm{e}^{-\varepsilon}) < 0$ by (4.1.32) and Lemma 4.6(ii), while $A(\varepsilon) \geq 0$, and it follows from (4.1.56) that **whp** $T_2 < \varepsilon$. Hence, $T_2 \xrightarrow{\mathbb{P}} 0$. We apply Proposition 4.7 (which holds in this case too, with $\theta = 0$) and find that if $\mathcal{C}$ is the first component found, then $|E(\mathcal{C})|/n \xrightarrow{\mathbb{P}} 0$.

Let $\eta > 0$. If $|E(\mathcal{C}_{\max})| \geq \eta \ell_n$, then the probability that the first half-edge chosen by Step 1 belongs to $\mathcal{C}_{\max}$, and thus $\widetilde{\mathcal{C}} = \mathcal{C}_{\max}$, is $2|E(\mathcal{C}_{\max})|/(2\ell_n) \geq \eta$, and hence,

$$\mathbb{P}(|E(\mathcal{C}_{\max})| \geq \eta \ell_n) \leq \eta^{-1} \mathbb{P}(|E(\mathcal{C})| \geq \eta \ell_n) \to 0. \tag{4.1.57}$$

The results follows since $\ell_n = \Theta(n)$ by [I, Condition 7.7(b)] and $|\mathcal{C}_{\max}| \leq |E(\mathcal{C}_{\max})|+1$. This completes the proof of Theorem 4.1(ii), and thus that of Theorem 4.1. □

4.1.4 The giant component of related random graphs

In this section, we extend the results of Theorem 4.1 to some related models, such as uniform simple random graphs with a given degree sequence, as well as generalized random graphs.

Recall that $\mathrm{UG}_n(\boldsymbol{d})$ denotes a uniform simple random graph with degrees $\boldsymbol{d}$ (see [I, Section 7.5]). The results in Theorem 4.1 also hold for $\mathrm{UG}_n(\boldsymbol{d})$ when we assume that [I, Condition 7.7(a)-(c)] holds:

Theorem 4.8 (Phase transition in $\mathrm{UG}_n(\boldsymbol{d})$)**.** *Let $\boldsymbol{d}$ satisfy [I, Condition 7.7(a)-(c)]. Then, the results in Theorem 4.1 also hold for a* uniform *simple graph with degree sequence $\boldsymbol{d}$.*

Proof. By Corollary 7.15, and since $\boldsymbol{d} = (d_i)_{i\in[n]}$ satisfies [I, Condition 7.7(a)-(c)], any event $\mathcal{E}_n$ that occurs whp for $\mathrm{CM}_n(\boldsymbol{d})$, also occurs whp for $\mathrm{UG}_n(\boldsymbol{d})$. By [I, Theorem 4.1], the event $\mathcal{E}_n$ that $\{\big||\mathcal{C}_{\max}|/n - \zeta\big| \leq \varepsilon\}$ occurs whp for $\mathrm{CM}_n(\boldsymbol{d})$, so it also holds whp for $\mathrm{UG}_n(\boldsymbol{d})$. The proof for the other limits is identical. □

We next prove Theorem 2.7 for $\mathrm{GRG}_n(\boldsymbol{w})$:

Theorem 4.9 (Phase transition in $\mathrm{GRG}_n(\boldsymbol{w})$)**.** *Let $\boldsymbol{w}$ satisfy [I, Condition 6.4(a)-(c)]. Then, the results in Theorem 4.1 also hold for* $\mathrm{GRG}_n(\boldsymbol{w})$*,* $\mathrm{CL}_n(\boldsymbol{w})$ *and* $\mathrm{NR}_n(\boldsymbol{w})$*.*

Proof. Let d_i be the degree of vertex i in $\mathrm{GRG}_n(\boldsymbol{w})$ defined in [I, (6.2.1)], where we use a small letter to avoid confusion with D_n, which is the degree of a *uniform* vertex in $[n]$. By [I, Theorem 7.16], the law of $\mathrm{GRG}_n(\boldsymbol{w})$ conditionally on the degrees $\boldsymbol{d}$ and $\mathrm{CM}_n(\boldsymbol{d})$ conditionally on being simple agree. Assume that $(d_i)_{i\in[n]}$ satisfies that [I, Condition 7.7(a)-(c)] hold in probability. Then, by [I, Theorem 7.16] and Theorem 4.1, the results in Theorem 4.1 also hold for $\mathrm{GRG}_n(\boldsymbol{w})$. By [I, Theorem 6.20], the same result applies to $\mathrm{CL}_n(\boldsymbol{w})$, and by [I, Exercise 6.36], also to $\mathrm{NR}_n(\boldsymbol{w})$. Thus, we are left to prove that [I, Condition 6.4(a)-(c)] implies that [I, Condition 7.7(a)-(c)] holds for $\mathrm{GRG}_n(\boldsymbol{w})$, which is the content of the next proposition:

Proposition 4.10 (Relating the assumptions on weights and degrees)**.** *Let $\boldsymbol{d} = (d_i)_{i\in[n]}$ be the (random) degrees in $\mathrm{GRG}_n(\boldsymbol{w})$. Let $\boldsymbol{w}$ satisfy [I, Condition 6.4(a)-(c)]. Then, [I, Condition 7.7(a)-(c)] holds for $\boldsymbol{d}$, where the convergence holds in probability, and the limit law D equals a mixed Poisson random variable with mixing distribution W.*

Proof. [I, Condition 7.7(a)] follows from [I, Theorem 6.10]. To avoid confusion, we denote by $\mathbb{E}_n$ the conditional expectation given the graph $\mathrm{GRG}_n(\boldsymbol{w})$, so that $\mathbb{E}_n[D_n]$ denotes the degree in $\mathrm{GRG}_n(\boldsymbol{w})$ of a vertex chosen uniformly from $[n]$ *given the realization of* $\mathrm{GRG}_n(\boldsymbol{w})$. Thus, for [I, Condition 7.7(b)-(c)], we need to show that $\mathbb{E}_n[D_n] \xrightarrow{\mathbb{P}} \mathbb{E}[D] = \mathbb{E}[W]$, $\mathbb{E}_n[D_n^2] \xrightarrow{\mathbb{P}} \mathbb{E}[D^2] = \mathbb{E}[W(W+1)]$, where the latter equalities hold since W has a Poisson distribution with (random) parameter W.

We perform the proofs of [I, Condition 7.7(b)-(c)] simultaneously. Let $p \in \{1,2\}$. Then, we compute

$$\mathbb{E}_n[D_n^p] = \sum_{k=1}^{\infty} k^p P_k^{(n)}, \tag{4.1.58}$$

where we recall that $P_k^{(n)} = \frac{1}{n}\sum_{i\in[n]} \mathbb{1}_{\{d_i=k\}}$ denotes the proportion of vertices with degree k. Let $K \geq 1$ be a large constant, and split

$$\mathbb{E}_n[D_n^2] = \sum_{k=1}^{K} k^p P_k^{(n)} + \sum_{k=K+1}^{\infty} k^p P_k^{(n)}. \tag{4.1.59}$$

By [I, Theorem 6.10],

$$\sum_{k=1}^{K} k^p P_k^{(n)} \xrightarrow{\mathbb{P}} \sum_{k=1}^{K} k^p p_k = \mathbb{E}[D^p \mathbb{1}_{\{D\leq K\}}]. \tag{4.1.60}$$

By the Dominated Convergence Theorem ([I, Theorem A.17]), $\mathbb{E}[D^p\mathbb{1}_{\{D\leq K\}}] \to \mathbb{E}[D^p]$ when $K\to\infty$. Therefore, we are left to show that, when first $n\to\infty$ followed by $K\to\infty$,

$$\sum_{k=K+1}^{\infty} k^p P_k^{(n)} \xrightarrow{\mathbb{P}} 1. \tag{4.1.61}$$

We rewrite

$$\sum_{k=K+1}^{\infty} k^p P_k^{(n)} = \frac{1}{n} \sum_{i\in[n]:\, d_i\geq K+1} d_i^p. \tag{4.1.62}$$

We split the above sum, depending on whether $w_i \geq d_i/2$ or not, to obtain

$$\sum_{k=K+1}^{\infty} k^p P_k^{(n)} \leq \frac{1}{n}\sum_{i\in[n]} d_i^p \mathbb{1}_{\{d_i\geq K+1, w_i<d_i/2\}} + \frac{1}{n}\sum_{i\in[n]} d_i^p \mathbb{1}_{\{d_i\geq K+1, w_i\geq d_i/2\}}. \tag{4.1.63}$$

We bound each of these contributions separately. We start by bounding

$$\frac{1}{n}\sum_{i\in[n]} d_i^p \mathbb{1}_{\{d_i\geq K+1, w_i\geq d_i/2\}} \leq \frac{2^p}{n}\sum_{i\in[n]} w_i^p \mathbb{1}_{\{w_i\geq K/2\}} = 2^d \mathbb{E}[W_n^p \mathbb{1}_{\{W_n\geq K/2\}}], \tag{4.1.64}$$

the right-hand side being *deterministic*. By [I, Condition 7.7(b)-(c)], for $p\in\{1,2\}$,

$$\mathbb{E}[W_n^p \mathbb{1}_{\{W_n\geq K/2\}}] \to \mathbb{E}[W^p \mathbb{1}_{\{W\geq K/2\}}] \leq \varepsilon/2, \tag{4.1.65}$$

when we take $K = K(\varepsilon)$ sufficiently large. This bounds the second term in (4.1.63).

To bound the first term in (4.1.63), we continue to bound the expectation

$$\frac{1}{n}\mathbb{E}\Big[\sum_{i\in[n]} d_i^p \mathbb{1}_{\{d_i\geq K+1, w_i<d_i/2\}}\Big] = \frac{1}{n}\sum_{i\in[n]} \mathbb{E}[d_i^p \mathbb{1}_{\{d_i>(2w_i\vee K)\}}]. \tag{4.1.66}$$

For $\mathrm{NR}_n(\boldsymbol{w})$, the random degree d_i is stochastically bounded by a $\mathsf{Poi}(w_i)$ random variable (recall Proposition 3.10). Further, by [I, Corollary 6.20] and [I, Exercise 6.36], $\mathrm{GRG}_n(\boldsymbol{w}), \mathrm{CL}_n(\boldsymbol{w})$ and $\mathrm{NR}_n(\boldsymbol{w})$ are asymptotically equivalent when [I, Condition 6.4(a)-(c)] holds. Therefore,

$$\mathbb{E}[d_i^p \mathbb{1}_{\{d_i>(2w_i\vee K)\}}] \leq \mathbb{E}[\mathsf{Poi}(w_i)^p \mathbb{1}_{\{\mathsf{Poi}(w_i)>(2w_i\vee K)\}}]. \tag{4.1.67}$$

Let $Y\sim\mathsf{Poi}(\lambda)$. Then, for each $k\geq 2p$,

$$\begin{aligned}\mathbb{E}[Y^p\mathbb{1}_{\{Y\geq k\}}] &\leq 2^p\mathbb{E}[Y(Y-1)\cdots(Y-p+1)\mathbb{1}_{\{Y\geq k\}}]\\ &= 2^p\sum_{y=k}^{\infty} y(y-1)\cdots(y-p+1)\mathbb{P}(Y=y)\\ &= 2^p\sum_{y=k}^{\infty} y(y-1)\cdots(y-p+1)\frac{\lambda^y \mathrm{e}^{-\lambda}}{y!} = 2^p\lambda^p\mathbb{P}(Y\geq k-p).\end{aligned} \tag{4.1.68}$$

Therefore,

$$\mathbb{E}[\mathsf{Poi}(w_i)^p \mathbb{1}_{\{\mathsf{Poi}(w_i)>(2w_i\vee K)\}}] \leq 2^p w_i^p \mathbb{P}(\mathsf{Poi}(w_i)>(2w_i\vee K)), \tag{4.1.69}$$

and we arrive at

$$\begin{aligned}\frac{1}{n}\mathbb{E}\Big[\sum_{i\in[n]} d_i^p \mathbb{1}_{\{d_i\geq K+1, w_i<d_i/2\}}\Big] &\leq \frac{2^p}{n}\sum_{i\in[n]} w_i^p\mathbb{P}(\mathsf{Poi}(w_i)>(2w_i\vee K))\\ &= 2^p\mathbb{E}[W_n^p\mathbb{1}_{\{\mathsf{Poi}(W_n)\geq(2W_n\vee K)\}}].\end{aligned} \tag{4.1.70}$$

Again by [I, Condition 7.7(b)-(c)], for $p\in\{1,2\}$,

$$\mathbb{E}[W_n^p\mathbb{1}_{\{\mathsf{Poi}(W_n)\geq(2W_n\vee K)\}}] \to \mathbb{E}[W^p\mathbb{1}_{\{\mathsf{Poi}(W)\geq(2W\vee K)\}}], \tag{4.1.71}$$

which converges to zero when $K\to\infty$. Therefore, for each $\varepsilon>0$, we choose first K and then n sufficiently large, so that

$$\frac{1}{n}\mathbb{E}\Big[\sum_{i\in[n]} d_i^p \mathbb{1}_{\{d_i\geq K+1, w_i<d_i/2\}}\Big] \leq \varepsilon^2/2. \tag{4.1.72}$$

By the Markov inequality ([I, Theorem 2.15]),

$$\mathbb{P}\Big(\frac{1}{n}\sum_{i\in[n]} d_i^p \mathbb{1}_{\{d_i\geq K+1, w_i<d_i/2\}} \geq \varepsilon\Big) \leq \frac{1}{n\varepsilon}\mathbb{E}\Big[\sum_{i\in[n]} d_i^p \mathbb{1}_{\{d_i\geq K+1, w_i<d_i/2\}}\Big] \leq \varepsilon/2. \tag{4.1.73}$$

Combining (4.1.65) and (4.1.73), we see that,

$$\mathbb{P}\Big(\sum_{k\geq K+1} P_k^{(n)} \leq \varepsilon\Big) \geq 1-\varepsilon. \tag{4.1.74}$$

As a result, together with (4.1.60), for $p \in \{1,2\}$,

$$\sum_{k\geq 0} k^p |P_k^{(n)} - p_k| \xrightarrow{\mathbb{P}} 1, \tag{4.1.75}$$

as required. This completes the proof of Proposition 4.10, and thus also that of Theorem 4.9. □

Unfortunately, when $\nu = \infty$, we cannot rely on the fact that by [I, Theorem 7.16], the law of $\mathrm{GRG}_n(\boldsymbol{w})$ conditionally on the degrees $\boldsymbol{d}$ and $\mathrm{CM}_n(\boldsymbol{d})$ conditionally on being simple agree. Indeed, when $\nu = \infty$, the probability that $\mathrm{CM}_n(\boldsymbol{d})$ is simple vanishes. Therefore, we instead rely on a *truncation argument* to extend Theorem 4.9 to the case where $\nu = \infty$:

Theorem 4.11 (Phase transition in $\mathrm{GRG}_n(\boldsymbol{w})$)**.** *Let $\boldsymbol{w}$ satisfy [I, Condition 6.4(a)-(b)]. Then, the results in Theorem 4.1 also hold for $\mathrm{GRG}_n(\boldsymbol{w})$, $\mathrm{CL}_n(\boldsymbol{w})$ and $\mathrm{NR}_n(\boldsymbol{w})$.*

Proof. We only prove that $|\mathcal{C}_{\max}|/n \xrightarrow{\mathbb{P}} \zeta$, the other statements in Theorem 4.1 can be proved in a similar fashion. We prove Theorem 4.11 only for $\mathrm{NR}_n(\boldsymbol{w})$, the proof for $\mathrm{GRG}_n(\boldsymbol{w})$ and $\mathrm{CL}_n(\boldsymbol{w})$ being similar. Recall that $\mathcal{N}_m(i)$ denotes the set of vertices at graph distance m from $i \in [n]$, and let

$$|\mathcal{C}'_{\max}| = \#\{i\colon \mathcal{N}_m(i) \neq \varnothing\} \tag{4.1.76}$$

denote the number of vertices that survive to graph distance m. Then, by Corollary 3.13, $\mathbb{E}[|\mathcal{C}'_{\max}|/n] \to \zeta_m = \mathbb{P}(\mathcal{Z}_m > 0)$, and

$$\mathrm{Var}(|\mathcal{C}'_{\max}|/n) \to 0. \tag{4.1.77}$$

Therefore, $|\mathcal{C}'_{\max}|/n \xrightarrow{\mathbb{P}} \zeta_m$. Since $\mathcal{C}_{\max}$ has a diameter that tends to infinity (recall Exercise 3.24), $|\mathcal{C}_{\max}| \leq |\mathcal{C}'_{\max}|$, so that $|\mathcal{C}_{\max}|/n \leq \zeta_m + o_{\mathbb{P}}(1)$. This proves the required upper bound.

For the lower bound, we bound $\mathrm{NR}_n(\boldsymbol{w})$ from below by a random graph with edge probabilities

$$p_{ij}^{(K)} = 1 - \mathrm{e}^{-(w_i \wedge K)(w_j \wedge K)/\ell_n}. \tag{4.1.78}$$

Therefore, also $|\mathcal{C}_{\max}| \preceq |\mathcal{C}^{(K)}_{\max}|$, where $\mathcal{C}^{(K)}_{\max}$ is the largest connected component in the inhomogeneous random graph with edge probabilities $(p_{ij}^{(K)})_{i,j\in[n]}$. Let

$$w_i^{(K)} = (w_i \wedge K)\frac{1}{\ell_n}\sum_{j\in[n]}(w_j \wedge K), \tag{4.1.79}$$

so that the edge probabilities in (4.1.78) correspond to the Norros-Reittu model with weights $(w_i^{(K)})_{i\in[n]}$. It is not hard to see that when [I, Condition 6.4(a)] holds for

$(w_i)_{i\in[n]}$, then [I, Condition 6.4(a)-(c)] hold for $(w_i^{(K)})_{i\in[n]}$, where the limiting random variable equals $(W \wedge K)$. Therefore, Theorem 4.9 applies to $(w_i^{(K)})_{i\in[n]}$. We deduce that $|\mathcal{C}_{\max}^{(K)}|/n \xrightarrow{\mathbb{P}} \zeta^{(K)}$, which is the survival probability of the two-stage mixed-Poisson branching process with mixing variable $(W \wedge K)$. Since $\zeta^{(K)} \to \zeta$ when $K \to \infty$, we conclude that $|\mathcal{C}_{\max}|/n \xrightarrow{\mathbb{P}} \zeta$. □

Exercise 4.5 (Number of vertices with degree k). *Let $\boldsymbol{w}$ satisfy [I, Condition 6.4(a)-(b)]. Adapt the above proof to show that also $v_k(\mathcal{C}_{\max})/n \xrightarrow{\mathbb{P}} p_k(1-\xi^k)$ for $\mathrm{NR}_n(\boldsymbol{w})$.*

4.2 Connectivity of $\mathrm{CM}_n(\boldsymbol{d})$

Assume that $\mathbb{P}(D=2) < 1$. By Theorem 4.1, we see that $|\mathcal{C}_{\max}|/n \xrightarrow{\mathbb{P}} 1$ when $\mathbb{P}(D \geq 2) = 1$, as in this case the survival probability equals 1. In this section, we investigate conditions under which $\mathrm{CM}_n(\boldsymbol{d})$ is whp *connected*, i.e., $\mathcal{C}_{\max} = [n]$ and $|\mathcal{C}_{\max}| = n$.

We first show that $\mathrm{CM}_n(\boldsymbol{d})$ is with positive probability *disconnected* when either $n_1 \gg n^{1/2}$, or when $\mathbb{P}(D \geq 2) > 1$. The main result in this section is Theorem 4.14, which states that for *all* possible degree sequences with $n_1 = n_2 = 0$, $\mathrm{CM}_n(\boldsymbol{d})$ is **whp** connected. Remarkably, we do not even need [I, Condition 7.7(a)] for this result.

Proposition 4.12 (Disconnectivity of $\mathrm{CM}_n(\boldsymbol{d})$ when $n_1 \gg n^{1/2}$). *Let [I, Condition 7.7(a)-(b)] hold, and assume that $n_1 \gg n^{1/2}$. Then,*

$$\lim_{n\to\infty} \mathbb{P}(\mathrm{CM}_n(\boldsymbol{d}) \text{ } \textit{connected}) = 0. \tag{4.2.1}$$

Proof. We note that $\mathrm{CM}_n(\boldsymbol{d})$ is disconnected when there are two vertices of degree 1 whose half-edges are paired to each other. When the half-edges of two vertices of degree 1 are paired to each other, we say that a *2-pair* is created. Then,

$$\mathbb{P}(\mathrm{CM}_n(\boldsymbol{d}) \text{ contains no 2-pair}) = \prod_{i=1}^{n_1} \frac{\ell_n - n_1 - 2i + 1}{\ell_n - 2i + 1} = \prod_{i=1}^{n_1} \Big(1 - \frac{n_1}{\ell_n - 2i + 1}\Big). \tag{4.2.2}$$

Since, for each $i \geq 1$,

$$1 - \frac{n_1}{\ell_n - 2i + 1} \leq 1 - \frac{n_1}{\ell_n} \leq \mathrm{e}^{-n_1/\ell_n}, \tag{4.2.3}$$

we arrive at

$$\mathbb{P}(\mathrm{CM}_n(\boldsymbol{d}) \text{ contains no 2-pair}) \leq \mathrm{e}^{-n_1^2/\ell_n} = o(1), \tag{4.2.4}$$

since $\ell_n = \Theta(n)$ and $n_1 \gg n^{1/2}$. □

Proposition 4.13 (Disconnectivity of $\mathrm{CM}_n(\boldsymbol{d})$ when $p_2 > 0$). *Let [I, Condition 7.7(a)-(b)] hold, and assume that $p_2 > 0$. Then,*

$$\limsup_{n\to\infty} \mathbb{P}(\mathrm{CM}_n(\boldsymbol{d}) \text{ } \textit{connected}) < 1. \tag{4.2.5}$$

Proof. We perform a second moment method on the number $P(2)$ of connected components consisting of two vertices of degree 2. The expected number of such components equals

$$\mathbb{E}[P(2)] = \frac{2n_2(n_2-1)}{2(\ell_n-1)(\ell_n-3)}, \tag{4.2.6}$$

since there are $n_2(n_2-1)/2$ pairs of vertices of degree 2, and the probability that a fixed pair forms a connected component is equal to $2/(\ell_n-1)(\ell_n-3)$. By [I, Condition 7.7(a)-(b)], which implies that $n_2/n \to p_2$,

$$\mathbb{E}[P(2)] \to p_2^2/\mathbb{E}[D]^2 \equiv \lambda_2. \tag{4.2.7}$$

By assumption, $p_2 > 0$, so that also $\lambda_2 > 0$. We can use Theorem 2.6 to show that $P(2) \xrightarrow{d} \mathsf{Poi}(\lambda_2)$, so that

$$\mathbb{P}(\mathrm{CM}_n(\boldsymbol{d}) \text{ disconnected}) \geq \mathbb{P}(P(2) > 0) \to 1 - \mathrm{e}^{-\lambda_2} > 0, \tag{4.2.8}$$

as required. The proof that Theorem 2.6 can be applied is left as an exercise below. □

Exercise 4.6 (Factorial moments of $P(2)$)**.** *Let [I, Condition 7.7(a)-(b)] hold, and assume that $p_2 > 0$. Prove that, for every $k \geq 1$ and with $\lambda_2 = p_2^2/\mathbb{E}[D]^2$,*

$$\mathbb{E}[(P(2))_k] \to \lambda_2^k. \tag{4.2.9}$$

Conclude that $P(2) \xrightarrow{d} \mathsf{Poi}(\lambda_2)$.

We continue to investigate when the configuration model yields a connected graph:

Theorem 4.14 (Connectivity of $\mathrm{CM}_n(\boldsymbol{d})$)**.** *Assume that $d_i \geq 3$ for every $i \in [n]$. Then $\mathrm{CM}_n(\boldsymbol{d})$ is connected whp. More precisely, there exists a constant $C > 0$ such that for every $\boldsymbol{d}$,*

$$\mathbb{P}(\mathrm{CM}_n(\boldsymbol{d}) \text{ disconnected}) = O(1/n). \tag{4.2.10}$$

When [I, Condition 7.7(a)] holds with $p_1 = p_2 = 0$, then $\nu \geq 2 > 1$ is immediate, so we are always in the supercritical regime. Also, $\zeta = 1$ when $p_1 = p_2 = 0$, since survival of the two-stage branching process occurs with probability 1. Therefore, Theorem 4.1 implies that the largest connected component has size $n(1+o_{\mathbb{P}}(1))$ when [I, Condition 7.7(a)-(b)] hold. Theorem 4.14 extends this to the statement that $\mathrm{CM}_n(\boldsymbol{d})$ is with high probability *connected*. However, we do not assume that [I, Condition 7.7] holds here.

We note that Theorem 4.14 yields an important difference between the generalized random graph and the configuration model, also from a practical point of view. Indeed, for the generalized random graph to be whp connected, the degrees must tend to infinity. This can be observed already for $\mathrm{ER}_n(p)$ in Theorem 5.8. For the configuration model, it is possible that the graph is connected while the average degree is bounded. Many real-world networks are connected, which makes the configuration model often more suitable than inhomogeneous random graphs.

Proof. We recall that a *configuration* denotes a pairing of all the half-edges. We note that the probability of a configuration equals $1/(\ell_n-1)!!$. On the event that $\mathrm{CM}_n(\boldsymbol{d})$ is disconnected, there exists a set of indices $\mathcal{I}\subset[n]$ with $|\mathcal{I}|\leq\lfloor n/2\rfloor$ such that all half-edges incident to vertices in $\mathcal{I}$ are *only* paired to other half-edges incident to other vertices in $\mathcal{I}$. For $\mathcal{I}\subseteq[n]$, we let

$$\ell_n(\mathcal{I})=\sum_{i\in\mathcal{I}}d_i. \tag{4.2.11}$$

Clearly, in order for the half-edges incident to vertices in $\mathcal{I}$ to be paired only to other half-edges incident to vertices in $\mathcal{I}$, $\ell_n(\mathcal{I})$ needs to be even. The number of configurations for which this happens is bounded above by

$$(\ell_n(\mathcal{I})-1)!!(\ell_n(\mathcal{I}^c)-1)!!. \tag{4.2.12}$$

As a result,

$$\begin{aligned}\mathbb{P}(\mathrm{CM}_n(\boldsymbol{d})\text{ disconnected})&\leq\sum_{\mathcal{I}\subset[n]}\frac{(\ell_n(\mathcal{I})-1)!!(\ell_n(\mathcal{I}^c)-1)!!}{(\ell_n-1)!!}\\&=\sum_{\mathcal{I}\subset[n]}\prod_{j=1}^{\ell_n(\mathcal{I})/2}\frac{\ell_n(\mathcal{I})-2j+1}{\ell_n-2j+1},\end{aligned} \tag{4.2.13}$$

where the sum over $\mathcal{I}\subset[n]$ is restricted to $\mathcal{I}$ for which $|\mathcal{I}|\leq\lfloor n/2\rfloor$.

Exercise 4.7 (Isolated vertex)**.** *Use the above bound to show that, when $d_i\geq 3$ for all $i\in[n]$,*

$$\mathbb{P}(\textit{there exists an isolated vertex})\leq\frac{3n}{(2\ell_n-1)(2\ell_n-3)}. \tag{4.2.14}$$

Define

$$f(x)=\prod_{j=1}^{x}\frac{2x-2j+1}{\ell_n-2j+1}. \tag{4.2.15}$$

We can rewrite

$$f(x)=\frac{\prod_{j=1}^{x}(2x-2j+1)}{\prod_{j=1}^{x}(\ell_n-2j-1)}=\frac{\prod_{i=0}^{x-1}(2i+1)}{\prod_{k=0}^{x-1}(\ell_n-2k+1)}=\prod_{j=0}^{x-1}\frac{2i+1}{\ell_n-2i-1}, \tag{4.2.16}$$

where we write $i=x-j$ and $k=j-1$ in the second equality. Thus, for $x\leq\ell_n/4$, $x\mapsto f(x)$ is decreasing, since

$$\frac{f(x+1)}{f(x)}=\frac{2x+1}{\ell_n-2x-1}\leq 1. \tag{4.2.17}$$

Now, for every $\mathcal{I}$, since $d_i\geq 3$ for every $i\in[n]$ and since $\ell_n(\mathcal{I})$ is *even*,

$$\ell_n(\mathcal{I})\geq 2\lceil 3|\mathcal{I}|/2\rceil, \tag{4.2.18}$$

which only depends on the number of vertices in $\mathcal{I}$. Since there are precisely $\binom{n}{m}$ ways of choosing m vertices out of $[n]$, we conclude that

$$\mathbb{P}(\mathrm{CM}_n(\boldsymbol{d}) \text{ disconnected}) \leq \sum_{\mathcal{I}\subset[n]} f(\lceil 3|\mathcal{I}|/2\rceil) = \sum_{m=1}^{\lfloor n/2\rfloor} \binom{n}{m} f(\lceil 3m/2\rceil), \tag{4.2.19}$$

with $m = |\mathcal{I}|$.

Exercise 4.8 (Isolated vertex)**.** *Use* (4.2.19) *to reprove Exercise 4.8. Hence, the above bound is quite sharp.*

Exercise 4.9 (A cluster of size two)**.** *Use* (4.2.19) *to prove that*

$$\mathbb{P}(\textit{there exists a cluster of size 2}) \leq \frac{15n(n-1)}{(2\ell_n-1)(2\ell_n-3)(2\ell_n-5)}. \tag{4.2.20}$$

Note that, for m odd,

$$\frac{f(2\lceil 3(m+1)/2\rceil)}{f(2\lceil 3m/2\rceil)} = \frac{f((3m+1)/2+1)}{f((3m+1)/2)} = \frac{3m+3}{\ell_n-3m-2}. \tag{4.2.21}$$

while, for m even,

$$\frac{f(\lceil 3(m+1)/2\rceil)}{f(\lceil 3m/2\rceil)} = \frac{f(3m/2+2)}{f(3m/2)} = \frac{3m+5}{\ell_n-3m-5}\frac{3m+3}{\ell_n-3m-3}, \tag{4.2.22}$$

Define

$$h_n(m) = \binom{n}{m} f(\lceil 3m/2\rceil), \tag{4.2.23}$$

so that

$$\mathbb{P}(\mathrm{CM}_n(\boldsymbol{d}) \text{ disconnected}) \leq \sum_{m=1}^{\lfloor n/2\rfloor} h_n(m). \tag{4.2.24}$$

Then,

$$\frac{h_n(m+1)}{h_n(m)} = \frac{n-m}{m+1}\frac{f(\lceil 3(m+1)/2\rceil)}{f(\lceil 3m/2\rceil)}, \tag{4.2.25}$$

so that, for m odd and using $\ell_n \geq 3n$,

$$\frac{h_n(m+1)}{h_n(m)} = \frac{3(n-m)}{\ell_n-3m-3} \leq \frac{n-m}{n-m-1}, \tag{4.2.26}$$

while, for m even and using $\ell_n \geq 3n$,

$$\frac{h_n(m+1)}{h_n(m)} = \frac{3(n-m)}{\ell_n-3m-5}\frac{3m+5}{\ell_n-3m-3} \leq \frac{n-m}{n-m-1}\frac{m+2}{n-m-2}. \tag{4.2.27}$$

Thus, we obtain that, for $m \leq n/2$, there exists a $c > 0$ such that

$$\frac{h_n(m+1)}{h_n(m)} \leq 1 + \frac{c}{n}. \tag{4.2.28}$$

We conclude that, for $m \leq n/2$ such that $m \geq 3$,

$$h_n(m) = h_n(3)\prod_{j=3}^{m}\frac{h_n(j+1)}{h_n(j)} \leq h_n(3)\prod_{j=3}^{\lfloor n/2\rfloor}(1+c/n) \tag{4.2.29}$$
$$\leq h_n(3)(1+c/n)^{\lfloor n/2\rfloor} \leq h_n(3)\mathrm{e}^{c/2},$$

so that

$$\mathbb{P}(\mathrm{CM}_n(\boldsymbol{d}) \text{ disconnected}) \leq \sum_{m=1}^{\varepsilon n} h_n(m) \leq h_n(1)+h_n(2)+\sum_{m=3}^{\lfloor n/2\rfloor} h_n(m) \tag{4.2.30}$$
$$\leq h_n(1)+h_n(2)+nh_n(3)\mathrm{e}^{c/2}/2.$$

By Exercises 4.7 and 4.9, $h_n(1), h_n(2) = O(1/n)$, so we are left to compute $h_n(3)$. For this, we note that $\lceil 3m/2\rceil = 5$ when $m = 3$, so that

$$h_n(3) = \binom{n}{3} f(5) = \frac{9!!n(n-1)(n-2)}{6(\ell_n-1)(\ell_n-3)(\ell_n-5)(\ell_n-7)(\ell_n-9)} = O(1/n^2), \tag{4.2.31}$$

so that $nh_n(3) = O(1/n)$. We conclude that

$$\mathbb{P}(\mathrm{CM}_n(\boldsymbol{d}) \text{ disconnected}) = O(1/n), \tag{4.2.32}$$

as required. □

4.3 Related results for the configuration model

In this section, we discuss related results for the configuration model. We start by discussing the subcritical behavior of the configuration model.

The largest subcritical cluster. When $\nu < 1$, so that in particular $\mathbb{E}[D^2] < \infty$, the largest connected component for $\mathrm{CM}_n(\boldsymbol{d})$ is closely related to the largest degree:

Theorem 4.15 (Subcritical phase for $\mathrm{CM}_n(\boldsymbol{d})$)**.** *Let $\boldsymbol{d}$ satisfy [I, Condition 7.7(a)-(c)] with $\nu = \mathbb{E}[D(D-1)]/\mathbb{E}[D] < 1$. Suppose further that there exists $\tau > 3$ and $c_2 > 0$ such that*

$$[1-F_n](x) \leq c_2 x^{-(\tau-1)}. \tag{4.3.1}$$

Then, for $\mathrm{CM}_n(\boldsymbol{d})$ with $\Delta = \max_{j\in[n]} d_j$,

$$|\mathcal{C}_{\max}| = \frac{\Delta}{1-\nu} + o_{\mathbb{P}}(n^{1/(\tau-1)}). \tag{4.3.2}$$

Theorem 4.15 is closely related to Theorem 3.25. In fact, we can use Theorem 4.15 to prove Theorem 3.25:

Exercise 4.10 (Proof of Theorem 3.25)**.** *Use Theorem 4.15 and Proposition 4.10 to prove Theorem 3.25.*

The near-critical supercritical behavior in the configuration model. In [78], also partial results appear on the near-critical behavior of $\mathrm{CM}_n(\boldsymbol{d})$:

Theorem 4.16 (Near-critical behavior $\mathrm{CM}_n(\boldsymbol{d})$)**.** *Let $\boldsymbol{d}$ satisfy [I, Condition 7.7(a)-(c)] with $\nu = \mathbb{E}[D(D-1)]/\mathbb{E}[D] < 1$. Assume further that $\alpha_n = \nu_n - 1 = \mathbb{E}[D_n(D_n-2)]/\mathbb{E}[D_n] > 0$ is such that $n^{1/3}\alpha_n \to \infty$, and that*

$$\mathbb{E}[D_n^{4+\varepsilon}] = O(1) \tag{4.3.3}$$

for some $\varepsilon > 0$. Let $\beta = \mathbb{E}[D(D-1)(D-2)]/\mathbb{E}[D] > 0$. Then, $\mathrm{CM}_n(\boldsymbol{d})$ satisfies

$$\begin{aligned} |\mathcal{C}_{\max}| &= \frac{2}{\mathbb{E}[D]\beta} n\alpha_n + o_{\mathbb{P}}(n\alpha_n), \\ |v_k(\mathcal{C}_{\max})| &= \frac{2\mathbb{E}[D]}{\beta} k p_k n\alpha_n + o_{\mathbb{P}}(n\alpha_n), \text{ for every } k \geq 0, \\ |E(\mathcal{C}_{\max})| &= \frac{2\mathbb{E}[D]^2}{\beta} n\alpha_n + o_{\mathbb{P}}(n\alpha_n), \end{aligned}$$

while $|\mathcal{C}_{(2)}| = o_{\mathbb{P}}(n\alpha_n)$ and $|E(\mathcal{C}_{(2)})| = o_{\mathbb{P}}(n\alpha_n)$.

The asymptotics of $|\mathcal{C}_{\max}|$ in Theorem 4.16 can be understood by the fact that, for a branching process with offspring distribution X having mean $\mathbb{E}[X] = 1+\varepsilon$ where ε is small, the survival probability ζ satisfies $\zeta = 2\varepsilon/\mathrm{Var}(X)(1+o(1))$. Therefore, the survival probability ζ^* of the branching process with offspring distribution $D^* - 1$ is close to $2\varepsilon/\beta$, where we note that $\beta = \mathrm{Var}(D^* - 1) = \mathrm{Var}(D^*)$. Since the limit of $|\mathcal{C}_{\max}|/n$ ζ satisfies

$$\zeta = \sum_{k=1}^{\infty} p_k(1 - (1-\zeta^*)^k), \tag{4.3.4}$$

we further obtain that

$$\zeta = \zeta^* \mathbb{E}[D](1+o(1)). \tag{4.3.5}$$

The results on $|v_k(\mathcal{C}_{\max})|$ and $|E(\mathcal{C}_{\max})|$ can be understood in a similar way.

The critical behavior in the configuration model. We continue to study the critical case of $\mathrm{CM}_n(\boldsymbol{d})$ for i.i.d. degrees.

Theorem 4.17 (Weak convergence of the ordered critical clusters)**.** *Let $\boldsymbol{d} = (d_i)_{i\in[n]}$ be a sequence of i.i.d. random variables having the same distribution as D satisfying $\nu = \mathbb{E}[D(D-1)]/\mathbb{E}[D] = 1$. Let $(|\mathcal{C}_{(i)}|)_{i\geq 1}$ denote the clusters of $\mathrm{CM}_n(\boldsymbol{d})$, ordered in size.*
(a) Let $\mathbb{E}[D^3] < \infty$. Then, as $n \to \infty$,

$$\left(n^{-2/3}|\mathcal{C}_{(i)}|\right)_{i\geq 1} \xrightarrow{d} (\gamma_i)_{i\geq 1}, \tag{4.3.6}$$

in the product topology, for some non-degenerate limit $(\gamma_i)_{i\geq 1}$.
(b) Let the distribution function F of D satisfy that there exists a $\tau \in (3,4)$ and $0 < c_F < \infty$ such that

$$\lim_{x\to\infty} x^{\tau-1}[1 - F(x)] = c_F. \tag{4.3.7}$$

Then, as $n \to \infty$,

$$\big(n^{-(\tau-2)/(\tau-1)}|\mathcal{C}_{(i)}|\big)_{i\geq 1} \xrightarrow{d} (\gamma_i)_{i\geq 1}, \tag{4.3.8}$$

in the product topology, for some non-degenerate limit $(\gamma_i)_{i\geq 1}$.

Theorem 4.17 implies that as long as $\mathbb{E}[D^3] < \infty$, the scaling of the large clusters in $\mathrm{CM}_n(\boldsymbol{d})$ for i.i.d. degrees is similar to that for the Erdős-Rényi random graph $\mathrm{ER}_n(\lambda/n)$ (recall [I, Theorem 5.7 in Section 5.2.4]), and the effect of large degrees is negligible. When $\mathbb{E}[D^3] = \infty$, on the other hand, the critical scaling changes rather dramatically, and the largest critical cluster has size n^ρ, where $\rho = (\tau-2)/(\tau-1) \in (1/2, 2/3)$.

One would expect Theorem 4.17 to hold for fixed degrees, under similar (but stronger) assumptions as in [I, Condition 7.7(a)-(c)]. In particular, for Theorem 4.17(a), one would expect to need $\mathbb{E}[D_n^3] \to \mathbb{E}[D^3] < \infty$.

4.4 Similar results in related random graph models

In this section, we describe a few related models that have been investigated in the literature. See [I, Section 7.8] where these models where first introduced.

The configuration model with household structure. The configuration model has low clustering, which often makes it inappropriate in applied contexts. Indeed, in Chapter 1, we have seen that many real-world networks have a high amount of clustering instead. For example, in modeling a social network, one can expect a large amount of clustering. A possible solution to overcome this low clustering, is by introducing a *community* or *household* structure. Consider the configuration model $\mathrm{CM}_n(\boldsymbol{d})$ with a degree sequence $\boldsymbol{d} = (d_i)_{i\in[n]}$ satisfying [I, Condition 7.7(a)-(b)]. Now we replace each of the vertices by a small graph. Thus, vertex i is replaced by a local graph G_i. We assign each of the d_i half-edges incident to vertex i to a uniform vertex in G_i. As a result, we obtain a graph with two levels of hierarchy, whose local structure is described by the local graphs G_i, whereas its global structure is described by the configuration model $\mathrm{CM}_n(\boldsymbol{d})$.

The number of vertices in the configuration model with household structure is given by

$$N = \sum_{i=1}^{n} n_i, \tag{4.4.1}$$

where n_i denotes the number of vertices in G_i. Further, assume that the empirical distribution of the graph sizes converges, i.e., there exists a distribution function F_{H} such that, for every $x \in \mathbb{R}$,

$$F_{\mathrm{H},n}(x) = \frac{1}{n}\sum_{i\in[n]} \mathbb{1}_{\{n_i\leq x\}} \to F_{\mathrm{H}}(x), \tag{4.4.2}$$

and $N/n \to \mathbb{E}[H]$, where H has distribution function F_{H}. Assume further that the local graphs G_i are all *connected*. Then, assuming (4.4.2) and $N/n \to \mathbb{E}[H]$, the size

of the largest connected component $|\mathcal{C}_{\max}|$ of the configuration model with household structure satisfies

$$|\mathcal{C}_{\max}|/N \xrightarrow{\mathbb{P}} \zeta, \tag{4.4.3}$$

where ζ is the asymptotic proportion of vertices in the giant component of $\mathrm{CM}_n(\boldsymbol{d})$.

Exercise 4.11 (Law of large number for $|\mathcal{C}_{\max}|$ in CM with households)**.** *Use Theorem 4.1 to prove* (4.4.3).

It would be of interest to investigate typical distances within the configuration model with household structure. It seems reasonable to predict that $H_N/\log N$ converges to a limit when [I, Condition 7.7(a)-(c)] holds, where the limit equals $\alpha/\log\nu$, where α denotes the typical distance in a household drawn uniformly at random. Similar results should hold for infinite variance degrees, where it can be expected that $H_N/\log\log N \xrightarrow{\mathbb{P}} 2\alpha/|\log(\tau-2)|$.

Due to the fact that a vertex in the configuration model is a household, which in itself is a small subgraph, it seems plausible that the fluctuations of the typical distances *within* the different households give rise to larger fluctuations of the typical distances for the configuration model with household structure than for $\mathrm{CM}_n(\boldsymbol{d})$ as presented in Theorem 5.28–5.29. In particular, one might expect that $(H_N - \alpha\log_\nu N)/\sqrt{\log N}$ converges to a normal random variable when [I, Condition 7.7(a)-(c)] holds, while $(H_n - 2\alpha\log\log N/|\log(\tau-2)|)/\sqrt{\log\log N}$ converges to a normal random variable when a condition as in (5.1.7) holds with $\delta = 0$.

Configuration model with clustering. The low clustering of $\mathrm{CM}_n(\boldsymbol{d})$ can be resolved by introducing households as described above. Alternatively, and in the spirit of clustered inhomogeneous random graphs as described in Section 2.6, we can also introduce clustering directly. In the configuration model with clustering, we assign two numbers to a vertex $i \in [n]$. We let $d_i^{(\mathrm{si})}$ denote the number of simple half-edges incident to vertex i, and we let $d_i^{(\mathrm{tr})}$ denote the number of triangles that vertex i is part of. We say that there are $d_i^{(\mathrm{si})}$ half-edges incident to vertex i, and $d_i^{(\mathrm{tr})}$ third-triangles.

The graph is built by (a) recursively choosing two half-edges uniformly at random without replacement, and pairing them into edges (as for $\mathrm{CM}_n(\boldsymbol{d})$); and (b) choosing triples of third-triangles uniformly at random and without replacement, and drawing edges between the three vertices incident to the third-triangles that are chosen.

Let $(D_n^{(\mathrm{si})}, D_n^{(\mathrm{tr})})$ denote the number of simple edges and triangles incident to a uniform vertex in $[n]$, and assume that $(D_n^{(\mathrm{si})}, D_n^{(\mathrm{tr})}) \xrightarrow{d} (D^{(\mathrm{si})}, D^{(\mathrm{tr})})$ for some limiting distribution $(D^{(\mathrm{si})}, D^{(\mathrm{tr})})$. In [88], Newman performs a generating function analysis of when a giant component is expected to exist. The criterion Newman finds is that

$$\Big(\frac{\mathbb{E}[(D^{(\mathrm{si})})^2]}{\mathbb{E}[D^{(\mathrm{si})}]} - 2\Big)\Big(\frac{2\mathbb{E}[(D^{(\mathrm{tr})})^2]}{\mathbb{E}[D^{(\mathrm{tr})}]} - 3\Big) < \frac{2\mathbb{E}[D^{(\mathrm{si})}D^{(\mathrm{tr})}]}{\mathbb{E}[D^{(\mathrm{si})}]\mathbb{E}[D^{(\mathrm{tr})}]}. \tag{4.4.4}$$

When $D^{(\mathrm{tr})} = 0$ a.s., so that there are no triangles, this reduces to

$$\frac{\mathbb{E}[(D^{(\mathrm{si})})^2]}{\mathbb{E}[D^{(\mathrm{si})}]} - 2 > 0, \tag{4.4.5}$$

which is equivalent to $\nu = \mathbb{E}[D^{(\mathrm{si})}(D^{(\mathrm{si})}-1)]/\mathbb{E}[D^{(\mathrm{si})}] > 1$. It would be of interest to analyze this model mathematically.

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

One way to obtain a directed version of $\mathrm{CM}_n(\boldsymbol{d})$ 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 $\boldsymbol{d}^{(\mathrm{in})} = (d_i^{(\mathrm{in})})_{i\in[n]}$ to be a sequence of in-degrees, where $d_i^{(\mathrm{in})}$ denotes the in-degree of vertex i. Similarly, we let $\boldsymbol{d}^{(\mathrm{out})} = (d_i^{(\mathrm{out})})_{i\in[n]}$ be a sequence of out-degrees. Naturally, we need that

$$\sum_{i\in[n]} d_i^{(\mathrm{in})} = \sum_{i\in[n]} d_i^{(\mathrm{out})} \tag{4.4.6}$$

in order for a graph with in- and out-degree sequence $\boldsymbol{d} = (\boldsymbol{d}^{(\mathrm{in})}, \boldsymbol{d}^{(\mathrm{out})})$ to exist. We think of $d_i^{(\mathrm{in})}$ as the number of in-half-edges incident to vertex i and $d_i^{(\mathrm{out})}$ as the number of out-half-edges incident to vertex i. The directed configuration model $\mathrm{DCM}_n(\boldsymbol{d})$ 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 $d_i^{(\mathrm{in})}$ and out-degree $d_i^{(\mathrm{out})}$. Similarly to $\mathrm{CM}_n(\boldsymbol{d})$, $\mathrm{DCM}_n(\boldsymbol{d})$ 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 $(D_n^{(\mathrm{in})}, D_n^{(\mathrm{out})})$ denote the in- and out-degree of a vertex chosen uniformly at random from $[n]$. The following exercise investigates the limiting distribution of the number of self-loops in $\mathrm{DCM}_n(\boldsymbol{d})$:

Exercise 4.12 (Self-loops and multiple edges in $\mathrm{DCM}_n(\boldsymbol{d})$)**.** *Adapt the proof of Proposition 7.12 to show that when* $(D_n^{(\mathrm{in})}, D_n^{(\mathrm{out})}) \xrightarrow{d} (D^{(\mathrm{in})}, D^{(\mathrm{out})})$ *and*

$$\mathbb{E}[D_n^{(\mathrm{in})} D_n^{(\mathrm{out})}] \to \mathbb{E}[D^{(\mathrm{in})} D^{(\mathrm{out})}], \tag{4.4.7}$$

then the number of self-loops in $\mathrm{DCM}_n(\boldsymbol{d})$ *converges to a Poisson random variable with parameter* $\mathbb{E}[D^{(\mathrm{in})} D^{(\mathrm{out})}]$*. What can you say about the number of multiple edges in* $\mathrm{DCM}_n(\boldsymbol{d})$*?*

We continue to investigate the strongly connected component of $\mathrm{DCM}_n(\boldsymbol{d})$. Assume, similarly to [I, Condition 7.7(a)-(b)], that $(D_n^{(\mathrm{in})}, D_n^{(\mathrm{out})}) \xrightarrow{d} (D^{(\mathrm{in})}, D^{(\mathrm{out})})$, and that $\mathbb{E}[D_n^{(\mathrm{in})}] \to \mathbb{E}[D^{(\mathrm{in})}]$ and $\mathbb{E}[D_n^{(\mathrm{out})}] \to \mathbb{E}[D^{(\mathrm{in})}]$. Naturally, by (4.4.6), this implies that $\mathbb{E}[D^{(\mathrm{out})}] = \mathbb{E}[D^{(\mathrm{in})}]$:

Exercise 4.13 (Equivalence of convergence in- and out-degree in $\mathrm{DCM}_n(\boldsymbol{d})$)**.** *Show that* (4.4.6) *implies that* $\mathbb{E}[D^{(\mathrm{out})}] = \mathbb{E}[D^{(\mathrm{in})}]$ *when* $(D_n^{(\mathrm{in})}, D_n^{(\mathrm{out})}) \xrightarrow{d} (D^{(\mathrm{in})}, D^{(\mathrm{out})})$, $\mathbb{E}[D_n^{(\mathrm{in})}] \to \mathbb{E}[D^{(\mathrm{in})}]$ *and* $\mathbb{E}[D_n^{(\mathrm{out})}] \to \mathbb{E}[D^{(\mathrm{in})}]$.

Let

$$p_{k,l} = \mathbb{P}(D^{(\mathrm{in})} = k, D^{(\mathrm{out})} = l) \tag{4.4.8}$$

denote the asymptotic joint in- and out-degree distribution. We refer to $(p_{k,l})_{k,l\geq 0}$ simply as the asymptotic degree distribution of $\mathrm{DCM}_n(\boldsymbol{d})$. The distribution $(p_{k,l})_{k,l\geq 0}$ plays a similar role for $\mathrm{DCM}_n(\boldsymbol{d})$ as $(p_k)_{k\geq 0}$ does for $\mathrm{CM}_n(\boldsymbol{d})$. We further define

$$g_k^{(\mathrm{in})} = \sum_l l p_{k,l}/\mathbb{E}[D^{(\mathrm{out})}], \qquad g_l^{(\mathrm{out})} = \sum_k k p_{k,l}/\mathbb{E}[D^{(\mathrm{in})}]. \tag{4.4.9}$$

The distributions $(g_k^{(\mathrm{in})})_{k\geq 0}$ and $(g_k^{(\mathrm{out})})_{k\geq 0}$ correspond to the asymptotic forward in- and out-degree of a uniformly chosen edge in $\mathrm{CM}_n(\boldsymbol{d})$. Let $\theta^{(\mathrm{in})}$ and $\theta^{(\mathrm{out})}$ be the survival probabilities of the branching processes with offspring distributions $(g_k^{(\mathrm{in})})_{k\geq 0}$ and $(g_k^{(\mathrm{out})})_{k\geq 0}$, respectively, and define

$$\zeta^{(\mathrm{in})} = 1 - \sum_{k,l} p_{k,l}(1-\theta^{(\mathrm{in})})^l, \qquad \zeta^{(\mathrm{out})} = 1 - \sum_{k,l} p_{k,l}(1-\theta^{(\mathrm{out})})^k. \tag{4.4.10}$$

Then, $\zeta^{(\mathrm{out})}$ has the interpretation of the asymptotic probability that a uniform vertex has a large forward cluster, while $\zeta^{(\mathrm{in})}$ has the interpretation of the asymptotic probability that a uniform vertex has a large backward cluster. Here, the backward cluster of a vertex v consists of all vertices u that are connected to v, and the forward cluster of v consists of all vertices u for which v is connected to u.

Further, let

$$\psi = \sum_{k,l} p_{k,l}(1-\theta^{(\mathrm{in})})^l(1-\theta^{(\mathrm{out})})^k \tag{4.4.11}$$

so that ψ has the interpretation of the asymptotic probability that a uniform vertex has finite forward and backward cluster. We conclude that $1-\psi$ is the probability that a uniform vertex has either a large forward or backward cluster, and thus

$$\zeta = \zeta^{(\mathrm{out})} + \zeta^{(\mathrm{in})} - (1-\psi) \tag{4.4.12}$$

has the interpretation of the asymptotic probability that a uniform vertex has both a large forward and backward cluster. Recall that the *strongly connected component* of a vertex v is the set of u for which there are directed paths from v to u and from u to v, so that u is both in the forward and backward cluster of v. We let $\mathcal{C}_{\max}$ denote the size of the largest strongly connected component in $\mathrm{DCM}_n(\boldsymbol{d})$. Finally, we let

$$\nu = \sum_{k=0}^{\infty} k g_k^{(\mathrm{in})} = \sum_{k,l} k l p_{k,l}/\mathbb{E}[D^{(\mathrm{out})}], \tag{4.4.13}$$

Alternatively, $\nu = \sum_{k=0}^{\infty} k g_k^{(\mathrm{out})}$. Then, the main results is as follows:

Theorem 4.18 (Phase transition in $\mathrm{DCM}_n(\boldsymbol{d})$)**.** *Suppose that* $\boldsymbol{d}$ *satisfies that*

$$(D_n^{(\mathrm{in})}, D_n^{(\mathrm{out})}) \xrightarrow{d} (D^{(\mathrm{in})}, D^{(\mathrm{out})}), \quad \mathbb{E}[D_n^{(\mathrm{in})}] \to \mathbb{E}[D^{(\mathrm{in})}], \quad \mathbb{E}[D_n^{(\mathrm{out})}] \to \mathbb{E}[D^{(\mathrm{in})}], \tag{4.4.14}$$

and

$$\mathbb{E}[D_n^{(\mathrm{in})} D_n^{(\mathrm{out})}] \to \mathbb{E}[D^{(\mathrm{in})} D^{(\mathrm{out})}]. \tag{4.4.15}$$

Further, assume that $\boldsymbol{d}$ *is* proper*, as explained below.*

(a) When $\nu > 1$*,* ζ *in* (4.4.12) *satisfies* $\zeta \in (0,1]$ *and*

$$|\mathcal{C}_{\max}|/n \xrightarrow{\mathbb{P}} \zeta, \tag{4.4.16}$$

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

(b) When $\nu < 1$*,* ζ *in* (4.4.12) *satisfies* $\zeta = 0$ *and* $|\mathcal{C}_{\max}|/n \xrightarrow{\mathbb{P}} 0$ *and* $|E(\mathcal{C}_{\max})|/n \xrightarrow{\mathbb{P}} 0$*.*

Theorem 4.18 is the adaptation to $\mathrm{DCM}_n(\boldsymbol{d})$ of Theorem 4.1 for $\mathrm{CM}_n(\boldsymbol{d})$. In the statement of Theorem 4.18, we have assumed that $\boldsymbol{d}$ is *proper*, which is a technical requirement on the degree sequence stating that (a) $\mathbb{E}[(D_n^{(\mathrm{in})})^2] = O(1)$, $\mathbb{E}[(D_n^{(\mathrm{out})})^2] = O(1)$; (b) $\mathbb{E}[D_n^{(\mathrm{in})}(D_n^{(\mathrm{out})})^2] = o(n^{1/12}\log n)$. In view of the fact that such conditions do not appear in Theorem 4.1, these conditions are likely to be suboptimal for Theorem 4.18 to hold.

Random intersection graph with prescribed degrees and groups. In Section 2.6, we have studied random intersection graphs, where connections are randomly and independently formed between individuals and groups. We now describe a model in which vertex $v \in [n]$ belongs to $d_v^{(\mathrm{ve})}$ groups, while group $g \in [m]$ has size $d_g^{(\mathrm{gr})}$. Here n is the number of individuals, while m is the number of groups. Naturally, in order for the model to be well defined, we need that

$$\sum_{v\in[n]} d_v^{(\mathrm{ve})} = \sum_{g\in[m]} d_g^{(\mathrm{gr})}. \tag{4.4.17}$$

We call two vertices v_1 and v_2 neighbors when they are connected to the same group, so that the degree of a vertex v is the total number of other vertices u for which there exists a group of which both u and v are members. This model might be appropriate as a simple model for collaboration graphs such as the IMDb and the collaboration graph among mathematicians. In the above setting, this model has not received much attention in the mathematical community.

4.5 Notes and discussion

Notes on Section 4.1 This section is adapted from [78], which, in turn, generalizes the results in [84, 85]. The results in [84, 85] are not phrased in terms of branching processes, which makes them a bit more difficult to grasp. We have chosen to reformulate the results using branching process terminology.

Notes on Section 4.2. These results are folklore. A version of Theorem 4.14 can be found in [31, Lemma 1.2]. We could not find the precise version stated in Theorem 4.14.

Notes on Section 4.4. The configuration model with household structure was investigated in [13, 14] in the context of epidemics on social networks. Particularly when studying epidemics on networks, clustering is highly relevant, as clustering slows down the spread of infectious diseases. Random intersection graph with prescribed degrees and groups are studied in a non-rigorous way in [91, 92].

The directed configuration model was investigated in [39], where the results discussed here are proved. In fact, the results in [39] are much more detailed than the one in Theorem 4.18, and also include detailed bounds on the strongly connected component in the subcritical regime, as well as precise bounds on the number of vertices whose forward and backward clusters are large and the asymptotic size of forward and backward clusters.

Chapter 5
Small-world phenomena in the configuration model

Abstract

In this chapter, we investigate the connectivity structure of the configuration model by investigating its typical distances.

Organization of this chapter. This chapter is organized as follows. In Section 5.1, we study the typical graph distance in the configuration model. In Section 5.2, we prove these distance results, using path counting techniques and comparisons to branching processes. In Section 5.3, we identify the diameter of the configuration model when it has infinite-variance degrees. In Section 5.4, we study infinite-mean branching processes, as these arise in the configuration model with infinite-variance degrees. In Section 5.5, we state further results in the configuration model.

5.1 The small-world phenomenon in $\mathrm{CM}_n(\boldsymbol{d})$

In this section, we study distances in the configuration model, both in the case of finite-variance degrees as well as in the case of infinite variance degrees.

Finite-variance degrees. We start by analyzing the typical graph distance in the case where the configuration model $\mathrm{CM}_n(\boldsymbol{d})$ when [I, Condition 7.7(a)-(c)] holds:

Theorem 5.1 (Typical distances in $\mathrm{CM}_n(\boldsymbol{d})$ for finite-variance degrees)**.** *In the configuration model $\mathrm{CM}_n(\boldsymbol{d})$, where the degrees $\boldsymbol{d} = (d_i)_{i\in[n]}$ satisfy [I, Condition 7.7(a)-(c)] and where $\nu > 1$, conditionally on $H_n < \infty$,*

$$H_n/\log n \xrightarrow{\mathbb{P}} 1/\log \nu. \tag{5.1.1}$$

Theorem 5.1 shows that the typical distances in $\mathrm{CM}_n(\boldsymbol{d})$ are of order $\log_\nu n$, and is thus similar in spirit as Theorem 3.3. We shall see that also its proof is quite similar.

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. Recall that $G^\star_D(x)$ is defined in (4.1.9) as the probability generating function of $g = (g_k)_{k\geq 0}$ defined in (4.1.2). We recall that ξ is the extinction probability of the branching process with offspring distribution g defined in (4.1.6) and further define

$$\mu = G^\star_D(\xi) = \sum_{k=1}^{\infty} k\xi^{k-1} g_k. \tag{5.1.2}$$

When $\xi < 1$, we also have that $\mu \leq 1$. Then, the main result is as follows:

Theorem 5.2 (Diameter of the configuration model)**.** *Let [I, Condition 7.7(a)-(b)] hold. Assume that $n_1 = 0$ when $p_1 = 0$, and that $n_2 = 0$ when $p_2 = 0$. Then,*

$$\frac{\mathrm{diam}(\mathrm{CM}_n(\boldsymbol{d}))}{\log n} \xrightarrow{\mathbb{P}} \frac{1}{\log \nu} + \frac{(2 - \mathbb{1}_{\{p_1=0\}} - \mathbb{1}_{\{p_2=0\}})}{|\log \mu|}. \tag{5.1.3}$$

We note that, by Theorem 5.1 and Theorem 5.2, the diameter of the configuration model is strictly larger than the typical graph distance, except when $p_1 = p_2 = 0$. In the latter case, the degrees are at least three, so that thin lines are not possible, and the configuration model is whp *connected* (recall Theorem 4.14). We also remark that Theorem 5.2 applies not only to the finite variance case, but also to the finite mean and infinite variance case. In the latter case, the diameter is of order $\log n$ unless $p_1 = p_2 = 0$, in which case Theorem 5.2 implies that the diameter is $o_{\mathbb{P}}(\log n)$. By [I, Corollary 7.15], Theorem 5.2 also applies to uniform random graphs with a given degree sequence. This shall be used in the examples below:

Random regular graphs. Let r be the degree of the random regular graph, where $r \geq 3$. By Corollary 7.15, the diameter of a random regular r-graph has with high probability the same asymptotics as the diameter of $\mathrm{CM}_n(\boldsymbol{d})$, where $d_i = r$ with probability 1. Thus, $p_r = 1$ and $p_i = 0$ for any $i \neq r$. We assume that nr is even, so that the degree sequence is feasible. It is not hard to see that all assumptions of Theorem 5.2 are satisfied. Moreover, $\nu = r - 1$. When $r \geq 3$, we thus obtain that

$$\frac{\mathrm{diam}(\mathrm{CM}_n(\boldsymbol{d}))}{\log n} \xrightarrow{\mathbb{P}} \frac{1}{\log (r-1)}. \tag{5.1.4}$$

When $r = 2$, on the other hand, the graph is *critical*, so that there is no giant component. Since $\nu = 1$, we have that $\mu = \nu = 1$, so that $\mathrm{diam}(\mathrm{CM}_n(\boldsymbol{d})) \gg \log n$. This is quite reasonable, since the graph will consist of a collection of *cycles*. The diameter of such a graph is equal to half the longest cycle.

Exercise 5.1 (Diameter of soup of cycles)**.** *Prove that in a graph consisting solely of cycles, the diameter is equal to the longest cycle divided by 2.*

Exercise 5.2 (Longest cycle 2-regular graph)**.** *What is the size of the longest cycle of the 2-regular graph?*

Erdős-Rényi random graph. We next study the diameter of $\mathrm{ER}_n(\lambda/n)$. We let $\lambda > 1$. By Proposition 4.10, [I, Condition 7.7(a)-(b)] holds with $p_k = \mathrm{e}^{-\lambda}\frac{\lambda^k}{k!}$. Also, $\mu = \mu_\lambda$, the dual parameter in [I, (3.6.6)].

Exercise 5.3 (Parameters for $\mathrm{ER}_n(\lambda/n)$)**.** *Prove that $\nu = \lambda$ and $\mu = \mu_\lambda$.*

We again make essential use of [I, Theorem 7.16], which relates the configuration model and the generalized random graph. We note that $\mathrm{ER}_n(\lambda/n)$ is the same as $\mathrm{GRG}_n(\boldsymbol{w})$, where (recall [I, Exercise 6.1])

$$w_i = \frac{n\lambda}{n-\lambda}. \tag{5.1.5}$$

Clearly, $\boldsymbol{w} = (n\lambda/(n-\lambda))_{i\in[n]}$ satisfies [I, Conditions 6.4(a)-(c)], so that also the degree sequence of $\mathrm{ER}_n(\lambda/n)$ satisfies [I, Conditions 7.7(a)-(c)], where the convergence holds in probability (recall Proposition 4.10). From the above identifications and using [I, Theorem 7.16], we find that

$$\frac{\mathrm{diam}(\mathrm{ER}_n(\lambda/n))}{\log n} \xrightarrow{\mathbb{P}} \frac{1}{\log \lambda} + \frac{2}{|\log \mu_\lambda|}. \tag{5.1.6}$$

This identifies the diameter of the Erdős-Rényi random graph.

Finite mean, infinite variance degrees. We next study the typical distance of the configuration model with degrees having finite mean and infinite variance. We start by formulating the precise condition on the degrees that we shall work with. This condition is identical to the condition on F_n for $\mathrm{NR}_n(\boldsymbol{w})$ formulated in (3.2.2). Recall that $F_n(x)$ denotes the proportion of vertices having degree at most x. Then, we assume that there exists a $\tau \in (2,3)$ and for all $\delta > 0$, there exist $c_1 = c_1(\delta)$ and $c_2 = c_2(\delta)$ such that, uniformly in n,

$$c_1 x^{-(\tau-1+\delta)} \le [1-F_n](x) \le c_2 x^{-(\tau-1-\delta)}, \tag{5.1.7}$$

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

Theorem 5.3 (Typical distances in $\mathrm{CM}_n(\boldsymbol{d})$ for $\tau \in (2,3)$)**.** *Let the degrees $\boldsymbol{d} = (d_i)_{i\in[n]}$ in the configuration model $\mathrm{CM}_n(\boldsymbol{d})$ satisfy [I, Condition 7.7(a)-(b)] and (5.1.7). Then, conditionally on $H_n < \infty$,*

$$\frac{H_n}{\log\log n} \xrightarrow{\mathbb{P}} \frac{2}{|\log(\tau-2)|}. \tag{5.1.8}$$

Theorem 5.3 is similar in spirit to Theorem 3.4 for $\mathrm{NR}_n(\boldsymbol{w})$.

5.2 Proofs of small-world results $\mathrm{CM}_n(\boldsymbol{d})$

In this section, we give the proofs of Theorems 5.1 and 5.3 describing the small-world properties in $\mathrm{CM}_n(\boldsymbol{d})$. These proofs are adaptations of the proofs of Theorems 3.3 and 3.4, and we focus on the differences in the proofs. This section is organized as follows. In Section 5.2.1 we give a branching process approximation for the neighborhoods of a pair of uniform vertices in $\mathrm{CM}_n(\boldsymbol{d})$. In Section 5.2.2 we perform similar path counting techniques as in Section 3.5.1.

5.2.1 Branching process approximation

In this section, we give a convenient description of the breadth-first exploration in $\mathrm{CM}_n(\boldsymbol{d})$, and relate this to a branching process. We start by describing the algorithm.

Breadth-first search in $\mathrm{CM}_n(\boldsymbol{d})$. Suppose we start in a vertex i having degree d_i, so that there are i half-edges incident to vertex i. Set $S_0 = d_i$. In the course of our exploration, S_t will denote the number of unpaired half-edges incident to the vertices found in the exploration after having paired t half-edges.

At a time step $t \geq 1$ in our exploration, we pair one half-edge, say x_t, to its brother, which we denote by y_t. For a half-edge x, we let V_x denote the vertex to which x is incident. Let this half-edge be incident to vertex V_{y_t}. Let $X_t + 1$ denote the change in the number of unpaired half-edges incident to the vertices found in the exploration after pairing x_t, so that, for $t \geq 1$,

$$\begin{cases} X_t = d_{V_{y_t}} - 1 & \text{when } V_{y_t} \notin \{i, V_{y_1}, \ldots, V_{y_{t-1}}\}, \\ X_t = -1 & \text{otherwise.} \end{cases} \tag{5.2.1}$$

Then, the number S_t of unpaired half-edges incident to $\{i, V_{y_1}, \ldots, V_{y_t}\}$ satisfies

$$S_t = S_{t-1} + X_t - 1. \tag{5.2.2}$$

We perform this exploration in a breadth-first way, so that the half-edge x_t that is paired at time t is chosen from that half-edges that are incident to vertices that minimize the graph distance between i and the vertices in $\{V_{y_1}, \ldots, V_{y_{t-1}}\}$ that still have unpaired half-edges incident to it. In our breadth-first exploration, it will be convenient to identify the graph distance between a half-edge x and a vertex $j \in [n]$ by the graph distance between j and the vertex V_x to which x is incident. The precise order in which the half-edges at the same graph distance from the root vertex i are paired is arbitrary.

Let, by convention, $Z_0(i) = 1$ and, for $m \geq 1$, let $Z_m(i)$ denote the number of unpaired half-edges incident to vertices at graph distance $m-1$, so that $Z_1(i) = d_i = S_0$. Thus, $Z_2(i)$ is obtained after pairing all the $Z_1(i)$ half-edges at distance 1 from the root. Then, $Z_m(i)$ equals $S_{T_m(i)}$, where $T_m(i)$ is the time where all the $Z_{m-1}(i)$ half-edges at distance $m-2$ from vertex i have been paired. The above describes the breadth-first exploration from a single vertex i. Let, as usual, V_1 and V_2 be two vertices chosen uniformly at random from $[n]$, and denote $Z_m^{(i)} = Z_m(V_i)$, so that $Z_m^{(i)}$ is the number of unpaired half-edges at distance $m-1$ from vertex V_i. The following proposition shows that, for some $m_n \to \infty$ sufficiently slowly, the processes $(Z_l^{(1)}, Z_l^{(2)})_{k=0}^{m_n}$ are close to two independent two-stage branching processes:

Proposition 5.4 (Coupling of neighborhoods of two vertices)**.** *Let the degrees $\boldsymbol{d} = (d_i)_{i\in[n]}$ satisfy [I, Condition 7.7(a)-(b)]. Let $(\mathcal{Z}_l^{(1)}, \mathcal{Z}_l^{(2)})_{l\geq 0}$ be two independent branching processes with offspring distribution D in the first generation, and offspring distribution $D^* - 1$ in all further generations. Then, there exists $m_n \to \infty$ such that*

$$\mathbb{P}(H_n \leq 2m_n) = o(1), \tag{5.2.3}$$

and a coupling $\big((\hat{Z}_l^{(1)}, \hat{Z}_l^{(2)})_{l=0}^{m_n}, (\hat{\mathcal{Z}}_l^{(1)}, \hat{\mathcal{Z}}_l^{(2)})_{l=0}^{m_n}\big)$ *of* $(Z_l^{(1)}, Z_l^{(2)})_{l=0}^{m_n}$ *and* $(\mathcal{Z}_l^{(1)}, \mathcal{Z}_l^{(2)})_{l=0}^{m_n}$*, such that*

$$\mathbb{P}\big((\hat{Z}_l^{(1)}, \hat{Z}_l^{(2)})_{l=0}^{m_n} \neq (\hat{\mathcal{Z}}_l^{(1)}, \hat{\mathcal{Z}}_l^{(2)})_{l=0}^{m_n}\big) = o(1). \tag{5.2.4}$$

In words, (5.2.4) states that, whp, we can perfectly couple the neighborhoods of V_1 and V_2 up to distance at most m_n to two independent two-stage branching processes. As explained below Theorem 4.1 (see (4.1.4)), when pairing a single half-edge, the probability that it connects to a vertex of degree $k+1$, this adding k unpaired half-edges to the exploration process, is close to $g_k = (k+1)p_{k+1}/\mathbb{E}[D]$ (recall (4.1.2)). Here we recall that $(p_k)_{k\geq 0}$ is the probability mass function of D, so that $p_k = \mathbb{P}(D = k)$, and note that $(g_k)_{k\geq 0}$ is the probability mass function of $D^* - 1$ in (4.1.2).

When we pair only a few half-edges, most half-edges are incident to vertices that are have not been found in the exploration process so far, and thus this distribution hardly changes. When this distribution would be unchanged, then $(X_t)_{t\geq 2}$ in (5.2.1) would be an i.i.d. sequence, thus leading us to a two-stage branching process. However, the random variables $(X_t)_{t\geq 2}$ are *not* i.i.d., since the half-edges are being used up in the process of their pairing.

There are three effects that make the exploration in $\mathrm{CM}_n(\boldsymbol{d})$ different from the limiting two-stage branching process: (a) the fact that the number of unpaired half-edges decreases, which is similar to the depletion-of-points effect for $\mathrm{ER}_n(\lambda/n)$; (b) sometimes we pair to a half-edge that is already incident to a vertex in the exploration process, so that $X_t = -1$ and we have created a cycle; and (c) the degree distribution in $\mathrm{CM}_n(\boldsymbol{d})$ is not equal to the limiting degree distribution $(p_k)_{k\geq 0}$. The above intuition argues that these effects are negligible in the first few generations. To make this intuition precise, we need to investigate the cluster exploration in more detail. In order to describe the degrees of vertices that occur during our exploration, the notion of a *size-biased reordering* is crucial:

Definition 5.5 (Size-biased reordering)**.** *Given an index set* $\mathcal{I}$ *containing* N *elements, and a sequence* $\boldsymbol{d} = (d_i)_{i\in\mathcal{I}}$*, so that element* i *has weight* d_i*, the* size-biased re-ordering *of* $\mathcal{I}$ *is the random order of the elements* $(v(1), v(2), \ldots, v(N))$ *where we choose vertex* $v(1)$ *randomly from* $[N]$ *with*

$$\mathbb{P}(v(1) = l) = \frac{d_l}{\sum_{j\in\mathcal{I}} d_j}, \tag{5.2.5}$$

and then select $v(2)$ *from the set* $\mathcal{I}\setminus\{v(1)\}$ *again with probability proportional to the weights of the remaining elements and so on.*

The connection to the exploration of the neighborhood of a vertex in $\mathrm{CM}_n(\boldsymbol{d})$ is made in the following lemma. In its statement, we consider the exploration in (5.2.1)–(5.2.2), where, if $S_{t-1} = 0$, we draw an unpaired half-edge x uniformly at random, take $X_t = d_{V_x} - 1$ and $S_t = S_{t-1} + X_t + 1$ instead of (5.2.2). Then, the recursions (5.2.1)–(5.2.2) can be continued as long as there are unpaired half-edges available, and we explore the connected components one by one, the first being special as it starts in vertex i.

Make notation consistent with Chapter 8!

Lemma 5.6 (Size-biased reordering in $\mathrm{CM}_n(\boldsymbol{d})$)**.** *Let B_j be the jth non-zero value of $(X_t+1)_{t\geq 1}$. Then,*
(a) $(B_j)_{j\in[n-1]}$ is a size-biased reordering of $[n-1]$ with weight sequence $(d_j)_{j\in[n]\setminus\{i\}}$.
(b) $(B_j)_{j\in[n-1]}$ can be coupled to an i.i.d. sequence $(\tilde{B}_j)_{j\in[n-1]}$, where $\tilde{B}_j$ has distribution D_n^*, the size-biased version of D_n, such that

$$\mathbb{P}(\exists j\leq l\colon B_j\neq \tilde{B}_j)\leq ld_i/\ell_n + l^2\mathbb{E}[D_n^*]/\ell_n. \tag{5.2.6}$$

Proof. Since B_j is the jth non-zero value of $(X_t+1)_{t\geq 1}$, it means that $B_j = d_{V_{x_{t_j}}}$ where x_{t_j} is incident to the jth vertex found in the exploration process and t_j is the time this occurs. Each time we choose a new vertex, we do so with probability proportional to its degree, since we choose a half-edge uniformly at random from the unpaired ones. This proves the claim for all j for which $S_{t_j}\geq 1$. However, by construction, when $S_{t_j}=0$, $X_{t_j}+1 = d_{V_x}$, where the half-edge x is chosen uniformly at random from all unpaired half-edges. Since $S_{t_j}=0$ implies that there are no half-edges incident to vertices found already, this means that we find a *new* vertex, which completes the proof of the first part.

For the second part, we let $(\tilde{B}_j)_{j\in[n-1]}$ be an i.i.d. sequence of random variables with distribution D_n^*. Then, the sequence $(B_j)_{j\in[n-1]}$ corresponds to drawing from the distribution in (5.2.5) *without replacement from* $[n]$, while $(\tilde{B}_j)_{j\in[n-1]}$ corresponds to drawing *with replacement from* $[n]\setminus\{i\}$. Therefore, the distribution of $(B_j)_{j=1}^l$ is equal to that of $(\tilde{B}_j)_{j=1}^l$ conditioned on $v(s)\neq i$ for all $s\in[l]$ and $v(s)\neq v(t)$ for every $s,t\in[l]$ with $s\neq t$. Take $(\tilde{v}(s))_{s\in[n-1]}$ an i.i.d. sequence of random variables with distribution

$$\mathbb{P}(\tilde{v}(1)=j) = d_j/\ell_n, \tag{5.2.7}$$

and let $\tilde{B}_j = d_{\tilde{v}(j)}$. Thus, we can couple $(B_j)_{j=1}^l$ and $(\tilde{B}_j)_{j=1}^l$ by taking $B_j = d_{\tilde{v}(s_j)}$, where s_j is the jth distinct element in $(\tilde{v}(s))_{s\in[n-1]}$ that is unequal to i.

The above provides a perfect coupling between $(B_j)_{j=1}^l$ and $(\tilde{B}_j)_{j=1}^l$ as long as i is not drawn and no repetition occurs before time l. Now, for every $s,t\in[l]$ with $s\neq t$, the probability that $\tilde{v}(s)=i$ is equal to

$$\mathbb{P}(\tilde{v}(s)=i) = d_i/\ell_n, \tag{5.2.8}$$

while the probability that $\tilde{v}(s)=\tilde{v}(t)$ is at most

$$\begin{aligned}\mathbb{P}(\tilde{v}(s)=\tilde{v}(t)) &= \sum_{j\in[n]}\mathbb{P}(\tilde{v}(s)=j)\mathbb{P}(\tilde{v}(s)=j) = \sum_{j\in[n]}\mathbb{P}(v(1)=j)^2 \\ &= \sum_{j\in[n]}\frac{d_j^2}{\ell_n^2} = \mathbb{E}[D_n^*]/\ell_n.\end{aligned} \tag{5.2.9}$$

Boole's inequality yields the result. □

In the following lemma, we couple the sequence $(X_t)_{t\geq 1}$ in (5.2.1) to a sequence of i.i.d. random variables $(Y_t)_{t\geq 1}$ with distribution D^*-1. In order to state the result,

and for each $l \geq 1$, we define the stopping time

$$T_l^X = \inf\{t\colon S_t + t \geq l\} = \inf\{t\colon d_{V_1} + \sum_{s=1}^{t}(X_s + 1) \geq l\}, \tag{5.2.10}$$

so that for all $t < T_l^X$, the number of half-edges found in the exploration process is at most l.

Lemma 5.7 (Coupling to a branching process for $\mathrm{CM}_n(\boldsymbol{d})$)**.** *Assume that [I, Condition 7.7(a)-(b)] holds. Take* $a_n = (\ell_n/\mathbb{E}[D_n^*])^{1/3} \wedge (d_{\mathrm{TV}}(g^{(n)}, g))^{-1/2} \to \infty$. *The random variables* $(X_t)_{t=0}^{T_{a_n}^X}$ *in (5.2.1) can be coupled to a sequence of i.i.d. random variables* $(Y_t)_{t=1}^{T_{a_n}^Y}$ *with distribution* $D^\star - 1$, *such that*

$$\mathbb{P}\Big((X_t)_{t=0}^{T_{a_n}^X} \neq (Y_t)_{t=1}^{T_{a_n}^Y}\Big) = o(1). \tag{5.2.11}$$

Proof. We note that $S_t \leq l$ for all $t < T_l^X$, and the number of unpaired half-edges at time $t < T_l^X$ is at least $\ell_n - l$, so that

$$\mathbb{P}(X_t = -1, t < T_l^X) \leq l/(\ell_n - l). \tag{5.2.12}$$

We let $a_n \to \infty$ in a way to be determined later on. Then,

$$\mathbb{P}(\exists t < T_{a_n}^X\colon X_t = -1) \leq a_n^2/(\ell_n - a_n), \tag{5.2.13}$$

which converges to zero as long as $a_n = o(\sqrt{n})$. As a result, whp, no cycles are formed up to the time where the number of edges in the cluster of a vertex is of the order $a_n = o(\sqrt{n})$. Denote the stopping time T by

$$T = \inf\{t\colon S_t = 0\}, \tag{5.2.14}$$

so that T equals the number of edges in $\mathcal{C}(i)$. By (5.2.13), whp,

$$(X_t)_{t=1}^{T\wedge T_{a_n}^X} = (d_{V_{x_t}} - 1)_{t=1}^{T\wedge T_{a_n}^X} = (B_t - 1)_{t=1}^{T\wedge T_{a_n}^X}. \tag{5.2.15}$$

By Lemma 5.6, when i is chosen uniformly from $[n]$,

$$\mathbb{P}(\exists j \leq l\colon B_j \neq \tilde{B}_j) \leq \frac{l}{n}\sum_{i\in[n]}\frac{d_i}{\ell_n} + \frac{l^2\mathbb{E}[D_n^*]}{\ell_n} = \frac{l}{n} + \frac{l^2\mathbb{E}[D_n^*]}{\ell_n}. \tag{5.2.16}$$

Note that [I, Condition 7.7(a)-(b)] implies that $\mathbb{E}[D_n^*]/\ell_n = o(1)$, so that, when $a_n \leq (\ell_n/\mathbb{E}[D_n^*])^{1/3}$, the probability that the coupling between $(X_t)_{t=1}^{T\wedge T_{a_n}^X}$ and the i.i.d. sequence $(B_t - 1)_{t=1}^{T\wedge T_{a_n}^X}$ fails is bounded from above by

$$\mathbb{P}(\exists t \leq a_n\colon X_t \neq B_t - 1) \leq (\ell_n/\mathbb{E}[D_n^*])^{1/3}/n + (\mathbb{E}[D_n^*]/\ell_n)^{-1/3} = o(1). \tag{5.2.17}$$

Recall that $(\tilde{B}_t - 1)_{t\geq 1}$ is an i.i.d. sequence of random variables with distribution $D_n^* - 1$. By [I, Condition 7.7(a)-(b)], $D_n^* - 1 \xrightarrow{d} D^* - 1$. The probability mass

function of $D_n^* - 1$ is equal to $g^{(n)}$ as defined in (4.1.4). Since $D_n^* - 1$ is an integer-valued random variable, also $d_{\mathrm{TV}}(g^{(n)}, g) \to 0$. By Theorem 2.9, there exists a coupling $(\hat{D}_n^* - 1, \hat{D}^* - 1)$ of $D_n^* - 1$ and $D^* - 1$ such that

$$\mathbb{P}(\hat{D}_n^* - 1 \neq \hat{D}^* - 1) = d_{\mathrm{TV}}(g^{(n)}, g). \tag{5.2.18}$$

As a result, there exists a coupling $\big((\hat{X}_t)_{t=1}^{a_n}, (\hat{Y}_t)_{t=1}^{a_n}\big)$ of random variables with distribution $D^* - 1$ such that

$$\mathbb{P}((\hat{X}_t)_{t=1}^{a_n} \neq (\hat{Y}_t)_{t=1}^{a_n}) \leq (\ell_n/\mathbb{E}[D_n^*])^{1/3}/n + (\mathbb{E}[D_n^*]/\ell_n)^{-1/3} + a_n d_{\mathrm{TV}}(g^{(n)}, g) = o(1). \tag{5.2.19}$$

This completes the proof. □

Proof of Proposition 5.4. We start by showing that there exists $m_n \to \infty$ such that and a coupling $\big((\hat{Z}_l^{(1)})_{l=0}^{m_n}, (\hat{\mathcal{Z}}_l^{(1)})_{l=0}^{m_n}\big)$ of $(Z_l^{(1)})_{l=0}^{m_n}$ and $(\mathcal{Z}_l^{(1)})_{l=0}^{m_n}$ such that

$$\mathbb{P}\big((\hat{Z}_l^{(1)})_{l=0}^{2m_n} \neq (\hat{\mathcal{Z}}_l^{(1)})_{l=0}^{m_n}\big) = o(1). \tag{5.2.20}$$

By construction $S_0 = d_V$, where $V \in [n]$ is chosen uniformly at random from $[n]$, so that d_V has the same distribution as D_n. By [I, Condition 7.7(a)], $D_n \xrightarrow{d} D$, so that also $d_{\mathrm{TV}}(p^{(n)}, p) = o(1)$. By Lemma 5.7, $(S_t)_{t=0}^{a_n}$ in (5.2.2) can be whp perfectly coupled to $(S_t^Y)_{t=0}^{T_{a_n}^X}$, where

$$S_t^Y = D + \sum_{s=1}^{t} (Y_s - 1). \tag{5.2.21}$$

Define

$$T^Y = 1 + \inf\{t\colon S_t^Y = 0\}, \tag{5.2.22}$$

so that T^Y is the total progeny of a two-stage branching process where the root has offspring D, and each other individual has offspring $D^* - 1$. This gives a perfect coupling of the breadth-first exploration of the neighborhood of a uniform vertex in the random graph to a branching process up to the moment that the branching process tree has size at most a_n. Pick a_n such that $a_n = (\ell_n/\mathbb{E}[D_n^*])^{1/3} \wedge (d_{\mathrm{TV}}(g^{(n)}, g))^{-1/2} \to \infty$ and take m_n such that

$$\mathbb{P}\Big(\sum_{l=0}^{2m_n} \mathcal{Z}_l^{(1)} > a_n\Big) = o(1). \tag{5.2.23}$$

Then, (5.2.20) follows.

To prove (5.2.3), we apply the above to $2m_n$, and note that whp, we can perfectly couple $(Z_l^{(1)})_{l=0}^{2m_n}$ to $(\mathcal{Z}_l^{(1)})_{l=0}^{2m_n}$. Note that $H_n \leq 2m_n$ implies that V_2 equals one of the vertices found in the exploration of $(Z_l^{(1)})_{l=0}^{2m_n}$. By (5.2.23) and $a_n = o(n)$ and since $V_2 \in [n]$ is chosen uniformly, this has conditional probability given $(Z_l^{(1)})_{l=0}^{2m_n}$ equal to

$$\frac{1}{n}\sum_{l=0}^{2m_n} Z_l^{(1)} \xrightarrow{\mathbb{P}} 0. \tag{5.2.24}$$

As a result, $\mathbb{P}(H_n \leq 2m_n) = o(1)$, as required.

To prove (5.2.4), we note that $\mathbb{P}(H_n \leq 2m_n) = o(1)$, so that, whp, $H_n > 2m_n$. By (5.2.20), we can whp perfectly couple $(Z_l^{(1)})_{l=0}^{m_n}$ and $(\mathcal{Z}_l^{(1)})_{l=0}^{m_n}$. Conditionally on $(Z_l^{(1)})_{l=0}^{m_n}$ and on $H_n > 2m_n$, $(Z_l^{(2)})_{l=0}^{m_n}$ describes the exploration of a uniform vertex in $[n] \setminus \mathcal{N}_{\leq m_n}(V_1)$ with degrees $(d_i)_{i\in[n]\setminus\mathcal{N}_{\leq m_n}(V_1)}$. The degree sequence $(d_i)_{i\in[n]\setminus\mathcal{N}_{\leq m_n}(V_1)}$ satisfies [I, Condition 7.7(a)-(b)] precisely when $(d_i)_{i\in[n]}$ does. As a result, all the arguments used in the proof of (5.2.20) apply to this setting as well, so that we obtain

$$\mathbb{P}\Big((Z_l^{(2)})_{l=0}^{m_n} \neq (\mathcal{Z}_l^{(2)})_{l=0}^{m_n} \mid (Z_l^{(1)})_{l=0}^{m_n}\Big) \xrightarrow{\mathbb{P}} 0. \tag{5.2.25}$$

Since $(\mathcal{Z}_l^{(2)})_{l=0}^{m_n}$ is independent of $(Z_l^{(1)})_{l=0}^{m_n}$, also $(\mathcal{Z}_l^{(1)})_{l=0}^{m_n}$ and $(\mathcal{Z}_l^{(2)})_{l=0}^{m_n}$ are independent, which completes the proof of (5.2.4), and thus that of Proposition 5.4. □

With Proposition 5.4 at hand, we can show that $Z_{m_n}^{(1)} \xrightarrow{\mathbb{P}} \infty$ when $Z_{m_n}^{(1)} \geq 1$ and $(\mathcal{Z}_l^{(1)})_{l\geq 0}$ is *supercritical*:

Corollary 5.8 (Coupling to a branching process for $\mathrm{CM}_n(\boldsymbol{d})$)**.** *Assume that [I, Condition 7.7(a)-(b)] holds. Take m_n as in Proposition 5.4, and assume that $\mathbb{E}[D^*-1] > 1$. Then,*

$$\mathbb{P}(Z_{m_n}^{(1)} \geq 1, Z_{m_n}^{(2)} \geq 1) \to \zeta^2, \tag{5.2.26}$$

and, conditionally on $Z_{m_n}^{(1)} \geq 1, Z_{m_n}^{(2)} \geq 1$,

$$Z_{m_n}^{(1)} \xrightarrow{\mathbb{P}} \infty, \qquad Z_{m_n}^{(2)} \xrightarrow{\mathbb{P}} \infty. \tag{5.2.27}$$

Proof. Equation (5.2.26) immediately follows from the fact that (5.2.4) in Proposition 5.4, as well as the fact that

$$\mathbb{P}(\mathcal{Z}_{m_n}^{(1)} \geq 1, \mathcal{Z}_{m_n}^{(2)} \geq 1) = \mathbb{P}(\mathcal{Z}_{m_n}^{(1)} \geq 1)\mathbb{P}(\mathcal{Z}_{m_n}^{(2)} \geq 1) \to \zeta^2. \tag{5.2.28}$$

We prove (5.2.27) when $\mathbb{E}[D^* - 1] < \infty$, the case where $\mathbb{E}[D^* - 1] = \infty$ is left as an exercise. By Proposition 5.4, $(Z_l^{(1)}, Z_l^{(2)})_{l=0}^{m_n}$ and $(\mathcal{Z}_l^{(1)}, \mathcal{Z}_l^{(2)})_{l=0}^{m_n}$ can be coupled such that whp these vectors are equal. Therefore, (5.2.27) follows when $\mathcal{Z}_{m_n}^{(1)} \xrightarrow{\mathbb{P}} \infty$ conditionally on $\mathcal{Z}_{m_n}^{(1)} \geq 1$. By Theorems 3.9 and 3.10 and the fact that the branching process is supercritical when $\mathbb{E}[D^* - 1] > 1$, on the event of survival, $\mathcal{Z}_m^{(1)}$ grows exponentially in m. Further, $\mathbb{P}(\mathcal{Z}_{m_n}^{(1)} \geq 1)$ converges to the survival probability ζ since $m_n \to \infty$ when $n \to \infty$ by Proposition 5.4. This completes the proof. □

Exercise 5.4 (Infinite variance degrees)**.** *Complete the proof of (5.2.27) in Corollary 5.8 in the case where $\mathbb{E}[D^* - 1] = \infty$.*

Exercise 5.5 (Random regular graph)**.** *Fix $r \geq 2$ and consider the r-regular graph on n vertices, where nr is even. Show that $d_{\mathrm{TV}}(g^{(n)}, g) = 0$, and conclude that we can take $m_n = a \log_{r-1}(n)$ for any $a < 1/6$ in Proposition 5.4. Is this optimal?*

5.2.2 Path counting techniques

In this section, we present path counting techniques similar to those in Section 3.5.1. Since $\mathrm{CM}_n(\boldsymbol{d})$ is a multigraph, and not a simple graph as $\mathrm{NR}_n(\boldsymbol{w})$, we need to be precise what a path in $\mathrm{CM}_n(\boldsymbol{d})$ is. We start by introducing some notation.

A path π of length k in $\mathrm{CM}_n(\boldsymbol{d})$ means a sequence

$$\pi = \{(\pi_0, s_0), (\pi_1, s_1, t_1), \dots, (\pi_{k-1}, s_{k-1}, t_{k-1}), (\pi_k, t_k)\}, \tag{5.2.29}$$

where $\pi_i \in [n]$ denotes the ith vertex along the path, and $s_i \in [d_{\pi_i}]$ denotes the label of the half-edge incident to π_i and $t_{i+1} \in [d_{\pi_{i+1}}]$ denotes the label of the half-edge incident to π_{i+1}. In particular, multiple edges between π_i and π_{i+1} give rise to *distinct* paths through the same vertices. For a path π, we write $\pi \subset \mathrm{CM}_n(\boldsymbol{d})$ when the path π in (5.2.29) is present in $\mathrm{CM}_n(\boldsymbol{d})$, so that the half-edge corresponding to s_i is paired with the half-edge corresponding to t_{i+1} for $i = 0, \dots, k-1$. We assume throughout that the path π is *simple*, i.e., $\pi_0, \dots, \pi_k$ are distinct vertices.

In this section, we perform first and second moment computations on the number of paths present in $\mathrm{CM}_n(\boldsymbol{d})$. We start by proving upper bounds on the expected number of paths.

Upper bounds on the expected number of paths in $\mathrm{CM}_n(\boldsymbol{d})$. For $a, b \in [n]$, $\mathcal{I} \subseteq [n]$ and $k \geq 1$, we let $\mathcal{P}_k(a, b) = \mathcal{P}_k(a, b; \mathcal{I})$ denote the set of k-paths that only use vertices in $\mathcal{I}$, and we let

$$N_k(a, b) = N_k(a, b; \mathcal{I}) = \#\{\pi \in \mathcal{P}_k(a, b) \colon \pi \subseteq \mathrm{CM}_n(\boldsymbol{d})\} \tag{5.2.30}$$

denote the number of paths of length k between the vertices a and b. Then, we prove the following upper bound on the expected number of paths connecting a and b:

Proposition 5.9 (Expected numbers of paths)**.** *For any $k \geq 1$, $a, b \in [n]$ and $(d_i)_{i \in [n]}$,*

$$\mathbb{E}[N_k(a, b)] \leq \frac{d_a d_b \ell_n}{(\ell_n - 2k + 1)(\ell_n - 2k)} \nu_{\mathcal{I}}^{k-1}, \tag{5.2.31}$$

where

$$\nu_{\mathcal{I}} = \sum_{i \in \mathcal{I} \setminus \{a,b\}} \frac{d_i(d_i - 1)}{\ell_n}. \tag{5.2.32}$$

Proof. The probability that the path π in (5.2.29) is present in $\mathrm{CM}_n(\boldsymbol{d})$ is equal to

$$\prod_{i=1}^{k} \frac{1}{\ell_n - 2i + 1}, \tag{5.2.33}$$

and the number of paths with fixed vertices $\pi_0, \dots, \pi_k$ is equal to

$$d_{\pi_0} \Big(\prod_{i=1}^{k-1} d_{\pi_i}(d_{\pi_i} - 1) \Big) d_{\pi_k}. \tag{5.2.34}$$

Substituting $\pi_0 = a, \pi_k = b$, we arrive at

$$\mathbb{E}[N_k(a, b)] = \frac{d_a d_b}{\ell_n - 2k + 1} \sum_{\pi_1, \dots, \pi_{k-1}}^{*} \prod_{i=1}^{k-1} \frac{d_{\pi_i}(d_{\pi_i} - 1)}{\ell_n - 2i + 1}, \tag{5.2.35}$$

where the sum is over distinct elements of $\mathcal{I} \setminus \{a, b\}$. Let R denote the subset of vertices of $\mathcal{I} \setminus \{a, b\}$ for which $d_i \geq 2$. Then,

$$\mathbb{E}[N_k(a,b)] = \frac{d_a d_b}{\ell_n - 2k + 1} \sum_{\pi_1,\ldots,\pi_{k-1} \in R}^{*} \prod_{i=1}^{k-1} \frac{d_{\pi_i}(d_{\pi_i} - 1)}{\ell_n - 2i + 1}, \tag{5.2.36}$$

By an inequality of Maclaurin [61, Theorem 52], for $r = |R|$, $2 \leq k \leq r + 1$ and any $(a_i)_{i \in R}$ with $a_i \geq 0$,

$$\frac{(r-k+1)!}{r!} \sum_{\pi_1,\ldots,\pi_{k-1} \in R}^{*} \prod_{i=1}^{k-1} a_i \leq \Big(\frac{1}{r} \sum_{i \in R} a_i\Big)^{k-1}. \tag{5.2.37}$$

Let $a_i = d_i(d_i - 1)$, so that

$$\sum_{i \in R} a_i = \ell_n \nu_{\mathcal{I}}. \tag{5.2.38}$$

We arrive at

$$\begin{aligned} \mathbb{E}[N_k(a,b)] &= \frac{d_a d_b}{\ell_n - 2k + 1} (\ell_n \nu_{\mathcal{I}}/r)^{k-1} \prod_{i=1}^{k-1} \frac{(r - i + 1)}{(\ell_n - 2i + 1)} \\ &\leq \frac{d_a d_b}{\ell_n - 2k + 1} \frac{\ell_n}{\ell_n - 2k} \nu_{\mathcal{I}}^{k-1} \prod_{i=0}^{k-2} \frac{(1 - \frac{i}{r})}{(1 - \frac{2i}{\ell_n})}. \end{aligned} \tag{5.2.39}$$

Further, $\ell_n = \sum_{i \in [n]} d_i \geq 2r$, so that $1 - \frac{i}{r} \leq 1 - \frac{2i}{\ell_n}$. Substitution yields the required bound. □

Logarithmic lower bound typical distances $\mathrm{CM}_n(\boldsymbol{d})$. With Proposition 5.9 at hand, we can immediately prove the lower bound on the typical graph distance in the case where the degrees have finite second moment (as in Theorem 3.5):

Theorem 5.10 (Logarithmic lower bound typical distances $\mathrm{CM}_n(\boldsymbol{d})$)**.** *Assume that*

$$\limsup_{n \to \infty} \nu_n > 1, \tag{5.2.40}$$

where

$$\nu_n = \mathbb{E}[D_n(D_n - 1)]/\mathbb{E}[D_n]. \tag{5.2.41}$$

Then, for any $\varepsilon > 0$,

$$\mathbb{P}(H_n \leq (1 - \varepsilon) \log_{\nu_n} n) = o(1). \tag{5.2.42}$$

We leave the proof of Theorem 5.10, which similar to that of Theorem 3.5, as an exercise:

Exercise 5.6 (Proof Theorem 5.10)**.** *Let V_1, V_2 be two independent vertices chosen uniformly at random from $[n]$. Use Proposition 5.9 with $a = V_1, b = V_2, \mathcal{I} = [n]$ to prove Theorem 5.10.*

Truncated first moment method and log log **lower bound for** $\tau \in (2,3)$**.** We next extend the above upper bounds on the expected number of paths to deal with the case where $\tau \in (2,3)$, where similarly to the setting in Section 3.3.2 where $\mathrm{NR}_n(\boldsymbol{w})$ was investigated, we need to *truncate* the degrees occurring in the arising paths. Our main result is as follows:

Theorem 5.11 (Loglog lower bound on typical distances in $\mathrm{CM}_n(\boldsymbol{d})$)**.** *Suppose that the weights* $\boldsymbol{d} = (d_i)_{i\in[n]}$ *satisfy [I, Condition 7.7(a)] and that there exists a* $\tau \in (2,3)$ *and* c_2 *such that, for all* $x \geq 1$*,*

$$[1 - F_n](x) \leq c_2 x^{-(\tau-1)}, \tag{5.2.43}$$

Then, for every $\varepsilon > 0$*,*

$$\mathbb{P}\Big(H_n \leq (1-\varepsilon)\frac{2\log\log n}{|\log(\tau-2)|}\Big) = o(1). \tag{5.2.44}$$

The proof of Theorem 5.11 is identical to that of Theorem 3.7, and we discuss the changes only. For a fixed set of distinct vertices $(\pi_0,\ldots,\pi_k)$, (5.2.33)-(5.2.34) yield that the probability that there exists edges between π_{i-1} and π_i for all $i = 1,\ldots,k$ in $\mathrm{CM}_n(\boldsymbol{d})$ is bounded above by

$$\frac{d_{\pi_0}d_{\pi_k}}{\ell_n - 2k + 1}\Big(\prod_{i=1}^{k-1}\frac{d_{\pi_i}(d_{\pi_i}-1)}{\ell_n - 2i + 1}\Big). \tag{5.2.45}$$

Equation (5.2.45) replaces the similar identity (3.3.7) for $\mathrm{NR}_n(\boldsymbol{w})$. We see that w_{π_0} and w_{π_k} in (3.3.7) are replaced with d_{π_0} and d_{π_k} in (5.2.45), and, for $i = 1,\ldots,k-1$, the factors $w^2_{\pi_i}$ in (3.3.7) are replaced with $d_{\pi_i}(d_{\pi_i}-1)$ in (5.2.45), while the factors ℓ_n in (3.3.7) is replaced with $(\ell_n - 2i + 1)$ in (5.2.45).

Define, as in (3.3.34),

$$\nu_n(b) = \frac{1}{\ell_n}\sum_{i\in[n]} d_i(d_i-1)\mathbb{1}_{\{d_i\leq b\}}. \tag{5.2.46}$$

Then, we can adapt the arguments in Section 3.3.2 to obtain that (see in particular Exercise 3.20),

$$\begin{aligned}\mathbb{P}(\mathrm{dist}_{\mathrm{CM}_n(\boldsymbol{d})}(a,b) \leq k_n) \leq{}& \frac{d_a d_b}{\ell_n}\sum_{k=1}^{k_n}\frac{\ell_n^k(\ell_n-2k-1)!!}{(\ell_n-1)!!}\prod_{l=1}^{k-1}\nu_n(b_l\wedge b_{k-l}) \\ &+ (d_a+d_b)\sum_{k=1}^{k^*}\frac{\ell_n^k(\ell_n-2k-1)!!}{(\ell_n-1)!!}[1-F_n^*](b_k)\prod_{l=1}^{k}\nu_n(b_l),\end{aligned} \tag{5.2.47}$$

i.e., the bound in (3.3.41) is changed by factors $\frac{\ell_n^k(\ell_n-2k-1)!!}{(\ell_n-1)!!}$ in the sum. For $k = O(\log\log n)$ and when [I, Condition 7.7(a)-(b)] holds,

$$\frac{\ell_n^k(\ell_n-2k-1)!!}{(\ell_n-1)!!} = \prod_{i=1}^{k}\frac{\ell_n}{\ell_n-2i+i} = 1 + O(k^2/\ell_n) = 1 + o(1), \tag{5.2.48}$$

so this change has only minor effect. Since Lemma 3.9 applies under the conditions of Theorem 5.11, we can follow the proof of Theorem 3.7 verbatim. This completes the proof of Theorem 5.11. □

Second moment method for the number of paths in $\mathrm{CM}_n(\boldsymbol{d})$. We next extend the above first moment bounds on the number of paths in $\mathrm{CM}_n(\boldsymbol{d})$ to second moment methods. Define

$$\bar{n}_k(a,b) = \frac{\ell_n^k(\ell_n-2k-1)!!}{(\ell_n-1)!!}\frac{d_a d_b}{\ell_n}\Big(\sum_{i\in\mathcal{I}\setminus\{a,b\}}\frac{d_i(d_i-1)}{\ell_n}\Big)^{k-1}, \tag{5.2.49}$$

$$\underline{n}_k(a,b) = \frac{d_a d_b}{\ell_n}\Big(\sum_{i\in\mathcal{I}_{a,b,k}}\frac{d_i(d_i-1)}{\ell_n}\Big)^{k-1}, \tag{5.2.50}$$

where $\mathcal{I}_{a,b,k}$ is the subset of $\mathcal{I}$ in which a and b, as well as the $k-1$ indices with highest degrees have been removed. Let

$$\nu_{\mathcal{I}} = \frac{1}{\ell_n}\sum_{i\in\mathcal{I}} d_i(d_i-1), \qquad \gamma_{\mathcal{I}} = \frac{1}{\ell_n^{3/2}}\sum_{i\in\mathcal{I}} d_i(d_i-1)(d_i-2). \tag{5.2.51}$$

Proposition 5.12 (Variance of number of paths)**.** *For any $k \geq 1$, $a,b \in \mathcal{I}$ and $(u_i)_{i\in\mathcal{I}}$,*

$$\mathbb{E}[N_k(a,b)] \geq \underline{n}_k(a,b), \tag{5.2.52}$$

while, assuming that $\nu_{\mathcal{I}} > 1$,

$$\mathrm{Var}(N_k(a,b)) \leq n_k(a,b) + \bar{n}_k(a,b)^2\Big(\frac{\gamma_{\mathcal{I}}\nu_{\mathcal{I}}^2}{\nu_{\mathcal{I}}-1}\Big(\frac{1}{d_a}+\frac{1}{d_b}\Big) + \frac{\gamma_{\mathcal{I}}^2\nu_{\mathcal{I}}}{d_a d_b(\nu_{\mathcal{I}}-1)^2} + e_k'\Big), \tag{5.2.53}$$

where

$$\begin{aligned} e_k' = &\Big(\prod_{i=1}^{k}\frac{\ell_n-2i+1}{\ell_n-2i-2k+1}-1\Big) \\ &+ k\frac{\ell_n^{2k}(\ell_n-4k-1)!!}{(\ell_n-1)!!}\Big(1+\frac{\gamma_{\mathcal{I}}}{d_a\nu_{\mathcal{I}}}\Big)\Big(1+\frac{\gamma_{\mathcal{I}}}{d_b\nu_{\mathcal{I}}}\Big)\frac{\nu_{\mathcal{I}}}{\nu_{\mathcal{I}}-1}\Big(\mathrm{e}^{2k^3\gamma_{\mathcal{I}}^2/\nu_{\mathcal{I}}^3}-1\Big). \end{aligned} \tag{5.2.54}$$

Proof. The proof of (5.2.52) follows immediately from (5.2.35), together with the fact that $1/(\ell_n-2i+1) \geq 1/\ell_n$.

For the proof of (5.2.53), we follow the proof of (3.5.7), and discuss the differences only. We recall that

$$N_k(a,b) = \sum_{\pi\in\mathcal{P}_k(a,b)} \mathbb{1}_{\{\pi\subseteq\mathrm{CM}_n(\boldsymbol{d})\}} \tag{5.2.55}$$

is the number of paths π of length k between the vertices a and b, where a path is defined in (5.2.29). Since $N_k(a,b)$ is a sum of indicators, its variance can be written as

$$\mathrm{Var}(N_k(a,b)) = \sum_{\pi,\rho\in\mathcal{P}_k(a,b)}\big[\mathbb{P}(\pi,\rho\subseteq\mathrm{CM}_n(\boldsymbol{d})) - \mathbb{P}(\pi\subseteq\mathrm{CM}_n(\boldsymbol{d}))\mathbb{P}(\rho\subseteq\mathrm{CM}_n(\boldsymbol{d}))\big]. \tag{5.2.56}$$

Equation (5.2.56) replaces (3.5.12) for $\mathrm{NR}_n(\boldsymbol{w})$. We say that two paths π and ρ are *disjoint* when they use distinct sets of half-edges. Thus, it is possible that the vertex sets $\{\pi_1, \ldots, \pi_{k-1}\}$ and $\{\rho_1, \ldots, \rho_{k-1}\}$ have a non-empty intersection, but then the half-edges leading in and out of the joint vertices for π and ρ must be distinct. For $\mathrm{NR}_n(\boldsymbol{w})$, pairs of paths using different edges are *independent,* so that these pairs do not contribute to $\mathrm{Var}(N_k(a,b))$. For $\mathrm{CM}_n(\boldsymbol{d})$, instead,

$$\mathbb{P}(\pi, \rho \subseteq \mathrm{CM}_n(\boldsymbol{d})) = \prod_{i=1}^{k} \frac{\ell_n - 2i + 1}{\ell_n - 2i - 2k + 1} \mathbb{P}(\pi \subseteq \mathrm{CM}_n(\boldsymbol{d})) \mathbb{P}(\rho \subseteq \mathrm{CM}_n(\boldsymbol{d})), \quad (5.2.57)$$

which explains the first contribution to e'_k. For the other contributions, we follow the proof of (3.5.12) for $\mathrm{NR}_n(\boldsymbol{w})$, and omit further details. □

With Proposition 5.12 at hand, we can adapt the proof of Theorem 3.19 to $\mathrm{CM}_n(\boldsymbol{d})$:

Theorem 5.13 (Logarithmic upper bound graph distances $\mathrm{CM}_n(\boldsymbol{d})$)**.** *Assume that [I, Condition 7.7(a)-(c)] hold, where* $\nu = \mathbb{E}[D(D-1)]/\mathbb{E}[D] \in (1, \infty)$*. Then, for any* $\varepsilon > 0$*,*

$$\mathbb{P}(H_n \leq (1+\varepsilon) \log_\nu n \mid H_n < \infty) = 1 + o(1). \quad (5.2.58)$$

We leave the proof of Theorem 5.13 as an exercise:

Exercise 5.7 (Proof Theorem 5.13)**.** *Use Proposition 5.12 to prove Theorem 5.13 by adapting the proof of Theorem 3.19.*

5.2.3 A $\log\log$ upper bound on the diameter core $\tau \in (2,3)$

In order to prove the upper bound on the typical distance for $\mathrm{CM}_n(\boldsymbol{d})$ in Theorem 5.3, we use a different approach compared to the one in the proof of Theorem 3.14. Our proof for the upper bound on the typical distance for $\mathrm{CM}_n(\boldsymbol{d})$ in Theorem 5.3 is organized as follows. We first prove an upper bound on the *core* of $\mathrm{CM}_n(\boldsymbol{d})$, which consists of all vertices of degree at least $(\log n)^\sigma$ for some $\sigma > 0$. This is the content of Theorem 5.14 below. Followed by the proof of Theorem 5.14, we use a second moment method to prove that any vertex that survives to sufficient large graphs distance is whp quickly connected to the core. The bound on the diameter of the core is also useful in studying the diameter of $\mathrm{CM}_n(\boldsymbol{d})$ when $\tau \in (2,3)$ and $d_{\min} \geq 3$, which we perform in Section 5.3 below.

We take $\sigma > 1/(3-\tau)$ and define the *core* Core_n of the configuration model to be

$$\mathrm{Core}_n = \{i \colon d_i \geq (\log n)^\sigma\}, \quad (5.2.59)$$

i.e., the set of vertices with degree at least $(\log n)^\sigma$. Then, the diameter of the core is bounded in the following theorem:

Theorem 5.14 (Diameter of the core)**.** *Fix* $\tau \in (2,3)$ *and assume that (5.1.7) holds. For any* $\sigma > \frac{1}{3-\tau}$*, the diameter of* Core_n *is with high probability bounded above by*

$$\frac{2 \log\log n}{|\log(\tau - 2)|} + 1. \quad (5.2.60)$$

Proof. We note that (5.1.7) implies that, for some $\alpha \in (1/2, 1/(\tau-1))$,

$$\max_{i\in[n]} d_i \geq u_1, \quad \text{where} \quad u_1 = n^\alpha. \tag{5.2.61}$$

Define

$$\Gamma_1 = \{i\colon d_i \geq u_1\}, \tag{5.2.62}$$

so that $\Gamma_1 \neq \varnothing$. For some constant $C > 0$ to be determined later on, and for $k \geq 2$, we recursively define

$$u_k = C \log n \big(u_{k-1}\big)^{\tau-2}. \tag{5.2.63}$$

Then, we define

$$\Gamma_k = \{i\colon d_i \geq u_k\}. \tag{5.2.64}$$

We identify u_k in the following lemma:

Lemma 5.15 (Identification $(u_k)_{k\geq 1}$)**.** *For every $k \geq 1$,*

$$u_k = C^{a_k} (\log n)^{b_k} n^{c_k}, \tag{5.2.65}$$

where

$$c_k = \alpha(\tau-2)^{k-1}, \quad a_k = b_k = \frac{1}{3-\tau}[1-(\tau-2)^{k-1}]. \tag{5.2.66}$$

Proof. We note that c_k, b_k, a_k satisfy the recursions, for $k \geq 2$,

$$c_k = (\tau-2)c_{k-1}, \quad b_k = 1+(\tau-2)b_{k-1}, \quad a_k = 1+(\tau-2)a_{k-1}, \tag{5.2.67}$$

with initial conditions $c_1 = \alpha, a_1 = b_1 = 0$. Solving the recursions yields our claim. □

In order to study connectivity of sets in $\mathrm{CM}_n(\boldsymbol{d})$, we rely on the following lemma, which is of independent interest:

Lemma 5.16 (Connectivity sets in $\mathrm{CM}_n(\boldsymbol{d})$)**.** *For any two sets of vertices $A, B \subseteq [n]$,*

$$\mathbb{P}(A \text{ not directly connected to } B) \leq \mathrm{e}^{-d_A d_B/(2\ell_n)}, \tag{5.2.68}$$

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

$$d_A = \sum_{i\in A} d_i \tag{5.2.69}$$

denotes the total degree of vertices in A.

Proof. There are d_A half-edges incident to the set A, which we pair one by one. After having paired k half-edges, all to half-edges that are not incident to B, the probability to pair the next half-edge to a half-edge that is not incident to B equals

$$1 - \frac{d_B}{\ell_n - 2k + 1} \leq 1 - \frac{d_B}{\ell_n}. \tag{5.2.70}$$

Some half-edges incident to A may attach to other half-edges incident to A, so that possibly fewer than d_A half-edges need to be paired to pair all half-edges incident to

A. However, since each pairing uses up at most 2 half-edges incident to A, we need to pair at least $d_A/2$ half-edges, so that

$$\mathbb{P}(A \text{ not directly connected to } B) \leq \Big(1 - \frac{d_B}{\ell_n}\Big)^{d_A/2} \leq \mathrm{e}^{-d_A d_B/(2\ell_n)}, \tag{5.2.71}$$

where we use that $1 - x \leq \mathrm{e}^{-x}$. □

Exercise 5.8 (Γ_1 is a complete graph)**.** *Use Lemma 5.16 and $\alpha > 1/2$ to show that, whp, Γ_1 in (5.2.62) forms a* complete graph*, i.e., whp, every $i, j \in \Gamma_1$ are direct neighbors in* $\mathrm{CM}_n(\boldsymbol{d})$.

The key step in the proof of Theorem 5.14 is the following proposition showing that whp every vertex in Γ_k is connected to a vertex in Γ_{k-1}:

Proposition 5.17 (Connectivity between Γ_{k-1} and Γ_k)**.** *Fix $\tau \in (2, 3)$ and assume that [I, Condition 7.7(a)-(b)] and (5.1.7) hold. Fix $k \geq 2$, and take $C > 2\mathbb{E}[D]/c$. Then, the probability that there exists an $i \in \Gamma_k$ that is not directly connected to Γ_{k-1} is $o(n^{-\delta})$, for some $\delta > 0$ independent of k.*

Proof. We note that, by definition,

$$\sum_{i \in \Gamma_{k-1}} d_i \geq u_{k-1}|\Gamma_{k-1}| = u_{k-1} n [1 - F_n](u_{k-1}). \tag{5.2.72}$$

By (5.1.7), and since $k \mapsto u_k$ is decreasing with $u_1 = n^\alpha$,

$$[1 - F_n](u_{k-1}) \geq c(u_{k-1})^{1-\tau}. \tag{5.2.73}$$

As a result, we obtain that for every $k \geq 2$,

$$\sum_{i \in \Gamma_{k-1}} d_i \geq cn(u_{k-1})^{2-\tau}. \tag{5.2.74}$$

By (5.2.74) and Lemma 5.16, using Boole's inequality, the probability that there exists an $i \in \Gamma_k$ that is not directly connected to Γ_{k-1} is bounded by

$$n\mathrm{e}^{-\frac{u_k n u_{k-1}[1-F(u_{k-1})]}{2\ell_n}} \leq n\mathrm{e}^{-\frac{cu_k(u_{k-1})^{2-\tau}}{2\mathbb{E}[D_n]}} = n^{1-\frac{cC}{2\mathbb{E}[D_n]}}. \tag{5.2.75}$$

By [I, Condition 7.7(a)-(b)], $\mathbb{E}[D_n] \to \mathbb{E}[D]$, so that, as $n \to \infty$ and taking $C > 2\mathbb{E}[D]/c$, we obtain the claim for any $\delta < \frac{cC}{2\mathbb{E}[D]} - 1$. □

We now complete the proof of Theorem 5.14.

Proof of Theorem 5.14. Fix

$$k^* = \frac{\log\log n}{|\log(\tau - 2)|}. \tag{5.2.76}$$

As a result of Proposition 5.17, whp, the diameter of Γ_{k^*} is at most $2k^* + 1$, because the distance between any vertex in Γ_{k^*} and Γ_1 is at most k^*, while, by Exercise 5.8, Γ_1 forms a complete graph. Therefore, it suffices to prove that

$$\mathrm{Core}_n \subseteq \Gamma_{k^*}. \tag{5.2.77}$$

By (5.2.63), in turn, this is equivalent to $u_{k^*} \geq (\log n)^\sigma$, for any $\sigma > 1/(3-\tau)$. According to Lemma 5.15,

$$u_{k^*} = C^{a_{k^*}} (\log n)^{b_{k^*}} n^{c_{k^*}}. \tag{5.2.78}$$

We note that $n^{c_{k^*}} = \mathrm{e}^{\log n (\tau-2)^{k^*}}$. Since, for $2 < \tau < 3$,

$$x(\tau-2)^{\frac{\log x}{|\log(\tau-2)|}} = x \cdot x^{-1} = 1, \tag{5.2.79}$$

we find with $x = \log n$ that $n^{c_{k^*}} = \mathrm{e}$. Further, $b_k \to 1/(\tau-3)$ as $k \to \infty$, so that $(\log n)^{b_{k^*}} = (\log n)^{1/(3-\tau)+o(1)}$, and $a_k = b_k$, so that also $C^{a_{k^*}} = C^{1/(\tau-3)+o(1)}$. We conclude that

$$u_{k^*} = (\log n)^{1/(3-\tau)+o(1)}, \tag{5.2.80}$$

so that, by picking n sufficiently large, we can make $1/(3-\tau)+o(1) \leq \sigma$. This completes the proof of Theorem 5.14. □

We continue to use Theorem 5.14 to prove a $\log\log n$ upper bound on H_n in the case where $\tau \in (2,3)$. We start by describing the setting. We assume that there exist $\tau \in (2,3)$, $\alpha > 1/2$ and c_1 such that, uniformly in n and $x \leq n^\alpha$,

$$[1-F_n](x) \geq c_1 x^{-(\tau-1)}. \tag{5.2.81}$$

Theorem 5.18 (A $\log\log$ upper bound on typical distance for $\tau \in (2,3)$)**.** *Suppose that the empirical distribution function F_n of the degrees $\boldsymbol{d} = (d_i)_{i\in[n]}$ satisfies [I, Condition 7.7(a)-(b)] and (5.2.81). Then, for every $\varepsilon > 0$,*

$$\lim_{n\to\infty} \mathbb{P}\Big(H_n \leq \frac{2(1+\varepsilon)\log\log n}{|\log(\tau-2)|} \mid H_n < \infty\Big) = 1. \tag{5.2.82}$$

Proof. We make crucial use of the branching process approximation in Section 5.2.1. We let V_1, V_2 denote two vertices chosen uniformly at random from $[n]$, and we recall that $Z_m^{(i)}$ denote the number of unpaired or free half edges incident to vertices in $\mathcal{N}_m(V_i)$. By Proposition 5.4, $(Z_l^{(1)}, Z_l^{(2)})_{l=0}^{m_n}$ can be whp perfectly coupled to $(\mathcal{Z}_l^{(1)}, \mathcal{Z}_l^{(2)})_{l=0}^{m_n}$, which are two independent two-stage branching processes where the root has offspring distribution D, and individuals in all further generations have offspring distribution $D^* - 1$.

We condition on $\mathcal{N}_{\leq m_n}(V_1), \mathcal{N}_{\leq m_n}(V_2)$ which are such that $Z_{m_n}^{(1)} \geq 1, Z_{m_n}^{(2)} \geq 1$. Further, ny Corollary 3.13, conditionally on $Z_{m_n}^{(1)} \geq 1, Z_{m_n}^{(2)} \geq 1$, $Z_{m_n}^{(1)} \xrightarrow{\mathbb{P}} \infty, Z_{m_n}^{(2)} \xrightarrow{\mathbb{P}} \infty$ occurs.

We will condition on $\mathcal{N}_{\leq m_n}(V_1), \mathcal{N}_{\leq m_n}(V_2)$, and denote the conditional distribution by $\widetilde{\mathbb{P}}_{m_n}$, and the expectation and variance under the measure $\widetilde{\mathbb{P}}_{m_n}$ by $\widetilde{\mathbb{E}}_{m_n}$ and $\widetilde{\mathrm{Var}}_{m_n}$, respectively. We collapse $\mathcal{N}_{m_n}(V_1)$ to a single vertex a_1 and $\mathcal{N}_{m_n}(V_2)$ to a single vertex a_2. The distribution of the resulting random graph is again a configuration model, with degrees $d_{a_1} = Z_{m_n}^{(1)}, d_{a_2} = Z_{m_n}^{(2)}$ and vertex set $R = [n] \cup \{a_1, a_2\} \setminus (\mathcal{N}_{\leq m_n}(V_1) \cup \mathcal{N}_{\leq m_n}(V_2))$.

We apply Proposition 5.12 with $k = \varepsilon \log\log n$, a_1, $b = \mathrm{Core}_n$ and with $\mathcal{I} = \{i \in R\colon d_i \le K\}$. Then, Proposition 5.12 gives that, conditionally on $\mathcal{N}_{\le m_n}(V_1), \mathcal{N}_{\le m}(V_2)$ such that $Z^{(1)}_{m_n} \ge 1, Z^{(2)}_{m_n} \ge 1$,

$$\widetilde{\mathbb{P}}_{m_n}(N_k(a_i,b)=0) \le \widetilde{\mathrm{Var}}_{m_n}(N_k(a_i,b))/\widetilde{\mathbb{E}}_{m_n}[N_k(a_i,b)]^2 \le O(K)\big(1/Z^{(1)}_{m_n}+1/Z^{(2)}_{m_n}\big) \xrightarrow{\mathbb{P}} 0, \tag{5.2.83}$$

where convergence in probability follows from (5.2.27) in Corollary 5.8. As a result, conditionally on $\mathcal{N}_{\le m_n}(V_1), \mathcal{N}_{\le m_n}(V_2)$ such that $Z^{(1)}_{m_n} \ge 1, Z^{(2)}_{m_n} \ge 1$, with probability at least $1-o(1)$, $N_k(a_i,b) \ge 1$, so that, on this event,

$$H_n \le \mathrm{diam}_{\mathrm{CM}_n(\boldsymbol{d})}(\mathrm{Core}_n) + 2k \le \frac{2(1+\varepsilon)\log\log n}{|\log(\tau-2)|}. \tag{5.2.84}$$

We further use that, by Theorem 4.1,

$$\mathbb{P}(H_n < \infty) \to \zeta^2, \tag{5.2.85}$$

while, (5.2.26) in Corollary 5.8,

$$\mathbb{P}(Z^{(1)}_{m_n} \ge 1, Z^{(2)}_{m_n} \ge 1) \to \zeta^2. \tag{5.2.86}$$

As a result,

$$\begin{aligned}
&\mathbb{P}\Big(H_n \le \frac{2(1+\varepsilon)\log\log n}{|\log(\tau-2)|} \mid H_n < \infty\Big) \qquad (5.2.87)\\
&\quad \ge \frac{\mathbb{P}(H_n \le \frac{2(1+\varepsilon)\log\log n}{|\log(\tau-2)|}, Z^{(1)}_{m_n} \ge 1, Z^{(2)}_{m_n} \ge 1)}{\mathbb{P}(H_n < \infty)}\\
&\quad = \frac{\mathbb{P}(Z^{(1)}_{m_n} \ge 1, Z^{(2)}_{m_n} \ge 1) - o(1)}{\mathbb{P}(H_n < \infty)} = 1 - o(1),
\end{aligned}$$

as required. This completes the proof of Theorem 5.18. □

Exercise 5.9 (Alternative proof Theorem 5.18)**.** *Give an alternative proof of Theorem 5.18 by adapting the proof of Theorem 3.14.*

5.3 Diameter of $\mathrm{CM}_n(\boldsymbol{d})$ for $\tau \in (2,3)$

In this section, we use Theorem 5.14 to study the diameter of $\mathrm{CM}_n(\boldsymbol{d})$ when $\tau \in (2,3)$. Note that the diameter is equal to a positive constant times $\log n$ by Theorem 5.2 when $p_1 + p_2 > 0$. Therefore, we turn to the case where $p_1 = p_2 = 0$. When $d_{\min} \ge 3$, we know by Theorem 4.14 that $\mathrm{CM}_n(\boldsymbol{d})$ is whp connected. The main result is as follows:

Theorem 5.19 (Diameter of $\mathrm{CM}_n(\boldsymbol{d})$ for $\tau \in (2,3)$)**.** *Suppose that the empirical distribution function F_n of the degrees $\boldsymbol{d} = (d_i)_{i\in[n]}$ satisfies [I, Condition 7.7(a)-(b)] and that (5.1.7) holds. Assume further that $d_{\min} = \min_{i\in[n]} d_i \ge 3$ and $p_{d_{\min}} = \mathbb{P}(D = d_{\min}) > 0$. Then,*

$$\frac{\mathrm{diam}(\mathrm{CM}_n(\boldsymbol{d}))}{\log\log n} \xrightarrow{\mathbb{P}} \frac{2}{|\log(\tau-2)|} + \frac{2}{\log(d_{\min}-1)}. \tag{5.3.1}$$

When comparing Theorem 5.19 to Theorem 5.3, we see that for $d_{\min} \geq 3$, the diameter is of the same order $\log\log n$ as the typical distance, but that the constant differs. The diameter is due to pairs of vertices that have small local neighborhoods. Indeed, by assumption, there is a positive proportion of vertices of degree $d_{\min}$. As a result, we will see that the expected number of vertices whose $(1-\varepsilon)\log\log n/\log(d_{\min}-1)$ neighborhood only contains vertices with degree $d_{\min}$ tends to infinity. The minimal path between two such vertices then consists of three parts: the two paths from the two vertices to leave their minimally connected neighborhood, and then the path between the boundaries of these minimally connected neighborhoods. These minimal neighborhoods are at the typical distance $2\log\log n/|\log(\tau-2)|$, as in Theorem 5.3. This explains Theorem 5.19.

We prove Theorem 5.19 by proving an upper and a lower bound on the diameter. We start with the lower bound, which is the easier part:

Lower bound on the diameter. We call a vertex v *minimally-k-connected* when all $i \in \mathcal{N}_{\leq k}(v)$ satisfy $d_i = d_{\min}$, so that all vertices at distance at most k have the minimal degree. Let M_k denote the number of minimally-k-connected vertices. To prove the lower bound in on the diameter, we show that $M_k \xrightarrow{\mathbb{P}} \infty$ for $k = (1-\varepsilon)\log\log n/\log(d_{\min}-1)$. Followed by this, we show that two minimally k-connected vertices v_1, v_2 are such that whp the distance between $\mathcal{N}_k(v_1)$ and $\mathcal{N}_k(v_2)$ is at least $2\log\log n/|\log(\tau-2)|$. For this,w e start by computing the first and second moment of M_k in the following lemma:

Lemma 5.20 (Moments of number of minimally-k-connected vertices)**.** *Let* $\mathrm{CM}_n(\boldsymbol{d})$ *satisfy that* $d_{\min} \geq 3$, $n_{d_{\min}} \geq d_{\min}(d_{\min}-1)^{k-1}$ *and there exists at least one* v *with* $d_v > d_{\min}$. *Then, for all* $k \geq 1$,

$$\mathbb{E}[M_k] = n_{d_{\min}} \prod_{i=1}^{d_{\min}(d_{\min}-1)^{k-1}} \frac{d_{\min}(n_{d_{\min}} - (i-1))}{\ell_n - 2i + 1}, \tag{5.3.2}$$

and, for k *such that* $d_{\min}(d_{\min}-1)^{k-1} \leq \ell_n/8$,

$$\mathbb{E}[M_k^2] \leq \mathbb{E}[M_k]^2 + \mathbb{E}[M_k]\Big[\frac{d_{\min}}{d_{\min}-2}(d_{\min}-1)^k + \frac{2n_{d_{\min}}d_{\min}^2(d_{\min}-1)^{2k}}{(d_{\min}-2)\ell_n}\Big]. \tag{5.3.3}$$

Proof. We start by proving (5.3.2). We note that each vertex of degree $d_{\min}$ has the same probability of being minimally-k connected, and that there are precisely $n_{d_{\min}}$ vertices of degree $d_{\min}$, so that

$$\mathbb{E}[M_k] = n_{d_{\min}}\mathbb{P}(v \text{ with } d_v = d_{\min} \text{ is minimally-}k\text{-connected}). \tag{5.3.4}$$

Vertex v with $d_v = d_{\min}$ is minimally-k-connected when all its half-edges at distance at most k are paired to half-edges incident to a distinct vertex having minimal degree $d_{\min}$, and no cycles occur in $\mathcal{N}_{\leq k}(v)$. When $i-1$ half-edges are paired to distinct vertices of degree $d_{\min}$, then the probability that the ith half-edge is again paired to a distinct vertex of degree $d_{\min}$ equals

$$\frac{d_{\min}(n_{d_{\min}} - (i-1))}{\ell_n - 2i + 1}. \tag{5.3.5}$$

Since for v to be minimally-k-connected, there are $d_{\min}(d_{\min}-1)^{k-1}$ half-edges that need to be paired to distinct vertices of degree $d_{\min}$, this proves (5.3.2).

To prove (5.3.3), we note that

$$\mathbb{E}[M_k^2] = \sum_{v_1,v_2\in[n]} \mathbb{P}(v_1, v_2 \text{ with } d_{v_1}, d_{v_2} = d_{\min} \text{ are minimally-}k\text{-connected}). \quad (5.3.6)$$

We split the above probability depending on whether $\mathcal{N}_{\le k}(v_1)\cap\mathcal{N}_{\le k}(v_2)=\varnothing$ or not. The contribution to $\mathbb{E}[M_k^2]$ due to $\mathcal{N}_{\le k}(v_1)\cap\mathcal{N}_{\le k}(v_2)=\varnothing$ is, similarly to the proof of (5.3.2), equal to

$$n_{d_{\min}}(n_{d_{\min}}-i_{k-1})\prod_{i=1}^{i_k^2}\frac{d_{\min}(n_{d_{\min}}-(i-1))}{\ell_n-2i+1}, \quad (5.3.7)$$

where we abbreviate $i_k = d_{\min}(d_{\min}-1)^{k-1}$ and note that $n_{d_{\min}}-i_{k-1}>0$, since, by assumption, $n_{d_{\min}} > d_{\min}(d_{\min}-1)^{k-1}$.

We use that $i\mapsto d_{\min}(n_{d_{\min}}-(i-1))/(\ell_n-2i+1)$ is decreasing since

$$\frac{(n_{d_{\min}}-(i-1))}{\ell_n-2i+1} \ge \frac{n_{d_{\min}}-i}{\ell_n-2i-1} \quad (5.3.8)$$

precisely when $\ell_n \ge 2n_{d_{\min}}+1$, and which is true since by assumption there exists at least one v with $d_v > d_{\min}$. Therefore, the contribution to $\mathbb{E}[M_k^2]$ from v_1 and v_2 satisfying $\mathcal{N}_{\le k}(v_1)\cap\mathcal{N}_{\le k}(v_2)=\varnothing$ is at most

$$n_{d_{\min}}^2\Big(\prod_{i=1}^{i_k}\frac{d_{\min}(n_{d_{\min}}-(i-1))}{\ell_n-2i+1}\Big)^2 = \mathbb{E}[M_k^2], \quad (5.3.9)$$

which is the first contribution to the r.h.s. of (5.3.3).

We are left to deal with the contribution to $\mathbb{E}[M_k^2]$ from v_1 and v_2 such that $\mathcal{N}_{\le k}(v_1)\cap\mathcal{N}_{\le k}(v_2)\neq\varnothing$. When v_1 is k-minimally connected,

$$\begin{aligned}|\mathcal{N}_{\le k}(v_1)| &= 1+\sum_{l=1}^{k} d_{\min}(d_{\min}-1)^{l-1} \quad (5.3.10)\\ &= 1+d_{\min}\frac{(d_{\min}-1)^k-1}{d_{\min}-2} \le d_{\min}\frac{(d_{\min}-1)^k}{d_{\min}-2}.\end{aligned}$$

Therefore, the contribution due to $v_2\in\mathcal{N}_{\le k}(v_1)$ is bounded by

$$\mathbb{E}[M_k]\frac{d_{\min}}{d_{\min}-2}(d_{\min}-1)^k, \quad (5.3.11)$$

which is the second contribution to the r.h.s. of (5.3.3).

Finally, we study the case where $\mathcal{N}_{\le k}(v_1)\cap\mathcal{N}_{\le k}(v_2)\neq\varnothing$, but $v_2\notin\mathcal{N}_{\le k}(v_1)$. When $\mathcal{N}_{\le k}(v_1)\cap\mathcal{N}_{\le k}(v_2)\neq\varnothing$, but $v_2\notin\mathcal{N}_{\le k}(v_1)$, then one of the $d_{\min}(d_{\min}-1)^k$ half-edges in $\mathcal{N}_k(v_1)$ needs to be connected to one of the $d_{\min}(d_{\min}-1)^{l-k}$ half-edges in $\mathcal{N}_{l-k}(v_2)$, where $l=\mathrm{dist}_{\mathrm{CM}_n(\boldsymbol{d})}(v_1,v_2)\in[2k]\setminus[k]$. Conditionally on v_1 being k-minimally

connected and v_2 being $l-k$-minimally connected, the probability that this occurs is at most

$$\frac{d_{\min}(d_{\min}-1)^k d_{\min}(d_{\min}-1)^{l-k}}{\ell_n - 2i_k - 2i_{l-k}+1} \leq \frac{2d_{\min}(d_{\min}-1)^k d_{\min}(d_{\min}-1)^{l-k-1}}{\ell_n}, \tag{5.3.12}$$

where, in the last inequality, we used that $d_{\min}(d_{\min}-1)^{k-1} \leq \ell_n/8$. Therefore, this contribution is bounded by

$$\mathbb{E}[M_k] \sum_{l=k+1}^{2k} \mathbb{E}[M_{l-k}] \frac{2d_{\min}(d_{\min}-1)^k d_{\min}(d_{\min}-1)^{l-k-1}}{\ell_n}. \tag{5.3.13}$$

We bound $\mathbb{E}[M_{l-k}] \leq n_{d_{\min}}$ and sum

$$\sum_{l=k+1}^{2k} (d_{\min}-1)^{l-k-1} \leq \frac{(d_{\min}-1)^{2k}}{d_{\min}-2}, \tag{5.3.14}$$

to arrive at the third and final contribution to the r.h.s. of (5.3.3). □

To complete the proof of the lower bound on the diameter, we fix $\varepsilon > 0$ sufficiently small, and take $k^* = (1-\varepsilon)\frac{\log\log n}{\log(d_{\min}-1)}$. Clearly,

$$d_{\min}(d_{\min}-1)^{k^*-1} \leq (\log n)^{1-\varepsilon} \leq \ell_n/8, \tag{5.3.15}$$

so that, in particular, we may use Lemma 5.20.

We note that $n_{d_{\min}}/n \to p_{d_{\min}}$ by [I, Condition 7.7(a)] and $p_{d_{\min}} > 0$ by assumption. Therefore, by [I, Condition 7.7(a)-(b)], $d_{\min} n_{d_{\min}}/\ell_n \to \lambda_{d_{\min}}$, where we define $\lambda_{d_{\min}} = d_{\min} p_{d_{\min}}/\mathbb{E}[D]$. By (5.3.2) in Lemma 5.20,

$$\mathbb{E}[M_k] \geq n(p_{d_{\min}} - \delta)(\lambda_{d_{\min}} - \delta)^{d_{\min}(d_{\min}-1)^{k-1}}. \tag{5.3.16}$$

As a result, $\mathbb{E}[M_{k^*}] \geq n(p_{d_{\min}} - \delta)(\lambda_{d_{\min}} - \delta)^{(\log n)^{1-\varepsilon}}$. Further, by (5.3.3) in Lemma 5.20,

$$\mathrm{Var}(M_{k^*}) = o(\mathbb{E}[M_{k^*}]^2), \tag{5.3.17}$$

so that

$$M_{k^*}/\mathbb{E}[M_{k^*}] \xrightarrow{\mathbb{P}} 1. \tag{5.3.18}$$

We conclude that, whp, $M_{k^*} \geq n^{1-o(1)}$. Since each minimally-k^*-connected vertex uses up at most

$$1 + \sum_{l=1}^{k^*} d_{\min}(d_{\min}-1)^{l-1} = n^{o(1)} \tag{5.3.19}$$

vertices of degree $d_{\min}$, whp there must be at least two minimally-k^*-connected vertices whose k^*-neighborhoods are disjoint. We fix two such vertices and denote them by v_1 and v_2. We note that v_1 and v_2 have precisely $d_{\min}(d_{\min}-1)^{k^*-1}$ unpaired half-edges in $\mathcal{N}_{k^*}(v_1)$ and $\mathcal{N}_{k^*}(v_2)$. Let $\mathcal{A}_{12}$ denote the event that v_1, v_2 are minimally-k^*-connected with their k^*-neighborhoods being disjoint.

Conditionally on $\mathcal{A}_{12}$, the random graph obtained by collapsing the half-edges in $\mathcal{N}_{k^*}(v_1)$ to a single vertex a and the half-edges in $\mathcal{N}_{k^*}(v_1)$ to a single vertex b is a configuration model on the vertex set $\{a,b\} \cup [n] \setminus (\mathcal{N}_{\leq k^*}(v_1) \cup \mathcal{N}_{\leq k^*}(v_1))$, having degrees $\tilde{\boldsymbol{d}}$ given by $\tilde{d}_a = \tilde{d}_b = d_{\min}(d_{\min}-1)^{k^*-1}$ and $\tilde{d}_i = d_i$ for every $i \in [n] \setminus (\mathcal{N}_{\leq k^*}(v_1) \cup \mathcal{N}_{\leq k^*}(v_1))$.

By the truncated first moment method on paths, performed in the proof of Theorem 5.11 (recall (5.2.47)), it follows that, for any $\varepsilon > 0$,

$$\mathbb{P}\Big(\mathrm{dist}_{\mathrm{CM}_n(\boldsymbol{d})}(\mathcal{N}_{k^*}(v_1), \mathcal{N}_{k^*}(v_2)) \leq (1-\varepsilon)\frac{2\log\log n}{|\log(\tau-2)|} \mid \mathcal{A}_{12}\Big) = o(1). \tag{5.3.20}$$

Therefore, whp,

$$\begin{aligned}\mathrm{diam}(\mathrm{CM}_n(\boldsymbol{d})) &\geq (1-\varepsilon)\frac{2\log\log n}{|\log(\tau-2)|} + 2k^* \\ &= (1-\varepsilon)\log\log n\Big[\frac{2}{|\log(\tau-2)|} + \frac{2}{\log(d_{\min}-1)}\Big].\end{aligned} \tag{5.3.21}$$

Since $\varepsilon > 0$ is arbitrary, this proves the lower bound on $\mathrm{diam}(\mathrm{CM}_n(\boldsymbol{d}))$ in Theorem 5.19. □

Upper bound on the diameter. For the upper bound, it is convenient to explore the neighborhood of a vertex v by only pairing up the first $d_{\min}$ half-edges incident to v and the $d_{\min}-1$ half-edges incident to any other vertex appearing in the tree. We call this exploration graph the k-exploration tree. Our main result is the following proposition that shows that, for $k = (1+\varepsilon)\log\log n/\log(d_{\min}-1)$ and whp, the k-exploration tree quickly connects to Core_n:

Proposition 5.21 (Connecting the exploration tree to the core)**.** *Let $k^* = (1+\varepsilon)\frac{\log\log n}{\log(d_{\min}-1)}$. Then, for each $\varepsilon > 0$ and under the conditions in Theorem 5.19, the probability that there exists a vertex at distance at least k^* to the core is $o(1)$.*

By Proposition 5.21 and Theorem 5.14, for each $\varepsilon > 0$ and whp,

$$\begin{aligned}\mathrm{diam}(\mathrm{CM}_n(\boldsymbol{d})) &\leq 2(1+\varepsilon)\frac{\log\log n}{\log(d_{\min}-1)} + \mathrm{diam}(\mathrm{Core}_n) \\ &\leq 2(1+\varepsilon)\frac{\log\log n}{\log(d_{\min}-1)} + 2(1+\varepsilon)\frac{\log\log n}{|\log(\tau-2)|} \\ &= 2(1+\varepsilon)\log\log n\Big[\frac{1}{\log(d_{\min}-1)} + \frac{1}{|\log(\tau-2)|}\Big].\end{aligned} \tag{5.3.22}$$

Since $\varepsilon > 0$ is arbitrary, this proves the upper bound on $\mathrm{diam}(\mathrm{CM}_n(\boldsymbol{d}))$ in Theorem 5.19. □

The proof of Proposition 5.21 is organized as follows. We start by showing that whp the k^*-exploration tree does not contain more than one collisions. Here, by a collision, we mean a cycle in the k^*-exploration tree, so that one of the half-edges that is paired to form the k^*-exploration tree is paired to a half-edge incident to it as well. Thus,

whp, the k-exploration trees incident to all vertices have at least $(d_{\min}-2)(d_{\min}-1)^k$ half-edges incident to it. Secondly, we show that when we do not pair too many half-edges, the probability of connecting to a sleeping vertex remains on being substantial. Thirdly, we use that bound to show that each half-edge has a reasonable probability of connecting to Core_n within $o(\log\log n)$ steps. We complete the proof of Proposition 5.21 by shoing that whp at least one of the at least $(d_{\min}-2)(d_{\min}-1)^k$ half-edges incident to the k-exploration tree succeeds in connecting to Core_n quickly.

For k of order $\log\log n$, the probability that more than *one* collisions occur in the k-exploration tree before hitting Core_n is small, as we prove now:

Lemma 5.22 (At most one collision)**.** *Let $k = (1+\varepsilon/2)\frac{\log\log n}{\log(d_{\min}-1)}$. Then, under the conditions in Theorem 5.19, the probability that there exists a vertex whose k-exploration tree has at least two collisions before hitting the core Core_n, is bounded by $(d_{\min})^4(\log n)^d\ell_n^{-2}$, where $d=(1+\varepsilon)+2\sigma$.*

Proof. In this proof, we abbreviate $m = d_{\min}$. For any half-edge in the k-exploration tree, the probability that it create a collision before hitting the core is bounded above by $(m+1)m^{k-1}(\log n)^\sigma \ell_n^{-1}$. The probability that two half-edges both create a collision is, by similar arguments, bounded above by $\big[(m+1)m^{k-1}(\log n)^\sigma\ell_n^{-1}\big]^2$. The total number of possible pairs of half-edges in the k-exploration tree is bounded by

$$[(m+1)(1+m+\ldots+m^{k-1})]^2 \le [(m+1)m^k]^2,$$

so that, by Boole's inequality, the probability that the k-exploration tree has at least two collisions is bounded by

$$\big[(m+1)m^k\big]^4(\log n)^{2\sigma}\ell_n^{-2}. \tag{5.3.23}$$

When $k \le (1+\varepsilon)\frac{\log\log n}{\log m}$, this is bounded by $\big[(m+1)m^k\big]^4(\log n)^{2\sigma} \le (m+1)^4(\log n)^d$, where $d=(1+\varepsilon)+2\sigma$. □

In our proof, we often rely on the following lemma that shows that the probability that we pair a half-edge to a half-edge incident to a vertex of high degree does not decrease too fast throughout the pairing procedure:

Lemma 5.23 (Finding fresh vertices of high degree)**.** *Let $\mathcal{F}_l$ be the σ-algebra generated by the pairing of the first l half-edges, and denote by x_j the jth half-edge, y_j the half-edge to which x_j is paired and V_{y_j} the vertex to which y_j is incident. Then, for all z such that $l \le n[1-F_n](z)/4$,*

$$\mathbb{P}(d_{V_{y_{l+1}}} > z, V_{y_{l+1}} \notin \{V_{x_1}, V_{y_1}, \ldots, V_{x_l}, V_{y_l}\} \mid \mathcal{F}_l) \ge z[1-F_n](z)\frac{n}{2\ell_n}. \tag{5.3.24}$$

Proof. Let $\mathcal{A}_l$ denote the awake vertices given $\mathcal{F}_l$, i.e., those vertices for which at least one of its incident half-edges have been paired in the pairing of the first k half-edges.

Clearly, $|\mathcal{A}_l| \leq 2l$. Then, for each l,

$$\mathbb{P}(d_{V_{y_{l+1}}} > z, V_{y_{l+1}} \notin \{V_{x_1}, V_{y_1}, \ldots, V_{x_l}, V_{y_l}\} \mid \mathcal{F}_k) = \frac{1}{\ell_n - 2l + 1} \sum_{v \in [n] \setminus \mathcal{A}_k} d_v \mathbb{1}_{\{d_v > z\}} \tag{5.3.25}$$

$$\geq z \frac{n}{\ell_n} \Big([1 - F_n](z) - |\mathcal{A}_l|/n).$$

By assumption, $|\mathcal{A}_l| \leq 2l \leq n[1 - F_n](z)/2$, so that

$$\mathbb{P}(d_{V_{y_{l+1}}} > z, V_{y_{l+1}} \notin \{V_{x_1}, V_{y_1}, \ldots, V_{x_l}, V_{y_l}\} \mid \mathcal{F}_l) \geq z[1 - F_n](z) \frac{n}{2\ell_n}. \tag{5.3.26}$$

□

Denote the different half-edges incident to the kth layer of the k-exploration tree by $x_1, \ldots, x_N$, where $N \in (\log n)^{1+\varepsilon/2}[(d_{\min} - 1)/d_{\min}, (d_{\min} + 1)/d_{\min}]$. We call a half-edge a *success* when it reaches Core_n in at most $h = A \log\log\log n$ steps. Denote the event that x_j is a success by $\mathcal{S}_{x_j}$. Let $\mathcal{F}'_k$ denote the σ-algebra generated by the k-exploration tree. We start by giving a lower bound on the probability of $\mathcal{S}_{x_j}$:

Lemma 5.24 (Substantial probability of connecting to Core_n quickly)**.** *For $k_- = (1 + \varepsilon/2) \log\log n / \log(d_{\min} - 1)$, there exists $\eta > 0$ such that, uniformly in $n \geq 1$,*

$$\mathbb{P}(\mathcal{S}_{x_1} \mid \mathcal{F}'_k) \geq \eta \tag{5.3.27}$$

and, for each $j = 1, \ldots, N = (\log n)^a$,

$$\mathbb{P}(\mathcal{S}_{x_j} \mid \mathcal{F}'_k, \mathcal{S}^c_{x_1}, \ldots, \mathcal{S}^c_{x_{j-1}}) \geq \eta. \tag{5.3.28}$$

Proof. We first prove (5.3.27), and start by defining some notation. We pair x_1 to y_1, say, and let v_0 be the vertex incident to y_1. We recall that $h = A \log\log\log n$, and recursively define $v_0, \ldots, v_h$. Let $v_l \in [n]$ denote the vertex (if any) with maximal degree that is paired to a half-edge incident to v_{l-1}. Let $\varepsilon > 0$ be so small that $(1-\varepsilon)/(\tau-2) > 1$ and let $A \geq 1$ be a large constant. Denote $p = (1-\varepsilon)/(\tau - 2 + \delta)$, and pick $\varepsilon > 0$ and $\delta > 0$ so small that $p > 1$. If $d_{v_j} \geq A^{p^j}$ for each $j = 1, \ldots, h$, then

$$d_{v_h} \geq A^{p^{A \log\log\log n}} = \mathrm{e}^{\log A (\log\log n)^{A \log p}}, \tag{5.3.29}$$

which is much larger than $(\log n)^\sigma$ when $A > 1/\log p$. Therefore, $v_h \in \mathrm{Core}_n$, so that the half-edge x_i is a success and $\mathcal{S}_{x_1}$ occurs. As a result,

$$\mathbb{P}(\mathcal{S}_{x_1} \mid \mathcal{F}'_k) \geq \mathbb{P}(d_{v_j} \geq A^{p^j} \ \forall j = 0, \ldots, h, d_{v_j} \leq (\log n)^\sigma \ \forall j = 1, \ldots, h-1 \mid \mathcal{F}'_k). \tag{5.3.30}$$

In order to study the probability to pair a half-edge to a half-edge incident to a sleeping vertex having large degree, we use Lemma 5.23.

To apply Lemma 5.23, we rely on the lower bound on the empirical distribution function F_n in (5.1.7), which states that for all $\delta > 0$, there exists $c_1 = c_1(\delta)$ such that, uniformly in n and for all $z \leq n^\alpha$ for some $\alpha > 1/2$,

$$[1 - F_n](z) \geq c_1 z^{-(\tau - 1 + \delta)}. \tag{5.3.31}$$

As a result, we may apply Lemma 5.23 as long as l, z and n satisfy $l \le c_1 n z^{-(\tau-1+\delta)}/4$. By [I, Condition 7.7(a)-(b)], $\ell_n/n \to \mathbb{E}[D]$, so, for n sufficiently large, $\ell_n/n \le 2\mathbb{E}[D]$. We conclude that, with $c_1' = c_1/(4\mathbb{E}[D])$ and for any l, z and n satisfying $l \le c_1 n z^{-(\tau-1+\delta)}/4$

$$\mathbb{P}(d_{V_{y_{l+1}}} > z, V_{y_{l+1}} \notin \{V_{x_1}, V_{y_1}, \ldots, V_{x_l}, V_{y_l}\} \mid \mathcal{F}_l) \ge c_1' z^{-(\tau-2+\delta)}. \tag{5.3.32}$$

In the construction of the k-exploration tree, at most $l_1 = d_{\min}(d_{\min}-1)^{k-1}$ half-edges have been paired. Therefore, in particular, for any $A \ge 1$ fixed,

$$\mathbb{P}(d_{v_0} \ge A \mid \mathcal{F}_k') \ge c_1' A^{-(\tau-2+\delta)}. \tag{5.3.33}$$

Further, let $\mathcal{F}_{k,j}'$ denote the σ-algebra generated by $\mathcal{F}_k'$ and the pairing of the half-edges incident to $v_0, \ldots, v_j$. Then, as long as $d_{v_0}, \ldots, d_{v_j} \le (\log n)^\sigma$, the number of pairings involved in $\mathcal{F}_{k,j}'$ is at most $l_1 + l_2$, where $l_1 = d_{\min}(d_{\min}-1)^{k-1}$ and $l_2 = (j+1)(\log n)^\sigma$.

For $t = 1, \ldots, d_{v_j}$, let $x_{j,t}$ be the half-edge that needs to be paired, and let $y_{j,t}$ be the half-edge to which $x_{j,t}$ is paired. Then, by Lemma 5.23 and uniformly for $z \le (\log n)^\sigma$, and on the event that $d_{v_0}, \ldots, d_{v_j} \le (\log n)^\sigma$,

$$\mathbb{P}(d_{V_{y_{j,t}}} > z \mid \mathcal{F}_{k,j}') \ge c_1' z^{-(\tau-2+\delta)}. \tag{5.3.34}$$

Therefore, on the event that $d_{v_j} \ge A^{p^j}$ and $d_{v_0}, \ldots, d_{v_j} \le (\log n)^\sigma$, we conclude that, for $j \le h-1$,

$$\mathbb{P}(d_{v_{j+1}} \le z \mid \mathcal{F}_{k_-,j}') \le \prod_{t=1}^{A^{p^j}} (1 - c_1' z^{-(\tau-2+\delta)}) \le \mathrm{e}^{-cz^{-(\tau-2+\delta)}A^{p^j}}. \tag{5.3.35}$$

Taking $z = A^{p^{j+1}}$ and recalling that $p = (1-\varepsilon)/(\tau-2+\delta)$, we see that

$$\begin{aligned} \mathbb{P}(d_{v_{j+1}} > z \mid \mathcal{F}_{k,j}') &= 1 - \mathbb{P}(d_{v_{j+1}} \le x \mid \mathcal{F}_{k,j}') \\ &\ge 1 - \mathrm{e}^{-cA^{-(\tau-2+\delta)p^{j+1}}A^{p^j}} = 1 - \mathrm{e}^{-cA^{\varepsilon p^j}}. \end{aligned} \tag{5.3.36}$$

Therefore,

$$\begin{aligned} \mathbb{P}(\mathcal{S}_{x_1} \mid \mathcal{F}_k') &= \mathbb{E}\Big[\prod_{j=0}^{h} \mathbb{P}(d_{v_{i,j+1}} > A^{p^j} \mid \mathcal{F}_{k,j}') \mid \mathcal{F}_k'\Big] \\ &\ge c_1' A^{-(\tau-2+\delta)} \prod_{j=1}^{h} [1 - \mathrm{e}^{-c_1' A^{\varepsilon p^j}}] \ge c_1' A^{-(\tau-2+\delta)}/2, \end{aligned} \tag{5.3.37}$$

where we have used (5.3.33) and (5.3.36), and we assume that $A = A(\varepsilon)$ is sufficiently large. Denoting $\eta = c_1' A^{-(\tau-2+\delta)}/2 > 0$ completes the proof of (5.3.27).

To prove (5.3.28), we note that in the proof of (5.3.27), we only relied on an upper bound on the number of pairings that have been performed. This was necessary to

apply Lemma 5.23. When $\mathcal{S}^c_{x_1}, \ldots, \mathcal{S}^c_{x_{j-1}}$ occur, then the number of pairings is at most $l = l_1 + l_2$, where $l_1 = d_{\min}(d_{\min}-1)^{k-1}$ and $l_2 = Nh(\log n)^\sigma$. This number is bounded by $A(\log n)^{a+\sigma}\log\log\log n$. Therefore, (5.3.33) and (5.3.36) still apply, and we obtain (5.3.28) with $\eta = c_1' A^{-(\tau-2+\delta)}/2 > 0$. This completes the proof of Lemma 5.24. □

Now we are ready to complete the proof of Proposition 5.21:

Proof of Proposition 5.21. By Lemma 5.22, the probability that there exists a vertex $i \in [n]$ whose k-exploration tree has at least 2 collisions before hitting the core is $o(1)$.

Fix $i \in [n]$, and assume that its k-exploration tree has at most 2 collisions. Let N denote the number of half-edges in the kth layer of the k-exploration tree. Then $N \geq (m-1)m^{k-1}$, where we recall that $m = d_{\min} - 1$. For $k = (1+\varepsilon/2)\log\log n/\log(d_{\min}-1)$, $N \geq (\log n)^{1+\varepsilon/2}/2$. Throughout this proof, A denotes a large but finite constant.

By Lemma 5.24, the probability that none of the $N = (\log n)^{1+\varepsilon/2}/2$ half-edges in the kth layer of the k-exploration tree is a success is at most

$$\begin{aligned}\mathbb{P}(\mathcal{S}^c_{x_1} \cap \ldots \cap \mathcal{S}^c_{x_N} \mid \mathcal{F}'_k) &= \mathbb{P}(\mathcal{S}^c_{x_1} \mid \mathcal{F}'_k)\prod_{j=2}^{N}\mathbb{P}(\mathcal{S}^c_{x_j} \mid \mathcal{F}'_k, \mathcal{S}^c_{x_1} \cap \ldots \cap \mathcal{S}^c_{x_{j-1}}) \qquad (5.3.38)\\ &\leq (1-\eta)^N = o(1/n),\end{aligned}$$

since $N \geq (\log n)^{1+\varepsilon/2}/2$. The distance between vertex i and Core_n is at most $h+k = A\log\log\log n + (1+\varepsilon/2)\log\log n/\log(d_{\min}-1) \leq (1+\varepsilon)\log\log n/\log(d_{\min}-1) = k^*$ when at least one of the half-edges of vertex i is a success. Therefore, the probability that the distance between vertex i and Core_n is larger than k^* is $o(1/n)$, as required. □

5.4 Branching processes with infinite mean

When $\tau \in (2,3)$, the branching processes $(\mathcal{Z}^{(1)}_j)_{j\geq 0}$ and $(\mathcal{Z}^{(2)}_j)_{j\geq 0}$ are well-defined, but has infinite mean in generations 2, 3, etc. This leads us to consider *branching processes with infinite mean*. In this section, we give a scaling result for the generation sizes for such branching processes. This result will be crucial to describe the *fluctuations* of the typical distances in $\mathrm{CM}_n(\boldsymbol{d})$. The main result is the following theorem:

Theorem 5.25 (Branching processes with infinite mean)**.** *Let $(Z_n)_{n\geq 0}$ be a branching process with offspring distribution $Z_1 = X$ having distribution function F_X. Assume that there exist $\alpha \in (0,1)$ and a non-negative, non-increasing function $x \mapsto \gamma(x)$, such that*

$$x^{-\alpha-\gamma(x)} \leq 1 - F_X(x) \leq x^{-\alpha+\gamma(x)}, \qquad \text{for large } x, \qquad (5.4.1)$$

where $x \mapsto \gamma(x)$ satisfies

(i) *$x \mapsto x^{\gamma(x)}$ is non-decreasing,*

(ii) $\int_0^\infty \gamma(\mathrm{e}^{\mathrm{e}^x})\,dx < \infty$, *or, equivalently,* $\int_{\mathrm{e}}^\infty \frac{\gamma(y)}{y\log y}\,dy < \infty$.

Then $\alpha^n \log(Z_n \vee 1) \xrightarrow{a.s.} Y$, *with* $\mathbb{P}(Y = 0)$ *equal to the extinction probability of* $(Z_n)_{n\geq 0}$, *whereas* Y *admits a density 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 5.25 covers the case where the branching process has an offspring which has very thick tails. Indeed, it is not hard to show that Theorem 5.25 implies that $\mathbb{E}[X^s] = \infty$ for every $s > \alpha \in (0,1)$ (see Exercise 5.11 below).

We do not prove Theorem 5.25 in full generality. Rather, we prove it in a simpler, yet still quite general case, in which $\gamma(x) = (\log x)^{\gamma-1}$ for some $\gamma \in [0,1)$.

Exercise 5.10 (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 5.25.*

Proof of Theorem 5.25 for $\gamma(x) = (\log x)^{\gamma-1}$. The proof is divided into four main steps. Define

$$M_n = \alpha^n \log(Z_n \vee 1). \tag{5.4.2}$$

We shall first assume that $\mathbb{P}(Z_1 \geq 1) = 1$, so that $\eta = 1$. We start by splitting M_n in a suitable way.

The split. For $i \geq 1$, we define

$$Y_i = \alpha^i \log\Big(\frac{(Z_i \vee 1)}{(Z_{i-1} \vee 1)^{1/\alpha}}\Big). \tag{5.4.3}$$

We can write

$$M_n = Y_1 + Y_2 + \cdots + Y_n. \tag{5.4.4}$$

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

$$\sum_{i=1}^\infty \mathbb{E}\big[|Y_i|\big] < \infty. \tag{5.4.5}$$

This is what we prove in the following three steps.

Inserting normalization sequences. We next investigate $\mathbb{E}\big[|Y_i|\big]$. We prove by induction on i that there exist constants $\kappa < 1$ and $C > 0$ such that

$$\mathbb{E}\big[|Y_i|\big] \leq K\kappa^i. \tag{5.4.6}$$

For $i = 0$, this follows from the fact that, when (5.4.1) holds, the random variable $Y_1 = \alpha \log(Z_1 \vee 1)$ has a bounded absolute expectation. This initializes the induction hypothesis. We next turn to the advancement of the induction hypothesis. For this, we recall the definition of u_n in (2.6.7) and define

$$U_i = \alpha^i \log\Big(\frac{u_{Z_{i-1}\vee 1}}{(Z_{i-1} \vee 1)^{1/\alpha}}\Big), \qquad V_i = \alpha^i \log\Big(\frac{Z_i \vee 1}{u_{Z_{i-1}\vee 1}}\Big). \tag{5.4.7}$$

Then, $Y_i = U_i + V_i$, so that

$$\mathbb{E}\big[|Y_i|\big] \le \mathbb{E}\big[|U_i|\big] + \mathbb{E}\big[|V_i|\big]. \tag{5.4.8}$$

We bound each of these terms separately.

Bounding the normalizing constants. In this step, we analyse the normalizing constants $n \mapsto u_n$, assuming (5.4.1), and use this, as well as the induction hypothesis, to bound $\mathbb{E}\big[|U_i|\big]$.

When (5.4.1) holds and since $\lim_{x\to\infty} \gamma(x) = 0$, there exists a constant $C_\varepsilon \ge 1$ such that, for all $n \ge 1$,

$$u_n \le C_\varepsilon n^{1/\alpha+\varepsilon}, \tag{5.4.9}$$

This gives a first bound on $n \mapsto u_n$. We next substitute this bound into (5.4.1) and use that $x \mapsto x^{\gamma(x)}$ is non-decreasing together with $\gamma(x) = (\log x)^{\gamma-1}$, to obtain that

$$1 + o(1) = n[1 - F_X(u_n)] \ge n\big[u_n^{-(\tau-1)-\gamma(u_n)}\big] \ge n\Big[u_n^{-(\tau-1)} \mathrm{e}^{\log\left(C_\varepsilon n^{\frac{1}{\alpha}+\varepsilon}\right)^\gamma}\Big], \tag{5.4.10}$$

which, in turn, implies that there exists a constant $c > 0$ such that

$$u_n \le n^{1/\alpha} \mathrm{e}^{c(\log n)^\gamma}. \tag{5.4.11}$$

In a similar way, we can show the matching lower bound $u_n \ge n^{1/\alpha}\mathrm{e}^{-c(\log n)^\gamma}$. As a result,

$$\mathbb{E}\big[|U_i|\big] \le c\alpha^i \mathbb{E}\big[(\log(Z_{i-1} \vee 1))^\gamma\big]. \tag{5.4.12}$$

Using the concavity of $x \mapsto x^\gamma$ for $\gamma \in [0,1)$, as well as Jensen's Inequality, we arrive at

$$\mathbb{E}\big[|U_i|\big] \le c\alpha^i \Big(\mathbb{E}\big[(\log(Z_{i-1} \vee 1))\big]\Big)^\gamma = \alpha^{i(1-\gamma)} \mathbb{E}[M_{i-1}]^\gamma. \tag{5.4.13}$$

By (5.4.4) and (5.4.6), which implies that $\mathbb{E}[M_{i-1}] \le K\kappa/(1-\kappa)$, we arrive at

$$\mathbb{E}\big[|U_i|\big] \le \alpha^{i(1-\gamma)} c\Big(\frac{K\kappa}{1-\kappa}\Big)^\gamma. \tag{5.4.14}$$

Bounding the logarithmic moment of an asymptotic stable random variable. In this step, we bound $\mathbb{E}\big[|V_i|\big]$. We note that by [I, Theorem 2.31] and for Z_i quite large, the random variable $(Z_i \vee 1)/(u_{Z_{i-1}\vee 1})$ should be close to a stable random variable. We make use of this fact by bounding

$$\mathbb{E}\big[|V_i|\big] \le \alpha^i \sup_{m\ge 1} \mathbb{E}\big[|\log\left(S_m/u_m\right)|\big], \tag{5.4.15}$$

where $S_m = X_1 + \cdots + X_m$, and $(X_i)_{i=1}^m$ are i.i.d. copies of the offspring distribution X. We shall prove that there exists a constant $C > 0$ such that, for all $m \ge 1$,

$$\mathbb{E}\big[|\log\left(S_m/u_m\right)|\big] \le C. \tag{5.4.16}$$

In order to prove (5.4.16), we note that it suffices to bound

$$\mathbb{E}\big[\big(\log\big(S_m/u_m\big)\big)_+\big] \le C_+, \qquad \mathbb{E}\big[\big(\log\big(S_m/u_m\big)\big)_-\big] \le C_-, \tag{5.4.17}$$

where, for $x \in \mathbb{R}$, $x_+ = \max\{x, 0\}$ and $x_- = \max\{-x, 0\}$. Since $|x| = x_+ + x_-$, we then obtain (5.4.16) with $C = C_+ + C_-$. In order to prove (5.4.16), we start by investigating $\mathbb{E}\big[\big(\log\big(S_m/u_m\big)\big)_-\big]$. We note that $(\log x)_- = \log\big(x^{-1} \vee 1\big)$, so that

$$\mathbb{E}\big[\big(\log\big(S_m/u_m\big)\big)_-\big] = \mathbb{E}\big[\log\big(u_m/(S_m \wedge u_m)\big)\big], \tag{5.4.18}$$

where $x \wedge y = \min\{x, y\}$. The function $x \mapsto \log\big((u_m/(x \wedge u_m)\big)$ is non-increasing, and, since $S_m \ge X_{(m)}$, where $X_{(m)} = \max_{1\le i\le m} X_i$, we arrive at

$$\mathbb{E}\big[\log\big(u_m/(S_m \wedge u_m)\big)\big] \le \mathbb{E}\big[\log\big(u_m/(X_{(m)} \wedge u_m)\big)\big]. \tag{5.4.19}$$

We next use that, for $x \ge 1$, $x \mapsto \log(x)$ is concave, so that, for every s,

$$\begin{aligned}\mathbb{E}\big[\log\big(u_m/(X_{(m)} \wedge u_m)\big)\big] &= \frac{1}{s}\mathbb{E}\big[\log\big((u_m/(X_{(m)} \wedge u_m))^s\big)\big] \\ &\le \frac{1}{s}\log\Big(\mathbb{E}\big[\big(u_m/(X_{(m)} \wedge u_m)\big)^s\big]\Big) \\ &\le \frac{1}{s} + \frac{1}{s}\log\Big(u_m^s\mathbb{E}\big[X_{(m)}^{-s}\big]\Big),\end{aligned} \tag{5.4.20}$$

where, in the last step, we made us of the fact that $u_m/(x \wedge u_m) \le 1 + u_m/x$. Now rewrite $X_{(m)}^{-s} = (-Y_{(m)})^s$, where $Y_j = -X_j^{-1}$ and $Y_{(m)} = \max_{1\le j\le m} Y_j$. Clearly, $Y_j \in [-1, 0]$ since $X_i \ge 1$, so that $\mathbb{E}[(-Y_1)^s] < \infty$. Also, $u_m Y_{(m)} = -u_m/X_{(m)}$ converges in distribution to $-E^{-1/\alpha}$, where E is exponential with mean 1, so it follows from [95, Theorem 2.1] that, as $m \to \infty$,

$$\mathbb{E}\left[(u_m Y_{(m)})^p\right] \to \mathbb{E}[E^{-1/\alpha}] < \infty. \tag{5.4.21}$$

We proceed with $\mathbb{E}\big[\big(\log\big(S_m/u_m\big)\big)_+\big]$, for which the proof is a slight adaptation of the above argument. Now we make use of the fact that $(\log x)_+ = \log\big(x \vee 1\big) \le 1 + x$ for $x \ge 0$, so that we must bound

$$\mathbb{E}\big[\log\big(S_m \vee u_m/u_m\big)\big] = \frac{1}{s}\mathbb{E}\big[\log\big((S_m \vee u_m/u_m)\big)^s\big)\big] \le \frac{1}{s} + \log\Big(\mathbb{E}\big[\big(S_m/u_m\big)^s\big]. \tag{5.4.22}$$

The discussion on [60, Page 565 and Corollary 1] yields, for $s < \alpha$, $\mathbb{E}[S_m^s] = \mathbb{E}[|S_m|^s] \le 2^{s/2}\lambda_s(t)$, for some function $\lambda_s(m)$ depending on s, m and F_X. Using the discussion on [60, Page 564], we have that $\lambda_s(m) \le C_s m^{s/\alpha} l(m^{1/\alpha})^s$, where $l(\,\cdot\,)$ is a slowly varying function. With some more effort, it can be shown that we can replace $l(m^{1/\alpha})$ by $\ell(m)$, which gives

$$\mathbb{E}\big[\log\big(S_m \vee u_m/u_m\big)\big] \le \frac{1}{s} + \log\mathbb{E}\big[\big(\frac{S_m}{u_m}\big)^s\big] \le \frac{1}{s} + \frac{C_s}{s}m^{s/\alpha}\ell(m)^s u_m^{-s} = \frac{1}{s} + 2^{s/2}\frac{C_s}{s}, \tag{5.4.23}$$

and which proves the first bound in (5.4.17) with $C_+ = \frac{1}{s} + 2^{s/2}\frac{C_s}{s}$.

Completion of the proof of Theorem 5.25 when $X \geq 1$. Combining (5.4.8) with (5.4.14) and (5.4.15)–(5.4.16), we arrive at

$$\mathbb{E}\big[|Y_i|\big] \leq \alpha^{i(1-\gamma)} c\Big(\frac{K\kappa}{1-\kappa}\Big)^{\gamma} + C\alpha^i \leq K\kappa^i, \tag{5.4.24}$$

when we take $\kappa = \alpha^{1-\gamma}$ and we take K to be sufficiently large, for example $K \geq 2C$ and $K \geq 2c\Big(\frac{K\kappa}{1-\kappa}\Big)^{\gamma}$. We shall not prove that Y admits a density. This completes the proof when the offspring distribution X satisfies $X \geq 1$.

Completion of the proof of Theorem 5.25. We finally extend the result to the setting where $X = 0$ with positive probability. Since $\mathbb{E}[X] = \infty$, the survival probability ζ satisfies $\zeta > 0$. Conditionally on extinction, clearly $Z_n \xrightarrow{a.s.} 0$, so that, on the survival event, $\alpha^n \log(Z_n \vee 1) \xrightarrow{a.s.} Y$, where, conditionally on extinction, $Y = 0$.

It remains to prove that $\alpha^n \log(Z_n \vee 1) \xrightarrow{a.s.} Y$ on the survival event. By Theorem 3.12, we have that, conditionally on survival,

$$\frac{Z_n^{(\infty)}}{Z_n} \xrightarrow{a.s.} \xi > 0, \tag{5.4.25}$$

where we recall that $Z_n^{(\infty)}$ are the individuals in the nth generation which have an infinite line of descent. By [I, Theorem 3.11] and conditionally on survival, $(Z_n^{(\infty)})_{n\geq 0}$ is again a branching process, now with offspring distribution $p^{(\infty)}$ given in [I, (3.4.2)]. Note that, in particular, $\mathbb{P}(Z_1^{(\infty)} \geq 1) = 1$, and we wish to apply Theorem 5.25 to $Z_n^{(\infty)}$ instead of Z_n. It is not hard to show that also $p^{(\infty)}$ in [I, (3.4.2)] satisfies the conditions in Theorem 5.25 with the function $x \mapsto \gamma^*(x)$, given by $\gamma^*(x) = \gamma(x) + c/\log x$. Thus, conditionally on survival,

$$\alpha^n \log(Z_n^{(\infty)} \vee 1) \xrightarrow{a.s.} Y^{(\infty)}, \tag{5.4.26}$$

and combining (5.4.25) and (5.4.26), it immediately follows that, conditionally on survival,

$$\alpha^n \log(Z_n \vee 1) \xrightarrow{a.s.} Y^{(\infty)}. \tag{5.4.27}$$

We conclude that Theorem 5.25 holds, where $Y = 0$ with probability $\eta = 1 - \zeta$ and $Y = Y^{(\infty)}$ with probability ζ. □

Exercise 5.11 (Infinite mean under conditions Theorem 5.25)**.** *Prove that $\mathbb{E}[X] = \infty$ when the conditions in Theorem 5.25 are satisfied. Extend this to show that $\mathbb{E}[X^s] = \infty$ for every $s > \alpha \in (0, 1)$.*

Exercise 5.12 (Conditions in Theorem 5.25 for individuals with infinite line of descent)**.** *Prove that $p^{(\infty)}$ in [I, (3.4.2)] satisfies the conditions in Theorem 5.25 with the function $x \mapsto \gamma^*(x)$, given by $\gamma^*(x) = \gamma(x) + c/\log x$.*

Exercise 5.13 (Convergence for Z_n+1)**.** *Show that, under the conditions of Theorem 5.25, also $\alpha^n \log(Z_n + 1)$ converges to Y almost surely.*

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

Theorem 5.26 (Limiting variable for infinite-mean branching processes)**.** *Under the conditions of Theorem 5.25,*

$$\lim_{x\to\infty} \frac{\log \mathbb{P}(Y > x)}{x} = -1, \tag{5.4.28}$$

where is the a.s. limit of $\alpha^n \log(Z_n \wedge 1)$.

Theorem 5.26 can be understood from the fact that by (5.4.2)–(5.4.3),

$$Y = \sum_{n=1}^{\infty} Y_i, \tag{5.4.29}$$

where

$$Y_1 = \alpha \log\big(Z_1 \vee 1\big). \tag{5.4.30}$$

By (5.4.1),

$$\mathbb{P}(Y_1 > x) = \mathbb{P}(Z_1 > \mathrm{e}^{x^{1/\alpha}}) = \mathrm{e}^{-x(1+o(1))}, \tag{5.4.31}$$

which shows that Y_1 satisfies (5.4.28). The equality in (5.4.29) together with (5.4.3) suggests that the tails of Y_1 are equal to those of Y, which heuristically explains (5.4.28).

5.5 Related results for the configuration model

In this section, we discuss related results for the configuration model. We start by discussing the distances in infinite-mean configuration models.

Infinite mean degrees. In this section, we assume that there exist $\tau \in (1, 2)$ and $c > 0$ such that

$$\lim_{x\to\infty} x^{\tau-1}[1 - F](x) = c. \tag{5.5.1}$$

We study the configuration model $\mathrm{CM}_n(\boldsymbol{d})$ where the degrees $\boldsymbol{d} = (d_i)_{i\in[n]}$ are an i.i.d. sequence of random variables with distribution F satisfying (5.5.1).

We make heavy use of the notation used in [I, Theorem 7.21], which we first recall. Recall that the random probability distribution $P = (P_i)_{i\geq 1}$ is given by

$$P_i = Z_i/Z, \tag{5.5.2}$$

where $Z_i = \Gamma_i^{-1/(\tau-1)}$ and $\Gamma_i = \sum_{j=1}^{i} E_i$ with $(E_i)_{i\geq 1}$ an i.i.d. sequence of exponential random variables with parameter 1, and where $Z = \sum_{i\geq 1} \Gamma_i$. Recall further that $M_{P,k}$ is a multinomial distribution with parameters k and probabilities $P = (P_i)_{i\geq 1}$. Thus, $M_{P,k} = (B_1, B_2, \ldots)$, where, conditionally on $P = (P_i)_{i\geq 1}$, B_i is the number of outcomes i in k independent trials such that each outcome is equal to i with

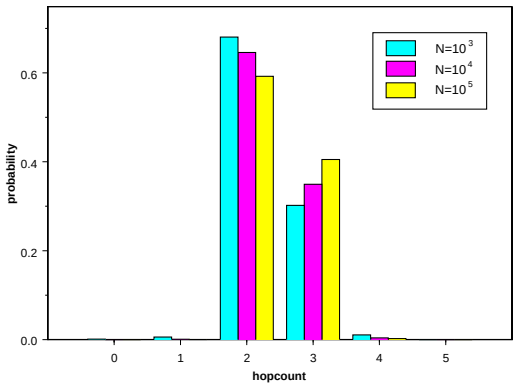

Figure 5.1: Empirical probability mass function of the hopcount for $\tau = 1.8$ and $N = 10^3, 10^4, 10^5$.

probability P_i. In Theorem 7.21, the random variable M_{P,D_1} appears, where D_1 is independent of $P = (P_i)_{i\geq 1}$. We let $M^{(1)}_{P,D_1}$ and $M^{(2)}_{P,D_2}$ be two random variables which are conditionally independent given $P = (P_i)_{i\geq 1}$. In terms of this notation, the main result on distances in $\mathrm{CM}_n(\boldsymbol{d})$ when the degrees have infinite mean is the following:

Theorem 5.27 (Distances in $\mathrm{CM}_n(\boldsymbol{d})$ with i.i.d. infinite mean degrees)**.** *Fix $\tau \in (1,2)$ in (5.5.1) and let $(d_i)_{i\in[n]}$ be a sequence of i.i.d. copies of D. Then, $\mathrm{CM}_n(\boldsymbol{d})$ satisfies*

$$\lim_{n\to\infty} \mathbb{P}(H_n = 2) = 1 - \lim_{n\to\infty} \mathbb{P}(H_n = 3) = p_F \in (0,1). \tag{5.5.3}$$

The probability p_F can be identifies as the probability that $M^{(1)}_{P,D_1}$ and $M^{(2)}_{P,D_2}$ have an identical outcome, i.e., there is an outcome that occurs both in $M^{(1)}_{P,D_1}$ and $M^{(2)}_{P,D_2}$, where D_1 and D_2 are two i.i.d. copies of D.

Proof. We sketch the proof of Theorem 5.27. First, whp, both $d_1 \leq \log n$ and $d_2 \leq \log n$. The event that $H_n = 1$ occurs precisely when one of the d_1 half-edges of vertex 1 is attached to one of the d_2 half-edges of vertex 2. Also, with high probability, $\ell_n \geq n^{1/(\tau-1)-\varepsilon}$. Therefore, on the event that $\ell_n \geq n^{1/(\tau-1)-\varepsilon}$ and $d_1 \leq \log n$ and $d_2 \leq \log n$, the probability that $H_n = 1$ is bounded above by

$$\frac{(\log n)^2}{n^{1/(\tau-1)-\varepsilon}} = o(1). \tag{5.5.4}$$

Exercise 5.14 (Typical distance is at least 2 whp)**.** *Complete the argument that $\mathbb{P}(H_n = 1) = o(1)$.*

We note that the proof of Theorem 7.21 implies that $M^{(1)}_{P,d_1}$ denotes the number of edges between vertex 1 and the largest order statistics. Indeed, $M^{(1)}_{P,d_1} = (B^{(1)}_1, B^{(1)}_2, \ldots)$, where $B^{(1)}_i$ is the number of edges between vertex i and the vertex with degree $d_{(n+1-i)}$. The same applies to vertex 2. As a result, when $M^{(1)}_{P,d_1}$ and

$M^{(2)}_{P,d_2}$ have an identical outcome, then the typical graph distance equals 2. We are left to prove that the typical graph distance is bounded by 3 with high probability. By (2.6.17), we have that $\xi_k k^{1/(\tau-1)} \xrightarrow{\mathbb{P}} 1$ as $k \to \infty$. Thus, when K is large, the probability that vertex 1 is not connected to any of the vertices corresponding to $(d_{(n+1-i)})_{i=1}^K$ converges to 0 when K tends to infinity.

Let $\mathbb{P}_n$ denote the conditional probability given the degrees $(d_i)_{i\in[n]}$. For $i \in [n]$, we let v_i be the vertex corresponding to the ith order statistic $d_{(n+1-i)}$. By Lemma 5.16,

$$\mathbb{P}_n(v_i \text{ not directly connected to } v_j) \le \mathrm{e}^{-\frac{d_{(n+1-i)}d_{(n+1-j)}}{2\ell_n}}. \tag{5.5.5}$$

Moreover, $d_{(n+1-i)}, d_{(n+1-j)} \ge n^{1/(\tau-1)-\varepsilon}$ with high probability for n sufficiently large and any $\varepsilon > 0$, while whp $\ell_n \le n^{1/(\tau-1)+\varepsilon}$. As a result, whp,

$$\mathbb{P}_n(v_i \text{ not directly connected to } v_j) \le \mathrm{e}^{-n^{1/(\tau-1)-3\varepsilon}}. \tag{5.5.6}$$

Therefore, for fixed K and for every $i, j \in [K]$, the vertices v_i and v_j are whp neighbors. This implies that the vertices corresponding to the high order statistics form a *complete graph*. We have already concluded that 1 is connected to v_i for some $i \le K$. In the same way, we conclude that vertex 2 is connected to v_j for some $j \le K$. Since v_i is whp connected to v_j, we conclude that

$$\mathbb{P}_n\big(H_n \le 3\big) = 1 - o(1). \tag{5.5.7}$$

This completes the proof. □

Exercise 5.15 (Typical distance equals 2 whp for $\tau = 1$)**.** *Let the $(d_i)_{i\in[n]}$ be a sequence of i.i.d. copies of D with distribution function F satisfying that $x \mapsto [1 - F](x)$ is slowly varying at ∞. Prove that $\mathrm{CM}_n(\boldsymbol{d})$ satisfies that $H_n \xrightarrow{\mathbb{P}} 2$.*

Fluctuation of distances for finite-variance degrees. We continue to study the *fluctuations* of the distances in the configuration model, starting with the case where the degrees have finite variance. We need a limit result from branching process theory before we can identify the limiting random variables $(R_a)_{a\in(-1,0]}$. Recall that $(\mathcal{Z}_k)_{k\ge0}$ denotes the two-stage branching process where in the first generation, the offspring has distribution D with distribution function F and in the second and further generations, the offspring has distribution $D^* - 1$, where D^* is the size-biased distribution of D. The process $(\mathcal{Z}_k/\mathbb{E}[D]\nu^{k-1})_{k\ge1}$ is a martingale with uniformly bounded expectation and consequently converges almost surely to a limit (see e.g., [I, Theorem 2.22 and Exercise 2.23]):

$$\lim_{n\to\infty} \frac{\mathcal{Z}_n}{\mathbb{E}[D]\nu^{n-1}} = \mathcal{W} \qquad a.s. \tag{5.5.8}$$

In the theorem below we need two independent copies $\mathcal{W}^{(1)}$ and $\mathcal{W}^{(2)}$ of $\mathcal{W}$.

Theorem 5.28 (Limit law for typical distance in $\mathrm{CM}_n(\boldsymbol{d})$)**.** *Let $(d_i)_{i\in[n]}$ be a sequence of i.i.d. copies of a random variable D, and assume that there exist $\tau > 3$ and $c < \infty$ such that, for all $x \geq 1$,*

$$[1 - F](x) \leq cx^{-(\tau-1)}, \tag{5.5.9}$$

and let $\nu > 1$. For $k \geq 1$, let $a_k = \lfloor \log_\nu k \rfloor - \log_\nu k \in (-1, 0]$. Then, $\mathrm{CM}_n(\boldsymbol{d})$ satisfies that there exist random variables $(R_a)_{a\in(-1,0]}$ such that as $n \to \infty$,

$$\mathbb{P}\big(H_n - \lfloor \log_\nu n \rfloor = k \mid H_n < \infty\big) = \mathbb{P}(R_{a_n} = k) + o(1), \qquad k \in \mathbb{Z}. \tag{5.5.10}$$

The random variables $(R_a)_{a\in(-1,0]}$ can be identified as

$$\mathbb{P}(R_a > k) = \mathbb{E}\big[\exp\{-\kappa\nu^{a+k}\mathcal{W}^{(1)}\mathcal{W}^{(2)}\}\big|\mathcal{W}^{(1)}\mathcal{W}^{(2)} > 0\big], \tag{5.5.11}$$

where $\mathcal{W}^{(1)}$ and $\mathcal{W}^{(2)}$ are independent limit copies of $\mathcal{W}$ in (5.5.8) and where $\kappa = \mathbb{E}[D](\nu - 1)^{-1}$.

In words, Theorem 5.28 states that for $\tau > 3$, the graph distance H_n between two randomly chosen connected vertices grows like the $\log_\nu n$, where n is the size of the graph, and that the fluctuations around this mean remain uniformly bounded in n.

The law of R_a is involved, and can in most cases not be computed exactly. The reason for this is the fact that the random variables $\mathcal{W}$ that appear in its statement are hard to compute explicitly (see also Chapter 3).

There are two examples where the law of $\mathcal{W}$ is known. The first is when all degrees in the graph are equal to some $r > 2$, and we obtain the r-regular graph. In this case, $\mathbb{E}[D] = r, \nu = r - 1$, and $\mathcal{W} = 1$ a.s. In particular, $\mathbb{P}(H_n < \infty) = 1 + o(1)$. Therefore,

$$\mathbb{P}(R_a > k) = \exp\{-\frac{r}{r-2}(r-1)^{a+k}\}, \tag{5.5.12}$$

and H_n is asymptotically equal to $\log_{r-1} n$. The second example is when g is the probability mass function of a geometric random variable, in which case the branching process with offspring g conditioned to be positive converges to an exponential random variable with parameter 1. This example corresponds to

$$g_j = p(1-p)^{j-1}, \quad \text{so that} \quad p_j = \frac{1}{jc_p}p(1-p)^{j-2}, \quad \forall j \geq 1, \tag{5.5.13}$$

and c_p is a normalization constant. For $p > \frac{1}{2}$, the law of $\mathcal{W}$ has the same law as the sum of D_1 copies of a random variable $\mathcal{Y}$, where $\mathcal{Y} = 0$ with probability $\frac{1-p}{p}$ and equal to an exponential random variable with parameter 1 with probability $\frac{2p-1}{p}$. Even in this simple case, the computation of the exact law of R_a is non-trivial.

Exercise 5.16 (Convergence along subsequences [66])**.** *Fix an integer n_1. Prove that, under the assumptions in Theorem 5.28, and conditionally on $H_n < \infty$, along the subsequence $n_k = \lfloor n_1\nu^{k-1} \rfloor$, the sequence of random variables $H_{n_k} - \lfloor \log_\nu n_k \rfloor$ converges in distribution to $R_{a_{n_1}}$ as $k \to \infty$.*

Exercise 5.17 (Tightness of the hopcount [66])**.** *Prove that, under the assumptions in Theorem 5.28,*

(i) *with probability* $1-o(1)$ *and conditionally on* $H_n < \infty$, *the random variable* H_n *is in between* $(1 \pm \varepsilon) \log_\nu n$ *for any* $\varepsilon > 0$;

(ii) *conditionally on* $H_n < \infty$, *the random variables* $H_n - \log_\nu n$ *form a tight sequence, i.e.,*

$$\lim_{K\to\infty} \limsup_{n\to\infty} \mathbb{P}\big(|H_n - \log_\nu n| \le K \big| H_n < \infty\big) = 1. \tag{5.5.14}$$

As a consequence, prove that the same result applies to a uniform random graph with degrees $(d_i)_{i\in[n]}$. *Hint: Make use of [I, Theorem 7.19].*

Fluctuation of distances for infinite-variance degrees. We next study the fluctuations of typical distances in $\mathrm{CM}_n(\boldsymbol{d})$ in the setting where the degrees are i.i.d. and satisfy that there exist $\tau \in (2,3)$, $\gamma \in [0,1)$ and $C < \infty$ such that

$$x^{-\tau+1-C(\log x)^{\gamma-1}} \le 1 - F(x) \le x^{-\tau+1+C(\log x)^{\gamma-1}}, \qquad \text{for large } x. \tag{5.5.15}$$

The condition in (5.5.15) is such that the results in Theorem 5.25 apply. Then, we can identify the fluctuations of the typical graph distance in $\mathrm{CM}_n(\boldsymbol{d})$ as follows:

Theorem 5.29 (Fluctuations graph distance $\mathrm{CM}_n(\boldsymbol{d})$ for infinite variance degrees)**.** *Let* $(d_i)_{i\in[n]}$ *be a sequence of i.i.d. copies of a random variable* D. *Fix* $\tau \in (2,3)$ *and assume that* (5.5.15) *holds. Then,* $\mathrm{CM}_n(\boldsymbol{d})$ *satisfies that there exist random variables* $(R_a)_{a\in(-1,0]}$ *such that, as* $n \to \infty$,

$$\mathbb{P}\Big(H_n = 2\Big\lfloor \frac{\log\log n}{|\log(\tau-2)|}\Big\rfloor + l \,\Big|\, H_n < \infty\Big) = \mathbb{P}(R_{a_n} = l) + o(1), \qquad l \in \mathbb{Z}, \tag{5.5.16}$$

where $a_n = \lfloor \frac{\log\log n}{|\log(\tau-2)|}\rfloor - \frac{\log\log n}{|\log(\tau-2)|} \in (-1,0]$. *Here, the random variables* $(R_a)_{a\in(-1,0]}$ *are given by*

$$\mathbb{P}(R_a > l) = \mathbb{P}\Big(\min_{s\in\mathbb{Z}} \big[(\tau-2)^{-s} Y^{(1)} + (\tau-2)^{s-c_l} Y^{(2)}\big] \le (\tau-2)^{\lceil l/2\rceil + a} \big| Y^{(1)} Y^{(2)} > 0\Big),$$

where $c_l = 1$ *if* l *is even, and zero otherwise, and* $Y^{(1)}, Y^{(2)}$ *are two independent copies of the limit random variable in Theorem 5.25.*

In words, Theorem 5.3 states that for $\tau \in (2,3)$, the graph distance H_n between two randomly chosen connected vertices grows proportional to log log of the size of the graph, and that the fluctuations around this mean remain uniformly bounded in n.

Simulations indicating the properties of the typical graph distance for $\tau \in (2,3)$ can bee seen in Figure 5.2. In it, the distances of the AS-graph in Figure 1.2, and these distances are compared to the ones in $\mathrm{CM}_n(\boldsymbol{d})$ where n is equal to the number of AS and the best approximation to the exponent of the power-law for the degree sequence of the AS-graph, which is $\tau = 2.25$.

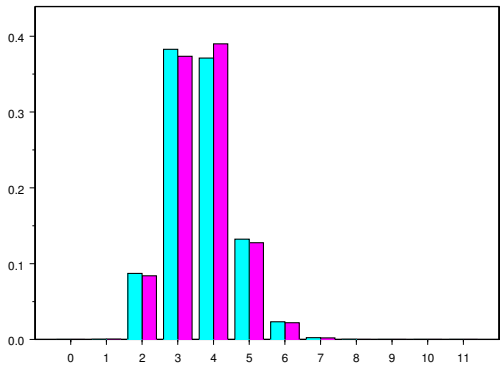

Figure 5.2: Number of AS traversed in hopcount data (blue) compared to the model (purple) with $\tau = 2.25, n = 10,940$.

5.6 Notes and discussion

Notes on Section 5.1. Distances in the configuration model were first obtained in a non-rigorous way in [89, 90]. Theorem 5.2 is [56, Theorem 4.1]. A $\log n$ lower bound on the diameter is also proved in [70]. Theorem 5.28 is proved in [66]. Theorem 5.29 is proved in [69]. The proof of Theorem 5.14 is close in spirit to the analysis in [98], the only difference being that we have simplified the argument slightly.

Notes on Section 5.2. Proposition 5.9 is adapted from [77, Lemma 5.1]. The path counting techniques used in Section 5.2 are novel. Comparisons to branching processes appear in many papers on the configuration model (see, in particular, [17, 66, 70]). We have strived for a construction that is most transparent and complete.

Notes on Section 5.3. Theorem 5.19 is novel, as far as we are aware.

Notes on Section 5.4. Theorem 5.25 is proved in [41]. A related result, under stronger conditions, appeared in [40]. Branching processes with infinite mean have attracted considerable attention, see e.g., [101, 102] and the references therein. There is a balance between the generality of the results and the conditions on the offspring distribution, and in our opinion Theorem 5.25 strikes a nice balance in that the result is relatively simple and the conditions fairly general.

Notes on Section 4.3. Theorem 4.16 is improved to the case where $\mathbb{E}[D_n^3] \to \mathbb{E}[D^3]$ in [71]. Interestingly, the behavior is *different* when [I, Condition 7.7(a)-(b)] hold, but $\mathbb{E}[D_n^3] \to \infty$ sufficiently fast. Also this case is studied in in [71].

Theorem 4.17 is proved in [80]. For related (but much weaker) results in the case of fixed degrees satisfying an assumption as in [I, Condition 7.7], see [63].

Theorem 5.27 is proved in [55]. The explicit identification of $\mathbb{P}(H_n = 2)$ is novel. One might argue that including degrees larger than $n-1$ is artificial in a network with n vertices. In fact, in many real networks, the degree is bounded by a physical constant. Therefore, in [55], also the case where the degrees are conditioned to be smaller than n^α is considered, where α is an arbitrary positive number. Of course, we cannot condition on the degrees to be at most M, where M is fixed and independent on n, since in this case, the degrees are uniformly bounded, and this case is treated in [55] as well. Therefore, [55] considers cases where the degrees are conditioned to be at most a given power of n. In this setting, it turns out that the average distance is equal to $k+3$ with high probability, whenever $\alpha \in (1/(\tau+k), 1/(\tau+k-1))$.

Chapter 6
Preferential attachment revisited

Abstract

In this chapter, we further investigate preferential attachment models. In Section 6.1 we start by discussing an important tool in this chapter: exchangeable random variables and their distribution described in De Finetti's Theorem. We apply these results to Polya urn schemes, which, in turn, we use to describe the distribution of the degrees in preferential attachment models.

Organization of this chapter. In Section 6.2 we investigate the connectivity of $\mathrm{PA}_t(m,\delta)$.

Throughout this chapter, we work with the preferential attachment model defined in [I, Section 8.2] and denoted by $(\mathrm{PA}_t(m,\delta))_{t\geq 1}$, unless stated otherwise. We recall that this model starts with a single vertex with m self-loops at time $t=1$ and at each time a vertex is added with m edges which are attached to the vertices in the graph with probabilities given in [I, (8.2.1)] for $m=1$, and as described on [I, page 209] for $m\geq 2$. This model can also be obtained by identifying blocks of m vertices in $(\mathrm{PA}_{1,\delta/m}(t))_{t\geq 1}$. We sometimes also discuss other variants of the model, such as $(\mathrm{PA}^{(b)}_{m,\delta}(t))_{t\geq 1}$, in which the $m=1$ model does not have any self-loops.

6.1 Exchangeable random variables and Polya urn schemes

In this section, we discuss the distribution of infinite sequences of exchangeable random variables and their applications to Polya urn schemes. We start by discussing De Finetti's Theorem.

De Finetti's Theorem for infinite sequences of exchangeable random variables. A sequence of random variables $(X_i)_{i\geq 1}$ is called *exchangeable* when, for every $n\geq 1$, the distribution of $(X_i)_{i=1}^n$ is the same as the one of $(X_{\sigma(i)})_{i=1}^n$ for any permutation $\sigma\colon [n]\to[n]$. Clearly, when a sequence of random variables is i.i.d., then it is also exchangeable. Remarkably, however, the distribution of an infinite sequence of random variables is always a *mixture* of independent and identically distributed random variables. This is the content of De Finetti's Theorem, which we state and prove here in the case where $(X_i)_{i\geq 1}$ are indicator variables:

Theorem 6.1 (De Finetti's Theorem)**.** *Let $(X_i)_{i\geq 1}$ be an infinite sequence of exchangeable random variables, and assume that $X_i\in\{0,1\}$. Then there exists a*

random variable U with $\mathbb{P}(U \in [0,1]) = 1$ such that, for all $1 \leq k \leq n$

$$\mathbb{P}(X_1 = \cdots = X_k = 1, X_{k+1} = \cdots = X_n = 0) = \mathbb{E}[U^k(1-U)^{n-k}]. \tag{6.1.1}$$

De Finetti's Theorem (Theorem 6.1) states that an infinite exchangeable sequence of indicators has the same distribution as an independent Bernoulli sequence with a *random* success probability U. Thus, the different elements of the sequence are *not* independent, but their dependence enters only through the random success probability U.

In the proof we will see that Theorem 6.1 in fact holds more generally, for example, when X_i takes on at most a *finite* number of values. Since we will use Theorem 6.1 only for indicator variables, we refrain from stating this version.

Define S_n to be the number of ones in $(X_i)_{i=1}^n$, i.e.,

$$S_n = \sum_{k=1}^{n} X_k. \tag{6.1.2}$$

Then Theorem 6.1 is equivalent to the statement that

$$\mathbb{P}(S_n = k) = \mathbb{E}\Big[\mathbb{P}\big(\mathsf{Bin}(n,U) = k\big)\Big]. \tag{6.1.3}$$

Equation (6.1.3) also allows us to compute the distribution of U. Indeed, when we would have that

$$\lim_{n\to\infty} n\mathbb{P}(S_n = \lceil un \rceil) = f(u), \tag{6.1.4}$$

where f is a density, then (6.1.3) implies that f is in fact the density of the random variable U. This will be useful in applications of De Finetti's Theorem (Theorem 6.1). Furthermore, $S_n/n \xrightarrow{a.s.} U$ by the strong law of large numbers applied to the conditional law given U.

Exercise 6.1 (The number of ones in $(X_i)_{i=1}^n$)**.** *Prove* (6.1.3).

Proof of Theorem 6.1. The proof makes use of Helly's Theorem, which states that any sequence of *bounded* random variables has a weakly converging subsequence. We fix $m \geq n$ and condition on S_m to write

$$\begin{aligned}\mathbb{P}(X_1 = \cdots = X_k = 1, X_{k+1} = \cdots = X_n = 0)& \\ = \sum_{j=k}^{m} \mathbb{P}\big(X_1 = \cdots = X_k = 1, X_{k+1} = \cdots = X_n = 0 \big| S_m = j\big)\mathbb{P}(S_m = j).&\end{aligned} \tag{6.1.5}$$

By exchangeability and conditionally on $S_m = j$, each sequence $(X_i)_{i=1}^m$ containing precisely j ones is equally likely. Since there are precisely $\binom{m}{j}$ such sequences, and precisely $\binom{n}{k}\binom{m-n}{j-k}$ of them start with k ones and $n-k$ zeros, we obtain

$$\mathbb{P}\big(X_1 = \cdots = X_k = 1, X_{k+1} = \cdots = X_n = 0 \big| S_m = j\big) = \frac{\binom{n}{k}\binom{m-n}{j-k}}{\binom{m}{j}}. \tag{6.1.6}$$

We therefore arrive at

$$\begin{aligned}&\mathbb{P}(X_1 = \cdots = X_k = 1, X_{k+1} = \cdots = X_n = 0)\\ &\quad = \sum_{j=k}^{m} \frac{j \cdots (j-k+1) \cdot (m-j) \cdots (m-j-(n-k)+1)}{m \cdots (m-n+1)} \mathbb{P}(S_m = j). \quad (6.1.7)\end{aligned}$$

When $m \to \infty$ and for $k \leq n$ fixed,

$$\frac{j \cdots (j-k+1) \cdot (m-j) \cdots (m-j-(n-k)+1)}{m \cdots (m-n+1)} = \Big(\frac{j}{m}\Big)^k \Big(1 - \frac{j}{m}\Big)^{n-k} + o(1). \quad (6.1.8)$$

Recall that $S_m = j$, so that

$$\mathbb{P}(X_1 = \cdots = X_k = 1, X_{k+1} = \cdots = X_n = 0) = \lim_{m \to \infty} \mathbb{E}\big[Y_m^k (1 - Y_m)^{n-k}\big], \quad (6.1.9)$$

where $Y_m = S_m/m$. Note that it is here that we make use of the fact that $(X_i)_{i \geq 1}$ is an *infinite* exchangeable sequence of random variables.

We have that $0 \leq Y_m \leq 1$ since $0 \leq S_m \leq m$, so that the sequence of random variables $(Y_m)_{m \geq 1}$ is a bounded sequence. By Helly's Theorem, it contains a weakly converging subsequence, i.e., there exists a $(Y_{m_l})_{l \geq 1}$ with $\lim_{l \to \infty} m_l = \infty$ and a random variable U such that $Y_{m_l} \xrightarrow{d} U$. Since the random variable $Y_m^k (1 - Y_m)^{n-k}$ is uniformly bounded for each k, n, Lebegues Dominated Convergence Theorem (Theorem A.17) gives that

$$\lim_{m \to \infty} \mathbb{E}\big[Y_m^k (1 - Y_m)^{n-k}\big] = \lim_{l \to \infty} \mathbb{E}\big[Y_{m_l}^k (1 - Y_{m_l})^{n-k}\big] = \mathbb{E}\big[U^k (1 - U)^{n-k}\big] \quad (6.1.10)$$

This completes the proof. □

De Finetti's Theorem implies that when X_k and X_n are coordinates of an infinite exchangeable sequence of indicators, then they are *positively correlated*:

Exercise 6.2 (Positive correlation of exchangeable random variables)**.** *Let $(X_i)_{i \geq 1}$ be an infinite sequence of exchangeable random variables. Prove that*

$$\mathbb{P}(X_k = X_n = 1) \geq \mathbb{P}(X_k = 1)\mathbb{P}(X_n = 1). \quad (6.1.11)$$

Prove that equality holds if and only if there exists a p such that $\mathbb{P}(U = p) = 1$.

In the proof of De Finetti's Theorem, it is imperative that the sequence $(X_i)_{i \geq 1}$ is infinite. This is not mere a technicality of the proof. Rather, there are finite exchangeable sequences of random variables for which the equality (6.1.1) does *not* hold. Indeed, take an urn filled with b blue and r red balls and draw balls successively without replacement. Let X_i denote the indicator that the ith ball drawn is blue. Then, clearly, the sequence $(X_i)_{i=1}^{r+b}$ is exchangeable. However,

$$\mathbb{P}(X_1 = X_2 = 1) = \frac{b(b-1)}{(b+r)(b+r-1)} < \big(\frac{b}{b+r}\big)^2 = \mathbb{P}(X_1 = 1)\mathbb{P}(X_2 = 1), \quad (6.1.12)$$

so that X_1 and X_2 are *negatively* correlated. As a result, (6.1.1) fails.

Polya urn schemes. An important application of De Finetti's Theorem (Theorem 6.1) arises in so-called Polya 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 $W_b, W_r \colon \mathbb{N} \to (0, \infty)$ be two weight functions. Then, at time $n+1$, the probability of drawing a blue ball, conditionally on the number B_n of blue balls at time n, is proportional to the weight of the blue balls at time n, i.e., the conditional probability of drawing a blue ball is equal to

$$\frac{W_b(B_n)}{W_b(B_n) + W_r(R_n)}. \tag{6.1.13}$$

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

$$W_b(k) = a_b + k, \qquad W_r(k) = a_r + k, \tag{6.1.14}$$

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

Theorem 6.2 (Limit theorem for linear Polya urn schemes)**.** *Let $((B_n, R_n))_{n=1}^{\infty}$ be a Polya urn scheme with linear weight functions W_b and W_r as in (6.1.14). Then, as $n \to \infty$,*

$$\frac{B_n}{B_n + R_n} \xrightarrow{a.s.} U, \tag{6.1.15}$$

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

$$\mathbb{P}(B_n = B_0 + k) = \mathbb{E}\Big[\mathbb{P}\big(\mathsf{Bin}(n, U) = k\big)\Big]. \tag{6.1.16}$$

Proof of Theorem 6.2. Let X_n denote the indicator that the n^{th} ball drawn is blue. We first show that $(X_n)_{n \geq 1}$ is an infinite exchangeable sequence. For this, we note that

$$B_n = b_0 + \sum_{j=1}^{n} X_j, \qquad R_n = r_0 + \sum_{j=1}^{n} (1 - X_j) = r_0 - b_0 + n - B_n. \tag{6.1.17}$$

Now, for any sequence $(x_t)_{t=1}^n$,

$$\mathbb{P}\big((X_t)_{t=1}^n = (x_t)_{t=1}^n\big) = \prod_{t=1}^{n} \frac{W_b(b_t)^{x_t} W_r(r_t)^{1-x_t}}{W_b(b_t) + W_r(r_t)}, \tag{6.1.18}$$

where $b_t = b_0 + \sum_{j=1}^{t} x_j$ and $r_t = R_0 - B_0 + t - b_t$. Denote $k = \sum_{t=1}^{n} x_t$. Then, by (6.1.14) and (6.1.17),

$$\prod_{t=1}^{n} (W_b(b_t) + W_r(r_t)) = \prod_{t=0}^{n-1} (b_0 + r_0 + a_b + a_r + t), \tag{6.1.19}$$

while

$$\prod_{t=1}^{n} W_b(b_t)^{x_t} = \prod_{m=0}^{k-1}(b_0 + a_b + m), \qquad \prod_{t=1}^{n} W_r(r_t)^{1-x_t} = \prod_{j=0}^{n-k-1}(r_0 + a_r + j). \quad (6.1.20)$$

Thus, we arrive at

$$\mathbb{P}\big((X_t)_{t=1}^n = (x_t)_{t=1}^n\big) = \frac{\prod_{m=0}^{k-1}(b+m)\prod_{j=0}^{n-k-1}(r+j)}{\prod_{t=0}^{n-1}(b+r+t)}, \quad (6.1.21)$$

where $b = b_0 + a_b$ and $r = r_0 + a_r$. In particular, (6.1.21) does not depend on the *order* in which the elements of $(x_t)_{t=1}^n$ appear, so that the sequence $(X_n)_{n\geq 1}$ is an infinite exchangeable sequence. Thus, by De Finetti's Theorem (Theorem 6.1), the sequence $(X_n)_{n\geq 1}$ is a mixture of Bernoulli random variables with a random success probability U, and we are left to compute the distribution of U. We also observe that the distribution of depends only on b_0, r_0, a_b, a_r through $b = b_0 + a_b$ and $r = r_0 + a_r$.

We next verify (6.1.4). For fixed $0 \leq k \leq n$, there are $\binom{n}{k}$ sequences of k ones and $n-k$ zeros. Each sequence has the same probability given by (6.1.21). Thus,

$$\begin{aligned}\mathbb{P}(S_n = k) &= \binom{n}{k}\frac{\prod_{m=0}^{k-1}(b+m)\prod_{j=0}^{n-k-1}(r+j)}{\prod_{t=0}^{n-1}(b+r+t)} \\ &= \frac{\Gamma(n+1)}{\Gamma(k+1)\Gamma(n-k+1)} \times \frac{\Gamma(k+b)}{\Gamma(b)} \times \frac{\Gamma(n-k+r)}{\Gamma(r)} \times \frac{\Gamma(b+r)}{\Gamma(n+b+r)} \\ &= \frac{\Gamma(b+r)}{\Gamma(r)\Gamma(b)} \times \frac{\Gamma(k+b)}{\Gamma(k+1)} \times \frac{\Gamma(n-k+r)}{\Gamma(n-k+1)} \times \frac{\Gamma(n+1)}{\Gamma(n+b+r)}. \end{aligned} \quad (6.1.22)$$

For k and $n-k$ large, by (8.3.8),

$$\mathbb{P}(S_n = k) = \frac{\Gamma(b+r)}{\Gamma(r)\Gamma(b)}\frac{k^{b-1}(n-k)^{r-1}}{n^{b+r-1}}(1+o(1)). \quad (6.1.23)$$

Taking $k = \lceil un \rceil$ (recall [I, (6.1.4)])

$$\lim_{n\to\infty} n\mathbb{P}(S_n = \lceil un \rceil) = \frac{\Gamma(b+r)}{\Gamma(r)\Gamma(b)}u^{b-1}(1-u)^{r-1}, \quad (6.1.24)$$

which is the density of a Beta-distribution with parameters b and r. □

Applications to scale-free trees. We close this section by discussing applications of Polya urn schemes to scale-free trees. We start at time $t = 2$ with two vertices of which vertex 1 has degree d_1 and vertex 2 has degree d_2. After this, we successively attach vertices to older vertices with probability proportional to the degree plus $\delta > -1$. We do not allow for self-loops, so that indeed we obtain a tree. This is a generalization of $(\mathrm{PA}_t(1,\delta))_{t=2}^{\infty}$, in which we are are more flexible in choosing the initial graph.

We now decompose the growing tree in two trees. For $i = 1, 2$, we let $T_i(t)$ be the tree of vertices which are closer to i than to $3 - i$. Thus, the tree $T_2(t)$ consists of those vertices for which the path in the tree from the vertex to the root passes through vertex 2, and $T_1(t)$ consists of the remainder of the scale-free tree. Let $S_i(t) = |T_i(t)|$ denote the number of vertices in $T_i(t)$. Clearly, $S_1(t) + S_2(t) = t$, which is the total number of vertices of the tree at time t.

Theorem 6.3 (Tree decomposition for scale-free trees)**.** *As $t \to \infty$,*

$$\frac{S_1(t)}{S_1(t) + S_2(t)} \xrightarrow{a.s.} U, \tag{6.1.25}$$

where U has a Beta-distribution with parameters $a = \frac{d_1+\delta}{2+\delta}$ and $b = \frac{d_2+\delta}{2+\delta}$, and

$$\mathbb{P}(S_1(t) = k) = \mathbb{E}\Big[\mathbb{P}\big(\mathsf{Bin}(t-1, U) = k-1\big)\Big]. \tag{6.1.26}$$

By Theorem 6.3, we can decompose a scale-free tree into two disjoint scale-free trees each of which contains a positive proportion of the vertices that converges almost surely to a Beta-distribution with parameters $a = \frac{d_1+\delta}{2+\delta}$ and $b = \frac{d_2+\delta}{2+\delta}$.

Proof of Theorem 6.3. The evolution of $(S_1(t))_{t\geq 2}$ can be viewed as a Polya urn scheme. Indeed, when $S_1(t) = s_1(t)$, then the probability of attaching the $(t+1)$st vertex to $T_1(t)$ is equal to

$$\frac{(2s_1(t) + d_1 - 2) + \delta s_1(t)}{(2s_1(t) + d_1 - 2) + \delta s_1(t) + 2(s_2(t) + d_2) + \delta s_2(t)}, \tag{6.1.27}$$

since the number of vertices in $T_i(t)$ equals $S_i(t)$, while the total degree of $T_i(t)$ equals $(2S_i(t) + d_i - 2)$. We can rewrite this as

$$\frac{s_1(t) + \frac{d_1-2}{2+\delta}}{s_1(t) + s_2(t) + \frac{d_1+d_2-4}{2+\delta}}, \tag{6.1.28}$$

which is equal to (6.1.13) in the case (6.1.14) when $r_0 = b_0 = 1$ and $a_b = \frac{d_1-2}{2+\delta}, a_r = \frac{d_2-2}{2+\delta}$. Therefore, the proof of Theorem 6.3 follows directly from Theorem 6.2. □

Exercise 6.3 (Uniform recursive trees)**.** *A uniform recursive tree is obtained by starting with a single vertex, and successively attaching the $(n+1)$st vertex to a uniformly chosen vertex in $[n]$. Prove that for uniform recursive trees the tree decomposition described above is such that*

$$\frac{S_1(n)}{S_1(n) + S_2(n)} \xrightarrow{a.s.} U, \tag{6.1.29}$$

where U is uniform on $[0, 1]$. Use this to prove that $\mathbb{P}(S_1(n) = k) = 1/n$ for each $k \in [n]$.

We continue by discussing an application of Polya urn schemes to the relative sizes of the initial degrees. For this, we fix an integer $k \geq 2$, and only regard times $t \geq k$ at which an edge is attached to one of the k initial vertices. We work with $(\mathrm{PA}_t(1,\delta))_{t\geq 1}$, so that we start at time $t = 1$ with one vertices with one self-loop, after which we successively attach vertices to older vertices with probability proportional to the degree plus $\delta > -1$, allowing for self-loops. The main result is as follows:

Theorem 6.4 (Relative degrees in scale-free trees)**.** *For* $(\mathrm{PA}_t(1,\delta))_{t\geq 1}$, *as* $t \to \infty$,

$$\frac{D_k(t)}{D_1(t)+\cdots+D_k(t)} \xrightarrow{a.s.} B_k, \tag{6.1.30}$$

where B_k *has a Beta-distribution with parameters* $a = 1+\delta$ *and* $b = (k-1)(2+\delta)$.

By [I, Theorem 8.13], $B_k = \xi_k/(\xi_1+\cdots+\xi_k)$ (where we also use that these random variables are positive almost surely). Theorem 6.4 allows to identify properties of the law of the limiting degrees.

Proof of Theorem 6.4. Denote the sequence of stopping times $(\tau_k(n))_{n\geq 2k-1}$, by $\tau_k(2k-1) = k-1$, and

$$\tau_k(n) = \inf\{t\colon D_1(t)+\cdots+D_k(t) = n\}, \tag{6.1.31}$$

i.e., $\tau_k(n)$ is the time where the total degree of vertices $1,\ldots,k$ equals n. Since $D_j(t) \xrightarrow{a.s.} \infty$ for every j, $\tau_k(n) < \infty$ for every n. Moreover, since $\lim_{n\to\infty}\tau_k(n) = \infty$,

$$\lim_{t\to\infty}\frac{D_k(t)}{D_1(t)+\cdots+D_k(t)} = \lim_{n\to\infty}\frac{D_k(\tau_k(n))}{D_1(\tau_k(n))+\cdots+D_k(\tau_k(n))} = \lim_{n\to\infty}\frac{D_k(\tau_k(n))}{n}. \tag{6.1.32}$$

Now, the random variables $\big((D_k(\tau_k(n)), D_1(\tau_k(n))+\cdots+D_{k-1}(\tau_k(n)))\big)_{n\geq 2k-1}$ form a Polya urn scheme, with $D_k(\tau_k(2k-1)) = 1$, and $D_1(\tau_k(2k-1))+\cdots+D_{k-1}(\tau_k(2k-1)) = 2k-2$. The edge at time $\tau_k(n)$ is attached to vertex k with probability

$$\frac{D_k(t)+\delta}{n+k\delta}, \tag{6.1.33}$$

which are the probabilities of a Polya urn scheme in the linear weight case in (6.1.14) when $a_b = \delta, a_r = (k-1)\delta$, $b_0 = 1$, $r_0 = 2(k-1)$. Thus, the statement follows from Theorem 6.2. □

Exercise 6.4 (Relative degrees of vertices 1 and 2)**.** *Compute* $\lim_{t\to\infty}\mathbb{P}(D_2(t) \geq xD_1(t))$ *for* $(\mathrm{PA}_t(1,\delta))_{t\geq 1}$.

6.2 Connectivity of preferential attachment models

In this section we investigate the connectivity of $(\mathrm{PA}_t(m,\delta))_{t\geq 1}$. We start by describing the connectivity when $m = 1$, which is special. For $m = 1$, the number of connected components of $\mathrm{PA}_t(1,\delta)$ N_t has distribution given by

$$N_t = I_1 + I_2 + \cdots + I_t, \tag{6.2.1}$$

where I_i is the indicator that the ith edge connects to itself, so that $(I_i)_{i\geq 1}$ are independent indicator variables with

$$\mathbb{P}(I_i = 1) = \frac{1+\delta}{(2+\delta)(i-1)+1+\delta}. \tag{6.2.2}$$

It is not hard to see that this implies that $N_t/\log t$ converges in probability to $(1+\delta)/(2+\delta) < 1$, so that whpthere exists a largest connected component of size at least $t/\log t$. As a result, whp $\mathrm{PA}_t(1,\delta)$ is not connected, but has few connected components which are almost all quite large. We do not elaborate more on the connectivity properties for $m = 1$ and instead leave the asymptotics of the number of connected components as an exercise:

Exercise 6.5 (CLT for number of connected components for $m = 1$)**.** *Show that the number of connected components N_t in $\mathrm{PA}_t(1,\delta)$ satisfies a central limit theorem with equal asymptotic mean and variance given by*

$$\mathbb{E}[N_t] = \frac{1+\delta}{2+\delta}\log t(1+o(1)), \qquad \mathrm{Var}(N_t) = \frac{1+\delta}{2+\delta}\log t(1+o(1)). \tag{6.2.3}$$

For $m \geq 2$ the situation is entirely different since then $\mathrm{PA}_t(m,\delta)$ is *connected* whp:

Theorem 6.5 (Connectivity of $\mathrm{PA}_t(m,\delta)$ for $m \geq 2$)**.** *Fix $m \geq 2$. Then, with high probability, $\mathrm{PA}_t(m,\delta)$ is connected.*

Proof of Theorem 6.5. Again we let N_t denote the number of connected components of $\mathrm{PA}_t(m,\delta)$. We note that, $I_t = N_t - N_{t-1} = 1$ precisely when all m edges of vertex t are attached to vertex t. Thus

$$\mathbb{P}(I_t = 1) = \prod_{e=1}^{m} \frac{2e-1+\delta}{(2m+\delta)t + (2e-1+\delta)}. \tag{6.2.4}$$

For $m \geq 2$

$$\sum_{t=2}^{\infty} \mathbb{P}(I_t = 1) < \infty, \tag{6.2.5}$$

so that, almost surely, $I_t = 1$ only occurs *finitely* often. As a result, $\lim_{t\to\infty} N_t < \infty$ almost surely since $N_t \leq 1 + \sum_{t=2}^{\infty} I_t$. This implies that, for $m \geq 2$, $\mathrm{PA}_t(m,\delta)$ almost surely contains only finitely many connected components. However, $\mathrm{PA}_t(m,\delta)$ has a *positive* probability of being disconnected at a certain time $t \geq 2$ (see Exercise 6.6 below). However, for $m \geq 2$, $I_t = N_t - N_{t-1}$ can also be negative, since the edges of the vertex v_t can be attached to two distinct connected components. We will see tat this hapens with high probability, which explains why $N_t = 1$ whp for t large, as we next show.

We first fix $K \geq 1$ large. Then, with probability converging to 1 as $K \to \infty$, $\sum_{t=K}^{\infty} I_t = 0$. We condition on $\sum_{t=K}^{\infty} I_t = 0$, so that no new connected components are formed after time K, and we are left to prove that for t sufficiently large, the vertices $1,\dots,K$ are whp all connected in $\mathrm{PA}_t(m,\delta)$. This proof proceeds in two

steps. In the first step, we fix $i \in [K]$ and prove that whp the number of vertices that are connected to $i \in [K]$ is large when $t \to \infty$. In the next step, we show that this implies that for t sufficiently large and whp in fact *all* vertices in $[K]$ are connected to i in $\mathrm{PA}_t(m, \delta)$.

To prove a lower bound on the number of vertices attached to vertex $i \in [K]$, we lower bound this number by a Polya urn scheme. We denote $b_0 = 1$ and $r_0 = i$. We recursively define the Polya urn scheme $((B_n, R_n))_{n \geq 1}$ as follows. We draw a ball, where the color of the ball is blue with probability proportional to the number of blue balls. If we draw a blue ball, then we replace it with an extra blue ball. If we draw a red ball, then we replace it with an extra red ball. We claim that the number of vertices $C_i(t)$ in the connected component of vertex i at time t, is bounded below by B_{t-i}. We prove this by induction on $t \geq i$. At time i, the number of vertices connected to vertex i is at least 1 (namely, vertex i itself), and this proves the claim since $B_0 = b_0 = 1$. Tis initializes the induction hypothesis. To advance the induction hypothesis, suppose that $C_i(t) \geq B_{t-i}$ and we will prove that also $C_i(t+1) \geq B_{t+1-i}$. When $C_i(t) \geq B_{t-i}$, the total weight of the vertices connected to vertex i at time t is *at least* $B_{t-i}(2m + \delta)$, while the total weight of all vertices is equal to $t(2m + \delta) + (1 + \delta) \leq (t + 1)(2m + \delta) = (2m + \delta)(B_{t-i} + R_{t-i})$, since $B_{t-i} + R_{t-i} = (t - i) + b_0 + r_0 = t + 1$. Now, we say that we draw a red ball at time $t + 1$ when the first edge of vertex $t + 1$ is attached to a vertex which is in the connected component of vertex i at time t. If this happens, then indeed vertex $t + 1$ *will* be in the connected component of vertex i, and $C_i(t)$ is increased by (at least) one. Thus, this advances the induction hypothesis, and proves the claim.

Now, $B_t/t \xrightarrow{a.s.} U$ where U has a Beta-distribution with parameters $a = 1$ and $b = i$ by Theorem 6.2. Since $C_i(t) \geq B_{t-i}$ and $\mathbb{P}(U = 0) = 0$ whp $C_i(t) \geq \varepsilon(t - i)$ for all t sufficiently large when $\varepsilon > 0$ is small. We conclude that, as $t \to \infty$,

$$\mathbb{P}(\liminf_{t\to\infty} C_i(t)/t \geq \varepsilon \ \ \forall i \in [K]) = 1 - o(1) \tag{6.2.6}$$

as $\varepsilon \downarrow 0$. This completes the first step of the analysis, and shows that the size of the connected components of each of the vertices in $[K]$ is whp at least $\varepsilon > 0$.

For the second step, to show that whp all vertices in $[K]$ are connected to one another, we claim that when $C_i(t)/t \geq \varepsilon$ and $C_j(t)/t \geq \varepsilon$ for $i \neq j$, then whp i is connected to j in $\mathrm{PA}_{2t}(m, \delta)$. Indeed, let $l \in [2t] \setminus [t]$. Then, when the first edge of vertex l is attached to a vertex in the connected component of vertex i, and the second to a vertex in the connected component of vertex j, then i and j are connected to one another in $\mathrm{PA}_{2t}(m, \delta)$. In this case, we say that l is a *t-connector* for vertices i and j. Independently of the attachment of the edges of vertices $t+1, \ldots, l-1$, and conditionally on $\mathrm{PA}_{2t}(m, \delta)$, the probability that l is a t-connector for the vertices i and j is at least

$$\frac{C_i(t)(2m + \delta)}{l(2m + \delta) + 1 + \delta} \frac{C_j(t)(2m + \delta)}{l(2m + \delta) + 2 + \delta} \geq \frac{C_i(t)C_j(t)}{(2t + 1)^2}. \tag{6.2.7}$$

When $C_i(t)/t \geq \varepsilon$ and $C_j(t)/t \geq \varepsilon$, this is at least $\varepsilon^2/16$ independently of all previous connections. Thus, the probability that there is no t-connector in $[2t] \setminus [t]$ for vertices

i and j is at most

$$\left(1-\frac{\varepsilon^2}{16}\right)^t \leq \mathrm{e}^{-\varepsilon^2 t/16}. \tag{6.2.8}$$

This tends to 0 exponentially when $t \to \infty$, so that we obtain that the probability that the probability that there exists $i, j \in [K]$ that are not connected in $\mathrm{PA}_{2t}(m, \delta)$ tends to 0 as $t \to \infty$. This proves that $\mathrm{PA}_{2t}(m, \delta)$ is whp connected for t large, which implies Theorem 6.5. □

Exercise 6.6 (All-time connectivity for $(\mathrm{PA}_t(m, \delta))_{t\geq 1}$)**.** *Fix $m \geq 2$. Compute the probability that $(\mathrm{PA}_t(m, \delta))_{t\geq 1}$ is connected for all times $t \geq 1$, and show that this probability is in $(0, 1)$.*

6.3 Notes and discussion

Notes on Section 6.1. The proof of Theorem 6.1 is adapted from [100]. More recent discussions on exchangeable random variables and their properties can be found in [5] and [94], the latter focussing on random walks with self-interaction, where exchangeability is a crucial tool. There is a lot of work on urn schemes, also in cases where the weight functions are not linear with equal slope, in which case the limits can be seen to obey rather different characteristics. See e.g., [12, Chapter 9].

Notes on Section 6.2.

Chapter 7
Small-world phenomena in preferential attachment models

Abstract

In this chapter, we further investigate preferential attachment models. In Section 7.2 we investigate graph distances in $\mathrm{PA}_t(m,\delta)$.

Throughout this chapter, we work with the preferential attachment model defined in [I, Section 8.2] and denoted by $(\mathrm{PA}_t(m,\delta))_{t\geq 1}$, unless stated otherwise. We recall that this model starts with a single vertex with m self-loops at time $t=1$ and at each time a vertex is added with m edges which are attached to the vertices in the graph with probabilities given in [I, (8.2.1)] for $m=1$, and as described on page [I, 209] for $m\geq 2$. This model can also be obtained by identifying blocks of m vertices in $(\mathrm{PA}_{1,\delta/m}(t))_{t\geq 1}$. We sometimes also discuss other variants of the model, such as $(\mathrm{PA}^{(b)}_{m,\delta}(t))_{t\geq 1}$, in which the $m=1$ model does not have any self-loops.

7.1 Logarithmic distances in preferential attachment trees

In this section, we investigate distances in scale-free trees, arising for $m=1$:

Theorem 7.1 (Typical distance in scale-free trees)**.** *Fix $m=1$ and $\delta>-1$. Then*

$$\frac{H_t}{\log t}\xrightarrow{\mathbb{P}}\frac{2(1+\delta)}{(2+\delta)}. \tag{7.1.1}$$

Theorem 7.2 (Diameter of scale-free trees)**.** *Fix $m=1$ and $\delta>-1$. Let γ be the non-negative solution of*

$$\gamma+(1+\delta)(1+\log\gamma)=0. \tag{7.1.2}$$

Then

$$\frac{\mathrm{diam}(\mathrm{PA}_t(1,\delta))}{\log t}\xrightarrow{\mathbb{P}}\frac{2(1+\delta)}{(2+\delta)\gamma}. \tag{7.1.3}$$

The proof of Theorems 7.1–7.2 rely on the fact that $\mathrm{PA}_t(1,\delta)$ consists of a collection of trees with precisely one self-loop. There is a close analogy between these trees and so-called *uniform recursive trees*. In uniform recursive trees, we grow a tree such that at time 1, we have a unique vertex called the root, with label 1. At time t, we add a vertex and connect it to a uniformly chosen vertex in the tree. See [104] for a survey of recursive trees.

A variant of a uniform recursive tree is the case where the probability that a newly added vertex is attached to a vertex is proportional to the degree of the vertices (and, for the root, the degree of the root plus one). This process is called a *random plane-oriented recursive tree*. For a uniform recursive tree of size t, it is proved by Pittel in [97] that the maximal distance between the root and any other vertex is with high probability equal to $\frac{1}{2\gamma}\log t(1+o(1))$, where γ satisfies (7.1.2) with $\delta = 0$. It is not hard to see that this implies that the maximal graph distance between any two vertices in the uniform recursive tree is equal to $\frac{1}{\gamma}\log t(1+o(1))$. We shall make use of similar arguments below to prove Theorem 7.2.

Exercise 7.1 (Relation θ and γ). *Prove that the solution γ of (7.1.2) satisfies $\gamma < 1$. What does this imply for the diameter and typical distances in scale-free trees?*

Exercise 7.2 (Bound on γ). *Prove that the solution γ of (7.1.2) satisfies $\gamma \in (0, \mathrm{e}^{-1})$.*

In the proof of Theorem 7.2 it will be useful to work with $\mathrm{PA}_t^{(b)}(1,\delta)$ instead of $\mathrm{PA}_t(1,\delta)$, for which the same result holds:

Theorem 7.3 (Distances in of scale-free trees $\mathrm{PA}_t^{(b)}(1,\delta)$). *Fix $m = 1$ and $\delta > -1$, and let γ be the solution of (7.1.2). Then*

$$\frac{H_t}{\log t} \xrightarrow{\mathbb{P}} \frac{2(1+\delta)}{(2+\delta)}. \tag{7.1.4}$$

and

$$\frac{\mathrm{diam}(\mathrm{PA}_t^{(b)}(1,\delta))}{\log t} \xrightarrow{\mathbb{P}} \frac{2(1+\delta)}{(2+\delta)\gamma}. \tag{7.1.5}$$

In order to prove Theorem 7.3, we make use of a result on the *height* of scale-free trees, which is the maximal distance between any of the vertices of the tree to its root. For a tree T, we denote the height of T by $\mathrm{height}(T)$. Further, we let V denote a vertex in $[t]$ chosen uniformly at random, and we let G_t denote the height of V. Then the asymptotics of heights in scale-free trees is as follows:

Theorem 7.4 (Height of scale-free trees). *Fix $m = 1$ and $\delta > -1$, and let γ be the solution of (7.1.2). Then*

$$\frac{G_t}{\log t} \xrightarrow{\mathbb{P}} \frac{(1+\delta)}{(2+\delta)} \tag{7.1.6}$$

and

$$\frac{\mathrm{height}\big(\mathrm{PA}_t^{(b)}(1,\delta)\big)}{\log t} \xrightarrow{a.s.} \frac{(1+\delta)}{(2+\delta)\gamma}. \tag{7.1.7}$$

We start by proving the upper bound in Theorem 7.4. We remark that the almost sure limit of the height in Theorem 7.4 does not depend on the precise starting configuration of the graph $\mathrm{PA}_2^{(b)}(1,\delta)$.

In the proof of the upper bound, we make use of the following result which computes the probability mass function of the distance between vertex v_t and the root v_1. Before stating the result, we need some more notation. We write $t \longrightarrow s$ when in $(\mathrm{PA}_t^{(b)}(1,\delta))_{t\geq 1}$ one of the edges of vertex v_t is connected to vertex v_s. Note that for

this to happen, we need that $t > s$. For $s_1 = t > s_2 > \cdots > s_k = 1$, and denoting $\vec{s}_k = (s_1, s_2, \ldots, s_k)$, we write

$$E_{\vec{s}_k} = \bigcap_{i=1}^{k-1} \{s_i \longrightarrow s_{i+1}\}. \tag{7.1.8}$$

For a configuration of $\mathrm{PA}_t^{(b)}(1,\delta)$, we let $\mathrm{dist}(v_t, v_s)$ denote the unique value of k such that $t = s_1 \longrightarrow s_2 \longrightarrow \cdots \longrightarrow s_{k-1} \longrightarrow s_k = s$. Then the probability mass function of $\mathrm{dist}(v_t, v_s)$ can be identified as follows:

Proposition 7.5 (Distribution of $\mathrm{dist}(v_t, v_s)$ in $\mathrm{PA}_t^{(b)}(1,\delta)$)**.** *Fix $m = 1$ and $\delta > -1$. Then, for all $t > s$,*

$$\mathbb{P}(\mathit{dist}(v_t, v_s) = k) = \Big(\frac{1+\delta}{2+\delta}\Big)^k \frac{\Gamma(t+\frac{1}{2+\delta})\Gamma(s)}{\Gamma(s+\frac{1}{2+\delta})\Gamma(t+1)} \sum_{\vec{s}_k} \prod_{i=1}^{k-1} \frac{1}{s_i}, \tag{7.1.9}$$

where the sum is over ordered vectors $\vec{s}_k = (s_0, \ldots, s_k)$ of length $k+1$ with $s_0 = t$ and $s_k = s$. Further,

$$\mathbb{P}(v_t \longrightarrow v_{s_1} \longrightarrow \cdots \longrightarrow v_{s_{k-1}} \longrightarrow v_s) = \Big(\frac{1+\delta}{2+\delta}\Big)^k \frac{\Gamma(t+\frac{1}{2+\delta})\Gamma(s)}{\Gamma(s+\frac{1}{2+\delta})\Gamma(t+1)} \prod_{i=1}^{k-1} \frac{1}{s_i}, \tag{7.1.10}$$

Proof of Proposition 7.5. Since the path between vertex v_t and v_s is *unique*

$$\mathbb{P}(\mathrm{dist}(v_t, v_s) = k) = \sum_{\vec{s}_k} \mathbb{P}\Big(\bigcap_{i=0}^{k-1} \{s_i \longrightarrow s_{i+1}\}\Big), \tag{7.1.11}$$

where again the sum is over all ordered vectors $\vec{s}_k = (s_0, \ldots, s_k)$ of length $k+1$with $s_0 = t$ and $s_k = s$. Therefore, (7.1.9) follows immediately from (7.1.10).

We claim that the events $\{s_i \longrightarrow s_{i+1}\}$ are *independent*, i.e., for every sequence $\vec{s}_k = (s_0, \ldots, s_k)$

$$\mathbb{P}\Big(\bigcap_{i=0}^{k-1} \{s_i \longrightarrow s_{i+1}\}\Big) = \prod_{i=0}^{k-1} \mathbb{P}(s_i \longrightarrow s_{i+1}). \tag{7.1.12}$$

We prove the independence in (7.1.12) by induction on k. For $k = 0$, there is nothing to prove, and this initializes the induction hypothesis. To advance the induction hypothesis in (7.1.12), we condition on $\mathrm{PA}_{s_1 0-1}^{(b)}(1,\delta)$ to obtain

$$\begin{aligned}\mathbb{P}\Big(\bigcap_{i=0}^{k-1} \{s_i \longrightarrow s_{i+1}\}\Big) &= \mathbb{E}\Big[\mathbb{P}\Big(\bigcap_{i=0}^{k-1} \{s_i \longrightarrow s_{i+1}\} \Big| \mathrm{PA}_{s_1-1}^{(b)}(1,\delta)\Big)\Big] \\ &= \mathbb{E}\Big[\mathbb{1}_{\{\bigcap_{i=1}^{k-1}\{s_i \longrightarrow s_{i+1}\}\}} \mathbb{P}\big(s_0 \longrightarrow s_1 \mid \mathrm{PA}_{s_0-1}^{(b)}(1,\delta)\big)\Big],\end{aligned} \tag{7.1.13}$$

since the event $\bigcap_{i=1}^{l-1}\{s_i \longrightarrow s_{i+1}\}$ is measurable with respect to $\mathrm{PA}^{(b)}_{s_0-1}(1,\delta)$. Furthermore, from (8.2.2),

$$\mathbb{P}\big(s_0 \longrightarrow s_1 \mid \mathrm{PA}^{(b)}_{s_0-1}(1,\delta)\big) = \frac{D_{s_1}(s_0-1)+\delta}{(2+\delta)(s_0-1)}. \tag{7.1.14}$$

In particular,

$$\mathbb{P}\big(s_0 \longrightarrow s_1\big) = \mathbb{E}\Big[\frac{D_{s_1}(s_0-1)+\delta}{(2+\delta)(s_0-1)}\Big]. \tag{7.1.15}$$

Therefore,

$$\begin{aligned}\mathbb{P}\Big(\bigcap_{i=0}^{k-1}\{s_i \longrightarrow s_{i+1}\}\Big) &= \mathbb{E}\Big[\mathbb{1}_{\{\bigcap_{i=1}^{k-1}\{s_i\longrightarrow s_{i+1}\}\}}\frac{D_{s_2}(s_1-1)+\delta}{(2+\delta)(s_1-1)}\Big]\\ &= \mathbb{P}\Big(\bigcap_{i=1}^{k-1}\{s_i \longrightarrow s_{i+1}\}\Big)\mathbb{E}\Big[\frac{D_{s_1}(s_0-1)+\delta}{(2+\delta)(s_0-1)}\Big],\end{aligned} \tag{7.1.16}$$

since the random variable $D_{s_1}(s_0-1)$ only depends on how many edges are connected to s_1 after time s_1, and is thus independent of the event $\bigcap_{i=1}^{l-1}\{s_i \longrightarrow s_{i+1}\}$, which only depends on the attachment of the edges *up to and including* time s_1. We conclude that

$$\mathbb{P}\Big(\bigcap_{i=0}^{k-1}\{s_i \longrightarrow s_{i+1}\}\Big) = \mathbb{P}\big(s_1 \longrightarrow s_2\big)\mathbb{P}\Big(\bigcap_{i=0}^{k-1}\{s_i \longrightarrow s_{i+1}\}\Big). \tag{7.1.17}$$

The claim in (7.1.12) for k follows from the induction hypothesis.

Combining (7.1.11) with (7.1.12) and (7.1.15), we obtain that

$$\mathbb{P}(v_t \longrightarrow v_{s_1} \longrightarrow \cdots \longrightarrow v_{s_{k-1}} \longrightarrow v_s) = \prod_{i=0}^{k-1}\mathbb{E}\Big[\frac{D_{s_{i+1}}(s_i-1)+\delta}{(2+\delta)(s_i-1)}\Big]. \tag{7.1.18}$$

By (8.3.12),

$$\mathbb{E}\Big[\frac{D_i(t)+\delta}{(2+\delta)t}\Big] = (1+\delta)\frac{\Gamma(t+\frac{1}{2+\delta})\Gamma(i)}{(2+\delta)t\Gamma(t)\Gamma(i+\frac{1}{2+\delta})} = \frac{1+\delta}{2+\delta}\frac{\Gamma(t+\frac{1}{2+\delta})\Gamma(i)}{\Gamma(t+1)\Gamma(i+\frac{1}{2+\delta})}, \tag{7.1.19}$$

so that

$$\begin{aligned}\mathbb{P}(v_t \longrightarrow v_{s_1} \longrightarrow \cdots \longrightarrow v_{s_{k-1}} \longrightarrow v_s) &= \Big(\frac{1+\delta}{2+\delta}\Big)^k\prod_{i=0}^{k-1}\frac{\Gamma(s_i+\frac{1}{2+\delta})\Gamma(s_{i+1})}{\Gamma(s_i+1)\Gamma(s_{i+1}+\frac{1}{2+\delta})}\\ &= \Big(\frac{1+\delta}{2+\delta}\Big)^k\frac{\Gamma(s_0+\frac{1}{2+\delta})\Gamma(s_k+1)}{\Gamma(s_k+\frac{1}{2+\delta})\Gamma(s_0)}\prod_{i=1}^{k-1}\frac{1}{s_i}\\ &= \Big(\frac{1+\delta}{2+\delta}\Big)^k\frac{\Gamma(t+\frac{1}{2+\delta})\Gamma(s)}{\Gamma(1+\frac{1}{2+\delta})\Gamma(t+1)}\prod_{i=1}^{k-1}\frac{1}{s_i}.\end{aligned} \tag{7.1.20}$$

This completes the proof of Proposition 7.5. □

Proof of the upper bounds in Theorem 7.4. We first use Proposition 7.3 to prove that, in probability,

$$\limsup_{t\to\infty} \frac{\mathrm{dist}(v_t, v_1)}{\log t} \leq \frac{(1+\delta)}{(2+\delta)\theta}. \tag{7.1.21}$$

and, almost surely,

$$\limsup_{t\to\infty} \frac{\mathrm{dist}(v_t, v_1)}{\log t} \leq \frac{(1+\delta)}{(2+\delta)\gamma}. \tag{7.1.22}$$

We use (7.1.9) and symmetry to obtain

$$\mathbb{P}(\mathrm{dist}(v_t, v_1) = k) = \Big(\frac{1+\delta}{2+\delta}\Big)^k \frac{\Gamma(t+\frac{1}{2+\delta})}{\Gamma(1+\frac{1}{2+\delta})\Gamma(t+1)} \sum_{\vec{t}_{k-1}}^{*} \frac{1}{(k-1)!} \prod_{i=1}^{k-1} \frac{1}{t_i}, \tag{7.1.23}$$

where the sum now is over all vectors $\vec{t}_{k-1} = (t_1, \ldots, t_{k-1})$ with $1 < t_i < t$ with *distinct* coordinates. We can upper bound this sum by leaving out the restriction that the coordinates of $\vec{t}_{k-1}$ are distinct, so that

$$\mathbb{P}(\mathrm{dist}(v_t, v_1) = k) \leq \Big(\frac{1+\delta}{2+\delta}\Big)^k \frac{\Gamma(t+\frac{1}{2+\delta})}{\Gamma(1+\frac{1}{2+\delta})\Gamma(t+1)} \frac{1}{(k-1)!} \Big(\sum_{s=2}^{t-1} \frac{1}{s}\Big)^{k-1}. \tag{7.1.24}$$

Since $x \mapsto 1/x$ is monotonically decreasing

$$\sum_{s=2}^{t-1} \frac{1}{s} \leq \int_1^t \frac{1}{x} dx = \log t. \tag{7.1.25}$$

Also, we use [I, (8.3.8)] to bound, for some constant $C_\delta > 0$,

$$\mathbb{P}(\mathrm{dist}(v_t, v_1) = k) \leq C_\delta t^{-\frac{1+\delta}{2+\delta}} \frac{\Big(\frac{1+\delta}{2+\delta}\log t\Big)^{k-1}}{(k-1)!} = C\mathbb{P}\Big(\mathsf{Poi}\big(\frac{1+\delta}{2+\delta}\log t\big) = k-1\Big). \tag{7.1.26}$$

Now we are ready to prove (7.1.9). We note that V is chosen uniformly in $[t]$, so that

$$\begin{aligned}
\mathbb{P}(G_t = k) &= \frac{1}{t}\sum_{s=1}^{t} \mathbb{P}(\mathrm{dist}(v_s, v_1) = k) \leq \frac{1}{t}\sum_{s=1}^{t} C_\delta s^{-\frac{1+\delta}{2+\delta}} \frac{\Big(\frac{1+\delta}{2+\delta}\log s\Big)^{k-1}}{(k-1)!} \\
&\leq \frac{\Big(\frac{1+\delta}{2+\delta}\log t\Big)^{k-1}}{t(k-1)!} \sum_{s=1}^{t} C_\delta s^{-\frac{1+\delta}{2+\delta}} \\
&\leq C\frac{\Big(\frac{1+\delta}{2+\delta}\log t\Big)^{k-1}}{(k-1)!} t^{-\frac{1+\delta}{2+\delta}} = C\mathbb{P}\Big(\mathsf{Poi}\big(\frac{1+\delta}{2+\delta}\log t\big) = k-1\Big).
\end{aligned} \tag{7.1.27}$$

Therefore,

$$\mathbb{P}(G_t > k) \le C\mathbb{P}\Big(\mathsf{Poi}\big(\frac{1+\delta}{2+\delta}\log t\big) \ge k\Big). \tag{7.1.28}$$

Now we fix $\varepsilon > 0$ and take $k = k_t = \frac{(1+\varepsilon)(1+\delta)}{(2+\delta)}\log t$, to arrive at

$$\mathbb{P}(G_t > k_t) \le C\mathbb{P}\Big(\mathsf{Poi}\big(\frac{1+\delta}{2+\delta}\log t\big) \ge \frac{(1+\varepsilon)(1+\delta)}{(2+\delta)}\log t\Big) = o(1), \tag{7.1.29}$$

by the law of large numbers and for any $\varepsilon > 0$, as required.

We continue to prove (7.1.22). By (7.1.26)

$$\mathbb{P}(\mathrm{dist}(v_t, v_1) > k) \le C_\delta\mathbb{P}\Big(\mathsf{Poi}\big(\frac{1+\delta}{2+\delta}\log t\big) \ge k\Big). \tag{7.1.30}$$

Now take $k = a\log t$ with $a > (1+\delta)/(2+\delta)$, and use the large deviation bounds for Poisson random variables in Exercise 2.17 with $\lambda = (1+\delta)/(2+\delta)$ to obtain that

$$\mathbb{P}(\mathrm{dist}(v_t, v_1) > a\log t) \le C_\delta t^{-[a(\log(a(2+\delta)/(1+\delta))-1)+\frac{1+\delta}{2+\delta}]}. \tag{7.1.31}$$

Let x be the solution of

$$x(\log(x(2+\delta)/(1+\delta)) - 1) + \frac{1+\delta}{2+\delta} = 1, \tag{7.1.32}$$

so that $x = \frac{(1+\delta)}{(2+\delta)\gamma}$. Then, for every $a > x$,

$$\mathbb{P}(\mathrm{dist}(v_t, v_1) > a\log t) = O(t^{-p}), \tag{7.1.33}$$

where

$$p = [a(\log(a(2+\delta)/(1+\delta)) - 1) + (1+\delta)(2+\delta)] > 1. \tag{7.1.34}$$

As a result, by the Borel-Cantelli Lemma, the event $\{\mathrm{dist}(v_t, v_1) > k_t\}$ occurs only finitely often, and we conclude that (7.1.22) holds. □

Proof of the lower bound on G_t in Theorem 7.4. We use (7.1.27) to obtain that

$$\mathbb{P}(G_t \le k) \le C\mathbb{P}\Big(\mathsf{Poi}\big(\frac{1+\delta}{2+\delta}\log t\big) \le k\Big). \tag{7.1.35}$$

Fix $k_t = \frac{(1+\delta)(1-\varepsilon)}{(2+\delta)}\log t$, and note that $\mathbb{P}(G_t \le k_t) = o(1)$ by the law of large numbers. □

To complete the proof of Theorem 7.4, we use the second moment method to prove that $\mathrm{height}\big(\mathrm{PA}_t^{(b)}(1,\delta)\big) \le \frac{(1+\delta)(1-\varepsilon)}{(2+\delta)\gamma}\log t$ has vanishing probability. This is formalized in the following proposition:

Proposition 7.6 (Height of $\mathrm{PA}_t^{(b)}(1,\delta)$ converges in probability)**.** *For every $\varepsilon > 0$ there exists a $\eta = \eta(\varepsilon) > 0$ such that*

$$\mathbb{P}\Big(\mathrm{height}\big(\mathrm{PA}_t^{(b)}(1,\delta)\big) \le \frac{(1+\delta)(1-\varepsilon)}{(2+\delta)\gamma}\log t\Big) \le O(t^{-\eta}). \tag{7.1.36}$$

Proof of lower bound on $\mathrm{height}\big(\mathrm{PA}_t^{(b)}(1,\delta)\big)$ *in Theorem 7.4 subject to Proposition 7.6.* Fix $\alpha > 0$, and take $t_k = t_k(\alpha) = \mathrm{e}^{\alpha k}$. For any $\alpha > 0$, by Proposition 7.6 and the fact that $t_k^{-\eta}$ is summable, almost surely, $\mathrm{height}\big(\mathrm{PA}_{t_k}^{(b)}(1,\delta)\big) \geq \frac{(1+\delta)(1-\varepsilon)}{(2+\delta)\gamma}\log t_k$. This proves the almost sure lower bound on $\mathrm{height}\big(\mathrm{PA}_t^{(b)}(1,\delta)\big)$ along the subsequence $(t_k)_{k\geq 0}$. To extend this to an almost sure lower bound when $t\to\infty$, we use that $t \mapsto \mathrm{height}\big(\mathrm{PA}_t^{(b)}(1,\delta)\big)$ is non-decreasing, so that, for every $t \in [t_{k-1}, t_k]$,

$$\begin{aligned}\mathrm{height}\big(\mathrm{PA}_t^{(b)}(1,\delta)\big) &\geq \mathrm{height}\big(\mathrm{PA}_{t_{k-1}}^{(b)}(1,\delta)\big)\\ &\geq \frac{(1+\delta)(1-\varepsilon)}{(2+\delta)\gamma}\log t_{k-1}\\ &\geq (1-\varepsilon)(1-\alpha)\frac{(1+\delta)}{(2+\delta)\gamma}\log t,\end{aligned} \tag{7.1.37}$$

where the third inequality follows from the almost sure lower bound on $\mathrm{height}\big(\mathrm{PA}_{t_{k-1}}^{(b)}(1,\delta)\big)$. The above bound holds for all $\varepsilon, \alpha > 0$, so that letting $\varepsilon, \alpha \downarrow 0$ proves our claim. □

Proof of Proposition 7.6. We perform a path counting argument. We fix $T \in [t]$ and $k \in \mathbb{N}$. Recall that a path $\pi = (\pi_0, \dots, \pi_k)$ is a sequence of vertices. In this section we assume that $\pi_i > \pi_{i+1}$, since our paths will be part of the scale-free tree $\mathrm{PA}_t^{(b)}(1,\delta)$. We write $\pi \in \mathrm{PA}_t^{(b)}(1,\delta)$ for the event that the edge from π_i is connected to π_{i+1} for all $i = 0, \dots, k-1$. We let

$$N_k(t) = \#\{\pi \subseteq \mathrm{PA}_t^{(b)}(1,\delta)\colon \pi_k \in [T]\} \tag{7.1.38}$$

denote the number of k-step paths in $\mathrm{PA}_t^{(b)}(1,\delta)$ with an endpoint in $[T]$. By Proposition 7.5

$$\begin{aligned}\mathbb{E}[N_k(t)] &= \Big(\frac{1+\delta}{2+\delta}\Big)^k \sum_{s_k=1}^{T}\sum_{s_0=s_k}^{t} \frac{\Gamma(s_0+\frac{1}{2+\delta})\Gamma(s_k)}{\Gamma(s_k+\frac{1}{2+\delta})\Gamma(s_0+1)} \sum_{\vec{s}_k}\prod_{i=1}^{k-1}\frac{1}{s_i}\\ &\geq \Big(\frac{1+\delta}{2+\delta}\Big)^k \sum_{s_k=1}^{T}\frac{\Gamma(s_k)}{\Gamma(s_k+\frac{1}{2+\delta})}\sum_{s_0=s_k}^{t}\frac{\Gamma(s_0+\frac{1}{2+\delta})}{\Gamma(s_0+1)} \sum_{\vec{s}_k}\prod_{i=1}^{k-1}\frac{1}{s_i},\end{aligned} \tag{7.1.39}$$

where again the sum is over all ordered $\vec{s}_k = (s_0, \dots, s_k)$ with $s_k \in [T]$ and $s_0 \in [t]$. We can bound this from below by

$$\mathbb{E}[N_k(t)] \geq \Big(\frac{1+\delta}{2+\delta}\Big)^k \sum_{s_k=1}^{T}\frac{\Gamma(s_k)}{\Gamma(s_k+\frac{1}{2+\delta})}\sum_{s_0=s_k}^{t}\frac{\Gamma(s_0+\frac{1}{2+\delta})}{\Gamma(s_0+1)}\frac{1}{(k-1)!}\sum_{\vec{t}_k}\prod_{i=1}^{k-1}\frac{1}{t_i}, \tag{7.1.40}$$

where now the sum is over all vectors $\vec{t}_k = (t_1, \dots, t_{k-1})$ with distinct coordinates with $t_i \in [s_k+1, s_0-1]$. For fixed s_0, s_k,

$$\sum_{\vec{t}_k}\prod_{i=1}^{k-1}\frac{1}{t_i} \geq \Big(\sum_{s=s_k+k}^{s_0-1}\frac{1}{s}\Big)^{k-1}. \tag{7.1.41}$$

We can lower bound

$$\sum_{s=s_k+k}^{s_0-1} \frac{1}{s} \geq \int_{s_k+k}^{s_0} \frac{1}{x} dx = \log(s_0/(s_k+k)) \geq (1-\varepsilon)\log t \tag{7.1.42}$$

when $s_0 \geq t/2$, $s_k \leq T$ and $\log[2(T+k)] \leq \varepsilon \log t$. Thus, we conclude that

$$\mathbb{E}[N_k(t)] \geq \Big(\frac{1+\delta}{2+\delta}\Big)^k \sum_{s_k=1}^{T} \frac{\Gamma(s_k)}{\Gamma(s_k+\frac{1}{2+\delta})} \sum_{s_0=t/2}^{t} \frac{\Gamma(s_0+\frac{1}{2+\delta})}{\Gamma(s_0+1)} \frac{1}{(k-1)!} [(1-\varepsilon)\log t]^{k-1}. \tag{7.1.43}$$

Using (8.3.8), we therefore arrive at

$$\begin{aligned}
\mathbb{E}[N_k(t)] &\geq (1+o(1))\Big(\frac{1+\delta}{2+\delta}\Big)^k \sum_{s_k=1}^{T} s_k^{-1/(2+\delta)} \sum_{s_0=t/2}^{t} s_0^{-(1+\delta)/(2+\delta)} \frac{1}{(k-1)!} [(1-\varepsilon)\log t]^k \\
&\geq c\Big(\frac{1+\delta}{2+\delta}\Big)^k T^{(1+\delta)/(2+\delta)} t^{1/(2+\delta)} \frac{1}{(k-1)!} [(1-\varepsilon)\log t]^{k-1} \\
&= cT^{(1+\delta)/(2+\delta)} t^{[1+(1-\varepsilon)(1+\delta)]/(2+\delta)} \mathbb{P}\Big(\mathsf{Poi}\big(\frac{1+\delta}{2+\delta}(1-\varepsilon)\log t\big) = k-1\Big).
\end{aligned} \tag{7.1.44}$$

When we take $k = k_t = \frac{(1+\delta)(1-\varepsilon)}{(2+\delta)\gamma} \log t$, then

$$\mathbb{P}\Big(\mathsf{Poi}\big(\frac{1+\delta}{2+\delta}(1-\varepsilon)\log t\big) = k-1\Big) \geq ct^{-(1-\varepsilon)}/\sqrt{\log t}, \tag{7.1.45}$$

so that

$$\mathbb{E}[N_k(t)] \geq T^{(1+\delta)/(2+\delta)} t^{\varepsilon/(2+\delta)+o(1)}. \tag{7.1.46}$$

We next take $T = t^{\varepsilon}$ to arrive at

$$\mathbb{E}[N_k(t)] \geq t^{\varepsilon+o(1)}. \tag{7.1.47}$$

This provides the required lower bound on $\mathbb{E}[N_k(t)]$. We defer the proof of an upper bound on $\mathrm{Var}(N_k(t))$ to Section 7.3.1, where we prove that ther exist $C > 0$ and $\eta = \eta(\varepsilon) > 0$ such that $\mathrm{Var}(N_k(t)) \leq C\mathbb{E}[N_k(t)]^2 t^{-\eta}$ (see Lemma 7.14). As a result, by the Chebychev inequality ([I, Theorem 2.16])

$$\mathbb{P}(N_k(t) = 0) \leq \frac{\mathrm{Var}(N_k(t))}{\mathbb{E}[N_k(t)]^2} \leq Ct^{-\eta}. \tag{7.1.48}$$

Since $\mathrm{height}(\mathrm{PA}_t^{(b)}(1,\delta)) \geq k$ when $N_k(t) \geq 1$, this proves the claim in Proposition 7.6. □

We complete this section by proving Theorems 7.2–7.3:

Proof of Theorems 7.2 and 7.3. We first prove the upper bound on the diameter of $\mathrm{PA}_t^{(b)}(1,\delta)$ in Theorem 7.3, for which we use that

$$\mathrm{diam}(\mathrm{PA}_t^{(b)}(1,\delta)) \leq 2 \cdot \mathrm{height}\big(\mathrm{PA}_t^{(b)}(1,\delta)\big). \tag{7.1.49}$$

Equation (7.1.49) together with the upper bound in Theorem 7.4 imply that

$$\limsup_{t\to\infty} \frac{\mathrm{diam}(\mathrm{PA}_t^{(b)}(1,\delta))}{\log t} \leq \frac{2(1+\delta)}{\gamma(2+\delta)}. \tag{7.1.50}$$

For the lower bound, we use the lower bound on $\mathrm{diam}(\mathrm{PA}_t^{(b)}(1,\delta))$ in Theorem 7.3 and the decomposition of scale-free trees in Theorem 6.3. Theorem 6.3 states that the scale-free tree $\mathrm{PA}_t^{(b)}(1,\delta)$ can be decomposed into two scale-free trees, having a similar distribution as copies $\mathrm{PA}_{S_1(t)}^{(b1)}(1,\delta)$ and $\mathrm{PA}_{t-S_1(t)}^{(b2)}(1,\delta)$, where $(\mathrm{PA}_t^{(b1)}(1,\delta))_{t\geq 1}$ and $(\mathrm{PA}_t^{(b2)}(1,\delta))_{t\geq 1}$ are *independent* scale-free tree processes, and the law of $S_1(t)$ is described in (6.1.26). By this tree decomposition

$$\mathrm{diam}\big(\mathrm{PA}_t^{(b)}(1,\delta)\big) \geq \mathrm{height}\big(\mathrm{PA}_{S_1(t)}^{(b1)}(1,\delta)\big) + \mathrm{height}\big(\mathrm{PA}_{t-S_1(t)}^{(b2)}(1,\delta)\big). \tag{7.1.51}$$

The two trees $(\mathrm{PA}_t^{(b1)}(1,\delta))_{t\geq 1}$ and $(\mathrm{PA}_t^{(b2)}(1,\delta))_{t\geq 1}$ are not *exactly* equal in distribution to $(\mathrm{PA}_t^{(b)}(1,\delta))_{t\geq 1}$, because the initial degree of the starting vertices at time $t=2$ is different. However, the precise almost sure scaling in Theorem 6.3 does not depend in a sensitive way on d_1 and d_2, and also the height of the scale-free tree in Theorem 7.4 does not depend on the starting graphs $\mathrm{PA}_2^{(b1)}(1,\delta)$ and $\mathrm{PA}_2^{(b1)}(1,\delta)$ (see the remark below Theorem 7.4). Since $S_1(t)/t \xrightarrow{a.s.} U$, with U having a Beta-distribution with parameters $a=\frac{3+\delta}{2+\delta}$ and $b=\frac{1+\delta}{2+\delta}$, we obtain that $\mathrm{height}\big(\mathrm{PA}_{S_1(t)}^{(b1)}(1,\delta)\big)/\log t \xrightarrow{a.s.} \frac{(1+\delta)}{(2+\delta)\gamma}$ and $\mathrm{height}\big(\mathrm{PA}_{t-S_1(t)}^{(b2)}(1,\delta)\big)/\log t \xrightarrow{a.s.} \frac{(1+\delta)}{(2+\delta)\gamma}$. Thus, we conclude that

$$\limsup_{t\to\infty} \frac{\mathrm{diam}\big(\mathrm{PA}_t^{(b)}(1,\delta)\big)}{\log t} \geq \frac{2(1+\delta)}{(2+\delta)\gamma}. \tag{7.1.52}$$

Combining (7.1.50) and (7.1.52) proves Theorem 7.3.

To prove Theorem 7.2, we note that the connected components of $\mathrm{PA}_t(1,\delta)$ are similar in distribution to single scale-free tree $\mathrm{PA}_{t_1}^{(b1)}(1,\delta),\ldots,\mathrm{PA}_{t_{N_t}}^{(bN_t)}(1,\delta)$, apart from the initial degree of the root. Here t_i denotes the size of the ith tree at time t, and we recall that N_t denotes the total number of trees at time t. Since $N_t/\log t \xrightarrow{d} (1+\delta)/(2+\delta)$ (recall Exercise 6.5), whp the largest connected component has size at least $\varepsilon t/\log t$. Since

$$\log\big(\varepsilon t/\log t\big) = \log t(1+o(1)), \tag{7.1.53}$$

the result follows along the same lines as in the proof of Theorem 7.3. □

7.2 Small-world effect in preferential attachment models

In the next sections we investigate distances in preferential attachment models for $m \geq 2$. These results are not as complete as those for inhomogeneous random graphs or the configuration model as discussed in Chapter 3 and 4, respectively. We investigate both the diameter as well as typical distances. By Theorem 6.5, $\mathrm{PA}_t(m,\delta)$ is whp connected when $m \geq 2$. Recall that in a connected graph, the typical distance

or hopcount H_t is the graph distance between two vertices chosen uniformly at random from $[t]$. Recall further that the power-law degree exponent for $\mathrm{PA}_t(m,\delta)$ is equal to $\tau = 3 + \delta/m$. Therefore, $\tau > 3$ precisely when $\delta > 0$. For the generalized random graph and the configuration model, we have seen that distaces are logarithmic in the size of the graph when $\tau > 3$, and doubly logarithmic when $\tau \in (2,3)$. We will see that similar behavior is true for $\mathrm{PA}_t(m,\delta)$. We start by analyzing the case where $\delta > 0$.

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 7.7 (A $\log t$ bound for typical distances in PAMs)**.** *Fix $m \geq 1$ and $\delta > 0$. For $\mathrm{PA}_t(m,\delta)$ there exist $0 < a_1 < a_2 < \infty$ such that, as $t \to \infty$,*

$$\mathbb{P}(a_1 \log t \leq H_t \leq a_2 \log t) = 1 - o(1). \tag{7.2.1}$$

Theorem 7.8 (A $\log t$ bound for the diameter in PAMs)**.** *Fix $m \geq 1$ and $\delta > 0$. For $\mathrm{PA}_t(m,\delta)$ there exist $0 < b_1 < b_2 < \infty$ such that, as $t \to \infty$,*

$$\mathbb{P}(b_1 \log t \leq \mathrm{diam}(\mathrm{PA}_{m,\delta}(t)) \leq b_2 \log t) = 1 - o(1). \tag{7.2.2}$$

While we believe that there are constants a and b with $a < b$ such that

$$H_t/\log t \xrightarrow{\mathbb{P}} a, \qquad \mathrm{diam}(\mathrm{PA}_{m,\delta}(t))/\log t, \tag{7.2.3}$$

we have no proof for this fact.

Distances in preferential attachment models with $m \geq 2$ and $\delta = 0$. For $\delta = 0$, $\tau = 3$. For $\mathrm{NR}_n(\boldsymbol{w})$, distances grow as $\log n/\log\log n$ in this case (recall Theorem 3.22). The same turns out to be true for $\mathrm{PA}_t(m,\delta)$:

Theorem 7.9 (Diameter of $\mathrm{PA}_t(m,\delta)$ for $\delta = 0$)**.** *Fix $m \geq 2$ and $\delta = 0$. For $\mathrm{PA}_t(m,\delta)$, as $t \to \infty$,*

$$H_t \frac{\log\log t}{\log t} \xrightarrow{\mathbb{P}} 1, \tag{7.2.4}$$

and

$$\mathrm{diam}(\mathrm{PA}_t(m,\delta)) \frac{\log\log t}{\log t} \xrightarrow{\mathbb{P}} 1. \tag{7.2.5}$$

Theorem 7.9 shows that distances for $\tau = 3$ are similar in $\mathrm{PA}_t(m,\delta)$ as in $\mathrm{NR}_n(\boldsymbol{w})$. Interestingly, for $\mathrm{PA}_t(m,\delta)$ with $\delta = 0$, the diameter and the typical distances are close to being equal. For $\mathrm{NR}_n(\boldsymbol{w})$ and $\mathrm{CM}_n(\boldsymbol{d})$ with power-law exponent $\tau = 3$, this fact is not known.

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 douby logarithmically in the size of the graph:

Theorem 7.10 ($\log\log t$ asymptotics for the diameter for $\delta < 0$)**.** *Fix $m \geq 2$ and assume that $\delta \in (-m, 0)$. For $\mathrm{PA}_t(m, \delta)$, as $t \to \infty$,*

$$\frac{H_t}{\log\log t} \xrightarrow{\mathbb{P}} \frac{4}{|\log(\tau - 2)|}. \tag{7.2.6}$$

Interestingly, the term $4/|\log(\tau - 2)|$ appearing in Theores 7.10 replaces the term $2/|\log(\tau - 2)|$ in Theorems 5.3 and 5.19 for the configuration model $\mathrm{CM}_n(\boldsymbol{d})$ with power-law exponent $\tau \in (2, 3)$. Thus, typical distances are twice as big for $\mathrm{PA}_t(m, \delta)$ compared to $\mathrm{CM}_n(\boldsymbol{d})$ with the same power-law exponent. This can be intuitively explianed as follows. For the configuration model $\mathrm{CM}_n(\boldsymbol{d})$, vertices with high degrees are likely to be directly connected (see e.g. Lemma 5.16). For $\mathrm{PA}_t(m, \delta)$, this is *not* the case. However, vertices with high degrees are likely to be at distance two. This makes distances in $\mathrm{PA}_t(m, \delta)$ about twice as big as those for $\mathrm{CM}_n(\boldsymbol{d})$ with the same degree sequence. This effect is special for $\delta < 0$ and is studied in more detail in the next exercises:

Exercise 7.3 (Early vertices are whp at distance 2 for $\delta < 0$)**.** *Let $\delta \in (-m, 0)$ and $m \geq 2$. Show that for i, j fixed*

$$\lim_{t\to\infty} \mathbb{P}(\mathrm{dist}_{\mathrm{PA}_t(m,\delta)}(v_i, v_j) \leq 2) = 1. \tag{7.2.7}$$

Exercise 7.4 (Early vertices are not at distance 2 when $\delta > 0$)**.** *Let $\delta > 0$ and $m \geq 2$. Show that for i, j fixed*

$$\lim_{t\to\infty} \mathbb{P}(\mathrm{dist}_{\mathrm{PA}_t(m,\delta)}(v_i, v_j) = 2) = 0. \tag{7.2.8}$$

Universality in distances for scale-free graphs. The available results are all consistent with the prediction that distances in preferential attachment models have the same asymptotics as distances in the configuration model with the same degree sequence. This suggest a strong form of *universality*, which is interesting in its own right.

7.3 Small-world effect in PA models: lower bounds

In this section we prove lower bounds on distances in $\mathrm{PA}_t(m, \delta)$ with $m \geq 2$. In Section 7.3.1 we start by proving an upper bound on the probability that a path exists in $\mathrm{PA}_t(m, \delta)$, which is our main tool in this section. After this we prove the lower bounds on distances for $\delta > 0$ in Section 7.3.2, for $\delta = 0$ in Section 7.3.3, and for $\delta < 0$ in Section 7.3.4.

7.3.1 Path counting in preferential attachment models

In this section we study the probability that a certain path is present in $\mathrm{PA}_t(m,\delta)$. Recall that we call a path $\pi = (s_0, s_1, \dots, s_l)$ *self-avoiding* when $s_i \neq s_j$ for all $1 \leq i < j \leq l$. The following proposition studies the probability that a path is present in $\mathrm{PA}_t(m,\delta)$:

Proposition 7.11 (Path counting in $\mathrm{PA}_t(m,\delta)$)**.** *Denote* $\gamma = \frac{m}{2m+\delta}$. *Let* $\pi = (\pi_0, \pi_1, \dots, \pi_l)$ *be a self-avoiding path of length* l *consisting of the* $l+1$ *unordered vertices* $\pi_0, \pi_1, \dots, \pi_l$. *Then*

$$\mathbb{P}(\pi \subseteq \mathrm{PA}_t(m,\delta)) \leq (Cm^2)^l \prod_{i=0}^{l-1} \frac{1}{(\pi_i \wedge \pi_{i+1})^{\gamma} (\pi_i \vee \pi_{i+1})^{1-\gamma}}. \tag{7.3.1}$$

Paths are formed by repeatedly forming edges. When $m = 1$, paths go from later vertices to older vertices. When $m \geq 2$, this monotonicity property of paths is lost, which makes the proof harder. We start by investigating intersections of events that specify which edges are present in $\mathrm{PA}_t(m,\delta)$. We start by introducing some notation. Denote by

$$\{g(t,j) = s\}, \quad 1 \leq j \leq m, \tag{7.3.2}$$

the event that the jth edge of vertex t is attached to the earlier vertex s. For $\mathrm{PA}_t(m,\delta)$, this event means that in $\{\mathrm{PA}_{mt}(1,\delta)\}$ the edge from vertex $m(t-1)+j$ is attached to one of the vertices $m(s-1)+1, \dots, ms$.

It is a direct consequence of the definition of PA-models that the event (7.3.2) increases the preference for vertex s, and hence decreases (in a relative way) the preference for the vertices u, $1 \leq u \leq t$, $u \neq s$. It should be intuitively clear that another way of expressing this effect is to say that, for different $s_1 \neq s_2$, the events $\{g(t_1, j_1) = s_1\}$ and $\{g(t_2, j_2) = s_2\}$ are negatively correlated. We now formalize this result For integer $n_s \geq 1$, we denote by

$$E_s = \bigcap_{i=1}^{n_s} \big\{g(t_i^{(s)}, j_i^{(s)}) = s\big\}, \tag{7.3.3}$$

the event that the j_ith edge of vertex t_i is attached to the earlier vertex s, for $i = 1, \dots, n_s$. We start by proving that the events E_s, for different s, are *negatively correlated* for each choice of $k \geq 1$ and all possible choices of $t_i^{(s)}, j_i^{(s)}$.

Lemma 7.12 (Negative correlation for connection of edges)**.** *For distinct* $s_1, s_2, \dots, s_k$, *both for* $\mathrm{PA}_t(m,\delta)$ *and for* $\mathrm{PA}_t^{(b)}(m,\delta)$,

$$\mathbb{P}\Big(\bigcap_{i=1}^{k} E_{s_i}\Big) \leq \prod_{i=1}^{k} \mathbb{P}(E_{s_i}). \tag{7.3.4}$$

Proof. We only prove the statement for $\mathrm{PA}_t(m,\delta)$, the proof for $\mathrm{PA}_t^{(b)}(m,\delta)$ is identical. We use induction on the largest edge number present in the events $E_{s_1}, \dots, E_{s_k}$. Here, we define the edge number of the event $\{g(t,j) = s\}$ to be $m(t-1)+j$, which

is the order of the edge when we consider the edges as being attached in sequence in $\mathrm{PA}_{mt}(1, \delta/m)$.

The induction hypothesis is that (7.3.4) holds for all k and all choices of $t_i^{(s)}, j_i^{(s)}$ such that $\max_{i,s} m(t_i^{(s)} - 1) + j_i^{(s)} \leq e$, where induction is performed with respect to e.

To initialize the induction, we note that for $e = 1$, the induction hypothesis holds trivially, since $\bigcap_{i=1}^{k} E_{s_i}$ can be empty or consist of exactly one event, and in the latter case there is nothing to prove. This initializes the induction.

To advance the induction, we assume that (7.3.4) holds for all k and all choices of $t_i^{(s)}, j_i^{(s)}$ such that $\max_{i,s} m(t_i^{(s)} - 1) + j_i^{(s)} \leq e - 1$, and we extend it to all k and all choices of $t_i^{(s)}, j_i^{(s)}$ such that $\max_{i,s} m(t_i^{(s)} - 1) + j_i^{(s)} \leq e$. Clearly, for k and $t_i^{(s)}, j_i^{(s)}$ such that $\max_{i,s} m(t_i^{(s)} - 1) + j_i^{(s)} \leq e - 1$, the bound follows from the induction hypothesis, so we may restrict attention to the case that $\max_{i,s} m(t_i^{(s)} - 1) + j_i^{(s)} = e$. We note that there is a unique choice of t, j such that $m(t-1) + j = e$. There are two possibilities: (1) Either there is exactly one choice of s and $t_i^{(s)}, j_i^{(s)}$ such that $t_i^{(s)} = t, j_i^{(s)} = j$, or (2) there are at least *two* of such choices. In the latter case, $\bigcap_{s=1}^{k} E_s = \varnothing$, since the eth edge is connected to a *unique* vertex. Hence, there is nothing to prove.

We are left to investigate the case where there exists a unique s and $t_i^{(s)}, j_i^{(s)}$ such that $t_i^{(s)} = t, j_i^{(s)} = j$. Denote by

$$E'_s = \bigcap_{i=1:(t_i^{(s)}, j_i^{(s)}) \neq (t,j)}^{n_s} \big\{ g(t_i^{(s)}, j_i^{(s)}) = s \big\} \tag{7.3.5}$$

the restriction of E_s to all *other* edges. Then we can write

$$\bigcap_{i=1}^{k} E_{s_i} = \big\{ g(t,j) = s \big\} \cap E'_s \cap \bigcap_{i=1:\, s_i \neq s}^{k} E_{s_i}. \tag{7.3.6}$$

By construction, all the edge numbers of the events in $E'_s \cap \bigcap_{i=1:\, s_i \neq s}^{k} E_{s_i}$ are at most $e - 1$. Thus we obtain

$$\mathbb{P}\Big(\bigcap_{i=1}^{k} E_{s_i} \Big) \leq \mathbb{E}\Big[\mathbb{1}_{E'_s \cap \bigcap_{i=1:\, s_i \neq s}^{k} E_{s_i}} \mathbb{P}_{e-1}(g(t,j) = s) \Big], \tag{7.3.7}$$

where $\mathbb{P}_{e-1}$ denotes the conditional probability given the edge attachments up to the $(e-1)$st edge connection, or, equivalently, given $\mathrm{PA}_{e-1}(1, \delta/m)$, and we have used that the event $E'_s \cap \bigcap_{i=1:\, s_i \neq s}^{k} E_{s_i}$ is measurable with respect to $\mathrm{PA}_{e-1}(1, \delta/m)$.

We compute

$$\mathbb{P}_{e-1}(g(t,j) = s) = \frac{D_s(t-1, j-1) + \delta}{(2m + \delta)(t-1) + (j-1)(2 + \delta/m) + 1 + \delta}, \tag{7.3.8}$$

where we recall that $D_s(t-1, j-1)$ is the degree of vertex s after $j-1$ edges of vertex t have been attached. We wish to use the induction hypothesis. For this, we note that

$$D_s(t-1, j-1) = m + \sum_{(t',j'):\, mt' + j' \leq e-1} \mathbb{1}_{\{g(t',j') = s\}}, \tag{7.3.9}$$

where we recall that $e-1 = m(t-1)+j-1$. Each of the events $\{g(t',j') = s\}$ in (7.3.9) has edge number strictly smaller than e and occurs with a non-negative multiplicative constant. As a result, we may use the induction hypothesis for each of these terms. Thus, we obtain, using also $m+\delta \geq 0$, that,

$$\mathbb{P}\Big(\bigcap_{i=1}^{k} E_{s_i}\Big) \leq \frac{m+\delta}{(2m+\delta)(t-1)+(j-1)(2+\delta/m)+1+\delta}\mathbb{P}(E'_s)\prod_{i=1:\, s_i\neq s}^{k}\mathbb{P}(E_{s_i}) \tag{7.3.10}$$
$$+\sum_{(t',j'):\, mt'+j'\leq e-1}\frac{\mathbb{P}(E'_s\cap\{g(t',j')=s\})}{(2m+\delta)(t-1)+(j-1)(2+\delta/m)+1+\delta}\prod_{i=1:\, s_i\neq s}^{k}\mathbb{P}(E_{s_i}).$$

We use (7.3.9) to recombine the above as

$$\mathbb{P}\Big(\bigcap_{i=1}^{k} E_{s_i}\Big) \leq \mathbb{E}\Big[\mathbb{1}_{E'_s}\frac{D_s(t-1,j-1)+\delta}{(2m+\delta)(t-1)+(j-1)(2+\delta/m)+1+\delta}\Big]\prod_{i=1:s_i\neq s}^{k}\mathbb{P}(E_{s_i}), \tag{7.3.11}$$

and the advancement is completed when we note that

$$\mathbb{E}\Big[\mathbb{1}_{E'_s}\frac{D_s(t-1,j-1)+\delta}{(2m+\delta)(t-1)+(j-1)(2+\delta/m)+1+\delta}\Big] = \mathbb{P}(E_s). \tag{7.3.12}$$

The claim in Lemma 7.12 follows by induction. □

Exercise 7.5 (Negative correlations for $m=1$)**.** *Show that when $m=1$, Lemma 7.12 implies that when $(\pi_0,\dots,\pi_k)$ contains different coordinates as $(\rho_0,\dots,\rho_k)$, then*

$$\mathbb{P}\Big(\bigcap_{i=0}^{k-1}\{\pi_i\longrightarrow\pi_{i+1}\}\cap\bigcap_{i=0}^{k-1}\{\rho_i\longrightarrow\rho_{i+1}\}\Big) \leq \mathbb{P}\Big(\bigcap_{i=0}^{k-1}\{\pi_i\longrightarrow\pi_{i+1}\}\Big)\mathbb{P}\Big(\bigcap_{i=0}^{k-1}\{\rho_i\longrightarrow\rho_{i+1}\}\Big). \tag{7.3.13}$$

We next study the probabilities of the events E_s when $n_s \leq 2$:

Lemma 7.13 (Edge connection events for at most two edges)**.** *Denote $\gamma = \frac{m}{2m+\delta}$. There exist absolute constants $M_1 = M_1(\delta,m), M_2 = M_2(\delta,m)$, such that*
(i) for $m=1$ and any $t>s$,

$$\mathbb{P}(g(t,1)=s) = (1+\delta)\frac{\Gamma(t)\Gamma(s+\frac{1+\delta}{2+\delta})}{\Gamma(t+\frac{1+\delta}{2+\delta})\Gamma(s+1)} \leq \frac{M_1}{t^{1-\gamma}s^{\gamma}}. \tag{7.3.14}$$

Consequently, for each $1\leq j\leq m$ and $t>s$,

$$\mathbb{P}\big(g(t,j)=s\big) \leq \frac{M_1}{t^{1-\gamma}s^{\gamma}}. \tag{7.3.15}$$

(ii) for $m=1$ and any $t_2>t_1>s$,

$$\mathbb{P}\big(g(t_1,1)=s, g(t_2,1)=s\big) = \frac{\Gamma(t_2)\Gamma(t_1+\frac{1}{2+\delta})\Gamma(s+\frac{1+\delta}{2+\delta})}{\Gamma(t_2+\frac{1+\delta}{2+\delta})\Gamma(t_1+1)\Gamma(s+\frac{3+\delta}{2+\delta})}(1+\delta). \tag{7.3.16}$$

Consequently, for any $1 \le j_1, j_2 \le m$ *and* $t_2 > t_1 > s$,

$$\mathbb{P}\Big(g(t_1, j_1) = s, g(t_2, j_2) = s\Big) \le \frac{M_2}{(t_1 t_2)^{1-\gamma} s^{2\gamma}}. \tag{7.3.17}$$

Proof. We only prove (7.3.14) and (7.3.16), (7.3.15) and (7.3.17) follow immediately from [I, (8.3.8)].

Throughout this proof, we abbreviate $g(t) = g(t, 1)$. By the definition of $(\mathrm{PA}_t(m, \delta))_{t\ge 1}$ in terms of $(\mathrm{PA}_t(1, \delta/m))_{t\ge 1}$, this implies the result for general $m \ge 1$, where the factors of m follow from the fact that vertex s in $\mathrm{PA}_t(m, \delta)$ corresponds to vertices $ms, \dots, m(s+1) - 1$ in $\mathrm{PA}_{mt}(1, \delta/m)$, which are all at least ms. Note, in particular, that $g(t, j) = s$ for $m \ge 2$ in $\mathrm{PA}_t(m, \delta)$ is equivalent to $g(m(t-1) + j) \in \{m(s-1) + 1, \dots, ms\}$ in $\mathrm{PA}_{mt}(1, \delta/m)$.

For (7.3.14), we use [I, Theorem 8.1] to compute

$$\begin{aligned}\mathbb{P}\big(g(t) = s\big) &= \mathbb{E}\big[\mathbb{E}[\mathbb{1}_{\{g(t)=s\}} \,|\mathrm{PA}_{t-1}(1, \delta)]\big] = \mathbb{E}\big[\frac{D_s(t-1) + \delta}{(2+\delta)(t-1) + 1 + \delta}\big] \\ &= (1+\delta)\frac{\Gamma(t)\Gamma(s + \frac{1+\delta}{2+\delta})}{\Gamma(t + \frac{1+\delta}{2+\delta})\Gamma(s+1)}.\end{aligned} \tag{7.3.18}$$

Take $t_2 > t_1$. We proceed with the proof of (7.3.16) by computing

$$\begin{aligned}&\mathbb{P}\Big(g(t_1) = s, g(t_2) = s\Big) \qquad (7.3.19)\\ &= \mathbb{E}\Big[\mathbb{P}\big(g(t_1) = s, g(t_2) = s\big|\mathrm{PA}_{t_2-1}(m, \delta)\big)\Big] \\ &= \mathbb{E}\left[\mathbb{1}_{\{g(t_1)=s\}}\left(\frac{D_s(t_2-1) + \delta}{(t_2-1)(2+\delta) + 1 + \delta}\right)\right] \\ &= \frac{1}{(t_2-1)(2+\delta) + 1 + \delta}\frac{\Gamma(t_2)\Gamma(t_1 + \frac{1+\delta}{2+\delta})}{\Gamma(t_2 - 1 + \frac{1+\delta}{2+\delta})\Gamma(t_1+1)}\mathbb{E}\left[\mathbb{1}_{\{g(t_1)=s\}}\left(D_s(t_1) + \delta\right)\right] \\ &= \frac{\Gamma(t_2)\Gamma(t_1 + \frac{1+\delta}{2+\delta})}{\Gamma(t_2 + \frac{1+\delta}{2+\delta})\Gamma(t_1+1)}\mathbb{E}\left[\mathbb{1}_{\{g(t_1)=s\}}\left(D_s(t_1) + \delta\right)\right] \qquad (7.3.20)\end{aligned}$$

where we use the iteration, for $t_1 < u \le t_2 - 1$,

$$\begin{aligned}&\mathbb{E}\left[\mathbb{1}_{\{g(t_1)=s\}}\left(D_s(u) + \delta\right)\right] \qquad (7.3.21)\\ &\quad = \Big(1 + \frac{1}{(2+\delta)(u-1) + 1 + \delta}\Big)\mathbb{E}\left[\mathbb{1}_{\{g(t_1)=s\}}\left(D_s(u-1) + \delta\right)\right].\end{aligned}$$

We are lead to compute $\mathbb{E}\left[\mathbb{1}_{\{g(t_1)=s\}}\left(D_s(t_1) + \delta\right)\right]$. We use recursion to obtain

$$\begin{aligned}&\mathbb{E}\left[\mathbb{1}_{\{g(t_1)=s\}}\left(D_s(t_1) + \delta\right)\big|\mathrm{PA}_{t_1-1}(m, \delta)\right] \qquad (7.3.22)\\ &\quad = \mathbb{E}\left[\mathbb{1}_{\{g(t_1)=s\}}\left(D_s(t_1) - D_s(t_1-1)\right)\big|\mathrm{PA}_{t_1-1}(m, \delta)\right] \\ &\qquad + \mathbb{E}\left[\mathbb{1}_{\{g(t_1)=s\}}\left(D_s(t_1-1) + \delta\right)\big|\mathrm{PA}_{t_1-1}(m, \delta)\right] \\ &\quad = \frac{(D_s(t_1-1) + \delta)(D_s(t_1-1) + 1 + \delta)}{(t_1-1)(2+\delta) + 1 + \delta}.\end{aligned}$$

By [I, Proposition 8.14],

$$\mathbb{E}[(D_s(t)+\delta)(D_s(t)+1+\delta)] = \frac{2}{c_2(t)}\mathbb{E}[Z_{s,2}(t)] = \frac{c_2(s)}{c_2(t)}(2+\delta)(1+\delta). \tag{7.3.23}$$

Recalling that $c_k(j) = \Gamma(j+\frac{1+\delta}{2+\delta})/\Gamma(j+\frac{k+1+\delta}{2+\delta})$, this brings us to

$$\mathbb{E}[(D_s(t)+\delta)(D_s(t)+1+\delta)] = \frac{\Gamma(t+\frac{3+\delta}{2+\delta})\Gamma(s+\frac{1+\delta}{2+\delta})}{\Gamma(t+\frac{1+\delta}{2+\delta})\Gamma(s+\frac{3+\delta}{2+\delta})}(2+\delta)(1+\delta). \tag{7.3.24}$$

where M_4 is a uniform constant. Consequently,

$$\begin{aligned}
&\mathbb{E}\left[\mathbb{1}_{\{g(t_1)=s\}}\left(D_s(t_1)+\delta\right)\right] \qquad (7.3.25)\\
&\quad= \frac{\Gamma(t_1+\frac{1}{2+\delta})\Gamma(s+\frac{1+\delta}{2+\delta})}{[(t_1-1)(2+\delta)+1+\delta]\Gamma(t_1-\frac{1}{2+\delta})\Gamma(s+\frac{3+\delta}{2+\delta})}(2+\delta)(1+\delta)\\
&\quad= \frac{\Gamma(t_1+\frac{1}{2+\delta})\Gamma(s+\frac{1+\delta}{2+\delta})}{\Gamma(t_1+\frac{1+\delta}{2+\delta})\Gamma(s+\frac{3+\delta}{2+\delta})}(1+\delta).
\end{aligned}$$

Combining (7.3.20), (A.2.9) and (7.3.23), we arrive at

$$\begin{aligned}
&\mathbb{P}\big(g(t_1)=s, g(t_2)=s\big) \qquad (7.3.26)\\
&\quad= \frac{\Gamma(t_2)\Gamma(t_1+\frac{1+\delta}{2+\delta})}{\Gamma(t_2+\frac{1+\delta}{2+\delta})\Gamma(t_1+1)} \times \frac{\Gamma(t_1+\frac{1}{2+\delta})\Gamma(s+\frac{1+\delta}{2+\delta})}{\Gamma(t_1+\frac{1+\delta}{2+\delta})\Gamma(s+\frac{3+\delta}{2+\delta})}(1+\delta)\\
&\quad= \frac{\Gamma(t_2)\Gamma(t_1+\frac{1}{2+\delta})\Gamma(s+\frac{1+\delta}{2+\delta})}{\Gamma(t_2+\frac{1+\delta}{2+\delta})\Gamma(t_1+1)\Gamma(s+\frac{3+\delta}{2+\delta})}(1+\delta),
\end{aligned}$$

as required. □

Exercise 7.6 (Extension of (7.3.16) to $\mathrm{PA}_t^{(b)}(1,\delta)$)**.** *Prove that for* $\mathrm{PA}_t^{(b)}(1,\delta)$, (7.3.16) *is replaced with*

$$\mathbb{P}\big(g(t_1)=s, g(t_2)=s\big) = (1+\delta)\frac{\Gamma(t_1-\delta/(2+\delta))\Gamma(t_2-(1+\delta)/(2+\delta))\Gamma(s)}{\Gamma(t_1+1/(2+\delta))\Gamma(t_2)\Gamma(s+2/(2+\delta))}. \tag{7.3.27}$$

With Lemmas 7.12 and 7.13 at hand, we are ready to prove Proposition 7.11:

Proof of Proposition 7.11. Since π is self-avoiding, we can write $\{\pi \subseteq \mathrm{PA}_t(m,\delta)\} = \cap_{i=1}^k E_{s_i}$, where either

$$E_s = \{g(t,j) = s\} \tag{7.3.28}$$

for some $t > s$ and some $1 \le j \le m$, or

$$E_s = \{g(t_1,j_1) = g(t_2,j_2) = s\}, \tag{7.3.29}$$

for some $t_1, t_2 > s$ and some $1 \le j_1, j_2 \le m$. In the first case, by (7.3.14),

$$\mathbb{P}(E_s) = \mathbb{P}\big(g(t,j) = s\big) \le \frac{M_1}{t^{1-\gamma} s^{\gamma}}, \tag{7.3.30}$$

whereas in the second case, according to (7.3.16),

$$\mathbb{P}(E_s) = \mathbb{P}(g(t_1,j_1) = s, g(t_2,j_2) = s) \le \frac{M_2}{(t_1 t_2)^{1-\gamma} s^{2\gamma}} = \frac{M_2}{t_1^{1-\gamma} s^{\gamma} t_2^{1-\gamma} s^{\gamma}}. \tag{7.3.31}$$

In both cases M_i, $i = 1, 2$, is an absolute constant. Lemma 7.12 then yields (7.3.1), where the factor m^{2l} originates from the number of possible choices of $j_i \in [m]$ for $i = 1, \ldots, k$ and the possible s_i that are collapsed to the same vertex. □

Lemma 7.14 (A variance estimate on $N_k(t)$)**.** *Recall the definition of $N_k(t)$ in (7.1.38) and let $T = t^{\varepsilon}$ for some $\varepsilon > 0$. Then there exists constants $C > 0$ and $\eta = \eta(\varepsilon) > 0$ such that*

$$\mathrm{Var}(N_k(t)) \le C \mathbb{E}[N_k(t)]^2 t^{-\eta}. \tag{7.3.32}$$

Proof. By Exercise 7.5, when the path $(\pi_0, \ldots, \pi_k)$ is completely disjoint from $(\rho_0, \ldots, \rho_k)$,

$$\mathbb{P}\Big(\bigcap_{i=0}^{k-1}\{\pi_i \longrightarrow \pi_{i+1}\} \cap \bigcap_{i=0}^{k-1}\{\rho_i \longrightarrow \rho_{i+1}\}\Big) \le \mathbb{P}\Big(\bigcap_{i=0}^{k-1}\{\pi_i \longrightarrow \pi_{i+1}\}\Big)\mathbb{P}\Big(\bigcap_{i=0}^{k-1}\{\rho_i \longrightarrow \rho_{i+1}\}\Big). \tag{7.3.33}$$

Therefore, the indicators of disjoint paths are negatively correlated. As a result, we can bound

$$\mathrm{Var}(N_k(t)) \le \sum_{\pi,\rho:\ \pi \cap \rho \neq \varnothing} \mathbb{P}(\pi, \rho \subseteq \mathrm{PA}_t^{(b)}(1,\delta)). \tag{7.3.34}$$

Since $m = 1$, the paths π, ρ of length k must merge at some point, before moving off to their common end point in $[T]$. When $\rho = \pi$, then we obtain a contribution $\mathbb{E}[N_k(t)]$, so that from now on we assume that $\pi \neq \rho$.

Write $\pi = (\pi_0, \ldots, \pi_k)$ and $\rho = (\rho_0, \ldots, \rho_k)$ and $\pi \neq \rho$. Then there must be an $l \in [k-1]$ such that $\pi_j = \rho_j$ for all $j = l, \ldots, k$. For two fixed paths π and ρ for which $\pi_j = \rho_j$ for all $j = l, \ldots, k$, while $\pi_{l-1} \neq \rho_{l-1}$. By Lemma 7.12, for such paths π, ρ,

$$\begin{aligned}\mathbb{P}(\pi, \rho \subseteq \mathrm{PA}_t^{(b)}(1,\delta)) \le \Big(\prod_{i=1}^{l-1} \mathbb{P}(\pi_{i-1} \longrightarrow \pi_i)\mathbb{P}(\rho_{i-1} \longrightarrow \rho_i)\Big) \\ \times \mathbb{P}(\pi_{l-1}, \rho_{l-1} \longrightarrow \pi_l)\Big(\prod_{j=l+1}^{k} \mathbb{P}(\pi_{j-1} \longrightarrow \pi_j)\Big).\end{aligned} \tag{7.3.35}$$

By (7.1.10) in Proposition 7.5,

$$\prod_{i=1}^{l-1} \mathbb{P}(\pi_{i-1} \longrightarrow \pi_i) = \Big(\frac{1+\delta}{2+\delta}\Big)^{l-1} \frac{\Gamma(\pi_0 + \frac{1}{2+\delta})\Gamma(\pi_{l-1})}{\Gamma(\pi_{l-1} + \frac{1}{2+\delta})\Gamma(\pi_0 + 1)} \prod_{i=1}^{l-2} \frac{1}{\pi_i} \tag{7.3.36}$$

By symmetry, we may assume without loss of generality that $\pi_{l-1} > \rho_{l-1}$. Then, by (7.3.16),

$$\mathbb{P}(\pi_{l-1}, \rho_{l-1} \longrightarrow \pi_l) = (1+\delta)\frac{\Gamma(\rho_{l-1} - \frac{\delta}{2+\delta})\Gamma(\pi_{l-1} - \frac{1+\delta}{2+\delta})\Gamma(\pi_l)}{\Gamma(\rho_{l-1} + \frac{1}{2+\delta})\Gamma(\pi_{l-1})\Gamma(\pi_l + \frac{2}{2+\delta})}. \tag{7.3.37}$$

As a result,

$$\begin{aligned}\mathbb{P}(\pi, \rho \subseteq \mathrm{PA}_t^{(b)}(1,\delta)) &\leq \Big(\frac{1+\delta}{2+\delta}\Big)^{l+k-3} \frac{\Gamma(\pi_0 + \frac{1}{2+\delta})\Gamma(\pi_{l-1})}{\Gamma(\pi_{l-1} + \frac{1}{2+\delta})\Gamma(\pi_0+1)} \frac{\Gamma(\rho_0 + \frac{1}{2+\delta})\Gamma(\rho_{l-1})}{\Gamma(\rho_{l-1} + \frac{1}{2+\delta})\Gamma(\rho_0+1)} \\ &\quad \times (1+\delta)\frac{\Gamma(\rho_{l-1} - \frac{\delta}{2+\delta})\Gamma(\pi_{l-1} - \frac{1+\delta}{2+\delta})\Gamma(\pi_l)}{\Gamma(\rho_{l-1} + \frac{1}{2+\delta})\Gamma(\pi_{l-1})\Gamma(\pi_l + \frac{2}{2+\delta})} \prod_{i=1}^{l-2} \frac{1}{\pi_i \rho_i} \\ &\quad \times \frac{\Gamma(\pi_l + \frac{1}{2+\delta})\Gamma(\pi_k)}{\Gamma(\pi_k + \frac{1}{2+\delta})\Gamma(\pi_l+1)} \prod_{i=l+1}^{k-1} \frac{1}{\pi_i} \\ &= (1+\delta)\Big(\frac{1+\delta}{2+\delta}\Big)^{l+k-3} \frac{\Gamma(\pi_0 + \frac{1}{2+\delta})}{\Gamma(\pi_0+1)} \frac{\Gamma(\rho_0 + \frac{1}{2+\delta})}{\Gamma(\rho_0+1)} \frac{\Gamma(\rho_{l-1})\Gamma(\rho_{l-1} - \frac{\delta}{2+\delta})}{\Gamma(\rho_{l-1} + \frac{1}{2+\delta})\Gamma(\rho_{l-1} + \frac{1}{2+\delta}))} \\ &\quad \times \frac{\Gamma(\pi_l + \frac{1}{2+\delta})}{\Gamma(\pi_l + \frac{2}{2+\delta})} \frac{\Gamma(\pi_k)}{\Gamma(\pi_k + \frac{1}{2+\delta})} \frac{1}{(\pi_{l-1} + \frac{1}{2+\delta})} \prod_{i=1}^{l-2} \frac{1}{\pi_i \rho_i} \prod_{i=l}^{k-1} \frac{1}{\pi_i}.\end{aligned} \tag{7.3.38}$$

By (8.3.8), this can be bounded by

$$C\Big(\frac{1+\delta}{2+\delta}\Big)^{l+k} (\pi_0\rho_0)^{-(1+\delta)/(2+\delta)} (\pi_l \pi_k)^{-1/(2+\delta)} \prod_{i=1}^{l-1} \frac{1}{\rho_i} \prod_{i=1}^{k-1} \frac{1}{\pi_i}, \tag{7.3.39}$$

where C is a uniform constant. We need to sum the above over $l < k$, all decreasing $\pi = (\pi_0, \dots, \pi_k)$ with $\pi_k \in [T]$ and all decreasing $(\rho_0, \dots, \rho_{l-1})$ with $\rho_{l-1} > \pi_l$. The sum can be bounded from above by summing over *all* decreasing $(\rho_0, \dots, \rho_{l-1})$ and bounding $\pi_l^{-1/(2+\delta)} \leq 1$ to obtain an upper bound as in (7.1.40)

$$\mathbb{E}[N_k(t)]\Big(\frac{1+\delta}{2+\delta}\Big)^{l} \frac{(\log t)^l}{l!} \sum_{s \in [t]} s^{-(1+\delta)/(2+\delta)}. \tag{7.3.40}$$

Therefore,

$$\mathrm{Var}(N_k(t)) \leq \mathbb{E}[N_k(t)] + \mathbb{E}[N_k(t)] \sum_{l=1}^{k-1} \Big(\frac{1+\delta}{2+\delta}\Big)^{l} \frac{(\log t)^l}{l!} \sum_{s \in [t]} s^{-(1+\delta)/(2+\delta)}. \tag{7.3.41}$$

When $T = t^{\varepsilon}$, it is not hard to adapt the arguments in (7.1.40)–(7.1.46) to show that this is at most $\mathbb{E}[N_k(t)]^2 t^{-\eta}$ for some $\eta = \eta(\varepsilon) > 0$.

□

7.3.2 Logarithmic lower bounds on distances for $\delta > 0$

In this section we investigate lower bounds on the distances when $\delta > 0$, in which case $\gamma = m/(2m+\delta) < 1/2$. By Proposition 7.11,

$$\mathbb{P}\Big(\mathrm{dist}_{\mathrm{PA}_t(m,\delta)}(1,t) = k\Big) \le c^k \sum_{\pi} \prod_{j=0}^{k-1} \frac{1}{(\pi_j \wedge \pi_{j+1})^{\gamma} (\pi_j \vee \pi_{j+1})^{1-\gamma}}, \tag{7.3.42}$$

where $c = m^2 C$, and where the sum is over all self-avoiding paths $\pi = (\pi_0, \ldots, \pi_k)$ with $\pi_k = t, \pi_0 = 1$. Define

$$f_k(i,t) = \sum_{\vec{s}} \prod_{j=0}^{k-1} \frac{1}{(\pi_j \wedge \pi_{j+1})^{\gamma} (\pi_j \vee \pi_{j+1})^{1-\gamma}}, \tag{7.3.43}$$

where now the sum is over all self-avoinding $\pi = (\pi_0, \ldots, \pi_k)$ with $pi_k = t, \pi_0 = i$, so that

$$\mathbb{P}\Big(\mathrm{dist}_{\mathrm{PA}_t(m,\delta)}(i,t) = k\Big) \le c^k f_k(i,t). \tag{7.3.44}$$

We study the function $f_k(i,t)$ in the following lemma:

Lemma 7.15 (A bound on f_k)**.** *Fix $\gamma < 1/2$. Then, for every $b > \gamma$ such that $\gamma + b < 1$, there exists a $C_{\gamma,b} > 0$ such that, for every $1 \le i < t$ and all $k \ge 1$,*

$$f_k(i,t) \le \frac{C_{\gamma,b}^k}{i^b t^{1-b}}. \tag{7.3.45}$$

Proof. We prove the lemma using induction on $k \ge 1$. To initialize the induction hypothesis, we note that, for $1 \le i < t$ and every $b \ge a$,

$$f_1(i,t) = \frac{1}{(i \wedge t)^{\gamma} (i \vee t)^{1-\gamma}} = \frac{1}{i^{\gamma} t^{1-\gamma}} = \frac{1}{t} \Big(\frac{t}{i}\Big)^{\gamma} \le \frac{1}{t} \Big(\frac{t}{i}\Big)^{b} = \frac{1}{i^b t^{1-b}}. \tag{7.3.46}$$

This initializes the induction hypothesis when $C_{\gamma,b} \ge 1$. To advance the induction hypothesis, note that

$$f_k(i,t) \le \sum_{s=1}^{i-1} \frac{1}{s^{\gamma} i^{1-\gamma}} f_{k-1}(s,t) + \sum_{s=i+1}^{\infty} \frac{1}{i^{\gamma} s^{1-\gamma}} f_{k-1}(s,t). \tag{7.3.47}$$

We now bound each of these two contributions, making use of the induction hypothesis. We bound the first sum by

$$\begin{aligned} \sum_{s=1}^{i-1} \frac{1}{s^{\gamma} i^{1-\gamma}} f_{k-1}(s,t) &\le C_{\gamma,b}^{k-1} \sum_{s=1}^{i-1} \frac{1}{s^{\gamma} i^{1-\gamma}} \frac{1}{s^b t^{1-b}} = \frac{C_{\gamma,b}^{k-1}}{i^{1-\gamma} t^{1-b}} \sum_{s=1}^{i-1} \frac{1}{s^{\gamma+b}} \\ &\le \frac{1}{1-\gamma-b} \frac{C_{\gamma,b}^{k-1}}{i^b t^{1-b}}, \end{aligned} \tag{7.3.48}$$

since $\gamma + b < 1$. We bound the second sum by

$$\begin{aligned}\sum_{s=i+1}^{\infty} \frac{1}{i^\gamma s^{1-\gamma}} f_{k-1}(s,t) &\le C_{\gamma,b}^{k-1} \sum_{s=i+1}^{t-1} \frac{1}{i^\gamma s^{1-\gamma}} \frac{1}{s^b t^{1-b}} + C_{\gamma,b}^{k-1} \sum_{s=t+1}^{\infty} \frac{1}{i^\gamma s^{1-\gamma}} \frac{1}{t^b s^{1-b}} \\ &= \frac{C_{\gamma,b}^{k-1}}{i^\gamma t^{1-b}} \sum_{s=i+1}^{t-1} \frac{1}{s^{1-\gamma+b}} + \frac{C_{\gamma,b}^{k-1}}{i^\gamma t^b} \sum_{s=t+1}^{\infty} \frac{1}{s^{2-\gamma-b}} \\ &\le \frac{1}{b-\gamma} \frac{C_{\gamma,b}^{k-1}}{i^b t^{1-b}} + \frac{1}{1-\gamma-b} \frac{C_{\gamma,b}^{k-1}}{i^b t^{1-b}},\end{aligned} \tag{7.3.49}$$

since $1 + b - a > 1$, $2 - \gamma - b > 1, b > \gamma$ and $(t/i)^\gamma \le (t/i)^b$. We conclude that

$$f_k(i,t) \le \frac{C_{\gamma,b}^{k-1}}{i^b t^{1-b}} \Big(\frac{1}{b-\gamma} + \frac{2}{1-\gamma-b} \Big) \le \frac{C_{\gamma,b}^{k}}{i^b t^{1-b}}, \tag{7.3.50}$$

when

$$C_{\gamma,b} = \frac{1}{b-\gamma} + \frac{2}{1-\gamma-b} \ge 1. \tag{7.3.51}$$

This advances the induction hypothesis, and completes the proof of Lemma 7.15. □

We next prove the upper bound on the diameter of $\mathrm{PA}_t(m,\delta)$. By Lemma 7.15 and (7.3.44),

$$\mathbb{P}\Big(\mathrm{dist}_{\mathrm{PA}_t(m,\delta)}(1,t) = k \Big) \le \frac{(cC_{\gamma,b})^k}{t^{1-b}}. \tag{7.3.52}$$

As a result,

$$\mathbb{P}\big(\mathrm{diam}(\mathrm{PA}_t(m,\delta)) \le k \big) \le \mathbb{P}\Big(\mathrm{dist}_{\mathrm{PA}_t(m,\delta)}(1,t) \le k \Big) \le \frac{(cC_{\gamma,b})^{k+1}}{t^{1-b}(cC_{\gamma,b}-1)} = o(1), \tag{7.3.53}$$

whenever $k \le \frac{1-b}{\log(cC_{\gamma,b})} \log t$. We conclude that there exists $c_2 = c_2(m,\delta)$ such that $\mathrm{diam}(\mathrm{PA}_t(m,\delta)) \ge c_2 \log t$ whp. □

We next extend the above discussion to *typical* distances:

Lemma 7.16 (Typical distances for $\delta > 0$)**.** *Fix $m \ge 1$ and $\delta > 0$. Let $H_t = \mathrm{dist}_{\mathrm{PA}_t(m,\delta)}(V_1, V_2)$ be the distance between two uniformly chosen vertices in $[t]$. Then, whp, for $c_2 = c_2(m,\delta) > 0$ sufficiently small, $H_t \ge c_2 \log t$.*

Proof. By Lemma 7.15, with $K = \log(cC_{a,b} \vee 2)$ and $\gamma < b < 1 - \gamma$, and for all $1 \le i < j \le t$,

$$\mathbb{P}\Big(\mathrm{dist}_{\mathrm{PA}_t(m,\delta)}(i,j) = k \Big) \le c^k f_k(i,j) \le \frac{\mathrm{e}^{Kk}}{i^b j^{1-b}}. \tag{7.3.54}$$

As a result,

$$\mathbb{P}\Big(\mathrm{dist}_{\mathrm{PA}_t(m,\delta)}(i,j) \le c_2 \log t \Big) \le \frac{t^{Kc_2}}{i^b j^{1-b}} \frac{\mathrm{e}^K}{\mathrm{e}^K - 1}, \tag{7.3.55}$$

and thus, using also $\sum_{i=1}^{j-1} i^{-b} \le j^{1-b}/(1-b)$,

$$\mathbb{P}(H_t \le c_2 \log t) = \frac{1}{t^2}\mathbb{P}\Big(\mathrm{dist}_{\mathrm{PA}_t(m,\delta)}(i,j) \le c_2 \log t\Big) \le 2 \sum_{1\le i<j\le t} \frac{t^{Kc_2}}{i^b j^{1-b}} = O(t^{Kc_2-1}) = o(1), \tag{7.3.56}$$

for every $c_2 > 0$ such that $Kc_2 + 1 < 2$. □

7.3.3 Lower bounds on distances for $\delta = 0$ and $m \ge 2$

In this section we investigate lower bounds on the distances in $\mathrm{PA}_t(m,\delta)$ when $m \ge 2$ and $\delta = 0$ and prove the lower bound in Theorem 7.9.

We again start from Proposition 7.11, which as we will show implies that for $\delta = 0$,

$$k = \frac{\log(t-1)}{\log(3Cm^2 \log t)} \tag{7.3.57}$$

is a lower bound for the diameter of $\mathrm{PA}_t(m,\delta)$. Consider a path π of length l consisting of the vertices $\pi_0, \pi_1, \dots, \pi_l$, then (**??**) implies that

$$\mathbb{P}(\pi \subseteq \mathrm{PA}_t(m,\delta)) \le (Cm^2)^l \prod_{j=0}^{l-1} \frac{1}{\sqrt{\pi_j \pi_{j+1}}} = \frac{(Cm^2)^l}{\sqrt{\pi_0 \pi_l}} \prod_{j=1}^{l-1} \frac{1}{\pi_j}. \tag{7.3.58}$$

Thus, the expected number of paths π of length l between $\pi_0 = t$ and $\pi_l = t-1$ is bounded by

$$\begin{aligned} &\frac{(Cm^2)^l}{\sqrt{t(t-1)}} \sum_{1\le \pi_1,\dots,\pi_{l-1}\le t-2} \prod_{j=1}^{l-1} \frac{1}{\pi_j} = \frac{(Cm^2)^l}{\sqrt{t(t-1)}} \left(\sum_{s=1}^{t-2} \frac{1}{s}\right)^{l-1} \\ &\quad \le \frac{(Cm^2)^l}{t-1} (\log t)^{l-1} \le (1/2)^l (\log t)^{-1} \to 0 \end{aligned}$$

precisely when $(2Cm^2 \log t)^l \le t-1$, or, equivalently,

$$l \le \frac{\log(t-1)}{\log((2Cm^2)\log t)}. \tag{7.3.59}$$

Equality in (7.3.59) holds for k in (7.3.57). This implies that the diameter is at least L in (7.3.57), and completes the proof of Theorem 7.9. □

7.3.4 Typical distance in PA-models: log log-lower bound

In this section we prove the lower bound in Theorem 7.10. We do so in a more general setting assuming an upper bound on the existence of paths in the model:

Assumption 7.17. *There exist κ and γ such that, for all t and pairwise distinct vertices $\pi_0, \dots, \pi_l \in [t]$,*

$$\mathbb{P}(\pi_0 \leftrightarrow \pi_1 \leftrightarrow \dots \leftrightarrow \pi_l) \le \prod_{i=1}^{l} \kappa(\pi_{i-1} \wedge \pi_i)^{-\gamma} (\pi_i \vee \pi_{i-1})^{\gamma-1}. \tag{7.3.60}$$

By Proposition 7.11, Assumption 7.17 is satisfied for $\mathrm{PA}_t(m,\delta)$ with $\gamma = m/(2m+\delta)$. We expect log log-distances in such networks if and only if $\delta \in (-m, 0)$, so that $\frac{1}{2} < \gamma < 1$. Theorem 7.18, which is the main result in this section, gives a lower bound on the typical distance in this case:

Theorem 7.18 (Doubly logarithmic lower bound on distances PAMs)**.** *Let* $(\mathrm{PA}_t)_{t\in\mathbb{N}}$ *be a random graph model that satisfies Assumption 7.17 for some* γ *satisfying* $\frac{1}{2} < \gamma < 1$*. Then, for random vertices* V_1 *and* V_2 *chosen independently and uniformly from* $[t]$ *and whp as* K *grows large,*

$$\mathrm{dist}_{\mathrm{PA}_t}(V_1, V_2) \geq \frac{4\log\log t}{\log(\gamma/(1-\gamma))} - K. \tag{7.3.61}$$

For $\mathrm{PA}_t(m,\delta)$, $\gamma = m/(2m+\delta)$, so that

$$\frac{\gamma}{1-\gamma} = \frac{m}{m+\delta} = \frac{1}{\tau-2}, \tag{7.3.62}$$

where we recall that $\tau = 3 + \delta/m$. Therefore, Theorem 7.18 proves the lower bound in Theorem 7.10.

The proof of Theorem 7.18 is based on a *constrained* or *truncated first order method*, similar to the ones used for $\mathrm{NR}_n(\boldsymbol{w})$ in Theorem 3.7 and for $\mathrm{CM}_n(\boldsymbol{d})$ in Theorem 5.11. Due to the fact that the probability for existence of paths satisfies a rather different bound compared to the bounds for $\mathrm{NR}_n(\boldsymbol{w})$ and $\mathrm{CM}_n(\boldsymbol{d})$, this truncated first order method looks rather different compared to the ones presented in the proof of Theorems 3.7 and 5.11.

Let us now briefly explain the truncated first moment method. We start with an explanation of the (unconstrained) first moment bound and its shortcomings. Let v, w be distinct vertices of PA_t. Then for $k_n \in \mathbb{N}$

$$\begin{aligned}\mathbb{P}(\mathrm{dist}_{\mathrm{PA}_t(m,\delta)}(v,w) \leq 2k_n) &= \mathbb{P}\left(\bigcup_{k=1}^{2k_n}\bigcup_{\pi}\{v \leftrightarrow \pi_1 \leftrightarrow \pi_2 \leftrightarrow \ldots \leftrightarrow \pi_{k-1} \leftrightarrow w\}\right)\\ &\leq \sum_{k=1}^{2k_n}\sum_{\pi}\prod_{j=1}^{k} p(\pi_{j-1},\pi_j),\end{aligned} \tag{7.3.63}$$

where $\pi = (\pi_0, \ldots, \pi_k)$ is any collection of pairwise distinct vertices in PA_t with $\pi_0 = v$ and $\pi_k = w$ and, for n, $m \in \mathbb{N}$, we define

$$p(n,m) = \kappa(n \wedge m)^{-\gamma}(n \vee m)^{\gamma-1}. \tag{7.3.64}$$

We assign to each path $\pi = (\pi_0, \ldots, \pi_k)$ the weight

$$p(\pi) = \prod_{j=1}^{k} p(\pi_{j-1}, \pi_j), \tag{7.3.65}$$

and the upper bound is just the sum over the weights of all paths from v to w of length no more than $2k_n$. The shortcoming of this bound is that the paths that contribute

most to the total weight are those that connect v, resp. w, quickly to vertices with extremely small indices and thus extremely high degree. Since such paths are quite unlikely, they have to be removed in order to get a reasonable estimate.

To this end we define a decreasing sequence $\ell = (\ell_k)_{k=0,\dots,k}$ of positive integers and consider a tuple of vertices $\pi = (\pi_0, \dots, \pi_k)$ as *good* if $v_l \wedge v_{k-l} \geq \ell_l$ for all $l \in \{0, \dots, k\}$. We denote the probability that there exists a good path of length k between v and w by $\mathcal{E}_k(v, w)$. We further denote by $\mathcal{F}_k(v)$ the event that there exists a bad path of length k in the network starting at v, i.e., a path $v = \pi_0 \leftrightarrow \dots \leftrightarrow \pi_l$ such that $\pi_0 \geq \ell_0, \dots, \pi_{l-1} \geq \ell_{l-1}$, but $\pi_l < \ell_l$, i.e., a path that traverses the threshold after exactly l steps. For fixed vertices $v, w \geq \ell_0$, the truncated first moment estimate is the estimate that

$$\mathbb{P}(\mathrm{dist}_{\mathrm{PA}_t}(v, w) \leq 2k) \leq \sum_{l=1}^{k} \mathbb{P}(\mathcal{F}_l(v)) + \sum_{l=1}^{k} \mathbb{P}(\mathcal{F}_l(w)) + \sum_{k=1}^{2k} \mathbb{P}(\mathcal{E}_l(v, w)) \tag{7.3.66}$$

where the good paths in the last sum start with $v_0 = v$ and end with $v_k = w$. Equation (7.3.66) is identical to the inequality (3.3.33) used in the proof of Theorem 3.7. However, the notion of *good* has changed due to the fact that vertices no longr have a *weight*, but rather an *age*, and vertices that have appeared early in PA_t are the most likely to have large degrees. This explains why good vertices have high indices for PA_t, while good vertices have high weights for $\mathrm{NR}_n(\boldsymbol{w})$.

By assumption,

$$\mathbb{P}(\pi_0 \leftrightarrow \dots \leftrightarrow \pi_k) \leq p(\pi_0, \dots, \pi_k) \tag{7.3.67}$$

so that for $v \geq \ell_0$ and $l = 1, \dots, k$, and with $\pi = (\pi_0, \dots, \pi_)$ with $\pi_0 = v$,

$$\mathbb{P}(\mathcal{F}_l(v)) \leq \sum_{\pi_1=\ell_1}^{t} \dots \sum_{\pi_{l-1}=\ell_{l-1}}^{t} \sum_{\pi_l=1}^{\ell_l-1} p(\pi). \tag{7.3.68}$$

Given $\varepsilon > 0$ we choose $\ell_0 = \lceil \varepsilon t \rceil$ and $(\ell_j)_{j=0,\dots,k}$ decreasing fast enough so that the first two summands on the right hand side of (7.3.66) together are no larger than 2ε. For $l \in [k]$, we set

$$f_{l,t}(v, u) := \mathbb{1}_{\{v \geq \ell_0\}} \sum_{\pi_1=\ell_1}^{t} \dots \sum_{\pi_{l-1}=\ell_{l-1}}^{t} p(v, \pi_1, \dots, \pi_{l-1}, u), \tag{7.3.69}$$

and set $f_{0,t}(v, u) = \mathbb{1}_{\{v=u\}} \mathbb{1}_{\{u \leq t\}}$. To rephrase the truncated moment estimate in terms of f, note that p is symmetric so that for all $l \leq 2k$

$$\begin{aligned} \mathbb{P}(\mathcal{E}_l(v, w)) &\leq \sum_{\pi_1=\ell_1}^{t} \dots \sum_{\pi_{\lfloor l/2 \rfloor}=\ell_{\lfloor l/2 \rfloor}}^{t} \dots \sum_{\pi_{l-1}=\ell_1}^{t} p(v, \pi_1, \dots, \pi_{\lfloor l/2 \rfloor}) p(\pi_{\lfloor l/2 \rfloor}, \dots, \pi_{l-1}, w) \\ &= \sum_{\pi_{\lfloor l/2 \rfloor}=\ell_{\lfloor l/2 \rfloor}}^{t} f_{\lfloor l/2 \rfloor,t}(v, \pi_{\lfloor l/2 \rfloor}) f_{\lceil l/2 \rceil,t}(w, \pi_{\lfloor l/2 \rfloor}). \end{aligned} \tag{7.3.70}$$

Using the recursive representation

$$f_{k+1,t}(v,n) = \sum_{m=\ell_k}^{t} f_{k,t}(v,m)p(m,n), \tag{7.3.71}$$

we establish upper bounds for $f_{k,t}(v,u)$ and use these to show that the rightmost term in (7.3.66) remains small if $k = k_n$ is chosen sufficiently small. This leads to the lower bounds for the typical distance in Theorem 7.18. Let us now make these ideas precise:

Proof of Theorem 7.18. We assume that Assumption 7.17 holds for a $\gamma \in (\frac{1}{2}, 1)$ with a fixed constant κ. Recall the definition of $f_{k,t}$ and the key estimates (7.3.66), (7.3.68) and (7.3.70), which combined give

$$\begin{aligned}\mathbb{P}(\mathrm{dist}_{\mathrm{PA}_t}(v,w) \le 2k_n) \le \sum_{k=1}^{k_n}\sum_{l=1}^{\ell_k-1} f_{k,t}(v,l) + \sum_{k=1}^{k_n}\sum_{l=1}^{\ell_k-1} f_{k,t}(w,l) \\ + \sum_{k=1}^{2k_n}\sum_{v_{\lfloor k/2\rfloor}=\ell_{\lfloor n/2\rfloor}}^{t} f_{\lfloor n/2\rfloor,t}(v,v_{\lfloor n/2\rfloor}) f_{\lceil n/2\rceil,t}(w,v_{\lfloor n/2\rfloor}).\end{aligned} \tag{7.3.72}$$

The remaining task of the proof is to choose $k_n \in \mathbb{N}$ and $2 \le \ell_{k_n} \le \ldots \le \ell_0 \le t$ which allow the required estimates for the right-hand side. Denote the truncated version of $f_{k,t}(v,m)$ by $\bar{f}_{k,t}(v,m) = \mathbb{1}_{\{m\ge\ell_k\}} f_{k,t}(v,m)$. Our aim is to provide a majorant of the form

$$f_{k,t}(v,m) \le \alpha_k m^{-\gamma} + \mathbb{1}_{\{m>\ell_{k-1}\}}\beta_k m^{\gamma-1} \tag{7.3.73}$$

for suitably chosen parameters $\alpha_k, \beta_k \ge 0$. Key to this choice is the following lemma:

Lemma 7.19 (A recursive bound on $f_{k,t}(v,m)$ for $\gamma \in (1/2,1)$)**.** *Let $\gamma \in (1/2,1)$ and suppose that $2 \le \ell \le t$, $\alpha, \beta \ge 0$ and $q\colon [t] \to [0,\infty)$ satisfies*

$$q(m) \le \mathbb{1}_{\{m\ge\ell\}}(\alpha m^{-\gamma} + \beta m^{\gamma-1}) \quad \text{for all } m \in [t]. \tag{7.3.74}$$

Then there exists a constant $c > 1$ depending only on γ and κ such that

$$\sum_{k=1}^{t} q(k)p(k,m) \le c\left(\alpha\log(t/\ell) + \beta t^{2\gamma-1}\right)m^{-\gamma} + c\mathbb{1}_{\{m>\ell\}}\left(\alpha\ell^{1-2\gamma} + \beta\log(t/\ell)\right)m^{\gamma-1} \tag{7.3.75}$$

for all $m \in [t]$.

Proof. We use (7.3.64) to rewrite

$$\begin{aligned}\sum_{k=1}^{t} q(k)p(k,m) &= \sum_{k=m\vee\ell}^{t} q(k)p(k,m) + \mathbb{1}_{\{m>\ell\}}\sum_{k=\ell}^{m-1} q(k)p(k,m) \\ &= \sum_{k=m\vee\ell}^{t} \kappa(\alpha k^{-\gamma} + \beta k^{\gamma-1})k^{\gamma-1}m^{-\gamma} + \mathbb{1}_{\{m>\ell\}}\sum_{k=\ell}^{m-1}\kappa(\alpha k^{-\gamma} + \beta k^{\gamma-1})k^{-\gamma}m^{\gamma-1}.\end{aligned}$$

Simplifying the sums leads to

$$\begin{aligned}\sum_{k=1}^{t} q(k)p(k,m) &\le \kappa\Big(\alpha \sum_{k=m\vee\ell}^{t} k^{-1} + \beta \sum_{k=m\vee\ell}^{t} k^{2\gamma-2}\Big)m^{-\gamma} \\ &\quad + \kappa \mathbb{1}_{\{m>\ell\}}\Big(\alpha \sum_{k=\ell}^{m-1} k^{-2\gamma} + \beta \sum_{k=\ell}^{m-1} k^{-1}\Big)m^{\gamma-1} \\ &\le \kappa\Big(\alpha \log\Big(\frac{m}{\ell-1}\Big) + \frac{\beta}{2\gamma-1} t^{2\gamma-1}\Big)m^{-\gamma} \\ &\quad + \kappa \mathbb{1}_{\{m>\ell\}}\Big(\frac{\alpha}{1-2\gamma}(\ell-1)^{1-2\gamma} + \beta \log\Big(\frac{m}{\ell-1}\Big)\Big)m^{\gamma-1}. \quad (7.3.76)\end{aligned}$$

This immediately implies the assertion since $\ell \ge 2$ by assumption. □

We apply Lemma 7.19 iteratively. We use induction to prove that there exist $(\ell_k)_{k\ge0}$, $(\alpha_k)_{k\ge1}$ and $(\beta_k)_{k\ge1}$ such that

$$f_{k,t}(v,m) \le \alpha_k m^{-\gamma} + \beta_k m^{\gamma-1} \quad \text{for all } m \in [t]. \quad (7.3.77)$$

The sequences $(\ell_k)_{k\ge0}$, $(\alpha_k)_{k\ge1}$ and $(\beta_k)_{k\ge1}$ are chosen as follows. We let $\ell_0, \alpha_1, \beta_1$ be determined by

$$\ell_0 = \lceil \varepsilon t \rceil, \quad \alpha_1 = \kappa(\varepsilon t)^{\gamma-1} \quad \text{and} \quad \beta_1 = \kappa(\varepsilon t)^{-\gamma}, \quad (7.3.78)$$

where $\varepsilon > 0$ is small.

For higher values of k, $\alpha_{k+1}, \beta_{k+1}$ and ℓ_k satisfy the recursions

$$\frac{6\varepsilon}{\pi^2 k^2} \ge \frac{1}{1-\gamma}\alpha_k \ell_k^{1-\gamma}, \quad (7.3.79)$$

where we let ℓ_k be the largest integer satisfying the above inequality under the assumption that $\ell_k \ge 2$, and

$$\alpha_{k+1} = c\left(\alpha_k \log\left(t/\ell_k\right) + \beta_k t^{2\gamma-1}\right), \qquad \beta_{k+1} = c\left(\alpha_k \ell_k^{1-2\gamma} + \beta_k \log\left(t/\ell_k\right)\right), \quad (7.3.80)$$

Check whether equality is okay??

where c is the constant in Lemma 7.19.

We start by initializing the induction. Fix $v \ge \ell_0$. Then, for all $m \in [t]$,

$$\begin{aligned} f_{1,t}(v,m) = p(v,m) &\le k\ell_0^{\gamma-1} m^{-\gamma} + \mathbb{1}_{\{m>\ell_0\}} \kappa \ell_0^{-\gamma} m^{\gamma-1} \\ &\le \alpha_1 m^{-\gamma} + \mathbb{1}_{\{m>\ell_0\}} \beta_1 m^{\gamma-1}. \end{aligned} \quad (7.3.81)$$

Now suppose, for some $k \in \mathbb{N}$, that we have chosen α_k, β_k and an integer ℓ_{k-1} such that

$$f_{k,t}(v,m) \le \alpha_k m^{-\gamma} + \beta_k m^{\gamma-1} \quad \text{for all } m \in [t]. \quad (7.3.82)$$

By the induction hypothesis we can apply Lemma 7.19 with $\ell = \ell_k$ and $q(m) = \bar{f}_{k,t}(v,m) = \mathbb{1}_{\{m \geq \ell_k\}} f_{k,t}(v,m)$. Then Lemma 7.19 yields that

$$f_{k+1,t}(v,m) \leq \alpha_{k+1} m^{-\gamma} + \mathbb{1}_{\{m > \ell_k\}} \beta_{k+1} m^{\gamma-1} \quad \text{for all } m \in [t], \tag{7.3.83}$$

showing that the induction can be advanced up to the point where $\ell_k < 2$. Having advanced the induction hypothesis, we obtain that (7.3.77) holds with the given choices of $(\ell_k)_{k\geq 0}$, $(\alpha_k)_{k\geq 1}$ and $(\beta_k)_{k\geq 1}$.

We next use (7.3.77) to prove Theorem 7.18. Summing over (7.3.83) and using (7.3.77) and (7.3.79) we obtain

$$\sum_{l=1}^{\ell_k - 1} f_{k,t}(v,l) \leq \frac{1}{1-\gamma} \alpha_k \ell_k^{1-\gamma} \leq \frac{6\varepsilon}{\pi^2 k^2}, \tag{7.3.84}$$

which, when summed over all $k \geq 1$ is bounded by ε. Hence the first two summands on the right-hand side in (7.3.72) together are smaller than 2ε. It remains to choose $k_n = k_n(t)$ as large as possible while ensuring that $\ell_{k_n} \geq 2$ and

$$\lim_{t\to\infty} \sum_{k=1}^{2k_n} \sum_{\pi_{\lfloor k/2 \rfloor} = \ell_{\lfloor k/2 \rfloor}}^{t} f_{\lfloor k/2 \rfloor, t}(v, \pi_{\lfloor k/2 \rfloor}) f_{\lceil k/2 \rceil, t}(w, \pi_{\lfloor k/2 \rfloor}) = 0. \tag{7.3.85}$$

To this end recall that ℓ_k is the largest integer satisfying (7.3.79) and the parameters α_k, β_k are defined via equalities in (7.3.80). To establish lower bounds for the decay of ℓ_k we investigate the growth of $\eta_k = t/\ell_k > 0$. Going backwards through the definitions yields, for $k \geq 1$, that there exists a constant $C > 0$ such that

$$\eta_{k+2}^{1-\gamma} \leq C \Big[\frac{(k+2)^2}{k^2} \eta_k^{\gamma} + \frac{(k+2)^2}{(k+1)^2} \eta_{k+1}^{1-\gamma} \log \eta_{k+1} \Big], \tag{7.3.86}$$

with $\eta_1, \eta_2 \leq C_0$ for some constant $C_0 > 0$ (which, as all constants in this paragraph, may depend on ε). Indeed, writing C for a constant that may change from line to line, using first the relation for ℓ_{k+2} in (7.3.79), followed by the equality for α_{k+2} in (7.3.80),

$$\eta_{k+2}^{1-\gamma} = t^{1-\gamma} \ell_{k+2}^{\gamma-1} = C t^{1-\gamma} \alpha_{k+2} (k+2)^2 = C t^{1-\gamma} (k+2)^2 (\alpha_{k+1} \log(\eta_{k+1}) + \beta_{k+1} t^{2\gamma-1}). \tag{7.3.87}$$

The first term on the right-hand side of (7.3.87) equals $C \frac{(k+2)^2}{(k+1)^2} \eta_{k+1}^{1-\gamma} \log(\eta_{k+1})$, which is the second term on the right-hand side of (7.3.86). For the second term on the right-hand side of (7.3.87), we use the equality for β_{k+1} in (7.3.80) to write it as

$$C t^{\gamma} (k+2)^2 \beta_{k+1} = C t^{\gamma} (k+2)^2 \big(\alpha_k \ell_k^{1-2\gamma} + \beta_k \log(\eta_k) \big). \tag{7.3.88}$$

Now the first term is equal to $C \frac{(k+2)^2}{k^2} \eta_{k+1}^{\gamma}$, which is the first term in (7.3.86). We are left with $C t^{\gamma} (k+2)^2 \beta_k \log(\eta_k)$. Since $c \beta_k t^{2\gamma-1} \leq \alpha_{k+1}$,

$$\begin{aligned} C t^{\gamma} (k+2)^2 \beta_k \log(\eta_k) &\leq C t^{1-\gamma} (k+2)^2 \alpha_{k+1} \log(\eta_k) \leq C t^{1-\gamma} \frac{(k+2)^2}{(k+1)^2} \ell_{k+1}^{-(1-\gamma)} \\ &= C \eta_{k+1}^{1-\gamma} \log(\eta_k) \leq C \eta_{k+1}^{1-\gamma} \log(\eta_{k+1}), \end{aligned} \tag{7.3.89}$$

since $k \mapsto \ell_k$ is degreasing, so that $k \mapsto \eta_k$ is increasing. This completes the proof of (7.3.86).

Lemma 7.20 (Inductive bounds on η_k)**.** *Let $(\eta_k)_{k\geq 1}$ satisfy $\eta_0 = 1/\varepsilon$ and $\eta_1 =$, and (7.3.86) holds for $k \geq 2$. Then there exist constants $b, B > 0$ such that*

$$\eta_k \leq b \exp\Big(B\big(\frac{\gamma}{1-\gamma}\big)^{k/2}\Big). \tag{7.3.90}$$

Proof. By assumption, (7.3.90) follows for $k = 0$ by the fact that $\eta_0 = t/\ell_0 \leq 1/\varepsilon$. By (7.3.86), we can obtain similar bounds for η_k with $k \leq 2$, Suppose that we know that (7.3.90) holds for all $l \leq k-1$ where $k \geq 3$. Then (7.3.86) together with the fact that $k/(k-1)$ and $k/(k-2)$ are bounded yields that there exists a constant $C > 0$ such that

$$\eta_k^{1-\gamma} \leq C\eta_{k-2}^{\gamma} + C\eta_{k-1}^{1-\gamma} \log \eta_{k-1}, \tag{7.3.91}$$

and using that $(x+y)^{1/(1-\gamma)} \leq 2^{1/(1-\gamma)}(x^{1/(1-\gamma)} + y^{1/(1-\gamma)})$ leads us to

$$\eta_k \leq C\eta_{k-2}^{\gamma/(1-\gamma)} + C\eta_{k-1}(\log \eta_{k-1})^{1/(1-\gamma)}. \tag{7.3.92}$$

Iterating the above inequality once more yields

$$\eta_k \leq C(2C)^{\gamma/(1-\gamma)}\eta_{k-4}^{(\gamma/(1-\gamma))^2} + C(2C)^{\gamma/(1-\gamma)}\eta_{k-3}^{\gamma/(1-\gamma)}(\log \eta_{k-3})^{\gamma/(1-\gamma)^2} + C\eta_{k-1}(\log \eta_{k-1})^{1/(1-\gamma)}. \tag{7.3.93}$$

Iterating indefinitely yields

$$\eta_k \leq C(2C)^{\sum_{l=1}^{k/2}(\gamma/(1-\gamma))^l}\eta_0^{(\gamma/(1-\gamma))^{k/2}} + C\sum_{i=1}^{k/2}(2C)^{\sum_{l=1}^{i-1}(\gamma/(1-\gamma))^l}\eta_{k-2i+1}^{(\gamma/(1-\gamma))^{i-1}}(\log \eta_{k-2i+1})^{\gamma^{i-1}/(1-\gamma)^i}. \tag{7.3.94}$$

We now prove by induction that there exists constants b, B such that

$$\eta_k \leq b \exp\Big(B\big(\frac{\gamma}{1-\gamma}\big)^{k/2}\Big). \tag{7.3.95}$$

Using the induction hypothesis, we can bound η_k by

Continue from here!!

$$\eta_k^{1-\gamma} \leq CA^{\gamma(k-1)} \exp\Big(B(1-\gamma)\big(\frac{\gamma}{1-\gamma}\big)^{k/2}\Big) + Cb^{1-\gamma} \exp\Big(B\sqrt{(1-\gamma)^3/\gamma}\big(\frac{\gamma}{1-\gamma}\big)^{k/2}\Big) B\big(\frac{\gamma}{1-\gamma}\big)^{(k-1)/2}. \tag{7.3.96}$$

Since $\gamma > 1/2$, we have that $(1-\gamma)/\gamma < 1$, so that for large k the second term is negligible compared to the first term for every $k \geq 2$ when B is sufficiently large. For $\gamma > 1/2$... Taking b large enough, we can bound the right-hand side of (7.3.96) by $b^{1-\gamma} \exp\Big(B(1-\gamma)\big(\frac{\gamma}{1-\gamma}\big)^{k/2}\Big)$ □

We now use (7.3.83) and (7.3.90) to estimate

$$\begin{aligned}\sum_{n=1}^{2k_n}\sum_{u=\ell_k}^{t}\mu^{(v)}_{n^*}(u)\mu^{(w)}_{n-n^*}(u) &\le 2\sum_{k=1}^{k_n}\sum_{u=\ell_k}^{t}(\alpha_k u^{-\gamma}+\beta_k u^{\gamma-1})^2\\ &\le \frac{4}{2\gamma-1}\sum_{k=1}^{k_n}(\alpha_k^2\ell_k^{1-2\gamma}+\beta_k^2 t^{2\gamma-1})\\ &\le \frac{4}{2\gamma-1}k_n(\alpha_{k_n}^2\ell_{k_n}^{1-2\gamma}+\beta_{k_n}^2 t^{2\gamma-1}).\end{aligned} \tag{7.3.97}$$

Using (7.3.79) and (7.3.90) the first summand in the bracket can be estimated by

$$\alpha_{k_n}^2\ell_{k_n}^{1-2\gamma}\le\left(k_n^{-2}\frac{6\varepsilon}{\pi^2}(1-\gamma)\right)^2\ell_{k_n}^{-1}\le\left(\frac{6\varepsilon}{b\pi^2}(1-\gamma)\right)^2\frac{1}{tk_n^4}\exp\left(B\left(\frac{\gamma}{1-\gamma}\right)^{k_n/2}\right). \tag{7.3.98}$$

Using equality in (7.3.80) we get $\beta_{k_n}\le c(\alpha_{k_n}\ell_{k_n}^{1-2\gamma}+\alpha_{k_n}t^{1-2\gamma}\log(t/\ell_{k_n}))$. Noting that the second summand on the right-hand side is bounded by a multiple of the first, we find a constant $C_1>0$ such that $\beta_{k_n}^2t^{2\gamma-1}\le C_1\alpha_{k_n}^2\ell_{k_n}^{1-2\gamma}$, and thus, for a suitable constant $C_2>0$,

$$\sum_{n=1}^{2k_n}\sum_{u=\ell_k}^{t}\mu^{(v)}_{n^*}(u)\mu^{(w)}_{n-n^*}(u)\le C_2\frac{1}{tk_n{}^3}\exp\left(B\left(\frac{\gamma}{1-\gamma}\right)^{k_n/2}\right). \tag{7.3.99}$$

Hence, for a suitable constant $C>0$, choosing

$$k_n\le\frac{\log\log t}{\log\sqrt{\frac{\gamma}{1-\gamma}}}-K, \tag{7.3.100}$$

we obtain that the term we consider goes to zero of order $O((\log\log t)^{-3})$. Note from (7.3.90) that this choice also ensures that $\ell_{k_n}\ge 2$. We have thus shown that

$$\mathbb{P}(\mathrm{dist}_{\mathrm{PA}_t(m,\delta)}(v,w)\le 2k_n)\le 2\varepsilon+O\left((\log\log t)^{-3}\right), \tag{7.3.101}$$

whenever $v,w\ge\ell_0=\lceil\varepsilon t\rceil$, which implies the statement of Theorem 7.18. □

7.4 Small-world effect in PA models: upper bounds

7.4.1 Logarithmic upper bounds for $\delta>0$

In this section we prove lower bounds on distances in $\mathrm{PA}_t(m,\delta)$. We start by proving that the logarithmic bounds on the diameter for $(\mathrm{PA}_t(m,\delta))_{t\ge1}$:

Proof of Theorem 7.7. We start by proving the claim for $(\mathrm{PA}^{(b)}_{m,\delta}(t))_{t\ge1}$. Since $(\mathrm{PA}^{(b)}_{m,\delta}(t))_{t\ge1}$ is obtained from $(\mathrm{PA}^{(b)}_{1,\delta/m}(mt))_{t\ge1}$ by collapsing m successive vertices, $\mathrm{diam}(\mathrm{PA}^{(b)}_{m,\delta}(t))\preceq\mathrm{diam}(\mathrm{PA}^{(b)}_{1,\delta/m}(mt))$, and the result follows from Theorem 7.3.

□

7.4.2 The diameter of the core

In this section we investigate $\mathrm{PA}_t(m,\delta)$ with $m \geq 2$ and $\delta \in (-m, 0)$ and prove the upper bounds in Theorem 7.10.

The proof of Theorem **??** is divided into two key steps. In the first, in Theorem 7.21, we give a bound on the diameter of the *core* which consists of the vertices with degree at least a certain power of $\log t$. This argument is close in spirit to the argument in [98] used to prove bounds on the average distance for the configuration model, but substantial adaptations are necessary to deal with preferential attachment. After this, in Theorem 7.25, we derive a bound on the distance between vertices with a small degree and the core. We start by defining and investigating the core of the preferential attachment model. In the sequel, it will be convenient to prove Theorem **??** for $2t$ rather than for t. Clearly, this does not make any difference for the results.

We adapt the proof of Theorem 5.14 to $\mathrm{PA}_t(m,\delta)$. We recall that

$$\tau = 3 + \frac{\delta}{m}, \tag{7.4.1}$$

so that $-m < \delta < 0$ corresponds to $\tau \in (2,3)$. Throughout this section, we fix $m \geq 2$.

We take $\sigma > \frac{1}{3-\tau} = -\frac{m}{\delta} > 1$ and define the *core* Core_t of the PA-model $\mathrm{PA}_{m,\delta}(2t)$ to be

$$\mathrm{Core}_t = \big\{i \in [t] : D_i(t) \geq (\log t)^\sigma\big\}, \tag{7.4.2}$$

i.e., all the vertices which at time t have degree at least $(\log t)^\sigma$.

For a graph G with vertex set $[t]$ and a given edge set, we write $\mathrm{dist}_G(i,j)$ for the shortest-path distance between i and j in the graph G. Also, for $A \subseteq [t]$, we write

$$\mathrm{diam}_t(A) = \max_{i,j\in A} \mathrm{dist}_{\mathrm{PA}_{m,\delta(t)}}(i,j). \tag{7.4.3}$$

Then, the diameter of the core in the graph $\mathrm{PA}_{m,\delta}(2t)$, which we denote by $\mathrm{diam}_{2t}(\mathrm{Core}_t)$, is bounded in the following theorem:

Theorem 7.21 (The diameter of the core)**.** *Fix $m \geq 2$. For every $\sigma > \frac{1}{3-\tau}$, whp,*

$$\mathrm{diam}_{2t}(\mathrm{Core}_t) \leq (1+o(1))\frac{4\log\log t}{|\log(\tau-2)|}. \tag{7.4.4}$$

The proof of Theorem 7.21 is divided into several smaller steps. We start by proving that the diameter of the *inner core* Inner_t, which is defined by

$$\mathrm{Inner}_t = \big\{i \in \{1,2,\ldots,t\} : D_i(t) \geq t^{\frac{1}{2(\tau-1)}}(\log t)^{-\frac{1}{2}}\big\}, \tag{7.4.5}$$

is, whpbounded by $C_\delta < \infty$. After this, we will show that the distance from the *outer core*, which is defined to be equal to $\mathrm{Outer}_t = \mathrm{Core}_n \backslash \mathrm{Inner}_t$, to the inner core can be bounded by a fixed constant times $\log\log t$. This also shows that the diameter of the outer core is bounded by a different constant times $\log\log t$. We now give the details.

Proposition 7.22 (The diameter of the inner core)**.** *Fix $m \geq 2$ and $\delta \in (-m, 0)$. Then, whp,*

$$\mathrm{diam}_{2t}(\mathrm{Inner}_t) < C_\delta. \tag{7.4.6}$$

Proof. We first introduce the important notion of a t-connector between a vertex $i \in \{1, 2, \ldots, t\}$ and a set of vertices $A \subseteq \{1, 2, \ldots, t\}$, which plays a crucial role throughout the proof. Fix a set of vertices A and a vertex i. We say that the vertex $j \in \{t+1, t+2, \ldots, 2t\}$ is a *t-connector* between i and A if one of the edges incident to j connects to i and another edge incident to j connects to a vertex in A. Thus, when there exists a t-connector between i and A, the distance between i and A in $\mathrm{PA}_{m,\delta}(2t)$ is at most 2.

We note that for a set of vertices A and a vertex i with degree at time t equal to $D_i(t)$, we have that, conditionally on $\mathrm{PA}_{m,\delta}(t)$, the probability that $j \in \{t+1, t+2, \ldots, 2t\}$ is a t-connector for i and A is at least

$$\frac{(D_A(t) + \delta|A|)(D_i(t) + \delta)}{[2t(2m+\delta)]^2}, \tag{7.4.7}$$

independently of the fact whether the other vertices are t-connectors or not, and where, for any $A \subseteq \{1, 2, \ldots, t\}$, we write

$$D_A(t) = \sum_{i \in A} D_i(t). \tag{7.4.8}$$

Since $d_i(t) + \delta \geq m + \delta > 0$ for every $i \leq t$, and $\delta < 0$, we have that

$$D_i(t) + \delta = D_i(t)\Big(1 + \frac{\delta}{D_i(t)}\Big) \geq D_i(t)(1 + \frac{\delta}{m}) = D_i(t)\frac{m+\delta}{m}, \tag{7.4.9}$$

and, thus, also $D_A(t) + \delta|A| \geq D_A(t)\frac{m+\delta}{m}$. As a result, for $\eta = (m+\delta)^2/(2m(2m+\delta))^2 > 0$, the probability that $j \in \{t+1, t+2, \ldots, 2t\}$ is a t-connector for i and A is at least $\frac{\eta D_A(t) D_i(t)}{t^2}$, independently of the fact whether the other vertices are t-connectors or not. Therefore, the probability that there is no t-connector for i and A is, conditionally on $\mathrm{PA}_{m,\delta}(t)$, bounded above by

$$\Big(1 - \frac{\eta D_A(t) D_i(t)}{t^2}\Big)^t \leq \exp\left\{-\frac{\eta D_A(t) D_i(t)}{t}\right\}. \tag{7.4.10}$$

We shall make use of (7.4.10) in several places throughout the proof.

From [I, Theorem 8.2] whp, Inner_t contains at least $\sqrt{t}$ vertices and denote the first $\sqrt{t}$ vertices of Inner_t by I. Observe that for $\tau > 2$ we have $t^{(\tau-1)^{-1}-1} \downarrow 0$ so that, for any $i, j \in I$, the probability that there exists a t-connector for i and j is bounded below by

$$1 - \exp\{-\eta t^{\frac{1}{\tau-1}-1}(\log t)^{-1}\} \geq p_t \equiv t^{\frac{1}{\tau-1}-1}(\log t)^{-2}, \tag{7.4.11}$$

for t sufficiently large.

We wish to couple Inner_t with an Erdős-Rényi random graph with $n_t = \sqrt{t}$ vertices and edge probability p_t, which we denote by $\mathrm{ER}(n_t, p_t)$. For this, for $i, j \in \{1, 2, \ldots, n_t\}$, we say that an edge between i and j is present when there exists a t-connector connecting the ith and jth vertex in I. We now prove that this graph is bounded below by $\mathrm{ER}(n_t, p_t)$. Note that (7.4.11) does not guarantee this coupling,

instead we should prove that the lower bound holds uniformly, when i and j belong to I.

For this, we order the $n_t(n_t-1)/2$ edges in an arbitrary way, and bound the conditional probability that the lth edge is present conditionally on the previous edges from below by p_t, for every l. This would prove the claimed stochastic domination by $\mathrm{ER}(n_t, p_t)$.

Indeed, the lth edge is present precisely when there exists a t-connector connecting the corresponding vertices which we call i and j in I. Moreover, we shall not make use of the first vertices which were used to t-connect the previous edges. This removes at most $n_t(n_t-1)/2 \le t/2$ possible t-connectors, after which at least another $t/2$ remain. The probability that one of them is a t-connector for the ith and jth vertex in I is bounded below by, for t sufficiently large,

$$1-\exp\{-\eta t^{\frac{1}{\tau-1}-2}(\log t)^{-1}t/2\} = 1-\exp\{-\eta t^{\frac{1}{\tau-1}-1}(\log t)^{-1}/2\} \ge p_t \equiv t^{\frac{1}{\tau-1}-1}(\log t)^{-2}, \tag{7.4.12}$$

using $1-\mathrm{e}^{-x} \ge x/2$ for $x \in [0,1]$ and $\eta/2 \ge \log t^{-1}$ for t sufficiently large.

This proves the claimed stochastic domination of the random graph on the vertices I and $\mathrm{ER}(n_t, p_t)$. Next, we show that $\mathrm{diam}(\mathrm{ER}(n_t, p_t))$ is, whp, bounded by a uniform constant.

For this we use the result in [22, Corollary 10.12], which gives sharp bounds on the diameter of an Erdős-Rényi random graph. Indeed, this result implies that if $p^d n^{d-1} - 2\log n \to \infty$, while $p^{d-1}n^{d-2} - 2\log n \to -\infty$, then $\mathrm{diam}(\mathrm{ER}(n,p)) = d$, whp. In our case, $n = n_t = t^{1/2}$ and $p = p_t = t^{\frac{1}{\tau-1}-1}(\log t)^{-2}$, which implies that, whp, $\frac{\tau-1}{3-\tau} < d \le \frac{\tau-1}{3-\tau} + 1$. Thus, we obtain that the diameter of I in $\mathrm{PA}_{m,\delta}(2t)$ is whp bounded by $2(\frac{\tau-1}{3-\tau}+1)$ in this case.

We finally show that for any $i \in \mathrm{Inner}_t \setminus I$, the probability that there does not exist a t-connector connecting i and I is small. Indeed, this probability is, since $D_I(t) \ge \sqrt{t}t^{\frac{1}{\tau-1}}(\log t)^{-1/2}$, and $D_i(t) \ge t^{\frac{1}{2(\tau-1)}}(\log t)^{-1/2}$, the probability of there not existing a t-connector is bounded above by $\mathrm{e}^{-\eta t^{1/(\tau-1)-1/2}(\log t)^{-1}}$, which is tiny since $\tau < 3$. This proves that whpthe distance between any vertex $i \in \mathrm{Inner}_t \setminus I$ and I is bounded by 2, and, together with the fact that $\mathrm{diam}_{2t}(I) \le 2(\frac{\tau-1}{3-\tau}+1)$ thus implies that $\mathrm{diam}_{2t}(\mathrm{Inner}_t) \le 2(\frac{\tau-1}{3-\tau}+2)$. □

Proposition 7.23 (Distance between outer and inner core)**.** *Fix $m \ge 2$. With high probability, the inner core* Inner_t *can be reached from any vertex in the outer core* Outer_t *using no more than* $\frac{2\log\log t}{|\log(\tau-2)|}$ *edges in* $\mathrm{PA}_{m,\delta}(2t)$*. More precisely, whp*

$$\max_{i\in\mathrm{Outer}_t}\min_{j\in\mathrm{Inner}_t} \mathrm{dist}_{\mathrm{PA}_{m,\delta}(2t)}(i,j) \le \frac{2\log\log t}{|\log(\tau-2)|}. \tag{7.4.13}$$

Proof. Recall that

$$\mathrm{Outer}_t = \mathrm{Core}_t \setminus \mathrm{Inner}_t. \tag{7.4.14}$$

and define

$$\mathcal{N}_1 = \mathrm{Inner}_t = \{i : D_i(t) \ge \mathcal{W}_1\}, \tag{7.4.15}$$

where

$$\mathcal{W}_1 = l_t = t^{\frac{1}{2(\tau-1)}}(\log t)^{-\frac{1}{2}}. \tag{7.4.16}$$

We now recursively define a sequence u_k, for $k \geq 2$, so that for any vertex $i \in \{1, 2, \ldots, t\}$ with degree at least u_k, the probability that there is no t-connector for the vertex i and the set

$$\mathcal{N}_{k-1} = \{j : D_j(t) \geq \mathcal{W}_{k-1}\}, \tag{7.4.17}$$

conditionally on $\mathrm{PA}_t(m, \delta)$ is tiny. According to (7.4.10) and Exercise 8.18, this probability is at most

$$\exp\Big\{-\frac{\eta B t[u_{k-1}]^{2-\tau}u_k}{t}\Big\} = o(t^{-1}), \tag{7.4.18}$$

where we define

$$\mathcal{W}_k = D\log t\big(\mathcal{W}_{k-1}\big)^{\tau-2}, \tag{7.4.19}$$

with D exceeding $(\eta B)^{-1}$. By Lemma 5.15 we have

$$\mathcal{W}_k = D^{a_k}(\log t)^{b_k}t^{c_k}, \tag{7.4.20}$$

where

$$c_k = \frac{(\tau-2)^{k-1}}{2(\tau-1)}, \qquad b_k = \frac{1-(\tau-2)^{k-1}}{3-\tau} - \frac{1}{2}(\tau-2)^{k-1}, \qquad a_k = \frac{1-(\tau-2)^{k-1}}{3-\tau}. \tag{7.4.21}$$

Then, the key step in the proof of Proposition 7.23 is the following lemma:

Lemma 7.24 (Connectivity between $\mathcal{N}_{k-1}$ and $\mathcal{N}_k$)**.** *Fix $m, k \geq 2$. Then the probability that there exists an $i \in \mathcal{N}_k$ that is not at distance two from $\mathcal{N}_{k-1}$ in $\mathrm{PA}_{m,\delta}(2t)$ is $o(t^{-1})$.*

Proof. We note that, by [I, Exercise 8.18], with probability exceeding $1 - o(t^{-1})$, for all k,

$$\sum_{i\in\mathcal{N}_{k-1}} D_i(t) \geq Bt[\mathcal{W}_{k-1}]^{2-\tau}. \tag{7.4.22}$$

On the event that the bounds in (7.4.22) hold, we obtain by (7.4.10) that the conditional probability, given $\mathrm{PA}m, \delta(t)$, that there exists an $i \in \mathcal{N}_k$ such that there is no t-connector between i and $\mathcal{N}_{k-1}$ is bounded, using Boole's inequality, by

$$t\mathrm{e}^{-\eta B[\mathcal{W}_{k-1}]^{2-\tau}\mathcal{W}_k} = t\mathrm{e}^{-\eta BD\log t} = o(t^{-1}), \tag{7.4.23}$$

where we have used (7.4.19) and we have taken $D > 2(\eta B)^{-1}$. □

We now complete the proof of Proposition 7.23. Fix

$$k^* = \Big\lfloor \frac{\log\log t}{|\log(\tau-2)|} \Big\rfloor. \tag{7.4.24}$$

As a result of Lemma 7.24, we have that the distance between $\mathcal{N}_{k^*}$ and Inner_t is at most $2k^*$. Therefore, we are done when we can show that

$$\text{Outer}_t \subseteq \{i : D_i(t) \geq (\log t)^\sigma\} \subseteq \mathcal{N}_{k^*} = \{i : D_i(t) \geq \mathcal{W}_{k^*}\}, \tag{7.4.25}$$

so that it suffices to prove that $(\log t)^\sigma \geq \mathcal{W}_{k^*}$, for any $\sigma > \frac{1}{3-\tau}$. For this, we note that, by Lemma 5.15, we have that

$$\mathcal{W}_{k^*} = D^{a_{k^*}} (\log t)^{b_{k^*}} t^{c_{k^*}}. \tag{7.4.26}$$

We have that $t^{c_{k^*}} = O(1) = (\log t)^{o(1)}$, $(\log t)^{b_{k^*}} = (\log t)^{\frac{1}{3-\tau}+o(1)}$, and $D^{a_{k^*}} = (\log t)^{o(1)}$. Thus,

$$\mathcal{W}_{k^*} = (\log t)^{\frac{1}{3-\tau}+o(1)}, \tag{7.4.27}$$

so that, by picking t sufficiently large, we can make $\sigma \geq \frac{1}{3-\tau} + o(1)$. This completes the proof of Proposition 7.23. □

Proof of Theorem 7.21. We note that whp $\text{diam}_{2t}(\text{Core}_t) \leq C_\delta + 2k^*$, where k^* is the upper bound on $\max_{i \in \text{Outer}_t} \min_{j \in \text{Inner}_t} d_{\text{PA}_{m,\delta}(2t)}(i,j)$ in Proposition 7.23, and we have made use of Proposition 7.22. This proves Theorem 7.21. □

Proof of the upper bound in Theorem 7.10. □

7.4.3 Connecting the periphery to the core

In this section, we extend the results of the previous section and, in particular, study the distance between the vertices not in the core Core_n and the core. The main result in this section is the following theorem:

Theorem 7.25 (Connecting the periphery to the core)**.** *Fix $m \geq 2$. For every $\sigma > \frac{1}{3-\tau}$, whp, the maximal distance between any vertex and Core_t in $G_m(2t)$ is bounded from above by $2\sigma \log\log t / \log m$.*

Together with Theorem 7.21, Theorem 7.25 proves the main result in Theorem **??**.

The proof of Theorem 7.25 again consists of two key steps. The first in Proposition 7.26 states that, for any two vertices $i, j \in \{1, 2, \ldots, t\}$, the distance $d_{\text{PA}_{m,\delta}(2t)}(i,j)$ is bounded by a constant times $\log\log t$, i.e., $\text{diam}_{2t}(\text{PA}_t(m,\delta))$ is bounded by some constant times $\log\log t$. The second in Proposition 7.29 shows that the distance between any vertex in $\{t+1, t+2, \ldots, 2t\}$ and $\{1, 2, \ldots, t\}$ is bounded by another constant times $\log\log t$.

Proposition 7.26 (Connecting half of the periphery to the core)**.** *Fix $m \geq 2$. For every $\sigma > \frac{1}{3-\tau}$, whp, the maximal distance between any vertex in $\{1, 2, \ldots, t\}$ and the core Core_t in $\text{PA}_{m,\delta}(2t)$ is bounded from above by $\sigma \log\log t / \log m$.*

Proof. We start from a vertex $i \in \{1, 2, \ldots, t\}$ and will show that the probability that the distance between i and Core_t is at least $C \log\log t$ is $o(t^{-1})$ where $C = \sigma / \log m$. This proves the claim. For this, we explore the neighborhood of i as follows. From i, we connect its $m \geq 2$ edges. Then, successively, we connect the m edges from each of

the at most m vertices that i has connected to and have not yet been explored. We continue in the same fashion. We call the arising process when we have explored up to distance k from the initial vertex i the *k-exploration tree*.

When we never connect two edges to the same vertex, then the number of vertices we can reach within k steps is *precisely* equal to m^k. We call an event where an edge connects to a vertex which already was in the exploration tree a *collision*. When k increases, the probability of a collision increases. However, the probability that there exists a vertex for which *many* collisions occur in the k-exploration tree before it hits the core is small, as we prove now:

Lemma 7.27 (A bound on the probability of multiple collisions)**.** *Fix $m \geq 2$ and $\delta \in (-m, 0)$. Fix $C = \sigma/\log m$, $l \geq 1$, $b \in (0, 1]$ and take $k \leq C \log\log t$. Then, for every vertex $i \in \{1, 2, \ldots, t\}$, the probability that its k-exploration tree has at least l collisions before it hits $\mathrm{Core}_t \cup \{j : j \leq t^b\}$ is bounded above by*

$$\Big((\log t)^d t^{-b}\Big)^l = (\log t)^{dl} t^{-bl},$$

for some $d > 0$.

Proof. Take $i \in \{\lceil t^b \rceil + 1, \lceil t^b \rceil + 2, \ldots, t\}$ and consider its k-exploration tree $\mathcal{T}ki$. Since we add edges after time t^b the denominator in [I, (8.2.1)] is at least t^b. Moreover, before hitting the core, any vertex in the k-exploration tree has degree at most $(\log t)^\sigma$. Hence, for $l = 1$, the probability mentioned in the statement of the lemma is at most

$$\sum_{v \in \mathcal{T}ki} \frac{D_v(t) + \delta}{t^b} \leq \sum_{v \in \mathcal{T}ki} \frac{(\log t)^\sigma}{t^b} \leq \frac{m^k (\log t)^\sigma}{t^b} \tag{7.4.28}$$

where the bound follows from $\delta < 0$ and $|\mathcal{T}ki| \leq m^k$. For general l this upper bound becomes

$$\left(\frac{m^k (\log t)^\sigma}{t^b}\right)^l$$

When $k = C \log\log t$ with $C = \sigma/\log m$, we have that $m^{lk} = (\log t)^{l\sigma}$. Therefore, the claim in Lemma 7.27 holds with $d = 2\sigma$. □

Lemma 7.27 shall prove to be extremely useful, as it will imply that the shortest path graph from any vertex is, with high probability, close to a tree up to the moment when $\mathrm{Core}_t \cup \{j : j \leq t^b\}$ is hit. We recall that Core_t consists of the vertices with high degree, and it is reasonable to suspect that the early vertices are precisely the ones which have high degree. This suggests that $\{j : j \leq t^b\} \subseteq \mathrm{Core}_t$. We next prove that indeed, whp, $\{j : j \leq t^b\}$ is a subset of the core Core_t:

Lemma 7.28 (Early vertices have large degrees whp)**.** *Fix $m \geq 1$. There exists a $b > 0$ such that, whp, $\min_{j \leq t^b} d_j(t) \geq (\log t)^\sigma$, for some $\sigma > \frac{1}{3-\tau}$. As a result, whp, $\{j : j \leq t^b\} \subseteq \mathrm{Core}_t$.*

Proof. Note that, by Exercise 8.19, for all $m \geq 1$ and for $\delta < 0$,

$$\mathbb{P}(D_i(t) \leq j) \leq j \frac{\Gamma(mt)\Gamma\big(mi + \frac{m+\delta}{2m+\delta}\big)}{\Gamma\big(mt + \frac{m+\delta}{2m+\delta}\big)\Gamma(mi)}. \tag{7.4.29}$$

Indeed, note that by [I, (8.7.4)], and when $\delta \leq 0$, we have that $C_j \leq 1$. Thus, for $m = 1$, the bound in (7.4.29) follows in this case. Furthermore, to conclude (7.4.29) from Exercise 8.19 for $m \geq 2$, we note that, by the relation between $\mathrm{PA}_t(m, \delta)$ and $\mathrm{PA}_{1,\delta/m}(mt)$, the degree of vertex i in $\mathrm{PA}_t(m, \delta)$ is bounded from below by the degree of vertex im in $\mathrm{PA}_{1,\delta}(mt)$. As a result,

$$\mathbb{P}(D_i^{(m)}(t) \leq j) \leq \mathbb{P}(D_{im}^{(1)}(mt) \leq j), \tag{7.4.30}$$

after which the result follows from (7.4.29) for $m = 1$.

Below, we shall rely on the obvious consequence of (7.4.29) that

$$\mathbb{P}(D_i(t) \leq j) \leq j \frac{\Gamma(mt)\Gamma\big(mi + \frac{m+\delta}{2m+\delta}\big)}{\Gamma\big(mt + \frac{m+\delta}{2m+\delta}\big)\Gamma(mi)}. \tag{7.4.31}$$

Obviously, for t and i large, we have that

$$\mathbb{P}(D_i(t) \leq j) \leq j t^{-\frac{m+\delta}{2m+\delta}} i^{\frac{m+\delta}{2m+\delta}} (1 + o(1)). \tag{7.4.32}$$

We finally use (7.4.31) to complete the proof of Lemma 7.28. Take $0 < b < \frac{\frac{m+\delta}{2m+\delta}}{\frac{m+\delta}{2m+\delta}+1} = \frac{m+\delta}{3m+2\delta}$. Then, by Boole's inequality,

$$\begin{aligned}
\mathbb{P}(\exists i \leq t^b : D_i(t) \leq (\log t)^\sigma) &\leq \sum_{i=1}^{t^b} \mathbb{P}(d_i(t) \leq (\log t)^\sigma) \\
&\leq (\log t)^\sigma \frac{\Gamma(mt)}{\Gamma\big(mt + \frac{m+\delta}{2m+\delta}\big)} \sum_{i=1}^{t^b} \frac{\Gamma\big(mi + \frac{m+\delta}{2m+\delta}\big)}{\Gamma(mi)} \\
&\leq (\log t)^\sigma \big(\frac{m+\delta}{2m+\delta} + 1\big)^{-1} \frac{\Gamma(mt)}{\Gamma\big(mt + \frac{m+\delta}{2m+\delta}\big)} \frac{\Gamma\big(mt^b + \frac{m+\delta}{2m+\delta} + 1\big)}{\Gamma(mt^b)} \\
&= o(1),
\end{aligned} \tag{7.4.33}$$

by a similar equality as in (7.4.32). This completes the proof of Lemma 7.28. □

Now we are ready to complete the proof of Proposition 7.26:

Proof of Proposition 7.26. By combining Lemmas 7.27 and 7.28, the probability that there exists an $i \in \{1, 2, \ldots, t\}$ for which the exploration tree $\mathcal{T}ki$ has at least l collisions before hitting the core is $o(1)$, whenever $l > 1/b$, since, by Boole's inequality, it is bounded by

$$\sum_{i=1}^{t} (\log t)^{dl} t^{-bl} = (\log t)^{2\sigma l} t^{-bl+1} = o(1), \tag{7.4.34}$$

precisely when $l > \frac{1}{b}$. When the k-exploration tree hits the core, then we are done by Theorem 7.21. When the k-exploration tree from a vertex i does not hit the core, but has less than l collisions, then there are at least m^{k-l} vertices in k-exploration tree. Indeed, when there are at most l collisions, the minimal size of the tree is obtained by identifying at most l vertices and their complete offspring, and the size of the pruned tree has size at least m^{k-l}.

When $k = C \log\log t$ with $C = \sigma/\log m$, this number is at least equal to $(\log t)^{\sigma+o(1)}$. The total weight of the core is, by Exercise 8.18, at least

$$\sum_{i\in \mathrm{Core}_t} (D_i(t)+\delta) \geq Bt(\log t)^{-(\tau-2)\sigma}. \tag{7.4.35}$$

The probability that there does not exist a t-connector between the k-exploration tree and the core is, by (7.4.10) bounded above by

$$\exp\left\{-\frac{\eta Bt(\log t)^{-(\tau-2)\sigma}(\log t)^{\sigma+o(1)}}{t}\right\} = o(t^{-1}), \tag{7.4.36}$$

by picking B sufficiently large, since $\sigma > 1/(3-\tau)$. This completes the proof. □

Proposition 7.29. *Fix $m \geq 2$ and $\delta \in (-m, 0)$. For every $\sigma > \frac{1}{3-\tau}$, whp, the maximal distance between any vertex and $\mathrm{Core}_n \cup \{1, 2, \ldots, t\}$ in $\mathrm{PA}_{m,\delta}(2t)$ is bounded from above by $\frac{\sigma \log\log t}{\log m}$.*

Proof. Denote $k = \lfloor \frac{\sigma \log\log t}{\log m} \rfloor - 1$. We again grow the k-exploration trees from the vertices $i \in \{t+1, t+2, \ldots, 2t\}$.

By Lemma 7.27 for $b = 1$, the probability that there exists a vertex whose k-exploration tree contains at least two collisions before hitting the vertex set $\mathrm{Core}_t \cup \{1, 2, \ldots, t\}$ is bounded above by $t^{-2}(\log t)^{d_1}$ for some d_1 sufficiently large. When the k-exploration tree contains a vertex in $\mathrm{Core}_t \cup \{1, 2, \ldots, t\}$, then we are done by Proposition 7.26 and Theorem 7.21. If not, and there are at most 2 collisions, then there are at least $m_k = (m-1)m^{k-1}$ vertices in $\{t+1, t+2, \ldots, 2t\}$ at distance precisely equal to k from the original vertex. Denote these vertices by $i_1, \ldots, i_{m_k}$, and denote the k-exploration tree of vertex $i \in \{t+1, t+2, \ldots, 2t\}$ by $\mathcal{T}ki$. We write

$$\begin{aligned} &\mathbb{P}(\nexists j \in \{1, 2, \ldots, m_k\} \text{ such that } i_j \longrightarrow \{1, 2, \ldots, t\} \big| \mathcal{T}ki) \\ &= \prod_{j=1}^{m_k} \mathbb{P}\big(i_j \not\longrightarrow \{1, 2, \ldots, t\} \big| i_s \not\longrightarrow \{1, 2, \ldots, t\} \forall s < j, \mathcal{T}ki\big). \end{aligned} \tag{7.4.37}$$

Now we note that, uniformly in the way all edges in $\mathrm{PA}_{m,\delta}(2t)$ are formed, we have that for every $s \in \{t+1, t+2, \ldots, 2t\}$,

$$\frac{\sum_{i=1}^{t}(D_i(s)+\delta)}{(2m+\delta)s} \geq \frac{1}{2}. \tag{7.4.38}$$

Thus, for any vertex i_j in the boundary of $\mathcal{T}ki$, the probability that it will be directly connected to $\{1, 2, \ldots, t\}$ is at least $1/2$. As a result, we have that, uniformly in t, i

and j,

$$\mathbb{P}\big(i_j \not\longleftrightarrow \{1,2,\ldots,t\} \big| i_s \not\longleftrightarrow \{1,2,\ldots,t\} \forall s<j, \mathcal{T}ki\big) \leq \frac{(2m+\delta)t}{(2m+\delta)(2t)} = \frac{1}{2}. \tag{7.4.39}$$

Therefore, we obtain that

$$\mathbb{P}(\nexists j = 1,\ldots,m_k \text{ such that } i_j \longleftrightarrow \{1,2,\ldots,t\} | \mathcal{T}ki) \leq 2^{-m_k}. \tag{7.4.40}$$

Since $m_k = \frac{m-1}{m}(\log t)^\sigma$, with $\sigma = \frac{1}{3-\tau} > 1$, we have that $2^{-m_k} = o(t^{-1})$. Therefore, any vertex $i \in \{t+1, t+2, \ldots, 2t\}$ is, whp, within distance $k+1$ from $\{1,2,\ldots,t\}$. □

Proof of Theorem 7.25. Proposition 7.29 states that whpevery vertex in $\mathrm{PA}_{m,\delta}(2t)$ is within distance $k+1 = \lfloor \frac{\sigma \log\log t}{\log m} \rfloor$ of $\mathrm{Core}_t \cup \{1,2,\ldots,t\}$. Proposition 7.26 states that whpevery vertex in $\{1,2,\ldots,t\}$ is at most distance $k+1$ from the core Core_t. This shows that every vertex in $\mathrm{PA}_{m,\delta}(2t)$ is whpwithin distance $2(k+1)$ from the core. □

Proof of Theorem **??**. Theorem 7.25 states that every vertex in $\mathrm{PA}_{m,\delta}(2t)$ is within distance $\frac{2\sigma \log\log t}{\log m}$ of the core Core_t. Theorem 7.21 states that the diameter of the core is at most $\frac{4\log\log t}{|\log(\tau-2)|}(1+o(1))$, so that the diameter of $\mathrm{PA}_{m,\delta}(2t)$ is at most $C_G \log\log t$, where C_G is given in (**??**). This completes the proof of Theorem **??**. □

7.5 Notes and discussion

Notes on Section 7.1. Scale-free trees have received substantial attention in the literature, we refer to [25, 97] and the references therein. Theorem **??** is one of the main results in [97].

There is a beautiful result on the height of trees using branching processes due to Kingman [83], which Pittel [97] makes crucial use of. This approach is based on exponential martingales, and allows for a relatively short proof of the lower bound on the height of the tree.

Notes on Section 7.2. A weaker version of Theorem 7.10 is proved in [45]. The current theorem is inspired by [42].

Notes on Section 7.3.

Chapter 8
Optimal flows on random graphs

Abstract

In this chapter, we investigate the effect of independent and identically distributed edge weights on the metric structure of the graph. For this, we investigat the asymptotic properties of shortest-weight paths between vertices in complete and random graphs. We start by motivating the problem.

8.1 Motivation for shortest-weight routing

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 performed by shortest-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 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 shortest-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 minimal-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 8.1, which we repeat here. 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 [103] 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 [113].

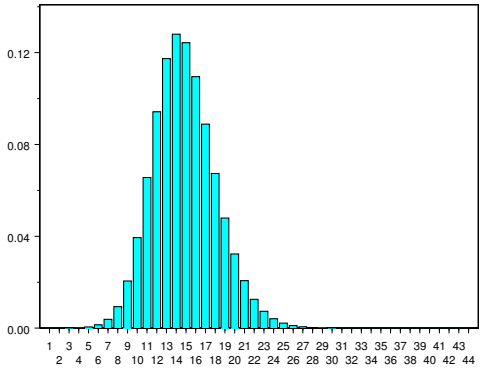

Figure 8.1: Internet hopcount data. Data courtesy of H. Tangmunarunkit.

First-passage percolation has many more 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, then first-passage percolation gives rise to a caricature model of an epidemic. When we think of two competing species trying to explore new territory, then first-passage percolation can be used to describe which species wins the most of the available territory.

We now formally introduce minimal-weight routing on graphs. Let $G = (V(G), E(G))$ be a graph with vertex set $V(G)$ and edge set $E(G)$. The graphs 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 G, where we assume that $Y_e > 0$ a.s. Then, for $i, j \in V(G)$, we let $\mathcal{C}_n(i, j)$ denote the minimal weight over all paths from i to j, i.e.,

$$\mathcal{C}_n(i,j) = \min_{\pi:\, i\to j} \sum_{e\in\pi} Y_e. \tag{8.1.1}$$

Denote the optimal path realizing the minimum in (8.1.1) 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

$$H_n(i,j) = |\pi_n(i,j)|, \tag{8.1.2}$$

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

$$\mathcal{C}_n = \mathcal{C}_n(U_1, U_2), \qquad H_n = H_n(U_1, U_2), \tag{8.1.3}$$

where U_1, U_2 are two vertices chosen uniformly at random, independent of each other. When the graph is disconnected, we will often condition on $H_n < \infty$. Further objects of interest are the *flooding time* $\max_{j\in V(G)} \mathcal{C}_n(i, j)$ and *weight diameter* $\max_{i,j\in V(G)} \mathcal{C}_n(i, j)$.

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. 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.

Organization of Chapter 8. This chapter is organised as follows. In Section 8.2, 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 8.3, we investigate the configuration model with exponential edge weights. In Section 8.4, we introduce continuous-time branching processes, which serve as a main tool in the remainder of this chapter. We continue in Section 8.5 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. We continue in Section 8.6 by describing related results on routing on random graphs. We close in Section 8.7 with notes and discussion.

8.2 Markovian flow 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}, $1 \le i < j \le n$, are i.i.d. exponential random variables with parameter 1, so that $\mathbb{P}(Y_{ij} > x) = \mathrm{e}^{-x}$ for $x \ge 0$. The main theorem of this section is a set of three different asymptotic results for $\mathcal{C}_n(i,j)$:

Theorem 8.1 (One, two and three times $\log n/n$)**.** *For the complete graphs with exponential edge weights, as $n \to \infty$*

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

$$\frac{\mathcal{C}_n(i,j)}{\log n/n} \xrightarrow{\mathbb{P}} 1; \tag{8.2.1}$$

(ii) *for any fixed i,*

$$\frac{\max_{j\le n} \mathcal{C}_n(i,j)}{\log n/n} \xrightarrow{\mathbb{P}} 2; \tag{8.2.2}$$

(iii)

$$\frac{\max_{i,j\le n} \mathcal{C}_n(i,j)}{\log n/n} \xrightarrow{\mathbb{P}} 3. \tag{8.2.3}$$

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 8.1(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 8.1:

Exercise 8.1 (Minimal weights from a given vertex)**.** *Let* $X_i = \min_{j\neq i} Y_{ij}$. *Show that, as* $n\to\infty$*:*

$$nX_i \xrightarrow{d} \mathsf{Exp}(1), \quad n^2 \min_{i\in[n]} X_i = n^2 \min_{i,j\in[n]:\, i\neq j} Y_{ij} \xrightarrow{d} \mathsf{Exp}(2), \quad \frac{\max_{i\in[n]} X_i}{\log n/n} \xrightarrow{\mathbb{P}} 1. \tag{8.2.4}$$

Using Exercise 8.1, we can give a simple informal interpretation of the three parts of Theorem 8.1 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 8.1. 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 $\mathsf{Exp}(1)$-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 $n-1$ neighbours of the initially infected vertex, the time T_1 until the second vertex is infected is exponentially distributed with expectation $1/(n-1)$. More generally, when $k < n$ vertices have been infected, there are $k(n-k)$ edges connecting the infected and non-infected vertices, and thus the time T_k until the next vertex is infected is $\mathsf{Exp}(k(n-k))$; 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

$$S_m = \sum_{k=1}^{m-1} T_k, \tag{8.2.5}$$

where $T_1,\dots,T_{m-1}$ are independent with $T_k \stackrel{d}{=} \mathsf{Exp}(k(n-k))$.

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,

$$\max_{j\in[n]} \mathcal{C}_n(1,j) = S_n = \sum_{k=1}^{n-1} T_k. \tag{8.2.6}$$

Hence, using that

$$\frac{1}{k(n-k)} = \frac{1}{n}\left(\frac{1}{k} + \frac{1}{n-k}\right), \tag{8.2.7}$$

we arrive at

$$\begin{aligned}\mathbb{E}[\max_{j\in[n]} \mathcal{C}_n(1,j)] &= \sum_{k=1}^{n-1} \mathbb{E}[T_k] = \sum_{k=1}^{n-1} \frac{1}{k(n-k)} = \frac{1}{n}\sum_{k=1}^{n-1}\left(\frac{1}{k} + \frac{1}{n-k}\right) \\ &= \frac{2}{n}\sum_{k=1}^{n-1}\frac{1}{k} = \frac{2\log n}{n} + O\left(1/n\right), \end{aligned} \tag{8.2.8}$$

and, similarly,

$$\begin{aligned}\mathrm{Var}(\max_{j\in[n]} \mathcal{C}_n(1,j)) &= \sum_{k=1}^{n-1} \mathrm{Var}(T_k) = \sum_{k=1}^{n-1}\left(\frac{1}{k(n-k)}\right)^2 \\ &\le 2\sum_{k=1}^{n/2} \frac{1}{k^2(n-k)^2} \le \frac{8}{n^2}\sum_{k=1}^{n/2}\frac{1}{k^2} = O(1/n^2). \end{aligned} \tag{8.2.9}$$

Part (ii) now follows by Chebyshev's inequality (see Theorem 2.16).

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

$$\mathcal{C}_n(1,2) = S_{N+1} = \sum_{k=1}^{N} T_k, \tag{8.2.10}$$

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

$$\mathcal{C}_n(1,2) = \sum_{k=1}^{n-1} \mathbb{1}_{\{N\ge k\}} T_k, \tag{8.2.11}$$

using indicator functions to eliminate the random summation limit. This rewrite allows us to compute the first moment of $\mathcal{C}_n(1,2)$ as

$$\begin{aligned}\mathbb{E}[\mathcal{C}_n(1,2)] &= \sum_{k=1}^{n-1} \mathbb{E}[\mathbb{1}_{\{N\ge k\}} T_k] = \sum_{k=1}^{n-1} \mathbb{P}(N \ge k)\mathbb{E}[T_k] \\ &= \sum_{k=1}^{n-1} \frac{n-k}{n-1}\frac{1}{k(n-k)} = \sum_{k=1}^{n-1}\frac{1}{k(n-1)} \\ &= \frac{\log n}{n} + O\left(1/n\right). \end{aligned} \tag{8.2.12}$$

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

$$\begin{aligned}\mathcal{C}_n(1,2) &= \sum_{k=1}^{N}(T_k - \mathbb{E}[T_k]) + \sum_{k=1}^{N}\frac{1}{n}\left(\frac{1}{k} + \frac{1}{n-k}\right) \\ &= \sum_{k=1}^{N}(T_k - \mathbb{E}[T_k]) + \frac{1}{n}\left(\log N + \log n - \log(n-N)\right) + O\left(1/n\right). \end{aligned} \tag{8.2.13}$$

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

$$\begin{aligned}\mathrm{Var}\left(\sum_{k=1}^{N}(T_k - \mathbb{E}[T_k])\right) &= \mathbb{E}\left[\left(\sum_{k=1}^{N}(T_k - \mathbb{E}[T_k])\right)^2\right] = \mathbb{E}\left[\sum_{k=1}^{N}\mathrm{Var}(T_k)\right] \\ &\le \sum_{k=1}^{n-1}\mathrm{Var}(T_k) = \sum_{k=1}^{n-1}\frac{1}{k^2(n-k)^2} \\ &\le \sum_{k=1}^{n/2}\frac{4}{k^2n^2} + \sum_{k=n/2}^{n-1}\frac{4}{n^2(n-k)^2} = O\left(1/n^2\right). \end{aligned} \tag{8.2.14}$$

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

$$\mathbb{E}[(\log N - \log(n-1))^2] = \mathbb{E}\left[\left(\log\frac{N}{n-1}\right)^2\right] \to \int_0^1 (\log x)^2 dx < \infty. \tag{8.2.15}$$

Hence $\mathrm{Var}(\log N) = \mathrm{Var}(\log(n-N)) = O(1)$, and it follows that the variance of the second term in (8.2.13) is also $O(1/n^2)$. The same is trivially true for the third term. Consequently, $\mathrm{Var}(\mathcal{C}_n(1,2)) = O(1/n^2)$, which together with (8.2.12) 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 (8.2.6), for $-\infty \le t < 1 - 1/n$.

$$\mathbb{E}[\mathrm{e}^{nt\max_{j\in[n]}\mathcal{C}_n(1,j)}] = \prod_{k=1}^{n-1}\mathbb{E}[\mathrm{e}^{ntT_k}] = \prod_{k=1}^{n-1}\left(1 - \frac{nt}{k(n-k)}\right)^{-1}. \tag{8.2.16}$$

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

$$\begin{aligned}\mathbb{P}(n\max_{j\in[n]}\mathcal{C}_n(1,j) > a\log n) &\le \mathbb{E}[\mathrm{e}^{nt\max_{j\in[n]}\mathcal{C}_n(1,j) - ta\log n}] \\ &= \mathrm{e}^{-ta\log n}\prod_{k=1}^{n-1}\left(1 - \frac{nt}{k(n-k)}\right)^{-1} \\ &= \left(1 - \frac{nt}{n-1}\right)^{-2}\exp\Big(-ta\log n - \sum_{k=2}^{n-2}\log\left(1 - \frac{nt}{k(n-k)}\right)\Big). \end{aligned} \tag{8.2.17}$$

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

$$\begin{aligned}&\mathbb{P}(n\max_{j\in[n]}\mathcal{C}_n(1,j) > a\log n) \\ &\le \left(1 - \frac{nt}{n-1}\right)^{-2}\exp\Big(-ta\log n + \sum_{k=2}^{n-2}\Big(\frac{nt}{k(n-k)} + \Big(\frac{nt}{k(n-k)}\Big)^2\Big)\Big) \\ &= (1 - t + O(1/n))^{-2}\exp(-ta\log n + 2t\log n + O(1)) = O(n^{2-a}\log^2 n). \end{aligned} \tag{8.2.18}$$

This evidently implies that

$$\begin{aligned}\mathbb{P}(\max_{i,j\in[n]} \mathcal{C}_n(i,j) > a\log n/n) &= \mathbb{P}(\max_i \max_{j\in[n]} \mathcal{C}_n(i,j) > a\log n/n) \\ &\le n\mathbb{P}(\max_{j\in[n]} \mathcal{C}_n(1,j) > a\log n/n) = O(n^{3-a}\log^2 n),\end{aligned} \tag{8.2.19}$$

which tends to 0 as $n\to\infty$, 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 whpthere 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 $X_i \ge (1-\varepsilon)\log n/n$.

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

$$\mathcal{N} = \{i\colon X_i \ge (1-\varepsilon)\log n/n\}, \tag{8.2.20}$$

where $X_i = \min_{j\neq i} Y_{ij}$ is the minimal edge weight from vertex i as studied in Exercise 8.1.

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

$$\mathbb{E}[|\mathcal{N}|] = n\mathbb{P}(nX_1 \ge (1-\varepsilon)\log n) = n\mathrm{e}^{-(1-\varepsilon)\frac{n-1}{n}\log n} = n^{\varepsilon}(1+o(1)). \tag{8.2.21}$$

Further

$$\mathbb{E}[|\mathcal{N}|(|\mathcal{N}|-1)] = n(n-1)\mathbb{P}\big(nX_1 \ge (1-\varepsilon)\log n, nX_2 \ge (1-\varepsilon)\log n\big), \tag{8.2.22}$$

where

$$\mathbb{P}\big(nX_1 \ge (1-\varepsilon)\log n, nX_2 \ge (1-\varepsilon)\log n\big) = \mathrm{e}^{-(1-\varepsilon)\frac{(2n-3)}{n}\log n}. \tag{8.2.23}$$

Therefore,

$$\mathbb{E}[|\mathcal{N}|(|\mathcal{N}|-1)] \le \mathbb{E}[|\mathcal{N}|]^2 \mathrm{e}^{(1-\varepsilon)\log n/n}. \tag{8.2.24}$$

We conclude that $\mathrm{Var}(|\mathcal{N}|) = o\big(\mathbb{E}[|\mathcal{N}|]^2\big)$, so that by the Chebychev inequality (Theorem 2.16),

$$n^{-\varepsilon}|\mathcal{N}| \xrightarrow{\mathbb{P}} 1. \tag{8.2.25}$$

This completes the first step in the proof.

We next show that $n\mathcal{C}_n(i,j) \ge 3(1-\varepsilon)\log n$ for most pairs of vertices $i,j\in\mathcal{N}$. For this, we let

$$\mathcal{P} = \{(i,j)\in\mathcal{N}^2\colon nY_{ij} \ge 2(1-\varepsilon)\log n, n\mathcal{C}_n(i,j) \ge 3(1-\varepsilon)\log n\} \tag{8.2.26}$$

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

$$0 \leq |\mathcal{N}|(|\mathcal{N}|-1) - |\mathcal{P}| \leq N_1 + N_2, \tag{8.2.27}$$

where

$$N_1 = |\{(i,j) \in \mathcal{N}^2 \colon nY_{ij} < 2(1-\varepsilon)\log n\}|, \tag{8.2.28}$$
$$N_2 = |\{(i,j) \in \mathcal{N}^2 \colon nY_{ij} \geq 2(1-\varepsilon)\log n, n\mathcal{C}_n(i,j) < 3(1-\varepsilon)\log n\}|. \tag{8.2.29}$$

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

$$\begin{aligned}\mathbb{E}[N_1] &= \mathbb{E}|\{(i,j) \in \mathcal{N}^2 \colon nY_{ij} < 2(1-\varepsilon)\log n\}| \\ &\leq n(n-1)[1 - \mathrm{e}^{-2(1-\varepsilon)\log n/n}]\mathrm{e}^{-2(1-\varepsilon)\frac{n-1}{n}\log n} = o(n^{2\varepsilon}),\end{aligned} \tag{8.2.30}$$

so that by the Markov inequality (Theorem 2.15) indeed $n^{-2\varepsilon}N_1 \xrightarrow{\mathbb{P}} 0$. For N_2, we compute

$$\mathbb{E}[N_2] = n(n-1)\mathbb{P}(1,2 \in \mathcal{N}, nY_{12} \geq 2(1-\varepsilon)\log n, n\mathcal{C}_n(1,2) < 3(1-\varepsilon)\log n) \tag{8.2.31}$$
$$\begin{aligned}&= n(n-1)\mathbb{P}(1,2 \in \mathcal{N}, nY_{12} \geq 2(1-\varepsilon)\log n) \\ &\quad \times \mathbb{P}(n\mathcal{C}_n(1,2) < 3(1-\varepsilon)\log n \mid 1,2 \in \mathcal{N}, nY_{12} \geq 2(1-\varepsilon)\log n).\end{aligned}$$

A minor adaptation of (8.2.23) yields that

$$n^{2(1-\varepsilon)}\mathbb{P}(1,2 \in \mathcal{N}, nY_{12} \geq 2(1-\varepsilon)\log n) \to 1, \tag{8.2.32}$$

so that

$$\mathbb{E}[N_2] = n^{2\varepsilon}\mathbb{P}(n\mathcal{C}_n(1,2) < 3(1-\varepsilon)\log n \mid 1,2 \in \mathcal{N}, nY_{12} \geq 2(1-\varepsilon)\log n)(1+o(1)). \tag{8.2.33}$$

We are left to investgate the conditional probability. For this, we note that conditionally on $Y_{ij} > x$, the distribution of $Y_{ij} - x$ is again $\mathsf{Exp}(1)$. 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 $\mathsf{Exp}(1)$ 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 $\mathsf{Exp}(1)$ 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

$$\sum_{e\in\pi} E_e + 2(1-\varepsilon)\log n/n, \tag{8.2.34}$$

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

$$\begin{aligned}\mathbb{P}(n\mathcal{C}_n(1,2) < 3(1-\varepsilon)\log n \mid 1,2 \in \mathcal{N}, nY_{12} \geq 2(1-\varepsilon)\log n) \\ = \mathbb{P}(n\mathcal{C}_n(1,2) < (1-\varepsilon)\log n) = o(1),\end{aligned} \tag{8.2.35}$$

by part (i). This completes the lower bound in part (iii). □

Exercise 8.2 (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 (8.2.5) and (8.2.9) also yields their asymptotic distributions:

Theorem 8.2 (Asymptotic distributions of minimal weights)**.** *As* $n \to \infty$*, for every* $i \neq j$ *with* $i,j \in [n]$ *fixed,*

$$n\mathcal{C}_n(i,j) - \log n \xrightarrow{d} \Lambda_1 + \Lambda_2 - \Lambda_3, \tag{8.2.36}$$

and, for every $i \in [n]$ *fixed,*

$$n \max_{j\in[n]} \mathcal{C}_n(i,j) - 2\log n \xrightarrow{d} \Lambda_1 + \Lambda_2, \tag{8.2.37}$$

where $\Lambda_1, \Lambda_2, \Lambda_3$ *are independent Gumbel random variables, i.e.,* $\mathbb{P}(\Lambda_i \leq x) = \mathrm{e}^{-\mathrm{e}^{-x}}$ *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 $n \to \infty$. Equation (8.2.10) implies that with $(E_i)_{i\geq 1}$ denoting independent $\mathsf{Exp}(1)$ random variables and N a discrete uniform random variable in the set $[n-1]$ independent of $(E_i)_{i\geq 1}$,

$$\begin{aligned} n\mathcal{C}_n(i,j) &\stackrel{d}{=} \sum_{k=1}^{N} \frac{n}{k(n-k)} E_k = \sum_{k=1}^{N} \frac{n}{k(n-k)}(E_k - 1) + \sum_{k=1}^{N} \left(\frac{1}{k} + \frac{1}{n-k}\right) \\ &\approx \sum_{k=1}^{N} (E_k - 1)/k + \log N + \gamma + \log n - \log(n-N) \\ &\approx \sum_{k=1}^{\infty} (E_k - 1)/k + \log\left(\frac{N/n}{1-N/n}\right) + \log n + \gamma, \end{aligned} \tag{8.2.38}$$

where γ is the Euler-Mascheroni constant.

Similarly, with E'_k denoting i.i.d. exponentials with mean 1 independent of $(E_k)_{k\geq 1}$, (8.2.6) implies that

$$\begin{aligned} n\mathsf{Flood}_n(i) &\stackrel{d}{=} \sum_{k=1}^{n-1} \frac{n}{k(n-k)} E_k = \sum_{k=1}^{n-1} \frac{n}{k(n-k)}(E_k - 1) + 2\sum_{1}^{n-1} \frac{1}{k} \\ &\approx \sum_{k=1}^{\lfloor n/2 \rfloor} (E_k - 1)/k + \sum_{k=\lfloor n/2 \rfloor + 1}^{n-1} \frac{1}{n-k}(E_k - 1) + 2\log n + 2\gamma \\ &\stackrel{d}{=} \sum_{k=1}^{\lfloor n/2 \rfloor} (E_k - 1)/k + \sum_{k=1}^{\lceil n/2 \rceil - 1} (E'_k - 1)/k + 2\log n + 2\gamma. \end{aligned} \tag{8.2.39}$$

Since $N/n \xrightarrow{d} U$, where U has a uniform distribution on $[0,1]$, the random variable $L = \log\left(U/(1-U)\right)$ has the logistic distribution, i.e., for every $x \in \mathbb{R}$,

$$\mathbb{P}(L \leq x) = \mathbb{P}\big(\log\big(U/(1-U)\big) \leq x\big) = \mathbb{P}\big(U \leq \mathrm{e}^x/(1+\mathrm{e}^x)\big) = \mathrm{e}^x/(1+\mathrm{e}^x). \tag{8.2.40}$$

Therefore,

$$n\mathcal{C}_n(i,j) - \log n \xrightarrow{d} \sum_{k=1}^{\infty} (E_k - 1)/k + \gamma + L, \tag{8.2.41}$$

and

$$n \max_{j\in[n]} \mathcal{C}_n(i,j) - 2\log n \xrightarrow{d} \sum_{k=1}^{\infty} (E_k - 1)/k + \sum_{k=1}^{\infty} (E'_k - 1)/k + 2\gamma. \tag{8.2.42}$$

We are left to show that

$$\Lambda = \sum_{k=1}^{\infty} (E_k - 1)/k + \gamma \tag{8.2.43}$$

has a Gumbel distribution, and that L has the same distribution as $\Lambda_1 - \Lambda_2$.

We see that

$$\Lambda = \sum_{k=1}^{\infty} \frac{1}{k}(E_k - 1) + \gamma = \lim_{n\to\infty} \sum_{k=1}^{n} E_k/k - \log n. \tag{8.2.44}$$

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 8.3 (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 $\big(\sum_{k=n}^{n-i+1} E_k/k\big)_{i\in[n]}$. Conclude that $E_{(n:n)} = \max_{i\in[n]} E_i \stackrel{d}{=} \sum_{k=1}^{n} E_k/k$.*

By Exercise 8.3 and (8.2.44),

$$\begin{aligned}\mathbb{P}(\Lambda \le x) &= \lim_{n\to\infty} \mathbb{P}(\max_{i=1}^{n} E_i - \log n \le x) = \lim_{n\to\infty} \mathbb{P}(E_1 \le x + \log n)^n \\ &= \lim_{n\to\infty} [1 - \mathrm{e}^{-(x+\log n)}]^n = [1 - \mathrm{e}^{-x}/n]^n = \mathrm{e}^{-\mathrm{e}^{-x}},\end{aligned} \tag{8.2.45}$$

so that Λ indeed has the Gumbel distribution. Further, we compute the distribution function of $\Lambda_1 - \Lambda_2$ as

$$\mathbb{P}(\Lambda_1 - \Lambda_2 \le x) = \mathbb{E}[\mathbb{P}(\Lambda_1 \le x + \Lambda_2 \mid \Lambda_2)] = \mathbb{E}[\mathrm{e}^{-\mathrm{e}^{-(x+\Lambda_2)}}] = \mathbb{E}[\mathrm{e}^{-\mathrm{e}^{-\Lambda_2}\mathrm{e}^{-x}}]. \tag{8.2.46}$$

When Λ has a Gumbel distribution, $E = \mathrm{e}^{-\Lambda}$ has an exponential distribution (see Exercise 8.4), so that

$$\mathbb{P}(\Lambda_1 - \Lambda_2 \le x) = \mathbb{E}[\mathrm{e}^{-E\mathrm{e}^{-x}}] = (1 + \mathrm{e}^{-x})^{-1} = \mathrm{e}^{x}/(1 + \mathrm{e}^{x}). \tag{8.2.47}$$

This completes the proof. □

Exercise 8.4 (Distribution of $\mathrm{e}^{-\Lambda}$)**.** *Let Λ have a Gumbel distribution. Show that $E = \mathrm{e}^{-\Lambda}$ has an exponential distribution with parameter 1.*

Exercise 8.5 (Integral representation of Euler-Mascheroni)**.** *Use the fact that*

$$\sum_{k=1}^{\infty} \frac{1}{k}(E_k - 1) + \gamma \tag{8.2.48}$$

has a Gumbel distribution to conclude the following integral representation for the Euler-Mascheroni constant γ:

$$\gamma = \int_{-\infty}^{\infty} x\mathrm{e}^{-x}\mathrm{e}^{-\mathrm{e}^{-x}}dx = \int_{0}^{\infty} \mathrm{e}^{-y}\log y dy. \tag{8.2.49}$$

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 (8.1.3) 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 8.3 (Central limit theorem for the hopcount)**.** *Consider first-passage percolation on the complete graph K_n with i.i.d. exponential edge weights. As $n \to \infty$, for any fixed $i, j \in [n]$,*

$$\frac{H_n(i,j) - \log n}{\sqrt{\log n}} \xrightarrow{d} Z, \tag{8.2.50}$$

where Z is standard normal.

If we compare Theorem 8.1 and 8.3, 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 8.1, where the empirical mean and variance of the number of edges in the minimal-weight paths are close.

Proof. The proof of Theorem 8.1 shows that the collection of minimal weight paths from a given vertex, 1 say, form a 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. 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 8.4 (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

$$\mathbb{P}(H_n = k) = \frac{1}{n}\mathbb{P}(H_n = k \mid V_n = n) + (1 - \frac{1}{n})\mathbb{P}(H_n = k \mid V_n \neq n). \tag{8.2.51}$$

Now, $\mathbb{P}(H_n = k \mid V_n \neq n) = \mathbb{P}(H_{n-1} = k)$, since conditonally 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

$$\mathbb{P}(H_n = k) = \frac{1}{n}\mathbb{P}(H_{n-1} = k) + (1 - \frac{1}{n})\mathbb{P}(H_{n-1} = k-1). \tag{8.2.52}$$

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 $1/n$. This advances the induction hypothesis, and completes the proof. □

Theorem 8.3 follows immediately from Lemma 8.4, together with the fact that $(\sum_{i=2}^n I_i - \log n)/\sqrt{\log n} \xrightarrow{d} Z$, where Z is standard normal (see also Exercise 8.8). □

Exercise 8.6 (Mean hopcount)**.** *Show that $\mathbb{E}[H_n] = \sum_{k=2}^n 1/k$, which is the harmonic series.*

Exercise 8.7 (Variance hopcount)**.** *Show that $\mathrm{Var}(H_n) = \sum_{k=2}^n 1/k - \sum_{k=2}^n 1/k^2 = \mathbb{E}[H_n] - \zeta(2) + O(1/n)$, where $\zeta(2) = \sum_{k=1}^\infty 1/k^2$ is equal to the zeta-function at 2.*

Exercise 8.8 (CLT for sums of independent indicators). *Let $(I_i)_{i\geq 1}$ be a collection of independent $\mathsf{Be}(p_i)$ random variables, and assume that*

$$\sum_{i=1}^{m} p_i \to \infty, \qquad \sum_{i=1}^{m} p_i^2 = o\Big(\sum_{i=1}^{m} p_i\Big). \tag{8.2.53}$$

Show that, as $m \to \infty$,

$$\frac{\sum_{i=1}^{m}(I_i - p_i)}{\sqrt{\sum_{i=1}^{m} p_i}} \xrightarrow{d} Z, \tag{8.2.54}$$

where Z is standard normal. Use this to show that $(H_n - \log n)/\sqrt{\log n} \xrightarrow{d} Z$, where H_n be the height in a random recursive tree of a uniformly chosen vertex in $[n]$.

Exercise 8.9 (Joint convergence). *Show that the convergence of $H_n(i,j)$ and $\mathcal{C}_n(i,j)$ in Theorems 8.2 and 8.3, respectively, occurs* jointly, *i.e.,*

$$\Big(\frac{H_n(i,j) - \log n}{\sqrt{\log n}}, n\mathcal{C}_n(i,j) - \log n\Big) \xrightarrow{d} (Z, \Lambda_1 + \Lambda_2 - \Lambda_3), \tag{8.2.55}$$

where Z is standard normal and $\Lambda_1, \Lambda_2, \Lambda_3$ are three independent Gumbel variables independent of Z.

8.3 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 $\mathrm{CM}_n(\boldsymbol{d})$ where the degree distribution has finite variance. Our main result of this section is the following theorem:

Theorem 8.5 (Hopcount and minimal path weight for CM with exponential edge weights). *Let $\mathrm{CM}_n(\boldsymbol{d})$ satisfy [I, Condition 7.7(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*

$$\lim_{n\to\infty} \mathbb{E}[D_n^{2+\varepsilon}] = \mathbb{E}[D^{2+\varepsilon}]. \tag{8.3.1}$$

Then, there exists a random variable Q such that, conditionally on $\mathcal{C}_n < \infty$,

$$\frac{H_n - \alpha \log n}{\sqrt{\alpha \log n}} \xrightarrow{d} Z, \qquad \mathcal{C}_n - \frac{1}{\lambda}\log n \xrightarrow{d} Q, \tag{8.3.2}$$

where Z has a standard normal distribution, and

$$\alpha = \nu/(\nu-1), \qquad \lambda = \nu - 1. \tag{8.3.3}$$

By Theorem 5.1, we know that the graph distance $d_{\mathrm{CM}_n(\boldsymbol{d})}(U_1, U_2)$ between two uniformly chosen vertices satisfies

$$d_{\mathrm{CM}_n(\boldsymbol{d})}(U_1, U_2)/\log n \xrightarrow{\mathbb{P}} 1/\log \nu. \tag{8.3.4}$$

By Theorem 8.5, $H_n/\log n \xrightarrow{\mathbb{P}} \alpha = \nu/(\nu-1) > 1/\log\nu$, so that the minimal-weight path is substantially longer than the shortest path realizing the graph distance. In turn,

$$\mathcal{C}_n/\log n \xrightarrow{\mathbb{P}} 1/(\nu-1) < 1/\log\nu, \tag{8.3.5}$$

while the expected weight of any shortest path has the same asymptotics as in (8.3.4), so that the minimal-weight path has a significantly smaller weight. Thus, the topology of the configuration model $\mathrm{CM}_n(\boldsymbol{d})$ 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 8.5. By Theorem 5.28, the centered graph distance $d_{\mathrm{CM}_n(\boldsymbol{d})}(U_1,U_2) - \log n/\log\nu$ is a tight sequence of random variables that does *not* converge in distribution. The latter is due to the discrete nature of $d_{\mathrm{CM}_n(\boldsymbol{d})}(U_1,U_2)$. In Theorem 8.5, 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. Further, the centered shortest-weight $\mathcal{C}_n - \log n/\lambda$, which is a continuous random variables and thus does not suffer from the discreteness as $d_{\mathrm{CM}_n(\boldsymbol{d})}(U_1,U_2)$, converges in distribution. In order to intuitively understand this asymptotics, we see that, for first-passage percolation on $\mathrm{CM}_n(\boldsymbol{d})$, the number of vertices that can be reached with weight at most t grows proportional to $\mathrm{e}^{\lambda t}$. As a result, when we draw an alive vertex uniformly at random from all the alive vertices at time t, then its lifetime 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 (8.3.2) can be identified as

$$Q = Q_1 + Q_2 - \frac{\Lambda}{\nu-1} + \frac{\log\mu(\nu-1)}{\nu-1}, \tag{8.3.6}$$

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 now start by describing the key steps in the proof of Theorem 8.5:

Local structure first-passage percolation on $\mathrm{CM}_n(\boldsymbol{d})$. Proposition 5.4 shows that the local structure around a uniform vertex in $\mathrm{CM}_n(\boldsymbol{d})$ is close to a two-stage branching process, where the root has offspring distribution D appearing in [I, Condition 7.7(a)], while all other individuals have offspring $D^\star - 1$, where $D^\star$ 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 8.1, 8.2 and 8.3. We start by describng this construction on *determinstic trees* where the ith explored vertex has degree d_i. Later, we shall take $d_1 = D$, and $d_i = D_i^\star - 1$ 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 $d_1, d_2, \ldots$. Consider the following construction of a branching tree:

Construction 8.6 (Flow from root of tree and split-times)**.** *The shortest-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 mth vertex dies, it leaves behind d_m alive offspring.

In Construction 8.6, 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 $s_i = d_1 + \cdots + d_i - (i-1) \geq 1$ for every $i \geq 1$:

Exercise 8.10 (The number of alive vertices)**.** *Show that the number of alive vertices after the ith step in Construction 8.6 is equal to*

$$s_i = d_1 + \cdots + d_i - (i-1). \tag{8.3.7}$$

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 8.6 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 8.6 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. We continue to investigate the *generation* and *weight* to the mth chosen alive vertex in our first-passage percolation problem on the tree:

Proposition 8.7 (Shortest-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 minimal-weight path between the root of the tree and the vertex chosen at time m is equal in distribution to

$$T_m \stackrel{d}{=} \sum_{i=1}^{m} E_i/s_i, \tag{8.3.8}$$

where $(E_i)_{i\geq 1}$ are i.i.d. exponential random variables with mean 1.
(b) the generation of the vertex chosen at time m is equal in distribution to

$$G_m \stackrel{d}{=} \sum_{i=1}^{m} I_i, \tag{8.3.9}$$

where $(I_i)_{i\geq 1}$ *are independent Bernoulli random variables with*

$$\mathbb{P}(I_i = 1) = d_i/s_i, \tag{8.3.10}$$

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 $(T_i)_{i=1}^{m-1}$,

$$T_m - T_{m-1} \stackrel{d}{=} E_m/s_m, \tag{8.3.11}$$

where E_m is independent from $(T_i)_{i=1}^{m-1}$. This proves the claim for 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 (8.3.9) holds for m, where $(I_i)_{i=1}^{m}$ are independent Bernoulli random variables satisfying (8.3.10). We then advance the induction hypothesis to $m + 1$, by showing that G_{m+1} has the distribution in (8.3.9).

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 $1 \leq k \leq m$,

$$\mathbb{P}(G_{m+1} = k) = \frac{d_{m+1}}{s_{m+1}}\mathbb{P}(G_m = k - 1) + \Big(1 - \frac{d_{m+1}}{s_{m+1}}\Big)\mathbb{P}(G_m = k). \tag{8.3.12}$$

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. □

Exercise 8.11 (Gumbel law for deterministic trees)**.** *Let* $d_1 = r - 2$ *and* $d_i = r - 1$ *for every* $i \geq 2$*. Use Proposition 8.7 to conclude that*

$$T_m - \frac{\log m}{r-2} \xrightarrow{d} \Lambda/(r-2), \tag{8.3.13}$$

where Λ *has a Gumbel distribution.*

Exercise 8.12 (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 8.7 to conclude that*

$$\frac{G_m - \frac{r-1}{r-2}\log m}{\sqrt{\frac{r-1}{r-2}\log m}} \xrightarrow{d} Z, \tag{8.3.14}$$

where Z *has a standard normal distribution.*

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 B_i denote the *forward degree* (i.e., the degree minus 1) of the vertex incident to the half-edge to which the ith half-edge is paired. By the results in Chapter 4, see in particular Proposition 5.4, $(B_i)_{i\geq 2}$ are close to being i.i.d., and have distribution given by $D^\star - 1$. Therefore, we are lead to studying random variables of the form (8.3.9)–(8.3.10), where $(B_i)_{i\geq 1}$ are independent random variables with $B_1 \stackrel{d}{=} D$ and $B_i \stackrel{d}{=} D^\star - 1$. Thus, we study the *unconditional* law of G_m in (8.3.9), in the setting where the vector $(d_i)_{i\geq 1}$ is replaced by a sequence of independent random variables $(B_i)_{i\geq 1}$. 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 $S_i^{(\mathrm{ind})} = B_1 + \cdots + B_i - (i-1)$, where $B_1 \stackrel{d}{=} D$ and $B_i \stackrel{d}{=} D_i^\star - 1$, and $(B_i)_{i\geq 1}$ are independent. Then, define

$$T_m = \sum_{i=1}^{m} E_i/S_i^{(\mathrm{ind})}, \qquad G_m \stackrel{d}{=} \sum_{i=1}^{m} I_i, \tag{8.3.15}$$

where, conditionally on $(S_i^{(\mathrm{ind})})_{i\geq 1}$, $(I_i)_{i\geq 1}$ are independent Bernoulli random variables with

$$\mathbb{P}(I_i = 1) = B_i/S_i^{(\mathrm{ind})}. \tag{8.3.16}$$

Corollary 8.8 (Asymptotics for shortest-weight paths on trees)**.** *Let $(B_i)_{i\geq 1}$ be a sequence of independent non-degenerate integer-valued random variables satisfying $\mathbb{E}[B_i^{1+\varepsilon}] < \infty$ for some $\varepsilon > 0$. Denote $\nu = \mathbb{E}[B_1]$. Then, conditionally on $S_i^{(\mathrm{ind})} = B_1 + \cdots + B_i - (i-1) \geq 1$ for all $i \geq 1$,*
(a) for T_m given by (8.3.15), there exists a random variable X such that

$$T_m - \frac{1}{\lambda}\log m \xrightarrow{d} X, \tag{8.3.17}$$

where $\lambda = \nu - 1$.
(b) for G_m given in (8.3.15)–(8.3.16), with $\alpha = \nu/(\nu-1) \geq 1$, as $m \to \infty$,

$$\frac{G_m - \alpha \log m}{\sqrt{\alpha \log m}} \xrightarrow{d} Z, \tag{8.3.18}$$

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 (8.3.8), and rewrite

$$T_m = \sum_{i=1}^{m} (E_i - 1)/S_i^{(\mathrm{ind})} + \sum_{i=1}^{m} \Big(\frac{1}{S_i^{(\mathrm{ind})}} - \frac{1}{(\nu-1)i}\Big) + \frac{1}{\nu-1}\sum_{i=1}^{m} 1/i. \tag{8.3.19}$$

On the event that $S_i^{(\mathrm{ind})} \geq 1$ for every $i \geq 1$, by the Strong Law of Large Numbers,

$$S_i^{(\mathrm{ind})}/i \xrightarrow{a.s.} \nu - 1, \tag{8.3.20}$$

As a result, there are a.s. only finitely many values for which $S_i \leq (1-\varepsilon)(\nu-1)i$. Denote $M_m = \sum_{i=1}^m (E_i - 1)/S_i^{(\mathrm{ind})}$, and note that, conditionally on $(S_i^{(\mathrm{ind})})_{i\geq 0}$, $(M_m)_{m\geq 1}$ is a martingale. with conditional second moment

$$\mathbb{E}[M_m^2 \mid (S_i^{(\mathrm{ind})})_{i\geq 0}] = \sum_{m=1}^{\infty} 1/(S_i^{(\mathrm{ind})})^2, \tag{8.3.21}$$

which is a.s. bounded on the event that $S_i^{(\mathrm{ind})} \geq 1$ for every $i \geq 1$ and that $S_i^{(\mathrm{ind})} \leq (1-\varepsilon)(\nu-1)i$ except for finitely many i. As a result, for a.e. $(S_i^{(\mathrm{ind})})_{i\geq 0}$, M_n and converges a.s. to its limit, which equals $M_\infty = \sum_{i=1}^{\infty}(E_i - 1)/S_i^{(\mathrm{ind})}$. For the second term, we can rewrite

$$\sum_{i=1}^{m} \Big(\frac{1}{S_i^{(\mathrm{ind})}} - \frac{1}{(\nu-1)i}\Big) = \sum_{i=1}^{m} \Big(\frac{1}{S_i^{(\mathrm{ind})}} - \frac{1}{\mathbb{E}[S_i^{(\mathrm{ind})}]}\Big) + \sum_{i=1}^{\infty} \Big(\frac{1}{(\mu + (\nu-1)(i-1)} - \frac{1}{(\nu-1)i}\Big), \tag{8.3.22}$$

and

$$\sum_{i=1}^{\infty} \Big(\frac{1}{(\mu + (\nu-1)(i-1)} - \frac{1}{(\nu-1)i}\Big) = \frac{1}{\nu-1} \sum_{i=1}^{\infty} \frac{\nu - \mu - 1}{i(\mu + (\nu-1)(i-1))}, \tag{8.3.23}$$

which is finite. Further, on the event that $S_i^{(\mathrm{ind})} \geq 1$ for every $i \geq 1$ and that $S_i \leq (1-\varepsilon)(\nu-1)i$ except for finitely many i,

$$\begin{aligned} \sum_{i=1}^{m} \Big(\frac{1}{S_i^{(\mathrm{ind})}} - \frac{1}{\mathbb{E}[S_i^{(\mathrm{ind})}]}\Big) &= -\sum_{i=1}^{m} \frac{S_i^{(\mathrm{ind})} - \mathbb{E}[S_i^{(\mathrm{ind})}]}{S_i^{(\mathrm{ind})}\mathbb{E}[S_i^{(\mathrm{ind})}]} \\ &= \Theta_{\mathbb{P}}(1) \sum_{i=1}^{m} \frac{|S_i^{(\mathrm{ind})} - \mathbb{E}[S_i^{(\mathrm{ind})}]|}{i^2}. \end{aligned} \tag{8.3.24}$$

Since $S_i^{(\mathrm{ind})} = D + \sum_{j=2}^{i}(D_j^\star - 1) + (i-1)$ and the summands have a $(1+\varepsilon)$ bounded moment, $|\mathbb{E}[S_i^{(\mathrm{ind})} - \mathbb{E}[S_i^{(\mathrm{ind})}]| = O(i^{1/(1+\varepsilon)})$, which implies that the random variable on the r.h.s. of (8.3.24) has a finite mean. Finally,

$$\sum_{i=1}^{m} 1/i - \log m \to \gamma, \tag{8.3.25}$$

where γ is Euler-Mascheroni's constant. This identifies the limit of $Q_1 = T_m - (\log m)/(\nu-1)$ as

$$Q_1 = \sum_{i=1}^{\infty}(E_i - 1)/S_i^{(\mathrm{ind})} + \sum_{i=1}^{\infty}(1/S_i^{(\mathrm{ind})} - 1/[(\nu-1)i) + \gamma/(\nu-1). \tag{8.3.26}$$

In turn, (8.3.26) identifies the limit law Q_1 appearing in the limit law of $\mathcal{C}_n - \gamma \log n$ in (8.3.6).

The flow on $\mathrm{CM}_n(\boldsymbol{d})$ from one vertex. We next discuss the first-passage percolation problem on $\mathrm{CM}_n(\boldsymbol{d})$, started from a single vertex U_1. Naturally, this is not quite the same as first-passage percolation on a tree. Differences arise since

(1) the degrees of vertices found by the first-passage percolation flow are not quite independent (but they are quite close to i.i.d);

(2) in the flow, sometimes we encounter vertices that have already been found by the flow, thus creating a *cycle* (but this is rare when few vertices have been found);

(3) the offspring distribution depends on n, the size of the graph (but the empirical distribution converges).

The aim of this paragraph is to quantify that these differences are only negligible.

Let us start by describing how we explore the first-passage percolation neighborhood of a vertex in the graph. We perform the exploration of the first-passage percolation flow from U_1 and build the neighborhood of U_1 in $\mathrm{CM}_n(\boldsymbol{d})$ *at the same time*. The root U_1 immediately dies, and has $d_{U_1} \stackrel{d}{=} D_n$ half-edges incident to it. We check, one by one, whether these half-edges are paired to one of the other d_{U_1} half-edges incident to the root U_1, thus creating a self-loop. We let $B_1^{(n)}$ be the number of half-edges left, i.e., the number of half-edges incident to U_1 that are not part of self-loops. We denote $S_1^{(n)} = B_1^{(n)}$, and we let $T_1^{(n)} = E_1/S_1^{(n)}$.

We now recursively continue the above construction. Suppose we have constructed $S_1^{(n)}, \ldots, S_{m-1}^{(n)}$, the random variables $B_1^{(n)}, \ldots, B_{m-1}^{(n)}$ as well as the vertices $U_1, V_1, \ldots, V_{m-1}$ where V_i is incident to the half-edge to which the mth chosen half-edge is paired. Then, we draw one of the $S_{m-1}^{(n)}$ alive half-edges at time $m-1$ uniformly at random, and pair it to one of the available half-edges in the graph. Let V_m denote the vertex to which it is incident, and let d_{V_m} denote its degree. By construction, V_m cannot be indicent to any of the half-edges found so far, so that $V_m \in [n] \setminus \{U_1, V_1, \ldots, V_{m-1}\}$. Moreover, by Lemma 5.6, the random variables $(d_{V_i})_{i \geq 2}$ form a size-biased reordering of $[n] \setminus \{U_1\}$. We let $B_m^{(n)}$ denote the number of half-edges from the $d_{V_m} - 1$ half-edges incident to V_m that are paired to vertices outside of $\{U_1, V_1, \ldots, V_m\}$, and let $S_m^{(n)}$ denote the number of alive half-edges after the edges involved in cycles have been removed. Define $S_m^{(n)} = S_{m-1}^{(n)} + B_m^{(n)} - 1$.

We note that $S_m^{(n)}$ is not quite the number of unpaired half-edges at time m. Indeed, let $Q_m^{(n)}$ be the number of half-edges counted in $S_{m-1}^{(n)}$ that are paired to a half-edge incident to V_m. Then, at time m, there are

$$S_m^{(n)} - \sum_{i=2}^{m} Q_m^{(n)} \tag{8.3.27}$$

unpaired half-edges. We call the $\sum_{i=2}^m Q_m^{(n)}$ half-edges *artificial*.

In order to correct for the mistake created due to the presence of artificial half-edges, we note that with probability $(\sum_{i=2}^m Q_m^{(n)})/S_m^{(n)}$, we draw one of these half-edges. If we do so, then we let $B_m^{(n)} = 0$. Then, when we only observe $(S_m^{(n)}, B_m^{(n)}, V_m)_{m \geq 2}$

at the times when we do not use one of the artificial half-edges, this describes the first-passage flow from vertex U_1. Since, when we draw an artificial half-edge, we set $B_m^{(n)} = 0$, these half-edges can be thought of as dangling ends that do not really change the dynamics of the first-passage percolation flow. We let R_m denote the time where we find the mth real vertex, i.e., $R_0 = 0$ and $R_1 = 1$ unless vertex U_1 only has self-loops. This leads us to our main coupling result:

Proposition 8.9 (Coupling shortest-weight graphs on a tree and CM)**.** *Jointly for all $m \geq 1$, the set of real vertices in SWG_{R_m} is equal in distribution to the set of ith closest vertices to vertex 1, for $i = 1, \ldots, m$. Consequently, the weight and generation of the mth closest vertex to vertex 1 have distribution (T_{R_m}, G_{R_m}), where $(T_m, G_m)_{m\geq 0}$ are defined in (8.3.15)–(8.3.16).*

We fix $a_n = \sqrt{n}$. We can combine Proposition 8.9 with Proposition 8.8, as well as Lemma 5.6, to obtain the following corollary:

Corollary 8.10 (Asymptotics for shortest-weight paths on CM)**.** *Conditionally on vertex $S_i^{(n)} = B_1^{(n)} + \cdots + B_i^{(n)} - (i-1) \geq 1$ for every $i \geq 1$, and with $a_n = \sqrt{n}$,*

$$T_{a_n} - \frac{1}{\lambda} \log a_n \xrightarrow{d} X, \tag{8.3.28}$$

where $\lambda = \nu - 1$ and X is defined in Proposition 8.8.
(b) for G_{a_n} the generation of the vertex incident to the a_nth added half edge in $(\mathsf{SWG}_t)_{t\geq 0}$,

$$\frac{G_{a_n} - \beta \log a_n}{\sqrt{\beta \log a_n}} \xrightarrow{d} Z, \tag{8.3.29}$$

where Z is a standard normal variable.

Corollary 8.10 describes the limiting split-time and the height of the a_nth half-edge. We now extend this result to th flow from *two* vertices, which is necessary to obtain the hopcount and weight result on $\mathrm{CM}_n(\boldsymbol{d})$.

Flow clusters started from two vertices. To compute the hopcount, we first grow the SWG from vertex U_1 until time $a_n = \sqrt{n}$, followed by the growth of the SWG from vertex U_2 until the two SWGs meet, as we now explain in more detail. Denote by $(\mathsf{SWG}_m^{(i)})_{m\geq 0}$ the SWG from the vertex U_i with $i \in \{1, 2\}$, and, for $m \geq 0$, let

$$\mathsf{SWG}_m^{(1,2)} = \mathsf{SWG}_{a_n}^{(1)} \cup \mathsf{SWG}_m^{(2)}, \tag{8.3.30}$$

the union of the SWGs of vertices U_1 and U_2. We only consider values of m where $\mathsf{SWG}_{a_n}^{(1)}$ and $\mathsf{SWG}_m^{(2)}$ are *disjoint*, i.e., they do not contain any common vertices. We discuss the moment when $\mathsf{SWG}_{a_n}^{(1)}$ and $\mathsf{SWG}_m^{(2)}$ connect below.

We start by describing the flow from vertex U_2, conditionally on $\mathsf{SWG}_{a_n}^{(1)}$. By the memoryless property of the exponential distribution, all alive and real half-edges in $\mathsf{SWG}_{a_n}^{(1)}$ have a remaining lifetime that has an exponential distribution with prameter 1. In first-passage percolation on $\mathrm{CM}_n(\boldsymbol{d})$, these weights are associated to *edges*, while

for our flow clusters, we associate these weights to the half-edge that is incident to the flow cluster at that time. For the flow from a single vertex, this makes no difference, as we associate weight 0 on the half-edge to which it is paired. We have grown $\mathsf{SWG}_m^{(1)}$ to size $m = a_n$, and, by construction, all unpaired half-edges in $\mathsf{SWG}_{a_n}^{(1)}$ are connected to vertices outside of $\mathsf{SWG}_m^{(1)}$.

We can now, for all half-edges that are paiired to the unpaired half-edges in $\mathsf{SWG}_{a_n}^{(1)}$, instead associate the remaining weight to the *other* end of the edge, which is incident to a vertex that is not part of $\mathsf{SWG}_{a_n}^{(1)}$. When applying this to edges that connect $\mathsf{SWG}_{a_n}^{(1)}$ and $\mathsf{SWG}_m^{(2)}$ for some $m \geq 0$, this restores the weight distribution on alive half-edges in $\mathsf{SWG}_m^{(2)}$ to independent exponential random variables with parameter 1. As a result, the weights of unpaired half-edges incident to $\mathsf{SWG}_m^{(2)}$ are independent on whether these half-edges are paired to unpaired half-edges incident to $\mathsf{SWG}_{a_n}^{(1)}$ or not. In effect, we have

transferred the remaining lifetimes of half-edges in $\mathsf{SWG}_{a_n}^{(1)}$
to weights of half-edges incident to $\mathsf{SWG}_m^{(2)}$.

We now give the details of this subtle argument. Let the half-edge x be incident to $\mathsf{SWG}_{a_n}^{(1)}$. Let the vertex incident to x be V_x. Let y be the half-edge incident to $\mathsf{SWG}_m^{(2)}$ to which x is paired, and let V_y be the vertex incident to y. Then, conditionally on $\mathsf{SWG}_{a_n}^{(1)}$ and $(T_i^{(1)})_{i=1}^{a_n}$, the weight of xy is at least $T_{a_n}^{(1)} - W_{V_x}^{(1)}$, where $W_{i_s}^{(1)}$ is the weight of the shortest path from 1 to V_x. By the memoryless property of the exponential distribution, therefore, the weight on edge $e = xy$ equals $T_{a_n}^{(1)} - W_{V_x}^{(1)} + E_e$, where the collection (E_e), for all e incident to $\mathsf{SWG}_{a_n}^{(1)}$ are i.i.d. Exp(1) random variables. This means that the remaining lifetime of the edge is E_e, which is an exponentila random variable with mean 1.

Alternatively, we can redistribute the weight by saying that the half-edge y has weight E_e, and the half-edge x has weight $T_{a_n}^{(1)} - W_{V_x}^{(1)}$. Further, in the growth of $(\mathsf{SWG}_m^{(2)})_{m\geq 0}$, we can also think of the exponential weights of the edges incident to $\mathsf{SWG}_m^{(2)}$ being positioned on the *half-edges* incident to $\mathsf{SWG}_m^{(2)}$. Hence, there is no distinction between the half-edges that are part of edges connecting $\mathsf{SWG}_{a_n}^{(1)}$ and $\mathsf{SWG}_m^{(2)}$ and the half-edges that are part of edges incident to $\mathsf{SWG}_m^{(2)}$, but not to $\mathsf{SWG}_{a_n}^{(1)}$. Therefore, in the growth of $(\mathsf{SWG}_m^{(2)})_{m\geq 0}$, we can think of the minimal weight *half-edge* incident to $\mathsf{SWG}_m^{(2)}$ being chosen uniformly at random, and then a uniform free half-edge is chosen to pair it with. As a result, the distribution of the half-edges chosen *at the time of connection* is equal to any of the other (real) half-edges chosen along the way. This is a crucial ingredient to prove the scaling of the shortest-weight path between vertices U_1 and U_2.

To grow $(\mathsf{SWG}_m^{(2)})_{m\geq 0}$, we follow a strategy close to the one for $(\mathsf{SWG}_m^{(1)})m \geq 0$, now living on the vertex set $[n]\cup\{\varnothing\}\setminus\{U_1, V_1, \ldots, V_{a_n}\}$. Here $\varnothing$ is a vertex that we think of as encoding the information of $\mathsf{SWG}_{a_n}^{(1)}$. In particular, we set $d_\varnothing = S_{a_n}^{(n;1)}$, the number of alive half-edges incident to $\mathsf{SWG}_{a_n}^{(1)}$. By the above remark, $(\mathsf{SWG}_m^{(2)})_{m\geq 0}$ evolves as the flow on configuration model with degree distribution formed by $(d_i)_{i\in[n]\setminus\{U_1,V_1,\ldots,V_{a_n}\}}$, until the flow finds the distuinguished vertex $\varnothing$ of degree $d_\varnothing = S_{a_n}^{(n;1)}$.

The above description shows how we can grow the SWG from vertex U_1 followed by the one of vertex U_2. In order to state an adaptation of Proposition 8.9 to the

setting where the SWGs of vertex U_1 is first grown to size a_n, followed by the growth of the SWG from vertex U_2 until the connecting edge appears, we let the random time $R_m^{(i)}$ be the first time l such that $\mathsf{SWG}_l^{(i)}$ consists of $m+1$ real vertices. Then, our main coupling result for two simultaneous SWGs is as follows:

Proposition 8.11 (Coupling SWGs from two vertices on CM)**.** *Jointly for $m \geq 0$, as long as the sets of real vertices in $(\mathsf{SWG}_{a_n}^{(1)}, \mathsf{SWG}_m^{(2)})$ are disjoint, these sets are equal in distribution to the sets of j_1th, respectively j_2th, closest vertices to vertex U_1 and U_2, respectively, for $j_1 = 1, \ldots, R_{a_n}^{(1)}$ and $j_2 = 1, \ldots, R_m^{(2)}$, respectively.*

Proposition 8.11 describes the flow from two vertices U_1 and U_2 up to the moment that the SWGs are connected to one another. We next investigate the moment the connection between the SWGs takes place, as well as the properties of the connecting edge.

The connecting edge. As described above, we grow the two SWGs until the first half-edge with minimal weight incident to $\mathsf{SWG}_m^{(2)}$ is paired to a half-edge incident to $\mathsf{SWG}_{a_n}^{(1)}$. We call the created edge linking the two SWGs the *connecting edge*. More precisely, let

$$\mathsf{C}_n = \min\{m \geq 0 \colon \mathsf{SWG}_{a_n}^{(1)} \cap \mathsf{SWG}_m^{(2)} \neq \varnothing\}, \tag{8.3.31}$$

be the first time that $\mathsf{SWG}_{a_n}^{(1)}$ and $\mathsf{SWG}_m^{(2)}$ share a vertex. When $m = 0$, this means that $U_2 \in \mathsf{SWG}_{a_n}^{(1)}$ (which we shall show happens with small probability), while when $m \geq 1$, this means that the mth half-edge of $\mathsf{SWG}^{(2)}$ which is chosen and then paired, is paired to a half-edge from $\mathsf{SWG}_{a_n}^{(1)}$. The path found then is the shortest-weight path between vertices 1 and 2, since $\mathsf{SWG}_{a_n}^{(1)}$ and $\mathsf{SWG}_m^{(2)}$ precisely consists of the closest real vertices to the root U_i, for $i = 1, 2$, respectively.

For $i \in \{1, 2\}$, let $H_n^{(i)}$ denote the length of the shortest-weight path between vertex U_i and the common vertex in $\mathsf{SWG}_{a_n}^{(1)}$ and $\mathsf{SWG}_{\mathsf{C}_n}^{(2)}$, so that

$$H_n = H_n^{(1)} + H_n^{(2)}. \tag{8.3.32}$$

Since at time C_n we have found the shortest-weight path,

$$(H_n^{(1)}, H_n^{(2)}) \stackrel{d}{=} (G_{a_n+1}^{(1)} - 1, G_{\mathsf{C}_n}^{(2)}), \tag{8.3.33}$$

where $(G_m^{(1)})_{m\geq 1}$ and $(G_m^{(2)})_{m\geq 1}$ are copies of the process in (8.3.9), which are *conditioned on drawing a real half-edge*. Indeed, at the time of the connecting edge, a uniform (real) half-edge of $\mathsf{SWG}_m^{(2)}$ is drawn, and it is paired to a uniform (real) half-edge of $\mathsf{SWG}_{a_n}^{(1)}$. The number of hops in $\mathsf{SWG}_{a_n}^{(1)}$ to the end of the attached edge is therefore equal in distribution to $G_{a_n+1}^{(1)}$ conditioned on drawing a real half-edge. The -1 in (8.3.33) arises since the connecting edge is counted twice in $G_{a_n+1}^{(1)} + G_{\mathsf{C}_n}^{(2)}$. The processes $(G_m^{(1)})_{m\geq 1}$ and $(G_m^{(2)})_{m\geq 1}$ are conditionally independent given the realizations of $(B_m^{(i)})_{m\geq 2}$.

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,

$$\mathcal{C}_n = T_{a_n}^{(1)} + T_{\mathsf{C}_n}^{(2)}, \tag{8.3.34}$$

where $(T^{(1)}_m)_{m\geq 1}$ and $(T^{(2)}_m)_{m\geq 1}$ are two copies of the process $(T_m)_{m\geq 1}$ in (8.3.8), again conditioned on drawing a real half-edge. Indeed, to see (8.3.34), we note that the weight of the connecting edge is equal to the sum of weights of its two half-edges. Therefore, the weight of the shortest weight path is equal to the sum of the weight within $\mathsf{SWG}^{(1)}_{a_n}$, which is equal to $T^{(1)}_{a_n}$, and the weight within $\mathsf{SWG}^{(2)}_{\mathsf{C}_n}$, which is equal to $T^{(2)}_{\mathsf{C}_n}$.

In the distributions in (8.3.33) and (8.3.34) above, we always condition on drawing a real half-edge. Since we shall show that this occurs whp, this conditioning plays a minor role.

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. Recall that $S^{(n;i)}_m$ denotes the number of allowed half-edges in the SWG from vertex i at time m. The total number of allowed half-edges incident to $\mathsf{SWG}^{(1)}_{a_n}$ is $S^{(1)}_{a_n}$, while the number incident to $\mathsf{SWG}^{(2)}_m$ is equal to $S^{(2)}_m$, and where

$$S^{(i)}_m = D_i + \sum_{l=2}^{m} (B^{(i)}_l - 1). \tag{8.3.35}$$

Proposition 8.12 (The time to connection)**.** *For $i \in \{1,2\}$, and with $\alpha = \nu/(\nu-1)$,*

$$\Big(\frac{G^{(1)}_{a_n+1} - \alpha \log a_n}{\sqrt{\alpha \log a_n}}, \frac{G^{(2)}_{\mathsf{C}_n} - \alpha \log a_n}{\sqrt{\alpha \log a_n}}\Big) \xrightarrow{d} (Z_1, Z_2), \tag{8.3.36}$$

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 (8.3.26) and an independent exponential random variable E such that,

$$\Big(T^{(1)}_{a_n} - \frac{1}{\nu-1}\log a_n, T^{(2)}_{\mathsf{C}_n} - \frac{1}{\nu-1}\log a_n\Big) \xrightarrow{d} (X_1, X_2 + \frac{1}{\nu-1}\log E + \frac{1}{\nu-1}\log(\mu/(\nu-1))). \tag{8.3.37}$$

Proof. Conditionally on $\mathsf{SWG}^{(1)}_{a_n}$ and $(S^{(2)}_l)_{l=1}^{m-1}$ and ℓ_n, and assuming that m and $S^{(i)}_m$ satisfy appropriate bounds

$$\mathbb{P}(\mathsf{C}_n = m | \mathsf{C}_n > m-1) \approx \frac{S^{(1)}_{a_n}}{\ell_n}. \tag{8.3.38}$$

We use this heuristic in order to identify the limit of C_n/a_n:

Lemma 8.13 (Weak convergence of connection time)**.** *Under the conditions of Theorem 8.5, with $a_n = \sqrt{n}$,*

$$S^{(1)}_{a_n}/a_n \xrightarrow{\mathbb{P}} \nu - 1. \tag{8.3.39}$$

Consequently,

$$\mathsf{C}_n/a_n \xrightarrow{d} E\mu/(\nu-1), \tag{8.3.40}$$

where E has an exponential distribution with mean 1.

We only sketch the proof of Lemma 8.13. The details are not very difficult, but somewhat tedious.

Sketch of proof of Lemma 8.13. We have that $S^{(1)}_{a_n} \approx \sum_{i=1}^{a_n}(B_i - 1)$, where $(B_i)_{i=1}^{a_n}$ is a collection of i.i.d. random variables with distribution $D^\star - 1$. Since $\mathbb{E}[(D^\star)^{1+\varepsilon}] < \infty$

$$\frac{1}{a_n}\sum_{i=1}^{a_n}(B_i - 1) \xrightarrow{a.s.} \nu - 1. \tag{8.3.41}$$

Naturally, $S^{(1)}_{a_n}$is not quite a sum of i.i.d. random variables, and, therefore, the above argument misses its rigorous details.

Each time we pair a half-edge, we have a probability close to $S^{(1)}_{a_n}/\ell_n \approx (\nu - 1)/(\mu\sqrt{n})$ of drawing one that is incident to $\mathsf{SWG}^{(1)}_{a_n}$. 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. □

Now we are ready to complete the proof of Proposition 8.12:

Proof of Proposition 8.12. We first complete the proof of (8.3.37). It is not hard to prove from (8.3.34) that

$$\big(T^{(1)}_{a_n} - \gamma \log a_n, T^{(2)}_{\mathsf{C}_n} - \frac{1}{\nu - 1}\log \mathsf{C}_n\big) \xrightarrow{d} (X_1, X_2), \tag{8.3.42}$$

where (X_1, X_2) are two independent random variables with distribution given by (8.3.26). By Lemma 8.13,

$$\log \mathsf{C}_n - \log a_n = \log\left(\mathsf{C}_n/a_n\right) \xrightarrow{d} \log E + \frac{1}{\nu - 1}\log(\mu/(\nu - 1))\big). \tag{8.3.43}$$

Also, the two limits are independent, since the limit in (8.3.42) is independent of the limit of C_n/a_n. This completes the proof for the weight of the shortest path in (8.3.37). The proof for (8.3.36) is similar. □

Proof of Theorem 8.5. The statement for $\mathcal{C}_n$ in (8.3.2) follows by (8.3.34), (8.3.37) in Proposition 8.12 and the fact that $-\log E = \log(1/E)$ has a Gumbel distribution.

By (8.3.33) and Proposition 8.12, 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,

$$\begin{aligned} H_n &\stackrel{d}{=} G^{(1)}_{a_n+1} + G^{(2)}_{\mathsf{C}_n} - 1 \\ &= 2\alpha \log a_n + Z_1\sqrt{\alpha \log a_n} + Z_2\sqrt{\alpha \log a_n} + o_{\mathbb{P}}(\sqrt{\log n}) \\ &= 2\alpha \log a_n + Z\sqrt{2\alpha \log a_n} + o_{\mathbb{P}}(\sqrt{\log n}) = \alpha \log n + Z\sqrt{\alpha \log n} + o_{\mathbb{P}}(\sqrt{\log n}). \end{aligned} \tag{8.3.44}$$

This completes the proof of (8.3.2) for the hopcount. □

8.4 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 minimal-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, the memoryless property no longer holds. It turns out that many of such settings can be described in terms of continuous-time branching processes (CTBPs).

In this section, we discuss such CTBPs, which are models for the evolution of a population in continuous time. In a CTBP, 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 couting process in continuous time, and we are interested in the evolution of $|\mathsf{BP}(t)|$, which is the number of individuals alive at a given time $t \geq 0$, as well as many related properties of the CTBP $\mathsf{BP}(t)$ when viewed as a tree. Here one can think of the heights of vertices or the residual lifetimes.

We now formally define CTBPs. Let $(\mathcal{P}(t))_{t\geq 0}$ be a birth process. Recall that $\mathcal{T}_\infty$ denotes the infinite Ulam-Harris tree of words, so that $v \in \mathcal{T}_\infty$ can be representated 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 $(\mathcal{P}^{(v)}(t))_{t\geq 0}$, where the birth processes for different v are independent. At time 0, the birth process of the root $\varnothing$, which is denoted as $(\mathcal{P}^{(\varnothing)}(t))_{t\geq 0}$, starts running. Each time 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 $\mathsf{BP}(t)$ and its cardinality or size by $|\mathsf{BP}(t)|$. Thus, $\mathsf{BP}(t)$ 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 CTBP $(\mathsf{BP}(t))_{t\geq 0}$ 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:

Example 8.14 (Yule process)**.** Let the birth process $(\mathcal{P}(t))_{t\geq 0}$ be a Poisson point process $(\mathsf{PP}(t))_{t\geq 0}$, so that $\mathsf{PP}[a, b] = \#\{i\colon Y_i \in [a, b]\}$, where $Y_i = E_1 + \ldots + E_i$ and $(E_i)_{i\geq 1}$ are i.i.d. exponential random variables with parameter 1. Individuals live forever. The process $(\mathsf{BP}(t))_{t\geq 0}$ where $\mathsf{BP}(t)$ 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 8.2. Let $T_m = \inf\{t\colon |\mathsf{BP}(t)| = m\}$ denote the birth time of the mth individual in the Yule process, where by convention $T_1 = 0$ so that the ancestor of the CTBP is born insantaneously. Then, by the memoryless property of the exponential distribution,

$$T_{m+1} \stackrel{d}{=} \sum_{k=1}^{m} E_k/k. \tag{8.4.1}$$

Thus, T_m has the same distribution as the limit as $n \to \infty$ of the birth time of the mth 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 mth 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 mth vertex in Theorem 8.3 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 $p_m(t) = \mathbb{P}(|\mathsf{BP}(t)| = m)$. Then, since the birth process of the root is a Poisson point process,

$$p_1(t) = \mathbb{P}(\mathsf{PP}[0,t] = 0) = \mathrm{e}^{-t}. \tag{8.4.2}$$

Further, for $|\mathsf{BP}(t)| = m$ to occur, the mth birth should occur at some time $u \in [0,t]$, while the $(m+1)$st birth should occur after time t. Conditionally on $|\mathsf{BP}(u)| = m-1$, the rate of birth of the mth individual is equal to $m-1$. This implies the following relation between $p_m(t)$ and $p_{m-1}(t)$:

$$p_m(t) = \int_0^t p_{m-1}(u)(m-1)\mathrm{e}^{-m(t-u)}du. \tag{8.4.3}$$

The solution to (8.4.3) is

$$p_m(t) = \mathrm{e}^{-t}(1-\mathrm{e}^{-t})^{m-1}, \qquad m \geq 1, t \geq 0, \tag{8.4.4}$$

which we now prove by induction. The claim for $m = 1$ is already proved in (8.4.2) above. To advance the induction hypothesis, we can rewrite

$$p_m(t) = (m-1)\mathrm{e}^{-mt}\int_0^t \mathrm{e}^{mu}p_{m-1}(u)du. \tag{8.4.5}$$

We use induction, which yields that

$$(m-1)\mathrm{e}^{mu}p_{m-1}(u) = (m-1)\mathrm{e}^{u}(\mathrm{e}^{u}-1)^{m-2} = \frac{d}{du}(\mathrm{e}^{u}-1)^{m-1}, \tag{8.4.6}$$

so that

$$p_m(t) = \mathrm{e}^{-mt}\int_0^t (m-1)\mathrm{e}^{mu}p_{m-1}(u)du = \mathrm{e}^{-mt}(\mathrm{e}^{t}-1)^{m-1} = \mathrm{e}^{-t}(1-\mathrm{e}^{-t})^{m-1}, \tag{8.4.7}$$

as required.

By (8.4.4), the distribution of $|\mathsf{BP}(t)|$ is Geometric with parameter e^{-t}. A geometric random variable with small success probability p is close to an exponential random variable:

Exercise 8.13 (Geometric with small success probability)**.** *Let X_p have a geometric distribution with parameter p. Show that, as $p \searrow 0$,*

$$pX_p \xrightarrow{d} E, \tag{8.4.8}$$

where E has an exponential distribution.

By Exercise 8.13, as $t \to \infty$,

$$\mathrm{e}^{-t}|\mathsf{BP}(t)| \xrightarrow{d} E. \tag{8.4.9}$$

In fact, we can improve this result to convergence almost surely using a martigale argument:

Exercise 8.14 (Yule process martingale). *Show that the process* $M(t) = \mathrm{e}^{-t}|\mathsf{BP}(t)|$ *is a continuous-time martingale, and conclude that the convergence in* (8.4.9) *occurs a.s.*

Example 8.15 (Bellman-Harris processes). In a Bellman-Harris process, the initial individual dies instantaneously, giving birth to a random number of children B. Each of these individuals has a lifetime 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 B. Each of these children again has and exponentila lifetime with mean 1, etc. This is a CTBP with birth process

$$\mathcal{P}[a,b] = B\mathbb{1}_{\{E\in[a,b]\}}. \tag{8.4.10}$$

Lifetimes 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 B and with exponential passage times.

To analyze a Bellman-Harris CTBP, we let $S_0 = 1$, and define $(S_i)_{i\geq 0}$ recursively by

$$S_i = S_{i-1} + B_i - 1, \tag{8.4.11}$$

as long as $S_{i-1} \geq 1$. Here $(B_i)_{i\geq 1}$ is an i.i.d. sequence of random variables whose law is called the *o*ffspring 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 $i \in [m-1]$,

$$T_m = \sum_{j=1}^{m} E_j/S_j, \tag{8.4.12}$$

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 $|\mathsf{BP}(t)|$ equal to the number of alive vertices at time t,

$$|\mathsf{BP}(T_m)| = S_{T_m}, \tag{8.4.13}$$

so that $|\mathsf{BP}(t)| = S_{T(t)}$, where $T(t) = \max\{m\colon T_m \leq t\}$.

Let us first make the relation to the first-passage percolation on $\mathrm{CM}_n(\boldsymbol{D})$ discussed in Section 8.3. There, we had that $S_1 = D$, while, for $i > 1$, the recursion relation in (8.4.11) is satisfied. We arrive at a process that is a Bellman-Harris CTBP, apart from the fact that the offspring distribution at the root is D instead of $B = D^\star - 1$.

Naturally, CTBPs 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 CTBP a *t*wo-stage CTBP. Indeed, let the number of children of the root be equal to D and the offspring of the ith individual to be equal to B_i where $(B_i)_{i\geq 2}$ is an i.i.d. sequence whose law is given by $B = D^\star - 1$. Then, we have the relation that

$$|\mathsf{BP}(t)| = \sum_{i=1}^{D} |\mathsf{BP}_i(t - E_i)|, \tag{8.4.14}$$

where $(E_i)_{i\geq 1}$ are i.i.d. exponentials, and $(|\mathsf{BP}_i(t)|)_{t\geq 0}$ are i.i.d. Bellmann-Harris processes with i.i.d. offsprings. By convention, $|\mathsf{BP}_i(t)| = 0$ when $t < 0$.

In Section 8.3, it was proved that, with T_m the death time of the mth individual and G_m its height,

$$T_m - \frac{1}{\lambda}\log n \xrightarrow{d} X, \qquad \frac{G_m - \frac{\lambda}{\lambda-1}\log m}{\sqrt{\frac{\lambda}{\lambda-1}\log m}} \xrightarrow{d} Z, \tag{8.4.15}$$

where $\lambda = \mathbb{E}[B]$ is the expected offspring, Z is standard normal and

$$X = \sum_{j=1}^{\infty}(E_j - 1)/S_j + \sum_{j=1}^{\infty}[1/S_j - 1/(\lambda j)] + \gamma/\lambda, \tag{8.4.16}$$

where γ is the Euler-Mascheroni constant and $\lambda = \nu - 1 = \mathbb{E}[B] - 1$. Therefore, $m\mathrm{e}^{-\lambda T_m} \xrightarrow{d} -\lambda X$, which immediately implies that

$$\mathrm{e}^{-\lambda t}|\mathsf{BP}(t)| \xrightarrow{d} W = \mathrm{e}^{-\lambda X}. \tag{8.4.17}$$

By the Markov property of this process, the random variable $M(t) = \mathrm{e}^{-\lambda t}|\mathsf{BP}(t)|$ is a continuous-time martingale, and therefore the convergence in (8.4.17) even holds almost surely.

From a first-passage percolation point of view, it may look more natural to let $|\mathsf{BP}(t)|$ 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 $|\mathsf{BP}(t)|$ 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.

General formulation. In Examples 8.14 and 8.15, we let $|\mathsf{BP}(t)|$ denote the number of alive indivuduals, and we see that there exists a λ and a limiting random variable W such that

$$\mathrm{e}^{-\lambda t}|\mathsf{BP}(t)| \xrightarrow{d} W. \tag{8.4.18}$$

Further, there are α, β such that

$$\frac{G_m - \alpha\log m}{\sqrt{\beta\log m}} \xrightarrow{d} Z, \tag{8.4.19}$$

where Z is standard normal. These results have been obtained in a direct way in Sections 8.2-8.3. We next turn to an example where such a direct computation is much more difficult:

Example 8.16 (Age-dependent branching processes)**.** Let $(\mathsf{BP}(t))_{t\geq 0}$ denote the CTBP where the birth process $\mathcal{P}$ is given by

$$\mathcal{P}[a, b] = B\mathbb{1}_{\{Y\in[a,b]\}}, \tag{8.4.20}$$

where 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 B_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 B_v children, where $(B_v)_v$ is a sequence of i.i.d. random variables with the same distribution as B_1. Lifetimes and number of offspring across individuals are independent.

We let $|\mathsf{BP}(t)|$ 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 B and passage-time distribution F_Y. When the lifetime distribution is exponential, we retrieve the Bellman-Harris process.

When the lifetime distribution Y is not exponential, the process $(|\mathsf{BP}(t)|)_{t\geq 0}$ is *not* Markovian. As a result, this process is much more difficult to study. Still, one would guess that $|\mathsf{BP}(t)|$ again grows exponentially in t.

In order to study general CTBPs, we define the following *Malthusian parameter:*

Definition 8.17 (Malthusian parameter)**.** *Let $(\mathsf{BP}(t))_{t\geq 0}$ 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*

$$\mathbb{E}\Big[\int_0^\infty \mathrm{e}^{-\lambda t}\mathcal{P}(dt)\Big] = 1. \tag{8.4.21}$$

The Mathusian parameter is the solution to (8.4.21) when this exists. The necesary and sufficient condition for this to hold is that $\mathbb{E}[|\mathcal{P}[0,\infty)|] > 1$ and $\lim_{t\to\infty}\mathbb{E}\Big[\int_0^\infty \mathrm{e}^{-\lambda t}\mathcal{P}(dt)\Big] < 1$. The following theorem shows that λ indeed describes the exponential growth of the CTBP:

Theorem 8.18 (Exponential growth CTBP)**.** *Assume the $X\log X$ condition in the form*

$$\mathbb{E}[\hat{X}(\lambda)\log(\hat{X}(\lambda))_+] < \infty, \tag{8.4.22}$$

where the random variable $\hat{X}(\lambda)$ is defined by

$$\hat{X}(\lambda) = \int_0^\infty \mathrm{e}^{-\lambda t}\mathcal{P}(dt). \tag{8.4.23}$$

Assume that $\mathbb{E}[\hat{X}(0)] > 1$ and $\lim_{\lambda\to\infty}\mathbb{E}[\hat{X}(\lambda)] < 1$.
Then, the Malthusian parameter $\lambda > 0$ given in Definition 8.17 exists and is unique, and there exists a random variable W such that

$$\mathrm{e}^{-\lambda t}|\mathsf{BP}(t)| \xrightarrow{d} W. \tag{8.4.24}$$

Furthermore, W is positive if and only if the survival event $\{|\mathsf{BP}(t)| > 0\ \forall t \geq 0\}$ occurs.

Example 8.19 (Examples of exponential growth CTBP)**.** *We now give three examples:*
(a) For the Yule process, we can compute that

$$\mathbb{E}\Big[\int_0^\infty \mathrm{e}^{-\lambda t}\mathcal{P}(dt)\Big] = \sum_{i\geq 1}\mathbb{E}\big[\mathrm{e}^{-\lambda Y_i}\big], \tag{8.4.25}$$

since a Poisson point process PP *can be written as* $\mathsf{PP} = \{Y_i\}_{i\geq 1}$ *with* $Y_i = E_1 + \cdots + E_i$ *and* $(E_i)_{i\geq 1}$ *are i.i.d. exponential random variables. Since* Y_i *has a Gamma-distribution with parameter* i*, we can further compute*

$$\begin{aligned}\sum_{i\geq 1}\mathbb{E}\big[\mathrm{e}^{-\lambda Y_i}\big] &= \sum_{i\geq 1}\mathbb{E}\big[\mathrm{e}^{-\lambda E_1}\big]^i \\ &= \sum_{i\geq 1}\Big(\frac{1}{\lambda+1}\Big)^i = 1/\lambda,\end{aligned} \tag{8.4.26}$$

so that $\lambda = 1$*. Thus,* (8.4.24) *reduces to* (8.4.9)*, 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:*

$$W \stackrel{d}{=} \sum_{i=1}^\infty \mathrm{e}^{-Y_i}W_i, \tag{8.4.27}$$

where again $Y_i = E_1 + \cdots + E_i$ *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* $(\mathrm{e}^{-Y_i}W_i)_{i\geq 1}$ *has the same distribution as* $\mathrm{e}^{-E_1}(\mathrm{e}^{-Y_{i-1}}W_i)_{i\geq 1}$*, we thus obtain that*

$$W \stackrel{d}{=} U(W_1 + W_2), \tag{8.4.28}$$

where W_1 *and* W_2 *have 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 $\mathbb{E}[B] \in (1,\infty)$*, we compute that*

$$\mathbb{E}\Big[\int_0^\infty \mathrm{e}^{-\lambda t}\mathcal{P}(dt)\Big] = \mathbb{E}[B]\mathbb{E}\big[\mathrm{e}^{-\lambda E}\big] = \frac{\mathbb{E}[B]}{\lambda+1}, \tag{8.4.29}$$

so that $\lambda = \mathbb{E}[B] - 1 > 0$*.*
(c) For general age-dependent branching processes with $\mathbb{E}[B] \in (1,\infty)$*, instead*

$$\mathbb{E}\Big[\int_0^\infty \mathrm{e}^{-\lambda t}\mathcal{P}(dt)\Big] = \mathbb{E}[B]\mathbb{E}\big[\mathrm{e}^{-\lambda Y}\big], \tag{8.4.30}$$

where Y *has the edge weight distribution. The equation* $\mathbb{E}[B]\mathbb{E}\big[\mathrm{e}^{-\lambda Y}\big] = 1$ *always has a solution, since* $\mathbb{E}[B] \in (1,\infty)$ *and* $\mathbb{E}\big[\mathrm{e}^{-\lambda Y}\big] \searrow 0$ *as* $\lambda \searrow 0$*. Thus, Theorem 8.18 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 8.15 ($X \log X$-condition Yule process)**.** *Show that the $X \log X$ condition holds for a Yule process. Hint: Show that $\hat{X}(1) = \sum_{i=1}^{\infty} \mathrm{e}^{-Y_i}$, where $Y_i = E_1 + \cdots + E_i$ are the points of a Poisson process and $(E_i)_{i \geq 1}$ are i.i.d. exponential random variables with parameter 1.*

Exercise 8.16 ($X \log X$-condition Bellman-Harris branching process)**.** *Show that the $X \log X$ condition holds for a Bellman-Harris process precisely when $\mathbb{E}[B \log(B)_+] < \infty$.*

Exercise 8.17 ($X \log X$-condition age-dependent branching process)**.** *Show that the $X \log X$ condition holds for age-dependent CTBP precisely when $\mathbb{E}[B \log(B)_+] < \infty$.*

Proof of Theorem 8.18. We will not give a full proof of Theorem 8.18. Instead, in the case of age-dependent branching processs, we prove that there exists an $A \in (0, \infty)$ such that

$$\mathrm{e}^{-\lambda t} \mathbb{E}[|\mathsf{BP}(t)|] \to A. \tag{8.4.31}$$

We assume that the expected offspring $\mathbb{E}[B] = \nu$ satisfies $\nu > 1$ and that $\mathbb{E}[B^2] < \infty$. Further, the lifetime distribution function F_Y has to be non-lattice, plus some small additional condition for the almost sure convergence (see [62] for the proof).

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 $|\mathsf{BP}(t)|$ at time t, by

$$F(s,t) = \sum_{k=0}^{\infty} s^k \mathbb{P}(|\mathsf{BP}(t)| = k). \tag{8.4.32}$$

The function $F(s,t)$ satisfies the equation

$$F(s,t) = s(1 - F_Y(t)) + \int_0^t G_B(F(s, t-u))\, F_Y(du), \tag{8.4.33}$$

where $G_B(s) = \mathbb{E}[s^B]$ is the generating function of the offspring B. Equation (8.4.33) follows by conditioning on the lifetime and number of offspring of the first individual. In particular $\mathbb{E}[|\mathsf{BP}(t)|]$ satisfies the renewal-like equation:

$$\mathbb{E}[|\mathsf{BP}(t)|] = 1 - F_Y(t) + \nu \int_0^t \mathbb{E}[|\mathsf{BP}(t-u)|]\, F_Y(du). \tag{8.4.34}$$

To make (8.4.34) into a *genuine* renewal equation, we define the *stable-age distribution:*

Definition 8.20 (Stable-age distribution)**.** *Let the* stable-age distribution *be the distribution function $\bar{F}_Y$ given by*

$$\bar{F}_Y(y) = \mathbb{E}\Big[\int_0^y \mathrm{e}^{-\lambda t} \mathcal{P}(dt) \Big]. \tag{8.4.35}$$

Let $\bar{Y}$ be a random variable with distribution function $\bar{F}_Y$.

Multiplying both sides of (8.4.34) by $\mathrm{e}^{-\lambda t}$, where λ is the Malthusian parameter, we obtain

$$\mathrm{e}^{-\lambda t}\mathbb{E}[|\mathsf{BP}(t)|] = \mathrm{e}^{-\lambda t}[1 - F_Y(t)] + \mathrm{e}^{-\lambda u}\nu \int_0^t \mathrm{e}^{-\lambda(t-u)}\mathbb{E}[\mathsf{BP}(t-u)]\, F_Y(du), \quad (8.4.36)$$

or, equivalently, using that by (8.4.35) $\bar{F}_Y(du) = \nu \mathrm{e}^{-\lambda u} F_Y(du)$,

$$K(t) = f(t) + \int_0^t K(t-u)\, \bar{F}_Y(du), \quad (8.4.37)$$

where

$$K(t) = \mathrm{e}^{-\lambda t}\mathbb{E}[|\mathsf{BP}(t)|], \qquad f(t) = \mathrm{e}^{-\lambda t}[1 - F_Y(t)]. \quad (8.4.38)$$

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

$$K(t) = f(t) + \int_0^t K(t-u)\, m(du), \quad (8.4.39)$$

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

$$\lim_{t\to\infty} K(t) = \frac{1}{\mu}\int_0^\infty f(u)du, \quad (8.4.40)$$

where $\mu = \int_0^\infty u m(du)$ is the mean of the probability measure m. The following exercises give an idea of why (8.4.40) could be true:

Exercise 8.18 (Key-renewal theorem (1))**.** *For $\alpha > 0$, let $\hat{K}(\alpha) = \int_0^\infty \mathrm{e}^{-\alpha t}K(t)dt$, where $(K(t))_{t\geq 0}$ is the solution to (8.4.39). Prove that*

$$\hat{K}(\alpha) = \frac{\int_0^\infty \mathrm{e}^{-\alpha t} f(t)dt}{1 - \int_0^\infty \mathrm{e}^{-\alpha t} m(dt)}. \quad (8.4.41)$$

Conclude that $\alpha\hat{K}(\alpha) \to \int_0^\infty f(u)du / \int_0^\infty u m(du)$ when $\alpha \searrow 0$.

Exercise 8.19 (Key-renewal theorem (2))**.** *Use the previous exercise to show that if $K(t) \to A$ when $t \to \infty$, then A must satisfy $A = \int_0^\infty f(u)du / \int_0^\infty u m(du)$.*

Applying the Key-Renewal Theorem to (8.4.37) yields

$$\lim_{t\to\infty} \mathrm{e}^{-\lambda t}\mathbb{E}[|\mathsf{BP}(t)|] = \lim_{t\to\infty} K(t) = \int_0^\infty f(y)\, dy/\bar{\nu}, \quad (8.4.42)$$

where $\bar{\nu} = \int_0^\infty [1 - \bar{F}_Y(t)]\, dt$ equals the mean of $\bar{F}_Y(t)$.

An easy computation verifies that (8.4.42) impies:

$$\lim_{t\to\infty} \mathrm{e}^{-\lambda t}\mathbb{E}[|\mathsf{BP}(t)|] = A, \quad (8.4.43)$$

or $\mathbb{E}[|\mathsf{BP}(t)|] = Ae^{\lambda t}(1+o(1))$, where

$$A = \frac{\nu - 1}{\lambda \nu \bar{\nu}} = \frac{\nu - 1}{\lambda \nu^2 \int_0^\infty y e^{-\lambda y} F_Y(dy)}. \tag{8.4.44}$$

This proves (8.4.31) for age-dependent CTBPs.

The proof of (8.4.24) in the more general case is more involved, and we give a sketch now.

Give a (heuristic) proof for (8.4.24)?

□

Exercise 8.20 (Second moment for CTBP*). *Assume that $\mathbb{E}[B^2] < \infty$ where B is the random amount of offspring of our CTBP. Show that*

$$\lim_{t\to\infty} e^{-2\lambda t}\mathbb{E}\big[|\mathsf{BP}(t)|^2\big] = \frac{A^2 a}{1-\eta}, \tag{8.4.45}$$

where

$$A = \lim_{t\to\infty} e^{-\lambda t}\mathbb{E}\big[|\mathsf{BP}(t)|\big], \quad \eta = \int_0^\infty e^{-\alpha y}\, d\bar{F}_Y(y) < 1, \quad a = \eta\mathbb{E}[B(B-1)]/\mathbb{E}[B]. \tag{8.4.46}$$

Hint: Adapt (8.4.33) to get a recursion formula for $\mathbb{E}\big[|\mathsf{BP}(t)|^2\big]$ by conditioning on the first generation.

In the following, we derive a results on the number of individuals in a CTBP satisfying certain properties. We investigate the residual lifetime distribution and the heights of alive individuals in a CTBP. In what follows, we restrict to age-dependent CTBPs.

The residual lifetime distribution. We let $|\mathsf{BP}[t, t+s)|$ denote the number of individuals in the CTBP at time t and with residual lifetime 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 lifetime distribution* F_R to have density f_R given by

$$f_R(x) = \frac{\int_0^\infty e^{-\lambda y} f_Y(x+y)\, dy}{\int_0^\infty e^{-\lambda y}[1 - F_Y(y)]\, dy}, \tag{8.4.47}$$

where f_Y is the density of the life time distribution F_Y.

Exercise 8.21 (Residual lifetime distribution). *Show that f_R in (8.4.47) is a denstity on $[0, \infty)$.*

The main result of this section is the following theorem:

Theorem 8.21 (The residual lifetime distribution). *Let $(\mathsf{BP}(t))_{t\geq 0}$ be an age-dependent CTBP with offspring B. Assume that the $X \log X$ condition holds, i.e., $\mathbb{E}[B\log(B)_+] < \infty$. Then, with $A = (\nu-1)/\lambda\nu\bar{\nu}$ and for all $s \geq 0$,*

$$\lim_{t\to\infty} e^{-\lambda t}\mathbb{E}\big[|\mathsf{BP}[t, t+s)|\big] = AF_R(s). \tag{8.4.48}$$

Proof. We adapt the proof of (8.4.31) in Theorem 8.18. We note that (8.4.34) is now replaced by

$$\mathbb{E}[|\mathsf{BP}[t,t+s)|] = F_Y(t+s) - F_Y(t) + \nu \int_0^t \mathbb{E}[|\mathsf{BP}[t-u,t+s-u)|]\, dF_Y(u). \quad (8.4.49)$$

Multiplying both sides of (8.4.49) by $\mathrm{e}^{-\lambda t}$, where λ is the Malthusian parameter, we obtain

$$\mathrm{e}^{-\lambda t}\mathbb{E}[|\mathsf{BP}[t,t+s)|] = \mathrm{e}^{-\lambda t}[F_Y(t+s)-F_Y(t)]+\nu \int_0^t \mathrm{e}^{-\lambda u}\mathrm{e}^{-\lambda(t-u)}\mathbb{E}[|\mathsf{BP}[t-u,t+s-u)|]\, F_Y(du), \quad (8.4.50)$$

or, equivalently,

$$K(t) = f(t) + \int_0^t K(t-u)\, \bar{F}_Y(du), \quad (8.4.51)$$

where again $\bar{F}_Y(du) = \nu \mathrm{e}^{-\lambda u} F_Y(du)$ and now

$$K(t) = \mathrm{e}^{-\lambda t}\mathbb{E}[|\mathsf{BP}[t,t+s)|], \qquad f(t) = \mathrm{e}^{-\lambda t}[F_Y(t+s) - F_Y(t)]. \quad (8.4.52)$$

By the Key-renewal theorem,

$$\lim_{t\to\infty} \mathrm{e}^{-\lambda t}\mathbb{E}[|\mathsf{BP}[t,t+s)|] = \frac{1}{\mu}\int_0^\infty f(u)du = \frac{\int_0^\infty \mathrm{e}^{-\lambda y}[F_Y(y+s) - F_Y(y)]dy}{\int_0^\infty u\bar{F}_Y(du)}. \quad (8.4.53)$$

Inserting the definition $A = (\nu-1)/(\lambda\nu\bar{\nu})$ and rewriting the above yields the result:

Exercise 8.22 (Completion proof Theorem 8.21.)**.** *Use* (8.4.53) *and thm-res-lifetime-CTBP and* (8.4.47) *to complete the proof of Theorem 8.21.*

□

The height of a random vertex. We next investigate the *heights* of vertices in the CTBP. For this, we let $|\mathsf{BP}_j(t)|$ denote the number of individuals alive at time t in generation j, and write

$$|\mathsf{BP}_{\le k}(t)| = \sum_{j=0}^{k} |\mathsf{BP}_j(t)|. \quad (8.4.54)$$

Theorem 8.22 (CLT for vertex heights in CTBP)**.** *Let $(\mathsf{BP}(t))_{t\ge 0}$ be an age-dependent CTBP with offspring B. Assume that $\mathbb{E}[B\log(B)_+] < \infty$, so that the $X\log X$ condition holds. Then, with $A = (\nu-1)/\lambda\nu\bar{\nu}$ and for all $x \in \mathbb{R}$,*

$$\lim_{t\to\infty} \mathrm{e}^{-\lambda t}\mathbb{E}\big[|\mathsf{BP}_{\le k_t(x)}(t)|\big] = A\Phi(x), \quad (8.4.55)$$

where

$$k_t(x) = \frac{t}{\bar{\nu}} + x\sqrt{t\frac{\bar{\sigma}^2}{\bar{\nu}^3}}, \quad (8.4.56)$$

and $\bar{\nu}$ and $\bar{\sigma}^2$ are the mean, respectively variance, of the stable-age distribution $\bar{F}_Y$.

Theorem 8.22 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 lifetime (with c.d.f. equal to F_Y) of the first individual, after which the individual dies and splits in a random number offspring with mean ν,

$$\mathbb{E}[|\mathsf{BP}_j[t,t+s)|] = \nu \int_0^t \mathbb{E}[|\mathsf{BP}_{j-1}(t-y)|]\, dF_Y(y). \tag{8.4.57}$$

As before, we multiply by $\mathrm{e}^{-\lambda t}$ and define

$$|\overline{\mathsf{BP}}_j(t)| = \mathrm{e}^{-\lambda t}|\mathsf{BP}_j(t)|. \tag{8.4.58}$$

Rewriting (8.4.57), we obtain the recursion

$$\mathbb{E}[|\overline{\mathsf{BP}}_j(t)|] = \int_0^t \mathbb{E}[|\overline{\mathsf{BP}}_{j-1}(t-y)|]\, d\bar{F}_Y(y), \tag{8.4.59}$$

where, as before $\bar{F}_Y(du) = \nu \mathrm{e}^{-\lambda u} F_Y(du)$ is the stable-age distribution. Hence, if we continue to iterate, then we get

$$\mathbb{E}[|\overline{\mathsf{BP}}_j(t)|] = \int_0^t \mathbb{E}[|\overline{\mathsf{BP}}(t-y)|]\, d\bar{F}_Y^{\star j}(y), \tag{8.4.60}$$

where $\bar{F}_Y^{\star j}$ 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 will use the CLT. For fixed $t > 0$ and $m \geq 0$, we define

$$|\overline{\mathsf{BP}}_{>m}(t)| = \sum_{j=m+1}^{\infty} |\overline{\mathsf{BP}}_j(t)|. \tag{8.4.61}$$

By Theorem 8.18,

$$\lim_{t\to\infty} \mathbb{E}[|\overline{\mathsf{BP}}(t)|] = \lim_{t\to\infty} \sum_{j=0}^{\infty} \mathbb{E}[|\overline{\mathsf{BP}}_j(t)|] = A. \tag{8.4.62}$$

Hence, (8.4.55) follows if we show that

$$\mathbb{E}[|\overline{\mathsf{BP}}_{>k_t(x)}(t)|] \to A - A\Phi(x) = A\Phi(-x). \tag{8.4.63}$$

Note that

$$\mathbb{E}[|\overline{\mathsf{BP}}_{>k_t(x)}(t)|] = \int_0^t \mathbb{E}[|\overline{\mathsf{BP}}_0(t-u)|]\, d\bar{F}_Y^{\star k_t(x)}(u). \tag{8.4.64}$$

Take an arbitrary $\varepsilon > 0$ and take t_0 so large so that for $t > t_0$,

$$|\mathbb{E}[|\overline{\mathsf{BP}}(t)|] - A| \leq \varepsilon. \tag{8.4.65}$$

Then,

$$\begin{aligned}&\Big|\mathbb{E}[|\overline{\mathsf{BP}}_{>k_t(x)}(t)|] - A\Phi(-x)\Big| \\ &\qquad \le \varepsilon \bar{F}_Y^{\star k_t(x)}(t) + A\big|\bar{F}_Y^{\star k_t(x)}(t) - \Phi(-x)\big| + \int_{t-t_0}^{t} \mathbb{E}[|\overline{\mathsf{BP}}(t)|]\, d\bar{F}_Y^{\star k_t(x)}(u).\end{aligned} \tag{8.4.66}$$

The last term vanishes since $\mathbb{E}[|\overline{\mathsf{BP}}(t)|]$ is uniformly bounded and $\bar{F}_Y^{\star k_t(x)}(t) - \bar{F}_Y^{\star k_t(x)}(t-t_0) = o(1)$. Furthermore, with $m = k_t(x) \to \infty$,

$$k_t(x) \sim \frac{t}{\bar{\nu}} + x\sqrt{t\frac{\bar{\sigma}^2}{\bar{\nu}^3}} \qquad \text{precisely when} \qquad t \sim m\bar{\nu} - x\bar{\sigma}\sqrt{m}. \tag{8.4.67}$$

As a result, by the CLT and the fact that $\bar{\nu}$ and $\bar{\sigma}^2$ are the mean and the variance of the distribution function $\bar{F}_Y$,

$$\lim_{t\to\infty} \bar{F}_Y^{\star k_t(x)}(t) = \Phi(-x). \tag{8.4.68}$$

Together with (8.4.66), this proves the claim in (8.4.63), and hence Theorem 8.22. □

Exercise 8.23 (Joint residual lifetime and height convergence)**.** *Adapt the proof of Theorem 8.22 to show that*

$$\lim_{t\to\infty} \mathrm{e}^{-\lambda t}\mathbb{E}\big[|\mathsf{BP}_{\le k_t(x)}[t, t+s)|\big] = A\Phi(x)F_R(s), \tag{8.4.69}$$

where $|\mathsf{BP}_{\le k}[t, t+s)|$ *is the number of alive individuals with residual lifetime at most* s *and with height at most* k*. Thus, the asymptotic height and residual lifetime are close to independent.*

Distributional convergence of alive vertices with residual lifetime and height. Theorems 8.21-8.22 investigate the mean number of vertices with given residual lifetime and height. Since we can interpret $A = \lim_{t\to\infty} \mathrm{e}^{-\lambda t}\mathbb{E}[|\mathsf{BP}(t)|]$, this *suggests* that on the event of survival, we also have that $\mathrm{e}^{-\lambda t}|\mathsf{BP}[t, t+s)| \xrightarrow{d} WF_R(s)$, as well as $\mathrm{e}^{-\lambda t}|\mathsf{BP}_{\le k_t(x)}(t)| \xrightarrow{d} W\Phi(x)$. However, Theorems 8.21-8.22 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 lifetimes are investigated in the following theorem:

Theorem 8.23 (Residual lifetime and heights in a CTBP)**.** *Let* $(\mathsf{BP}(t))_{t\ge 0}$ *be an age-dependent CTBP with offspring* B*. Assume that* $\mathbb{E}[B\log(B)_+] < \infty$ *so that the* $X\log X$ *condition holds. Then, with* W *denoting the limit in Theorem 8.18,* $A = (\nu-1)/\lambda\nu\bar{\nu}$ *and for all* $x \in \mathbb{R}, s \ge 0$,

$$\mathrm{e}^{-\lambda t}|\mathsf{BP}_{\le k_t(x)}(t)| \xrightarrow{d} \Phi(x)W, \qquad \mathrm{e}^{-\lambda t}|\mathsf{BP}[t, t+s)| \xrightarrow{d} F_R(s)W. \tag{8.4.70}$$

We will not give a proof of Theorem 8.23. Instead, we argue why the residual lifetime and height of individuals are approximately independent from the asymptotic growth of the CTBP that is described by W. Note that W is primarily determined by what happens early on in the CTBP, 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 $|\mathsf{BP}(t)|$ were born in a time that is close to t. Thus, the heights and residual lifetimes of most of the individuals in $\mathsf{BP}(t)$ are described by what happens close to time t. This explains why these random influences are close to independent.

Theorem 8.23 is an extremely powerful result. Unfortunately, the CLT for the height of alive individuals in a CTBP is not known in the most general setting, which is why we have only stated it for age-dependent branching processes.

Finite explosion times. So far, we have only dealt with CTBP satisfying the $X \log X$-condition. For age-dependent CTBPs with offspring B, this in particular implies that $\mathbb{E}[B] < \infty$. We now investigate the case where $\mathbb{E}[B] = \infty$, in which the CTBP may explode in finite time. The results are simplest in the case of exponential lifetimes, with which we start. In Section 8.3, we have seen that the mth split time T_m for a Bellman-Harris process is given by

$$T_m = \sum_{j=1}^{m} E_j/S_j, \tag{8.4.71}$$

where $S_j = \sum_{i=1}^{j} B_i - (i-1)$. There, we assumed that $\mathbb{E}[B^{1+\varepsilon}] < \infty$ for some $\varepsilon > 0$. In Theorem 8.18, the condition on B was weakened to $\mathbb{E}[B \log(B)_+] < \infty$ (see also Theorem 8.23). We now investigate cases where $\mathbb{E}[B] = \infty$, for which it is possible that $\lim_{m\to\infty} T_m = T_\infty < \infty$ a.s. This means that the CTBP *explodes in finite time*, i.e., after a finite amount of time, there are infinitely many individuals alive. The following theorem gives a precise condition when this can occur:

Theorem 8.24 (Finite explosion time)**.** *A Bellman-Harris process with offspring B almost surely explodes in finite time with explosion time given by*

$$T_\infty = \sum_{j=1}^{\infty} E_j/S_j < \infty \qquad a.s. \tag{8.4.72}$$

precisely when $\mathbb{P}(B \geq 1) = 1$ and the probability generating function G_B of the offspring distribution B satifies that there exists $\delta > 0$ such that

$$\int_0^\delta \frac{1}{1 - \mathbb{E}[\mathrm{e}^{-t(B-1)}]} ds < \infty. \tag{8.4.73}$$

The Bellman-Harris process with offspring B satisfying (8.4.73) can also have a finite explosion time when $\mathbb{P}(B \geq 1) < 1$, but then only on the event of survival. Theorem 8.24 also has implications for age-dependent CTBPs:

Exercise 8.24 (Explosion time for age-dependent CTBP). *Show that if the lifetime Y satisfies that there exists an $a > 0$ such that $Y \preceq aE$, then also the age-dependent CTBP with lifetime Y explodes in finite time when (8.4.73) holds.*

Proof of Theorem 8.24. Note that (8.4.73) and $\mathbb{P}(B \geq 1) = 1$ imply that $\mathbb{P}(B \geq 2) > 0$. Indeed, when $\mathbb{P}(B = 1) = 1$, then $\mathbb{E}[\mathrm{e}^{-t(B-1)}] = 1$, and thus (8.4.73) fails to hold.

The birth times $m \mapsto T_m$ are increasing in m, and thus it suffices to prove that

$$\mathbb{E}\Big[\sum_{j=1}^{\infty} E_j/S_j\Big] < \infty. \tag{8.4.74}$$

For this, we compute that

$$\mathbb{E}\Big[\sum_{j=1}^{\infty} E_j/S_j\Big] = \sum_{j=1}^{\infty} \mathbb{E}\big[1/S_j\big]. \tag{8.4.75}$$

We rewrite, for any $\delta > 0$,

$$\mathbb{E}\big[1/S_j\big] = \int_0^{\delta} \mathbb{E}[\mathrm{e}^{-tS_j}]dt + \mathbb{E}\big[\mathrm{e}^{-\delta S_j}/S_j\big]. \tag{8.4.76}$$

The contribution of the second term is bounded by

$$\begin{aligned}\sum_{j=1}^{\infty} \mathbb{E}\big[\mathrm{e}^{-\delta S_j}/S_j\big] \leq \sum_{j=1}^{\infty} \mathbb{E}\big[\mathrm{e}^{-\delta S_j}\big] &\leq \sum_{j=1}^{\infty} \big(\mathbb{E}[\mathrm{e}^{-\delta(B-1)}]\big)^j \\ &\leq 1/(1 - \mathbb{E}[\mathrm{e}^{-\delta(B-1)}]) < \infty,\end{aligned} \tag{8.4.77}$$

since $\mathbb{P}(B \geq 2) > 0$ implies that $S_j \geq 1$ a.s. for every $j \geq 1$, as well as $\mathbb{E}[\mathrm{e}^{-\delta(B-1)}] < 1$.

In a similar way, we compute the contribution of the first term as

$$\begin{aligned}\sum_{j=1}^{\infty} \int_0^{\delta} \mathbb{E}[\mathrm{e}^{-tS_j}]dt &= \int_0^{\delta} \sum_{j=1}^{\infty} \mathrm{e}^{-t}\big(\mathbb{E}[\mathrm{e}^{-t(B-1)}]\big)^j dt \\ &= \int_0^{\delta} \frac{\mathrm{e}^{-t}\mathbb{E}[\mathrm{e}^{-t(B-1)}]}{1 - \mathbb{E}[\mathrm{e}^{-t(B-1)}]} dt.\end{aligned} \tag{8.4.78}$$

By assumption, this is finite. Since $\mathbb{E}[T_\infty] < \infty$, certainly $T_\infty < \infty$ a.s. □

We next investigate two related theorems for age-dependent CTBPs, for which we do not know the precise conditions for explosion in finite time to hold. In their statements, we let $G_B(s) = \mathbb{E}[s^B]$ be the probability generating function of the offspring B:

Theorem 8.25 (Finite explosion time age-dependent CTBPs). *An age-dependent CTBP with offspring B and lifetime Y a.s. explodes in finite time precisely when $\mathbb{P}(B \geq 1) = 1$ and the probability generating function G_B of the offspring distribution B satifies that there exists $\delta > 0$ such that, for all $0 \leq t \leq \delta$,*

$$1 - G_B^{-1}(t) \leq \int_0^t G_Y(s)ds < \infty. \tag{8.4.79}$$

Theorem 8.26 (Infinite explosion time age-dependent CTBPs)**.** *For every B with $\mathbb{E}[B] = \infty$, there exists a non-negative random variable Y such that the age-dependent CTBP with offspring B and lifetime Y does not explode in finite time.*

8.5 Universality for first-passage percolation on CM

In this section, we use the general theory of age-dependent continuous-time branching processes in order to show that Theorem 8.5 extends to first-passage percolation with *general* continuous edge-weight distributions. We start by introducing the necessary notation. We investigate the configuration model $\mathrm{CM}_n(\boldsymbol{d})$ under the degree-regularity condition in [I, Condition 7.7(a)-(c)]. We need to extend [I, Condition 7.7(a)] slightly in order to have that the branching process approximation of local neighborhoods in $\mathrm{CM}_n(\boldsymbol{d})$ satisfies the $X \log X$ condition. We assume a *uniform $X^2 \log X$ condition* for the degrees of $\mathrm{CM}_n(\boldsymbol{d})$, stating that

$$\limsup_{n\to\infty} \mathbb{E}[D_n^2 \log (D_n)_+] = \mathbb{E}[D^2 \log(D)_+] < \infty. \tag{8.5.1}$$

The condition in (8.5.1) implies that the $X \log X$-condition is satisfied for the limiting CTBP with offspring distribution $D^\star - 1$, where $D^\star$ is the size-biased version of D. Indeed, define the size-biased distribution $F^\star$ of the random variable D with distribution function F by

$$F^\star(x) = \mathbb{E}[D\mathbb{1}_{\{D\le x\}}]/\mathbb{E}[D], \quad x \in \mathbb{R}. \tag{8.5.2}$$

Now let $(\mathsf{BP}(t))_{t\ge 0}$ denote the following CTBP:
(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 in Condition 7.7(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 $D_v^\star - 1$ children, where $D_v^\star \sim F^\star$. Lifetimes and number of offspring across individuals are independent.

Note that in the above construction, by [I, Condition 7.7(b)], if we let $X_v = D_v^\star - 1$ be the number of children of an individual then the expected number of children satisfies

$$\mathbb{E}[X_v] = \mathbb{E}[D_v^\star - 1] = \nu > 1, \tag{8.5.3}$$

The CTBP defined above is a *splitting* process, with lifetime distribution F_Y and offspring distribution $D^\star - 1$, except for the root, which has offspring distribution D. Thus, this is a *two-stage* CTBP. Naturally, the Malthusian parameter of the two-stage CTBP is equal to the one with offspring distribution $D^\star - 1$. By (8.4.30), the Malthusian parameter λ of the branching process $\mathsf{BP}(\cdot)$ is the unique solution of the equation

$$\nu \int_0^\infty \mathrm{e}^{-\lambda t} dF_Y(t) = 1. \tag{8.5.4}$$

Since $\nu > 1$, we obtain that $\lambda \in (0, \infty)$. We also let λ_n be the solution to (8.5.4) with ν replaced with $\nu_n = \mathbb{E}[D_n(D_n-1)]/\mathbb{E}[D_n]$. Clearly, $\lambda_n \to \lambda$, when Condition 7.7(c) holds, and $|\lambda_n - \lambda| = O(|\nu_n - \nu|)$.

By Theorem 8.18, there exists a random variable W such that

$$\mathrm{e}^{-\lambda t}|\mathsf{BP}(t)| \xrightarrow{d} W. \tag{8.5.5}$$

Recall that $\bar{F}_Y$ denotes the stable-age distribution in Definition 8.20, 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 8.20 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 8.27 (Joint convergence hopcount and weight)**.** *Consider the configuration model* $\mathrm{CM}_n(\boldsymbol{d})$ *with degrees satisfying [I, Condition 7.7(a)-(c)] and (8.5.1), 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* $\alpha_n \to \alpha, \lambda_n \to \lambda$*, such that the hopcount* H_n *and weight* W_n *of the optimal path between two uniformly selected vertices conditioned on being connected, satisfy*

$$\Big(\frac{H_n - \alpha_n \log n}{\sqrt{\beta \log n}},\ W_n - \frac{1}{\lambda_n}\log n\Big) \xrightarrow{d} (Z, Q), \tag{8.5.6}$$

as $n \to \infty$*, where*
(a)

$$\alpha_n = \frac{1}{\lambda_n \bar{\nu}_n}, \qquad \beta = \frac{\bar{\sigma}^2}{\bar{\nu}^3 \lambda}, \tag{8.5.7}$$

(b) Z *and* Q *are independent and* Z *has a standard normal distribution, while* Q *has a continuous distribution given by*

$$Q = \frac{1}{\lambda}\left(-\log W^{(1)} - \log W^{(2)} - \Lambda + c\right), \tag{8.5.8}$$

where $\mathbb{P}(\Lambda \le x) = \mathrm{e}^{-\mathrm{e}^{-x}}$*, so that* Λ *is a standard Gumbel random variable,* $W^{(1)}, W^{(2)}$ *are two independent copies of the variable* W *in (8.5.5), also independent from* Λ*, and* c *is the constant*

$$c = \log(\mu(\nu-1)^2/(\nu\alpha\bar{\nu})). \tag{8.5.9}$$

Theorem 8.27 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 8.28 (Asymptotic mean)**.** *We can replace* λ_n *and* α_n *by their limits* λ *and* $\alpha = 1/(\lambda\bar{\nu})$ *in (8.5.6) 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 8.5.1, we first use Theorem 8.27 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 8.5.2, we explain the key ingredients in the proof of Theorem 8.27.

8.5.1 Extensions and examples

Extension to related random graphs. We start by extending Theorem 8.27 to *uniform* random graphs with a prescribed degree sequence:

Theorem 8.29 (Extension to uniform random graphs with prescribed degrees)**.** *Under the conditions of Theorem 8.27 the results in Theorem 8.27 apply to uniform random graphs with prescribed degree sequence* $\boldsymbol{d}$ $\mathrm{UG}_n(\boldsymbol{d})$.

The proof of Theorem 8.29 follows rather directly from that of Theorem 8.27, by conditioning on simplicity. By [I, Theorem 7.11], under [I, Condition 7.7(a-c)],

$$\lim_{n\to\infty} \mathbb{P}(\mathrm{CM}_n(\boldsymbol{d}) \text{ simple}) = \mathrm{e}^{-\nu/2-\nu^2/4}. \tag{8.5.10}$$

The proof of Theorem 8.27 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, W_n). This explains Theorem 8.29.

We next extend Theorem 8.27 to rank-1 inhomogeneous random graphs:

Theorem 8.30 (Extension to rank-1 inhomogeneous random graphs)**.** *Let* $\boldsymbol{w}$ *satisfy [I, Condition 6.4(a)-(c)], and further assume that*

$$\lim_{n\to\infty} \mathbb{E}[W_n^2 \log (W_n)_+] = \mathbb{E}[W^2 \log(W)_+]. \tag{8.5.11}$$

Then, the results in Theorem 8.27 also hold for $\mathrm{GRG}_n(\boldsymbol{w})$, $\mathrm{CL}_n(\boldsymbol{w})$ *and* $\mathrm{NR}_n(\boldsymbol{w})$.

The proof of Theorem 8.30 is similar to the proof of Theorem 4.9 in Section 4.1.4. Indeed, we already know that [] implies that the degree sequence of $\mathrm{GRG}_n(\boldsymbol{w})$ satisfies [I, Condition 7.7(a)-(c)]. Therefore, the only thing left is to prove that (8.5.11) implies (8.5.1), which we omit.

Exercise 8.25 (Example of exponential weights)**.** *Show that Theorem 8.27 implies Theorem 8.5 in the case of exponential weights.*

Example 8.31 (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. By Theorem 5.28, $(H_n - \log_\nu(n))_{n\geq 1}$ is a tight sequence of random variables. This suggests that, as $k \to \infty$,

$$\lambda \to \log \nu, \qquad \bar{\nu} \to 1, \qquad \frac{\bar{\sigma}^2}{\bar{\nu}^3 \lambda} \to 0. \tag{8.5.12}$$

We now check this intuitive argument. Indeed,

$$\nu \int_0^\infty \mathrm{e}^{-\lambda x} dF_Y(x) = \nu k \int_1^\infty \mathrm{e}^{-\lambda x} \mathrm{e}^{-k(x-1)} dx = \frac{\nu k}{\lambda + k} \mathrm{e}^{-\lambda} = 1. \tag{8.5.13}$$

While solving this equation *explicitly* is hard, it is not too difficult to see that $k \to \infty$ implies that $\lambda \to \log \nu$.

Exercise 8.26 (Stable-age distribution)**.** *Show that for $Y = 1 + E/k$ and E an exponential with parameter 1, the stable-age distribution in Definition 8.20 is equal to $1 + \mathsf{Exp}(k + \lambda)$.*

By Exercise 8.26, the stable-age distribution is equal to $1 + \mathsf{Exp}(k + \lambda)$, so that $\bar{\nu} = 1 + 1/(k + \lambda)$, while $\bar{\sigma}^2 = 1/(k + \lambda)^2 \to 0$. Therefore, $\bar{\nu} \sim 1$, which in turn also implies that $\lambda\bar{\nu} \to \log \nu$. Further,

$$\frac{\bar{\sigma}^2}{\bar{\nu}^3 \lambda} = k^{-2} (\log \nu)^{-1} (1 + o(1)) \to 0. \tag{8.5.14}$$

This shows that the two settings of graph distances and FPP with weights $Y \stackrel{d}{=} 1 + \mathsf{Exp}(1)/k$ match up nicely when $k \to \infty$.

8.5.2 Overview of proof of Theorem 8.27

To understand the minimal-weight path between these vertices, 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 referred to the flow cluster or the minimal-weight graph of vertex U_i. When the two flows collide or create prospective collision edges, then these generate prospective minimal-weight paths.

In Section 8.3, we have performed these flows by first growing $(\mathsf{SWG}_t^{(1)})_{t\geq 0}$ until it contains $a_n = \sqrt{n}$ vertices, followed by the growth of $(\mathsf{SWG}_t^{(2)})_{t\geq 0}$ until an edge was created linking $\mathsf{SWG}_t^{(2)}$ to $\mathsf{SWG}^{(1)}_{T^{(1)}_{a_n}}$. This construction crucially depends on the memoryless property of the exponential distribution, which implies that the residual lifetimes of all edges incident to $\mathsf{SWG}^{(1)}_{T^{(1)}_{a_n}}$ 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 $(\mathsf{SWG}_t^{(1)})_{t\geq 0}$ and $(\mathsf{SWG}_t^{(2)})_{t\geq 0}$ *simultaneously*, and each time we find a vertex, we check whether any of its half-edges are paired to half-edges in the other SWG. When this is the case, however, this edge has not been completely filled by the fluid, so that it could create the minimal-weight path, but it might also not, as this depends on the residual lifetime of the edge in question, as well as all other edges linking to two SWG's.

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 $\mathsf{A}(t)$ of half-edges that are incident to a vertex in the flow cluster, but who have not yet been completely filled. The alive set $\mathsf{A}(t)$ only changes at random times $T_0 = 0 < T_1 < T_2 < \ldots$ 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 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 $\mathsf{A}(0)$ at time 0. For $y \in \mathsf{A}(0)$, we define $I(y) = i$ if the half-edge y is incident to U_i, $i = 1, 2$, and we define $(R_0(y))_{y\in\mathsf{A}(0)}$ as an i.i.d. sequence of lifetimes having distribution function F_Y.

We denote the set of alive half-edges at time t by $\mathsf{A}(t)$. For $y \in \mathsf{A}(t)$, 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 $y \in \mathsf{A}(0)$. When we introduce new half-edges in $\mathsf{A}(t)$ at later times we will specify the heights and labels of these half-edges. Now define $T_1 = \min_{y\in\mathsf{A}(0)} R_0(y)$ and denote by $y_0^\star$ the half-edge equal to the argument of this minimum, hence $R_0(y_0^\star) = \min_{y\in\mathsf{A}(0)} R_0(y)$. Since lifetimes have a continuous distribution, $y_0^\star$ is a.s. *unique*. Now set $\mathsf{A}(t) = \mathsf{A}(0)$, $0 \le t < T_1$, i.e., the set of alive half-edges remains the same during the interval $[0, T_1)$, and define the flow cluster $\mathsf{SWG}(t)$, for $0 \le t < T_1$, by

$$\mathsf{SWG}(t) = \{y, I(y), H(y), R_t(y)\}_{y\in\mathsf{A}(0)}, \tag{8.5.15}$$

where $I(y)$ and $H(y)$ are defined above and $R_t(y) = R_0(y) - t$, $0 \le t \le T_1$, denotes the remaining lifetime of half-edge y. This concludes the initial step in the recursion, where we defined $\mathsf{A}(t)$ and $\mathsf{SWG}(t)$ during the random interval $[0, T_1)$.

We continue recursively, by defining $\mathsf{A}(t)$ and $\mathsf{SWG}(t)$ during the random interval $[T_k, T_{k+1})$, given that the processes are defined on $[0, T_k)$. At time $t = T_k$,we let $y_{k-1}^\star$

be the argument of $\min_{y\in\mathsf{A}(t)} R_t(y)$, and we remove $y^\star_{k-1}$ from the set $\mathsf{A}(t-)$. We then pair $y^\star_{k-1}$ to a uniform available half-edge, which we denote by z_k. By construction, we know that $z_k \equiv P_{y^\star_{k-1}} \notin \mathsf{A}(t-)$, 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 $\mathsf{A}(t-)$. All half-edges incident to V_{z_k} that are part of a self-loop or are paired to a half-edge incident to $\mathsf{A}(t-)$ are removed from vertex V_{z_k}. We also remove the involved half-edges from the set $\mathsf{A}(t-)$. We will discuss the role of the half-edges incident to V_{z_k} that are paired to half-edges in $\mathsf{A}(t-)$ in more detail in the next paragraph below.

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 $\mathsf{A}(T_k)$, and we define $I(x) = I(y^\star_{k-1})$, $H(x) = H(y^\star_{k-1}) + 1$, while $R_{T_k}(x)$ is an i.i.d. lifetime with distribution F_Y. We now set $\mathsf{A}(t) = \mathsf{A}(T_k)$, $T_k \le t < T_{k+1}$, where $T_{k+1} = T_k + \min_{y\in\mathsf{A}(T_k)} R_{T_k}(y)$, and where the minimizing half-edge is called $y^\star_k$. Furthermore, for $t \in [T_k, T_{k+1})$, we can define $\mathsf{SWG}(t)$ by (8.5.15), 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 $\mathsf{A}(t-)$ by X_k. Later, it will also be convenient to introduce $B_k = d_{V_{z_k}} - 1$. Let $S_k = |\mathsf{A}(T_k)|$, so that $S_0 = X_0^{(1)} + X_0^{(2)}$, while S_k satisfies the recursion

$$S_k = S_{k-1} + X_k - 1. \tag{8.5.16}$$

This describes the evolution of $(\mathsf{SWG}(t))_{t\ge 0}$.

Cycle edges and collision edges. At the times T_k, $k \ge 1$, we find the half-edge $y^\star_{k-1}$ which is paired to $z_k = P_{y^\star_{k-1}}$, and for each of the other half-edges x incident to V_{z_k}, we check whether or not $P_x \in \mathsf{A}(T_k-)$, where P_x is the half-edge to which x is paired. The half-edges paired to alive half-edges in $\mathsf{A}(T_k-)$ 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 minimal-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

$$\Big(\big(T_k, I(z_k), H(z_k), H(P_{x_k}), R_{T_k}(P_{x_k})\big)\Big)_{k\ge 0}. \tag{8.5.17}$$

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 (8.5.17). In this definition it is tempting to write $I(x_k)$ and $H(x_k)$, but note that $x_k \notin \mathsf{A}(T_k)$, whereas its sibbling half-edge does satisfy $z_k \in \mathsf{A}(T_k)$, and, moreover, x_k and z_k have the same ancestor and the same height.

With some abuse of notation we denote the ith 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 $T_i^{(\mathrm{col})}$ be the time of creation of the ith 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 $2T_i^{(\mathrm{col})} + R_{T_i^{(\mathrm{col})}}(P_{x_i})$,

so that the smallest weight of all paths between U_1 and U_2 equals

$$W_n = \min_{i\geq 0}[2T_i^{(\mathrm{col})} + R_{T_i^{(\mathrm{col})}}(P_{x_i})]. \tag{8.5.18}$$

Let $I^\star$ denote the minimizer of $i \mapsto 2T_i^{(\mathrm{col})} + R_{T_i^{(\mathrm{col})}}(P_{x_i})$, then

$$H_n = H(z_{I^\star}) + H(P_{x_{I^\star}}) + 1. \tag{8.5.19}$$

Of course, (8.5.18) and (8.5.19) need a proof, which we give now.

Proof that W_n given by (8.5.18) yields the minimal weight. Observe that each path between U_1 and U_2 has a weight $\mathcal{C}$ that can be written in the form $2T_i^{(\mathrm{col})} + R_{T_i^{(\mathrm{col})}}(P_{x_i})$ for some $i \geq 0$. Indeed, let $(i_0 = U_1, i_1, i_2, \ldots, i_k = U_2)$ 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 $Y_{e_1} + \cdots + Y_{e_j} \leq \mathcal{C}/2$. Then, we write

$$\mathcal{C} = \begin{cases} 2\sum_{s=1}^{j} Y_{e_s} + [\sum_{s=j+1}^{k} Y_{e_s} - \sum_{s=1}^{j} Y_{e_s}], & \text{when } \sum_{s=1}^{j} Y_{e_s} < \sum_{s=j+1}^{k} Y_{e_s}, \\ 2\sum_{s=j+1}^{k} Y_{e_s} + [\sum_{s=1}^{j} Y_{e_s} - \sum_{s=j+1}^{k} Y_{e_s}], & \text{when } \sum_{s=1}^{j} Y_{e_s} > \sum_{s=j+1}^{k} Y_{e_s}, \end{cases} \tag{8.5.20}$$

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 (8.5.18), 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 to which . Thus, in fact, by the redistribution of the weights in (8.5.18) is an equality in distribution. This completes the proof of the claim. □

Basic constructions and properties. To state our main technical result concerning the appearance of collision edges, we need to define some new constructs. We start by defining a rescaled version of the point process corresponding to the points in (8.5.17). Let us first setup some notation. For $i \in \{1, 2\}$ and $t \geq 0$, we let

$$|\mathsf{SWG}(t)| = \#\{y \in \mathsf{A}(t)\}, \qquad |\mathsf{SWG}^{(i)}(t)| = \#\{y \in \mathsf{A}(t)\colon I(y) = i\}, \tag{8.5.21}$$

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, $\mathsf{SWG}^{(1)}(t)$ and $\mathsf{SWG}^{(2)}(t)$ 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 $\mathcal{F}_s = \sigma((\mathsf{SWG}(t))_{t\in[0,s]}$ denoting the sigma-algebra generated by the minimal-weight graph up to time s. This informally means that $\mathcal{F}_s = \sigma((\mathsf{SWG}(t))_{t\in[0,s]}$ contains all the information of the flow up to time s.

Fix a deterministic sequence $s_n \to \infty$ that will be chosen later on. Now let

$$t_n = \frac{1}{2\lambda_n}\log n, \qquad \bar{t}_n = \frac{1}{2\lambda_n}\log n - \frac{1}{2\lambda_n}\log\left(W_{s_n}^{(1)}W_{s_n}^{(2)}\right), \tag{8.5.22}$$

where, for $s \geq 0$,

$$W_s^{(i)} = \mathrm{e}^{-\lambda_n s}|\mathsf{SWG}^{(i)}(s)|. \tag{8.5.23}$$

Note that $\mathrm{e}^{\lambda_n t_n} = \sqrt{n}$, so that at time t_n, both $|\mathsf{SWG}^{(i)}(s)|$ 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 $s_n \to \infty$ such that $\mathsf{SWG}^{(i)}(t)$ for $t \leq s_n$ can be coupled with two independent two-stage branching processes $\mathsf{BP}^{(i)}(t)$ such that $\{\mathsf{BP}(t) = \mathsf{SWG}(t)\ \forall t \leq s_n\}$ whp (see (8.5.44) below).

Define the *residual lifetime distribution* F_R to have density f_R given by

$$f_R(x) = \frac{\int_0^\infty \mathrm{e}^{-\lambda y} f_Y(x+y)\,dy}{\int_0^\infty \mathrm{e}^{-\lambda y}[1 - F_Y(y)]\,dy}. \tag{8.5.24}$$

Recall that the ith 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 SWG. In terms of the above definitions, we define

$$\bar{T}_i^{(\mathrm{col})} = T_i^{(\mathrm{col})} - \bar{t}_n, \qquad \bar{H}_i^{(\mathrm{or})} = \frac{H(x_i) - t_n/\bar{\nu}_n}{\sqrt{\bar{\sigma}^2 t_n/\bar{\nu}^3}}, \qquad \bar{H}_i^{(\mathrm{de})} = \frac{H(P_{x_i}) - t_n/\bar{\nu}_n}{\sqrt{\bar{\sigma}^2 t_n/\bar{\nu}^3}}, \tag{8.5.25}$$

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

$$\Xi_i = \big(\bar{T}_i^{(\mathrm{col})}, I(x_i), \bar{H}_i^{(\mathrm{or})}, \bar{H}_i^{(\mathrm{de})}, R_{T_i}(P_{x_i})\big). \tag{8.5.26}$$

Then, for sets A in the Borel $\sigma-$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

$$\Pi_n(A) = \sum_{i\geq 1} \delta_{\Xi_i}(A), \tag{8.5.27}$$

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. We now come to our main result in the proof:

Theorem 8.32 (PPP limit of collision edges)**.** *Consider the distribution of the point process $\Pi_n \in \mathcal{M}(\mathcal{S})$ defined in (8.5.27) conditional on $(\mathsf{SWG}(s))_{s\in[0,s_n]}$ such that $W_{s_n}^{(1)} > 0$ and $W_{s_n}^{(2)} > 0$. Then Π_n converges in distribution as $n \to \infty$ to a Poisson Point Process (PPP) Π with intensity measure*

$$\lambda(dt \times i \times dx \times dy \times dr) = \frac{2\nu f_R(0)}{\mathbb{E}[D]} \mathrm{e}^{2\lambda t} dt \otimes \{1/2, 1/2\} \otimes \Phi(dx) \otimes \Phi(dy) \otimes F_R(dr). \tag{8.5.28}$$

Theorem 8.32 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 $\int_A \lambda(dt \times i \times dx \times dy \times dr)$.

[1] Let $\mathcal{M}(\mathcal{S})$ denote the space of all simple locally finite point processes on $\mathcal{S}$ equipped with the vague topology (see e.g. [81]). 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 8.32.

Completion of the proof of Theorem 8.27. Let us now prove Theorem 8.27 subject to Theorem 8.32. First of all, by (8.5.25), (8.5.18) and (8.5.19),

$$\Big(\frac{H_n - \frac{1}{\lambda_n \bar{\nu}} \log n}{\sqrt{\frac{\bar{\sigma}^2}{\bar{\nu}^3 \lambda} \log n}}, W_n - \frac{1}{\lambda_n} \log n\Big), \tag{8.5.29}$$

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).

Recall that $I^\star$ denotes the minimizer of $i \mapsto 2T_i^{(\mathrm{col})} + R_{T_i^{(\mathrm{col})}}(P_{x_i})$. By (8.5.18), the weight W_n as well as the value of $I^\star$, 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 (8.5.28), 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 (8.5.25),

$$\frac{H_n - \frac{1}{\lambda_n \bar{\nu}_n} \log n}{\sqrt{\frac{\bar{\sigma}^2}{\bar{\nu}^3 \alpha} \log n}} = \frac{1}{2}\sqrt{2}\bar{H}_{I^\star}^{(\mathrm{or})} + \frac{1}{2}\sqrt{2}\bar{H}_{I^\star}^{(\mathrm{de})} + o_{\mathbb{P}}(1). \tag{8.5.30}$$

By Theorem 8.32, the random variables $(\bar{H}_{I^\star}^{(\mathrm{or})}, \bar{H}_{I^\star}^{(\mathrm{de})})$ converge in distribution to two independent standard normals, so that also the left-hand side of (8.5.30) converges in distribution to a standard normal.

The limiting distribution of the weight W_n is slightly more involved. By (8.5.22), (8.5.18) and (8.5.25),

$$\begin{aligned} W_n - \frac{1}{\lambda_n} \log n &= W_n - 2t_n = W_n - 2\bar{t}_n - \frac{1}{\lambda_n} \log(W_{s_n}^{(1)} W_{s_n}^{(2)}) \\ &= -\frac{1}{\lambda_n} \log(W_{s_n}^{(1)} W_{s_n}^{(2)}) + \min_{i \geq 1}[2T_i^{(\mathrm{col})} + R_{T_i^{(\mathrm{col})}}(P_{x_i})] - 2\bar{t}_n \\ &= -\frac{1}{\lambda_n} \log(W_{s_n}^{(1)} W_{s_n}^{(2)}) + \min_{i \geq 1}[2\bar{T}_i^{(\mathrm{col})} + R_{T_i^{(\mathrm{col})}}(P_{x_i})]. \end{aligned} \tag{8.5.31}$$

By (8.5.44) 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 (8.4.24) in Theorem 8.18. Hence,

$$W_n - \frac{1}{\lambda_n} \log n \xrightarrow{d} -\frac{1}{\lambda} \log(W^{(1)} W^{(2)}) + \min_{i \geq 1}[2P_i + R_i], \tag{8.5.32}$$

where $(P_i)_{i \geq 1}$ form a PPP with intensity $\frac{2\nu f_R(0)}{\mu} \mathrm{e}^{2\lambda t} dt$, 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 $2c\mathrm{e}^{2\lambda t}$ is a rescaled Gumbul distribution:

Exercise 8.27. *Let (P_i) be a Poisson point proces with intensity measure $\lambda(t) = 2c\mathrm{e}^{2\lambda t}$. 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 8.27, 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 $\frac{2\nu f_R(0)}{\mathbb{E}[D]}\mathrm{e}^{2\lambda t}dt$ 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 $\frac{\nu f_R(0)}{\mathbb{E}[D]}\mathrm{e}^{\lambda t}dt$. According to [99, 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 $\mu(-\infty, x] = \int_{-\infty}^{x} \frac{\nu f_R(0)}{\mathbb{E}[D]}\mathrm{e}^{\lambda t}\,dt$ 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

$$\mathbb{P}(M \geq x) = \exp\{-\frac{\nu f_R(0)}{\mathbb{E}[D]}\mathrm{e}^{\lambda x}\int_0^\infty F_R(z)\mathrm{e}^{-\lambda z}\,dz\}. \tag{8.5.33}$$

Let Λ have a Gumbel distribution, i.e., $\mathbb{P}(\Lambda \leq x) = \mathrm{e}^{-\mathrm{e}^{-x}}$, $x \in \mathbb{R}$, then

$$\mathbb{P}(-a\Lambda + b \geq x) = \mathrm{e}^{-\mathrm{e}^{x/a}\mathrm{e}^{-b/a}}. \tag{8.5.34}$$

From the identity

$$\frac{\nu f_R(0)}{\mathbb{E}[D]}\mathrm{e}^{\lambda x}\int_0^\infty F_R(z)\mathrm{e}^{-\lambda z}\,dz = \mathrm{e}^{x/a}\mathrm{e}^{-b/a}, \tag{8.5.35}$$

we conclude that if we take

$$a = 1/\lambda \qquad \text{and} \qquad b = -\lambda^{-1}\log\Big((\nu f_R(0)/\mathbb{E}[D])\int_0^\infty F_R(z)\mathrm{e}^{-\lambda z}\,dz\Big), \tag{8.5.36}$$

then

$$\min_{i\geq 1}(2P_i + R_i) \stackrel{d}{=} -\lambda^{-1}\Lambda - \lambda^{-1}\log(\nu f_R(0)a/\mathbb{E}[D]), \tag{8.5.37}$$

with $a = \int_0^\infty F_R(z)\mathrm{e}^{-\lambda z}\,dz$. In the following lemma, we simplify the constants a and $f_R(0)$:

Lemma 8.33 (The constant)**.** *The constants $a = \int_0^\infty F_R(z)\mathrm{e}^{-\lambda z}\,dz$ and $f_R(0)$ are given by*

$$a = \bar{\nu}/(\nu - 1), \qquad f_R(0) = \lambda/(\nu - 1). \tag{8.5.38}$$

Consequently, the constant c in the limit variable (8.5.8) equals

$$c = -\log(\nu f_R(0)a/\mu) = \log(\mathbb{E}[D](\nu - 1)^2/(\lambda\nu\bar{\nu})). \tag{8.5.39}$$

Proof. We start by computing $f_R(0)$, for which we note that by (8.5.24) and the definition of the Malthusian parameter in Definition 8.17,

$$f_R(0) = \frac{\int_0^\infty \mathrm{e}^{-\lambda y}f_Y(y)\,dy}{\int_0^\infty \mathrm{e}^{-\lambda y}[1 - F_Y(y)]\,dy} = \Big(\nu\int_0^\infty \mathrm{e}^{-\lambda y}[1 - F_Y(y)]\,dy\Big)^{-1}. \tag{8.5.40}$$

Further, by partial integration,

$$\int_0^\infty \mathrm{e}^{-\lambda y}[1-F_Y(y)]\,dy = \Big[-\frac{1}{\lambda}\mathrm{e}^{-\lambda y}[1-F_Y(y)]\Big]_{y=0}^{\infty} - \frac{1}{\lambda}\int_0^\infty \mathrm{e}^{-\lambda y} f_Y(y)\,dy \tag{8.5.41}$$
$$= \frac{1}{\lambda} - \frac{1}{\lambda\nu} = \frac{\nu-1}{\lambda\nu},$$

where we again use the definition of the Malthusian parameter in Definition 8.17. Combining both equalities yields $f_R(0) = \lambda/(\nu-1)$.

For a, we again use partial integration, followed by the substitution of (8.5.24). This yields

$$a = \int_0^\infty F_R(z)\mathrm{e}^{-\lambda z}\,dz = \frac{1}{\lambda}\int_0^\infty f_R(z)\mathrm{e}^{-\lambda z}\,dz \tag{8.5.42}$$
$$= \frac{\nu}{\nu-1}\int_0^\infty \mathrm{e}^{-\lambda z}\int_0^\infty \mathrm{e}^{-\lambda y} f_Y(y+z)\,dy\,dz,$$

by (8.5.41). The final integral can be computed using

$$\int_{-\infty}^\infty \mathrm{e}^{-\lambda z}\mathbb{1}_{\{z\geq 0\}}\int_{-\infty}^\infty \mathrm{e}^{-\lambda y} f_Y(y+z)\mathbb{1}_{\{y\geq 0\}}\,dy\,dz$$
$$= \int_0^\infty s f_Y(s)\mathrm{e}^{-\lambda s}\,ds = \frac{1}{\nu}\int_0^\infty s\bar{F}_Y(ds) = \bar{\nu}/\nu. \tag{8.5.43}$$

This completes the proof of Theorem 8.27 subject to Theorem 8.32. □

Overview of the proof of Theorem 8.32. We next infomally explain how to prove Theorem 8.32. Recall the minimal-weight graph or flow cluster $\mathsf{SWG}(t)$ defined in the previous section as well as the associated filtration $(\mathcal{F}_t)_{t\geq 0}$. We shall couple these flow clusters from two points with $(\mathsf{BP}(t))_{t\geq 0}$ where $\mathsf{BP}(t) = (\mathsf{BP}^{(1)}(t), \mathsf{BP}^{(2)}(t))$ are two independent CTBPs starting with offspring distribution D, in such a way that for some $s_n \to \infty$

$$\mathbb{P}\Big((\mathsf{SWG}(s))_{s\in[0,s_n]} = (\mathsf{BP}(s))_{s\in[0,s_n]}\Big) = 1 - o(1). \tag{8.5.44}$$

Exercise 8.28 (Perfect coupling for increasing times)**.** *Prove the perfect coupling statement in* (8.5.44) *using Lemma 5.7.*

By (8.5.44), with $W_{s_n}^{(i)} = \mathrm{e}^{-\lambda_n s_n}|\mathsf{SWG}^{(i)}(s_n)|$,

$$\liminf_{\varepsilon\downarrow 0}\liminf_{n\to\infty}\mathbb{P}\Big(W_{s_n}^{(1)}\in[\varepsilon,1/\varepsilon], W_{s_n}^{(2)}\in[\varepsilon,1/\varepsilon]\Big|W_{s_n}^{(1)}>0, W_{s_n}^{(2)}>0\Big) = 1. \tag{8.5.45}$$

Further, the coupling satisfies that, conditionally on $\mathcal{F}_{s_n}$,

$$\mathbb{P}\Big(|\mathsf{SWG}(t_n+B^{(n)})\triangle\mathsf{BP}_{(n)}(t_n+B^{(n)})|\geq\varepsilon_n\sqrt{n}\mid\mathcal{F}_{s_n}\Big)\xrightarrow{\mathbb{P}}0. \tag{8.5.46}$$

For $i \in \{1,2\}$, $k \geq 0$, and $t \geq 0$, we define

$$|\mathsf{SWG}^{(i)}_{\leq k}[t, t+s)| = \#\{y \in \mathsf{A}(t) \colon I(y) = i, H(y) \leq k, R_t(y) \in [0,s)\}, \tag{8.5.47}$$

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 lifetime at most s. To formulate the CLT for the height of vertices, we further write

$$k_t(x) = \frac{t}{\bar{\nu}} + x\sqrt{t\frac{\bar{\sigma}^2}{\bar{\nu}^3}}. \tag{8.5.48}$$

Finally, for a half-edge $y \in \mathsf{A}(t)$, we let $X^\star_y = d_{V_y} - 1$.

We will argue that, for $\mathcal{I} = [a,b) \times \{j\} \times (-\infty, x] \times (-\infty, y] \times [0,s]$ a subset of $\mathcal{S}$,

$$\mathbb{P}(\Pi_n(\mathcal{I}) = 0 \mid \mathcal{F}_{s_n}) \xrightarrow{\mathbb{P}} \exp\Big\{ - \int_a^b \frac{2\nu f_{\scriptscriptstyle R}(0)}{\mu} \mathrm{e}^{2\alpha t} \Phi(x)\Phi(y) F_{\scriptscriptstyle R}(s) dt \Big\}. \tag{8.5.49}$$

By [81, 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 minimal-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

$$\mathbb{P}(\Pi_n(\mathcal{I}) \geq 1 \mid \mathcal{F}_{s_n}) \approx \varepsilon \frac{2\nu f_{\scriptscriptstyle R}(0)}{\mu} \mathrm{e}^{2\alpha a} \Phi(x)\Phi(y) F_{\scriptscriptstyle R}(s). \tag{8.5.50}$$

The number of half-edges z in $\mathsf{SWG}^{(j)}(\bar{t}_n + a)$ 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 $|\mathsf{SWG}^{(j)}[\bar{t}_n + a, \bar{t}_n + b)|$. These half-edges are in turn paired to other half-edges whose siblings can alse be found. 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 a half-edge P_z for which one of the sibbling half-edges pairs to a half-edge incident to the *other* SWG. On average, there are $\nu|\mathsf{SWG}^{(j)}[\bar{t}_n + a, \bar{t}_n + b)|$ sibbling half-edges to P_z, and each of them pairs to a half-edge incident to $\mathsf{SWG}^{(3-j)}(\bar{t}_n + a)$ with probability close to

$$\frac{|\mathsf{SWG}^{(3-j)}(\bar{t}_n + a)|}{\ell_n}.$$

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

$$\frac{\nu}{\ell_n}|\mathsf{SWG}^{(j)}[\bar{t}_n{+}a, \bar{t}_n{+}b)||\mathsf{SWG}^{(3-j)}(\bar{t}_n{+}a)| \approx \frac{\nu}{\ell_n} F_{\scriptscriptstyle R}(b{-}a)|\mathsf{SWG}^{(j)}(\bar{t}_n{+}a)||\mathsf{SWG}^{(3-j)}(\bar{t}_n{+}a)|,$$

where we use Theorem 8.21. We can approximate $F_{\scriptscriptstyle R}(b-a) \approx (b-a) f_{\scriptscriptstyle R}(0) = \varepsilon f_{\scriptscriptstyle R}(0)$. Also, replacing $|\mathsf{SWG}^{(j)}(\bar{t}_n + a)| \approx W^{(j)}_{s_n} \mathrm{e}^{\lambda_n(\bar{t}_n + a)}$, as suggested by Theorem 8.18, leads us to

$$\frac{\nu}{\ell_n}|\mathsf{SWG}^{(j)}[\bar{t}_n + a, \bar{t}_n + b)||\mathsf{SWG}^{(3-j)}(\bar{t}_n + a)| \approx \frac{\varepsilon f_{\scriptscriptstyle R}(0)}{\ell_n} W^{(j)}_{s_n} W^{(3-j)}_{s_n} \mathrm{e}^{2\lambda_n(\bar{t}_n + a)}.$$

By the definition of $\bar{t}_n$ in (8.5.22)

$$W_{s_n}^{(j)} W_{s_n}^{(3-j)} \mathrm{e}^{2\lambda_n(\bar{t}_n+a)} = W_{s_n}^{(1)} W_{s_n}^{(2)} \mathrm{e}^{2\lambda_n(\bar{t}_n+a)} = n\mathrm{e}^{2\lambda_n a}. \tag{8.5.51}$$

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

$$\frac{\nu\varepsilon f_R(0)}{\ell_n} n\mathrm{e}^{2\lambda_n a} = \frac{\nu\varepsilon f_R(0)}{\mathbb{E}[D_n]} \mathrm{e}^{2\lambda_n a} \to \frac{\nu\varepsilon f_R(0)}{\mathbb{E}[D]} \mathrm{e}^{2\lambda a}, \tag{8.5.52}$$

as required. Further, by Exercise 8.23 (which is based on Theorems 8.21 and 8.22) the height of the two half-edges that form the collision edge are close to normal, and the residual lifetime of the collision half-edge has distribution close to F_R. This explains Theorem 8.32.

8.6 Related 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 8.6.1 we discuss further results for FPP on the complete graph, and in Section 8.6.1 we discuss further results for FPP on random graphs.

8.6.1 Related results for first-passage percolation on the complete graph

Extremal functionals for FPP on the complete graph. We now investigate extremal functionals for FPP on the complete graph, by extending the results in Section 8.2. We start by investigating the lenths of the maximal paths. Indeed, recall that $H_n(i,j)$ is the length of the optimal path between vertices $i, j \in [n]$. 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 8.34 (Hopcount diameter)**.** *For FPP on the complete graph with exponential edge weights, as $n \to \infty$,*

$$\frac{\max_{j\in[n]} H_n(1,j)}{\log n} \xrightarrow{\mathbb{P}} \mathrm{e}, \qquad \frac{\max_{i,j\in[n]} H_n(i,j)}{\log n} \xrightarrow{\mathbb{P}} \alpha^\star, \tag{8.6.1}$$

where $\alpha^\star \approx 3.5911$ is the unique solution of the equation $\alpha \log \alpha - \alpha = 1$.

Exercise 8.29 (Equation for the height)**.** *Note that* e *is the unique solution to the equation $\alpha \log \alpha - \alpha = 0$.*

Exercise 8.30 (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*

$$\frac{\max_{j\in[n]} H_n(1,j)}{\log n} \xrightarrow{\mathbb{P}} \mathrm{e}. \tag{8.6.2}$$

We now give some intuition for the results in Theorem 8.34, relying on the analysis in the proof of Theorem 8.3. We only prove the upper bound. First note that, by Boole's inequality,

$$\mathbb{P}(\max_{i,j\in[n]} H_n(i,j) \geq \alpha \log n) \leq n\mathbb{P}(\max_{j\in[n]} H_n(1,j) \geq \alpha \log n). \tag{8.6.3}$$

Again by Boole's inequality,

$$\mathbb{P}(\max_{j\in[n]} H_n(1,j) \geq \alpha \log n) \leq n\mathbb{P}(H_n(1,2) \geq \alpha \log n), \tag{8.6.4}$$

so that, for the upper bound, it suffices to bound the tails of the distribution of $H_n(1,2)$. By Lemma 8.4, for every $t \geq 0$,

$$\mathbb{P}(H_n(1,2) \geq \alpha \log n) \leq \mathrm{e}^{-t\alpha \log n}\mathbb{E}[\mathrm{e}^{tH_n(1,2)}] = \mathrm{e}^{-t\alpha \log n}\prod_{i=2}^{n}\Big(1-\frac{1}{i}+\mathrm{e}^t\frac{1}{i}\Big). \tag{8.6.5}$$

We further bound, using $1+x \leq \mathrm{e}^x$ and $\sum_{i=2}^n 1/i \leq \int_1^n dx/x = \log n$,

$$\prod_{i=2}^{n}\Big(1-\frac{1}{i}+\mathrm{e}^t\frac{1}{i}\Big) = \prod_{i=2}^{n}\Big(1+(\mathrm{e}^t-1)\frac{1}{i}\Big) \leq \mathrm{e}^{\sum_{i=1}^n(\mathrm{e}^t-1)\frac{1}{i}} \leq \mathrm{e}^{(\mathrm{e}^t-1)\log n}. \tag{8.6.6}$$

This leads to

$$\mathbb{P}(H_n(1,2) \geq \alpha \log n) \leq \mathrm{e}^{\log n(-t\alpha+\mathrm{e}^t-1)}. \tag{8.6.7}$$

Opimizing over $t \geq 0$ yields $t = \log \alpha$, so that

$$\mathbb{P}(H_n(1,2) \geq \alpha \log n) \leq n^{-(\alpha\log\alpha-\alpha+1)}. \tag{8.6.8}$$

Thus,

$$\mathbb{P}(\max_{j\in[n]} H_n(1,j) \geq \alpha \log n) = o(1) \qquad \text{when} \qquad \alpha\log\alpha-\alpha+1>1, \tag{8.6.9}$$

which occurs when $\alpha > \mathrm{e}$, while

$$\mathbb{P}(\max_{i,j\in[n]} H_n(i,j) \geq \alpha \log n) = o(1) \qquad \text{when} \qquad \alpha\log\alpha-\alpha+1>2, \tag{8.6.10}$$

which occurs when $\alpha > \alpha^\star$ which is the solution of $\alpha\log\alpha-\alpha = 1$. We conclude that the upper bounds in Theorem 8.34 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 $\alpha < \mathrm{e}$, and that whp there really exist two vertices i,j for which $H_n(i,j) \geq \alpha \log n$ for any $\alpha < \alpha^\star$.

The lower bound on $\max_{i,j\in[n]} H_n(i,j)$ in (8.6.1) is the hard part, and is proved in [2] using a delicate conditional second moment method. The basic starting point is that *if* a path $\pi\colon v_0 \to v_k$ 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 [2, 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

$$\mathbb{E}[\mathcal{P}_{k,\varepsilon}(n)] \sim n^{k+1}\Big(\frac{(1-\varepsilon)\log n}{n}\Big)^k \frac{1}{k!} \sim \frac{n}{\sqrt{2\pi k}}\Big(\frac{\mathrm{e}\log n}{k}\Big)^k. \tag{8.6.11}$$

This tends to infinity when $\alpha < \alpha^\star$ 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, in [2] a second moment method is used 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.

We next consider the fluctuations of the weight diameter for first-passage percolation on the complete graph. Let us first define the limiting weighted random graph that arises in the description of our main result. The vertex set of this graph will be the set of positive integers $\mathbb{Z}_+ = \{1, 2, \ldots\}$. Let PP be a Poisson process on $\mathbb{R}$ with intensity measure having density

$$\lambda(y) = \exp(-y), \qquad y \in \mathbb{R}. \tag{8.6.12}$$

Exercise 8.31 implies that $\Xi' = \max\{x\colon x \in \mathsf{PP}\} < \infty$ a.s. Thus, we can order the points in PP as $Y_1 > Y_2 > \cdots$. Let $\{\Lambda_{st}\colon s, t \in \mathbb{Z}_+, s < t\}$ be a family of independent standard Gumbel random variables. Define

$$\Xi = \max\{Y_s + Y_t - \Lambda_{st}\colon s, t \in \mathbb{Z}_+, s < t\}. \tag{8.6.13}$$

It can be shown that $\Xi < \infty$ almost surely:

Exercise 8.31 (The random variables $\Xi, \Xi' < \infty$ almost surely)**.** *Show that* $\Xi' < \infty$ *and* $\Xi < \infty$ *almost surely.*

The main result concerning the weight diameter for FPP on the complete graph is as follows:

Theorem 8.35 (Fluctuations of the weight diameter)**.** *For FPP on the complete graph with exponential edge weights, as* $n \to \infty$*,*

$$n \max_{i,j\in[n]} \mathcal{C}_n(i,j) - 3\log n \xrightarrow{d} \Xi. \tag{8.6.14}$$

The proof of Theorem 8.35 is a close adaptation of the proof of the upper bound in (8.2.3) in Theorem 8.1. 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 $X_{(1)} < X_{(2)} < \cdots < X_{(n)}$. Further, let V_i be the vertex corresponding to $X_{(i)}$. Then, obviously,

$$\max_{i,j\in[n]} \mathcal{C}_n(i,j) = \max_{i,j\in[n]} \mathcal{C}_n(V_i, V_j). \tag{8.6.15}$$

The point process $(X_{(i)} - \log n)_{i\in[n]}$ converges to the point-process with intensity λ in (8.6.12). For $i < j$ fixed, we can write

$$\mathcal{C}_n(V_i, V_j) \stackrel{d}{=} X_{(i)} + X_{(j)} + \mathcal{C}_n(\{V_i, U_i\}, \{V_j, U_j\}), \tag{8.6.16}$$

where $\mathcal{C}_n(A,B)$ denotes the minimal weight between the subsets $A,B\subseteq[n]$ and U_i is such that $E_{V_iU_i}=X_{(i)}$ (and we have ignored the case where $E_{V_iV_j}=X_{(i)}$ or $E_{ij}=X_{(j)}$ for simplicity). The ideas used in Theorem 8.1

$$\mathcal{C}_n(\{i,n\},\{j,n-1\})-\log n \xrightarrow{d} \Lambda_i'+\Lambda_j'-\Lambda_{ij}, \tag{8.6.17}$$

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

$$\Lambda_i'=\lim_{m\to\infty}\sum_{k=2}^{m}E_k/k-\log m. \tag{8.6.18}$$

This suggests that

$$n\max_{i,j\in[n]}\mathcal{C}_n(V_i,V_j)-3\log n \xrightarrow{d} \max_{1\leq i<j} Y_i+Y_j+\Lambda_i'+\Lambda_j'-\Lambda_{ij}, \tag{8.6.19}$$

which is close to what we wish to prove, except the point process $(Y_i)_{i\geq 1}$ is replaced with the point process $(Y_i+\Lambda_i')_{i\geq 1}$. Interestingly, these have the *same* distribution, which explains Theorem 8.35.

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 $W_n(m)$ denote its weight, and $H_n(m)$ denote the number of edges in it. Here we focus on aspects of $W_n(m)$. In [67], the mean of the weight of the multicast tree is identified as

$$\mathbb{E}[W_n(m)]=\sum_{j=1}^{m}\frac{1}{n-j}\sum_{k=j}^{n-1}\frac{1}{k}. \tag{8.6.20}$$

In particular,

$$\mathbb{E}[W_n(n-1)]=\sum_{k=1}^{n-1}\frac{1}{k^2}. \tag{8.6.21}$$

In [68], the result is extended to a central limit theorem of the form

$$\sqrt{n}\big(W_n(n-1)-\zeta(2)\big)\xrightarrow{d}\sigma Z, \tag{8.6.22}$$

where $\sigma^2=4\zeta(3)\simeq 4.80823$. A crucial ingredient in this analysis is the computation of the variance of $W_n(n-1)$, which is explicitly given by

$$\mathrm{Var}(W_n(n-1))=\frac{4}{n}\sum_{k=1}^{n-1}\frac{1}{k^3}+4\sum_{j=1}^{n-1}\frac{1}{j^3}\sum_{k=1}^{j}\frac{1}{k}-5\sum_{j=1}^{n-1}\frac{1}{j^4}. \tag{8.6.23}$$

The fact that $n\mathrm{Var}(W_n(n-1))\to 4\zeta(3)$ is non-trivial and amounts to carefully compute the arsing 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 minimal-weight path between the two vertices. For $m \geq 3$, however, the Steiner tree can be different from the multicast tree. Let $W_n^{(\mathrm{Ste})}(m)$ denote the weight of the Steiner tree of size m on the complete graph with exponential edge weights. In [23], it is shown that, with $w_n(m) = (m-1)(\log(n/m))/n$

$$\frac{1}{w_n(m)} W_n^{(\mathrm{Ste})}(m) \xrightarrow{\mathbb{P}} 1. \tag{8.6.24}$$

In [67], the following analogous result is shown for the multicast tree, namely, for all $m = O(n^a)$ with $a < 1$,

$$\mathbb{E}[W_n(m)] = \frac{m}{n} \log \frac{n}{m+1} + O(n^{2(a-1)} \log n). \tag{8.6.25}$$

Exercise 8.32 (Comparison Steiner and multicast trees)**.** *Show that* (8.6.24) *and* (8.6.25) *imply that also*

$$\frac{1}{w_n(m)} W_n(m) \xrightarrow{\mathbb{P}} 1. \tag{8.6.26}$$

The complete graph with positive powers of exponential edge weights. We next study 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 $[n] \equiv \{1, \ldots, n\}$ 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 $n \to \infty$. 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 seting 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

$$P_i = \Big(\sum_{j=1}^{i} E_j \Big)^s. \tag{8.6.27}$$

Let PP be the above point process, i.e.,

$$\mathsf{PP} = (P_1, P_2, \ldots). \tag{8.6.28}$$

Now consider the continuous-time branching process (CTBP) 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 $\mathsf{PP}^{(v)} \sim \mathsf{PP}$ as in (8.6.28) independent of every other vertex. Let $(\mathsf{BP}_t)_{t \geq 0}$ denote the CTBP with the above offspring distribution.

Let us first investigate some of the properties of this CTBP. Let $\lambda = \lambda(s)$ be the *Malthusian parameter* in Definition 8.17, which determines the rate of exponential growth of this model. Indeed, by Theorem 8.18, there exists a strictly positive random variable W such that

$$\mathrm{e}^{-\lambda t}|\mathsf{BP}(t)| \xrightarrow{a.s.} W. \tag{8.6.29}$$

The constant λ satisfies the equation

$$\mathbb{E}\Big[\int_0^\infty \mathrm{e}^{-\lambda t}\mathsf{PP}(dt)\Big] = \sum_{i=1}^\infty \mathbb{E}\big(\mathrm{e}^{-\lambda P_i}\big) = 1. \tag{8.6.30}$$

In this case, since $P_i^{1/s}$ has a Gamma-distribution with parameter i, we can explicitly compute that

$$\begin{aligned}\sum_{i=1}^\infty \mathbb{E}\big(\mathrm{e}^{-\lambda P_i}\big) &= \sum_{i=1}^\infty \int_0^\infty \frac{y^{i-1}}{(i-1)!}\mathrm{e}^{-y}\mathrm{e}^{-\lambda y^s}dy \\ &= \int_0^\infty \mathrm{e}^{-\lambda y^s}\Big(\sum_{i=1}^\infty \frac{y^{i-1}}{(i-1)!}\mathrm{e}^{-y}\Big)dy \\ &= \int_0^\infty \mathrm{e}^{-\lambda y^s}dy = \lambda^{-1/s}\Gamma(1+1/s),\end{aligned} \tag{8.6.31}$$

so that

$$\lambda = \lambda(s) = \Gamma(1+1/s)^s. \tag{8.6.32}$$

In the next theorem, we investigate H_n and $\mathcal{C}_n$ in this setting:

Theorem 8.36 (The weight of and edges in the minimal-weight path)**.** *For FPP on the complete graph with i.i.d. edge weights with distribution E^s, where E has an exponential distribution with mean 1 and $s > 0$, as $n \to \infty$,*

$$\Big(\frac{H_n - s\log n}{\sqrt{s^2\log n}}, n^s\mathcal{C}_n - \frac{1}{\lambda}\log n\Big) \xrightarrow{d} (Z, Q), \tag{8.6.33}$$

where Z has a standard normal distribution, and

$$Q \stackrel{d}{=} \frac{1}{\lambda}\left(-\log W^{(1)} - \log W^{(2)} - \Lambda - \log(1/s)\right), \tag{8.6.34}$$

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 (8.6.29).

Theorem 8.36 shows that the behvior 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 Mathusian parameter of the corresponding CTBP 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 FPP on the configuration model in (8.5.8) in Theorem 8.27. These results suggest that if a CTBP approximation of the local neighborhood exploration holds, then the limiting distributions of the hopcount and weight of the minimal-weight path behave rather universally.

The proof of Theorem 8.36 relies on a CTBP 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 8.27.

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 will study the complete graph with i.i.d. edge weights E_{ij}^{-s}, $1 \le i \le j \le n$. 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

$$g_s(x) = \frac{x^{s+1}}{(x-1)^s}, \qquad x \ge 2. \tag{8.6.35}$$

Observe that, for $0 < s \le 1$, the function $g_s(x)$, $x \ge 2$, is increasing, while for $s > 1$, the function is strictly convex with unique minimum at $x = s + 1$. We shall be interested in minimizing this function only on the space $\mathbb{Z}_+$, the set of positive integers. Then, there is a sequence of values $s = s_j$, $j \ge 2$, 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

$$s_j = \frac{\log(1 + j^{-1})}{\log(1 + (j^2 - 1)^{-1})} \in (j-1, j), \qquad j = 2, 3, \ldots. \tag{8.6.36}$$

We will need to deal with these special points separately. When $s \notin \{s_2, \ldots\}$, then there is a unique integer which minimizes the function $g_s(x)$ on $\mathbb{Z}_+$. Let $\mathcal{S} = \{s_2, \ldots\}$.

Theorem 8.37 (Hopcount and weight asymptotics)**.** *Consider FPP 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 (8.6.35). Then, the hopcount $H_n = H_n(s) \xrightarrow{\mathbb{P}} k^*(s)$ as $n \to \infty$, and the optimal weight W_n, properly normalized, converges in distribution to a Gumbel distribution. More precisely, as $n \to \infty$, there exist constants b_k such that

$$\mathbb{P}\left(\frac{k-1}{s g_s(k)} (\log n)^{s+1} \big(W_n - \frac{g_s(k)}{(\log n)^s} \big) + \frac{k-1}{2} \log\log n + b_k > t \right) \to \mathrm{e}^{-\mathrm{e}^t}. \tag{8.6.37}$$

When $s \in \mathcal{S}$, $\mathbb{P}(H_n \in \{\lceil s+1 \rceil, \lfloor s+1 \rfloor\}) \to 1$ and again a linear transformation of W_n converges in distribution to a Gumbel random variable.

Theorem 8.37 states that the hopcount H_n converges to the optimal value of the function $x \mapsto g_s(x)$ defined in (8.6.35), while the rescaled and recentered minimal weight W_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 8.33 (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*

$$(\log n)^s \min_{i\in[n]} E_i^{-s} \xrightarrow{\mathbb{P}} 1. \tag{8.6.38}$$

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 8.34 (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 8.35 (Hopcount for inverse powers of exponentials)**.** *Use Exercises 8.33 and 8.34 to show that $\mathbb{P}(H_n \le 2^{2s+1}) \to 1$.*

8.6.2 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 the configuration model with exponential edge weights. In this section, 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 (4.1.2) that $g_k = \mathbb{P}(D^\star = k+1)$ denotes the asymptotic forward degree distribution in $\mathrm{CM}_n(\boldsymbol{d})$. Let η denote the extinction probability of the branching process with offspring distribution $D^\star - 1$, and recall that $\eta < 1$ precisely when $\mathbb{E}[D^\star - 1] = \nu > 1$. Let $\eta^\star = \mathbb{E}[(D^\star - 1)\eta^{D^\star - 2}] = G'_{D^\star - 1}(\eta)$. The crucial object to describe the weight-flooding and -diameter in $\mathrm{CM}_n(\boldsymbol{d})$ is the following quantity:

$$\Gamma(d) = d\mathbb{1}_{\{d\ge 3\}} + 2(1q_1)\mathbb{1}_{\{d=2\}} + (1\eta^\star)\mathbb{1}_{\{d=1\}}. \tag{8.6.39}$$

In terms of $\Gamma(d)$, the weight-flooding and -diameter in $\mathrm{CM}_n(\boldsymbol{d})$ can be quantified as follows:

Theorem 8.38 (Weight-diameter and flooding on CM)**.** *Let* $\mathrm{CM}_n(\boldsymbol{d})$ *satisfy [I, Condition 7.7(a)-(c)], assume that* $p_{d_{\min}} > 0$ *and that* $\limsup_{n\to\infty} \mathbb{E}[D_n^{2+\varepsilon}] < \infty$ *for some* $\varepsilon > 0$*. Equip* $\mathrm{CM}_n(\boldsymbol{d})$ *with i.i.d. exponential edge weights with mean 1. Then, as* $n \to \infty$*,*

$$\frac{\max_j \mathcal{C}_n(1,j)}{\log n} \xrightarrow{\mathbb{P}} \frac{1}{\nu-1} + \frac{1}{\Gamma(d_{\min})}, \tag{8.6.40}$$

and

$$\frac{\max_{i,j} \mathcal{C}_n(i,j)}{\log n} \xrightarrow{\mathbb{P}} \frac{1}{\nu-1} + \frac{2}{\Gamma(d_{\min})}, \tag{8.6.41}$$

where $\Gamma(d)$ *is defined in* (8.6.39)*.*

Theorem 8.38 can be best understood when $d_{\min} \geq 3$. In this case, $\Gamma(d_{\min}) = d_{\min}$, and we see that the flooding is $1/d_{\min}$ larger than the typical weight as described in Theorem 8.5, while the diameter is $2/d_{\min}$ larger than the typical weight. This is similar to the situation for the complete graph in Theorem 8.1. Indeed, for $i \in [n]$, let $e \in E(\mathrm{CM}_n(\boldsymbol{d}))$ denote the edges in $\mathrm{CM}_n(\boldsymbol{d})$ and define $X_i = \min_{j:\, ij \in E(\mathrm{CM}_n(\boldsymbol{d}))} E_{ij}$, where $(E_{ij})_{ij\in E(\mathrm{CM}_n(\boldsymbol{d}))}$ are the exponential edge weights with mean 1. Then, for every $t > 0$,

$$\mathbb{P}(X_i \geq t) = \mathrm{e}^{-td_i}, \tag{8.6.42}$$

so that, for every $a > 0$,

$$\mathbb{P}(X_i \geq a \log n) = n^{-ad_i}. \tag{8.6.43}$$

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 $p_d > 0$, there exist whp 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 whe $a > 1/d_{\min}$. 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 $\Gamma(d) < d$ when $d = 1, 2$. For $d_{\min} = 2$, such paths can be realized by long thin paths. In order to prove Theorem 8.38 in this case, it is instructive to consider a Markov branching process starting from k individuals having offspring distribution B. Denote the time it takes the CTBP to reach n vertices by T_n^k. Then,

$$-\lim_{n\to\infty} \frac{1}{\log n} \log \mathbb{P}\Big(\big(\frac{1}{\mathbb{E}[B]-1} + x\big) T_n^k < \infty \Big) = x g(B_{\min}, k), \tag{8.6.44}$$

where $B_{\min} = \min\{k \colon \mathbb{P}(B = k) > 0\}$. Note that $\Gamma(d_{\min}) = g((D^\star - 1)_{\min}, d_{\min})$.

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 $f\colon \mathbb{R}_+ \mapsto \mathbb{R}$ by

$$f(\alpha) = \alpha \log\left(\frac{d-2}{d-1}\alpha\right) - \alpha + \frac{1}{d-2}. \tag{8.6.45}$$

Theorem 8.39 (Hopcount-diameter on random regular graphs)**.** *Fix $d_i = d \geq 3$ for every $i \in [n]$. Consider the random-regular graph $\mathrm{CM}_n(\boldsymbol{d})$ equipped with i.i.d. exponential edge weights with mean 1. Then, as $n \to \infty$,*

$$\frac{\max_j H_n(1,j)}{\log n} \xrightarrow{\mathbb{P}} \alpha, \tag{8.6.46}$$

and

$$\frac{\max_{i,j} H_n(i,j)}{\log n} \xrightarrow{\mathbb{P}} \alpha^\star, \tag{8.6.47}$$

where α is the solution to $f(\alpha) = 0$ and $\alpha^\star$ is the solution to $f(\alpha) = 1$.

Configuration model with exponential weights and infinite-variance degrees. We next study the configuration model in the case where the degrees have infinite variance. The main resut is as follows:

Theorem 8.40 (Precise asymptotics for $\tau \in (2,3)$)**.** *Consider the configuration model $\mathrm{CM}_n(\boldsymbol{D})$ with i.i.d. degrees having distribution F satisfying $F(x) = 0$, $x < 2$, and satisfy that there exists $\tau \in (2,3)$ and $0 < c_1 \leq c_2 < \infty$ such that, for all $x \geq 0$,*

$$c_1 x^{-(\tau-1)} \leq 1 - F(x) \leq c_2 x^{-(\tau-1)}. \tag{8.6.48}$$

Then, there exists a random variable Q such that

$$\mathcal{C}_n \xrightarrow{d} Q, \qquad \frac{H_n - \alpha \log n}{\sqrt{\alpha \log n}} \xrightarrow{d} Z, \tag{8.6.49}$$

where $\alpha = \frac{2(\tau-2)}{\tau-1} \in (0,1)$, Z has a standard normal distribution and

$$Q = Q_1 + Q_2, \tag{8.6.50}$$

where Q_1, Q_2 are two independent copies of a random variable which is the explosion time of an infinite-mean continuous-time branching process.

Theorem 8.40 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 minimal-weight path is much larger than shortest paths, which by Theorem 5.3 consist of $2\log\log n/|\log(\tau-2)|$ edges instead of the $\alpha \log n$ edges in Theorem 8.40. 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. Define S_i recursively by $S_0 = D$ and $S_i = S_{i-1} + D^\star - 2$. Then, $(S_i)_{i\geq 0}$ is a random walk

starting in D with step distribution $D^\star - 2 \geq 0$. In particular, when (8.6.48) holds for some $\tau \in (2,3)$, then $D^\star$ is in the domain of attraction of a $(\tau-2)$-stable random variable. Therefore, $\mathbb{E}[(D^\star)^a] = \infty$ for all $a \geq \tau - 2$. In terms of this random variable, we have the description

$$Q_1 = \sum_{j \geq 1} E_j / S_j. \tag{8.6.51}$$

When $D^\star$ is in the domain of attraction of a $(\tau-2)$-stable random variable, $\mathbb{E}[Q_1] < \infty$. Comparing to Theorem 8.24, we see that Q_i is the explosion time of a two-stage Markov CTBP, where the offspring in the first generation is equal to D and the offpring in all other generations is equal to $D^\star - 1$.

Theorem 8.40, combined with the result on graph distances in Theorem 5.3, raises the question what the *universality classes* are for first-passage percolation on $\mathrm{CM}_n(\boldsymbol{D})$ with infinite-variance degrees. Indeed, contrary to the setting of finite-variance degrees in Theorem 8.27, we can show that there are continuous edge-weight distributions for which the hopcount and weight of the minimal-weight path scale like $\log\log n$:

Exercise 8.36 (Infinite variance degrees and another universality class)**.** *Consider the configuration model* $\mathrm{CM}_n(\boldsymbol{d})$ *where the degrees satisfy the conditions in Theorem 5.3. Let the edge weights be equal to* $(1+E_e)_{e\in E(\mathrm{CM}_n(\boldsymbol{d}))}$. *Use Theorem 5.3 to show that, whp,* $W_n \leq 4(1+\varepsilon)\log\log n/|\log(\tau-2)|$. *Use this, in turn, to show that* $H_n \leq 4(1+\varepsilon)\log\log n/|\log(\tau-2)|$

Exercise 8.36 shows that there are at least two universality classes for FPP on the CM with $\tau \in (2,3)$.

8.7 Notes and discussion

Notes on Section 8.2. The presentation in Section 8.2 follows Janson [74]. The restrictions on the weight distribution Y_{ij} in [74] 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 $\mathsf{Exp}(1)$ distributions.

We see from Theorem 8.1(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 in [74] 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 8.41 (Minimal-maximal weight)**.** *As* $n \to \infty$,

$$\frac{\min_{i\in[n]} \max_{j\in[n]} \mathcal{C}_n(i,j)}{\log n/n} \xrightarrow{\mathbb{P}} 2. \tag{8.7.1}$$

For the case of exponentially distributed Y_{ij}, the proof of Theorem 8.1 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. Ditances between vertices in random recursive trees as in Theorem 8.3 have been studied in [86], and the diameter in [97] as $n \to \infty$. See also the survey [104].

Notes on Section 8.3. Theorem 8.5 is proved in [16] under slightly more restrictive conditions on the degrees as the ones stated. Indeed, in [16], 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 [18] under slightly weaker conditions. We follow the proof in [16].

Proposition 8.7 is proved first by Bühler [30]. We present the probabilistic proof from [16], since there is some confusion between the definition s_i given here, and the definition of s_i given in [30, below Equation (3.1)]. More precisely, in [30], s_i is defined as $s_i = d_1 + \ldots + d_i - i$, which is our $s_i - 1$.

Notes on Section 8.4. A comprehensive study of continuous-time branching processes can be found in [62], [72] and [12]. We follow [62]. In particular, [62, Chapter V] deals with Markov branching processes, in which the evolution of $(|\mathsf{BP}(t)|)_{t\geq 0}$ is Markovian. Examples are Bellman-Harris processes. [62, Chapter VI] deals with general age-dependent CTBPs. The most general results on CTBPs can be found in [73]. The convergence in (8.4.31) is [62, Theorem 17.1]. Jagers and Nerman [73] have proved some of the strongest results for general continuous-time branching processes. In particular, Theorem 8.18 is [73, Theorem 5.3]. Theorems 8.25 and 8.26 are due to Grey [58].

Notes on Section 8.5. Theorem 8.27 is proved in [18].

Notes on Section 8.6. The first result in Theorem 8.34 is in [43] and [96], the second is the main result in [2]. In [2], it is proved further that $\max_{i,j} H_n(i,j) \leq \alpha^\star \log n + \omega(n)$ whp for any $\omega(n) \to \infty$, while $\max_{i,j} H_n(i,j) \geq \alpha^\star \log n + L \log\log n$ 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 8.35 is proved in [15] and solves [74, Problem 1]. Theorem 8.36 is proved in [21], where, erroneously, $-\Lambda$ in (8.6.34) was replaced by Λ.

It is yet unclear what the universality classes of FPP on the complete graph are. In [48], settings are investigated where the minimal-weight paths are short compared to $\log n$, in [47] settings where minimal-weight paths are long compared to $\log n$. The behavior of FPP 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 $s_n \to \infty$, sometimes also called the *strong disorder phase* as in [29, 64, 109], 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 $s_n \to 0$ is handled in [48], the case where $s_n \to \infty$ in [47]. 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 FPP on the complete graph is closely related to the properties of the minimal spanning tree, as investigated in [1, 3, 57]. There, the gaph distance between vertices grows like $n^{1/3}$, which suggests that FPP and the minimal spanning tree can be in the *same* universality class.

Theorem 8.38 is proved in [8], see also [7]. The result in (8.6.44), which is crucial in the understanding of Theorem 8.38, is proved in [9]. The special case of d-random regular graphs with $d \geq 3$ is also proved in [44], where it was used to investigate the diameter of near-critical supercritical Erdős-Rényi random graphs. Theorem 8.39 is proved in [10].

Theorem 8.40 is proved in [17]. In [16], FPP with exponential edge weights on $\mathrm{CM}_n(\boldsymbol{D})$ with $\tau \in (1,2)$ is studied, and it is shown that both W_n and H_n converge in distribution. Interestingly, the distributional limits of W_n and H_n depend on whether we work with the original CM (where we keep all multiple edges) or the erased CM (where all multiple edges are merged into a single edge).

A.1 Solutions to the exercises of Chapter 2.

Solution to Exercise 2.2. Let $\kappa(x,y) = [1-F]^{-1}(x)[1-F]^{-1}(y)/\mathbb{E}[W]$, which consists of the product of two functions that are continuous a.e. on $\mathcal{S}$, and therefore satisfies 2.3(a). To see 2.3 is satisfied, consider the double integral

$$\begin{aligned}\iint_{\mathcal{S}^2} \kappa(x,y)d\mu(x)d\mu(y) &= \frac{1}{\mathbb{E}[W]}\int_{\mathcal{S}}[1-F]^{-1}(y)\int_{\mathcal{S}}[1-F]^{-1}(x)d\mu(x)d\mu(y)\\ &= \frac{1}{\mathbb{E}[W]}\int_{\mathcal{S}}[1-F]^{-1}(y)\mathbb{E}[W]d\mu(y) = \mathbb{E}[W], \qquad \text{(A.1.1)}\end{aligned}$$

which is finite if and only if $\mathbb{E}[W]$ is finite.
The expected number of edges in the graph is half of the sum of expected degrees over the vertices, due to the handshaking lemma, which is $\frac{n\mu}{2}$ in this case. Therefore,

$$\frac{1}{n}\mathbb{E}[E(\mathrm{IRG}_n(\boldsymbol{p}(\kappa)))] \to \frac{1}{2}\iint_{\mathcal{S}^2}\kappa(x,y)d\mu(x)d\mu(y) = \frac{1}{2}\mu, \qquad \text{(A.1.2)}$$

which proves 2.3(c), and thus the kernel is graphical exactly when $\mathbb{E}[W]$ is finite.

For irreducibility, note that $[1-F]^{-1}(x)[1-F]^{-1}(y) = 0$ if and only if $x = 1$ or $y = 1$, which gives rise to two sets that both have measure 0 in $\mathcal{S}^2$. □

Solution to Exercise 2.3. Denote by $\tilde{\mathbf{p}}$ and $\mathbf{p}$ the edge occupation probabilities for $\mathrm{CL}_n(\tilde{\boldsymbol{w}})$ and $\mathrm{CL}_n(\boldsymbol{w})$, respectively, i.e.,

$$\tilde{p}_{ij} = \min\{1, \tilde{w}_i\tilde{w}_j/l_n\} = \min\{1, w_iw_j/(n\mu)\} \qquad p_{ij} = \min\{1, w_iw_j/l_n\}. \qquad \text{(A.1.3)}$$

We assume without loss of generality that $l_n \geq \mu n$, when $l_n \leq \mu n$ the argument can easily be adapted. Then, $p_{ij} \leq \tilde{p}_{ij}$, and

$$\tilde{p}_{ij} - p_{ij} \leq (\frac{\mu n}{l_n} - 1)\tilde{p}_{ij}. \qquad \text{(A.1.4)}$$

We note that

$$\sum_{1\leq i<j\leq n}(\tilde{p}_{ij}-p_{ij})^2/\tilde{p}_{ij} \leq (\frac{\mu n}{l_n}-1)^2\sum_{1\leq i<j\leq n}\tilde{p}_{ij} \leq (\frac{\mu n}{l_n}-1)^2 l_n = o(1), \qquad \text{(A.1.5)}$$

whenever $(\frac{\mu n}{l_n}-1)^2 = o(n)$. Theorem 6.18 then proves the claim. □

Solution to Exercise 2.4. Assume (2.2.13) holds. Note that this means that for each i and j, $\kappa(x_i,x_j)/n \to 0$ as well. Asymptotically, when considering $\min\{\kappa(x,y)/n, 1\}$, we get that $\kappa(x,y)/n < 1$, and we will write $\kappa_{ij} = \kappa(x_i,x_j)$. We now apply Theorem 6.18 with p_{ij} as in (2.2.7) and $q_{ij} = p_{ij}^{(\mathrm{NR})}(\kappa)$. To approximate $q_{ij} = 1-\mathrm{e}^{-\kappa_{ij}/n}$ we use the Taylor approximation $\kappa_{ij}/n - \kappa_{ij}^2/2n^2 + O(\kappa_{ij}^3/n^3)$.

$$\begin{aligned}\sum_{i<j}\frac{(p_{ij}-q_{ij})^2}{p_{ij}} &= \sum_{i<j}\frac{(\kappa_{ij}/n - \kappa_{ij}/n + \kappa_{ij}^2/2n^2 + O(\kappa_{ij}^3/n^3))^2}{\kappa_{ij}/n}\\ &= \sum_{i<j}\frac{\kappa_{ij}^4/4n^4 + O(\kappa_{ij}^3/n^3)}{\kappa_{ij}/n} \to 0, \qquad \text{(A.1.6)}\end{aligned}$$

as $n \to \infty$, by (2.2.13). Therefore, the random graphs are asymptotically equivalent by Theorem 6.18.
Now we apply Theorem 6.18 with p_{ij} as in (2.2.7) and $q_{ij} = p_{ij}^{(\mathrm{GRG})}(\kappa)$.

$$\begin{aligned}\sum_{i<j} \frac{(p_{ij}-q_{ij})^2}{p_{ij}} &= \sum_{i<j} \frac{\left(\kappa_{ij}/n - \frac{\kappa_{ij}}{n+\kappa_{ij}}\right)^2}{\kappa_{ij}/n} \\ &= \sum_{i<j} \frac{\left(\frac{\kappa_{ij}^2}{n^2+n\kappa_{ij}}\right)^2}{\kappa_{ij}/n} \\ &= \sum_{i<j} \frac{\kappa_{ij}^3}{n(n+\kappa_{ij})^2} \to 0, \end{aligned} \tag{A.1.7}$$

as $n \to \infty$, by (2.2.13). Therefore, the random graphs are asymptotically equivalent by Theorem 6.18, as desired. □

Solution to Exercise 2.5. Definition 2.2(i) holds because $\mathcal{S} = \{1, 2\}$ is finite and therefore a separable metric space, and because μ given by $\mu(\{1\}) = \mu(\{2\}) = 1/2$ is a Borel probability measure. Since $\boldsymbol{x}_n$ consists of $n/2$ vertices of each type, $\nu_n(\{1\}) = \nu_n(\{2\}) = (n/2)/n = 1/2 = \mu(\{1\}) = \mu(\{2\})$, and Definition 2.2(ii) holds. Now, κ can be written as an $|\mathcal{S}| \times |\mathcal{S}|$ matrix, which makes it a linear and therefore continuous function. Since continuity implies Borel measurability, Definition 2.2(iii) holds as well.

We have seen that κ is continuous, so κ satisfies Definition 2.3(i)(a). As for Definition 2.3(i)(b) and (c),

$$\iint_{\mathcal{S}^2} \kappa(x,y) d\mu(x) d\mu(y) = \lambda/2 < \infty \tag{A.1.8}$$

$$\frac{1}{n}\mathbb{E}[E(\mathrm{IRG}_n(\boldsymbol{p}(\kappa)))] = (\frac{n}{2})^2 \frac{\lambda}{n} = \frac{1}{2}\iint_{\mathcal{S}^2} \kappa(x,y) d\mu(x) d\mu(y). \tag{A.1.9}$$

There are only two subsets $A \subseteq S$ such that $0 < \mu(A) < 1$: $\{1\}$ and $\{2\}$, and $\kappa = \lambda \neq 0$ on $\{1\} \times \{2\}$. Therefore κ is irreducible.

Solution to Exercise 2.6. Consider the finite types case where $r = 2$, $s_1 = 1$, $s_2 = 2$, and $\mu(\{i\}) = 1/2$ for $i = 1, 2$. Now take n even and take $n_1 = n/2$ vertices of type 1 and $n_2 = n/2$ vertices of type 2. Then $n_i/n = 1/2$ for both $i = 1, 2$ and for all n. Let

$$\kappa = \begin{pmatrix} 0 & \lambda \\ \lambda & 0 \end{pmatrix}, \tag{A.1.10}$$

be the 2×2 matrix, such that vertices of the same type are not connected by an edge with high probability, and vertices of differing types are connected by an edge with probability λ/n, independently of the other edges. This is exactly the homogeneous bipartite random graph, as desired. □

Solution to Exercise 2.7. The proof goes by contraposition. Assume κ is reducible. This means there is some nonempty proper subset $A \subset \mathcal{S}$ such that $\kappa = 0$ a.e. on $A \times (\mathcal{S} \setminus A)$. Since $\mathcal{S}$ is finite, κ is an $r \times r$ matrix such that for each $x \in A$ and $y \in (\mathcal{S} \setminus A)$, the matrix element κ_{xy} is zero. Therefore, the submatrix of κ restricted to A and the submatrix of κ restricted to $(\mathcal{S} \setminus A)$ contain amongst them all nonzero elements of κ, the other elements must be zero. Under the right permutation of rows and columns, κ will be of the form

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix}, \tag{A.1.11}$$

where A is the $|A| \times |A|$ submatrix of κ restricted to A, D is the $|(\mathcal{S} \setminus A)| \times |(\mathcal{S} \setminus A)|$ submatrix of κ restricted to $(\mathcal{S} \setminus A)$, and B and C are zero matrices. Repeatedly multiplying this matrix with itself will not change the form of the matrix, because the blocks B and C will always remain zero. Therefore, there is no m such that κ^m contains no zeros, as desired. □

A.2 Solutions to the exercises of Chapter 3.

Solution to Exercise 3.18. Assume that κ is irreducible, $\sup_{x,y} \kappa(x,y) < \infty$ and $\nu = \|\mathbf{T}_\kappa\| > 1$. Now, by Theorem 2.7 and the irreducibility of κ, we have $\mathbb{P}(H_n < \infty) = \zeta_\kappa^2 + o(1)$. Furthermore, since $\sup_{x,y} \kappa(x,y) < \infty$, Theorem 2.8(i) gives $\mathbb{P}(H_n \leq (1-\varepsilon)\log_\nu n) = o(1)$. Finally, since κ is irreducible and $\nu = \|\mathbf{T}_\kappa\| > 1$, Theorem 2.8(ii) gives that $\mathbb{P}(H_n \leq (1+\varepsilon)\log_\nu n) = \zeta_\kappa^2 + o(1)$. We now rewrite the probability

$$\begin{aligned}
\mathbb{P}(1-\varepsilon \leq \frac{H_n}{\log_\nu n} \leq 1+\varepsilon | H_n \leq \infty) &= \mathbb{P}(\frac{H_n}{\log_\nu n} \leq 1+\varepsilon | H_n \leq \infty) - \mathbb{P}(\frac{H_n}{\log_\nu n} < 1-\varepsilon | H_n \leq \infty) \\
&= \frac{\mathbb{P}(\frac{H_n}{\log_\nu n} \leq 1+\varepsilon)}{\mathbb{P}(H_n < \infty)} - \frac{\mathbb{P}(\frac{H_n}{\log_\nu n} < 1-\varepsilon)}{\mathbb{P}(H_n < \infty)} \\
&= \frac{\zeta_\kappa^2 + o(1)}{\zeta_\kappa^2 + o(1)} - \frac{o(1)}{\zeta_\kappa^2 + o(1)} \to 1,
\end{aligned} \tag{A.2.1}$$

as $n \to \infty$. Because $\log_a b = \log b / \log a$, we now have that conditionally on $H_n < \infty$, $\frac{H_n}{\log n / \log \nu} \xrightarrow{\mathbb{P}} 1$, or equivalently $H_n / \log n \xrightarrow{\mathbb{P}} 1/\log \nu$, as desired. □

Solution to Exercise 3.19. Under the assumptions, and using Theorem 2.7 and the irreducibility of κ, we once again get that $\mathbb{P}(H_n < \infty) = \zeta_\kappa^2 + o(1)$. Furthermore, we have that there exists a function $f(n) = o(n)$ such that $\mathbb{P}(H_n \leq f(n)) = \zeta_\kappa^2 + o(1)$. When rewriting this, and conditioning on $H_n < \infty$, we get

$$\begin{aligned}
\mathbb{P}(H_n / \log n \leq f(n)/\log n | H_n < \infty) &= \mathbb{P}(H_n \leq f(n) | H_n < \infty) \\
&= \frac{\zeta_\kappa^2 + o(1)}{\zeta_\kappa^2 + o(1)} \to 1,
\end{aligned} \tag{A.2.2}$$

as $n \to \infty$. Since $f(n) = o(n)$, we get that $H_n / \log n \xrightarrow{\mathbb{P}} 0$, as desired.

Solutions to the exercises of Chapter 4.

Solution to Exercise ??. We note that $p_k \leq p_k(n)$ when $k \leq n^a$, while $p_k(n) = 0$ for $k > n^a$. Therefore, we obtain that

$$d_{\mathrm{TV}}(p, p(n)) = \frac{1}{2}\sum_{k=1}^{n^a} p_k\Big(\frac{1}{F(n^a)}-1\Big)+\frac{1}{2}\sum_{k>n^a} p_k = \frac{1}{2}F(n^a)\Big(\frac{1}{F(n^a)}-1\Big)+\frac{1}{2}[1-F(n^a)] = 1-F(n^a).$$

Solution to Exercise ??. We use Exercise **??** and a coupling argument. We take $\{D_i\}_{i=1}^n$ and $\{D_i^{(n)}\}_{i=1}^n$ to be two i.i.d. sequences of random variables, where D_i has probability mass function $\{p_k\}_{k=1}^\infty$, while $D_i^{(n)}$ has probability mass function $\{p_k(n)\}_{k=1}^\infty$.

By the coupling in Section 2.2, we can couple each of the $D_i, D_i^{(n)}$ with $(\hat{D}_i, \hat{D}_i^{(n)})$ such that

$$\mathbb{P}(\hat{D}_i \neq \hat{D}_i^{(n)}) = d_{\mathrm{TV}}(p, p(n)).$$

By Exercise **??**, and (2.6.8), we obtain that

$$\mathbb{P}(\hat{D}_i \neq \hat{D}_i^{(n)}) \leq n^{a(1-\tau)}.$$

Therefore, by Boole's inequality,

$$\begin{aligned}\mathbb{P}(\{\hat{D}_i^{(n)}\}_{i=1}^n \neq \{\hat{D}_i\}_{i=1}^n) &\leq \sum_{i=1}^n \mathbb{P}(\hat{D}_i \neq \hat{D}_i^{(n)}) \\ &= n\mathbb{P}(\hat{D}_1 \neq \hat{D}_1^{(n)}) \leq n^{1+a(1-\tau)}.\end{aligned}$$

Solutions to the exercises of Chapter 6.

Solution to Exercise 7.6. We follow the proof of (7.3.16), and adapt it to $\mathrm{PA}_t^{(b)}(1,\delta)$. We start by computing $\mathbb{E}[(D_i(t)+\delta)]$ and $\mathbb{E}[(D_i(t)+1+\delta)(D_i(t)+\delta)]$ for $\mathrm{PA}_t^{(b)}(1,\delta)$, as these expectations are needed in the proof. We rely on the recursion

$$\mathbb{E}[(D_i(t+1)+\delta) \mid \mathrm{PA}_t^{(b)}(1,\delta)] = (1+\frac{1}{(2+\delta)t})(D_i(t)+\delta), \tag{A.2.3}$$

which implies that

$$\begin{aligned}\mathbb{E}[(D_i(t)+\delta)] &= (D_i(i)+\delta)\prod_{s=i}^{t-1}(1+\frac{1}{(2+\delta)s}) = (1+\delta)\prod_{s=i}^{t-1}\frac{s+1/(2+\delta)}{s} \\ &= (1+\delta)\frac{\Gamma(i)\Gamma(t+1/(2+\delta)}{\Gamma(i+1/(2+\delta)\Gamma(t)}.\end{aligned} \tag{A.2.4}$$

Similarly,

$$\begin{aligned}
&\mathbb{E}[(D_i(t+1)+1+\delta)(D_i(t+1)+\delta) \mid \mathrm{PA}_t^{(b)}(1,\delta)] \qquad \text{(A.2.5)}\\
&\quad = (D_i(t)+1+\delta)(D_i(t)+\delta)\big(1-\frac{D_i(t)+\delta}{(2+\delta)t}\big)+(D_i(t)+2+\delta)(D_i(t)+1+\delta)\big(\frac{D_i(t)+\delta}{(2+\delta)t}\big)\\
&\quad = (1+\frac{2}{(2+\delta)t})(D_i(t)+1+\delta)(D_i(t)+\delta),
\end{aligned}$$

which implies that

$$\begin{aligned}
\mathbb{E}[(D_i(t)+1+\delta)(D_i(t)+\delta)] &= (D_i(i)+\delta)(D_i(i)+1+\delta)\prod_{s=i}^{t-1}(1+\frac{2}{(2+\delta)s}) \qquad \text{(A.2.6)}\\
&= (2+\delta)(1+\delta)\prod_{s=i}^{t-1}\frac{s+2/(2+\delta)}{s}\\
&= (2+\delta)(1+\delta)\frac{\Gamma(i)\Gamma(t+2/(2+\delta)}{\Gamma(i+2/(2+\delta)\Gamma(t)}.
\end{aligned}$$

Take $t_2 > t_1$. We proceed with the proof of (7.3.27) by computing

$$\begin{aligned}
&\mathbb{P}\Big(g(t_1)=s, g(t_2)=s\Big) \qquad \text{(A.2.7)}\\
&\quad = \mathbb{E}\Big[\mathbb{P}\big(g(t_1)=s, g(t_2)=s\big|\mathrm{PA}_{t_2-1}^{(b)}(1,\delta)\big)\Big]\\
&\quad = \mathbb{E}\left[\mathbb{1}_{\{g(t_1)=s\}}\left(\frac{D_s(t_2-1)+\delta}{(t_2-1)(2+\delta)}\right)\right]\\
&\quad = \frac{1}{(t_2-1)(2+\delta)}\frac{\Gamma(t_1)\Gamma(t_2-1+1/(2+\delta))}{\Gamma(t_1+1/(2+\delta))\Gamma(t_2-1)}\mathbb{E}\left[\mathbb{1}_{\{g(t_1)=s\}}\left(D_s(t_1)+\delta\right)\right]\\
&\quad = \frac{\Gamma(t_1)\Gamma(t_2-(1+\delta)/(2+\delta))}{(2+\delta)\Gamma(t_1+1/(2+\delta))\Gamma(t_2)}\mathbb{E}\left[\mathbb{1}_{\{g(t_1)=s\}}\left(D_s(t_1)+\delta\right)\right],
\end{aligned}$$

where we use the iteration, for $t_1 < u \le t_2 - 1$,

$$\begin{aligned}
&\mathbb{E}\left[\mathbb{1}_{\{g(t_1)=s\}}\left(D_s(u)+\delta\right)\right] \qquad \text{(A.2.8)}\\
&\qquad = \Big(1+\frac{1}{(2+\delta)(u-1)+1+\delta}\Big)\mathbb{E}\left[\mathbb{1}_{\{g(t_1)=s\}}\left(D_s(u-1)+\delta\right)\right].
\end{aligned}$$

We are lead to compute $\mathbb{E}\left[\mathbb{1}_{\{g(t_1)=s\}}\left(D_s(t_1)+\delta\right)\right]$. We again use recursion to obtain

$$\begin{aligned}
&\mathbb{E}\left[\mathbb{1}_{\{g(t_1)=s\}}\left(D_s(t_1)+\delta\right)\big|\mathrm{PA}_{t_1-1}^{(b)}(1,\delta)\right] \qquad \text{(A.2.9)}\\
&\quad = \mathbb{E}\left[\mathbb{1}_{\{g(t_1)=s\}}\left(D_s(t_1)-D_s(t_1-1)\right)\big|\mathrm{PA}_{t_1-1}^{(b)}(1,\delta)\right]\\
&\qquad + \mathbb{E}\left[\mathbb{1}_{\{g(t_1)=s\}}\left(D_s(t_1-1)+\delta\right)\big|\mathrm{PA}_{t_1-1}^{(b)}(1,\delta)\right]\\
&\quad = \frac{(D_s(t_1-1)+\delta)(D_s(t_1-1)+1+\delta)}{(t_1-1)(2+\delta)}.
\end{aligned}$$

By (A.2.6),

$$\mathbb{E}[(D_s(t)+\delta)(D_s(t)+1+\delta)] = (2+\delta)(1+\delta)\frac{\Gamma(s)\Gamma(t+2/(2+\delta)}{\Gamma(s+2/(2+\delta)\Gamma(t)}. \tag{A.2.10}$$

Consequently,

$$\begin{aligned}
&\mathbb{E}\left[\mathbb{1}_{\{g(t_1)=s\}}\left(D_s(t_1)+\delta\right)\right] \\
&\quad = \mathbb{E}[\frac{(D_s(t_1-1)+\delta)(D_s(t_1-1)+1+\delta)}{(t_1-1)(2+\delta)}] \\
&\quad = (2+\delta)(1+\delta)\frac{\Gamma(s)\Gamma(t_1-1+2/(2+\delta)}{(t_1-1)(2+\delta)\Gamma(s+2/(2+\delta)\Gamma(t_1-1)} \\
&\quad = (1+\delta)\frac{\Gamma(t_1-\delta/(2+\delta))\Gamma(s)}{\Gamma(t_1)\Gamma(s+2/(2+\delta))}.
\end{aligned} \tag{A.2.11}$$

Combining (A.2.7) and (A.2.11), we arrive at

$$\begin{aligned}
&\mathbb{P}\big(g(t_1)=s, g(t_2)=s\big) \\
&\quad = \frac{\Gamma(t_1)\Gamma(t_2-(1+\delta)/(2+\delta))}{(2+\delta)\Gamma(t_1+1/(2+\delta))\Gamma(t_2)} \times (1+\delta)\frac{\Gamma(t_1-\delta/(2+\delta))\Gamma(s)}{\Gamma(t_1)\Gamma(s+2/(2+\delta))} \\
&\quad = (1+\delta)\frac{\Gamma(t_1-\delta/(2+\delta))\Gamma(t_2-(1+\delta)/(2+\delta))\Gamma(s)}{\Gamma(t_1+1/(2+\delta))\Gamma(t_2)\Gamma(s+2/(2+\delta))},
\end{aligned} \tag{A.2.12}$$

as required.

References

[1] L. Addario-Berry, N. Broutin, and C. Goldschmidt. The scaling limit of the minimal spanning tree on the complete graph. Preprint (2012).

[2] 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).

[3] 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).

[4] R. Albert and A.-L. Barabási. Statistical mechanics of complex networks. *Rev. Modern Phys.*, **74**(1):47–97, (2002).

[5] D. Aldous. Exchangeability and related topics. In *École d'été de probabilités de Saint-Flour, XIII—1983*, volume **1117** of *Lecture Notes in Math.*, pages 1–198. Springer, Berlin, (1985).

[6] N. Alon and J. Spencer. *The probabilistic method*. Wiley-Interscience Series in Discrete Mathematics and Optimization. John Wiley & Sons, New York, second edition, (2000).

[7] H. Amini, M. Draief, and M. Lelarge. Flooding in weighted sparse random graphs. *SIAM Journal on Discrete Mathematics*, **27**(1):1–26, (2013).

[8] H. Amini and M. Lelarge. The diameter of weighted random graphs. Available at `http://arxiv.org/pdf/1112.6330.pdf`, Preprint (2011).

[9] H. Amini and M. Lelarge. Large deviations for split times of branching processes. Preprint (2013).

[10] H. Amini and Y. Peres. Shortest-weight paths in random regular graphs. Available at `http://arxiv.org/pdf/1210.2657.pdf`, Preprint (2012).

[11] R. Arratia, AD Barbour, and S. Tavaré. *Logarithmic combinatorial structures: a probabilistic approach*. European Mathematical Society, (2003).

[12] K. Athreya and P. Ney. *Branching processes*. Springer-Verlag, New York, (1972). Die Grundlehren der mathematischen Wissenschaften, Band 196.

[13] F. Ball, D. Sirl, and P. Trapman. Threshold behaviour and final outcome of an epidemic on a random network with household structure. *Adv. in Appl. Probab.*, **41**(3):765–796, (2009).

[14] F. Ball, D. Sirl, and P. Trapman. Analysis of a stochastic SIR epidemic on a random network incorporating household structure. *Math. Biosci.*, **224**(2):53–73, (2010).

[15] S. Bhamidi and R. van der Hofstad. Diameter of the stochastic mean-field model of distance. Preprint (2013).

[16] S. Bhamidi, R. van der 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).

[17] S. Bhamidi, R. van der Hofstad, and G. Hooghiemstra. First passage percolation on random graphs with finite mean degrees. *Ann. Appl. Probab.*, **20**(5):1907–1965, (2010).

[18] S. Bhamidi, R. van der Hofstad, and G Hooghiemstra. First passage percolation on the Erdős-Rényi random graph. *Combin. Probab. Comput.*, **20**(5):683–707, (2011).

[19] S. Bhamidi, R. van der Hofstad, and J. van Leeuwaarden. Scaling limits for critical inhomogeneous random graphs with finite third moments. *Electronic Journal of Probability*, **15**:1682–1702, (2010).

[20] S. Bhamidi, R. van der Hofstad, and J. van Leeuwaarden. Novel scaling limits for critical inhomogeneous random graphs. Preprint (2009). To appear in *Annals of Probability*.

[21] 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).

[22] B. Bollobás. *Random graphs*, volume **73** of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, second edition, (2001).

[23] 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).

[24] B. Bollobás, S. Janson, and O. Riordan. The phase transition in inhomogeneous random graphs. *Random Structures Algorithms*, **31**(1):3–122, (2007).

[25] B. Bollobás and O. Riordan. Shortest paths and load scaling in scale-free trees. *Phys. Rev. E.*, **69**:036114, (2004).

[26] C. Borgs, J. Chayes, R. van der 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).

[27] C. Borgs, J. Chayes, R. van der 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).

[28] C. Borgs, J. Chayes, R. van der 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).

[29] L.A. Braunstein, S.V. Buldyrev, R. Cohen, S. Havlin, and H.E. Stanley. Optimal paths in disordered complex networks. *Physical review letters*, **91**(16):168701, (2003).

[30] W. Bühler. Generations and degree of relationship in supercritical Markov branching processes. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, **18**:141–152, (1971).

[31] S. Chatterjee and R. Durrett. Contact processes on random graphs with power law degree distributions have critical value 0. *Ann. Probab.*, **37**(6):2332–2356, (2009).

[32] F. Chung and L. Lu. The diameter of sparse random graphs. *Adv. in Appl. Math.*, **26**(4):257–279, (2001).

[33] 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).

[34] F. Chung and L. Lu. Connected components in random graphs with given expected degree sequences. *Ann. Comb.*, **6**(2):125–145, (2002).

[35] F. Chung and L. Lu. The average distance in a random graph with given expected degrees. *Internet Math.*, **1**(1):91–113, (2003).

[36] F. Chung and L. Lu. Coupling online and offline analyses for random power law graphs. *Internet Math.*, **1**(4):409–461, (2004).

[37] 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).

[38] 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).

[39] C. Cooper and A. Frieze. The size of the largest strongly connected component of a random digraph with a given degree sequence. *Combin. Probab. Comput.*, **13**(3):319–337, (2004).

[40] D. A. Darling. The Galton-Watson process with infinite mean. *J. Appl. Probability*, **7**:455–456, (1970).

[41] P. L. Davies. The simple branching process: a note on convergence when the mean is infinite. *J. Appl. Probab.*, **15**(3):466–480, (1978).

[42] S. Dereich, C. Mönch, and P. Mörters. Typical distances in ultrasmall random networks. *Adv. in Appl. Probab.*, **44**(2):583–601, (2012).

[43] L. Devroye. Branching processes in the analysis of the heights of trees. *Acta Inform.*, **24**(3):277–298, (1987).

[44] 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).

[45] S. Dommers, R. van der Hofstad, and G. Hooghiemstra. Diameters in preferential attachment graphs. *Journ. Stat. Phys.*, **139**:72–107, (2010).

[46] R. Durrett. *Random graph dynamics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, (2007).

[47] M. Eckhoff, J. Goodman, R. van der Hofstad, and F.R. Nardi. Long paths in first passage percolation on the complete graph. Preprint (2012).

[48] M. Eckhoff, J. Goodman, R. van der Hofstad, and F.R. Nardi. Short paths in first passage percolation on the complete graph. Preprint (2012).

[49] P. Erdős. Some remarks on the theory of graphs. *Bull. Amer. Math. Soc.*, **53**:292–294, (1947).

[50] P. Erdős and A. Rényi. On random graphs. I. *Publ. Math. Debrecen*, **6**:290–297, (1959).

[51] P. Erdős and A. Rényi. On the evolution of random graphs. *Magyar Tud. Akad. Mat. Kutató Int. Közl.*, **5**:17–61, (1960).

[52] P. Erdős and A. Rényi. On the evolution of random graphs. *Bull. Inst. Internat. Statist.*, **38**:343–347, (1961).

[53] P. Erdős and A. Rényi. On the strength of connectedness of a random graph. *Acta Math. Acad. Sci. Hungar.*, **12**:261–267, (1961).

[54] H. van den Esker, R. van der Hofstad, and G. Hooghiemstra. Universality for the distance in finite variance random graphs. *J. Stat. Phys.*, **133**(1):169–202, (2008).

[55] H. van den Esker, R. van der Hofstad, G. Hooghiemstra, and D. Znamenski. Distances in random graphs with infinite mean degrees. *Extremes*, **8**:111–140, (2006).

[56] D. Fernholz and V. Ramachandran. The diameter of sparse random graphs. *Random Structures Algorithms*, **31**(4):482–516, (2007).

[57] A. M. Frieze. On the value of a random minimum spanning tree problem. *Discrete Appl. Math.*, **10**(1):47–56, (1985).

[58] D. R. Grey. Explosiveness of age-dependent branching processes. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, **28**:129–137, (1973/74).

[59] G. Grimmett. *Percolation*. Springer, Berlin, 2nd edition, (1999).

[60] P. Hall. Order of magnitude of moments of sums of random variables. *J. London Math Soc.*, **24**(2):562–568, (1981).

[61] G. H. Hardy, J. E. Littlewood, and G. Pólya. *Inequalities*. Cambridge Mathematical Library. Cambridge University Press, Cambridge, (1988). Reprint of the 1952 edition.

[62] T. Harris. *The theory of branching processes*. Die Grundlehren der Mathematischen Wissenschaften, Bd. 119. Springer-Verlag, Berlin, (1963).

[63] H. Hatami and M. Molloy. The scaling window for a random graph with a given degree sequence. In *Proceedings of the Twenty-First Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1403–1411. Society for Industrial and Applied Mathematics, (2010).

[64] S. Havlin, L.A. Braunstein, S.V. Buldyrev, R. Cohen, T. Kalisky, S. Sreenivasan, and H. Eugene Stanley. Optimal path in random networks with disorder: A mini review. *Physica A: Statistical Mechanics and its Applications*, **346**(1-2):82–92, (2005).

[65] R. van der Hofstad. Random graphs and complex networks. Vol. I. In preparation, see `http://www.win.tue.nl/~rhofstad/NotesRGCN.pdf`, (2014).

[66] R. van der Hofstad, G. Hooghiemstra, and P. Van Mieghem. Distances in random graphs with finite variance degrees. *Random Structures Algorithms*, **27**(1):76–123, (2005).

[67] R. van der 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).

[68] R. van der Hofstad, G. Hooghiemstra, and P. Van Mieghem. The weight of the shortest path tree. *Random Structures Algorithms*, **30**(3):359–379, (2007).

[69] R. van der 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).

[70] R. van der Hofstad, G. Hooghiemstra, and D. Znamenski. Random graphs with arbitrary i.i.d. degrees. Preprint Mittag-Leffer Institute, (2004).

[71] R. van der Hofstad, S. Janson, and M. Luczak. The near-critical behavior for the configuration model with finite-variance degrees. In preparation.

[72] 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.

[73] P. Jagers and O. Nerman. The growth and composition of branching populations. *Adv. in Appl. Probab.*, **16**(2):221–259, (1984).

[74] 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).

[75] S. Janson. The largest component in a subcritical random graph with a power law degree distribution. *Ann. Appl. Probab.*, **18**(4):1651–1668, (2008).

[76] S. Janson. Standard representation of multivariate functions on a general probability space. *Electron. Commun. Probab.*, **14**:343–346, (2009).

[77] S. Janson. Susceptibility of random graphs with given vertex degrees. *J. Comb.*, **1**(3-4):357–387, (2010).

[78] S. Janson and M. Luczak. A new approach to the giant component problem. *Random Structures Algorithms*, **34**(2):197–216, (2009).

[79] S. Janson, T. Łuczak, and A. Rucinski. *Random graphs*. Wiley-Interscience Series in Discrete Mathematics and Optimization. Wiley-Interscience, New York, (2000).

[80] A. Joseph. The component sizes of a critical random graph with pre-described degree sequence. Available at `http://arxiv.org/pdf/1012.2352.pdf`, Preprint (2010).

[81] O. Kallenberg. *Random measures*. Akademie-Verlag, Berlin, fourth edition, (1986).

[82] O. Kallenberg. *Foundations of modern probability*. Probability and its Applications (New York). Springer-Verlag, New York, second edition, (2002).

[83] J. F. C. Kingman. The first birth problem for an age-dependent branching process. *Ann. Probability*, **3**(5):790–801, (1975).

[84] 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).

[85] 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).

[86] 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).

[87] M. E. J. Newman. The structure and function of complex networks. *SIAM Rev.*, **45**(2):167–256 (electronic), (2003).

[88] M. E. J. Newman. Random graphs with clustering. *Phys. Rev. Lett.*, **103**:058701, Jul (2009).

[89] M. E. J. Newman, S. Strogatz, and D. Watts. Random graphs with arbitrary degree distribution and their application. *Phys. Rev. E*, **64**:026118, 1–17, (2000).

[90] M. E. J. Newman, S. Strogatz, and D. Watts. Random graph models of social networks. *Proc. Nat. Acad. Sci.*, **99**:2566–2572, (2002).

[91] M.E.J. Newman. Properties of highly clustered networks. *Physical Review E*, **68**(2):026121, (2003).

[92] M.E.J. Newman and J. Park. Why social networks are different from other types of networks. *Physical Review E*, **68**(3):036122, (2003).

[93] I. Norros and H. Reittu. On a conditionally Poissonian graph process. *Adv. in Appl. Probab.*, **38**(1):59–75, (2006).

[94] R. Pemantle. A survey of random processes with reinforcement. *Probab. Surv.*, **4**:1–79 (electronic), (2007).

[95] J. Pickands III. Moment convergence of sample extremes. *Ann. Math. Statistics*, **39**:881–889, (1968).

[96] B. Pittel. On tree census and the giant component in sparse random graphs. *Random Structures Algorithms*, **1**(3):311–342, (1990).

[97] B. Pittel. Note on the heights of random recursive trees and random m-ary search trees. *Random Structures Algorithms*, **5**(2):337–347, (1994).

[98] H. Reittu and I. Norros. On the power law random graph model of massive data networks. *Performance Evaluation*, **55**(1-2):3–23, (2004).

[99] 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.

[100] S. M. Ross. *Stochastic processes*. Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons Inc., New York, second edition, (1996).

[101] H.-J. Schuh and A. D. Barbour. On the asymptotic behaviour of branching processes with infinite mean. *Advances in Appl. Probability*, **9**(4):681–723, (1977).

[102] E. Seneta. The simple branching process with infinite mean. I. *J. Appl. Probability*, **10**:206–212, (1973).

[103] Y. Shavitt and E. Shir. Dimes - let the internet measure itself. *ACM SIGCOMM Computer Communication Review*, **35**(5):74, (2005).

[104] R. Smythe and H. Mahmoud. A survey of recursive trees. *Teor. Ĭmovīr. Mat. Stat.*, (51):1–29, (1994).

[105] B. Söderberg. General formalism for inhomogeneous random graphs. *Phys. Rev. E (3)*, **66**(6):066121, 6, (2002).

[106] B. Söderberg. Properties of random graphs with hidden color. *Phys. Rev. E (3)*, **68**(2):026107, 12, (2003).

[107] B. Söderberg. Random graph models with hidden color. *Acta Physica Polonica B*, **34**:5085–5102, (2003).

[108] B. Söderberg. Random graphs with hidden color. *Phys. Rev. E (3)*, **68**(1):015102, 4, (2003).

[109] S. Sreenivasan, T. Kalisky, L.A. Braunstein, S.V. Buldyrev, S. Havlin, and H.E. Stanley. Effect of disorder strength on optimal paths in complex networks. *Physical Review E*, **70**(4):46133, (2004).

[110] T.S. Turova. The largest component in subcriticalinhomogeneous random graphs. *Combinatorics, Probability and Computing*, **20**(01):131–154, (2011).

[111] T.S. Turova. Diffusion approximation for the components in critical inhomogeneous random graphs of rank 1. Preprint (2009).

[112] T.S. Turova and T. Vallier. Merging percolation on $\mathbf{Z}^d$ and classical random graphs: Phase transition. *Random Structures Algorithms*, **36**(2):185–217, (2010).

[113] W. Willinger, D. Alderson, and J.C. Doyle. Mathematics and the internet: A source of enormous confusion and great potential. *Notices of the American Mathematical Society*, **56**(5):586–599, (2009).