# Critical percolation on random networks with prescribed degrees

*Document Version:*
Publisher's PDF, also known as Version of Record (includes final page, issue and volume numbers)

*Please check the document version of this publication:*

# Critical Percolation on Random Networks with Prescribed Degrees

Critical Percolation on Random Networks with Prescribed Degrees

# Critical Percolation on Random Networks with Prescribed Degrees

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de
Technische Universiteit Eindhoven, op gezag van de
rector magnificus, prof.dr.ir. F.P.T. Baaijens, voor een
commissie aangewezen door het College voor
Promoties in het openbaar te verdedigen
op dinsdag 28 augustus 2018 om 13.30 uur

door

Souvik Dhara

geboren te Kolkata, India

Dit proefschrift is goedgekeurd door de promotoren en de samenstelling van de promotiecommissie is als volgt:

voorzitter: prof.dr. M.A. Peletier
1<sup>e</sup> promotor: prof.dr. R.W. van der Hofstad
2<sup>e</sup> promotor: prof.dr. J.S.H. van Leeuwaarden
leden: prof.dr.ir. S.C. Borst
prof.dr. N. Broutin (Sorbonne Université)
prof.dr. W.Th.F. den Hollander (UL)
prof.dr. N.V. Litvak
dr. L. Warnke (Georgia Institute of Technology)

Het onderzoek dat in dit proefschrift wordt beschreven is uitgevoerd in overeenstemming met de TU/e Gedragscode Wetenschapsbeoefening.

# Acknowledgment

The journey from a student to a researcher requires development of several aspects: sharpening knowledge, becoming aware of contemporary research, developing communication skills, and identifying sources of inspiration. In the four years of my PhD, I have received enormous support from many people in nourishing each of the above aspects. Before going into the contents of this thesis, I would like to express my gratitude to them for their contribution to my academic career.

I am deeply indebted to my supervisors Remco van der Hofstad and Johan van Leeuwaarden for guiding me through this journey. It is an honor to be your successor in the math genealogy-tree. Remco, your unending energy and enthusiasm for research are and will always be my inspiration. You have taught me everything from simple to deep mathematical facts. Our 17:30 meetings will always be in my cherished memories. You have always led with a brilliant example and showed me the highest standard for research and teaching. Johan, your thoughts always provided me with a unique perspective. Your questions would always bring out the best from our research projects. You have always given me the best career advices, and shared your own invaluable experiences.

I sincerely thank the committee members Sem Borst, Nicolas Broutin, Frank den Hollander, Nelly Litvak, and Lutz Warnke for a careful reading of my thesis and giving me feedback for improvement.

I have had wonderful experiences during research visits outside TU/e. I would like to thank Shankar Bhamidi for welcoming me heartily at the University of North Carolina Chapel Hill, and for sharing his enthusiasm and passion for research. Shankar, our trip to Nantahala is memorable to me. Special thanks to Michel Mandjes for sharing ideas from a vibrant research field which was orthogonal to my research project, and for adding new perspectives to my research.

Working with Sem Borst was a fantastic experience. I learned a lot about

# Contents

# Contents

# Chapter 1

# Introduction

The study of large networks, arising from applications in social, physical and life sciences, has witnessed meteoric growth over the past two decades. Recent technological innovations allow practitioners to access and study network data of ever-increasing size. A thorough understanding of large networks can often provide deep insights into the workings of complex systems. These networks are usually composed of two key components:

(1) *Structure:* The units/individuals are represented by nodes/vertices in the network, whereas their connectivity represents some sort of interaction. The connectivity structure or topology of these large networks are often either unknown or highly complex. Therefore, probabilistic models of graphs or random graphs have been used extensively to model real-world networks, along with their diverse structural characteristics.

(2) *Functionality:* Each of these networks come with certain functionality in the real world, such as information exchange or the spread of rumors/diseases. The functionality of networks is often modeled as a stochastic process acting upon the network.

From a mathematical perspective, the understanding of processes on random networks is interesting due to the inherent double randomness: the random structure produced by the random graph model, and the stochastic process living on this random structure. The interplay between the random topology and the stochastic process gives rise to novel behavior in the large network regime in terms of scaling limits and their analysis demands new tools from probability theory.

## Network properties and universality.

Discovering fundamental principles that can describe large complex networks has been a celebrated theme of research in network science. Below we discuss the main conceptual strands related to this research area:

**Sparsity and power-law degree distributions.** Empirical analysis of a number of real-world systems such as the internet, citation networks, and protein-protein interaction networks, seem to suggest that these networks are inherently sparse in the sense that the number of edges scales linearly with the number of vertices. Moreover, the empirical degree distributions follow approximately power-law distributions. Write $N_k$ for the number of vertices

with degree $k$ and $p_k = N_k/n$ for the degree distribution with $n$ being the number of vertices in the network. Then for large $k$,

$$p_k \approx Ck^{-\tau}, \text{ for } C > 0, \text{ and } \tau > 2.$$

The constant $\tau$ is called the *degree exponent* of the corresponding network. This feature makes it impossible to model real-world networks by classical homogeneous random graph models such as the Erdős-Rényi random graph or the random regular graph. Extensive discussions of the veracity as well as limitations of these findings, and the multitude of network models proposed to understand real-world data can be found in [6, 55, 77, 98, 137, 140] and the references therein.

**Universality in the large network limit.**   The second major thread that has emerged in network science, especially in the probability community, is the notion of universality: asymptotics in the large network limit, for a wide range of functionals, often depend mainly on the degree exponent $\tau$. In turn, this suggests that the qualitative behavior, across a plethora of models, can be largely insensitive to the details of the network model.

**Phase transition.**

Many random graph models for real-world networks are observed to exhibit a *phase transition*. For a communication network, where the link between any two servers can fail with a certain probability, one may expect to observe a transition in the connectivity structure of networks depending on the link failure probability. One may also consider a computer virus spreading over the internet, which becomes an epidemic over a short window of time. The study of phase transitions is often interesting in numerous applications in statistical physics as well. From the perspective of network science, the objective is to provide a framework for describing universality laws governing the phase transition in terms of basic characteristics like the degree distribution. However, even defining rigorously the meaning of a phase transition is a non-trivial task for the simplest models. Identifying the point and nature of phase transitions has been a fundamental question in the development of the current existing theory of random graphs and complex network models.

**Aim of this chapter.**

We discuss how the phase transition occurs in random graphs when percolation acts on them. This is a fundamental model for analyzing the effect of link failure or spread of epidemics on a given network. A detailed analysis is provided for the critical behavior in random graphs that generate networks with arbitrary degree distributions. The idea is to establish key relations between the network statistics such as the power-law exponent, and the nature of the critical behavior of this phase transition. In particular, we investigate different universality classes for the critical behavior of percolation based on the degree exponent $\tau$. In particular, for $\tau > 4$, the behavior lies in the same universality class as classical homogeneous random graph models. This shows that the inhomogeneity in the degree distribution does not influence the percolation critical behavior as long as the degree distribution has a finite third-moment. The behavior is more intricate in the other regimes with $\tau \in (3, 4)$ and $\tau \in (2, 3)$. In Section 1.1, we give an introduction to some of the most studied random graph models that will be pivotal to the discussions in this thesis. In Section 1.2, we define the percolation phase transition on finite graphs. In Section 1.3, we provide an overview of the rich history of critical behavior of percolation on finite graphs, and formulate the key questions associated to this literature. Different universality classes are also described in this section. In the following sections, we then provide answers to the key questions, and present our contributions to this literature. The overall goal of this introductory chapter is to describe the results at a high level, and discuss the central ideas behind our methods. The main results, along with all the associated technical details, will be discussed in full detail in subsequent chapters.

## 1.1   Random graph preliminaries

A graph $G = (V, E)$ consists of a vertex set $V$, and a set of edges $E \subset \{\{i, j\} : i, j \in V\}$ specifying the connections between different vertices. For a multigraph, $E$ is a multi-set possibly consisting of multiple-edges between vertices, as well as self-loops. Throughout, we will assume that $|V|, |E| < \infty$. A random graph model specifies a probability distribution over the space of graphs. We will consider $n$ vertices labeled by $[n] := \{1, 2, ..., n\}$, which will serve as the vertex set of the random graph. We now discuss some classical random graph models, and some properties related to their connectivity structure.

### 1.1.1   Random graph models

**Erdős-Rényi random graph.**    The Erdős-Rényi random graph is the simplest and most widely studied random graph model, where any two vertices share an edge with some fixed probability $p$, independently across edges. We denote the graph generated by the above procedure on $n$ vertices by $\text{ERRG}_n(p)$. This model was studied in the earliest work on random graphs by [78, 79, 89]. The model in [78, 79] chooses $M$ edges uniformly at random from all possible $\binom{n}{2}$ edges, and thus is slightly different. However, for $M \approx np$, the two models are asymptotically equivalent [115]. Note that the degree of each vertex is distributed as a $\text{Bin}(n-1, p)$ random variable. Thus for $p = c/n$, the asymptotic degree of each vertex is $\text{Poisson}(c)$, with fixed average degree $c$.

**The configuration model.**    Consider a non-increasing sequence of degrees $\mathbf{d} = (d_i)_{i \in [n]}$ such that $\ell_n = \sum_{i \in [n]} d_i$ is even. The configuration model on $n$ vertices having degree sequence $\mathbf{d}$ is constructed as follows [22, 40]:

> Equip vertex $j$ with $d_j$ stubs, or *half-edges*. Two half-edges create an edge once they are paired. Therefore, initially we have $\ell_n = \sum_{i \in [n]} d_i$ half-edges. Pick any one half-edge and pair it with a uniformly chosen half-edge from the remaining unpaired half-edges and keep repeating the above procedure until all the unpaired half-edges are exhausted.

Let $\text{CM}_n(\mathbf{d})$ denote the graph constructed by the above procedure. Note that $\text{CM}_n(\mathbf{d})$ may contain self-loops or multiple edges. Given any degree sequence, let $\text{UM}_n(\mathbf{d})$ denote the graph chosen uniformly at random from the collection of all simple graphs with degree sequence $\mathbf{d}$. It can be shown that the law of $\text{CM}_n(\mathbf{d})$, conditioned on the graph being simple, is the same as that of $\text{UM}_n(\mathbf{d})$ (see [98, Proposition 7.13]). Thus, in order to sample a graph uniformly from the space of all simple graphs with a given degree sequence $\mathbf{d}$, we can keep on generating the configuration model until we obtain a simple graph. It was shown in [18, 110] that, under very general assumptions, the asymptotic probability of the graph being simple is positive, so that with high probability we need to repeat the above algorithm only a finite number of times to generate $\text{UM}_n(\mathbf{d})$. The graph $\text{UM}_n(\mathbf{d})$, in the special case $d_i = d$ for some fixed $d$, is in the literature also known as the random $d$-regular graph.

**Inhomogeneous random graph.**    An inhomogeneous random graph is generated by equipping each vertex $i$ with a weight a $w_i$, and creating an edge

between vertices $i$ and $j$ with probability $p_{ij} = \kappa(w_i, w_j)$ independently, for some function $\kappa$. Thus the special case where $\kappa$ is a constant, i.e., $p_{ij} = p$ gives rise to $\mathrm{ERRG}_n(p)$ defined above. A detailed analysis of the properties of this random graph model has been provided in [44] under a very general setup. Some choices of $p_{ij}$ have been popular for their special properties:

> ▷ *Norros-Reittu model* [138]. $p_{ij} = 1 - \exp(-w_i w_j / L_n)$, where we define $L_n := \sum_{i \in [n]} w_i$. This model is also referred to as Poissonian graph model, or the Norros-Reittu random graph. We denote this model by $\mathrm{NR}_n(\boldsymbol{w})$.

> ▷ *Chung-Lu model* [61, 62]. $p_{ij} = \min\{w_i w_j / L_n, 1\}$. This model will be denoted by $\mathrm{CL}_n(\boldsymbol{w})$.

> ▷ *Generalized random graph* [54]. $p_{ij} = w_i w_j / (L_n + w_i w_j)$. We denote this model by $\mathrm{GRG}_n(\boldsymbol{w})$. This model has the property that the distribution of this random graph, conditionally on the degree sequence $\boldsymbol{d}$, is the same as for $\mathrm{UM}_n(\boldsymbol{d})$.

The weight $w_i$ plays a similar role as degree $d_i$ for $\mathrm{CM}_n(\boldsymbol{d})$. In fact, the expected degree of vertex $i$ is asymptotically $w_i$ in all the random graphs $\mathrm{NR}_n(\boldsymbol{w}, 1)$, $\mathrm{CL}_n(\boldsymbol{w})$, and $\mathrm{GRG}_n(\boldsymbol{w})$ under some regularity conditions. These models are often *asymptotically equivalent*. We refer the reader to [98, Chapter 6] for a detailed account of these properties.

In the subsequent sections, we will consider a sequence of degrees sequences $(\boldsymbol{d}_n)_{n \geqslant 1}$ and weight sequences $(\boldsymbol{w}_n)_{n \geqslant 1}$ while generating sequences of graphs. For notational convenience, we suppress the dependence of the degree and the weight sequences on $n$.

### 1.1.2  Existence of a giant component

A giant component exists if, in the large network limit, the proportion of vertices in the largest connected component stays bounded away from zero. In the sparse regime, where the number of edges scales linearly as the number of vertices, a unique giant component exists with high probability for most random graph models. Interestingly, global properties such as the existence of the giant component can be approximated by local properties, owing to the well-behaved topology of random graphs. To understand this more precisely, let $\mathscr{C}(v)$ denote the component containing vertex $v$ and let $V_n$ denote a vertex chosen uniformly at random. Then the existence of the giant component can be characterized by the following two fundamental properties:

**Branching process approximation.**    For all random graph models described above, the proportion of vertices involved in a cycle of length at most $2k$ is negligible, so that for any fixed $k \geqslant 1$, the $k$-neighborhood of $V_n$ is a tree with high probability. Therefore the finite neighborhoods of a randomly chosen vertex can be *approximated* by the neighborhoods of the root of an infinite random rooted tree. In all the random graph models mentioned above, the random tree is a branching process with a suitable progeny distribution. However, there are examples of random graph models where the approximating tree is not a simple branching process [26, 44]. The above notion of approximation can be formalized in terms of *local weak convergence*, introduced by Benjamini and Schramm [25] (see also the survey [16]). We refer the reader to [100] for local weak convergence results for random graphs. Thus, the so-called *local events* (events depending on finite neighborhoods of $V_n$) can be described by functionals of a branching process that are often tractable.

**Approximating global events by local events.**    The components having size $[\omega(n), n/\omega(n)]$, for any $\omega(n) \to \infty$, span an asymptotically negligible proportion of vertices, so that with high probability, either $|\mathscr{C}(V_n)|$ must have finite size, or $\mathscr{C}(V_n)$ is the giant component. Therefore, when a long path exists from $V_n$, then $V_n$ must be in the giant component, and $\mathscr{C}(V_n)$ is the giant. This is the reason why for random graph models in this section a unique giant component exists with high probability if and only if the mean of the progeny distribution of the approximating branching process is larger than one.

For $\mathrm{ERRG}_n(c/n)$, it is not difficult to see that the progeny distribution of the approximating branching process is $\mathrm{Poisson}(c)$, and indeed a giant component exists precisely when $c > 1$ [79, 89]. For $\mathrm{CM}_n(\mathbf{d})$, note that while pairing the $k$-th half-edge, the probability of pairing to a vertex of degree $r$ is approximately $rn_r/\ell_n$. Thus the degree of a neighbor of a vertex is approximately given by the *size-biased distribution*

$$\mathbb{P}(D_n^* = r) = \frac{rn_r}{\ell_n}. \tag{1.1}$$

Therefore, the approximating branching process has progeny $D_n^* - 1$, since one edge is connected to the parent of a vertex. The expectation of $D_n^* - 1$ is

$$\nu_n := \frac{\sum_{r=1}^{\infty} r(r-1)n_r}{\ell_n} = \frac{\sum_{i \in [n]} d_i(d_i - 1)}{\ell_n}.$$

It was established in [114, 133] (see also the recent results [120]) that the giant component exists precisely when

$$\lim_{n \to \infty} \nu_n = \nu > 1, \tag{1.2}$$

again confirming the local weak-limit heuristics. It can also be shown that for the inhomogeneous random graph models, the mean of the approximating branching process turns out to be $\sum_{i \in [n]} w_i^2 / \sum_{i \in [n]} w_i$, and the giant component exists [44, 98] when

$$\lim_{n \to \infty} \frac{\sum_{i \in [n]} w_i^2}{\sum_{i \in [n]} w_i} > 1.$$

## 1.2   Percolation on finite graphs

**Percolation process and Harris coupling.**   Given a graph G, bond (site) percolation refers to deleting each edge (vertex) independently with probability $1 - p$. Throughout, we will be interested in bond percolation; thus we simply write percolation which refers to bond percolation, and the obtained graph is denoted by $G(p)$. In case of percolation on random graphs, the deletion of edges is also independent from the underlying random graph. With the *percolation process*, we refer to the graph-valued stochastic process $(G(p))_{p \in [0,1]}$ coupled through the so-called Harris coupling. More precisely:

> Associate an independent uniform $[0,1]$ random variable $U_e$ to each edge $e$ of the graph G. $G(p)$ can be generated by keeping edge $e$ iff $U_e \leqslant p$. Keeping the uniform random variables fixed while varying $p$, gives a coupling between the graphs $(G(p))_{p \in [0,1]}$.

Classically, percolation has been extensively studied on infinite connected graphs such as the hypercubic lattice. This is the simplest known model that exhibits a *phase transition*. If $p$ is small, then $G(p)$ consists of connected components of finite size only. On the other hand, if $p$ is close to 1, then $G(p)$ contains an infinite cluster. Since for $p_1 < p_2$, $G(p_1)$ is a subgraph of $G(p_2)$ under the Harris coupling, there exists a *unique* value $p_c$ such that $G(p)$ contains an infinite cluster if and only if $p > p_c$. Thus $p_c$ can also be defined as the unique point of discontinuity of the function $\mathbb{P}(\exists \text{ an infinite component in } G(p))$. The quantity $p_c$ is called the *critical value* for the phase transition of the percolation process. Several questions like finding the value of $p_c$, establishing the uniqueness of the infinite components and behavior of different functionals close to $p_c$ for infinite transitive graphs have been discussed extensively in [47, 92, 96, 125].

**Phase transition on finite graphs.**   It is not evident how to define the phase transition for a fixed finite graph $G_n$ with $n$ vertices. For any $f : G_n \mapsto \mathbb{R}$, the expectation $\mathbb{E}[f(G_n(p))]$ is a smooth function in $p$, and therefore none of the functionals of the graph experience a transition. The phase transition can only arise when the graphs become large, i.e., it should be related to the large-network limit. For this reason, the phase transition is defined for a sequence of graphs $(G_n)_{n \geqslant 1}$ rather than a given fixed graph, and the transition is captured in terms of the limit $\lim_{n \to \infty} \mathbb{E}[f(G_n(p))]$. Let $\mathscr{C}_{(k)}(p)$ denote the k-th largest connected component of $G_n(p)$. The *critical value* $p_c = p_c(n)$ is defined such that the following holds given any $\varepsilon > 0$:

$$\lim_{n \to \infty} \mathbb{P}(|\mathscr{C}_{(1)}(p)| > \delta n) = \begin{cases} 0 & \text{for } p < p_c(1-\varepsilon), \quad \text{(subcritical)} \\ 1 & \text{for } p > p_c(1+\varepsilon), \quad \text{(supercritical)} \end{cases} \tag{1.3}$$

where the first limit should hold for all $\delta > 0$, and the second one for some $\delta = \delta(p) > 0$ sufficiently small. In most cases, $|\mathscr{C}_{(1)}(p)|/n$ converges in probability to some positive constant $\eta$ that depends on the graph sequence $(G_n)_{n \geqslant 1}$ and $p$, and more importantly $|\mathscr{C}_{(1)}(p)| \gg |\mathscr{C}_{(2)}(p)|$, so that the giant component is *unique*. We will stick to the above definition of the phase transition and critical value throughout this thesis. It is worthwhile mentioning that there is a substantial literature on how to define the critical value, and the phase transition. See [48, 96, 99, 118, 134] for different definitions of critical probability and related discussions. One could also note that $p_c$ is not unique for finite graphs. This is due to the fact that $p_c$ is allowed to depend on $n$ and the phase transition is only an asymptotic notion.

The study of random graphs practically started with the question of identifying the critical value of $p_c$. Note that percolation on a complete graph yields an Erdős-Rényi random graph, and in the early works [79, 89], it was shown that $p_c = 1/(n-1)$. Over the past 60 years of development of the random graph literature, identifying the critical value and the asymptotics of the giant component has been one of the guiding questions, not only for the percolation process, but for any sequence of dynamically growing graph processes. The threshold has been identified under fairly general conditions, for example when the underlying graph is an expander [23], converges in a cut metric to an irreducible graphon [43], or even general sequences of graphs [10, 60]. For more details we refer to [10, 17, 23, 42–44, 60, 85, 98, 109, 115, 145] and the references therein.

**Percolation on random graphs.**    For a sequence of random graphs, the phase-transition of the percolation process occurs when (1.3) holds with high probability with respect to the joint distribution of the random graph and the percolation process. The percolation process is viewed as a dynamic process living on a disordered medium, i.e., the random graph. In a sense, this may appear paradoxical, since percolation itself often serves as a model for generating a disordered medium on which stochastic processes like random walks act. However, when viewing random graphs as models for real-world networks, percolation serves as a model for robustness of internet or communication networks when the nodes/edges of the underlying network experience random damage. Percolation has also been used to model the vaccination on a network to prevent the growth of an epidemic. A detailed account of these applications can be found in [21, 137].

Of particular interest to this thesis is the phase transition result for uniformly chosen graphs with given degree $UM_n(\mathbf{d})$, and the configuration model $CM_n(\mathbf{d})$. For random d-regular graphs [17], $p_c = 1/(d-1)$. The percolation phase transition on $UM_n(\mathbf{d})$ and on $CM_n(\mathbf{d})$ was studied in [85, 109] when the empirical degree distribution has a finite second moment in the large network limit, as $n \to \infty$. It was shown that $p_c = 1/\nu$, where the parameter $\nu$ is defined by (1.2). See also [86, 119] for some recent results on more general degree sequences.

**Relation to branching process approximation.**    The critical probability $p_c$ is intimately related to the branching process approximation. Indeed, when the number of edges in $G_n(p)$ scales linearly with $n$, the typical local neighborhoods of $G_n(p)$ can be approximated by a branching process in the sense of Section 1.1.2 for most sequences of (random) graphs $(G_n)_{n \geqslant 1}$. The value $p_c$ is then such that the mean of this approximating branching process is 1. Indeed, that turns out to be the case for $ERRG_n(c/n)$, random regular graphs, and $CM_n(\mathbf{d})$. Heuristically, these results complement Schramm's conjecture about infinite transitive graphs stating that the local weak limit determines the percolation threshold. Without going into further details, we refer the interested reader to [24, 27, 75] and the references therein for a beautiful line of work initiated with Schramm's conjecture.

**Formation of a complex structure.**    Around the critical value, the phase transition happens not only with respect to the size of $\mathscr{C}_{(1)}$, but also with respect to the *complexity* of its connectivity structure. To measure complexity,

let us define for a connected graph G the number of surplus edges as the number of edges to be deleted to turn G into a tree. Thus # surplus edges of G = # edges − # vertices + 1, denoted by SP(G). Note that SP(G) = 0 means that G is a tree, and a large value of SP(G) means that G has many (possibly overlapping) cycles with a more complex structure. In the subcritical regime, any component has at most one surplus edge [42, 71, 94, 115], and there are finitely many surplus edges in the whole graph, so that the subcritical components are mostly trees. On the other hand, the giant component in the supercritical regime satisfies $SP(\mathscr{C}_{(1)}(p)) \to \infty$ [114, 115] with high probability, so that the structure of the giant is highly complex. See [72] for a detailed result about the giant component of $ERRG_n(c/n)$. Thus the percolation process starts adding cycles and the complex structure of the giant component begins to form precisely around the critical value $p_c$. This explains the interest in the percolation critical behavior.

### 1.2.1   Some definitions and notation.

In the next section, we discuss the critical window of phase transition. We now define some basic notation used throughout this thesis. We will use the standard notation $\xrightarrow{\mathbb{P}}$, $\xrightarrow{d}$ to denote convergence in probability and in distribution or law, respectively. We often use the Bachmann Landau notation $O(\cdot)$, $o(\cdot)$, $\Theta(\cdot)$ for large $n$ asymptotics of real numbers. The topology needed for the distributional convergence will always be specified unless it is clear from the context. A sequence of events $(\mathcal{E}_n)_{n \geqslant 1}$ is said to occur with high probability with respect to probability measures $(\mathbb{P}_n)_{n \geqslant 1}$ if $\mathbb{P}_n(\mathcal{E}_n) \to 1$. Denote $f_n = O_{\mathbb{P}}(g_n)$ if $(|f_n|/|g_n|)_{n \geqslant 1}$ is tight; $f_n = o_{\mathbb{P}}(g_n)$ if $(|f_n|/|g_n|)_{n \geqslant 1}$ converges in probability to zero; $f_n = \Theta_{\mathbb{P}}(g_n)$ if $f_n = O_{\mathbb{P}}(g_n)$ and $g_n = O_{\mathbb{P}}(f_n)$. Denote by

$$\ell_{\downarrow}^p := \big\{ \mathbf{x} = (x_1, x_2, x_3, ...) : x_1 \geqslant x_2 \geqslant x_3 \geqslant ... \text{ and } \sum_{i=1}^{\infty} x_i^p < \infty \big\},$$

the subspace of non-negative, non-increasing sequences of real numbers with square norm metric $d(\mathbf{x}, \mathbf{y}) = (\sum_{i=1}^{\infty} (x_i - y_i)^p)^{1/p}$. Let $(\ell_{\downarrow}^2)^k$ denote the k-fold product space of $\ell_{\downarrow}^2$. With $\ell_{\downarrow}^2 \times \mathbb{N}^{\infty}$, we denote the product topology of $\ell_{\downarrow}^2$ and $\mathbb{N}^{\infty}$, where $\mathbb{N}^{\infty}$ denotes the collection of sequences on $\mathbb{N}$, endowed with the product topology. Define also

$$\mathbb{U}_{\downarrow} := \Big\{ ((x_i, y_i))_{i=1}^{\infty} \in \ell_{\downarrow}^2 \times \mathbb{N}^{\infty} : \sum_{i=1}^{\infty} x_i y_i < \infty \text{ and } y_i = 0 \text{ whenever } x_i = 0 \Big\}$$

with the metric

$$d_{\mathbb{U}}((\mathbf{x}_1, \mathbf{y}_1), (\mathbf{x}_2, \mathbf{y}_2)) := \left( \sum_{i=1}^{\infty} (x_{1i} - x_{2i})^2 \right)^{1/2} + \sum_{i=1}^{\infty} |x_{1i} y_{1i} - x_{2i} y_{2i}|.$$

Further, we introduce $\mathbb{U}_{\downarrow}^0 \subset \mathbb{U}_{\downarrow}$ as

$$\mathbb{U}_{\downarrow}^0 := \left\{ ((x_i, y_i))_{i=1}^{\infty} \in \mathbb{U}_{\downarrow} : \text{ if } x_k = x_m, k \leqslant m, \text{ then } y_k \geqslant y_m \right\}.$$

We usually use the boldface notation $\mathbf{X}$ for a time-dependent stochastic process $(X(s))_{s \geqslant 0}$, unless stated otherwise, $\mathbb{C}[0, t]$ denotes the set of all continuous functions from $[0, t]$ to $\mathbb{R}$ equipped with the topology induced by sup-norm $\| \cdot \|_t$. Similarly, $\mathbb{D}[0, t]$ (resp. $\mathbb{D}[0, \infty)$) denotes the set of all càdlàg functions from $[0, t]$ (resp. $[0, \infty)$) to $\mathbb{R}$ equipped with the Skorohod $J_1$ topology [37].

## 1.3    Critical window and emergence of the giant

The critical regime lies on the boundary between the subcritical and supercritical regimes, where the system exhibits an intermediate behavior. From a statistical physics perspective, this is the interesting regime to study because the properties in the critical regime help to answer the question "How did the phase transition happen?". Here, one tries to identify principles that govern the phase transition, which not only depend on the specifics of the system, but hold universally for a large class of systems. From a mathematical perspective, critical behavior often gives rise to novel scaling limit results. In this section, we first discuss the importance and relevance of studying the critical behavior for percolation processes. Then we state some key questions about the critical behavior of percolation in Section 1.3.1. In Section 1.3.2, we describe three fundamental types of critical behavior, i.e., universality classes, that will be crucial throughout this thesis. We finish this section with a review of the related literature and the relevance of our work; see Section 1.3.3.

**Critical window.**    To observe the critical behavior, one must take $p = p_c(1 \mp \varepsilon_n)$ in (1.3), for some $\varepsilon_n \to 0$ as $n \to \infty$. Interestingly, the critical behavior is not observed for any $\varepsilon_n$; there is a range of $\varepsilon_n$ where the graph shows qualitatively similar features as the sub/supercritical regimes and the critical behavior is observed only when $\varepsilon_n$ is chosen appropriately. In most situations, this means that $\varepsilon_n = \Theta(n^{-\eta})$, where $\eta > 0$ is a model-dependent constant. To be more precise, recall that $\mathscr{C}_{(i)}(p)$ denotes the $i$-th largest com-

ponent of $G_n(p)$. The following are classical results [42, 115] for $G_n(p) =$ Erdős-Rényi random graph (i.e., $G_n$ is the complete graph), where $\eta = 1/3$:

(a) *Barely subcritical regime:* $p = p_c(1 - \varepsilon_n)$ with $\varepsilon_n n^\eta \to \infty$. Then for each fixed $i \geqslant 1$, as $n \to \infty$,

$$\frac{|\mathscr{C}_{(i)}(p)|}{2\varepsilon_n^{-2}\log(n\varepsilon_n^3)} \xrightarrow{\mathbb{P}} 1, \quad \text{and} \quad \mathbb{P}(\exists i : \mathrm{SP}(\mathscr{C}_{(i)}(p)) > 1) \to 0.$$

Thus, $G_n(p)$ shows the two characteristic features of the subcritical regime: $|\mathscr{C}_{(1)}(p)|$ is not distinctively larger than $|\mathscr{C}_{(2)}(p)|$, and $G_n(p)$ is essentially a collection of trees. Thus, even if $p \approx p_c$, $G_n(p)$ is subcritical in this regime. This regime is often referred to as the barely subcritical regime in the literature.

(b) *Barely supercritical regime:* $p = p_c(1 + \varepsilon_n)$ with $\varepsilon_n n^\eta \to \infty$. Then, as $n \to \infty$,

$$\frac{|\mathscr{C}_{(1)}(p)|}{2n\varepsilon_n} \xrightarrow{\mathbb{P}} 1, \quad \frac{\mathrm{SP}(\mathscr{C}_{(1)}(p))}{n\varepsilon_n^3} \xrightarrow{\mathbb{P}} \frac{2}{3}, \quad \frac{|\mathscr{C}_{(2)}(p)|}{|\mathscr{C}_{(1)}(p)|} \xrightarrow{\mathbb{P}} 0,$$

$$\text{and} \quad \mathbb{P}(\exists i \geqslant 2 : \mathrm{SP}(\mathscr{C}_{(i)}(p)) > 1) \to 0.$$

See [112, Section 23], [42, 115]. Thus, $G_n(p)$ exhibits two characteristic features in the supercritical regime: $|\mathscr{C}_{(1)}(p)|$ is considerable larger than all other components, and $\mathscr{C}_{(1)}(p)$ is complex in the sense that there is a growing number of surplus edges, while all other components are trees.

Although the above formulations are stated for the Erdős-Rényi random graph, the recent literature has provided many interesting results about the barely subcritical regimes [31, 33, 108] and supercritical regimes [101, 114] for graphs with general degree sequence (see also [94, 123, 144] for results in both regimes). Now, the phase transition takes place between the barely subcritical and supercritical regimes when $\varepsilon_n \sim n^{-\eta}$. This regime is known as the *critical window* for the phase transition. More precisely, the critical window is defined to be the values of $p$ given by

$$p_c(\lambda) = p_c(1 + \lambda n^{-\eta}), \quad -\infty < \lambda < \infty. \tag{1.4}$$

In this regime, the largest components exhibit features that are completely different than the subcritical, or the supercritical regime: There exists a model-dependent exponent $\rho > 0$ such that

$$n^{-\rho}(|\mathscr{C}_{(i)}(p)|)_{i \geqslant 1} \text{ converges to a non-degenerate random vector.} \tag{1.5}$$

| $\varepsilon > 0$    $\varepsilon_n \gg n^{-\eta}$ | $\varepsilon_n \sim n^{-\eta}$ | $\varepsilon_n \gg n^{-\eta}$    $\varepsilon > 0$ |
|:---:|:---:|:---:|
| Subcritical | Critical window | Supercritical |
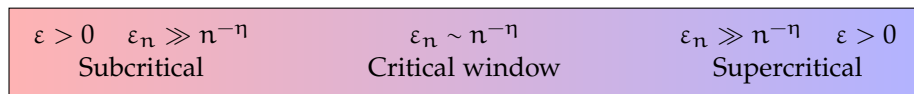
Mostly trees $\longrightarrow$ Components merge $\longrightarrow$ Birth of giant

Further, for any $i \geqslant 1$, $\liminf_{n\to\infty} \mathbb{P}(\mathrm{SP}(\mathscr{C}_{(i)}(p)) > 1) > 0$, but $\mathrm{SP}(\mathscr{C}_{(i)})$ is tight, so that the surplus-edge count for large components starts to grow in the critical window. The above two properties hold for all values of $\lambda$ in (1.4); in this sense there is not a single critical value, but a whole "window" of critical values over which the phase transition happens. This is due to finite-size effects and the joint scaling of $\varepsilon_n$ and $n$, a feature that is typically absent in the case of the phase transition on infinite graphs. The exponent $\eta$ in (1.4) is chosen as largest value such that the limit of (1.5) depends on $\lambda$, so that $\eta$ is uniquely defined.

Paul Erdős described the percolation process as the race between the components to become the giant [2]. The mental picture is that the collection of trees in the barely subcritical regime are the participants of this race and the component that outnumbers the other components in terms of the number of vertices wins the race. As the percolation parameter transitions through the critical window with $\lambda$ increasing with the Harris coupling in place, components grow in size and complexity, and the race is on. $\mathscr{C}_{(1)}(p_c(\lambda_1))$ and $\mathscr{C}_{(1)}(p_c(\lambda_2))$ can be completely disjoint sets of vertices for $\lambda_1 \neq \lambda_2$. However, at the end of the critical window, when $\lambda$ becomes sufficiently large, the leader $\mathscr{C}_{(1)}(p_c(\lambda))$ stops changing and this leader becomes the young giant component at the end of the critical window. At the barely supercritical phase, the race ends and the largest component stays the largest throughout the future of the percolation process. See [2] for a formalization of this picture under a general setup.

It is worthwhile to highlight the fact in (1.5) that the component sizes, after proper rescaling, converge to non-degenerate random variables. This is a special feature of the critical window that is never observed in the sub/supercritical regime. In fact, to the best of our knowledge, all dynamic graph processes that show phase transition with respect to its component sizes, exhibits this feature. Thus, this property could be considered as a potential definition of the critical window.

### 1.3.1    Key questions

We now describe the key questions about the percolation process in the critical window that we address in this thesis.

**(1) Component sizes and surplus edges.**    The phase transition typically happens with respect to functionals such as the size of the largest components and their surplus edges. Therefore, the most natural approach in this context is to find limit theorems for these functionals. For each fixed $\lambda$, consider $p_c(\lambda)$ defined in (1.4), and define

$$\mathbf{Z}_n(\lambda) := (n^{-\rho} |\mathscr{C}_{(i)}(p_c(\lambda))|, SP(\mathscr{C}_{(i)}(p_c(\lambda))))_{i \geqslant 1}$$

for some model-dependent constant $\rho > 0$. As discussed in (1.5), $\mathbf{Z}_n(\lambda)$ is expected to converge in distribution to some non-degenerate random vector. Since one deals with convergence of infinite-dimensional random vectors, the topology for the underlying distributional convergence turns out to be important, because one gets convergence of more functionals under a stronger topology. The results of this thesis will be discussed under the $\mathbb{U}_{\downarrow}^0$ topology, defined in Section 1.2.1.

**(2) Evolution when passing through the critical window.**    As mentioned before, there is not a single critical value here, but a whole window of critical values. It is thus interesting to explore the relation between the relevant component functionals for different values of $\lambda$. Now, under the Harris coupling, $\mathbf{Z}_n(\lambda)$ can be viewed as a stochastic process in $-\infty < \lambda < \infty$. As $\lambda$ increases, more and more edges get added in $G_n(p_c(\lambda))$, and $\mathbf{Z}_n(\lambda)$ evolves. In the context of the "race to become a giant", $(\mathbf{Z}_n(\lambda))_{\lambda \in \mathbb{R}}$ is the movie of this race. Therefore it is desirable to study the limit of the stochastic process $(\mathbf{Z}_n(\lambda))_{\lambda \in \mathbb{R}}$. This is a $\mathbb{U}_{\downarrow}^0$-valued process, and we will consider the topology $(\mathbb{U}_{\downarrow}^0)^{\mathbb{N}}$ for convergence of this process.

**(3) Global metric structure.**    A recent direction in this literature aims to find the global structure, and characterize the distance-related functionals of the components. The motivation comes from understanding the minimal spanning tree on a random network, which is important in many contexts like the spread of epidemics. Of course, the term *global structure* is a bit vague; however, this can be formalized. Each component $\mathscr{C} \subset G_n((p_c(\lambda))$ can be viewed as a metric space, equipped with a measure on the associated Borel

sigma algebra. The metric on $\mathscr{C}$ is the graph-distance where (i) each edge has length one, (ii) the measure being proportional to the counting measure, i.e., for any $A \subset \mathscr{C}$, the measure of $A$ is given by $\mu_{ct}(A) = |A|/|\mathscr{C}|$. Then, $\mathscr{C}_{(i)}(p_c(\lambda))$ can be viewed as a random element from $\mathscr{M}$, the space of metric spaces with an associated probability measure. For $M = (M, d, \mu) \in \mathscr{M}$ and $a > 0$, define $aM$ to be the measured metric space $(M, ad, \mu)$. Then the goal is to

$$\text{find the distributional limit of } \left(n^{-\delta}\mathscr{C}_{(i)}(p_c(\lambda))\right)_{i \geqslant 1}.$$

Since the limit is obtained after rescaling of graph-distances by $n^\delta$, and the limit is usually a compact metric space, the distances in $\mathscr{C}_{(i)}(p_c(\lambda)$ scale as $n^\delta$. The above quantity is an $\mathscr{M}$-valued sequence. Of course, the topology on $\mathscr{M}$ is important, and we will explore two different topologies, namely, the Gromov weak-topology and the Gromov-Hausdorff-Prokhov topology.

Another key question, which should have been stated as question (0) but will not be a topic of our discussion, is finding the value of $\eta$. For graphs, and in particular those with an underlying geometric structure, finding $\eta$ is a highly non-trivial task. An interested reader is referred to [82, 104] and the references therein. In the context of the models in this thesis, values of $\eta$ are described below while discussing the different universality classes.

### 1.3.2  Major universality classes

In a seminal work, Aldous [8] studied the first two questions above in the context of Erdős-Rényi random graphs. It turns out that $\eta = 1/3$, and $\rho = 2/3$. Along with identifying the limiting object for the component sizes, Aldous observed that the evolution of the rescaled component sizes can be described by a process called the multiplicative coalescent; see Chapter 2 for a precise definition. The first result about the convergence of the global structure was provided recently in [3] for the critical Erdős-Rényi random graphs with $\delta = 1/3$. Subsequently, there has been a surge in the literature to understand the most general cases under which one can establish qualitatively similar behavior as the Erdős-Rényi random graph, and identify the cases when the behavior is different. Following the above discussion, two universality classes have emerged in the literature. It turns out that, when the asymptotic degree distribution follows a power-law with exponent $\tau$, there is a transition in the critical behavior with respect to the exponent $\tau$.

**Erdős-Rényi universality class.**    For $\tau > 4$, the asymptotic empirical degree distribution has a finite third moment. In this case, the critical window turns out to be $p = p_c(1 + \lambda n^{-1/3})$, the maximal component sizes $|\mathscr{C}_{(i)}(p_c(\lambda))|$, for any fixed $i$, are of the order $n^{2/3}$ in the critical regime, whilst typical distances in these maximal connected components scale like $n^{1/3}$. Thus $\rho = 2/3$ and $\eta = \delta = 1/3$. Moreover, the scaling limits are the same as for the Erdős-Rényi random graphs up to constant factors of adjustment in the parameters.

**Heavy-tailed behavior.**    For $\tau \in (3, 4)$, the asymptotic degree distribution has an infinite third moment, but a finite second moment. Here the critical window turns out to be $p = p_c(1 + \lambda n^{-(\tau-3)/(\tau-1)})$, $|\mathscr{C}_{(i)}(p_c(\lambda))|$ is of the order $n^{(\tau-2)/(\tau-1)}$, whilst distances scale like $n^{(\tau-3)/(\tau-1)}$. Thus $\rho = (\tau-2)/(\tau-1)$ and $\eta = \delta = (\tau-3)/(\tau-1)$. The scaling limits turn out to be completely different in this regime. For example, it turns out that the high-degree vertices play a crucial role in the connectivity structure of $\mathscr{C}_{(i)}(p_c(\lambda))$ in the sense that a deliberate deletion of the $k$-th highest degree vertex changes the scaling limit completely. This is in sharp contrast with the behavior for the $\tau > 4$ regime.

**Universality in the evolution of the components.**    To intuitively understand the evolution of the component sizes and surplus edges, let us consider the Erdős-Rényi case. After increasing $p$ slightly, a new edge might appear in the graph, and due to the homogeneity in the connectivity structure of Erdős-Rényi random graphs, this edge selects two end-points uniformly at random. For this reason, two components $\mathscr{C}_{(i)}(p_c(\lambda))$ and $\mathscr{C}_{(j)}(p_c(\lambda))$ merge at rate $|\mathscr{C}_{(i)}(p_c(\lambda))| \times |\mathscr{C}_{(j)}(p_c(\lambda))|$ and create a component of size $|\mathscr{C}_{(i)}(p_c(\lambda))| + |\mathscr{C}_{(j)}(p_c(\lambda))|$. Moreover, a surplus edge is created in $\mathscr{C}_{(i)}(p_c(\lambda))$ at rate $|\mathscr{C}_{(i)}(p_c(\lambda))| \times (|\mathscr{C}_{(i)}(p_c(\lambda))| - 1)/2$. This merging dynamics of a collection of particles according to the product of their weights is known as the *multiplicative coalescent* [8, 9]. The creation of surplus edges can also be augmented in the evolution of the component sizes [30]. In both the $\tau > 4$ and $\tau \in (3, 4)$ regimes, the above merging dynamics describes the evolution of the component sizes and surplus edges over the critical window for a wide array of models, in-spite of the dependence in the connectivity structure. Thus, even if the scaling limits for $\tau > 4$ and $\tau \in (3, 4)$ are completely different for each fixed $\lambda$, the merging dynamics as $\lambda$ varies is guided by the same dynamics.

### 1.3.3    Literature review and the relevance of our work

Each of the key questions (1)–(3) in Section 1.3.1 have posed novel theoretical challenges in probability theory and combinatorics over the past decades. The study of critical random graphs began in the 1990's with the early works [41, 112, 129, 130] on critical Erdős-Rényi random graphs, where it was shown that the critical window is $p = n^{-1}(1 + \lambda n^{-1/3})$, and the component sizes are of the order $n^{2/3}$, whereas the surplus edges are $O(1)$. In a seminal work [8], Aldous derived the exact scaling limits of the rescaled component sizes and surplus edges, and showed that the evolution of the component sizes over the critical window can be described by the multiplicative coalescent process. This initiated the program for a large body of subsequent work [15, 30, 32, 57, 66, 83, 97, 103, 106, 121, 135, 136, 144], showing that the behavior of a wide array of random graphs at criticality is universal in the sense that it does not depend on the precise description of the model. Of particular relevance to this thesis are the works on $CM_n(\mathbf{d})$ [121, 135, 144]. The question (1), for the d-regular case, was extensively analyzed in [135], and the scaling limit for $\mathbf{Z}_n(\lambda)$ was derived for the critical $CM_n(\mathbf{d})$ with bounded maximum degree in [144] under the product topology. The results in [121] considered the special case that the degrees are an iid sample from a distribution having finite third-moment. Scaling limit results were derived for the component sizes; however there is no notion of "critical window" in this set up. In Section 1.4.2, we discuss the joint convergence of the component sizes and the surplus edges when the degree distribution satisfies a finite third-moment condition.

The second major universality class emerged with the study of dynamically evolving random networks given by the Norros-Reittu random graph model, with a heavy-tailed empirical distribution of average degrees. In [33, 97], the critical window was identified, along with scaling limit results for component sizes. In Section 1.4.3, we will show that the scaling limit of $\mathbf{Z}_n(\lambda)$ for $CM_n(\mathbf{d})$ under the heavy-tailed setup lies in the universality class of [33]. In fact, the results are stronger than [33] in terms of the topology of convergence. Joseph [121] studied the iid degree case, where the scaling limit turns out to be somewhat different than [33].

In the context of the evolution of the component sizes, Aldous [8] first studied the evolution of the component sizes. The evolution of the component sizes was also studied in the context of random graphs with immigrating vertices [15], and the Norros-Reittu random graph [33]. See also [57] for a construction of the multiplicative coalescent. A complete description of this process, along with its entrance boundary conditions, was provided in

[11]. This was generalized to augmented multiplicative coalescent processes in [30] to capture the evolution of the surplus edges as well. In Section 1.5, we describe the evolution of $\mathbf{Z}_n(\lambda)$ in both universality classes.

The study of the global metric structure is a recently emerging direction in this field, which started with the pioneering work [3] on critical Erdős-Rényi random graphs. The scaling limit identified in [3] was shown to be universal for the $\tau > 4$ regime in a recent line of work [29, 34, 35]. In the context of critical random graphs with degree-exponent $\tau \in (3, 4)$, candidate limit laws of maximal components with each edge rescaled to have length $1/n^{(\tau-3)/(\tau-1)}$ were established in [36]. In Section 1.6, we describe a "universality principle" for the $\tau \in (3, 4)$ regime, which yields the scaling limits for $CM_n(\mathbf{d}, p_c(\lambda))$.

### 1.3.4    A new universality class.

All the above literature assumes a finite second-moment condition on the degree distribution, and thus does not include the $\tau \in (2, 3)$ case, where the asymptotic degree distribution has an infinite second moment but a finite first moment. These networks are known in the literature as *scale-free networks* [21]. One of the popular features of scale-free networks is that these networks are *robust* under random edge-deletion, i.e., for any sequence $(p_n)_{n \geqslant 1}$ with $\liminf_{n \to \infty} p_n > 0$, the graph obtained by applying percolation with probability $p_n$ is always supercritical. This feature has been studied experimentally in [7], using heuristic arguments in [58, 63, 64, 74] (see also [51, 52, 95] in the context of optimal paths in the strong disorder regime), and mathematically in [45]. Thus, in order to observe the percolation critical behavior, one needs to take $p_c \to 0$ with the network size, even if the average degree of the network is finite. It was predicted from the physics literature that the critical value should be $p_c \sim n^{-(3-\tau)/(\tau-1)}$: Detailed properties of the component sizes and structures remained as open question.

In Section 1.7, we discuss the first mathematically rigorous results in the $\tau \in (2, 3)$ regime for component sizes and their complexity. The most striking thing about the results in the $\tau \in (2, 3)$ regime is that the critical value changes depending on whether the underlying random graph has the so-called *single-edge constraint*, i.e., the critical value when the underlying graph is a random multigraph generated by the configuration model is different than that under models like the erased configuration model $ECM_n(\mathbf{d})$ and the generalized random graph $GRG_n(\boldsymbol{w})$, where the underlying graph is simple. This feature was never observed in the finite second-moment scenario. For $CM_n(\mathbf{d})$ the critical value indeed turns out to be $p_c \sim n^{-(3-\tau)/(\tau-1)}$,

whereas for $\mathrm{ECM}_n(\mathbf{d})$ or $\mathrm{GRG}_n(\mathbf{w})$, we find that $p_c \sim n^{-(3-\tau)/2}$. The largest component sizes in both regimes are of the order $n^\alpha p_c$, and the scaling limits are in a completely different universality class than in the $\tau \in (3,4)$ and $\tau > 4$ cases.

### 1.3.5   Discussion

**Relation to branching process approximations.**   The distinction between the universality classes $\tau > 4$ and $\tau \in (3,4)$ can also be seen in terms of the branching process approximation. Recall that for $\mathrm{CM}_n(\mathbf{d})$ or $\mathrm{UM}_n(\mathbf{d})$, the local neighborhoods can be approximated by a branching process. The progeny distribution is $D_n^* - 1$, where $D_n^*$ is given by (1.1). This distribution has asymptotically finite variance if and only if the third moment of the asymptotic degree distribution is finite (i.e., $\tau > 4$). It is known that for critical branching processes the growth rate of the neighborhoods crucially depends on the variance [1, 126]. In fact the height scales as $|T|^{1/2}$ and $|T|^{(\tau-3)/(\tau-2)}$ in the $\tau > 4$ and $\tau \in (3,4)$ regimes, respectively, where $|T|$ is the total progeny of the branching process. Now, heuristically speaking, if $\mathscr{C}_{(1)}(p_c(\lambda))$ was a tree, then from the theory of branching processes, one would expect the following relations to be true:

$$\eta = \frac{\rho}{2}, \text{ for } \tau > 4, \quad \text{and} \quad \eta = \frac{\tau-3}{\tau-2}\rho, \text{ for } \tau \in (3,4). \tag{1.6}$$

Following [92], we refer to the identities in (1.6) as scaling relations. This aligns with the exponents suggested above for the two regimes. Intuitively, the above relations should hold since $\mathscr{C}_{(1)}(p_c(\lambda))$ is a tree, i.e., $\mathrm{SP}(\mathscr{C}_{(1)}(p_c(\lambda))) = 0$, with probability bounded away from zero. A disclaimer to the reader is that the bounds in [1, 126] are proved for a fixed branching process rather than a sequence of those. For a more rigorous explanation (1.6), an interested reader is referred to [97]. For $\tau \in (2,3)$, the branching process approximation does not work for $\mathrm{CM}_n(\mathbf{d})$ due to the presence of multiple edges.

**About the $\mathbb{U}_\downarrow^0$-topology.**   The distributional convergence of $\mathbf{Z}_n(\lambda)$ under the $\mathbb{U}_\downarrow^0$-topology implies convergence of many interesting functionals. Let $\mathscr{C}(v)$ denote the connected component containing vertex $v$ in $G_n(p)$, and let $V_n$ denote a vertex chosen uniformly at random from $[n]$, independently of $G_n(p)$. One example is a quenched version of the susceptibility function defined as

$\mathbb{E}[|\mathscr{C}(V_n)|\|G_n(p)]$. Note that

$$\mathbb{E}\big[\,|\mathscr{C}(V_n)|\,\big|\,G_n(p_c(\lambda))\big] = \frac{1}{n}\sum_{v\in[n]}|\mathscr{C}(v)| = \frac{1}{n}\sum_{i\geqslant 1}|\mathscr{C}_{(i)}(p_c(\lambda))|^2, \quad (1.7)$$

and therefore the convergence in $\mathbb{U}_\downarrow^0$-topology implies the convergence of $n^{-1/3}\mathbb{E}[|\mathscr{C}(V_n)|\|G_n(p_c(\lambda))]$. We also get the convergence of the quantity $n^{-2/3}\sum_{i\geqslant 1}|\mathscr{C}_{(i)}(p_c(\lambda))|\mathrm{SP}(\mathscr{C}_{(i)}(p_c(\lambda)))$, which in particular implies that the components of small size cannot contain too many surplus edges. The relevance of this topology is also discussed in [30] (see also [1]), because this turns up naturally in defining the augmented multiplicative coalescent, and establishing a version of the Feller property.

**Scaling relation.**    The following scaling relation is true in both the regimes $\tau > 4$ and $\tau \in (3,4)$:

$$1+\eta = 2\rho. \qquad (1.8)$$

This can be understood intuitively. Since the component sizes converge in the $\ell_\downarrow^2$-topology, one can expect that the expected value of $\sum_{i\geqslant 1}|\mathscr{C}_{(i)}(p_c(\lambda))|^2$ is of the order $n^{2\rho}$. One may also use (1.7) to calculate this expectation. In fact, (1.7) implies the scaling relation (1.8) if $\mathbb{E}[|\mathscr{C}(V_n)|] = O(n^\eta)$. Now $|\mathscr{C}(V_n)| \leqslant \sum_{l\geqslant 1}P_l$, where $P_l$ is the expected number of paths of length $l$ starting from vertex $v$. Using the branching process approximation for $CM_n(\mathbf{d})$, $\mathbb{E}[P_l] \leqslant C(p_c\nu_n)^l$, for some constant $C > 0$. Summing this estimate over $l$, one gets

$$\mathbb{E}[\mathscr{C}(V_n)] \leqslant C\sum_{l\geqslant 1}(p_c\nu_n)^l \leqslant C\sum_{l\geqslant 1}\left(1+\frac{\lambda}{n^\eta}\right)^l \leqslant Cn^\eta,$$

for $\lambda < 0$, which yields a heuristic derivation of (1.8). The above path counting technique has been formalized in [31, 111, 116].

**Effect of slowly-varying corrections.**    Suppose that the asymptotic degree distribution satisfies $\mathbb{P}(D \geqslant x) \sim L(x)x^{-(\tau-1)}$ with $L(\cdot)$ some slowly-varying function. For $\tau > 4$, the scaling limits, as well as the exponents, are insensitive to $L(\cdot)$. On the contrary, the component size, or even the critical window, depends crucially on the slowly-varying function for $\tau \in (3,4)$. The critical window becomes $p = p_c(1+\lambda L^*(n)^2n^{-\eta})$, and the component sizes turn out to be of the same order as $n^\rho/L^*(n)$, for some slowly varying $L^*$. However, the scaling limits lie in the same universality class; see Chapter 3.

## 1.4    Component sizes and surplus edges

In this section, we provide an outline of the proofs for establishing scaling limits of $\mathbf{Z}_n(\lambda)$ for the random graph $CM_n(\mathbf{d}, p_c(\lambda))$. We provide the key ideas, and the strategy of the proof, leaving many details for the later chapters. In Section 1.4.1, we start by describing a construction of $CM_n(\mathbf{d}, p)$ due to Janson [109], which is a key tool throughout this thesis. This construction allows us to treat $CM_n(\mathbf{d}, p_c(\lambda))$ as a configuration model with a suitable degree distribution, which can be easier to work with due to the sequential construction provided in Section 1.1. In Section 1.4.2, we consider the scaling limit for $\mathbf{Z}_n(\lambda)$ for the finite third-moment case, and outline a detailed proof strategy. The infinite third-moment case is considered in Section 1.4.3.

### 1.4.1    Janson's construction

Suppose that $\mathbf{d}'$ is the random degree sequence obtained after percolation. Fountoulakis [85] showed that, conditionally on $\mathbf{d}'$, the law of $CM_n(\mathbf{d}, p)$ is same as the law of $CM_n(\mathbf{d}')$. Often asymptotics of different functionals of $\mathbf{d}'$ can be calculated, which gives a powerful tool to deal with percolation on random graphs with general degree sequence. The following explicit construction of $CM_n(\mathbf{d}, p)$ is due to Janson [109], provided in the context of identifying the percolation phase transition on $CM_n(\mathbf{d})$. This construction will be crucial in what follows.

**Algorithm 1.1.** (S1) For each half-edge $e$, let $\nu_e$ be the vertex to which $e$ is attached. With probability $1 - \sqrt{p}$, one detaches $e$ from $\nu_e$ and associates $e$ to a new vertex $\nu'$ that we color red. This is done independently for every half-edge. Let $n_+$ be the number of red vertices created and $\tilde{n} = n + n_+$. Suppose that $\tilde{\mathbf{d}} = (\tilde{d}_i)_{i \in [\tilde{n}]}$ is the new degree sequence obtained by the above procedure, i.e., $\tilde{d}_i \sim \text{Bin}(d_i, \sqrt{p})$ for $i \in [n]$ and $\tilde{d}_i = 1$ for $i \in [\tilde{n}] \setminus [n]$.

(S2) Construct $CM_{\tilde{n}}(\tilde{\mathbf{d}})$, independently of (S1).

(S3) Delete all the red vertices. Alternatively, one can choose any $n_+$ degree-one vertices uniformly at random without replacement, independently of (S1) and (S2), and delete them.

An edge is kept by Algorithm 1.1 if both its endpoints are not red, which happens with probability $p$. Also, conditionally on the choice of non-red half-edges, the pairing between these half-edges is a uniform perfect matching.

Algorithm 1.1 indeed produces $CM_n(\mathbf{d}, p)$ using Fountoulakis' result [85] mentioned above. Due to the uniform matching, it does not matter whether we delete the red vertices, or $n_+$ degree-one vertices chosen uniformly at random. We end up with a sample from the same random graph distribution.

In what follows, directly setting up a technically tractable framework with exploration processes on $CM_n(\mathbf{d}, p_c(\lambda))$ turns out to be difficult even for simple d-regular graphs [135]. On the other hand, due to the sequential construction, the configuration model is often easier to handle. The above construction allows us to study $CM_n(\mathbf{d}, p_c(\lambda))$ via a suitable configuration model.

### 1.4.2    Finite third-moment case

This section is based on [70], where the asymptotics of $\mathbf{Z}_n(\lambda)$, under the finite third-moment assumption, has been treated. To ensure that $CM_n(\mathbf{d})$ has a giant component (otherwise there will be no phase transition for the percolation process), we must assume that (1.2) holds. In this case, $\eta = 1/3$, and $p_c = 1/v_n$, so that

$$p_c(\lambda) = \frac{1}{v_n}(1 + \lambda n^{-1/3}), \quad -\infty < \lambda < \infty.$$

Firstly, let us state the assumptions on the degree distribution, which includes the empirical degree distribution to obey a power law with exponent $\tau > 4$ as a special case.

**Assumption 1.1.** For each $n \geqslant 1$, let $\mathbf{d} = \mathbf{d}_n = (d_i)_{i \in [n]}$ be a degree sequence such that $\ell_n = \sum_{i \in [n]} d_i$ is even. We assume the following about $(\mathbf{d}_n)_{n \geqslant 1}$ as $n \to \infty$: Let $D_n$ denote the degree of a vertex chosen uniformly at random independently of the graph. Then,

(i) (*Weak convergence of* $D_n$) $D_n \xrightarrow{d} D$, for some random variable D such that $\mathbb{E}[D^3] < \infty$.

(ii) (*Uniform integrability of* $D_n^3$) $\mathbb{E}[D_n^3] = \frac{1}{n} \sum_{i \in [n]} d_i^3 \to \mathbb{E}[D^3]$.

We will use Algorithm 1.1 to reduce the analysis of $CM_n(\mathbf{d}, p_c(\lambda))$ to $CM_{\tilde{n}}(\tilde{\mathbf{d}})$. In fact, the following holds for $\tilde{\mathbf{d}}$ for $p = p_c(\lambda)$: Let $\mathbb{P}_p^n$ denote the probability measure induced on $\mathbb{N}^\infty$ by Algorithm 1.1 (S1). Denote the product measure of $(\mathbb{P}_p^n)_{n \geqslant 1}$ by $\mathbb{P}_p$.

**Lemma 1.4.1.** *The statements below are true* $\mathbb{P}_p$ *almost surely: Assumption 1.1 is satisfied by* $\tilde{\mathbf{d}}$ *and*

$$\frac{\sum_{i\in[\tilde{n}]} \tilde{d}_i(\tilde{d}_i - 1)}{\sum_{i\in[\tilde{n}]} \tilde{d}_i} = 1 + \lambda n^{-1/3} + o(n^{-1/3}).$$

The proof involves computing functionals of binomial distributions and their concentration, see Section 2.6 for a proof. Further, while performing (S3), the number of surplus edges within each component does not change, while the component size changes by the amount of deleted degree-one vertices. The latter can be estimated from the number of vertices of degree-one in each of the connected components. Thus without loss of generality, our study reduces to finding the scaling limit of $\mathbf{Z}_n(\lambda)$ on $CM_n(\mathbf{d})$ satisfying

$$\nu_n := \frac{\sum_{i\in[n]} d_i(d_i - 1)}{\sum_{i\in[n]} d_i} = 1 + \lambda n^{-1/3} + o(n^{-1/3}), \quad \text{for some } \lambda \in \mathbb{R}. \quad (1.9)$$

Another technical assumption that we make is that $\mathbb{P}(D = 1) > 0$, which is required for the phase transition result in [114, 133], as well as in some technical parts of our proof. Formally, we aim to prove the following theorem:

**Theorem 1.4.2.** *Consider* $CM_n(\mathbf{d})$ *satisfying Assumption 1.1, and* (1.9) *for some* $\lambda \in \mathbb{R}$. *As* $n \to \infty$,

$$\mathbf{Z}_n(\lambda) \xrightarrow{d} \mathbf{Z}(\lambda)$$

*with respect to the* $\mathbb{U}^0_{\downarrow}$ *topology. Here* $\mathbf{Z}(\lambda)$ *is some non-degenerate random vector which will be defined in the proof ideas and more formally in Theorem 2.2.2.*

In the subsequent sections, we describe the proof idea for Theorem 1.4.2.

**The exploration process**

The central idea to prove scaling limits of critical component sizes was introduced by Aldous [8] in the context of the Erdős-Rényi random graph. The idea is to explore the graph sequentially and encode the relevant information in terms of a walk called the *exploration process*. Then the idea is to establish scaling limits of the exploration process and then try to read off, if possible, the relevant property from the limit of the exploration process. Let us explore $CM_n(\mathbf{d})$ sequentially using depth-first exploration. At each step $k$, we find a new vertex with degree $d_{(k)}$. This vertex may create $c_{(k)}$ edges to the vertices which are already explored. Thus $d_{(k)} - c_{(k)}$ half-edges can give new vertices during the exploration. Once all the half-edges of a vertex are explored,
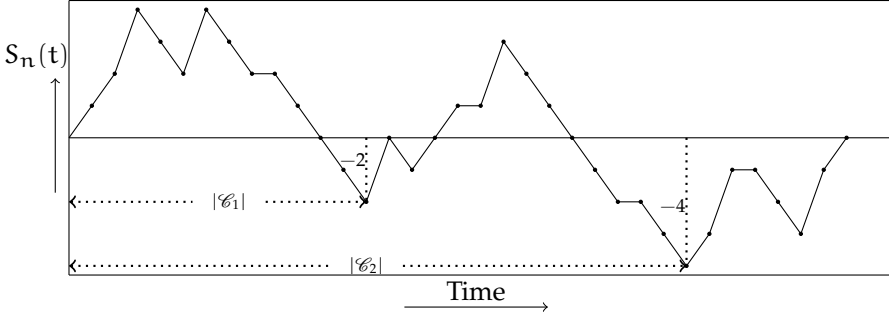
Figure 1.1: Component sizes as excursions of the exploration process

the vertex is declared dead, meaning that the complete neighborhood of that vertex has been identified. The precise description of the exploration algorithm is given in Section 2.4. Based on this exploration algorithm, define the exploration process by

$$S_n(0) = 0, \quad S_n(i) = \sum_{j=1}^{i} (d_{(j)} - 2 - 2c_{(j)}). \tag{1.10}$$

The minus two is due to the fact that an edge (i.e. two half-edges) is explored at each step. The process $\mathbf{S}_n = (S_n(i))_{i \in [n]}$ "encodes the component sizes as lengths of path segments above past minima" as discussed in [8]. Suppose $\mathscr{C}_i$ is the $i$-th connected component explored by the above exploration process. Define

$$\tau_k = \inf \{i : S_n(i) = -2k\}.$$

Then $\mathscr{C}_k$ is discovered between the times $\tau_{k-1} + 1$ and $\tau_k$ and $|\mathscr{C}_k| = \tau_k - \tau_{k-1}$. Therefore, the excursion lengths of the exploration process correspond to the sizes of the explored components, see Figure 1.1. This property allows one to recover the scaling limits of the component sizes from the scaling limit of the exploration process.

**Size-biased exploration**

During the above mentioned exploration process, the vertices are explored in a size-biased manner with sizes proportional to their degrees, i.e., if we denote by $v_{(i)}$ the $i$-th explored vertex, then

$$\mathbb{P}(v_{(i)} = j | v_{(1)}, v_{(2)}, ..., v_{(i-1)}) = \frac{d_j}{\sum_{k \notin \mathscr{V}_{i-1}} d_k}, \quad \forall j \in \mathscr{V}_{i-1},$$

where $\mathcal{V}_i$ denotes the first $i$ vertices to be discovered in the above exploration process. The following lemma is a consequence of the size-biased ordering, which keeps track of sum of the degrees and square of degrees of the explored vertices. This will allow us to track the drift and the quadratic variation of the exploration process (1.10):

**Lemma 1.4.3.** *Suppose that Assumption 1.1 holds and denote $\sigma_r = \mathbb{E}[D^r]$. Then for all $t > 0$, as $n \to \infty$,*

$$\sup_{u \leqslant t} \left| n^{-2/3} \sum_{i=1}^{\lfloor n^{2/3} u \rfloor} d_{(i)} - \frac{\sigma_2 u}{\sigma_1} \right| \xrightarrow{\mathbb{P}} 0, \quad \sup_{u \leqslant t} \left| n^{-2/3} \sum_{i=1}^{\lfloor n^{2/3} u \rfloor} d_{(i)}^2 - \frac{\sigma_3 u}{\sigma_1} \right| \xrightarrow{\mathbb{P}} 0.$$

$$(1.11)$$

**Analysis of the exploration process**

The next step is to obtain the scaling limit of the exploration process. Firstly, let us consider the simplified process

$$s_n(0) = 0, \quad s_n(i) = \sum_{j=1}^{i} (d_{(j)} - 2),$$

which ignores the effect of cycles in the exploration process. Due to the close relation to the size-biased exploration, we can more easily describe the scaling limit of $s_n(t)$:

**Proposition 1.4.4.** *Let $\bar{s}_n = (s_n(t))_{t \geqslant 0}$ be given by $\bar{s}_n(t) = n^{-1/3} s_n(\lfloor t n^{2/3} \rfloor)$. Under Assumption 1.1, as $n \to \infty$,*

$$\bar{s}_n \xrightarrow{\mathcal{L}} B \tag{1.12}$$

*with respect to the Skorohod $J_1$ topology, where $B(t) = c_0 W(t) + \lambda t - c_1 t^2$ with $W$ a standard Brownian motion.*

The contribution $\sum_{j=1}^{i} c_{(j)}$ counts the number of surplus edges created upto time $i$, and one can expect it to be negligible from earlier heuristics about small number of surplus edges. Thus Proposition 1.4.4 also provides the scaling limit of $\mathbf{S}_n$, after corresponding rescaling. The time scaling $n^{2/3}$ is due to our prior prediction that the component sizes are of this order. The space scaling $n^{1/3}$ is the usual square root fluctuation with respect to the time component that arises for Brownian scaling limits.

Let us now give some details as to how Proposition 1.4.4 can be proved. Let $\mathcal{F}_j$ denote the natural sigma algebra which contains all the information

about the explored graph up to time j. The Doob-Meyer decomposition yields that

$$s_n(i) = M_n(i) + A_n(i),$$

where

$$A_n(i) = \sum_{j=1}^{i} \mathbb{E}\big[d_{(j)} - 2|\mathscr{F}_{j-1}\big], \quad B_n(i) = \sum_{j=1}^{i} \mathrm{Var}\big(d_{(j)}|\mathscr{F}_{j-1}\big), \qquad (1.13)$$

$(B_n(i))_{i \geqslant 1}$ being the quadratic variation process of $(M_n(i))_{i \geqslant 1}$. We consider the convergence of the drift part $A_n$ and the martingale part $M_n$ separately.

**Convergence of the drift part.**    The negative quadratic drift in the limit of $\bar{s}_n$ is a consequence of the size-biased reordering stated in Lemma 1.4.3. Note that

$$\mathbb{E}\big[d_{(i)} - 2|\mathscr{F}_{i-1}\big] = \frac{\sum_{j \notin \mathscr{V}_{i-1}} d_j(d_j - 2)}{\sum_{j \notin \mathscr{V}_{i-1}} d_j}$$

$$= \frac{\sum_{j \in [n]} d_j(d_j - 2)}{\sum_{j \in [n]} d_j} - \frac{\sum_{j \in \mathscr{V}_{i-1}} d_j(d_j - 2)}{\sum_{j \in [n]} d_j} + \frac{\sum_{j \notin \mathscr{V}_{i-1}} d_j(d_j - 2)\sum_{j \in \mathscr{V}_{i-1}} d_j}{\sum_{j \notin \mathscr{V}_{i-1}} d_j \sum_{j \in [n]} d_j}$$

$$= \frac{\lambda}{n^{1/3}} - \frac{\sum_{j \in \mathscr{V}_{i-1}} d_j^2}{\sum_{j \in [n]} d_j} + \frac{\sum_{j \notin \mathscr{V}_{i-1}} d_j^2 \sum_{j \in \mathscr{V}_{i-1}} d_j}{\sum_{j \notin \mathscr{V}_{i-1}} d_j \sum_{j \in [n]} d_j} + o(n^{-1/3}).$$

Therefore,

$$A_n(k) = \sum_{i=1}^{k} \mathbb{E}\big[d_{(i)} - 2|\mathscr{F}_{i-1}\big]$$

$$= \frac{k\lambda}{n^{1/3}} - \sum_{i=1}^{k} \frac{\sum_{j \in \mathscr{V}_{i-1}} d_j^2}{\sum_{j \in [n]} d_j} + \sum_{i=1}^{k} \frac{\sum_{j \notin \mathscr{V}_{i-1}} d_j^2 \sum_{j \in \mathscr{V}_{i-1}} d_j}{\sum_{j \notin \mathscr{V}_{i-1}} d_j \sum_{j \in [n]} d_j} + o(kn^{-1/3}).$$

Now, $\sum_{j \in \mathscr{V}_{i-1}} d_j = o(\ell_n)$ uniformly over $i \leqslant tn^{2/3}$, since $\max_{i \in [n]} d_i = o(n^{1/3})$, by the uniform integrability of the third moment in Assumption 1.1. Therefore $\sum_{j \notin \mathscr{V}_{i-1}} d_j \approx \ell_n$, and a similar argument yields that $\sum_{j \notin \mathscr{V}_{i-1}} d_j^2 \approx \sum_{j \in [n]} d_j^2$. Combining this with (1.11), it follows that

$$n^{-1/3}A_n(\lfloor un^{2/3} \rfloor) \approx \lambda u - c_1 u^2, \qquad (1.14)$$

where $c_1 = (\sigma_3 \sigma_1 - \sigma_2^2)/\sigma_1^3$. Notice that $c_1 = \mathrm{Var}(D^*)/\mathbb{E}[D] > 0$, where $D^*$ is the size-biased version of the random variable $D$ appearing in Assumption 1.1. Thus, the drift term is negative and parabolic. In the above calculations, we see that the negative drift term arises from the depletion of degrees

in the size-biased exploration. As more vertices are explored, $\mathbb{E}[d_{(i)}|\mathscr{F}_{i-1}]$ decreases by an amount proportional to $\sum_{j\in\mathscr{V}_{i-1}} d_j^2$. Due to the finite third moment condition, Lemma 1.4.3 ensures that $\sum_{j\in\mathscr{V}_{i-1}} d_j^2$ increases linearly with time. Thus the negative part in the drift term, which is just the sum of $\sum_{j\in\mathscr{V}_{i-1}} d_j^2$, is quadratic. In this sense, the negative quadratic drift is related to the effect of depletion of degrees in sampling from the size-biased distribution without replacement.

**Convergence of the martingale part.** The proof relies on the celebrated Martingale Functional Central Limit theorem (FCLT). The Martingale FCLT ensures convergence of martingales to Brownian motion provided that the limiting process has continuous sample paths, and the quadratic variation converges to a constant multiple of t. The latter condition arises due to Lévy's characterization of Brownian motion as the unique process with quadratic variation t. In this case, it is enough to show that

$$n^{-2/3}B_n(\lfloor un^{2/3}\rfloor) \xrightarrow{\mathbb{P}} c_0^{1/2}u.$$

Again this can be deduced using Lemma 1.4.3. The increments of $B_n(\lfloor un^{2/3}\rfloor)$ in (1.13) are given by the asymptotic finite variance of the size-biased distribution, which is equivalent to the finite third moment of the degree distribution. Thus, the finite third moment is essential from the point of view of the functional invariance principle. The technical conditions for ensuring that the limiting process has continuous sample paths are explicitly stated in Section 2.4.

**Large components are explored early**

To learn about the largest component sizes from Proposition 1.4.4, one first needs to check that the ordered vector of excursion lengths is a continuous function on a set $\mathcal{G} \subset \mathbb{D}(\mathbb{R}_+, \mathbb{R})$, and the limiting process in (1.12) lies in $\mathcal{G}$ almost surely. This part of the argument follows using properties of Brownian motion with a negative parabolic drift, see [8, 11, 136]. In order to ensure that the largest excursions of **B** in (1.12) correspond to the largest components in the critical random graph, it must be ensured that the largest components are explored in $O(n^{2/3})$ time. This is because, due to the time scaling by $n^{2/3}$, we loose information about the components explored in $\Omega(n^{2/3})$ time. The following lemma ensures that no large component is explored after time $\Omega(n^{2/3})$:

**Lemma 1.4.5.** *Let $\mathscr{C}_{\max}^{\geqslant T}$ denote the largest component which is started to be explored after time $Tn^{2/3}$. Then, for any $\delta > 0$,*

$$\lim_{T\to\infty} \limsup_{n\to\infty} \mathbb{P}\left(|\mathscr{C}_{\max}^{\geqslant T}| > \delta n^{2/3}\right) = 0. \tag{1.15}$$

The idea for the proof of Lemma 1.4.5 is that due to the sequential matching of the half-edges after exploring the graph upto time $Tn^{2/3}$, the rest of the graph is again a configuration model. Moreover, the $\nu_n$ parameter for this new configuration model becomes

$$\nu_{n,Tn^{2/3}} = \nu_n - C_0 Tn^{-1/3} + o_{\mathbb{P}}(n^{-1/3}). \tag{1.16}$$

Thus, as we keep on exploring the graph, the rest of the graph becomes a configuration model that is more and more subcritical. Now the fact that the component sizes of a barely subcritical configuration model are $o(n^{2/3})$ can be leveraged. However, the iterated limit in (1.15) requires explicit bounds on the required functionals of a "slightly subcritical" configuration model. For a formal deduction, see Lemma 2.4.11.

**Component sizes and surplus edges in the product topology**

Let us now investigate how the exploration process can yield convergence of the surplus edges. At step $k+1$, we have discovered vertex $v_{(k+1)}$ with degree $d_{(k+1)}$, and since one half-edge has been used to discover $v_{(k+1)}$, $d_{(k+1)} - 1$ half-edges can create surplus edges. There are $A_k = S_n(k) - \min_{j\leqslant k} S_n(j)$ many half-edges associated to the vertices that are discovered, but not yet explored completely. Due to the uniform matching, $c_{(k+1)}$, defined in (1.10) satisfies

$$\mathbb{E}[c_{(k+1)}|\mathscr{F}_k, v_{(k+1)}] \approx \frac{(d_{(k+1)} - 1)A_k}{\ell_n}.$$

Now,

$$\mathbb{E}[d_{(k+1)} - 1|\mathscr{F}_k] = \frac{\sum_{j\notin \mathscr{V}_k} d_j(d_j - 1)}{\sum_{j\notin \mathscr{V}_k} d_j} \approx \frac{\sum_{j\in[n]} d_j(d_j - 1)}{\sum_{j\in[n]} d_j} + o_{\mathbb{P}}(1) \approx 1,$$

so that

$$\sum_{k\leqslant tn^{2/3}} \mathbb{E}[c_{(k+1)}|\mathscr{F}_k] \approx \sum_{k\leqslant tn^{2/3}} \frac{A_k}{\sigma_1 n} \approx \frac{1}{\sigma_1}\int_0^t \bar{A}_n(u)du, \tag{1.17}$$

where $\bar{A}_n(u) = n^{-1/3}A_n(\lfloor un^{2/3}\rfloor)$. Here $(\bar{A}_n(u))_{u\geqslant 0}$ converges to $(R(u))_{u\geqslant 0}$, where $R(u) = B(u) - \inf_{s\leqslant u} B(s)$ is the reflected version of the limit in (1.12). This proves the following lemma:

**Lemma 1.4.6.** *Let* $N_n(k)$ *be the number of surplus edges discovered up to time* $k$ *and* $\bar{N}_n(u) = N_n(\lfloor un^{2/3} \rfloor)$. *Then, as* $n \to \infty$, $\bar{N}_n \xrightarrow{d} N$, *where* $N$ *is the unique counting process such that the following is a martingale:*

$$N(t) - \frac{1}{\sigma_1} \int_0^t R(u)du.$$

Thus, the number of points $N_i$ in the $i$-th largest excursion $\gamma_i$ of $B$ is distributed as a mixed Poisson random variable with parameter $(\sigma_1)^{-1} \int_\gamma R(u)du$ (see Chapter 2 for a formal definition of excursions). At this moment, Proposition 1.4.4 and Lemma 1.4.6 yield the convergence of the component sizes and surplus edges that are explored before time $O(n^{2/3})$. The scaling limits of the component sizes are the largest excursions of $B$ in Proposition 1.4.4, and those of the surplus edges are given by mixed Poisson random variables with parameters being proportional to the areas under those excursions as given by Lemma 1.4.6. On the other hand, Lemma 1.4.5 ensures that the largest components are explored in time $O(n^{2/3})$ during the exploration process. This implies the finite-dimensional convergence of $Z_n(\lambda)$:

**Theorem 1.4.7.** *Consider* $CM_n(d)$ *satisfying Assumption 1.1 and (1.9) for some* $\lambda \in \mathbb{R}$. *As* $n \to \infty$,

$$Z_n(\lambda) \xrightarrow{d} Z(\lambda)$$

*with respect to the product topology, where* $Z(\lambda)$ *is the vector* $(\gamma_i, N_i)_{i \geqslant 1}$ *ordered as an element of* $\mathbb{U}_\downarrow^0$.

### Convergence in the $\mathbb{U}_\downarrow^0$ topology

In order to complete the proof of Theorem 1.4.2, it is now sufficient to show that $(Z_n(\lambda))_{n \geqslant 1}$ is tight in $\mathbb{U}_\downarrow^0$, owing to the convergence in product topology in Theorem 1.4.7. The tightness is more technical, and the details will be provided in Section 2.4.4. However, let us state here the conditions that we need to verify in order to complete the proof. Let $\mathscr{C}_i$ denote the $i$-th explored component, and $Y_i^n = n^{-2/3}|\mathscr{C}_i|$, $N_i^n = SP(\mathscr{C}_i)$. It is sufficient for the tightness of probability measures on $\mathbb{U}_\downarrow^0$ to prove that for any $\delta > 0$

$$\lim_{\varepsilon \to 0} \limsup_{n \to \infty} \mathbb{P}\left( \sum_{Y_i^n \leqslant \varepsilon} (Y_i^n)^2 > \delta \right) = 0,$$

$$\lim_{\varepsilon \to 0} \limsup_{n \to \infty} \mathbb{P}\left( \sum_{Y_i^n \leqslant \varepsilon} Y_i^n N_i^n > \delta \right) = 0.$$

(1.18)

**Degree distribution within components**

Define $v_k(G)$ as the number of vertices of degree $k$ in the connected graph $G$. Then,

$$v_k(\mathscr{C}_{(j)}) = \frac{kr_k}{\mathbb{E}[D]}|\mathscr{C}_{(j)}| + O_\mathbb{P}\big((k^{-1}n^{1/3})\big). \tag{1.19}$$

Again this can be deduced from the size-biased exploration process. If $N_k(t)$ denotes the number of vertices of degree $k$ discovered up to time $t$, then for any $t > 0$, uniformly over $k$,

$$\sup_{u \leqslant t}\Big|n^{-2/3}N_k(un^{2/3}) - \frac{kn_k}{\ell_n}u\Big| = O_\mathbb{P}((kn^{1/3})^{-1}).$$

This is due to the fact that, at each step during the exploration, we discover a vertex of degree $k$ with probability roughly $kn_k/\ell_n$. Obviously, there will be depletion in the total number of half-edges and the total number of half-edges attached to vertices of degree $k$, but that depletion does not matter in the $n^{2/3}$ scale. Now an application of Lemma 1.4.5 yields (1.19).

The above analysis provides a detailed picture of the size and complexity of the critical components for percolation on $CM_n(\mathbf{d})$. Whenever the degree distribution satisfies an asymptotic finite third-moment condition, the scaling limit lies in the same universality class as for the Erdős-Rényi random graph identified in [8]. For Erdős-Rényi random graphs the negative drift term takes a simpler form as the size-biased version of a Poisson random variable again has a Poisson distribution.

### 1.4.3 Infinite third-moment case

We now continue with the case where $\mathbb{E}[D^3] = \infty$. Since $\mathbb{E}[D^3]$ appears explicitly in the scaling limit of the exploration process in Section 1.4.2 (see e.g. (1.14)), the scaling limit must be different in this case. This section is based on the results for $\mathbf{Z}_n(\lambda)$ from [69]. Throughout this section we will use the notation

$$\alpha = 1/(\tau-1), \qquad \rho = (\tau-2)/(\tau-1), \qquad \eta = (\tau-3)/(\tau-1),$$
$$a_n = n^\alpha L(n), \qquad b_n = n^\rho(L(n))^{-1}, \qquad c_n = n^\eta(L(n))^{-2},$$

where $\tau \in (3,4)$ and $L(\cdot)$ is a slowly-varying function. The results for $\mathbf{Z}_n(\lambda)$ are derived under the following assumptions on the degree sequence:

**Assumption 1.2.** Fix $\tau \in (3,4)$. Let $\mathbf{d} = (d_1, \dots, d_n)$ be a degree sequence (ordered in a non-increasing manner) such that the following conditions hold:

(i) (*High-degree vertices*) For any fixed $i \geqslant 1$, $d_i/a_n \to \theta_i$, where $\theta = (\theta_1, \theta_2, \dots) \in \ell_{\downarrow}^3 \setminus \ell_{\downarrow}^2$.

(ii) (*Moment assumptions*) Let $D_n$ denote the degree of a vertex chosen uniformly at random from $[n]$, independently of $CM_n(\mathbf{d})$. Then, $D_n \xrightarrow{d} D$, for some integer-valued random variable $D$ and

$$\frac{1}{n} \sum_{i \in [n]} d_i \to \mu := \mathbb{E}[D], \quad \frac{1}{n} \sum_{i \in [n]} d_i^2 \to \mathbb{E}[D^2],$$

$$\lim_{K \to \infty} \limsup_{n \to \infty} a_n^{-3} \sum_{i=K+1}^{n} d_i^3 = 0.$$

(iii) (*Critical window*) For some $\lambda \in \mathbb{R}$,

$$\nu_n(\lambda) := \frac{\sum_{i \in [n]} d_i(d_i - 1)}{\sum_{i \in [n]} d_i} = 1 + \lambda c_n^{-1} + o(c_n^{-1}).$$

(iv) Let $n_1$ be the number of vertices of degree-one. Then $n_1 = \Theta(n)$, which is equivalent to assuming that $\mathbb{P}(D = 1) > 0$.

Assumption 1.2 can be understood intuitively. As in Section 1.4.2, we will set up an exploration process, which explores the components of $CM_n(\mathbf{d})$ in a size-biased manner. In this setting, we will see that the exploration process keeps on exploring vertices of high degree, resulting in jumps in the exploration process. Assumption 1.2 (i) is used to control the magnitude of these jumps. The scaling $a_n$ has the same order as $\max_{i \in [n]} D_i$, where $D_i$'s are i.i.d. random variables satisfying $\mathbb{P}(D_1 \geqslant x) \propto L_0(x) x^{-(\tau-1)}$ for some slowly-varying function $L_0(\cdot)$. The expectation and variance of the increments of the exploration process are governed by the moment assumptions in Assumption 1.2. Of particular interest is the assumption on the third moment, which basically says that the variance of the increments is dictated by the contributions from the high-degree vertices only. The condition in Assumption 1.2 (iii) is the same criticality condition as in (1.9). The fact that the above set-up covers $CM_n(\mathbf{d}, p_c(\lambda))$ can be established using an analogue of Lemma 1.4.1 in this setting. A key thing to note here is that if the degrees are an iid sample from a distribution $D$ with $\mathbb{P}(D \geqslant x) \propto L_0(x) x^{-(\tau-1)}$, for some $\tau \in (3, 4)$ and $L_0(\cdot)$ a slowly-varying function, then Assumption 1.2 is satisfied; see Section 3.3.

Recall that $\mathbf{Z}_n(\lambda)$ denotes the vector of rescaled component sizes and surplus edges, ordered as an element of $\mathbb{U}_{\downarrow}^0$. In this section, we rescale the

component sizes by $b_n$. The following theorem describes the scaling limit of $Z_n(\lambda)$ in the infinite third moment case:

**Theorem 1.4.8.** *Consider* $CM_n(\mathbf{d})$ *satisfying Assumption 1.2. As* $n \to \infty$, $Z_n(\lambda) \xrightarrow{d} Z(\lambda)$ *with respect to the* $\mathbb{U}_{\downarrow}^0$ *topology, where* $Z(\lambda)$ *is some non-degenerate random vector which will be defined in the proof ideas and more formally in Theorem 3.2.2.*

The proof of Theorem 1.4.8 can be approached by the steps outlined in Section 1.4.2. However, the techniques involved are substantially different, because, for example, the exploration process does not have a finite variance of the increment distribution. Below, we outline the analysis of the exploration process, and the necessary modifications to conclude that the largest components are explored in time $O(b_n)$. The asymptotics for the surplus edges follow identically, since (1.17) holds here as well, the only difference arises due to different scaling limit of the exploration process. We also discuss the scaling limit when the underlying graph is $UM_n(\mathbf{d})$, i.e., when the configuration model is conditioned to be simple. This problem was stated as a conjecture in [121] when the degrees are an iid sample from a power-law distribution with $\tau \in (3, 4)$.

**The size-biased exploration process**

For technical tractability, we modify the exploration process. We sequentially take active half-edges, pair them uniformly with an unpaired half-edge. If the new half-edge is incident to a new vertex, then we declare all of its half-edges to be active. The paired half-edges are killed. If there are no active half-edges in the system, then we choose one unexplored vertex with probability proportional to its degree and declare all its half-edges active. See Section 3.5 for an exact description. The only difference with the exploration process in Section 1.4.2 is that only one edge is created per step, and it is not necessary that new vertices are found at each step. Let $\mathscr{V}_l$ denote the set of vertices discovered up to time $l$ and $\mathcal{I}_i^n(l) := \mathbf{1}\{i \in \mathscr{V}_l\}$. The exploration process is given by $S_n(0) = 0$, and

$$S_n(l) = \#\text{active half-edges at } l - 2 \times \#\text{components explored upto } l$$
$$= \sum_{i \in [n]} d_i \mathcal{I}_i^n(l) - 2l. \tag{1.20}$$

Suppose that $\mathscr{C}_k$ is the $k^{\text{th}}$ connected component explored by the above exploration process and define $\tau_k = \inf\{i : S_n(i) = -2k\}$. Then $\mathscr{C}_k$ is discov-

ered between the times $\tau_{k-1} + 1$ and $\tau_k$, and $\tau_k - \tau_{k-1} - 1$ gives the total number of edges in $\mathscr{C}_k$. However, since the surplus edges will be shown to be tight, the number of edges and the component sizes are asymptotically the same, after rescaling by $b_n$. Note that we can write

$$S_n(l) = \sum_{i \in [n]} d_i \mathcal{I}_i^n(l) - 2l = \sum_{i \in [n]} d_i \left( \mathcal{I}_i^n(l) - \frac{d_i}{\ell_n} l \right) + (\nu_n(\lambda) - 1) l.$$

Define the re-scaled version $\bar{\mathbf{S}}_n$ of $\mathbf{S}_n$ by $\bar{S}_n(t) = a_n^{-1} S_n(\lfloor b_n t \rfloor)$. Then, by Assumption 1.2,

$$\bar{S}_n(t) = a_n^{-1} \sum_{i \in [n]} d_i \left( \mathcal{I}_i^n(tb_n) - \frac{d_i}{\ell_n} tb_n \right) + \lambda t + o(1). \qquad (1.21)$$

**Analysis of the exploration process**

The exploration process given by (1.21) has the following scaling limit:

**Theorem 1.4.9.** *As $n \to \infty$, $\bar{\mathbf{S}}_n \xrightarrow{d} \bar{\mathbf{S}}_\infty$ with respect to the Skorohod $J_1$ topology. The limit $\bar{\mathbf{S}}_\infty := (\bar{S}_\infty(t))_{t \geqslant 0}$ is given by*

$$\bar{S}_\infty(t) = \sum_{i=1}^{\infty} \theta_i \left( \mathcal{I}_i(t) - (\theta_i/\mu)t \right) + \lambda t, \qquad (1.22)$$

*where $\mathcal{I}_i(s) := \mathbf{1}\{\xi_i \leqslant s\}$ for $\xi_i \sim \mathrm{Exp}(\theta_i/\mu)$ independently, and $\mathrm{Exp}(r)$ denotes the exponential distribution with rate $r$.*

The limit (1.22) is a jump-process. The vertices of degree $\Theta(a_n)$ keep getting explored with time $O(b_n)$, and since the space has been rescaled by $a_n$, these create macroscopic jumps in the exploration process. Notice that $\bar{\mathbf{S}}_\infty$ does not have independent increments and therefore it is not a Lévy process. This was termed as *thinned Lévy process* in [33], since $\mathcal{I}_i(s)$ can be seen as a thinning of a Poisson process, with all points discarded except the first one. Due to the absence of independent increments, most standard techniques from the stochastic-process limits literature do not work here.

However from (1.22), it is evident that the scaling limit only depends on the asymptotics of the high-degree vertices given in Assumption 1.2 (i), and the contributions coming from the lower-degree vertices should be asymptotically negligible. With this in mind, define the truncated sum

$$M_n^K(l) = a_n^{-1} \sum_{i > K} d_i \left( \mathcal{I}_i^n(l) - \frac{d_i}{\ell_n} l \right). \qquad (1.23)$$

Recall that we have ordered the degree sequence in Assumption 1.2, so that the sum in (1.23) takes into account all the contributions in the exploration process except for the K largest degrees. With a proper estimate of the variance and the expectation of $M_n^K(l)$, along with maximal inequalities for supermartingales, it can be shown that (see Section 3.5) for any $\varepsilon > 0$ and $T > 0$,

$$\lim_{K \to \infty} \limsup_{n \to \infty} \mathbb{P}\Big( \sup_{l \leqslant Tb_n} |M_n^K(l)| > \varepsilon \Big) = 0.$$

This implies that if we truncate the sum in (1.21) at any fixed K, then it suffices to establish the iterated limit as $n \to \infty$ and then $K \to \infty$. Finally, to complete the proof of Theorem 1.4.9, it suffices to consider the joint distributional limit of the processes $(\mathcal{I}_i^n(tb_n))_{i \in [K], t \geqslant 0}$, since for any fixed $K \geqslant 1$,

$$a_n^{-1} \sum_{i=1}^K d_i \left( \mathcal{I}_i^n(tb_n) - \frac{d_i}{\ell_n} tb_n \right) = \sum_{i=1}^K \theta_i \left( \mathcal{I}_i^n(tb_n) - \frac{d_i}{\ell_n} tb_n \right) + o(1).$$

The following lemma characterizes the limit of $(\mathcal{I}_i^n(tb_n))_{i \in [K], t \geqslant 0}$:

**Lemma 1.4.10.** *Fix any $K \geqslant 1$. As $n \to \infty$,*

$$(\mathcal{I}_i^n(tb_n))_{i \in [K], t \geqslant 0} \xrightarrow{d} (\mathcal{I}_i(t))_{i \in [K], t \geqslant 0}.$$

*Proof.* By noting that $(\mathcal{I}_i^n(tb_n))_{t \geqslant 0}$ are indicator processes, it is enough to show that

$$\mathbb{P}\left( \mathcal{I}_i^n(t_i b_n) = 0, \ \forall i \in [K] \right) \to \mathbb{P}\left( \mathcal{I}_i(t_i) = 0, \ \forall i \in [K] \right) = \exp\Big( -\mu^{-1} \sum_{i=1}^K \theta_i t_i \Big)$$

for any $t_1, \ldots, t_K \in \mathbb{R}$. Now,

$$\mathbb{P}\left( \mathcal{I}_i^n(m_i) = 0, \ \forall i \in [K] \right) = \prod_{l=1}^\infty \Big( 1 - \sum_{i \leqslant K: l \leqslant m_i} \frac{d_i}{\ell_n - \Theta(l)} \Big),$$

$$= \exp\Big( -\sum_{l=1}^\infty \sum_{i \leqslant K: l \leqslant m_i} \frac{d_i}{\ell_n} + o(1) \Big) = \exp\Big( -\sum_{i \in [K]} \frac{d_i m_i}{\ell_n} + o(1) \Big).$$

Putting $m_i = t_i b_n$, Assumption 1.2 gives

$$\frac{m_i d_i}{\ell_n} = \frac{\theta_i t_i}{\mu}(1 + o(1)),$$

which completes the proof of Lemma 1.4.10.    □

**Convergence of component sizes and surplus edges**

As in the finite third moment case, one must ensure that the largest components are explored early during the exploration process, i.e., we need Lemma 1.4.5 to hold under Assumption 1.2 after replacing $n^{2/3}$ by $b_n$. One may try to adapt the argument of Lemma 1.4.5, but there is a more direct and simpler approach. The idea is that since the critical behavior is primarily governed by the asymptotics of the high-degree vertices, removing the vertices of high degree makes the graph more and more subcritical, a feature that is not present in the finite third moment case. This idea can be leveraged to obtain the proof that large connected components are with high probability explored in time $O(b_n)$, as well as showing the $\ell_\downarrow^2$ tightness for the vector of component sizes; see Proposition 3.5.1.

Let $\mathcal{G}^{[K]}$ be the random graph obtained by removing all edges attached to vertices $1, \ldots, K$ and let $\mathbf{d}'$ be the obtained degree sequence. Now, conditionally on the set of removed half-edges, $\mathcal{G}^{[K]}$ is still a configuration model with some degree sequence $\mathbf{d}'$ with $d_i' \leqslant d_i$ for all $i \in [n] \setminus [K]$ and $d_i' = 0$ for $i \in [K]$. Further, the criticality parameter of $\mathcal{G}^{[K]}$ satisfies

$$
\begin{aligned}
\nu_n^{[K]} &= \frac{\sum_{i \in [n]} d_i'(d_i' - 1)}{\sum_{i \in [n]} d_i'} \leqslant \frac{\sum_{i \in [n]} d_i(d_i - 1) - \sum_{i=1}^{K} d_i(d_i - 1)}{\ell_n - 2 \sum_{i=1}^{K} d_i} \\
&= \nu_n - C_1 n^{2\alpha - 1} L(n)^2 \sum_{i \leqslant K} \theta_i^2 = \nu_n - C_1 c_n^{-1} \sum_{i \leqslant K} \theta_i^2
\end{aligned}
\tag{1.24}
$$

for some constant $C_1 > 0$. Since $\theta \notin \ell_\downarrow^2$, $K$ can be chosen large enough such that $\nu_n^{[K]}$ becomes arbitrarily small uniformly for all $n$. (1.24) plays the same role in the infinite third-moment case as (1.16) in the finite third-moment case. We refer the reader to Lemma 3.5.5 for the exact details.

In Section 1.4.2, we have expressed the limiting number of surplus edges in Lemma 1.4.6 in terms of the reflected version of the scaling limit of the exploration process. This deduction holds in the infinite third moment case as well, but we have to replace $R(t)$ by the reflected version of $\mathbf{S}_\infty$. Thus, the finite-dimensional convergence of $\mathbf{Z}_n(\lambda)$ follows. The tightness argument involves establishing (1.18). Finally, an analogue of (1.19) also holds in this case; see Proposition 3.6.2.

**Proof for uniform simple graphs**

All the scaling limit results for component sizes and surplus edges hold for the critical configuration model. Let us now outline the strategy to transfer

those results to the case of $UM_n(\mathbf{d})$, the uniformly chosen simple graph with given degree distribution. Recall that the law of $UM_n(\mathbf{d})$ is the same as $CM_n(\mathbf{d})$, conditionally on the graph being simple. However, since the results about $\mathbf{Z}_n(\lambda)$ are distributional convergence results, it is not evident whether the results are still true conditionally on the graph being simple. Particularly, in the infinite third-moment case, a related question about critical $CM_n(\mathbf{d})$ with iid degree sequence was stated as a conjecture in [121].

**Theorem 1.4.11.** *Conditionally on* $CM_n(\mathbf{d})$ *being simple,* $\mathbf{Z}_n(\lambda) \xrightarrow{d} \mathbf{Z}(\lambda)$*, where* $\mathbf{Z}(\lambda)$ *is the scaling limit for* $CM_n(\mathbf{d})$*.*

We refer the reader to Section 3.7 for the technical details and only explain the idea here. Recall from [110, Theorem 1.1] that

$$\liminf_{n\to\infty} \mathbb{P}\left(CM_n(\mathbf{d}) \text{ is simple}\right) > 0.$$

Thus, the tightness of $\mathbf{Z}_n(\lambda)$ in the $\mathbb{U}_\downarrow^0$ topology follows directly, and we only need to prove that the finite-dimensional convergence remains valid. Note that the graph $UM_n(\mathbf{d})$ can be generated by sampling $CM_n(\mathbf{d})$ until we get a uniform simple graph. Suppose one can show that the exploration process does not encounter any self-loops or multiple edges in time $O(b_n)$ with high probability. Then, we can keep the graph explored up to time $O(b_n)$ and re-sample the uniform perfect matching on the half-edges explored after time $\Omega(b_n)$, until the latter one gives a simple graph. This will give us a sample from $UM_n(\mathbf{d})$ with high probability. However, the large components are explored in time $O(b_n)$, and these components remain fixed under re-sampling for the latter construction. Therefore, the finite-dimensional scaling limit for $\mathbf{Z}_n(\lambda)$ does not change, conditionally on the graph being simple. Let us now argue that no self-loops or multiple edges are explored before time $O(b_n)$.

Let $\ell_n' := \ell_n - 2Tb_n$. Let $v$ be a vertex being *explored* before time $O(b_n)$, and $(e_1, \dots, e_r)$ the set of half-edges of $v$. Note that, while pairing $e_i$, it creates a self-loop with probability at most $(d_v - i)/\ell_n'$ and creates a multiple edge with probability at most $(i - 1)/\ell_n'$. Therefore, conditionally on $\mathscr{F}_{l-1}$, the expected number of self-loops/multiple edges discovered while exploring

the vertex $\nu_l$ at the $l$-th step is at most $2d_{\nu_l}^2/\ell_n'$. Thus, for any $T > 0$,

$$\mathbb{E}\,[\#\{\text{self-loops or multiple edges discovered up to time } Tb_n]$$

$$\leqslant \frac{2}{\ell_n'}\mathbb{E}\Bigg[\sum_{i\in[n]} d_i^2 \mathcal{I}_i^n(Tb_n)\Bigg]$$

$$= \frac{2}{\ell_n'}\mathbb{E}\Bigg[\sum_{i=1}^K d_i^2 \mathcal{I}_i^n(Tb_n)\Bigg] + \frac{2}{\ell_n'}\mathbb{E}\Bigg[\sum_{i=K+1}^n d_i^2 \mathcal{I}_i^n(Tb_n)\Bigg],$$

where $\mathcal{I}_i^n(l) = \mathbf{1}\{i \in \mathcal{V}_l\}$. Now, for every fixed $K \geqslant 1$,

$$\frac{2}{\ell_n'}\mathbb{E}\Bigg[\sum_{i=1}^K d_i^2 \mathcal{I}_i^n(Tb_n)\Bigg] \leqslant \frac{2}{\ell_n'}\sum_{i=1}^K d_i^2 \to 0,$$

since $2\alpha - 1 < 0$. Moreover, the size-biased ordering of the vertices implies that

$$\mathbb{P}\,(\mathcal{I}_i^n(Tb_n) = 1) \leqslant Tb_n d_i/\ell_n'. \tag{1.25}$$

Therefore, for some constant $C > 0$,

$$\frac{2}{\ell_n'}\mathbb{E}\Bigg[\sum_{i=K+1}^n d_i^2 \mathcal{I}_i^n(Tb_n)\Bigg] \leqslant \frac{Tb_n}{\ell_n'^2}\sum_{i=K+1}^n d_i^3 \leqslant C\Bigg(a_n^{-3}\sum_{i=K+1}^n d_i^3\Bigg),$$

which, by Assumption 1.2, tends to zero if we first take $\limsup_{n\to\infty}$ and then take $\lim_{K\to\infty}$. Consequently, for any fixed $T > 0$, as $n \to \infty$,

$$\mathbb{P}\,(\text{at least one self-loop/multiple edge is discovered before time } Tb_n) \to 0,$$

which yields Theorem 1.4.11.

One may note that Theorem 1.4.11 implies that the scaling limit of a critical $CM_n(\mathbf{d})$ is the same as that of a critical $UM_n(\mathbf{d})$. Using Janson's construction, this yields the scaling limit of $CM_n(\mathbf{d}, p_c(\lambda))$, conditioned to be simple. This is different than $UM_n(\mathbf{d}, p_c(\lambda))$, since here first the graph is conditioned on simplicity, and after that percolation is performed. However, the above argument goes through if we perform the exploration process on $CM_n(\mathbf{d}, p_c(\lambda))$ directly.

## 1.5    Evolution over the critical window

In this section, we will aim to describe the evolution of the vector $\mathbf{Z}_n(\lambda)$ asymptotically. We discuss the following theorem:

**Theorem 1.5.1.** *Fix any* $k \geqslant 1$, $-\infty < \lambda_1 < \cdots < \lambda_k < \infty$. *Then, there exists a version* **AMC** $= (\mathrm{AMC}(\lambda))_{\lambda \in \mathbb{R}}$ *of the augmented multiplicative coalescent such that, as* $n \to \infty$,

$$(\mathbf{Z}_n(\lambda_1), \ldots \mathbf{Z}_n(\lambda_k)) \xrightarrow{d} (\mathrm{AMC}(\lambda_1), \ldots, \mathrm{AMC}(\lambda_k))$$

*with respect to the* $(\mathbb{U}_\downarrow^0)^k$ *topology.*

This will be based on the analysis in [69, 70]. We first introduce the candidate for the limit, the *augmented multiplicative coalescent* (AMC). Next we describe an alternative construction for the percolation process on $\mathrm{CM}_n(\mathbf{d})$. The main problem with the percolation process $(G_n(p))_{p \in [0,1]}$ on a random graph $G_n$ is that this process is non-Markovian, unless the information about $G_n$ is incorporated into the starting sigma-algebra. However, in the context of $\mathrm{CM}_n(\mathbf{d})$, this special construction allows us to compare the percolation process over the critical window with another dynamically growing Markovian graph process. Then we discuss how the new graph process gives rise to AMC. The proof outline in this section is illustrated for the infinite third-moment case. Although the proof for the finite third-moment case [70] was given for the evolution of the component sizes only, a similar proof holds there as well (See Remark 18). Also, for sake of simplicity, we will only consider the convergence of $(\mathbf{Z}_n(\lambda_1), \mathbf{Z}_n(\lambda_2))$, and the finite-dimensional convergence follows similarly.

### 1.5.1   Augmented multiplicative coalescent

Let us now describe in detail a Markov process $(\mathbf{Z}(\lambda))_{\lambda \in \mathbb{R}}$, called the *augmented multiplicative coalescent* (AMC) process. Think of a collection of particles in a system with $\mathbf{X}(\lambda)$ describing their masses (corresponding to limiting component sizes) and $\mathbf{Y}(\lambda)$ describing an additional attribute (corresponding to surplus edges) at time $\lambda$. Let $K_1, K_2 > 0$ be constants. The evolution of the system at time $\lambda$ takes place according to the following rule:

▷ For $i \neq j$, at rate $K_1 X_i(\lambda) X_j(\lambda)$, the $i$-th and $j$-th components merge and create a new component of mass $X_i(\lambda) + X_j(\lambda)$ and attribute $Y_i(\lambda) + Y_j(\lambda)$.

▷ For any $i \geqslant 1$, at rate $K_2 X_i^2(\lambda)$, $Y_i(\lambda)$ increases to $Y_i(\lambda) + 1$.

Of course, at each event time, the indices are re-organized to give a proper element of $\mathbb{U}_\downarrow^0$. The case when ignoring $\mathbf{Y}(\lambda)$ is called the multiplicative

coalescent which was studied extensively in [8, 11] in the context of understanding the evolution of the component sizes. The augmented version was proposed in [30] to study the joint evolution of component sizes and surplus edges. In [30], the authors showed in [30, Theorem 3.1] that this is a *nearly* Feller process. To understand this precisely, let $(T_\lambda)_{\lambda \in \mathbb{R}}$ denote the semigroup of operators corresponding to augmented multiplicative coalescent. Let $f : \mathbb{U}_\downarrow^0 \mapsto \mathbb{R}$ be a bounded continuous function, and $(z_n)_{n \geqslant 1}$ be a sequence in $\mathbb{U}_\downarrow^0$ such that $z_n \to z$. Further assume that $z = (x_i, y_i)_{i \geqslant 1} \in \mathbb{U}_\downarrow^0$ is such that $\sum_i x_i = \infty$. Then, as $n \to \infty$

$$T_\lambda(f(z_n)) \to T_\lambda(f(z)). \tag{1.26}$$

Thus, (1.26) does not hold for any $z \in \mathbb{U}_\downarrow^0$, which is the reason why this is called a *nearly Feller* property.

### 1.5.2   An alternative construction of the percolation process

Consider the percolation process $(CM_n(\boldsymbol{d}, p))_{p \in [0,1]}$, coupled through the Harris coupling. We give an alternative construction of the percolation process $(CM_n(\boldsymbol{d}, p))_{p \in [0,1]}$, that allows us to study the evolution of the percolated graphs.

**Algorithm 1.2.** Let $(U_e)_{e \in [\frac{\ell_n}{2}]}$ be a finite collection of iid uniform $[0, 1]$ random variables. Construct a collection of graphs $(G_n(p))_{p \in [0,1]}$ using the following two steps:

(S0) Construct the process of edge arrivals $E_n = (E_n(p))_{p \in [0,1]}$, where $E_n(p) = \#\{e : U_e \leqslant p\}$.

(S1) Initially, $G_n(0)$ is a graph only consisting of isolated vertices with no paired half-edges. At each time point $p$ where $E_n(p)$ has a jump, choose two unpaired half-edges uniformly at random and pair them. The graph $G_n(p)$ is obtained by adding this edge to $G_n(p-)$.

Algorithm 1.2 (S0) can be regarded as the birth of edges, and (S1) ensures that the edges of the graph $G_n(p)$ are obtained from a uniform perfect matching of the corresponding half-edges. The fact that $(G_n(p))_{p \in [0,1]}$ has the same distribution as the percolation process can be proved by showing that the finite-dimensional distributions are equal. The special case that $G_n(p)$ has the same distribution as $CM_n(\boldsymbol{d}, p)$ was proved in [85] for each fixed $p$. The finite-dimensional convergence requires generalizing those arguments, which is done in Section 2.7.1.

### 1.5.3    Comparison to a Markovian dynamic construction

The problem with the alternative construction in Algorithm 1.2 is that (S1) depends on the arrival of edges during the whole process. Thus the graph process is non-Markovian. Let us now describe a dynamically evolving graph process from [29] which is Markovian and at the same time approximates the percolation process over the critical window.

**Algorithm 1.3.** Let $s_1(t)$ denote the number of unpaired half-edges at time t. Thus $s_1(0) = \ell_n$. Let $\Xi_n$ be an inhomogeneous Poisson process with rate $s_1(t)$ at time t.

(S1) At each event time of $\Xi_n$, choose two unpaired half-edges uniformly at random and pair them. Thus $s_1(t)$ decreases by two. The graph $\mathcal{G}_n(t)$ is obtained by adding this edge to $\mathcal{G}_n(t-)$.

Note that $(\mathcal{G}_n(t))_{t \geqslant 0}$ is Markovian. In fact, many properties of this graph process such as the number of unpaired half-edges can be represented using the random time change of a unit-rate Poisson process [80], and thus can be shown to converge to some solution of a differential equation. The reader is referred [152] for an introduction to this differential equation method.

The graph process $(\mathcal{G}_n(t))_{t \geqslant 0}$ turns out to approximate the percolation process in the critical regime. To state this formally, let us define

$$t_c(\lambda) = \frac{1}{2} \log \left( \frac{\nu_n}{\nu_n - 1} \right) + \frac{1}{2(\nu_n - 1)} \frac{\lambda}{c_n}.$$

**Proposition 1.5.2.** *Fix $-\infty < \lambda_1 < \lambda_2 < \infty$. There exists a coupling such that with high probability*

$$\mathcal{G}_n(t_c(\lambda) - \varepsilon_n) \subset CM_n(\mathbf{d}, p_c(\lambda)) \subset \mathcal{G}_n(t_c(\lambda) + \varepsilon_n), \quad \forall \lambda \in [\lambda_1, \lambda_2],$$
$$CM_n(\mathbf{d}, p_c(\lambda) - \varepsilon_n) \subset \mathcal{G}_n(t_c(\lambda)) \subset CM_n(\mathbf{d}, p_c(\lambda) + \varepsilon_n), \forall \lambda \in [\lambda_1, \lambda_2],$$

*where $\varepsilon_n = n^{-\gamma_0}$, for some $\eta < \gamma_0 < 1/2$.*

Notice the similarity between Algorithm 1.2 (S1) and Algorithm 1.3 (S1). In both processes, two unpaired half-edges, chosen uniformly at random without replacement, are paired. We can couple the k-th uniform choice to be exactly the same for both processes. Thus, it is enough to compare the total number of edges, i.e., the total number of times (S1) has been executed in

both algorithms. Since one edge is created per execution of (S1), it is enough to show that with high probability the following holds: For all $\lambda \in [\lambda_1, \lambda_2]$

$$e(\mathcal{G}_n(t_c(\lambda) - \varepsilon_n)) \leqslant e(CM_n(\mathbf{d}, p_c(\lambda))) \leqslant e(\mathcal{G}_n(t_c(\lambda) + \varepsilon_n)). \qquad (1.27)$$

The quantity $e(\mathcal{G}_n(t_c(\lambda) \mp \varepsilon_n))$ can be estimated using the differential equation method. After some computations, it can be shown that the sandwiching inequality in (1.27) holds with expectation of $e(CM_n(\mathbf{d}, p_c(\lambda)))$ (see Proposition 2.7.4). Therefore, it is enough to establish suitable concentration inequalities for $e(CM_n(\mathbf{d}, p_c(\lambda)))$, uniformly over $\lambda \in [\lambda_1, \lambda_2]$. We can think of $\#\{U_e \leqslant p\}/(\ell_n/2)$ as an empirical distribution function on $[0, 1]$. Thus, concentration inequalities of empirical measures such as the Dvoretzky-Kiefer-Wolfowitz inequality [132] can be applied to conclude (1.27).

As a consequence of Proposition 1.5.2, it is also enough to prove Theorem 1.5.1 for $(\mathcal{G}_n(t_c(\lambda)))_{\lambda \in \mathbb{R}}$. This is very handy, because the Markovian nature allows us to keep track of our functionals of interest. Further, Theorem 1.4.8 also holds for $\mathcal{G}_n(t_c(\lambda))$. Therefore, in the later parts of this section, we consider $\mathbf{Z}_n(\lambda)$ for the graph $\mathcal{G}_n(t_c(\lambda))$ instead.

### 1.5.4   Convergence to AMC

We write $\mathscr{C}_{(i)}(\lambda)$ for the $i$-th largest component of $\mathcal{G}_n(t_n(\lambda))$, and define $\mathcal{O}_i(\lambda)$ to be the number of unpaired half-edges in $\mathscr{C}_{(i)}(\lambda)$. Think of $\mathcal{O}_i(\lambda)$ as the *mass* of $\mathscr{C}_{(i)}(\lambda)$. Let $\mathbf{Z}_n^o(\lambda)$ denote the vector in $\mathbb{U}_\downarrow^0$, where the $|\mathscr{C}_{(i)}(\lambda)|$'s are replaced by $\mathcal{O}_i(\lambda)$'s. Firstly, using the differential equation method, it can be shown that with high probability

$$\ell_n^o(\lambda) := \sum_{i \geqslant 1} \mathcal{O}_i(\lambda) = \frac{n\mu(\nu - 1)}{\nu}(1 + o_{\mathbb{P}}(1)).$$

Moreover, during the evolution of Algorithm 1.3, between time $[t_c(\lambda), t_c(\lambda + d\lambda)]$, the $i$-th and $j$-th largest components merge at rate

$$2\mathcal{O}_i(\lambda)\mathcal{O}_j(\lambda) \times \frac{1}{\ell_n^o(\lambda) - 1} \times \frac{1}{2(\nu_n - 1)c_n} \approx \frac{\nu}{\mu(\nu - 1)^2}\left(b_n^{-1}\mathcal{O}_i(\lambda)\right)\left(b_n^{-1}\mathcal{O}_j(\lambda)\right),$$

and create a component with open half-edges $\mathcal{O}_i(\lambda) + \mathcal{O}_j(\lambda) - 2$ and surplus edges $SP(\mathscr{C}_{(i)}(\lambda)) + SP(\mathscr{C}_{(j)}(\lambda))$. Also, a surplus edge is created in $\mathscr{C}_{(i)}(\lambda)$ at rate

$$\mathcal{O}_i(\lambda)(\mathcal{O}_i(\lambda) - 1) \times \frac{1}{\ell_n^o(\lambda) - 1} \times \frac{1}{2(\nu_n - 1)c_n} \approx \frac{\nu}{2\mu(\nu - 1)^2}\left(b_n^{-1}\mathcal{O}_i(\lambda)\right)^2,$$

and $\mathscr{C}_{(i)}(\lambda)$ becomes a component with surplus edges $SP(\mathscr{C}_{(i)}(\lambda)) + 1$ and open half-edges $\mathcal{O}_i(\lambda) - 2$. Thus $(\mathbf{Z}_n^o(\lambda))_{\lambda \in [\lambda_1, \lambda_2]}$ does *not* exactly evolve as an AMC process, but it is close. Let us now outline two key steps for reaching the proof of Theorem 1.5.1 from the above heuristics.

**Comparison to an exact AMC.**    If $(\mathbf{Z}_n^o(\lambda))_{\lambda \in [\lambda_1, \lambda_2]}$ would evolve as an exact multiplicative coalescent, then $(\mathbf{Z}_n^o(\lambda_1), \mathbf{Z}_n^o(\lambda_2))$ would converge by an application of the nearly Feller property in (1.26). Unfortunately, that is not the case, since two half-edges are lost after each pairing, which makes the masses deplete. If there were no such depletion of mass, then the vector of open half-edges, along with the surplus edges, would in fact merge as an exact AMC. Thus, one can modify the graph process, where after time $t_c(\lambda_1)$, the paired half-edges are replaced with a newly born half-edge to the corresponding vertex; see Section 2.7.3. Let $(\bar{\mathbf{Z}}_n^o(\lambda))_{\lambda \in [\lambda_1, \lambda_2]}$ denote the corresponding quantity under this modified algorithm. Then, $\mathbf{Z}_n^o(\lambda_1) = \bar{\mathbf{Z}}_n^o(\lambda_1)$ and the nearly Feller property of AMC yields that $(\bar{\mathbf{Z}}_n^o(\lambda_1), \bar{\mathbf{Z}}_n^o(\lambda_2))$ converges. Finally the convergence of $(\mathbf{Z}_n^o(\lambda_1), \mathbf{Z}_n^o(\lambda_2))$ is concluded by establishing that

$$d_U(\mathbf{Z}_n^o(\lambda_2), \bar{\mathbf{Z}}_n^o(\lambda_2))) \xrightarrow{\mathbb{P}} 0,$$

where $d_U$ denotes the metric corresponding to the $\mathbb{U}_\downarrow^0$-topology. We refer the reader to Section 3.10 for the formal deduction.

**Open half-edge vs component sizes.**    Finally, Theorem 1.5.1 is about the joint convergence of component sizes and surplus edges. Thus, in order to conclude Theorem 1.5.1, it needs to be shown that, for each fixed $\lambda$, as $n \to \infty$,

$$d_U(\mathbf{Z}_n^o(\lambda), \kappa \mathbf{Z}_n(\lambda)) \xrightarrow{\mathbb{P}} 0,$$

for some $\kappa > 0$. This can be proved using our exploration process and martingale arguments. See Section 3.10 for further details.

## 1.6    Global metric structure in the infinite third-moment case

In this section, we consider the metric structure of $\mathscr{C}_{(i)}(p_c(\lambda))$ for $CM_n(\mathbf{d})$ in the infinite third-moment setting. This section is based on the results in [31] Chapter 4. The global metric structure limit in the finite third-moment case has been derived in [29, 34]. Suppose that $CM_n(\mathbf{d})$ satisfies Assumption 1.2 (i), (ii). For simplicity, we ignore the slowly-varying functions here.

The percolation parameter satisfies

$$p_c(\lambda) = \frac{1}{\nu_n}\big(1 + \lambda n^{-\frac{\tau-3}{\tau-1}}\big), \quad -\infty < \lambda < \infty.$$

We simply write $\mathscr{C}_{(i)}(\lambda)$ to denote the $i$-th largest component of $CM_n(\mathbf{d}, p_c(\lambda))$. Recall from Section 1.3.1 that $\mathscr{C}_{(i)}(\lambda)$ can be viewed as a random measured metric space. Write $\mathscr{S}_*$ for the space of all measured metric spaces equipped with the Gromov weak topology (see Section 4.3) and let $\mathscr{S}_*^{\mathbb{N}}$ denote the corresponding product space with the accompanying product topology. The goal is to show the following theorem:

**Theorem 1.6.1.** *There exists a sequence of random measured metric spaces* $(\mathscr{M}_i(\lambda))_{i \geqslant 1}$ *such that on* $\mathscr{S}_*^{\mathbb{N}}$*, as* $n \to \infty$*,*

$$\big(n^{-\eta}\mathscr{C}_{(i)}(\lambda)\big)_{i \geqslant 1} \xrightarrow{d} \big(\mathscr{M}_i(\lambda)\big)_{i \geqslant 1}. \tag{1.28}$$

The description of the limiting metric space appearing in Theorem 1.6.1 requires several definitions and concepts. An interested reader is referred to Section 4.3 for an explicit description. The organization of this section is as follows: In Section 1.6.1 we start by describing the results and proof ideas from [36], where the above theorem was established in the context of Norros-Reittu random graphs. The results allow us to explain a *universality theorem* in Section 1.6.2, which identifies a domain of attraction for the same scaling limits as [36]. In Section 1.6.3, we argue how this universality theorem can be applied to deduce the scaling limit in Theorem 1.6.1. In Section 1.6.4, we describe the idea of establishing the so-called *global mass lower bound* which can be used to improve the underlying topology of convergence in Theorem 1.6.1. Due to the technical nature of these results, the proof ideas will be more sketchy than the previous sections, with the detailed treatment left to Chapters 4, and 5.

### 1.6.1   Scaling limit for Norros-Reittu model

Before going into the proof ideas in [31], let us briefly describe the results from [36], along with key proof ideas. In this section, we write $\mathscr{G}_n(\mathbf{x}, q)$ to denote the random graph obtained by keeping edge $(i, j)$ independently with probability $1 - e^{-q x_i x_j}$. Thus this is same as the Norros-Reittu model defined in Section 1.1, where the normalization has been changed for the sake of simplicity. The scaling limit result in (1.28) was derived for $\mathscr{G}_n(\mathbf{x}, q)$, and the candidate scaling limit was identified for the heavy-tailed regime.

**p-trees and their limit.**    To create the context, first let us describe a random tree known as a **p**-tree. Fix $m \geqslant 1$, and a non-increasing sequence $\mathbf{p} = (p_i)_{i \in [m]}$ with $p_i > 0$ for all $i \in [m]$, and $\sum_{i \in [m]} p_i = 1$. Then the law of the **p**-tree, denoted by $\mathscr{T}_m^{\mathbf{p}}$, is given by

$$\mathbb{P}(\mathscr{T}_m^{\mathbf{p}} = \mathbf{t}) = \prod_{v \in [m]} p_v^{d_v(\mathbf{t})}, \quad \text{for any tree } \mathbf{t} \text{ on } m \text{ vertices.}$$

In an ordered **p**-tree, children of each individual are assigned a uniform order. It was shown in [59, 142] that the random tree $\mathscr{T}_m^{\mathbf{p}}$, after assigning length $\sigma(\mathbf{p}) := (\sum_{i \in [m]} p_i^2)^{1/2}$ to each edge, converges in distribution to the so-called inhomogeneous continuum random tree in the Gromov-Hausdorff topology. If for each fixed $i \geqslant 1$, $p_i / \sigma(\mathbf{p}) \to \beta_i$, for some $(\beta_i)_{i \geqslant 1} \in \ell_{\downarrow}^2 \setminus \ell_{\downarrow}^1$, then the limiting object turns out to be structurally completely different from the classical Brownian continuum random tree. This case exhibits the *heavy-tail effect*, in the sense that the limiting structure contain vertices, so-called *hubs*, of infinite degree almost everywhere.

**A construction of the components of $\mathscr{G}_n(\mathbf{x}, q)$.**    A novel construction of the connected components of $\mathscr{G}_n(\mathbf{x}, q)$ in the critical regime was proposed in [35, 36]. Let $\mathscr{C}_{(i)}(q)$ denote the $i$-th largest component of $\mathscr{G}_n(\mathbf{x}, q)$. Define, for $i \geqslant 1$,

$$\mathbf{p}_n^{(i)} := \left( \frac{x_v}{\sum_{v \in \mathscr{C}_{(i)}(q)} x_v} \right)_{v \in \mathscr{C}_{(i)}(q)}, \quad a_n^{(i)} := q \left( \sum_{v \in \mathscr{C}_{(i)}(q)} x_v \right)^2. \tag{1.29}$$

Then, $\mathscr{C}_{(i)}(q)$ can be informally generated in the following two steps:

**Algorithm 1.4.** (S1) *Generate a tilted* **p**-*tree with* $\mathbf{p}_n^i$ *in (1.29), and some tilting function* $L(\cdot)$.

(S2) *Generate a mixed Poisson random variable* $N$, *and add* $N$ *many surplus edges.*

The formal description of the above algorithm is given in Section 4.3, which requires several technical definitions. It turns out that all the functionals involved in the formal construction of Algorithm 1.4 depend only on the quantities defined in (1.29). Thus, the graph in Algorithm 1.4 can be generated for any **p** and $a$ instead of specific choices in (1.29). We denote such a graph by $\mathcal{G}_m(\mathbf{p}, a)$.

**Scaling limits for critical $\mathcal{G}_n(\mathbf{x}, \mathbf{q})$.**    The graph $\mathcal{G}_m(\mathbf{p}, a)$ is generated from a
$\mathbf{p}$-tree $\mathscr{T}_m^{\mathbf{p}}$, after tilting the distribution by the function $L(\cdot) = L_m(\cdot)$, and then
adding only finitely many shortcuts. Thus, provided that $(L_m)_{m \geqslant 1}$ is uni-
formly integrable, it is not difficult to imagine that the distance in $\mathcal{G}_m(\mathbf{p}, a)$
should scale similarly as the $\mathbf{p}$-tree, which is $\sigma(\mathbf{p})$. However, one needs to es-
tablish a distributional convergence result which turns out to be significantly
harder. The following result was proved in [36], and the limiting object was
identified as a function of the inhomogeneous continuum random tree after
an appropriate tilt:

**Theorem 1.6.2.** *Suppose that $\sigma(\mathbf{p}) \to 0$, and for each fixed $i \geqslant 1$, $p_i/\sigma(\mathbf{p}) \to \beta_i$,
where $\boldsymbol{\beta} = (\beta_i)_{i \geqslant 1} \in \ell_{\downarrow}^2 \setminus \ell_{\downarrow}^1$. Moreover, there exists a constant $\gamma > 0$ such that
$a\sigma(\mathbf{p}) \to \gamma$. Then, as $m \to \infty$,*

$$\sigma(\mathbf{p})\mathcal{G}_m(\mathbf{p}, a) \textit{ converges in distribution in the Gromov-weak topology.}$$

The proof of Theorem 1.6.2 consists of showing that the tilting functions are
converging in distribution, and the operation of creating shortcuts on the
space of "trees" is continuous. Next, the idea is to apply Theorem 1.6.2
with the parameters in (1.29). Therefore, one needs to prove distributional
convergence results for these parameters and also obtain the asymptotics of
$\sigma(\mathbf{p}_n^{(i)})$. These asymptotics can be obtained using exploration processes on
$\mathcal{G}_n(\mathbf{x}, \mathbf{q})$. We refer to [36] for further details.

### 1.6.2    The universality theorem

In this section, we describe the *universality* theorem which forms the basis of
the results in [31]. A similar result for the Erdős-Rényi universality class was
established in [29]. Let us first describe the universality theorem, and then
discuss how this can be applied to obtain the proof of Theorem 1.6.1.

The idea is to replace each of the vertices in the graph $\mathcal{G}_m(\mathbf{p}, a)$ by so-
called *blobs*. Blobs are a collection $\{(M_i, d_i, \mu_i)\}_{i \in [m]}$ of connected, com-
pact measured metric spaces. Consider an independent collection of random
points $\mathbf{X} := (X_{i,j})_{i,j \in [m]}$ such that $X_{i,j} \sim \mu_i$ for all $i, j$. Further, $\mathbf{X}$ is indepen-
dent of $\mathcal{G}_m(\mathbf{p}, a)$. Put an edge of *length* one between the pair of points

$$\{(X_{i,j}, X_{j,i}) : (i, j) \text{ is an edge of } \mathcal{G}_m(\mathbf{p}, a)\},$$

and denote the resulting graph by $\tilde{\mathcal{G}}_m^{\mathrm{bl}}(\mathbf{p}, a)$. $\tilde{\mathcal{G}}_m^{\mathrm{bl}}(\mathbf{p}, a)$ inherits the metric
from the graph-distance and the distances within blobs; see Section 4.4 for

an exact description. Let $u_i := \mathbb{E}[d_i(X_i, X_i')]$ where $X_i, X_i' \sim \mu_i$ independently, $B_m := \sum_{i \in [m]} p_i u_i$ and $\Delta_i = \text{diameter}(M_i)$.

**Theorem 1.6.3** (Universality theorem). *Suppose that the assumptions of Theorem 1.6.2 hold, and additionally* $\lim_{m \to \infty} \frac{\sigma(\mathbf{p}) \max_{i \in [m]} \Delta_i}{B_m + 1} = 0$. *As* $m \to \infty$,

$$\frac{\sigma(\mathbf{p})}{B_m + 1} \tilde{\mathcal{G}}_m^{\text{bl}}(\mathbf{p}, a), \text{ and } \sigma(\mathbf{p})\mathcal{G}_m(\mathbf{p}, a) \text{ have the same limit}$$

*with respect to the Gromov-weak topology.*

The proof of Theorem 1.6.3 studies the effect of introducing surplus edges, or shortcuts in a $\mathbf{p}$-tree. The proof uses the birthday construction of $\mathbf{p}$-trees from [59]. A detailed proof is provided in Section 4.4.

To this end, let us observe that Norros-Reittu random graphs have a direct relation to multiplicative coalescence in the following sense: Consider a system of $n$ vertices, with vertex $i$ having mass $x_i$. Now, at rate $x_i x_i$, an edge is created between $i$ and $j$. The obtained graph at time $q$ is distributed as $\mathcal{G}_n(\mathbf{x}, q)$, defined in Section 1.6.1. Also, if we track the sum of $x_i$'s in each component, it evolves as an exact multiplicative coalescent. Thus the multiplicative coalescence evolution essentially gives rise to Norros-Reittu random graphs. Next consider Algorithm 1.3, and the modification of replacing open half-edges after time $t_c(\lambda_1)$ given in Section 1.5.4, where now $\lambda_1 = \lambda_1(n) \to -\infty$. Therefore, at the beginning of the modification, the graph process is in the barely subcritical regime. The modified process runs as an exact multiplicative coalescent, essentially giving rise to a *superstructure* of a Norros-Reittu graph on top of the barely subcritical components. Let us denote this graph by $\bar{H}_n$, and denote the graph produced by Algorithm 1.3 by $H_n$. Now one can apply Theorem 1.6.3 to the graph $\bar{H}_n$, with the components in the barely subcritical regime serving as blobs. Finally, we obtain the metric structure of $H_n$ by comparing its structures with $\bar{H}_n$.

### 1.6.3    Final steps in completing the proof

**Properties at the barely subcritical regime.**    In order to estimate several functionals like the quantities in (1.29), and in Theorem 1.6.3, we need to obtain precise asymptotics of functionals of a barely subcritical configuration model, whose components serve as blobs. This is an interesting question in its own right, which was not studied previously for $CM_n(\mathbf{d})$, in the $\tau \in (3, 4)$ universality class. In order to calculate the quantities in (1.29), we wish to verify the entrance boundary conditions for the "behavior at $-\infty$" of the

multiplicative coalescent [11], which characterize asymptotics of several functionals related to the multiplicative coalescent in terms of the asymptotics on the entrance boundary. The barely subcritical regime lies on the entrance boundary, and the verification of the Aldous-Limic entrance boundary yields the desired asymptotics. Further, one also needs to obtain bounds on the maximum diameter of the barely subcritical components $\max_{i\in[m]} \Delta_i$, and the within component average distance $B_m$ in order to apply Theorem 1.6.3.

**Structural comparison of $H_n$ and $\bar{H}_n$.** With the asymptotics obtained in the barely subcritical regime, Theorem 1.6.3 applies to $\bar{H}_n$. Finally, a structural comparison between the components of $H_n$ and $\bar{H}_n$ completes the proof of Theorem 1.6.1. Let $\mathscr{C}_{(i)}$ and $\bar{\mathscr{C}}_{(i)}$ denote the $i$-th largest component of $H_n$ and $\bar{H}_n$ respectively. Then the following structural comparisons allow us to conclude that the limit of largest connected components are the same in the Gromov-weak topology:

▷ For any $i \geqslant 1$, $\mathscr{C}_{(i)} \subset \bar{\mathscr{C}}_{(i)}$ with high probability. Note that, under the replacement scheme of open half-edges, $\cup_{j\leqslant i}\mathscr{C}_{(j)} \subset \cup_{j\leqslant i}\bar{\mathscr{C}}_{(j)}$ almost surely for any $i \geqslant 1$. Therefore, this statement can be concluded by showing that the component sizes $(|\mathscr{C}_{(i)}|)_{i\geqslant 1}$ and $(|\bar{\mathscr{C}}_{(i)}|)_{i\geqslant 1}$ have the same limit.

▷ The "mass" of $\bar{\mathscr{C}}_{(i)} \setminus \mathscr{C}_{(i)}$ converges to zero in probability.

▷ For any pair of vertices $u, v \in \mathscr{C}_{(i)}(\lambda)$, with high probability, the shortest path between them is exactly the same in $\mathscr{C}_{(i)}$ and $\bar{\mathscr{C}}_{(i)}$. This is obtained by showing that the number of surplus edges with at least one endpoint in $\bar{\mathscr{C}}_{(i)} \setminus \mathscr{C}_{(i)}$ converges to zero in probability.

For further details about the formal statements and the verification of the above properties, the reader is referred to Section 4.5.

### 1.6.4 Gromov-Hausdorff-Prokhorov convergence of the critical components

We now describe the so-called global lower mass-bound property of the critical components. The property basically establishes a lower bound on the number of vertices within small neighborhood of the connected components. If there is a single path of length $n^\eta \log(n)$, then since $\rho > \eta$ and $|\mathscr{C}_{(i)}(p_c(\lambda))| = \Theta_{\mathbb{P}}(n^\rho)$, we do not see any members from that path if we

sample finitely many points from $\mathscr{C}_{(i)}(p_c(\lambda))$ uniformly at random. For this reason, the Gromov-weak convergence does not take into whether there is a *thin long path* in the component. Further, since the Gromov-weak convergence is defined on the space of complete separable metric spaces, the limit may not be a compact metric space. The global lower mass bound rules out the existence of such thin long paths. Consequently, this implies that the scaling limit in Theorem 1.6.1 holds with respect to the Gromov-Hausdorff-Prokhorov (GHP) topology, the limiting metric space is compact, and the global distance related functionals (e.g. the diameter) converges after the appropriate rescaling. This relation between the Gromov-weak convergence and GHP convergence was studied in [20].

We consider a critical configuration model and denote the $i$-th largest connected component of $CM_n(\mathbf{d})$ by $\mathscr{C}_{(i)}$. For each $v \in [n]$ and $\delta > 0$, let $\mathcal{N}_v(\delta)$ denote the $\delta n^\eta$ neighborhood of $v$ in $CM_n(\mathbf{d})$. For each $i \geqslant 1$, define

$$\mathfrak{m}_i^n(\delta) = \inf_{v \in \mathscr{C}_{(i)}} n^{-\rho} |\mathcal{N}_v(\delta)|.$$

**Theorem 1.6.4** (Global lower mass-bound). *For any $\delta > 0$, $(\mathfrak{m}_i^n(\delta)^{-1})_{n \geqslant 1}$ is a tight sequence.*

In Chapter 5, Theorem 1.6.4 is proved under a more general setting, but the proof requires some additional technical assumptions on the degree distribution on top of Assumption 1.2 (see Assumption 5.1). The additional assumption is satisfied for power law degrees. Following the above heuristic description, Theorem 1.6.4 now yields several interesting corollaries. Using the results from [20], Theorem 1.6.4 establishes that the convergence in Theorem 1.6.1 holds with respect to the GHP topology. This in particular establishes that the limiting metric spaces in [31, 36] are compact almost surely. Due to technical reasons, some additional conditions are imposed on $\theta$. For example, the assumption is satisfied for $\theta_i \in [L_1(i)i^{-a_1}, L_2(i)i^{-a_2}]$, where $a_1, a_2 \in (1/3, 1/2)$, and $L_1, L_2$ are slowly varying functions. This is much less restrictive than assuming $\theta_i = i^{-\alpha}$ as in [36]. The compactness of the limiting metric spaces in [31, 36] was also established under some regularity conditions in a very recent preprint [56] using a completely independent method as in this paper. In addition to the compactness of the limiting metric space, we also have the convergence of the diameters, i.e.,

$$\left(n^{-\eta}\mathrm{diam}(\mathscr{C}_{(i)}(p_c(\lambda)))\right)_{i \geqslant 1} \xrightarrow{d} (X_i)_{i \geqslant 1}$$

with respect to the product topology, where $(X_i)_{i \geqslant 1}$ is a non-degenerate random vector. In fact $X_i$ corresponds to the diameter of the limiting object of $\mathscr{C}_{(i)}(p_c(\lambda))$ from [31].

Let us just briefly describe the key ideas in the proof of Theorem 1.6.4, which consists of two main steps. The first step is to show that the neighborhoods of the high-degree vertices, or *hubs*, have mass $\Theta(n^\rho)$. Secondly, the probability of any small $\varepsilon n^\eta$ neighborhood not containing hubs is arbitrarily small. These two facts, summarized in Propositions 5.2.1 and 5.2.2 below, together ensure that the total mass of any neighborhood of $\mathscr{C}_{(i)}$ of radius $\varepsilon n^\eta$ is bounded away from zero. These facts were proved in [36] in the context of inhomogeneous random graphs. The main advantage in [36] was that the breadth-first exploration of components could be dominated by a branching process with *mixed Poisson* progeny distribution that is *independent of* $n$. The above facts allow one to use existing literature and estimate the probabilities that a long path exists in the branching process in [36]. However, such a technique is specific to rank-one inhomogeneous random graphs and does not work in the cases where the above stochastic domination is not possible. This was partly a motivating reason for this work. Moreover, along the way we derive results about exponential bounds for the number of edges in the large components (Proposition 5.4.1), and a coupling of the neighborhood exploration with a branching process with stochastically larger progeny distribution (Section 5.4.2), which is interesting in its own right. The details are left to Chapter 5.

## 1.7    Analysis in the infinite second-moment case

We next discuss the critical behavior for percolation when the asymptotic empirical degree distribution is approximately a power law with exponent $\tau \in (2,3)$, i.e., the degree distribution has infinite second moment, but finite second moment. As canonical random graph models on which percolation acts, we take $CM_n(\boldsymbol{d})$, $ECM_n(\boldsymbol{d})$ and $GRG_n(\boldsymbol{w})$. The latter two models only allow for single edges, which is the reason for referring to them as models with a single-edge constraint. In Section 1.7.1, we describe the results for $CM_n(\boldsymbol{d})$, and in Section 1.7.2 those for $ECM_n(\boldsymbol{d})$ and $GRG_n(\boldsymbol{w})$. The results are based on ongoing work [68].

### 1.7.1    Results for the configuration model

We start by describing the assumptions on the degree distribution. Fix any $\tau \in (2,3)$. We denote

$$\alpha = 1/(\tau - 1), \quad \rho = (\tau - 2)/(\tau - 1), \quad \eta = (3 - \tau)/(\tau - 1). \tag{1.30}$$

Note that the $\eta$ in (1.30) is different than in Section 1.4.3. We assume the following about the degree sequences $(\mathbf{d}_n)_{n \geqslant 1}$:

**Assumption 1.3.**    (i) (*High-degree vertices*) For any $i \geqslant 1$, $n^{-\alpha} d_i \to \theta_i$, where the vector $\theta = (\theta_1, \theta_2, \dots) \in \ell_\downarrow^2 \setminus \ell_\downarrow^1$.

(ii) (*Moment assumptions*) $\lim_{n \to \infty} \frac{1}{n} \sum_{i \in [n]} d_i = \mu$, and

$$\lim_{K \to \infty} \limsup_{n \to \infty} n^{-2\alpha} \sum_{i=K+1}^{n} d_i^2 = 0.$$

Assumption 1.3 (i) fixes the asymptotics of the high-degree vertices in a similar manner as Assumption 1.2 (i), and characterizes jumps of an associated exploration process that we describe in detail below. As before, we will consider a size-biased exploration process. Assumption 1.3 (ii) says that the expectation of this size-biased distribution is carried predominantly by the contribution due to the hubs. Note that, under Assumption 1.3, $\alpha > 1/2$, and consequently the criticality parameter for $CM_n(\mathbf{d})$ satisfies $\nu_n = \Theta(n^{2\alpha - 1})$, which tends to infinity as $n \to \infty$. The critical behavior for percolation on $CM_n(\mathbf{d})$ is observed for values of $p$ given by

$$p_c = p_c(\lambda) := \frac{\lambda}{\nu_n}(1 + o(1)), \quad \lambda \in (0, \infty). \tag{1.31}$$

The nature of the critical window for $\tau \in (2,3)$ is fundamentally different than in the finite second-moment case. Here, the graph becomes more and more subcritical (or supercritical) as $\lambda \to 0$ (or $\lambda \to \infty$), contrary to the $\lambda \to \mp\infty$ scenario for $\tau > 4$ and $\tau \in (3,4)$.

To describe the results for the component sizes and the surplus edges, recall that $\mathbf{Z}_n(\lambda)$ denotes the vector $(n^{-\rho}|\mathscr{C}_{(i)}(p_c(\lambda))|, SP(\mathscr{C}_{(i)}(p_c(\lambda))))_{i \geqslant 1}$, ordered as an element of $\mathbb{U}_\downarrow^0$. A vertex is called isolated if it has degree zero in the graph $CM_n(\mathbf{d}, p_c(\lambda))$. We define the component size corresponding to an isolated vertex to be zero. This is required because in this case $2\rho < 1$. When we perform percolation with probability $p_c(\lambda)$, we see order $n$ isolated vertices and thus $n^{-2\rho} \times$ (# isolated vertices) tends to infinity, which destroys the $\ell_\downarrow^2$-tightness of the component sizes. The following theorem gives the asymptotics for the component sizes and the complexity for $CM_n(\mathbf{d}, p_c(\lambda))$:

**Theorem 1.7.1** (Component sizes and complexity). *Under Assumption 1.3, as $n \to \infty$,*

$$\mathbf{Z}_n(\lambda) \xrightarrow{d} \mathbf{Z}(\lambda)$$

*with respect to the $\mathbb{U}_{\downarrow}^0$ topology, where $\mathbf{Z}(\lambda)$ is some non-degenerate random described in detail in Theorem 6.1.1.*

Our next result shows that the diameter of the largest connected components is of constant order, which yields further insight into the distance structure of these critical components. Let $\Delta_i^n$ denote the diameter of $\mathscr{C}_{(i)}(p_c(\lambda))$.

**Theorem 1.7.2** (Diameter of largest clusters). *Under Assumption 1.3, $(\Delta_i^n)_{n \geqslant 1}$ is a tight sequence of random variables, for any $i \geqslant 1$.*

In order to establish that (1.31) gives the critical value, we further investigate the barely sub/supercritical regimes which are defined respectively by $p_n \ll p_c(\lambda)$ and $p_n \gg p_c(\lambda)$. We prove the following theorem for the barely-subcritical regime:

**Theorem 1.7.3** (Barely subcritical regime). *Suppose that $\frac{\log(n)}{\ell_n} \ll p_n \ll p_c(\lambda)$ and Assumption 1.3 holds. Then, for each fixed $i \geqslant 1$, as $n \to \infty$,*

$$\frac{|\mathscr{C}_{(i)}(p_n)|}{n^{\alpha}p_n} \xrightarrow{\mathbb{P}} \theta_i.$$

For the result about the barely supercritical regime, we need one further mild technical assumption, which is as follows: Let $D_n^*$ denote the degree of a vertex chosen in a size-biased manner with the sizes being $(d_i/\ell_n)_{i \in [n]}$. Then, there exists a constant $\kappa > 0$ such that

$$1 - \mathbb{E}[e^{-t p_n^{1/(3-\tau)} D_n^*}] = \kappa t^{\tau-2} p_n^{(\tau-2)/(3-\tau)}(1 + o(1)). \tag{1.32}$$

Condition (1.32) is related to the Abel-Tauberian theorem [84, Chapter XIII.5], but due to the joint scaling of $p_n$ and $D_n^*$, this has to be stated as an assumption. In Chapter 6, we show that (1.32) is satisfied if $d_i = (1 - F)^{-1}(i/n)$ for some distribution function $F$ supported on non-negative integers, and $(1 - F)(x) = Ck^{-(\tau-1)}$, for $k \leqslant x < k+1$. The next theorem considers the barely supercritical regime:

**Theorem 1.7.4** (Barely supercritical regime). *Suppose that $p_n \gg p_c(\lambda)$, and Assumption 1.3 and (1.32) holds. Then, as $n \to \infty$,*

$$\frac{|\mathscr{C}_{(1)}(p_n)|}{n p_n^{1/(3-\tau)}} \xrightarrow{\mathbb{P}} \frac{\mu\kappa^{1/(3-\tau)}}{2^{(\tau-2)/(3-\tau)}}, \qquad \frac{E(\mathscr{C}_{(1)}(p_n))}{n p_n^{1/(3-\tau)}} \xrightarrow{\mathbb{P}} \frac{\mu\kappa^{1/(3-\tau)}}{2^{(4-\tau)/(3-\tau)}},$$

*and for all* $i \geqslant 2$, $|\mathscr{C}_{(i)}(p_n)| = o_{\mathbb{P}}(np_n^{1/(3-\tau)})$, $E(\mathscr{C}_{(i)}(p_n)) = o_{\mathbb{P}}(np_n^{1/(3-\tau)})$, *where* $E(G)$ *denotes the number of edges in the graph* $G$.

In the next section, we will briefly discuss the main challenges in proving Theorem 1.7.1. The reader is referred to Chapter 6 for a further rigorous treatment of all the above mentioned results.

**Proof ideas for Theorem 1.7.1.**

The proof consists of two key steps: Set up a suitable exploration process which converges to a stochastic process; and analyze the scaling limit of the exploration process.

**The exploration process.**    In Section 1.4.1, Janson's construction in Algorithm 1.1 played a crucial role in representing the percolated configuration model as a configuration model, and thus one could use the exploration process on a configuration model to make conclusions about the percolated graph. Unfortunately, this technique does not work anymore when $p_c \to 0$, because in that case red vertices outnumber non-red vertices, which makes the discovery of the non-red vertices rare during the exploration process. However, we can still *approximate* $CM_n(d, p_c(\lambda))$ by a suitable configuration model, which is described as follows:

**Algorithm 1.5.**    (S0) Retain each half-edge with probability $p_n$. If the total number of retained half-edges is odd, attach a *dummy* half-edge with vertex 1.

(S1) Perform a uniform perfect matching between the half-edges retained in (S0). Pair unpaired half-edges sequentially with a uniformly chosen unpaired half-edge until all half-edges are paired. The paired half-edges create edges in the graph, and we call the resulting graph $\mathcal{G}_n(p_n)$.

Let $\tilde{d} = (\tilde{d}_1, \ldots, \tilde{d}_n)$ be the degree sequence induced by Algorithm 1.5 (S1). Then $\mathcal{G}_n(p_n)$ is distributed as $CM_n(\tilde{d})$. Moreover, the following proposition states that it is enough to consider the scaling limit of $\mathbf{Z}_n(\lambda)$ in $\mathcal{G}_n(p_n)$:

**Proposition 1.7.5.** *Let* $p_n$ *be such that* $\ell_n p_n \gg \log(n)$. *Then there exists a sequence* $(\varepsilon_n)_{n \geqslant 1}$ *with* $\varepsilon_n \to 0$, *and a coupling such that, with high probability,*

$$\mathcal{G}_n(p_n(1-\varepsilon_n)) \subset CM_n(d, p_n) \subset \mathcal{G}_n(p_n(1+\varepsilon_n)).$$

The proof of Proposition 1.7.5 is provided in Proposition 6.2.4. We can now set up the exploration process on $CM_n(\tilde{\mathbf{d}})$. Consider the same exploration algorithm and process as in the $\tau \in (3,4)$ case defined in (1.20) on the graph $CM_n(\tilde{\mathbf{d}})$. Recall that if $\mathcal{I}_i^n(l)$ denotes the indicator that vertex $i$ is discovered before time $l$, then the exploration process is given by

$$S_n(l) = \sum_{i \in [n]} \tilde{d}_i \mathcal{I}_i^n(l) - 2l.$$

Define the re-scaled version $\bar{\mathbf{S}}_n$ of $\mathbf{S}_n$ by $\bar{S}_n(t) = n^{-\rho} S_n(\lfloor tn^\rho \rfloor)$. Then,

$$\bar{S}_n(t) = n^{-\rho} \sum_{i \in [n]} \tilde{d}_i \mathcal{I}_i^n(tn^\rho) - 2t + o(1). \tag{1.33}$$

Now, using the estimate of the exploration probability in the above exploration process from (1.25)

$$n^{-\rho} \mathbb{E}\left[ \sum_{i > K} \tilde{d}_i \mathcal{I}_i^n(tn^\rho) \Big| \tilde{\mathbf{d}} \right] \leqslant \frac{t \sum_{i > K} \tilde{d}_i^2}{\bar{\ell}_n - 2tn^\rho}. \tag{1.34}$$

Using Assumption 1.3, along with the fact that $\tilde{d}_i \sim \mathrm{Bin}(d_i, p_n)$ independently over $i \in [n]$, it is not difficult to show that the probability that the final term in (1.34) is more than $\varepsilon$ tends to zero in the iterated limit $\lim_{K \to \infty} \limsup_{n \to \infty}$. See Lemma 6.2.5 for more details. Therefore, it is enough to find the scaling limit of (1.33) by truncating the sum upto the first $K$ terms and then taking the iterated limit as $n \to \infty$ and then $K \to \infty$. Upon a closer inspection, one can verify that an analogue of Lemma 1.4.10 is true for this case as well, which yields the following result:

**Theorem 1.7.6.** *Under Assumption 1.3, as $n \to \infty$,*

$$\bar{\mathbf{S}}_n \xrightarrow{d} \bar{\mathbf{S}}_\infty$$

*with respect to the Skorohod $J_1$ topology, where the limiting process is defined by*

$$S_\infty^\lambda(t) = \lambda \sum_{i=1}^\infty \theta_i \mathcal{I}_i(t) - 2t, \tag{1.35}$$

*for $\mathcal{I}_i(s) := \mathbf{1}\{\xi_i \leqslant s\}$ with $\xi_i \sim \mathrm{Exp}(\theta_i/\mu)$ independently,*

**Analysis of the limiting process.**    The limiting process (1.35) has turned up for the first time in the critical random graph literature, and its description is not covered by the general framework provided by Aldous and Limic [11].

One needs to establish several properties of the $\mathbf{S}_\infty^\lambda$ process to conclude that the rescaled component sizes converge to its ordered excursion lengths. For example, one first needs to show whether it is at all possible to order the excursion lengths. Further, the function mapping a càdlàg function to its largest excursion is only continuous on a subset of *good* càdlàg functions under the Skorohod $J_1$ topology, see Definition 1. Therefore, one needs to verify that the sample paths of (1.35) are good almost surely. The following proposition allows us to establish all those good properties:

**Proposition 1.7.7.** *(P1) As* $t \to \infty$, $S_\infty^\lambda(t) \xrightarrow{\text{a.s.}} -\infty$. *Thus,* $\mathbf{S}_\infty^\lambda$ *does not have an excursion of infinite length almost surely.*

*(P2) For any* $\delta > 0$, $\mathbf{S}_\infty^\lambda$ *has only finitely many excursions of length at least* $\delta$ *almost surely.*

*(P3) Let* $\mathcal{R}$ *denote the set of excursion end-points of* $\mathbf{S}_\infty^\lambda$. *Then* $\mathcal{R}$ *does not have an isolated point.*

*(P4) For any* $t > 0$, $\mathbb{P}(S_\infty^\lambda(t) = \inf_{u \leqslant t} S_\infty^\lambda(u)) = 0$.

The proof is mostly technical and is provided in Section 6.2.1.

**Completing the proof.**   Let us now briefly outline the final ingredients of the proof of Theorem 1.7.6. Firstly, if $A_l$ denotes the number of active half-edges after stage $l$ while implementing the exploration algorithm, then note that the probability of creating a surplus edge at time $i$ conditionally on $\mathscr{F}_{i-1}$ is given by

$$\frac{A_{i-1}-1}{\tilde{\ell}_n - 2i - 1} = \frac{A_{i-1}}{\tilde{\ell}_n}(1 + O(i/n)) + O(n^{-1}),$$

uniformly for $i \leqslant Tn^\rho$ for any $T > 0$. Therefore, the instantaneous rate of creating surplus edges at time $tn^\rho$, conditional on the past, is

$$n^\rho \frac{A_{\lfloor tn^\rho \rfloor}}{n^{2\rho} \frac{\mu^2}{\sum_{i \geqslant 1} \theta_i^2}} (1 + o(1)) + o(1) = \frac{\sum_{i \geqslant 1} \theta_i^2}{\mu^2} \text{refl}(\bar{S}_n(t)) (1 + o(1)) + o(1).$$

As in the $\tau > 4$ and $\tau \in (3,4)$ cases, this gives the asymptotics for the surplus edges within components. Finally, as in Sections 1.4.2, 1.4.3, to conclude that the largest component sizes and surplus edges converge to ordered excursion lengths of $\mathbf{S}_\infty^\lambda$, one needs to show that the largest components are explored before time $O(n^\rho)$, and $\mathbf{Z}_n(\lambda)$ is tight in $\mathbb{U}_\downarrow^0$. The reader is referred to Section 6.2 for the final details of this proof.

### 1.7.2    Effect of the single-edge constraint

In this section, we will consider two random graph models that do not allow for self-loops and multiple-edges: the generalized random graph $\mathrm{GRG}_n(w)$ and the erased configuration model $\mathrm{ECM}_n(d)$. The critical window for percolation is given by

$$p_c = p_c(\lambda) := \lambda n^{-\frac{3-\tau}{2}}(1 + o(1)), \quad \lambda \in (0, \infty). \tag{1.36}$$

Note that the critical value in (1.36) is strictly larger in order than (1.31). In fact, this is the only regime of $\tau$ where the exponent for the critical window changes after deleting the self-loops and multiple edges of $\mathrm{CM}_n(d)$.

Let us first state the result for $\mathrm{GRG}_n(w)$. We assume the following about the sequence of weights:

**Assumption 1.4.** For some $\tau \in (2, 3)$, consider the distribution function satisfying $(1 - F)(x) = Cx^{-(\tau-1)}$ and let $w_i = (1 - F)^{-1}(i/n)$.

In the above case, if $W_n$ denotes the weight of a vertex chosen uniformly at random from $[n]$, then

$$\mathbb{E}[W_n] = \frac{1}{n}\sum_{i \in [n]} w_i \to \mu = \mathbb{E}[W].$$

Moreover,

$$n^{-\alpha}w_i = c_F i^{-\alpha},$$

for some constant $c_F > 0$. Throughout $c_F$ will denote the special constant appearing above. Assumption 1.4 is strictly stronger than Assumption 1.3 in the sense that Assumption 1.4 specifies not only the high-degree vertices but all the $w_i$'s. This is required in the proofs as one needs precise estimates of quantities like $\mathbb{E}[W_n \mathbf{1}\{W_n \geqslant K_n\}]$. See Lemma 6.3.1 for many such required estimates.

Let $\mathscr{C}_{(i)}(p)$ denote the $i$-th largest component of $\mathrm{GRG}_n(w, p)$, and define $W_{(i)}(p) := \sum_{k \in \mathscr{C}_{(i)}(p)} w_k$. We will consider the scaling limits of $(W_{(i)}(p_c))_{i \geqslant 1}$ and $(\mathscr{C}_{(i)}(p_c))_{i \geqslant 1}$. To describe the limiting object, consider the graph $G_\infty(\lambda)$ on the vertex set $\mathbb{Z}_+$, where the vertices $i$ and $j$ are joined independently by $\mathrm{Poisson}(\lambda_{ij})$ many edges with $\lambda_{ij}$ given by

$$\lambda_{ij} := \lambda^2 \int_0^\infty \theta_i(x)\theta_j(x)dx, \quad \theta_i(x) := \frac{c_F^2 i^{-\alpha}x^{-\alpha}}{\mu + c_F^2 i^{-\alpha}x^{-\alpha}}.$$

Let $W^\infty_{(i)}(\lambda)$ denote the $i$-th largest element of the set

$$\left\{ \sum_{i \in \mathscr{C}} \theta_i : \mathscr{C} \text{ is a connected component of } G_\infty(\lambda) \right\}.$$

The following describes the component sizes of $\mathrm{GRG}_n(w, p_c(\lambda))$:

**Theorem 1.7.8** (Critical regime for **$\mathrm{GRG}_n(w)$**). *There exists an absolute constant $\lambda_0$ such that for any $\lambda \in (0, \lambda_0)$, under Assumption 1.4, as $n \to \infty$,*

$$n^{-\alpha}(W_{(i)}(p_c(\lambda)))_{i \geqslant 1} \xrightarrow{d} (W^\infty_{(i)}(\lambda))_{i \geqslant 1},$$

*and*

$$(n^\alpha p_c)^{-1}(|\mathscr{C}_{(i)}(p_c(\lambda))|)_{i \geqslant 1} \xrightarrow{d} (W^\infty_{(i)}(\lambda))_{i \geqslant 1},$$

*with respect to the $\ell^2_\downarrow$ topology.*

For the erased configuration model, we will assume that $d$ satisfies Assumption 1.4. Since $d$ can only take integer values, the support of $F$ is taken to be the set of non-negative integers. The limiting object for $\mathrm{ECM}_n(d)$ is similar to that in $\mathrm{GRG}_n(w)$ by now taking

$$\theta_i(x) := 1 - e^{-\frac{c_F^2 i^{-\alpha} x^{-\alpha}}{\mu}}. \tag{1.37}$$

**Theorem 1.7.9** (Critical regime for **$\mathrm{ECM}_n(d)$**). *There exists an absolute constant $\lambda_0$ such that for any $\lambda \in (0, \lambda_0)$, under Assumption 1.4, the scaling limit results in Theorem 1.7.8 holds for $\mathrm{ECM}_n(d, p_c(\lambda))$ with identical limit objects described by (1.37) above.*

Next, we state the result about the barely subcritical regime under the single-edge constraint. The following result holds for percolation on both $\mathrm{GRG}_n(w)$ and $\mathrm{ECM}_n(d)$:

**Theorem 1.7.10.** *Suppose that Assumption 1.4 holds and $p_n \ll p_c(\lambda)$. Then, for any fixed $i \geqslant 1$, as $n \to \infty$,*

$$\frac{|\mathscr{C}_{(i)}(p_n)|}{n^\alpha p_n} \xrightarrow{\mathbb{P}} c_F i^{-\alpha}, \quad \text{and} \quad \frac{W_{(i)}(p_n)}{n^\alpha} \xrightarrow{\mathbb{P}} c_F i^{-\alpha}.$$

Under the single-edge constraint, the exact asymptotics in the barely supercritical case is left to future work. In the proofs under the single-edge constraint, coming up with a tractable exploration process for the clusters seems

challenging. The only tool we have is an estimate of the connection proba-bilities of hubs via an intermediate vertex, which allows us to estimate ex-pectations of several moments of component sizes and total weights of those component. These are often referred to as susceptibility functions. The sus-ceptibility functions allow us to ignore negligible contributions on the total weights of cluster using the first-moment method. Unfortunately, the first-moment method does not work in the barely-supercritical regime, or for high values of $\lambda$ in Theorems 1.7.8, 1.7.9. This is the reason for assuming $\lambda \in (0, \lambda_0)$ in those theorems. The proof for general $\lambda$ is an open question.

The critical window changes due to the single-edge constraint, as noted in (1.31) and (1.36). However, in both cases, the critical window is the regime where hubs start getting connected. More precisely, the critical window is given by those values of $p$ such that for any fixed $i, j \geqslant 1$

$$\lim_{n \to \infty} \mathbb{P}(i, j \text{ are in the same component of the p-percolated graph}) \in (0, 1).$$

For the configuration model, hubs are connected directly with strictly pos-itive probability. In $CM_n(\mathbf{d})$, vertices $i$ and $j$ share $d_i d_j / (\ell_n - 1)$ edges in expectation. Thus for hubs with $d_i = O(n^\alpha)$ and $d_j = O(n^\alpha)$, $O(1)$ many edges survive after percolation in expectation in the critical window (1.31). On the other hand, whenever $p \to 0$, hubs are never connected directly under the single-edge constraint. We will see in Chapter 6 and in the heuristic argu-ments below that the value $p_c$ in (1.36) is such that the hubs are connected to each other via intermediate vertices which have degree $\Theta(n^\rho)$. Intuitively, in the barely subcritical regime, all the hubs are in different components. Hubs start forming the critical components as $p$ varies over the critical window, and finally in the barely super-critical regime the giant component is formed which contains all the hubs. This feature is also observed in the $\tau \in (3, 4)$ case [33].

In the next section, we only outline the proof of Theorem 1.7.8. The proof of Theorem 1.7.9 uses similar arguments, but additional complications arise due to the dependence between occurrence of edges. We refer the reader to Section 6.3 for rigorous derivations of the above results.

**Proof ideas for Theorem 1.7.8.**

The key idea of the proof is to first consider total weights of components. In this section, we will use the notation C as a generic notation for a positive con-stant that may only depend on F. Let $\mathscr{C}(i)$ denote the component containing

vertex $i$ and $W(i) = \sum_{v \in \mathscr{C}(i)} w_v$. Further, let $W_k(i) := \sum_{v \in \mathscr{C}(i),\, d(i,v)=k} w_v$, where $d(\cdot, \cdot)$ denotes the graph-distance. The idea is to show that the primary contribution to $W(i)$ comes from vertices at finite, even distance. This is because hubs are not connected directly, but via intermediate vertices. To identify negligible contributions to $W(i)$, we use the first-moment method. Note that $\mathbb{E}[W_k(i)] \leqslant \sum_{j \in [n]} w_j f_k(i,j)$, where $f_k(i,j)$ denotes the probability that there is a path of length $k$ from $i$ to $j$. The key ingredient in our proof is the following lemma which allows us to compute $f_k(i,j)$.

**Lemma 1.7.11** (Two-hop connection probabilities). *For all $n \geqslant 1$, and $i,j \in [n]$,*

$$p_{ij}(2) := p_c^2 \sum_{v \in [n]} \frac{w_i w_v^2 w_j}{(\ell_n + w_i w_v)(\ell_n + w_j w_v)} \leqslant \frac{C\lambda^2}{(i \wedge j)^{1-\alpha}(i \vee j)^{\alpha}}. \qquad (1.38)$$

Note that $p_{ij}(2)$ is the expected number of connections between $i$ and $j$ via an intermediate vertex. The upper bound in (1.38) is exactly the same as the connection probabilities in a preferential attachment model [46, 73] (See [73, Lemma 2.2]). Therefore, existing path-counting estimates for the preferential attachment model [73, Lemma 2.4] yield, for $1 - \alpha < b < \alpha$,

$$f_{2k}(i,j) \leqslant \frac{(C\lambda^2)^k}{(i \wedge j)^b (i \vee j)^{1-b}}. \qquad (1.39)$$

The geometric bound in (1.39) gives

$$\begin{aligned}
\mathbb{E}[W_{2k}(i)] &\leqslant \sum_{j \in [n]} w_j f_{2k}(i,j) \\
&\leqslant n^{\alpha}(c_0\lambda^2)^k \left[ \sum_{j<i} \frac{1}{j^{\alpha}} \frac{1}{j^b i^{1-b}} + \sum_{j>i} \frac{1}{j^{\alpha}} \frac{1}{i^b j^{1-b}} \right] \\
&\leqslant Cn^{\alpha}(c_0\lambda^2)^k \left[ \frac{1}{i^{1-b}} + \frac{1}{i^{\alpha}} \right] \leqslant C(c_0\lambda^2)^k \frac{n^{\alpha}}{i^{1-b}} \\
&\leqslant C(c_0\lambda^2)^k w_i i^{b-(1-\alpha)}.
\end{aligned}$$

The final term decays geometrically with $k$ when $c_0\lambda^2 < 1$. This is the precise reason why the condition $\lambda \in (0, \lambda_0)$ is needed in Theorem 1.7.8. Suitable upper bounds on $\mathbb{E}[W_{2k+1}(i)]$ can also be obtained using (1.39). Thus the next proposition follows using the first-moment method:

**Proposition 1.7.12.** *For any fixed* $i \geqslant 1$, $\varepsilon > 0$, *and* $\lambda \in (0, \lambda_0)$

$$\lim_{K \to \infty} \limsup_{n \to \infty} \mathbb{P}\left( \sum_{k > K} W_{2k}(i) > \varepsilon n^{\alpha} \right) = 0,$$

$$\lim_{n \to \infty} \mathbb{P}\left( \sum_{k=0}^{\infty} W_{2k+1}(i) > \varepsilon n^{\alpha} \right) = 0.$$

Hence, the primary contribution to $W(i)$ comes from weights of vertices at finite, even distance. The next proposition goes one step further and says that, even among the vertices at finite, even distance, the major contribution comes from the hubs. For $\delta > 0$, define $V_L(\delta) := \{v : w_v > \delta n^{\alpha}\}$, and $W_k(i, \delta) := \sum_{v \notin V_L(\delta), d(v,i)=k} w_v$.

**Proposition 1.7.13.** *For any fixed* $i \geqslant 1$, $K \geqslant 1$, *and* $\varepsilon > 0$,

$$\lim_{\delta \to 0} \limsup_{n \to \infty} \mathbb{P}\left( \sum_{k=1}^{K} W_{2k}(i, \delta) > \varepsilon n^{\alpha} \right) = 0.$$

Combining Propositions 1.7.12, 1.7.13, obtaining asymptotics of $W(i)$ now boils down to identifying the hubs which are in $\mathscr{C}(i)$. From (1.38), hubs are connected to each other via some intermediate vertices with probability bounded away from zero. Let $X_{ij}$ denote the number of paths of length 2 from $i$ to $j$. Note that, for any $i, j \in V_L(\delta)$ ($i \neq j$),

$$X_{ij} = \sum_{v \neq i, j} \mathrm{Ber}\left( \frac{w_i w_j w_v^2 p_c^2}{(\ell_n + w_i w_v)(\ell_n + w_j w_v)} \right), \tag{1.40}$$

with the different Bernoulli random variables in the sum (1.40) being independent. Now, the primary contribution in the sum (1.40) comes from vertices with weight $\Theta(n^{\rho})$. In fact using some estimates of the moments of $w$, we can show that

$$\sum_{v : w_v < \delta n^{\rho}} \mathrm{Ber}\left( \frac{w_i w_j w_v^2 p_c^2}{(\ell_n + w_i w_v)(\ell_n + w_j w_v)} \right) \leqslant C \delta^{3-\tau},$$

$$\sum_{v : w_v > \delta^{-1} n^{\rho}} \mathrm{Ber}\left( \frac{w_i w_j w_v^2 p_c^2}{(\ell_n + w_i w_v)(\ell_n + w_j w_v)} \right) \leqslant C \delta^{\tau-1},$$

see Section 6.3.3. Thus,

$$X_{ij} = \sum_{v : w_v \in [\delta n^{\rho}, \delta^{-1} n^{\rho}]} \mathrm{Ber}\left( \frac{w_i w_j w_v^2 p_c^2}{(\ell_n + w_i w_v)(\ell_n + w_j w_v)} \right) + E(\delta, n)$$

$$= X_{ij}(\delta) + E(\delta, n),$$

where for any $\varepsilon > 0$ $\lim_{\delta \to 0} \limsup_{n \to \infty} \mathbb{P}(E(\delta, n) > \varepsilon) = 0$. Define

$$\lambda_{ij}(\delta) = \sum_{v: w_v \in [\delta n^\rho, \delta^{-1} n^\rho]} \frac{w_i w_j w_v^2 p_c^2}{(\ell_n + w_i w_v)(\ell_n + w_j w_v)}.$$

Thus, the above term can be approximated by a Poisson random variable using Stein's method and

$$\lim_{\delta \to 0} \lim_{n \to \infty} \lambda_{ij}(\delta) = \lambda^2 \int_0^\infty \theta_i(x)\theta_j(x)dx, \quad \theta_i(x) := \frac{c_F^2 i^{-\alpha} x^{-\alpha}}{\mu + c_F^2 i^{-\alpha} x^{-\alpha}}.$$

Recall the description of the graph $G_\infty(\lambda)$ from Theorem 1.7.8. The above proves that the limit of $n^{-\alpha} W(i)$ is basically $\sum_{v \in C(i)} v^{-\alpha}$, $C(i)$ being the connected component containing $i$ in $G_\infty(\lambda)$. $|\mathscr{C}(i)| = p_c(\lambda)W(i)(1 + o_\mathbb{P}(1))$ is proved using the second-moment method. Finally, to prove the scaling limit of the ordered vector of component sizes and weights, we show that the vectors are tight in $\ell_\downarrow^2$ in Section 6.3.5. This completes the sketch of the proof of Theorem 1.7.8.

## 1.8    Summary of contributions

In summary, we analyze the critical window for the percolation process on random graph models such as $CM_n(\mathbf{d})$, $UM_n(\mathbf{d})$, $ECM_n(\mathbf{d})$ and $GRG_n(\mathbf{w})$. When the degree distribution satisfies a power law with exponent $\tau$, three universality classes arise depending on whether $\tau > 4$ (finite third moment), $\tau \in (3, 4)$ (infinite third moment) and $\tau \in (2, 3)$ (infinite second moment). Let us summarize the main contributions of this thesis below:

**Component sizes and complexity for finite third-moment case.**    In Chapter 2, we obtain precise asymptotics for the component sizes and the surplus edges for $CM_n(\mathbf{d}, p_c(\lambda))$ and $UM_n(\mathbf{d}, p_c(\lambda))$ in the *critical window* of the phase transition under a finite third-moment condition. The main contribution of this work is that we derive the strongest scaling limit results in the literature under optimal assumptions. This finite third-moment assumption is also necessary for Erdős-Rényi type scaling limits, since, amongst other reasons, the third-moment appears in the scaling limit. Also, we prove the joint convergence of the component sizes and the surplus edges under a strong topology namely the $\mathbb{U}_\downarrow^0$-topology, which improves the previously known results [121, 135, 144] substantially. The re-scaled vector of component sizes (ordered in a decreasing manner) is shown to converge to the ordered excursion lengths of a reflected inhomogeneous Brownian motion with a negative

parabolic drift. Moreover, the surplus edges converge jointly with the component sizes under $\mathbb{U}_{\downarrow}^0$-topology to Poisson random variables with parameters being the areas under the above mentioned ordered excursion lengths.

**Component sizes and complexity for infinite third-moment case.**   In Chapter 3, we consider the critical behavior for the component sizes and surplus edges in the infinite third-moment case. We consider a general set of assumptions, which include the case that the empirical degree distribution satisfies $\mathbb{P}\,(D_n \geqslant k) \sim L_0(k)/k^{\tau-1}$ for some $\tau \in (3,4)$ and $L_0(\cdot)$ a slowly-varying function. The largest connected components turn out to be of the order $n^{(\tau-2)/(\tau-1)}L(n)^{-1}$ and the width of the scaling window is of the order $n^{(\tau-3)/(\tau-1)}L(n)^{-2}$ for some slowly-varying function $L(\cdot)$. The joint distribution of the re-scaled component sizes and the surplus edges is shown to converge in distribution to a suitable limiting random vector under $\mathbb{U}_{\downarrow}^0$-topology. The scaling limits for the re-scaled ordered component sizes can be described in terms of the ordered excursions of a certain thinned Lévy process that only depends on the asymptotics of the *high*-degree vertices. This universality class was first identified in [33] in the context of Norros-Reittu random graphs. Further, the scaling limits for the surplus edges can be described by Poisson random variables with the parameters being the areas under the excursions of the thinned Lévy process. The results also hold conditioned on the graph being simple, thus solving an open question [121, Conjecture 8.5].

**Evolution of components and surplus edges.**   As $\lambda$ increases over the critical window, the component sizes and surplus edges jointly evolve, with components merging with each other, and more surplus edges getting created. In Chapters 2 and 3, the evolution of the component sizes and surplus edges is shown to converge to a version of the *augmented multiplicative coalescent* process both in the finite third-moment and infinite third-moment regimes. In fact, in the $\tau \in (3,4)$ case, our results imply that there exists a version of the augmented multiplicative coalescent process whose one-dimensional distribution can be described by the excursions of a thinned Lévy process and a Poisson process with the intensity being proportional to the thinned Lévy process, which is also novel.

**Metric structure of critical components in the infinite third-moment case.** In Chapter 4, we consider the metric structure of the critical components for

$CM_n(d, p_c(\lambda))$, with degree-exponent $\tau \in (3, 4)$. In this context, candidate limit law of maximal components with each edge rescaled to have length $1/n^{(\tau-3)/(\tau-1)}$ was established recently in [36]. In this work, we establish sufficient uniform asymptotic negligibility (UAN) conditions for a random graph model in the barely subcritical regime which, in combination with the appropriate merging dynamics of components as one increases edge density through the critical regime, implies convergence to limits obtained in [36]. This result identifies the domain of attraction for the limit laws established in [36], which holds for general sequences of dynamically evolving graphs. As a canonical example, we analyze the critical regime for percolation on the uniform random graph model (and the closely associated configuration model) with prescribed degree distribution that converges to a heavy-tailed degree distribution. In order to carry out the above analysis and in particular check the UAN assumptions, we establish refined bounds for various susceptibility functionals and diameter in the barely subcritical regime of the configuration model which are of independent interest. In Chapter 5, we prove the global lower mass-bound property for these critical components, which proves the convergence of the largest components with respect to the Gromov-Hausdorff-Prokhorov topology. The latter yields the compactness of the scaling limit in Chapter 4, as well as scaling limits of global functionals like diameter.

**Component sizes in the infinite second-moment case.** In Chapter 6, we consider a new universality class which corresponds to the degree exponent $\tau \in (2, 3)$. In this regime, the critical behavior is observed when the percolation probability tends to zero with the network size. We identify the critical window for the configuration model, the erased configuration model and the generalized random graph. The critical window for graphs with single-edges is given by $p_c \sim \lambda n^{-(3-\tau)/2}$, which is much larger than $p_c \sim n^{-(3-\tau)/(\tau-1)}$ for the multigraph $CM_n(d)$. This feature is unique to the critical behavior in the $\tau \in (2, 3)$ regime. The component sizes in both cases scale as $n^\alpha p_c$. For the configuration model multigraph, we obtain scaling limits for the largest component sizes and surplus edges under a strong topology. Further, the diameter of the largest components is shown to be a tight sequence of random variables. To establish that the scaling limits correspond to the critical behavior, we further look at the near-critical behavior and derive the asymptotics for the component sizes in the so-called barely sub/supercritical regimes. On the other hand, under the single-edge constraint, we identify the scaling

limit of the largest component sizes in the part of the critical window, where the criticality parameter is sufficiently small. The proof where the criticality parameter can be arbitrary is an ongoing work.

This is the first work on critical percolation on random graphs in the $\tau \in (2,3)$ setting, thus the techniques are novel. The primary difficulty in this setting is that the exploration process approach does not work. For the configuration model, this difficulty is circumvented by sandwiching the percolated graphs by two configuration models, which yield the same scaling limits for the component sizes. The main novelty in the proof of the configuration model is the analysis of the limiting exploration process. On the other hand, in the single-edge constraint scenario, the proofs require a more detailed understanding of the structure of the critical components. It turns out that the hubs do not connect to each other directly, but there are some special vertices that interconnect hubs. This interconnected structure forms the *core* of the critical components, and the 1-neighborhood of the core spans the critical components. We primarily use path counting techniques here since the exploration process approach does not seem to work anymore. For path-counting, we compare the connection probabilities between the hubs to the connection probabilities in a preferential attachment model, which is interesting in its own right.

# Chapter 2

# Critical window for the configuration model: finite third moment degrees

**Abstract.** We investigate the component sizes of the critical configuration model, as well as the related problem of critical percolation on a supercritical configuration model. We show that, at criticality, the finite third moment assumption on the asymptotic degree distribution is enough to guarantee that the component sizes are $O(n^{2/3})$ and the re-scaled component sizes converge to the excursions of an inhomogeneous Brownian Motion with a parabolic drift. This identifies the minimal condition for the critical behavior to be in the Erdős-Rényi universality class. We use percolation to study the evolution of these component sizes while passing through the critical window and show that the vector of percolation cluster-sizes, considered as a process in the critical window, converge to the multiplicative coalescent process in finite dimensions. This behavior was first observed for Erdős-Rényi random graphs by Aldous (1997) and our results provide support for the empirical evidences that the nature of the phase transition for a wide array of random-graphs are universal in nature. Further, we show that the re-scaled component sizes and surplus edges converge jointly under a strong topology, at each fixed location of the scaling window.

In this chapter, we focus on the critical behavior of the *configuration model*, and critical percolation on these graphs when the empirical degree distribution satisfies a finite third-moment condition. We include detailed proofs of all scaling limit results mentioned in Chapter 1 about the finite third-moment case. The scaling limit result for the component sizes and surplus edges holds for the critical configuration model which includes critical percolation on the configuration model as a special case. We also study percolation on a super-critical configuration model to show that the scaled vectors of component sizes at multiple locations of the percolation scaling window converge jointly to the finite-dimensional distributions of a multiplicative coalescent process. The scaling limit results show that component sizes and surplus edges of $CM_n(\mathbf{d})$ in the critical regime, for a large collection of possible degree sequences $\mathbf{d}$, lie in the same universality class as for the Erdős-Rényi random graph [8]. Before stating the main results, we need to introduce some notation and concepts.

## 2.1    Definitions and notation

Recall the definitions from Chapter 1.2.1. For a triangular array of random variables $(f_{k,n})_{k,n \geqslant 1}$, we write phrases like $f_{k,n} = O_{\mathbb{P}}(n^\alpha)$ (respectively $o_{\mathbb{P}}(n^\alpha)$), uniformly over $k \leqslant n^\beta$ to mean that $\sup_{k \leqslant n^\alpha} |f_{k,n}| = O_{\mathbb{P}}(n^\alpha)$ (respectively $o_{\mathbb{P}}(n^\alpha)$). We also write $f_n = O_E(a_n)$ (respectively $f_n = o_E(a_n)$) to denote that $\sup_{n \geqslant 1} \mathbb{E}\left[a_n^{-1} f_n\right] < \infty$ (respectively $\lim_{n \to \infty} \mathbb{E}\left[a_n^{-1} f_n\right] = 0$).

In this chapter, $\mathbf{B}_{\mu,\eta}^\lambda$ denotes an inhomogeneous Brownian motion with a parabolic drift, given by

$$B_{\mu,\eta}^\lambda(s) = \frac{\sqrt{\eta}}{\mu} B(s) + \lambda s - \frac{\eta s^2}{2\mu^3}, \tag{2.1}$$

where $\mathbf{B} = (B(s))_{s \geqslant 0}$ is a standard Brownian motion, and $\mu > 0$, $\eta > 0$ and $\lambda \in \mathbb{R}$ are constants. Define the reflected version of $\mathbf{B}_{\mu,\eta}^\lambda$ as

$$W^\lambda(s) = B_{\mu,\eta}^\lambda(s) - \min_{0 \leqslant t \leqslant s} B_{\mu,\eta}^\lambda(t). \tag{2.2}$$

For a function $f \in \mathbb{C}[0,\infty)$, an interval $\gamma = (l, r)$ is called an *excursion above past minima* or simply an *excursion* of $f$ if $f(l) = f(r) = \min_{u \leqslant r} f(u)$ and $f(x) > f(r)$ for all $l < x < r$. $|\gamma| = r(\gamma) - l(\gamma)$ will denote the length of the excursion $\gamma$.

Also, define the counting process of marks $\mathbf{N}^\lambda = (N^\lambda(s))_{s \geqslant 0}$ to be a unit-jump process with intensity $\beta W^\lambda(s)$ at time $s$ conditional on $(W^\lambda(u))_{u \leqslant s}$ so

that

$$N^\lambda(s) - \int_0^s \beta W^\lambda(u) \, du \tag{2.3}$$

is a martingale (see [8]). For an excursion $\gamma$, let $N(\gamma)$ denote the number of marks in the interval $[l(\gamma), r(\gamma)]$.

**Remark 1.** By [8, Lemma 25], the excursion lengths of $\mathbf{B}^\lambda_{\mu,\eta}$ can be rearranged in decreasing order of length and the ordered excursion lengths can be considered as a vector in $\ell^2_\downarrow$, almost surely. Let $\gamma^\lambda = (|\gamma^\lambda_j|)_{j \geqslant 1}$ be the ordered excursion lengths of $\mathbf{B}^\lambda_{\mu,\eta}$. Then, $(|\gamma^\lambda_j|, N(\gamma^\lambda_j))_{j \geqslant 1}$ can be ordered as an element of $\mathbb{U}^0_\downarrow$ almost surely by [30, Theorem 3.1 (iii)]. We denote this element of $\mathbb{U}^0_\downarrow$ by $\mathbf{Z}(\lambda) = ((Y^\lambda_j, N^\lambda_j))_{j \geqslant 1}$ obtained from $(|\gamma^\lambda_j|, N(\gamma^\lambda_j))_{j \geqslant 1}$.

Finally, we define a Markov process $\mathbb{X} := (\mathbf{X}(s))_{s \in \mathbb{R}}$ on $\mathbb{D}(\mathbb{R}, \ell^2_\downarrow)$, called the *multiplicative coalescent process*. Think of $\mathbf{X}(s)$ as a collection of masses of some particles (possibly infinite) in a system at time $s$. Thus the $i^{\text{th}}$ particle has mass $X_i(s)$ at time $s$. The evolution of the system takes place according to the following rule at time $s$: At rate $X_i(s)X_j(s)$, particles $i$ and $j$ merge into a new particle of mass $X_i(s) + X_j(s)$. This process has been extensively studied in [8, 11]. In particular, Aldous [8, Proposition 5] showed that this is a Feller process.

## 2.2    Main results

In this section, we discuss the main results in this chapter. We start by recalling the definition of the configuration model from Chapter 1.1, which is denoted by $\mathrm{CM}_n(\mathbf{d})$. Our results are twofold and concern (i) general $\mathrm{CM}_n(\mathbf{d})$ at criticality, and (ii) critical percolation on a super-critical configuration model, both under a finite third moment assumption.

### 2.2.1    Configuration model results

For each $n \geqslant 1$, let $\mathbf{d} = \mathbf{d}_n = (d_i)_{i \in [n]}$ be a degree sequence such that $\ell_n = \sum_{i \in [n]} d_i$ is even. We suppress $n$ in the notation of the degree sequence to simplify writing. We consider a sequence of configuration models $(\mathrm{CM}_n(\mathbf{d}))_{n \geqslant 1}$ satisfying the following conditions:

**Assumption 2.1.** Let $D_n$ denote the degree of a vertex chosen uniformly at random independently of the graph. Then the following holds as $n \to \infty$:

(i) (*Weak convergence of* $D_n$) $D_n \overset{d}{\to} D$ for some random variable $D$ such that $\mathbb{E}[D^3] < \infty$.

(ii) (*Uniform integrability of* $D_n^3$) $\mathbb{E}\left[D_n^3\right] = \frac{1}{n}\sum_{i\in[n]} d_i^3 \to \mathbb{E}[D^3]$.

(iii) (*Critical window*) $\nu_n := \frac{\sum_{i\in[n]} d_i(d_i-1)}{\sum_{i\in[n]} d_i} = 1 + \lambda n^{-1/3} + o(n^{-1/3})$, for some constant $\lambda \in \mathbb{R}$,

(iv) $\mathbb{P}(D = 1) > 0$.

Suppose that $\mathscr{C}_{(1)}, \mathscr{C}_{(2)},...$ are the connected components of $\mathrm{CM}_n(\mathbf{d})$ in decreasing order of size. In case of a tie, order the components according to the values of the minimum indices of vertices in those components. For a connected graph $G$, let $\mathrm{SP}(G) := $ (number of edges in $G$) $- (|G| - 1)$ denote the number of surplus edges. Intuitively, this measures the deviation of $G$ from a tree-like structure. Let $\sigma_r = \mathbb{E}[D^r]$ and consider the reflected Brownian motion, the excursions, and the counting process $\mathbf{N}^\lambda$ as defined in Section 2.1 with parameters

$$\mu := \sigma_1, \quad \eta := \sigma_3\mu - \sigma_2^2, \quad \beta := 1/\mu. \tag{2.4}$$

Let $\gamma^\lambda$ denote the vector of excursion lengths of the process $\mathbf{B}_{\mu,\eta}^\lambda$, arranged in non-increasing order. The next two theorems are our main results for the critical configuration model:

**Theorem 2.2.1.** *Fix any* $\lambda \in \mathbb{R}$. *Under Assumption 2.1,*

$$n^{-2/3}\big(|\mathscr{C}_{(j)}|\big)_{j\geqslant 1} \overset{\mathcal{L}}{\to} \gamma^\lambda$$

*with respect to the* $\ell_\downarrow^2$ *topology.*

Recall the definition of $\mathbf{Z}(\lambda)$ from Remark 1. Order the vector component sizes and surplus edges $\big(n^{-2/3}|\mathscr{C}_{(j)}|, \mathrm{SP}(\mathscr{C}_{(j)})\big)_{j\geqslant 1}$ as an element of $\mathbb{U}_\downarrow^0$ and denote it by $\mathbf{Z}_n(\lambda)$.

**Theorem 2.2.2.** *Fix any* $\lambda \in \mathbb{R}$. *Under Assumption 2.1,*

$$\mathbf{Z}_n(\lambda) \overset{\mathcal{L}}{\to} \mathbf{Z}(\lambda)$$

*with respect to the* $\mathbb{U}_\downarrow^0$ *topology.*

In words, Theorem 2.2.1 gives the precise asymptotic distribution of the component sizes re-scaled by $n^{2/3}$ and Theorem 2.2.2 gives the asymptotic number of surplus edges in each component jointly with their sizes.

**Remark 2.** The strength of Theorems 2.2.1 and 2.2.2 lies in Assumption 2.1. Clearly, Assumption 2.1 is satisfied when the distribution of D satisfies an asymptotic power-law relation with finite third moment, i.e., $\mathbb{P}(D \geqslant x) \sim x^{-(\tau-1)}(1 + o(1))$ for some $\tau > 4$. Also, if a random degree-sequence satisfies Assumption 2.1 with high probability, then Theorems 2.2.1 and 2.2.2 hold conditionally on the degrees. In particular, when the degree sequence consists of an i.i.d sample from a distribution with $\mathbb{E}[D^3] < \infty$ [121], then Assumption 2.1 is satisfied almost surely. We will later see that degree sequences in the percolation scaling window also satisfy Assumption 2.1.

### 2.2.2  Percolation results

Bond percolation on a graph G refers to deleting edges of G independently with equal probability p. In the case G is a random graph, the deletion of edges is also independent of G. Consider bond percolation on $CM_n(\mathbf{d})$ with probability $p_n$, yielding $CM_n(\mathbf{d}, p_n)$. We assume the following:

**Assumption 2.2.**  (i) Assumption 2.1 (i) and (ii) hold for the degree sequence and $CM_n(\mathbf{d})$ is super-critical, i.e.

$$\nu_n = \frac{\sum_{i \in [n]} d_i(d_i - 1)}{\sum_{i \in [n]} d_i} \to \nu = \frac{\mathbb{E}\left[D(D-1)\right]}{\mathbb{E}\left[D\right]} > 1.$$

(ii)  (*Critical window for percolation*) For some $\lambda \in \mathbb{R}$,

$$p_n = p_n(\lambda) := \frac{1}{\nu_n}\left(1 + \frac{\lambda}{n^{1/3}}\right).$$

Note that $p_n(\lambda)$, as defined in Assumption 2.2 ii, is always non-negative for n sufficiently large. Now, suppose $\tilde{d}_i \sim \text{Bin}(d_i, \sqrt{p_n})$, $n_+ := \sum_{i \in [n]}(d_i - \tilde{d}_i)$ and $\tilde{n} = n + n_+$. Consider the degree sequence $\tilde{\mathbf{d}}$ consisting of $\tilde{d}_i$ for $i \in [n]$ and $n_+$ additional vertices of degree 1, i.e. $\tilde{d}_i = 1$ for $i \in [\tilde{n}] \setminus [n]$. We will show later that the degree $\tilde{D}_n$ of a random vertex from this degree sequence satisfies Assumption 2.1 (i), (ii) almost surely for some random variable $\tilde{D}$ with $\mathbb{E}[\tilde{D}^3] < \infty$. Moreover, $\tilde{n}/n \to 1 + \mu(1 - \nu^{-1/2}) = \zeta$ almost surely. Now, using the notation in Section 2.1, define $\tilde{\gamma}_j^\lambda = \zeta^{2/3}\bar{\gamma}_j^\lambda$, where $\bar{\gamma}_j^\lambda$ is the $j^{\text{th}}$ largest excursion of the inhomogeneous Brownian motion $\mathbf{B}_{\mu,\eta}^\lambda$ with the parameters

$$\mu = \mathbb{E}[\tilde{D}], \quad \eta = \mathbb{E}[\tilde{D}^3]\mathbb{E}[\tilde{D}] - \mathbb{E}^2[\tilde{D}^2], \quad \beta = 1/\mathbb{E}[\tilde{D}]. \tag{2.5}$$

Define the process $\tilde{\mathbf{N}}$ as in (2.3) with the parameter values given by (2.5). Denote the $j^{\text{th}}$ largest cluster of $CM_n(\mathbf{d}, p_n(\lambda))$ by $\mathscr{C}_{(j)}^p(\lambda)$. Also, let $\mathbf{Z}_n^p(\lambda)$

denote the vector in $\mathbb{U}_\downarrow^0$ obtained by rearranging critical percolation clusters (re-scaled by $n^{2/3}$) and their surplus edges and $\tilde{\mathbf{Z}}(\lambda)$ denote the vector in $\mathbb{U}_\downarrow^0$ obtained by rearranging $((\sqrt{\nu}|\tilde{\gamma}_j^\lambda|, \tilde{N}(\tilde{\gamma}_j^\lambda)))_{j \geqslant 1}$.

**Theorem 2.2.3.** *Under Assumption 2.2,*

$$\mathbf{Z}_n^p(\lambda) \xrightarrow{\mathcal{L}} \tilde{\mathbf{Z}}(\lambda)$$

*with respect to the $\mathbb{U}_\downarrow^0$ topology.*

Next we consider the percolation clusters for multiple values of $\lambda$. There is a very natural way to couple $(CM_n(\mathbf{d}, p_n(\lambda))_{\lambda \in \mathbb{R}}$ described as follows: Suppose that each edge $(ij)$ of $CM_n(\mathbf{d})$ has an associated i.i.d uniform random variable $U_{ij}$, and the $U_{ij}$'s are also independent of $CM_n(\mathbf{d})$. Now, delete edge $(ij)$ if $U_{ij} > p_n(\lambda)$. The obtained graph is distributed as $CM_n(\mathbf{d}, p_n(\lambda))$. Moreover, if we fix the set of uniform random variables and change $\lambda$, this produces a coupling between the graphs $(CM_n(\mathbf{d}, p_n(\lambda))_{\lambda \in \mathbb{R}}$. The next theorem shows that the convergence of the component sizes holds jointly in finitely many locations within the critical window, under the above described coupling:

**Theorem 2.2.4.** *Let us denote $\mathbf{C}_n(\lambda) = (n^{-2/3}|\mathscr{C}_{(j)}^p(\lambda)|)_{j \geqslant 1}$. Suppose that Assumption 2.2 holds. For any $k \geqslant 1$ and $-\infty < \lambda_0 < \lambda_1 < \cdots < \lambda_{k-1} < \infty$,*

$$\left(\mathbf{C}_n(\lambda_0), \mathbf{C}_n(\lambda_1), \ldots, \mathbf{C}_n(\lambda_{k-1})\right) \xrightarrow{\mathcal{L}} \sqrt{\nu}(\tilde{\boldsymbol{\gamma}}^{\lambda_0}, \tilde{\boldsymbol{\gamma}}^{\lambda_1}, \ldots, \tilde{\boldsymbol{\gamma}}^{\lambda_{k-1}})$$

*with respect to the $(\ell_\downarrow^2)^k$ topology.*

**Remark 3.** The coupling for the limiting process in Theorem 2.2.4 is given by the multiplicative coalescent process described in Section 2.1. This will become more clear when we describe the ideas of the proof. To understand this intuitively, notice that the component $\mathscr{C}_{(i)}^p(\lambda)$ consists of some paired half-edges which form the edges of the percolated graph, and some *open half-edges* which were deleted due to percolation. Denote by $\mathcal{O}_i^p(\lambda)$, the total number of open half-edges of $\mathscr{C}_{(i)}^p(\lambda)$. One can think of $\mathcal{O}_i^p$ as the mass of $\mathscr{C}_{(i)}^p$. Now, as we change the value of the percolation parameter from $p_n(\lambda)$ to $p_n(\lambda + d\lambda)$, exactly one edge is added to the graph and the two endpoints are chosen proportional to the number of open half-edges of the components of $CM_n(\mathbf{d}, p_n(\lambda))$. By the above heuristics, $\mathscr{C}_{(i)}^p$ and $\mathscr{C}_{(j)}^p$ merge at rate proportional to $\mathcal{O}_i^p \mathcal{O}_j^p$ and creates a component of mass $\mathcal{O}_i^p + \mathcal{O}_j^p - 2$. Later, we will show that the mass of a component is approximately proportional to the

component size. Therefore, the component sizes merge *approximately* like the multiplicative coalescent over the critical scaling window.

**Remark 4.** Janson [109] studied the phase transition of the maximum component size for percolation on a super-critical configuration model. The critical value was shown to be $p = 1/\nu$. This is precisely the reason behind taking $p_n$ of the form given by Assumption 2.2 (ii). The width of the scaling window is intimately related to the asymptotics of the susceptibility function $\sum_i |\mathscr{C}_{(i)}|^2/n$. In fact, if $\sum_i |\mathscr{C}_{(i)}|^2 \sim n^{1+\eta}$, then the width of the critical window turns out to be $n^\eta$ and the largest component sizes are of the order $n^{(1+\eta)/2}$. This has been universally observed in the random graph literature [8, 32, 66, 121, 135, 144], even when the scaling limit is not in the same universality class as Erdős-Rényi random graphs [33, 69] and the same turns out to be the case in this chapter.

**Remark 5.** Theorems 2.2.1 and 2.2.2 also hold for configuration models conditioned on simplicity. We do not give a proof here. The arguments in [121, Section 7] can be followed verbatim to obtain a proof of this fact. As a result, Theorems 2.2.3 and 2.2.4 also hold, conditioned on simplicity.

The rest of the chapter is organized as follows: In Section 2.3.1, we give a brief overview of the relevant literature. This will enable the reader to understand better the relation of this work to the large body of literature already present. Also, it will become clear why the choices of the parameters in Assumption 2.1 (iii) and Assumption 2.2 (ii) should correspond to the critical scaling window. We prove Theorems 2.2.1 and 2.2.2 in Section 2.4. In Section 2.5 we find the asymptotic degree distribution in each component. This is used along with Theorem 2.2.2 to establish Theorem 2.2.3 in Section 2.6. In Section 2.7, we analyze the evolution of the component sizes over the percolation critical window and prove Theorem 2.2.4.

## 2.3   Discussion

### 2.3.1   Relation to precious work

**Erdős-Rényi type behavior.** We first explain what 'Erdős-Rényi type behavior' means. The study of *critical window* for random graphs started with the seminal paper [8] on Erdős-Rényi random graphs with $p = n^{-1}(1 + \lambda n^{-1/3})$. Aldous showed in this regime that the largest components are of asymptotic size $n^{2/3}$ and the ordered component sizes (scaled by $n^{2/3}$) asymptotically

have the same distribution as the ordered excursion lengths of a Brownian motion with a negative parabolic drift. Aldous also considered a natural coupling of the re-scaled vectors of component sizes as $\lambda$ varies, and viewed it as a dynamic $\ell^2_\downarrow$-valued stochastic process. It was shown that the dynamic process can be described by a process called the *standard multiplicative coalescent*, which has the Feller property. This implies the convergence of the component sizes jointly for different $\lambda$ values. In Theorem 2.2.4, we show that similar results hold for the configuration model under a very general set of assumptions. Of course, for general configuration models, there is no obvious way to couple the graphs such that the location parameter in the scaling window varies and percolation seems to be the most natural way to achieve this. By [85, 109], percolation on a configuration model can be viewed as a configuration model with a random degree sequence and this is precisely the reason for studying percolation in this chapter.

**Universality and optimal assumptions.** In [32] it was shown that, inside the critical scaling window, the ordered component sizes (scaled by $n^{2/3}$) of an inhomogeneous random graph with

$$p_{ij} = 1 - \exp\left(\frac{-(1+\lambda n^{-1/3})w_i w_j}{\sum_{k\in[n]} w_k}\right)$$

converge to the ordered excursion lengths of an inhomogeneous Brownian motion with a parabolic drift under only a finite third-moment assumption on the weight distribution. We establish a counterpart of this for the configuration model in Theorem 2.2.1. Later Nachmias and Peres [135] studied the case of percolation scaling window on random regular graphs; for percolation on the configuration model similar results were obtained by Riordan [144] for bounded maximum degrees. Joseph [121] obtained the same scaling limits as Theorem 2.2.1 for the component sizes when the degrees form an i.i.d. sample from a distribution having finite third moment. Theorems 2.2.2 and 2.2.3 prove stronger versions of all these existing results for the configuration model under the optimal assumptions. Further, in Theorem 2.2.4, we give a dynamic picture for percolation cluster sizes in the critical window and show that this dynamics can be approximated by the multiplicative coalescent.

**Comparison to branching processes.** In [114, 133] the phase transition for the component sizes of $CM_n(\mathbf{d})$ was identified in terms of the parameter $\nu = \mathbb{E}[D(D-1)]/\mathbb{E}[D]$. The local neighborhoods of the configuration model can be approximated by a branching process $\mathcal{X}$ which has $\nu$ as its expected

progeny and thus, when $\nu > 1$, $\mathrm{CM}_n(\mathbf{d})$ has a component $\mathscr{C}_{\max}$ of approximate size $\rho n$, where $\rho$ is the survival probability of $\mathcal{X}$. Further, the progeny distribution of $\mathcal{X}$ has finite variance when $\mathbb{E}[D^3] < \infty$. Now, for a branching process with mean $\approx 1 + \varepsilon$ and finite variance $\sigma^2$, the survival probability is approximately $2\sigma^{-2}\varepsilon$ for small $\varepsilon > 0$. This seems to suggest that the largest component size under Assumption 2.1 should be of the order $n^{2/3}$ since $\varepsilon = \Theta(n^{-1/3})$. Theorem 2.2.1 mirrors this intuition and shows that in fact all the largest component sizes are of the order $n^{2/3}$.

### 2.3.2 Proof ideas

The proof of Theorem 2.2.1 uses a standard functional central limit theorem argument. Indeed we associate a suitable semi-martingale with the graph obtained from an *exploration algorithm* used to explore the connected components of $\mathrm{CM}_n(\mathbf{d})$. The martingale part is then shown to converge to an inhomogeneous Brownian motion, and the drift part is shown to converge to a parabola. The fact that the component sizes can be expressed in terms of the hitting times of the semi-martingale implies the finite-dimensional convergence of the component sizes. The convergence with respect to $\ell^2_{\downarrow}$ is then concluded using size-biased point process arguments formulated by Aldous [8]. Theorem 2.2.2 requires a careful estimate of the tail probability of the distribution of surplus edges when the component size is small and we obtain this using martingale estimates in Lemma 2.4.17. Theorem 2.2.3 is proved by showing that the percolated degree sequence satisfies Assumption 2.1 almost surely. Finally, we prove Theorem 2.2.4 in Section 2.7. The key challenges here are that, for each fixed $n$, the components do not merge according to their component sizes, and that the components do not merge exactly like a multiplicative coalescent over the scaling window. Thus the main theme of the proof lies in approximating the evolution of the component sizes over the percolation scaling window with a suitable dynamic process that is an exact multiplicative coalescent.

## 2.4 Proofs of Theorems 2.2.1 and 2.2.2

### 2.4.1 The exploration process

Let us explore the graph sequentially using a natural approach outlined in [144]. At step $k$, divide the set of half-edges into three groups; sleeping

half-edges $\mathcal{S}_k$, active half-edges $\mathcal{A}_k$, and dead half-edges $\mathcal{D}_k$. The depth-first exploration process can be summarized in the following algorithm:

**Algorithm 2.1** (DFS exploration). At $k = 0$, $\mathcal{S}_k$ contains all the half-edges and $\mathcal{A}_k$, $\mathcal{D}_k$ are empty. While $(\mathcal{S}_k \neq \varnothing$ or $\mathcal{A}_k \neq \varnothing)$ we do the following at stage $k + 1$:

S1 If $\mathcal{A}_k \neq \varnothing$, then take the smallest half-edge $a$ from $\mathcal{A}_k$.

S2 Take the half-edge $b$ from $\mathcal{S}_k$ that is paired to $a$. Suppose $b$ is attached to a vertex $w$ (which is necessarily not discovered yet). Declare $w$ to be discovered, let $r = d_w - 1$ and $b_{w1}, b_{w2}, \ldots b_{wr}$ be the half-edges of $w$ other than $b$. Declare $b_{w1}, b_{w2},\ldots, b_{wr}, b$ to be smaller than all other half-edges in $\mathcal{A}_k$. Also order the half-edges of $w$ among themselves as $b_{w1} > b_{w2} > \cdots > b_{wr} > b$. Now identify $\mathcal{B}_k \subset \mathcal{A}_k \cup \{b_{w1}, b_{w2}, \ldots, b_{wr}\}$ as the collection of all half-edges in $\mathcal{A}_k$ paired to one of the $b_{wi}$'s and the corresponding $b_{wi}$'s. Similarly identify $\mathcal{C}_k \subset \{b_{w1}, b_{w2}, \ldots, b_{wr}\}$ which is the collection of self-loops incident to $w$. Finally, declare $\mathcal{A}_{k+1} = \mathcal{A}_k \cup \{b_{w1}, b_{w2}, \ldots, b_{wr}\} \setminus (\mathcal{B}_k \cup \mathcal{C}_k)$, $\mathcal{D}_{k+1} = \mathcal{D}_k \cup \{a, b\} \cup \mathcal{B}_k \cup \mathcal{C}_k$ and $\mathcal{S}_{k+1} = \mathcal{S}_k \setminus (\{b\} \cup \{b_{w1}, b_{w2}, \ldots, b_{wr}\})$. Go to stage $k + 2$.

S3 If $\mathcal{A}_k = \varnothing$ for some $k$, then take out one half-edge $a$ from $\mathcal{S}_k$ uniformly at random and identify the vertex $v$ incident to it. Declare $v$ to be discovered. Let $r = d_v - 1$ and assume that $a_{v1}, a_{v2},\ldots, a_{vr}$ are the half-edges of $v$ other than $a$ and identify the collection of half-edges involved in self-loops $\mathcal{C}_k$ as in Step 2. Order the half-edges of $v$ as $a_{v1} > a_{v2} > \cdots > a_{vr} > a$. Set $\mathcal{A}_{k+1} = \{a, a_{v1}, a_{v2},\ldots, a_{vr}\} \setminus \mathcal{C}_k$, $\mathcal{D}_{k+1} = \mathcal{D}_k \cup \mathcal{C}_k$, and $\mathcal{S}_{k+1} = \mathcal{S}_k \setminus \{a, a_{v1}, a_{v2},\ldots, a_{vr}\}$. Go to stage $k + 2$.

In words, we explore a new vertex at each stage and throw away all the half-edges involved in a loop/multiple edge/cycle with the vertex set already discovered before proceeding to the next stage. The ordering of the half-edges is such that the connected components of $\mathrm{CM}_n(\mathbf{d})$ are explored in the depth-first way. We call the half-edges of $\mathcal{B}_k \cup \mathcal{C}_k$ *cycle half-edges* because they create loops, cycles or multiple edges in the graph. Let

$$A_k := |\mathcal{A}_k|, \quad c_{(k+1)} := (|\mathcal{B}_k| + |\mathcal{C}_k|)/2, \quad U_k := |\mathcal{S}_k|. \tag{2.6}$$

Let $d_{(j)}$ be the degree of the $j^{\text{th}}$ explored vertex and define the following process:

$$S_n(0) = 0, \quad S_n(i) = \sum_{j=1}^{i} (d_{(j)} - 2 - 2c_{(j)}).$$

The process $\mathbf{S}_n = (S_n(i))_{i \in [n]}$ "encodes the component sizes as lengths of path segments above past minima" as discussed in [8]. Suppose $\mathscr{C}_i$ is the $i^{\text{th}}$ connected component explored by the above exploration process. Define

$$\tau_k = \inf \{i : S_n(i) = -2k\}. \tag{2.7}$$

Then $\mathscr{C}_k$ is discovered between the times $\tau_{k-1} + 1$ and $\tau_k$ and $|\mathscr{C}_k| = \tau_k - \tau_{k-1}$.

### 2.4.2   Size-biased exploration

The vertices are explored in a size-biased manner with sizes proportional to their degrees, i.e., if we denote by $v_{(i)}$ the $i^{\text{th}}$ explored vertex in Algorithm 2.1 and by $d_{(i)}$ the degree of $v_{(i)}$, then $\forall j \in \mathscr{V}_{i-1}$,

$$\mathbb{P}\left(v_{(i)} = j | v_{(1)}, v_{(2)}, ..., v_{(i-1)}\right) = \frac{d_j}{\sum_{k \notin \mathscr{V}_{i-1}} d_k} = \frac{d_j}{\sum_{k \in [n]} d_k - \sum_{k=1}^{i-1} d_{(k)}},$$

where $\mathscr{V}_i$ denotes the first $i$ vertices to be discovered in the above exploration process. The following lemma will be used crucially in the proof of Theorem 2.2.1:

**Lemma 2.4.1.** *Suppose that Assumption 2.1 holds and denote* $\sigma_r = \mathbb{E}[D^r]$ *and* $\mu = \mathbb{E}[D]$. *Then for all* $t > 0$, *as* $n \to \infty$,

$$\sup_{u \leqslant t} \left| n^{-2/3} \sum_{i=1}^{\lfloor n^{2/3} u \rfloor} d_{(i)} - \frac{\sigma_2 u}{\mu} \right| \xrightarrow{\mathbb{P}} 0, \tag{2.8}$$

*and*

$$\sup_{u \leqslant t} \left| n^{-2/3} \sum_{i=1}^{\lfloor n^{2/3} u \rfloor} d_{(i)}^2 - \frac{\sigma_3 u}{\mu} \right| \xrightarrow{\mathbb{P}} 0. \tag{2.9}$$

The proof of this lemma follows from the two lemmas stated below:

**Lemma 2.4.2** ([35, Lemma 8.2]). *Consider a weight sequence* $(w_i)_{i \in [n]}$ *and let* $m = m(n) \leqslant n$ *be increasing with* $n$. *Let* $\{v(i)\}_{i \in [n]}$ *be the size-biased reordering of indices* $[n]$, *where the size of index* $i$ *is* $d_i/\ell_n$. *Define* $\gamma_n = \sum_{i \in [n]} w_i d_i/\ell_n$

*and* $Y(t) = (m\gamma_n)^{-1} \sum_{i=1}^{\lfloor mt \rfloor} w_{\nu(i)}$. *Further, let* $d_{\max} = \max_{i \in [n]} d_i$, *and* $w_{\max} = \max_{i \in [n]} w_i$. *Assume that*

$$\lim_{n \to \infty} m d_{\max} / \ell_n = 0, \quad \text{and} \quad \lim_{n \to \infty} (m\gamma_n)^{-1} w_{\max} = 0. \tag{2.10}$$

*Then, for any* $t > 0$, *as* $n \to \infty$, $\sup_{u \leqslant t} |Y(t) - t| \xrightarrow{\text{P}} 0$.

**Lemma 2.4.3.** *Assumption 2.1 implies*

$$\lim_{k \to \infty} \lim_{n \to \infty} \frac{1}{n} \sum_{j \in [n]} \mathbf{1}_{\{d_j > k\}} d_j^r = 0, \quad r = 1, 2, 3.$$

*For* $r = 3$, *in particular, this implies* $d_{\max}^3 = o(n)$.

### 2.4.3   Estimate of cycle half-edges

The following lemma gives an estimate of the number of cycle half-edges created up to time t. This result is proved in [144] for bounded degrees. In our case, it follows from Lemma 2.4.1 as we show below:

**Lemma 2.4.4.** *For Algorithm 2.1, if* $A_k = |\mathcal{A}_k|$, $B_k := |\mathcal{B}_k|$, *and* $C_k := |\mathcal{C}_k|$, *then*

$$\mathbb{E}\big[B_k | \mathscr{F}_k\big] = (1 + o_\mathbb{P}(1)) \frac{2A_k}{U_k} + O_\mathbb{P}(n^{-2/3}) \tag{2.11}$$

*and*

$$\mathbb{E}\big[C_k | \mathscr{F}_k\big] = O_\mathbb{P}(n^{-1}) \tag{2.12}$$

*uniformly for* $k \leqslant tn^{2/3}$ *and any* $t > 0$, *where* $\mathscr{F}_k$ *is the sigma-field generated by the information revealed up to stage k. Further, all the* $O_\mathbb{P}$ *and* $o_\mathbb{P}$ *terms in (2.11) and (2.12) can be replaced by* $O_E$ *and* $o_E$.

*Proof.* Suppose $U_k := |\mathcal{S}_k|$. First note that by (2.8)

$$\frac{U_k}{n} = \frac{1}{n} \sum_{j \in [n]} d_j - \frac{1}{n} \sum_{j=1}^{k} d_{(j)} = \mathbb{E}[D] + o_\mathbb{P}(1)$$

uniformly over $k \leqslant tn^{2/3}$. Let $a$ be the half-edge that is being explored at stage $k + 1$. Now, each of the $(A_k - 1)$ half-edges of $\mathcal{A}_k \setminus \{a\}$ is equally likely to be paired with a half-edge of $\nu_{(k+1)}$, thus creating two elements of $\mathcal{B}_k$. Also, given $\mathscr{F}_k$ and $\nu_{(k+1)}$, the probability that a half-edge of $\mathcal{A}_k \setminus \{a\}$ is paired to one of the half-edges of $\nu_{(k+1)}$ is $(d_{(k+1)} - 1)/(U_k - 1)$. Therefore,

$$\mathbb{E}\big[B_k | \mathscr{F}_k, \nu_{(k+1)}\big] = 2(A_k - 1) \frac{d_{(k+1)} - 1}{U_k - 1} = 2\big(d_{(k+1)} - 1\big) \frac{A_k}{U_k - 1} - 2 \frac{d_{(k+1)} - 1}{U_k - 1}.$$

Hence,

$$\mathbb{E}\big[B_k|\mathscr{F}_k\big] = 2\mathbb{E}\big[d_{(k+1)}-1|\mathscr{F}_k\big]\frac{A_k}{U_k-1} - 2\frac{\mathbb{E}\big[d_{(k+1)}-1|\mathscr{F}_k\big]}{U_k-1}. \qquad (2.13)$$

Now, using (2.8) and (2.9),

$$\mathbb{E}\big[d_{(k+1)}-1|\mathscr{F}_k\big] = \frac{\sum_{j\notin\mathscr{V}_k} d_j(d_j-1)}{\sum_{j\notin\mathscr{V}_k} d_j} = \frac{\sum_{j\in[n]} d_j^2}{\sum_{j\in[n]} d_j} - 1 + o_{\mathbb{P}}(1) = 1 + o_{\mathbb{P}}(1).$$

uniformly over $k \leqslant tn^{2/3}$, where the last step follows from Assumption 2.1 (iii). Further, using the fact $\mathbb{P}(D=1) > 0$, $U_k \geqslant c_0 n$ for some constant $c_0 > 0$ uniformly over $k \leqslant tn^{2/3}$. Thus, (2.13) gives (2.11). The fact that all the $O_{\mathbb{P}}$, $o_{\mathbb{P}}$ can be replaced by $O_E$, $o_E$ follows from $\sum_{j\in[n]} d_j^r - kd_{max}^r \leqslant \sum_{j\notin\mathscr{V}_k} d_j^r \leqslant \sum_{j\in[n]} d_j^r$ for $r = 1, 2$, together with $d_{max} = o(n^{1/3})$. To prove (2.12), note that

$$\mathbb{E}\big[C_k|\mathscr{F}_k, \nu_{(k+1)}\big] = 2(d_{(k+1)}-2)\frac{d_{(k+1)}-1}{U_k-1}.$$

By Assumption 2.1 and (2.8)

$$\mathbb{E}[d_{(k+1)}^2|\mathscr{F}_k] = \frac{\sum_{j\notin\mathscr{V}_k} d_j^3}{\sum_{j\notin\mathscr{V}_k} d_j} \leqslant \frac{\sum_{j\in[n]} d_j^3}{\sum_{j\in[n]} d_j + o_{\mathbb{P}}(n^{2/3})} = O_{\mathbb{P}}(1),$$

uniformly for $k \leqslant tn^{2/3}$. Therefore,

$$\mathbb{E}\big[C_k|\mathscr{F}_k\big] = O_{\mathbb{P}}(n^{-1})$$

uniformly over $k \leqslant tn^{2/3}$. Again, the $O_{\mathbb{P}}$ term can be replaced by $O_E$, as argued before. $\qquad\square$

### 2.4.4   Key ingredients

For any $\mathbb{D}[0,\infty)$-valued process $\mathbf{X}_n$ define $\bar{X}_n(u) := n^{-1/3}X_n(\lfloor n^{2/3}u\rfloor)$ and $\bar{\mathbf{X}}_n := (\bar{X}_n(u))_{u\geqslant 0}$. The following result is the main ingredient for proving Theorem 2.2.1. Recall the definition of $\mathbf{B}^\lambda_{\mu,\eta}$ from (2.1) with parameters given in (2.4).

**Theorem 2.4.5** (Convergence of the exploration process). *Under Assumption 2.1, as $n \to \infty$,*

$$\bar{\mathbf{S}}_n \xrightarrow{\mathcal{L}} \mathbf{B}^\lambda_{\mu,\eta}$$

*with respect to the* Skorohod $J_1$ *topology.*

As in [121], we will prove this by approximating $\mathbf{S}_n$ by a simpler process defined as

$$s_n(0) = 0, \quad s_n(i) = \sum_{j=1}^{i} (d_{(j)} - 2). \tag{2.14}$$

Note that the difference between the processes $\mathbf{S}_n$ and $\mathbf{s}_n$ is due to the cycles, loops, and multiple-edges encountered during the exploration. Following the approach of [121], it will be enough to prove the following:

**Proposition 2.4.6.** *Under Assumption 2.1, as* $n \to \infty$,

$$\bar{\mathbf{s}}_n \xrightarrow{\mathcal{L}} \mathbf{B}^\lambda_{\mu,\eta}$$

*with respect to the* Skorohod $J_1$ *topology.*

**Remark 6.** It will be shown that the distributions of $\bar{\mathbf{S}}_n$ and $\bar{\mathbf{s}}_n$ are very close as $n \to \infty$, and therefore, Proposition 2.4.6 implies Theorem 2.4.5. This is achieved by proving that we will not see *too many* cycle half-edges up to the time $\lfloor n^{2/3}u \rfloor$ for any fixed $u > 0$.

From here onwards we will look at the continuous versions of the processes $\bar{\mathbf{S}}_n$ and $\bar{\mathbf{s}}_n$ by linearly interpolating between the values at the jump points and write it using the same notation. It is easy to see that these continuous versions differ from their càdlàg versions by at most $n^{-1/3}d_{max} = o(1)$ uniformly on $[0, T]$, for any $T > 0$. Therefore, the convergence in law of the continuous versions implies the convergence in law of the càdlàg versions and vice versa. Before proceeding to show that Theorem 2.4.5 is a consequences of Proposition 2.4.6, we will need to bound the difference of these two processes in a suitable way. We need the following lemma. Recall the definition of $c_{(k+1)} := (B_k + C_k)/2$ from (2.6).

**Lemma 2.4.7.** *Fix* $t > 0$ *and* $M > 0$ *(large). Define the event*

$$E_n(t, M) := \big\{ \max_{s \leqslant t}\{\bar{s}_n(s) - \min_{u \leqslant s} \bar{s}_n(u)\} < M \big\}.$$

*Then*

$$\limsup_{n \to \infty} \sum_{k \leqslant tn^{2/3}} \mathbb{E}\big[c_{(k)}\mathbf{1}_{E_n(t,M)}\big] < \infty.$$

*Proof.* Lemma 2.4.7 is similar to [121, Lemma 6.1]. We add a brief proof here. Note that, for all large $n$, $A_k \leqslant Mn^{1/3}$ on $E_n(t, M)$, because

$$A_k = S_n(k) - \min_{j \leqslant k} S_n(j) = s_n(k) - 2\sum_{j=1}^{k} c_{(j)} - \min_{j \leqslant k} S_n(j) \leqslant s_n(k) - \min_{j \leqslant k} s_n(j),$$

where the last step follows by noting that $\min_{j \leqslant k} s_n(j) \leqslant \min_{j \leqslant k} S_n(j) + 2 \sum_{j=1}^k c_{(j)}$. By Lemma 2.4.4,

$$\mathbb{E}\big[c_{(k)} \mathbf{1}_{E_n(t,M)}\big] \leqslant \frac{Mn^{1/3}}{\mu n} + o(n^{-2/3}) = \frac{M}{\mu} n^{-2/3} + o(n^{-2/3})$$

uniformly for $k \leqslant tn^{2/3}$. Summing over $1 \leqslant k \leqslant tn^{2/3}$ and taking the $\limsup$ completes the proof. $\qquad\square$

The proof of the fact that Theorem 2.4.5 follows from Proposition 2.4.6 and Lemma 2.4.7 is standard (see [121, Section 6.2]) and we skip the proof for the sake of brevity. From here onward the main focus of this section will be to prove Proposition 2.4.6. We use the martingale functional central limit theorem in a similar manner as [8].

*Proof of Proposition 2.4.6.* Let $\{\mathscr{F}_i\}_{i \geqslant 1}$ be the natural filtration defined in Lemma 2.4.4. Recall the definition of $s_n(i)$ from (2.14). By the Doob-Meyer decomposition [124, Theorem 4.10] we can write

$$s_n(i) = M_n(i) + A_n(i), \quad s_n^2(i) = H_n(i) + B_n(i),$$

where

$$M_n(i) = \sum_{j=1}^i \big(d_{(j)} - \mathbb{E}\big[d_{(j)}|\mathscr{F}_{j-1}\big]\big),$$

$$A_n(i) = \sum_{j=1}^i \mathbb{E}\big[d_{(j)} - 2|\mathscr{F}_{j-1}\big],$$

$$B_n(i) = \sum_{j=1}^i \big(\mathbb{E}\big[d_{(j)}^2|\mathscr{F}_{j-1}\big] - \mathbb{E}^2\big[d_{(j)}|\mathscr{F}_{j-1}\big]\big).$$

Recall that for a discrete-time stochastic process $(X_n(i))_{i \geqslant 1}$, we denote $\bar{X}_n(t) = n^{-1/3} X_n(\lfloor tn^{2/3} \rfloor)$. Our result follows from the martingale functional central limit theorem [151, Theorem 2.1] if we can prove the following four conditions: For any $u > 0$,

$$\sup_{s \leqslant u} \big|\bar{A}_n(s) - \lambda s + \frac{\eta s^2}{2\mu^3}\big| \xrightarrow{\mathbb{P}} 0, \qquad (2.16a)$$

$$n^{-1/3} \bar{B}_n(u) \xrightarrow{\mathbb{P}} \frac{\eta}{\mu^2} u, \qquad (2.16b)$$

$$\mathbb{E}\big[\sup_{s \leqslant u} |\bar{M}_n(s) - \bar{M}_n(s-)|^2\big] \to 0, \qquad (2.16c)$$

and

$$n^{-1/3}\mathbb{E}\big[\sup_{s\leqslant u}|\bar{B}_n(s) - \bar{B}_n(s-)|\big] \to 0. \qquad (2.16d)$$

Indeed (2.16a) gives rise to the quadratic drift term of the limiting distribution. Conditions (2.16b), (2.16c), (2.16d) are the same as [151, Theorem 2.1, Condition (ii)]. The facts that the jumps of both the martingale and the quadratic-variation process go to zero and that the quadratic variation process is converging to the quadratic variation of an inhomogeneous Brownian Motion, together imply the convergence of the martingale term. The validation of these conditions is given separately in the subsequent part of this section. □

**Lemma 2.4.8.** *The conditions* (2.16b), (2.16c), *and* (2.16d) *hold.*

*Proof.* Denote by $\sigma_r(n) = \frac{1}{n}\sum_{i\in[n]} d_i^r$, $r = 2, 3$ and $\mu(n) = \frac{1}{n}\sum_{i\in[n]} d_i$. To prove (2.16b), it is enough to prove that

$$n^{-2/3}B_n(\lfloor un^{2/3}\rfloor) \xrightarrow{\mathbb{P}} \frac{\sigma_3\mu - \sigma_2^2}{\mu^2}u.$$

Recall that $\mathbb{E}[d_{(i)}^2|\mathscr{F}_{i-1}] = \sum_{j\notin\mathscr{V}_{i-1}} d_j^3 / \sum_{j\notin\mathscr{V}_{i-1}} d_j$. Further, uniformly over $i \leqslant un^{2/3}$,

$$\sum_{j\notin\mathscr{V}_{i-1}} d_j = \sum_{j\in[n]} d_j + O_{\mathbb{P}}(d_{max}i) = \ell_n + o_{\mathbb{P}}(n). \qquad (2.17)$$

Assume that, without loss of generality, $j \mapsto d_j$ is non-increasing. Then, uniformly over $i \leqslant un^{2/3}$,

$$\left|\sum_{j\notin\mathscr{V}_{i-1}} d_j^3 - n\sigma_3(n)\right| \leqslant \sum_{j=1}^{un^{2/3}} d_j^3. \qquad (2.18)$$

For each fixed $k$,

$$\frac{1}{n}\sum_{j=1}^{un^{2/3}} d_j^3 \leqslant \frac{1}{n}\sum_{j=1}^{un^{2/3}} \mathbf{1}_{\{d_j\leqslant k\}}d_j^3 + \frac{1}{n}\sum_{j\in[n]} \mathbf{1}_{\{d_j>k\}}d_j^3$$

$$\leqslant k^3 un^{-1/3} + \frac{1}{n}\sum_{j\in[n]} \mathbf{1}_{\{d_j>k\}}d_j^3 = o(1),$$

where we first let $n \to \infty$ and then $k \to \infty$ and use Lemma 2.4.3. Therefore, the right-hand side of (2.18) is $o(n)$ and we conclude that, uniformly over $i \leqslant un^{2/3}$,

$$\mathbb{E}\big[d_{(i)}^2|\mathscr{F}_{i-1}\big] = \frac{\sigma_3}{\mu} + o_{\mathbb{P}}(1).$$

A similar argument gives

$$\mathbb{E}\big[d_{(i)}|\mathscr{F}_{i-1}\big] = \frac{\sigma_2}{\mu} + o_{\mathbb{P}}(1),$$

and (2.16b) follows by noting that the error term is $o_{\mathbb{P}}(1)$, uniformly over $i \leqslant un^{2/3}$. The proofs of (2.16c) and (2.16d) are rather short and we present them below. For (2.16c), we bound

$$\mathbb{E}\Big[\sup_{s\leqslant u}|\bar{M}_n(s) - \bar{M}_n(s-)|^2\Big] = n^{-2/3}\mathbb{E}\Big[\sup_{k\leqslant un^{2/3}}|M_n(k) - M_n(k-1)|^2\Big]$$

$$= n^{-2/3}\mathbb{E}\Big[\sup_{k\leqslant un^{2/3}}\big|d_{(k)} - \mathbb{E}[d_{(k)}|\mathscr{F}_{k-1}]\big|^2\Big]$$

$$\leqslant n^{-2/3}\mathbb{E}\Big[\sup_{k\leqslant un^{2/3}}d_{(k)}^2\Big] + n^{-2/3}\mathbb{E}\Big[\sup_{k\leqslant un^{2/3}}\mathbb{E}^2\big[d_{(k)}|\mathscr{F}_{k-1}\big]\Big]$$

$$\leqslant 2n^{-2/3}d_{max}^2.$$

Similarly, (2.16d) gives

$$n^{-1/3}\mathbb{E}\big[\sup_{s\leqslant u}|\bar{B}_n(s) - \bar{B}_n(s-)|^2\big] = n^{-2/3}\mathbb{E}\big[\sup_{k\leqslant un^{2/3}}|B_n(k) - B_n(k-1)|\big]$$

$$= n^{-2/3}\mathbb{E}\big[\sup_{k\leqslant un^{2/3}}\mathrm{var}\big(d_{(k)}|\mathscr{F}_{k-1}\big)\big]$$

$$\leqslant 2n^{-2/3}d_{max}^2,$$

and Conditions (2.16c) and (2.16d) follow from Lemma 2.4.3 using $d_{max} = o(n^{1/3})$. □

Next, we prove Condition (2.16a) which requires some more work. Note that

$$\mathbb{E}\big[d_{(i)} - 2|\mathscr{F}_{i-1}\big] = \frac{\sum_{j\notin\mathscr{V}_{i-1}}d_j(d_j-2)}{\sum_{j\notin\mathscr{V}_{i-1}}d_j}$$

$$= \frac{\sum_{j\in[n]}d_j(d_j-2)}{\sum_{j\in[n]}d_j} - \frac{\sum_{j\in\mathscr{V}_{i-1}}d_j(d_j-2)}{\sum_{j\in[n]}d_j} + \frac{\sum_{j\notin\mathscr{V}_{i-1}}d_j(d_j-2)\sum_{j\in\mathscr{V}_{i-1}}d_j}{\sum_{j\notin\mathscr{V}_{i-1}}d_j\sum_{j\in[n]}d_j}$$

$$= \frac{\lambda}{n^{1/3}} - \frac{\sum_{j\in\mathscr{V}_{i-1}}d_j^2}{\sum_{j\in[n]}d_j} + \frac{\sum_{j\notin\mathscr{V}_{i-1}}d_j^2\sum_{j\in\mathscr{V}_{i-1}}d_j}{\sum_{j\notin\mathscr{V}_{i-1}}d_j\sum_{j\in[n]}d_j} + o(n^{-1/3}), \tag{2.19}$$

where the last step follows from Assumption 2.1 (iii). Therefore,

$$A_n(k) = \sum_{i=1}^k \mathbb{E}\big[d_{(i)} - 2|\mathscr{F}_{i-1}\big]$$

$$= \frac{k\lambda}{n^{1/3}} - \sum_{i=1}^k \frac{\sum_{j\in\mathscr{V}_{i-1}} d_j^2}{\sum_{j\in[n]} d_j} + \sum_{i=1}^k \frac{\sum_{j\notin\mathscr{V}_{i-1}} d_j^2 \sum_{j\in\mathscr{V}_{i-1}} d_j}{\sum_{j\notin\mathscr{V}_{i-1}} d_j \sum_{j\in[n]} d_j} + o(kn^{-1/3}).$$

$$(2.20)$$

The following lemma estimates the sums on the right-hand side of (2.20):

**Lemma 2.4.9.** *For all $u > 0$, as $n \to \infty$,*

$$\sup_{s\leqslant u} \left| n^{-1/3} \sum_{i=1}^{\lfloor sn^{2/3}\rfloor} \sum_{j=1}^{i-1} \frac{d_{(j)}^2}{\ell_n} - \frac{\sigma_3 s^2}{2\mu^2} \right| \xrightarrow{\mathbb{P}} 0 \qquad (2.21)$$

*and*

$$\sup_{s\leqslant u} \left| n^{-1/3} \sum_{i=1}^{\lfloor sn^{2/3}\rfloor} \sum_{j=1}^{i-1} \frac{d_{(j)}}{\ell_n} - \frac{\sigma_2 s^2}{2\mu^2} \right| \xrightarrow{\mathbb{P}} 0. \qquad (2.22)$$

*Consequently,*

$$\sup_{s\leqslant u} \left| n^{-1/3} \sum_{i=1}^{\lfloor sn^{2/3}\rfloor} \frac{\sum_{j\notin\mathscr{V}_{i-1}} d_j^2 \sum_{j\in\mathscr{V}_{i-1}} d_j}{\sum_{j\notin\mathscr{V}_{i-1}} d_j \sum_{j\in[n]} d_j} - \frac{\sigma_2^2 s^2}{2\mu^3} \right| \xrightarrow{\mathbb{P}} 0. \qquad (2.23)$$

*Proof.* Notice that

$$\sup_{s\leqslant u} \left| n^{-1/3} \sum_{i=1}^{\lfloor sn^{2/3}\rfloor} \sum_{j=1}^{i-1} \frac{d_{(j)}^2}{\ell_n} - \frac{\sigma_3 s^2}{2\mu^2} \right| = \sup_{k\leqslant un^{2/3}} \left| n^{-1/3} \sum_{i=1}^k \sum_{j=1}^{i-1} \frac{d_{(j)}^2}{\ell_n} - \frac{\sigma_3 k^2}{2\mu^2 n^{4/3}} \right|$$

$$\leqslant \frac{1}{\ell_n} \sup_{k\leqslant un^{2/3}} \left| n^{-1/3} \sum_{i=1}^k \Big( \sum_{j=1}^{i-1} d_{(j)}^2 - \frac{\sigma_3(i-1)}{\mu} \Big) \right|$$

$$+ \sup_{k\leqslant un^{2/3}} \left| \frac{k\sigma_3}{2\mu\ell_n n^{1/3}} \right| + \sup_{k\leqslant un^{2/3}} \left| \frac{k^2\sigma_3}{2\mu\ell_n n^{1/3}} - \frac{k^2\sigma_3}{2\mu^2 n^{4/3}} \right|$$

$$\leqslant \frac{1}{\ell_n} n^{-1/3} u n^{2/3} \sup_{i\leqslant un^{2/3}} \left| \sum_{j=1}^i d_{(j)}^2 - \frac{\sigma_3 i}{\mu} \right| + o(1) + \frac{\sigma_3 n^{-1/3}}{2\mu} \left| \frac{1}{\ell_n} - \frac{1}{n\mu} \right| u^2 n^{4/3}$$

$$\leqslant \frac{u}{\mu + o(1)} \sup_{s\leqslant u} \left| \Big( n^{-2/3} \sum_{j=1}^{\lfloor sn^{2/3}\rfloor} d_{(j)}^2 - \frac{\sigma_3 s}{\mu} \Big) \right| + o(1).$$

and (2.21) follows from (2.9) in Lemma 2.4.1. The proof of (2.22) is similar and it follows from (2.8). We now show (2.23). Recall that $\sigma_2(n) = \frac{1}{n}\sum_{i\in[n]} d_i^2$ and observe

$$\frac{1}{n}\sum_{j\notin\mathscr{V}_{i-1}} d_j^2 = \sigma_2(n) - \frac{1}{n}\sum_{j\in\mathscr{V}_{i-1}} d_j^2 = \sigma_2(n) + o_{\mathbb{P}}(1)$$

uniformly over $i \leqslant un^{2/3}$ where we use Lemma 2.4.1 to conclude the uniformity. Similarly, (2.17) implies that $\sum_{j\notin\mathscr{V}_{i-1}} d_j = \ell_n + o_{\mathbb{P}}(n)$ uniformly over $i \leqslant un^{2/3}$. Therefore,

$$n^{-1/3}\sum_{i=1}^{k}\frac{\sum_{j\notin\mathscr{V}_{i-1}} d_j^2 \sum_{j\in\mathscr{V}_{i-1}} d_j}{\sum_{j\notin\mathscr{V}_{i-1}} d_j \sum_{j\in[n]} d_j} = \frac{n\sigma_2(n) + o_{\mathbb{P}}(n)}{\ell_n + o_{\mathbb{P}}(n)} n^{-1/3}\sum_{i=1}^{k}\frac{\sum_{j\in\mathscr{V}_{i-1}} d_j}{\ell_n}$$

and Assumption 2.1, combined with (2.22), completes the proof. $\qquad\square$

**Lemma 2.4.10.** *Condition* (2.16a) *holds.*

*Proof.* The proof follows by using Lemma 2.4.9 in (2.20). $\qquad\square$

### 2.4.5   Finite-dimensional convergence of the ordered component sizes

Note that the convergence of the exploration process in Theorem 2.4.5 implies that, for any large $T > 0$, the $k$-largest components explored up to time $Tn^{2/3}$ converge to the $k$-largest excursions above past minima of $\mathbf{B}_{\mu,\eta}^{\lambda}$ up to time $T$. Therefore, we can conclude the finite dimensional convergence of the ordered components sizes in the whole graph if we can show that the large components are explored *early* by the exploration process. The following lemma formalizes the above statement:

**Lemma 2.4.11.** *Let $\mathscr{C}_{\max}^{\geqslant T}$ denote the largest component which is started exploring after time $Tn^{2/3}$ in Algorithm 2.1. Then, for any $\delta > 0$,*

$$\lim_{T\to\infty} \limsup_{n\to\infty} \mathbb{P}\left(|\mathscr{C}_{\max}^{\geqslant T}| > \delta n^{2/3}\right) = 0. \qquad (2.24)$$

Let us first state the two main ingredients to complete the proof of Lemma 2.4.11:

**Lemma 2.4.12** ([111, Lemma 5.2])**.** *Consider $CM_n(\mathbf{d})$ with $\nu_n < 1$ and let $\mathscr{C}(V_n)$ denote the component containing the vertex $V_n$, where $V_n$ is a vertex chosen uniformly at random independently of the graph $CM_n(\mathbf{d})$. Then,*

$$\mathbb{E}\left[|\mathscr{C}(V_n)|\right] \leqslant 1 + \frac{\mathbb{E}[D_n]}{1 - \nu_n}.$$

**Lemma 2.4.13.** *Define,* $\nu_{n,i} = \sum_{j \notin \mathscr{V}_{i-1}} d_j(d_j - 1)/\sum_{j \notin \mathscr{V}_{i-1}} d_j$. *There exists some constant* $C_0 > 0$ *such that for any* $T > 0$,

$$\nu_{n,Tn^{2/3}} = \nu_n - C_0 T n^{-1/3} + o_{\mathbb{P}}(n^{-1/3}). \tag{2.25}$$

*Proof.* Using a similar split up as in (2.19), we have

$$\nu_{n,i} = \nu_n + \frac{\sum_{j \in \mathscr{V}_{i-1}} d_j(d_j - 1)}{\ell_n} - \frac{\sum_{j \notin \mathscr{V}_{i-1}} d_j(d_j - 1) \sum_{j \in \mathscr{V}_{i-1}} d_j}{\ell_n \sum_{j \notin \mathscr{V}_{i-1}} d_j}. \tag{2.26}$$

Now, (2.8) and (2.9) give that, uniformly over $i \leqslant Tn^{2/3}$,

$$\frac{\sum_{j \notin \mathscr{V}_{i-1}} d_j(d_j - 1)}{\sum_{j \notin \mathscr{V}_{i-1}} d_j} = \frac{\sum_{j \in [n]} d_j(d_j - 1) + o_{\mathbb{P}}(n^{2/3})}{\sum_{j \in [n]} d_j + o_{\mathbb{P}}(n^{2/3})} = 1 + o_{\mathbb{P}}(n^{-1/3}),$$

$$\sum_{j \in \mathscr{V}_{i-1}} d_j(d_j - 2) = \left(\frac{\sigma_3}{\mu} - 2\right)(i - 1) + o_{\mathbb{P}}(n^{2/3}).$$

Further, note that $\sigma_3 - 2\mu = \mathbb{E}[D(D-1)(D-2)] + \mathbb{E}[D(D-2)] > 0$, by Assumption 2.1 (iii), and (iv). Therefore, (2.26) gives (2.25). □

*Proof of Lemma 2.4.11.* Let $i_T := \inf\{\tau_k : \tau_k > Tn^{2/3}\}$, where $\tau_k$ is defined by (2.7). Thus, $i_T$ denotes the first time we finish exploring a component after time $Tn^{2/3}$. Note that, conditional on the explored vertices up to time $i_T$, the remaining graph $\bar{\mathcal{G}}$ is still a configuration model. Let $\bar{\nu}_n = \sum_{i \in \bar{\mathcal{G}}} d_i(d_i - 1)/\sum_{i \in \bar{\mathcal{G}}} d_i$ be the criticality parameter of $\bar{\mathcal{G}}$. Then, using (2.25), we can conclude that

$$\bar{\nu}_n \leqslant \nu_n - C_0 T n^{-1/3} + o_{\mathbb{P}}(n^{-1/3}). \tag{2.28}$$

Thus, as we explore more vertices, the graph becomes more subcritical. Take $T > 0$ such that $\lambda - C_0 T < 0$. Thus, with high probability, $\bar{\nu}_n < 1$. Denote the component corresponding to a randomly chosen vertex from $\bar{\mathcal{G}}$ by $\mathscr{C}^{\geqslant T}(V_n)$, and the $i^{\text{th}}$ largest component of $\bar{\mathcal{G}}$ by $\mathscr{C}^{\geqslant T}_{(i)}$. Also, let $\bar{\mathbb{P}}$ denote the probability measure conditioned on $\mathscr{F}_{i_T}$, and let $\bar{\mathbb{E}}$ denote the corresponding expectation. Now, for any $\delta > 0$,

$$\bar{\mathbb{P}}\left(\sum_{i \geqslant 1} |\mathscr{C}^{\geqslant T}_{(i)}|^2 > \delta^2 n^{4/3}\right) \leqslant \frac{1}{\delta^2 n^{4/3}} \sum_{i \geqslant 1} \bar{\mathbb{E}}\left(|\mathscr{C}^{\geqslant T}_{(i)}|^2\right)$$

$$\leqslant \frac{1}{\delta^2 n^{1/3}} \bar{\mathbb{E}}\left(|\mathscr{C}^{\geqslant T}(V_n)|\right) \leqslant \frac{1}{\delta^2(-\lambda + C_0 T + o_{\mathbb{P}}(1))},$$

where the second step follows from the Markov inequality and the last step follows by combining Lemma 2.4.12 and (2.28). Noting that $\bar{\nu}_n < 1$ with high

probability, we get

$$\limsup_{n \to \infty} \mathbb{P}\left( |\mathscr{C}_{\max}^{\geqslant T}| > \delta n^{2/3} \right) \leqslant \frac{C}{\delta^2 T},$$

for some constant $C > 0$ and large $T > 0$ and the proof follows.    □

**Theorem 2.4.14.** *The convergence in Theorem 2.2.1 holds with respect to the product topology.*

*Proof.* The proof follows from Theorem 2.4.5 and Lemma 2.4.11.    □

### 2.4.6    Proof of Theorem 2.2.1

The proof of Theorem 2.2.1 follows using a similar argument as [8, Section 3.3]. However, the proof is a bit tricky since the components are explored in a size-biased manner with sizes being the total degree in the components (not the component sizes as in [8]). For a sequence of random variables $\mathbf{Y} = (Y_i)_{i \geqslant 1}$ satisfying $\sum_{i \geqslant 1} Y_i^2 < \infty$ almost surely, define $\boldsymbol{\xi} := (\xi_i)_{i \geqslant 1}$ such that $\xi_i | \mathbf{Y} \sim \text{Exp}(Y_i)$ and the coordinates of $\boldsymbol{\xi}$ are independent conditional on $\mathbf{Y}$. For $a \geqslant 0$, let $\mathscr{S}(a) := \sum_{\xi_i \leqslant a} Y_i$. Then the *size biased point process* is defined to be the random collection of points $\Xi := \{(\mathscr{S}(\xi_i), Y_i)\}_{i \geqslant 1}$ (see [8, Section 3.3]). We will use Lemma 8, Lemma 14 and Proposition 15 from [8]. Let $\mathfrak{C} := \{\mathscr{C} : \mathscr{C} \text{ is a component of } CM_n(\mathbf{d})\}$. Consider the collection $\boldsymbol{\xi} := (\xi(\mathscr{C}))_{\mathscr{C} \in \mathfrak{C}}$ such that conditional on $(\sum_{k \in \mathscr{C}} d_k, |\mathscr{C}|)_{\mathscr{C} \in \mathfrak{C}}$, $\xi(\mathscr{C})$ has an exponential distribution with rate $n^{-2/3} \sum_{k \in \mathscr{C}} d_k$ independently over $\mathscr{C}$. Then the order in which Algorithm 2.1 explores the components can be obtained by ordering the components according to their $\xi$-value. Recall that $\mathscr{C}_i$ denotes the $i^{th}$ explored component by Algorithm 2.1 and let $D_i := \sum_{k \in \mathscr{C}_i} d_k$. Define the size-biased point process

$$\Xi_n := \left( n^{-2/3} \sum_{j=1}^{i} D_j, \ n^{-2/3} D_i \right)_{i \geqslant 1}.$$

Also define the point processes

$$\Xi_n' := \left( n^{-2/3} \sum_{j=1}^{i} |\mathscr{C}_j|, \ n^{-2/3} |\mathscr{C}_i| \right)_{i \geqslant 1},$$

$$\Xi_\infty := \left\{ (l(\gamma), |\gamma|) : \gamma \text{ an excursion of } \mathbf{B}_{\mu,\eta}^\lambda \right\},$$

where we recall that $l(\gamma)$ are the left endpoints of the excursions of $\mathbf{B}_{\mu,\eta}^\lambda$ and $|\gamma|$ is the length of the excursion $\gamma$ (see (2.2)). Note that $\Xi_n'$ is not a size-biased

point process. However, applying [8, Lemma 8] and Theorem 2.4.5, we get $\Xi'_n \xrightarrow{\mathcal{L}} \Xi_\infty$. We claim that

$$\Xi_n \xrightarrow{\mathcal{L}} 2\Xi_\infty. \tag{2.29}$$

To verify the claim, note that (2.8) and Assumption 2.1 (iii) together imply, for any $t > 0$,

$$\sup_{u \leqslant t} \left| n^{-2/3} \sum_{i=1}^{\lfloor un^{2/3} \rfloor} d_{(i)} - \frac{\sigma_2}{\mu} u \right| = \sup_{u \leqslant t} \left| n^{-2/3} \sum_{i=1}^{\lfloor un^{2/3} \rfloor} d_{(i)} - 2u \right| \xrightarrow{\mathbb{P}} 0, \tag{2.30}$$

since $\sigma_2/\mu = \mathbb{E}[D^2]/\mathbb{E}[D] = 2$. Thus, (2.29) follows using (2.30). Now, the point process $2\Xi_\infty$ satisfies all the conditions of [8, Proposition 15] as shown by Aldous. Thus, [8, Lemma 14] gives

$$\{D_{(i)}\}_{i \geqslant 1} \text{ is tight in } \ell^2_\downarrow. \tag{2.31}$$

This implies that $\left( n^{-2/3} |\mathscr{C}_{(i)}| \right)_{i \geqslant 1}$ is tight in $\ell^2_\downarrow$ by simply observing that $|\mathscr{C}_i| \leqslant \sum_{k \in \mathscr{C}_i} d_k + 1$. Therefore, the proof of Theorem 2.2.1 is complete using Theorem 2.4.14. □

### 2.4.7   Proof of Theorem 2.2.2

The proof of Theorem 2.2.2 is completed in two separate lemmas below. In Lemma 2.4.15 we first show that the convergence in Theorem 2.2.2 holds with respect to the $\ell^2_\downarrow \times \mathbb{N}^\infty$ topology. The tightness of $(\mathbf{Z}_n)_{n \geqslant 1}$ with respect to the $\mathbb{U}^0_\downarrow$ topology is ensured in Lemma 2.4.16.

**Lemma 2.4.15.** *Let $N^\lambda_n(k)$ be the number of surplus edges discovered up to time $k$ and $\bar{N}^\lambda_n(u) = N^\lambda_n(\lfloor un^{2/3} \rfloor)$. Then, as $n \to \infty$,*

$$\bar{N}^\lambda_n \xrightarrow{d} \mathbf{N}^\lambda,$$

*where $\mathbf{N}^\lambda$ is defined in (2.3).*

*Proof.* Recall the definitions of $a$, $b$, $\mathcal{A}_k$, $\mathcal{B}_k$, $\mathcal{C}_k$, $\mathcal{S}_k$ from Section 2.4.1. Recall also that $A_k := |\mathcal{A}_k|$, $B_k := |\mathcal{B}_k|$, $C_k := |\mathcal{C}_k|$, $U_k := |\mathcal{S}_k|$, $c_{(k+1)} := (|\mathcal{B}_k| + |\mathcal{C}_k|)/2$ from Section 2.4.1. Notice that $A_k = S_n(k) - \min_{j \leqslant k} S_n(j)$. From Lemma 2.4.4, we can conclude that, uniformly over $k \leqslant un^{2/3}$,

$$\mathbb{E}\left[ c_{(k+1)} | \mathscr{F}_k \right] = \frac{A_k}{\mu n} + O_{\mathbb{P}}(n^{-1}). \tag{2.32}$$

The counting process $\mathbf{N}_n^\lambda$ has conditional intensity (conditioned on $\mathscr{F}_{k-1}$) given by (2.32). Writing the conditional intensity in (2.32) in terms of $\bar{S}_n$, we get that the conditional intensity of the re-scaled process $\bar{\mathbf{N}}_n^\lambda$ is given by

$$\frac{1}{\mu}[\bar{S}_n(u) - \min_{\tilde{u} \leqslant u} \bar{S}_n(\tilde{u})] + o_{\mathbb{P}}(1).$$

Denote by $\bar{W}_n(u) := \bar{S}_n(u) - \min_{\tilde{u} \leqslant u} \bar{S}_n(\tilde{u})$ which is the reflected version of $\bar{S}_n$. By Theorem 2.2.1,

$$\bar{W}_n \xrightarrow{d} \mathbf{W}^\lambda,$$

where $\mathbf{W}^\lambda$ is defined in (2.2). Therefore, we can assume that there exists a probability space such that $\bar{W}_n \to \mathbf{W}^\lambda$ almost surely. Using [128, Theorem 1; Chapter 5.3], and the continuity of the sample paths of $\mathbf{W}^\lambda$, we conclude the proof. $\qquad\square$

**Lemma 2.4.16.** *The vector $(\mathbf{Z}_n)_{n \geqslant 1}$ is tight with respect to the $\mathbb{U}_\downarrow^0$ topology.*

The proof of Lemma 2.4.16 makes use of the following crucial estimate of the probability that a component with small size has a very large number of surplus edges:

**Lemma 2.4.17.** *Assume that $\lambda < 0$. Let $V_n$ denote a vertex chosen uniformly at random, independent of the graph $\mathrm{CM}_n(\mathbf{d})$ and let $\mathscr{C}(V_n)$ denote the component containing $V_n$. Let $\delta_k = \delta k^{-0.12}$. Then, for $\delta > 0$ (small),*

$$\mathbb{P}\left(\mathrm{SP}(\mathscr{C}(V_n)) \geqslant K, |\mathscr{C}(V_n)| \in (\delta_K n^{2/3}, 2\delta_K n^{2/3})\right) \leqslant \frac{C\sqrt{\delta}}{n^{1/3} K^{1.1}},$$

*where $C$ is a fixed constant independent of $n, \delta, K$.*

*Proof of Lemma 2.4.16.* To simplify the notation, we write $Y_i^n = n^{-2/3}|\mathscr{C}_{(i)}|$ and $N_i^n = \#\{$surplus edges in $\mathscr{C}_{(i)}\}$. Let $Y_i, N_i$ denote the distributional limits of $Y_i^n$ and $N_i^n$ respectively. Recall from Remark 1 that $\mathbf{Z}(\lambda)$ is almost surely $\mathbb{U}_\downarrow^0$-valued. Using Lemma 2.4.15, the proof of Lemma 2.4.16 is complete if we can show that, for any $\eta > 0$

$$\lim_{\varepsilon \to 0} \limsup_{n \to \infty} \mathbb{P}\left(\sum_{Y_i^n \leqslant \varepsilon} Y_i^n N_i^n > \eta\right) = 0. \tag{2.33}$$

First, consider the case $\lambda < 0$. For every $\eta, \varepsilon > 0$ sufficiently small

$$\mathbb{P}\Big( \sum_{Y_i^n \leqslant \varepsilon} Y_i^n N_i^n > \eta \Big) \leqslant \frac{1}{\eta} \mathbb{E}\Big[ \sum_{i=1}^{\infty} Y_i^n N_i^n \mathbf{1}_{\{Y_i^n \leqslant \varepsilon\}} \Big]$$

$$= \frac{n^{-2/3}}{\eta} \mathbb{E}\Big[ \sum_{i=1}^{\infty} |\mathscr{C}_{(i)}| N_i^n \mathbf{1}_{\{|\mathscr{C}_{(i)}| \leqslant \varepsilon n^{2/3}\}} \Big] = \frac{n^{1/3}}{\eta} \mathbb{E}\Big[ SP(\mathscr{C}(V_n)) \mathbf{1}_{\{|\mathscr{C}(V_n)| \leqslant \varepsilon n^{2/3}\}} \Big]$$

$$= \frac{n^{1/3}}{\eta} \sum_{k=1}^{\infty} \sum_{i \geqslant \log_2(1/(k^{0.12}\varepsilon))} \mathbb{P}\Big( SP(\mathscr{C}(V_n)) \geqslant k, |\mathscr{C}(V_n)| \in \Big( \frac{n^{2/3}}{2^{i+1}k^{0.12}}, \frac{n^{2/3}}{2^i k^{0.12}} \Big] \Big)$$

$$\leqslant \frac{C}{\eta} \sum_{k=1}^{\infty} \frac{1}{k^{1.1}} \sum_{i \geqslant \log_2(1/(k^{0.12}\varepsilon))} 2^{-(1/2)i} \leqslant \frac{C}{\eta} \sum_{k=1}^{\infty} \frac{\sqrt{\varepsilon}}{k^{1.04}} = O(\sqrt{\varepsilon}),$$

where we have used Lemma 2.4.17. Therefore, (2.33) holds when $\lambda < 0$. Now consider the case $\lambda > 0$. For $T > 0$ (large), let

$$\mathcal{K}_n := \{i : Y_i^n \leqslant \varepsilon, \mathscr{C}_{(i)} \text{ is explored before } Tn^{2/3}\}.$$

Then, by applying the Cauchy-Schwarz inequality,

$$\sum_{i \in \mathcal{K}_n} Y_i^n N_i^n \leqslant \Big( \sum_{i \in \mathcal{K}_n} (Y_i^n)^2 \Big)^{1/2} \times \Big( \sum_{i \in \mathcal{K}_n} (N_i^n)^2 \Big)^{1/2}$$

$$\leqslant \Big( \sum_{i \in \mathcal{K}_n} (Y_i^n)^2 \Big)^{1/2} \times (\# \text{ surplus edges explored before } Tn^{2/3})$$

(2.34)

For the case $\lambda > 0$, we can use similar ideas as the proof of Lemma 2.4.11, i.e., we can run the exploration process till $Tn^{2/3}$ and the unexplored graph becomes a configuration model with negative criticality parameter for large $T > 0$, by (2.25). Thus, the proof can be completed using (2.34), the $\ell_\downarrow^2$ convergence of the component sizes given by Theorem 2.2.1 and Lemma 2.4.15, and the proof for the case $\lambda < 0$. $\qquad \square$

*Proof of Lemma 2.4.17.* To complete the proof of Lemma 2.4.17, we will use martingale techniques coupled with Lemma 2.4.12. Fix $\delta > 0$ (small). First we describe another way of exploring $\mathscr{C}(V_n)$ which turns out to be convenient to work with.

**Algorithm 2.2** (Exploring components of uniform vertices). Consider the following exploration of $\mathscr{C}(V_n)$:

(S0) Initialize all half-edges to be alive. Choose a vertex from $[n]$ uniformly at random and declare all its half-edges active.

(S1) In the next step, take any active half-edge and pair it uniformly with another alive half-edge. Kill these paired half-edges. Declare all the half-edges corresponding to the new vertex (if any) active. Keep repeating (S1) until the set of active half-edges is empty.

Unlike Algorithm 2.1, we need not see a new vertex at each stage and we explore only two half-edges at each stage. In this proof, $\mathscr{F}_l$ denotes the sigma-field containing information revealed up to stage $l$ by Algorithm 2.2 and $\mathscr{V}_l$ denotes the vertex set discovered up to time $l$. Recall that we denote by $D_n$ the degree of $V_n$. Define the exploration process $\mathbf{s}'_n$ by,

$$s'_n(0) = D_n, \quad s'_n(l) = \sum_{i \in [n]} d_i \mathfrak{I}^n_i(l) - 2l,$$

where $\mathfrak{I}^n_i(l) = \mathbf{1}\{i \in \mathscr{V}_l\}$. Therefore, $s'_n(l)$ counts the number of active half-edges at time $l$, until $\mathscr{C}(V_n)$ is explored. Note that $\mathscr{C}(V_n)$ is explored when $\mathbf{s}'_n$ hits zero and the hitting time to zero gives the number of edges in $\mathscr{C}(V_n)$, since exactly one edge is being explored at each time step. We will use a generic constant $C$ to denote a positive constant that can be different in different equations. For $H > 0$, let

$$\gamma := \inf\{l \geqslant 1 : s'_n(l) \geqslant H \text{ or } s'_n(l) = 0\} \wedge 2\delta n^{2/3}.$$

Note that

$$\mathbb{E}[s'_n(l+1) - s'_n(l)|\mathscr{F}_l] = \sum_{i \in [n]} d_i \mathbb{P}\left(i \in \mathscr{V}_{l+1}|\mathscr{F}_l, \mathfrak{I}^n_i(l) = 0\right) - 2$$

$$= \frac{\sum_{i \notin \mathscr{V}_l} d_i^2}{\ell_n - 2l - 1} - 2 \leqslant \frac{\sum_{i \in [n]} d_i^2}{\ell_n - 2l - 1} - 2$$

$$= \frac{\lambda}{n^{1/3}} + o(n^{-1/3}) + \frac{2l+1}{\ell_n - 2l - 1} \times \frac{\sum_{i \in [n]} d_i^2}{\ell_n} \leqslant 0$$

uniformly over $l \leqslant 2\delta n^{2/3}$ for all small $\delta > 0$ and large $n$, where the last step follows from the fact that $\lambda < 0$. Therefore, $\{s'_n(l)\}_{l=1}^{2\delta n^{2/3}}$ is a super-martingale. The optional stopping theorem now implies

$$\mathbb{E}[D_n] \geqslant \mathbb{E}\left[s'_n(\gamma)\right] \geqslant H\mathbb{P}\left(s'_n(\gamma) \geqslant H\right).$$

Thus,

$$\mathbb{P}\left(s'_n(\gamma) \geqslant H\right) \leqslant \frac{\mathbb{E}[D_n]}{H}. \tag{2.35}$$

We put $H = n^{1/3}K^{1.1}/\sqrt{\delta}$. To simplify the notation, we write $s_n'[0,t] \in A$ to denote that $s_n'(l) \in A$, for all $l \in [0,t]$. Notice that, for $K \geqslant 1$,

$$\mathbb{P}\left(SP(\mathscr{C}(V_n)) \geqslant K, |\mathscr{C}(V_n)| \in (\delta_K n^{2/3}, 2\delta_K n^{2/3})\right)$$

$$\leqslant \mathbb{P}\left(s_n'(\gamma) \geqslant H\right) \tag{2.36}$$

$$+ \mathbb{P}\left(SP(\mathscr{C}(V_n)) \geqslant K, s_n'[0, 2\delta_K n^{2/3}] < H, s_n'[0, \delta_K n^{2/3}] > 0\right).$$

Here we have used the fact that if there is at least one surplus edge in $\mathscr{C}(V_n)$, the number of edges in $\mathscr{C}(V_n)$ is at least $\mathscr{C}(V_n)$. Therefore, $|\mathscr{C}(V_n)| > \delta_K n^{2/3}$ implies $s_n'[0, \delta_K n^{2/3}] > 0$. Let us denote the event that surplus edges appear at times $l_1, \ldots, l_K$, $s_n'[0, 2\delta_K n^{2/3}] < H$, and $s_n'[0, \delta_K n^{2/3}] > 0$ by $SPB(l_1, \ldots, l_K)$. Now,

$$\mathbb{P}\left(SP(\mathscr{C}(V_n)) \geqslant K, s_n'[0, 2\delta_K n^{2/3}] < H, s_n'[0, \delta_K n^{2/3}] > 0\right)$$

$$\leqslant \sum_{1 \leqslant l_1 < \cdots < l_K \leqslant 2\delta_K n^{2/3}} \mathbb{P}\left(SPB(l_1, \ldots, l_K)\right)$$

$$= \sum_{1 \leqslant l_1 < \cdots < l_K \leqslant 2\delta_K n^{2/3}} \mathbb{E}\left[\mathbf{1}\{0 < s_n'[0, l_K - 1] < H, \mathbf{SP}(l_K - 1) = K - 1\}Y\right],$$

where

$$Y = \mathbb{P}\left(K^{th} \text{ surplus occurs at } l_K, s_n'[l_K, 2\delta_K n^{2/3}] < H, s_n'[l_K, \gamma] > 0 \mid \mathscr{F}_{l_K - 1}\right)$$

$$\leqslant \frac{CK^{1.1}n^{1/3}}{l_n\sqrt{\delta}} \leqslant \frac{CK^{1.1}}{n^{2/3}\sqrt{\delta}}.$$

Therefore, using induction,

$$\mathbb{P}\left(SP(\mathscr{C}(V_n)) \geqslant K, s_n'[0, 2\delta_K n^{2/3}] < H, s_n'[0, \delta_K n^{2/3}] > 0\right)$$

$$\leqslant C\left(\frac{K^{1.1}}{\sqrt{\delta}n^{2/3}}\right)^K \frac{(2\delta n^{2/3})^{K-1}}{K^{0.12(K-1)}(K-1)!} \sum_{l_1=1}^{2\delta_K n^{2/3}} \mathbb{P}\left(|\mathscr{C}(V_n)| \geqslant l_1\right) \tag{2.37}$$

$$\leqslant C\frac{\delta^{K/2}}{K^{1.1}n^{2/3}}\mathbb{E}\left[|\mathscr{C}(V_n)|\right],$$

where we have used the fact that

$$\#\{1 \leqslant l_2 < \cdots < l_k \leqslant 2\delta n^{2/3}\} \leqslant (2\delta n^{2/3})^{K-1}/(K-1)!$$

and the Stirling approximation for $(K-1)!$ in the last step. Since $\lambda < 0$, we can use Lemma 2.4.12 to conclude that for all sufficiently large $n$

$$\mathbb{E}\left[|\mathscr{C}(V_n)|\right] \leqslant Cn^{1/3},$$

for some constant $C > 0$ and we get the desired bound for (2.36). The proof of Lemma 2.4.17 is now complete by applying (2.35) and (2.37) in (2.36). $\quad\square$

## 2.5    Degree distribution within components

In this section, we compute the number of vertices of degree $k$ in each connected component at criticality. This will be useful in Sections 2.6 and 2.7. Such an estimate was proved in [114, Theorem 2.4] for supercritical graphs under stronger moment assumptions.

**Lemma 2.5.1.** *Denote by* $N_k(t)$ *the number of vertices of degree $k$ discovered up to time* $t$. *For any* $t > 0$, *uniformly over* $k$,

$$\sup_{u \leqslant t} \left| n^{-2/3} N_k(un^{2/3}) - \frac{kn_k}{\ell_n} u \right| = O_{\mathbb{P}}((kn^{1/3})^{-1}).$$

*Proof.* By setting $w_i = \mathbf{1}_{\{d_i = k\}}$ in Lemma 2.4.2 we can directly conclude that

$$\sup_{u \leqslant t} \left| n^{-2/3} N_k(un^{2/3}) - \frac{kn_k}{\ell_n} u \right| \xrightarrow{\mathbb{P}} 0.$$

However, one can repeat same arguments as leading to the proof of Lemma 2.4.2 and obtain that

$$\mathbb{P}\left( \sup_{u \leqslant t} \left| n^{-2/3} N_k(un^{2/3}) - \frac{kn_k}{\ell_n} u \right| > \frac{A}{kn^{1/3}} \right)$$
$$\leqslant \frac{3\left( k^3 s^2 \frac{r_k}{(\mathbb{E}[D])^2} + \sqrt{s \frac{k^3 r_k}{\mathbb{E}[D]}} \right)}{A} + o(1).$$

Now, we can use the finite third-moment assumption to conclude that the numerator in the right hand side can be taken to be uniform over $k$. Thus, the proof follows. $\quad\square$

Define $\nu_k(G) :=$ the number of vertices of degree $k$ in the connected graph $G$. As a corollary to Lemma 2.5.1 and (2.24), we can deduce that

$$\nu_k(\mathscr{C}_{(j)}) = \frac{kr_k}{\mathbb{E}[D]} |\mathscr{C}_{(j)}| + O_{\mathbb{P}}((k^{-1}n^{1/3})). \tag{2.38}$$

Moreover, the following holds: Let $\mathrm{ord}(\mathbf{x})$ denote the vector with elements of $\mathbf{x}$ ordered in a non-increasing manner.

**Lemma 2.5.2.** *For each* $k \geqslant 1$ *denote by* $\mathbf{V}_k^n := (n^{-2/3} \nu_k(\mathscr{C}_j))_{j \geqslant 1}$. *Then, the sequence* $\{\mathrm{ord}(\mathbf{V}_k^n)\}_{n \geqslant 1}$ *is tight in* $\ell_{\downarrow}^2$.

*Proof.* Note that for any $j \geqslant 1$, $\nu_k(\mathscr{C}_{(j)}) \leqslant |\mathscr{C}_{(j)}|$ uniformly over $k$. The proof now follows from (2.38) and $\ell^2_\downarrow$ tightness of the component sizes given in Theorem 2.2.1. □

**Remark 7.** Define $\mathbf{V}^n := (n^{-2/3} \nu_k(\mathscr{C}_j))_{k,j\geqslant 1}$. Then $\{\mathrm{ord}(\mathbf{V}^n)\}_{n\geqslant 1}$ is also tight in $\ell^2_\downarrow$.

## 2.6    Critical percolation

Let $p = p_n \in (0,1)$ be the percolation parameter. Recall the notation $\mathrm{CM}_n(\mathbf{d}, p)$ for the random graph obtained after deleting edges of $\mathrm{CM}_n(\mathbf{d})$ independently with probability $1 - p$. Suppose, $\mathbf{d}'$ is the random degree sequence obtained after percolation. Fountoulakis [85] showed that, given $\mathbf{d}'$, the law of $\mathrm{CM}_n(\mathbf{d}, p)$ is same as the law of $\mathrm{CM}_n(\mathbf{d}')$. We will use the following construction of $\mathrm{CM}_n(\mathbf{d}, p)$ due to Janson [109]:

**Algorithm 2.3.**  (S1) For each half-edge $e$, let $\nu_e$ be the vertex to which $e$ is attached. With probability $1 - \sqrt{p}$, one detaches $e$ from $\nu_e$ and associates $e$ to a new vertex $\nu'$. Color the new vertex red. This is done independently for every existing half-edge. Let $n_+$ be the number of red vertices created and $\tilde{n} = n + n_+$. Suppose, $\tilde{\mathbf{d}} = (\tilde{d}_i)_{i\in[\tilde{n}]}$ is the new degree sequence obtained by the above procedure, i.e. $\tilde{d}_i \sim \mathrm{Bin}(d_i, \sqrt{p})$ for $i \in [n]$ and $\tilde{d}_i = 1$ for $i \in [\tilde{n}] \setminus [n]$.

 (S2) Construct $\mathrm{CM}_{\tilde{n}}(\tilde{\mathbf{d}})$, independently of (S1).

 (S3) Delete all the red vertices.

**Remark 8.** It was argued in [109] that the obtained multigraph also has the same distribution as $\mathrm{CM}_n(\mathbf{d}, p)$ if we replace (S3) by

(S3′) Instead of deleting red vertices, choose any $n_+$ degree-one vertices uniformly at random, independently of (S1) and (S2), and delete them.

**Remark 9.** The construction of $\mathrm{CM}_{\tilde{n}}(\tilde{\mathbf{d}})$ in Algorithm 2.3 consists of two stages of randomization, the first one is described by (S1), and the second one by (S2). We will consider the following probability space to describe the randomization arising from Algorithm 2.3 (S1): Suppose we have a sequence of degree sequences $(\mathbf{d})_{n\geqslant 1}$. Let $\mathbb{P}^n_p$ denote the probability measure induced on $\mathbb{N}^\infty$ by Algorithm 2.3 (S1). Denote the product measure of $(\mathbb{P}^n_p)_{n\geqslant 1}$ by $\mathbb{P}_p$. Thus (S1) is performed independently on $\mathbf{d} = \mathbf{d}(n)$ as $n$ varies. All the

almost sure statements in this section will be with respect to the probability measure $\mathbb{P}_p$.

**Remark 10.** The idea of the proof of Theorem 2.2.3 is as follows. We show that $\tilde{d}$, under Assumption 2.2, satisfies Assumption 2.1 $\mathbb{P}_p$ almost surely and then estimate the number of vertices to be deleted from each component using Lemma 2.5.1. Since deleting a degree-one vertex does not break up any component, we can just subtract this from the component sizes of $CM_{\tilde{n}}(\tilde{d})$ to get the component sizes of $CM_n(d, p_n(\lambda))$. Since the degree-one vertices do not get involved in surplus edges, deleting degree-one vertices does not change the number of surplus edges.

### 2.6.1   Proof of Theorem 2.2.3

We now consider the critical window corresponding to percolation. The goal is to prove Theorem 2.2.3. Let $n_j$ and $\tilde{n}_j$ be the number of vertices of degree $j$ before and after performing Algorithm 2.3 (S1) respectively. Further let

$$\tilde{\nu}_n = \frac{\sum_{i \in [\tilde{n}]} \tilde{d}_i (\tilde{d}_i - 1)}{\sum_{i \in [\tilde{n}]} \tilde{d}_i}.$$

For convenience we write $r_j = \mathbb{P}(D = j)$. Denote by $\tilde{n}_{jl}$, the number of vertices that had degree $l$ before and have degree $j$ after performing Algorithm 2.3 (S1). Therefore, $\tilde{n}_{jl} \sim \text{Bin}\big(n_l, b_{lj}(\sqrt{p_n})\big)$, where $b_{lj}(\sqrt{p_n}) = \binom{l}{j}(\sqrt{p_n})^j(1 - \sqrt{p_n})^{l-j}$. Using the strong law of large numbers for triangular arrays, note that $\mathbb{P}_p$ almost surely, $\tilde{n}_{jl} = n_l b_{lj}(\sqrt{p_n}) + o(n_l) = n r_l b_{lj}(\sqrt{p_n}) + o(n_l)$. Now, $\sum_{l \geqslant 1} |n_l/n - r_l| \to 0$ and therefore, for all $j \geqslant 2$, $\mathbb{P}_p$ almost surely

$$\frac{\tilde{n}_j}{n} = \frac{\sum_{l=j}^{\infty} \tilde{n}_{jl}}{n} = \sum_{l=j}^{\infty} r_l b_{lj}(\sqrt{p_n}) + o(1). \tag{2.39}$$

Also, $n_+ = \sum_{i \in [n]} \big(d_i - \tilde{d}_i\big) \sim \text{Bin}(\ell_n, 1 - \sqrt{p_n})$. Therefore, using similar arguments as (2.39) again, $\mathbb{P}_p$ almost surely,

$$\frac{n_+}{n} = \mathbb{E}(D)\big(1 - \sqrt{p_n}\big) + o(1),$$

$$\frac{\tilde{n}_1}{n} = \frac{\sum_{l=1}^{\infty} \tilde{n}_{1l} + n_+}{n} = \frac{\sum_{l=1}^{\infty} \tilde{n}_{1l}}{n} + \mathbb{E}(D)\big(1 - \sqrt{p_n}\big) + o(1),$$

and

$$\frac{\tilde{n}}{n} = 1 + \frac{n_+}{n} = 1 + \mathbb{E}(D)\big(1 - \sqrt{p_n}\big) + o(1). \tag{2.40}$$

Denote $\tilde{r}_l = \mathbb{P}(\tilde{D} = l) = \lim_{n \to \infty} \tilde{n}_l / \tilde{n}$. Let $\tilde{D}_n$ denote the degree of a uniformly chosen vertex from $[\tilde{n}]$, independently of the graph $CM_{\tilde{n}}(\tilde{d})$. Thus, (2.39) and (2.40) imply that $\tilde{D}_n \xrightarrow{d} \tilde{D}$. The following lemma verifies the rest of the conditions for $\tilde{d}$ in Assumption 2.1:

**Lemma 2.6.1.** *The statements below are true $\mathbb{P}_p$ almost surely:*

1. *Under Assumption 2.2 i and for $r = 1, 2, 3$,*

$$\frac{1}{\tilde{n}} \sum_{i \in [n]} \tilde{d}_i^r = \frac{1}{\tilde{n}} \sum_{j \in [n]} j^r \tilde{n}_j \xrightarrow{n \to \infty} \mathbb{E}[\tilde{D}^r].$$

2. *Under Assumption 2.2,*

$$\tilde{\nu}_n = 1 + \lambda n^{-1/3} + o(n^{-1/3}).$$

*Proof.* We will make use of [115, Corollary 2.27]. Suppose $Z_1, Z_2, ..., Z_N$ are independent random variables with $Z_i$ taking values in $\Lambda_i$ and $f : \prod_{i=1}^N \Lambda_i \to \mathbb{R}$ satisfies the following: If two vectors $z, z' \in \prod_{i=1}^N \Lambda_i$ differ only in the $i^{th}$ coordinate, then $|f(z) - f(z')| \leqslant c_i$ for some constant $c_i$. Then, for any $t > 0$, the random variable $X = f(Z_1, Z_2, \ldots, Z_N)$ satisfies

$$\mathbb{P}\Big(\big|X - \mathbb{E}[X]\big| > t\Big) \leqslant 2 \exp\Big(-\frac{t^2}{2 \sum_{i=1}^N c_i^2}\Big). \qquad (2.41)$$

Now let $I_{ij}$ denote the indicator of the $j^{th}$ half-edge corresponding to vertex $i$ to be kept after Algorithm 2.3 (S1). Then $I_{ij} \sim Ber(\sqrt{p_n})$ independently for $j \in [d_i]$, $i \in [n]$. Let

$$\mathbf{I} := (I_{ij})_{j \in [d_i], i \in [n]} \quad \text{and} \quad f_1(\mathbf{I}) := \sum_{i \in [n]} \tilde{d}_i(\tilde{d}_i - 1).$$

Note that $f_1(\mathbf{I}) = \sum_{i \in [\tilde{n}]} \tilde{d}_i(\tilde{d}_i - 1)$ since the degree-one vertices do not contribute to the sum. One can check that, by changing the status of one half-edge corresponding to vertex $k$, we can change $f_1(\cdot)$ by at most $2(d_k + 1)$. Therefore, (2.41) yields

$$\mathbb{P}_p\Big(\big|\sum_{i \in [n]} \tilde{d}_i(\tilde{d}_i - 1) - p_n \sum_{i \in [n]} d_i(d_i - 1)\big| > t\Big)$$
$$\leqslant 2 \exp\Big(-\frac{t^2}{8 \sum_{i \in [n]} d_i(d_i + 1)^2}\Big).$$

By setting $t = n^{1/2+\varepsilon}$ for some suitably small $\varepsilon > 0$, using the finite third moment conditions and the Borel-Cantelli lemma we conclude that $\mathbb{P}_p$ almost surely,

$$\sum_{i \in [n]} \tilde{d}_i(\tilde{d}_i - 1) = p_n \sum_{i \in [n]} d_i(d_i - 1) + O(n^{1/2+\varepsilon}),$$

and in particular,

$$\sum_{i \in [\tilde{n}]} \tilde{d}_i(\tilde{d}_i - 1) = \sum_{i \in [n]} \tilde{d}_i(\tilde{d}_i - 1) = p_n \sum_{i \in [n]} d_i(d_i - 1) + o(n^{2/3}). \quad (2.42)$$

Similarly, take $f_2(\mathbf{I}) = \sum_{i \in [n]} \tilde{d}_i(\tilde{d}_i - 1)(\tilde{d}_i - 2)$ and note that changing the status of one bond changes $f_2(\cdot)$ by at most $[2(d_k + 1)]^2$. Thus, (2.41) gives

$$\mathbb{P}_p\left(\left|f_2(\mathbf{I}) - p_n^{3/2} \sum_{i \in [n]} d_i(d_i - 1)(d_i - 2)\right| > t\right)$$

$$\leqslant 2 \exp\left(-\frac{t^2}{32 \sum_{i \in [n]} d_i(d_i + 1)^4}\right)$$

$$\leqslant \exp\left(-\frac{t^2}{32 d_{\max}(d_{\max} + 1) \sum_{i \in [n]}(d_i + 1)^3}\right),$$

which implies that, $\mathbb{P}_p$ almost surely,

$$\sum_{i \in [\tilde{n}]} \tilde{d}_i(\tilde{d}_i - 1)(\tilde{d}_i - 2) = \sum_{i \in [n]} \tilde{d}_i(\tilde{d}_i - 1)(\tilde{d}_i - 2)$$

$$= p_n^{3/2} \sum_{i \in [n]} d_i(d_i - 1)(d_i - 2) + o(n), \quad (2.43)$$

since $d_{\max}^2 \sum_{i \in [n]}(d_i + 1)^3 = o(n^{5/3})$. Now, to prove Lemma 2.6.1 (1), note that the case $r = 1$ follows by simply observing that $\sum_{i \in \tilde{n}} \tilde{d}_i = \sum_{i \in [n]} d_i$. The cases $r = 2, 3$ follow from (2.42) and (2.43). Finally, to see Lemma 2.6.1 (2), note that

$$\tilde{v}_n = \frac{\sum_{i \in [\tilde{n}]} \tilde{d}_i(\tilde{d}_i - 1)}{\sum_{i \in [\tilde{n}]} \tilde{d}_i} = \frac{p_n \sum_{i \in [n]} d_i(d_i - 1) + o(n^{2/3})}{\sum_{i \in [n]} d_i}$$

$$= \frac{p_n \sum_{i \in [n]} d_i(d_i - 1)}{\sum_{i \in [n]} d_i} + o(n^{-1/3}) = 1 + \frac{\lambda}{n^{1/3}} + o(n^{-1/3}),$$

by (2.42) and this completes the proof of Lemma 2.6.1.  □

We will denote by $\tilde{\mathscr{C}}_{(j)}$, the $j^{\text{th}}$ largest component of $CM_{\tilde{n}}(\check{\mathbf{d}})$. To conclude Theorem 2.2.3 we also need to estimate the number of deleted vertices

from each component. Recall from Remark 8 that $CM_n(\mathbf{d}, p_n(\lambda))$ can be obtained from $CM_{\tilde{n}}(\tilde{\mathbf{d}})$ by deleting the relevant number of degree-one vertices *uniformly* at random. Let $v_1^d(\tilde{\mathscr{C}}_{(j)})$ be the number of degree-one vertices of $\tilde{\mathscr{C}}_{(j)}$ that are deleted while creating $CM_n(\mathbf{d}, p_n(\lambda))$ from $CM_{\tilde{n}}(\tilde{\mathbf{d}})$. Since the vertices are to be chosen uniformly from all degree-one vertices, the number of vertices to be deleted from $\tilde{\mathscr{C}}_{(j)}$ is asymptotically the total number of degree-one vertices in $\tilde{\mathscr{C}}_{(j)}$ times the proportion of degree-one vertices to be deleted. Therefore,

$$
\begin{aligned}
v_1^d(\tilde{\mathscr{C}}_{(j)}) &= \frac{n_+}{\tilde{n}_1} v_1(\tilde{\mathscr{C}}_{(j)}) + o_{\mathbb{P}}(n^{2/3}) = \frac{n_+}{\tilde{n}_1} \frac{\tilde{n}_1}{\sum_{k=0}^{\infty} k\tilde{n}_k} |\tilde{\mathscr{C}}_{(j)}| + o_{\mathbb{P}}(n^{2/3}) \\
&= \frac{n_+}{\ell_n} |\tilde{\mathscr{C}}_{(j)}| + o_{\mathbb{P}}(n^{2/3}) = \frac{\mathbb{E}[D](1 - \sqrt{p_n})}{\mathbb{E}[D]} |\tilde{\mathscr{C}}_{(j)}| + o_{\mathbb{P}}(n^{2/3}) \quad (2.44) \\
&= (1 - \sqrt{p_n}) |\tilde{\mathscr{C}}_{(j)}| + o_{\mathbb{P}}(n^{2/3}),
\end{aligned}
$$

where the third equality follows from (2.38). The proof of Theorem 2.2.3 is now complete by using the $\ell_{\downarrow}^2$ convergence in Lemma 2.5.2, (2.44) and Remark 10.

## 2.7 Joint convergence at multiple locations in the critical window

We will prove Theorem 2.2.4 in this section. In Section 2.7.1, we give a construction of the joint distribution of the percolated graphs for different percolation parameters that are coupled in a way described in Theorem 2.2.4. In Section 2.7.2, we compare the process of percolated graphs with a different graph process that turns out to be easier to work with. As discussed in Remark 3, let the mass of a component be the number of open half-edges (rescaled by $n^{2/3}$). The alternatively constructed graph process can be modified in such a way that the vector of masses evolves according to an *exact* multiplicative coalescent as discussed in Section 2.7.3. Thus the joint convergence result at multiple locations of the scaling window can be deduced for the modified process using the Feller property of the multiplicative coalescent. Further, the modified process remains *close* to the dynamic construction. In Section 2.7.4, the vector of masses are shown to be asymptotically proportional to the component sizes and we combine all the above observations in Section 2.7.5 to complete the proof of Theorem 2.2.4.

### 2.7.1    Construction of the percolated graph process

We start by explaining a way to construct the graph process $(CM_n(\mathbf{d}, p))_{p \in [0,1]}$. Fix any $p_1 < p_2 < \cdots < p_m$ and consider $(CM_n(\mathbf{d}, p_i))_{i \in [m]}$. Recall that each edge $e$ of $CM_n(\mathbf{d})$ has an independent uniform $[0,1]$ random variable $U_e$ associated to it and $CM_n(\mathbf{d}, p_i)$ is obtained from $CM_n(\mathbf{d})$ by keeping only those edges $e$ with $U_e \leqslant p_i$. This couples the graphs $(CM_n(\mathbf{d}, p_i))_{i \in [m]}$. Moreover, under this coupling, $CM_n(\mathbf{d}, p_i)$ is distributed as the graph obtained from edge percolation on $CM_n(\mathbf{d}, p_{i+1})$ with probability $p_i/p_{i+1}$ for all $i < m$. The following two lemmas are modifications of [85, Lemmas 3.1, 3.2] that lead to the construction of Algorithm 2.4 below. For a graph G, let $E(G)$ denote the set of edges of G. For a sub-graph G of $CM_n(\mathbf{d})$, let $\mathcal{H}(G)$ denote the set of half-edges that are part of some edge in G and $\mathcal{H} = \mathcal{H}(CM_n(\mathbf{d}))$.

**Lemma 2.7.1.** *For $k_1 \leqslant \cdots \leqslant k_m$, conditionally on $\{|E(CM_n(\mathbf{d}, p_i))| = k_i : i \leqslant m\}$, the half-edges in $CM_n(\mathbf{d}, p_i)$ can be generated sequentially as follows: Let $k_0 = 0$, $\mathcal{H}(CM_n(\mathbf{d}, p_0)) = \varnothing$. For each $i \leqslant m$, declare $\mathcal{H}(CM_n(\mathbf{d}, p_i)) = \mathcal{H}(CM_n(\mathbf{d}, p_{i-1})) \cup \mathcal{H}_i$, where $\mathcal{H}_i$ is uniformly chosen among all the subsets of size $2k_i - 2k_{i-1}$ of $\mathcal{H} \setminus \cup_{j<i} \mathcal{H}_i$.*

**Lemma 2.7.2.** *Let $d_k(i, i+1)$ be the number of half-edges attached to vertex k in the graph $CM_n(\mathbf{d}, p_{i+1})$ that are not in $CM_n(\mathbf{d}, p_i)$. For any $i \geqslant 1$, conditionally on the event $\{\mathbf{d}(j, j+1) = \mathbf{d}_0(j, j+1) : j \leqslant m\}$ and $\mathcal{H}(CM_n(\mathbf{d}, p_{i-1}))$, the perfect matching of $\mathcal{H}(CM_n(\mathbf{d}, p_i)) \setminus \mathcal{H}(CM_n(\mathbf{d}, p_{i-1}))$ constituting the edges $E(CM_n(\mathbf{d}, p_i) \setminus CM_n(\mathbf{d}, p_{i-1}))$ is a uniform perfect matching, where we have assumed that $p_0 = 0$.*

**Algorithm 2.4.** Let $(U_i)_{i \geqslant 1}$ be a finite collection of i.i.d uniform $[0,1]$ random variables. Construct a collection of graphs $(G_n(\lambda))_{\lambda \in \mathbb{R}}$ using the following two steps:

(S0) Construct the process $E_n = (E_n(\lambda))_{\lambda \in \mathbb{R}}$, where $E_n(\lambda) = \#\{i : U_i \leqslant p_n(\lambda)\}$.

(S1) Initially, $G_n(-\infty)$ is a graph only consisting of isolated vertices with no paired half-edges. At each time point $\lambda$ where $E_n(\lambda)$ has a jump, choose two unpaired half-edges uniformly at random and pair them. The graph $G_n(\lambda)$ is obtained by adding this edge to $G_n(\lambda-)$.

Algorithm 2.4 (S0) can be regarded as the birth of edges and Algorithm 2.4 (S1) ensures that the edges of the graph $G_n(\lambda)$ are obtained from a uniform perfect matching of the corresponding half-edges. Using Lemmas 2.7.1 and 2.7.2,

$(G_n(\lambda))_{\lambda \in \mathbb{R}}$ and $(CM_n(\mathbf{d}, p_n(\lambda)))_{\lambda \in \mathbb{R}}$ have the same finite-dimensional distributions. Therefore, $(G_n(\lambda))_{\lambda \in \mathbb{R}}$ and $(CM_n(\mathbf{d}, p_n(\lambda)))_{\lambda \in \mathbb{R}}$ have the exact same distribution. We complete this section by adding proofs of Lemmas 2.7.1, and 2.7.2 which are in the same spirit as the arguments of [85, Lemmas 3.1, 3.2].

*Proof of Lemma 2.7.1.* Assume that $k = 2$ for the sake of simplicity. Observe that the total number of perfect matchings of $2k$ objects is given by $2k!/(k!2^k) = (2k-1)!!$. Let $H_1$, $H_2$ be two disjoint subsets of $\mathcal{H}$ with $|H_1| = 2k_1$, $|H_2| = 2k_2 - 2k_1$. Let $\mathcal{E}_1$ denote the event that a uniform perfect matching of all the half-edges contains also perfect matchings of the half-edges in $H_1$ and $H_2$. Then,

$$\mathbb{P}(\mathcal{E}_1) = \frac{(2k_1-1)!!(2k_2-2k_1-1)!!(\ell_n - 2k_2-1)!!}{(\ell_n - 1)!!}.$$

Also, for percolation on any (random) graph, conditional on the set of edges of the graph and the fact that $k$ edges have been retained by percolation, the choice of the retained edges is uniformly distributed among all subsets of size $k$ of the set of edges. Let $\mathcal{E}_2$ denote the event that $|\mathcal{H}(CM_n(\mathbf{d}, p_1))| = 2k_1$, and $|\mathcal{H}(CM_n(\mathbf{d}, p_2))| = 2k_2$. It follows that

$$\mathbb{P}(\mathcal{H}(CM_n(\mathbf{d}, p_2)) = H_1 \cup H_2 \mid \mathcal{E}_1, \mathcal{E}_2) = \frac{1}{\binom{\ell_n/2}{k_2}},$$

$$\mathbb{P}(\mathcal{H}(CM_n(\mathbf{d}, p_1)) = H_1 \mid \mathcal{E}_1, \mathcal{E}_2, \mathcal{H}(CM_n(\mathbf{d}, p_2)) = H_1 \cup H_2) = \frac{1}{\binom{k_2}{k_1}}.$$

Thus, conditional on the event $\mathcal{E}_2$, the probability that $\mathcal{H}(CM_n(\mathbf{d}, p_1)) = H_1$ and $\mathcal{H}(CM_n(\mathbf{d}, p_2)) \setminus \mathcal{H}(CM_n(\mathbf{d}, p_1)) = H_2$ is given by

$$\frac{(2k_1-1)!!(2k_2-2k_1-1)!!(\ell_n - 2k_2-1)!!}{(\ell_n - 1)!!} \frac{1}{\binom{\ell_n/2}{k_2}\binom{k_2}{k_1}} = \frac{1}{\binom{\ell_n}{2k_1}\binom{\ell_n - 2k_1}{2k_2 - 2k_1}}, \tag{2.45}$$

which does not depend on $H_1$ or $H_2$, and the proof follows. □

*Proof of Lemma 2.7.2.* Fix two disjoint subsets $H_1$, $H_2$ of $\mathcal{H}$ such that $|H_1| = 2k_1$, $|H_2| = 2k_2 - 2k_1$. As in the proof of Lemma 2.7.1, let $\mathcal{E}_2$ denote the event that $|\mathcal{H}(CM_n(\mathbf{d}, p_1))| = 2k_1$, and $|\mathcal{H}(CM_n(\mathbf{d}, p_2))| = 2k_2$. An identical argument as the proof of (2.45) now gives, conditionally on $\mathcal{E}_2$, the probability that $\mathcal{H}(CM_n(\mathbf{d}, p_1)) = H_1$, $\mathcal{H}(CM_n(\mathbf{d}, p_2)) \setminus \mathcal{H}(CM_n(\mathbf{d}, p_1)) = H_2$, and given

perfect matchings on $\mathcal{H}(CM_n(\mathbf{d}, p_1))$, $\mathcal{H}(CM_n(\mathbf{d}, p_2)) \setminus \mathcal{H}(CM_n(\mathbf{d}, p_1))$ have been observed, is given by

$$\frac{1}{\binom{\ell_n/2}{k_2}\binom{k_2}{k_1}} \frac{(\ell_n - 2k_2 - 1)!!}{(\ell_n - 1)!!}.$$

Let $D(H)$ denote the degree sequence induced by the set of half-edges $H$, and $S$ denote the collection of *disjoint* pairs $(H_1, H_2)$ such that $|H_1| = 2k_1$, $|H_2| = 2k_2 - 2k_1$, $D(H_1) = \mathbf{d}_0(0, 1)$, and $D(H_2) = \mathbf{d}_0(1, 2)$. Then, conditionally on $\mathcal{E}_2$, the probability that $\mathbf{d}(0, 1) = \mathbf{d}_0(0, 1)$, $\mathbf{d}(1, 2) = \mathbf{d}_0(1, 2)$, and given particular perfect matchings have been observed on $\mathcal{H}(CM_n(\mathbf{d}, p_1))$ and $\mathcal{H}(CM_n(\mathbf{d}, p_2)) \setminus \mathcal{H}(CM_n(\mathbf{d}, p_1))$, is

$$\sum_{(H_1, H_2) \in S} \frac{1}{\binom{\ell_n/2}{k_2}\binom{k_2}{k_1}} \frac{(\ell_n - 2k_2 - 1)!!}{(\ell_n - 1)!!} = \frac{|S|}{\binom{\ell_n/2}{k_2}\binom{k_2}{k_1}} \frac{(\ell_n - 2k_2 - 1)!!}{(\ell_n - 1)!!}. \quad (2.46)$$

Moreover, by Lemma 2.7.1, the probability that $\mathbf{d}(0, 1) = \mathbf{d}_0(0, 1)$, $\mathbf{d}(1, 2) = \mathbf{d}_0(1, 2)$, conditionally on $\mathcal{E}_2$, is given by

$$\frac{|S|}{\binom{\ell_n}{2k_1}\binom{\ell_n - 2k_1}{2k_2 - 2k_1}}. \quad (2.47)$$

Now, (2.46) and (2.47) together yield that the probability that two particular perfect matchings are observed on $\mathcal{H}(CM_n(\mathbf{d}, p_1))$ and $\mathcal{H}(CM_n(\mathbf{d}, p_2)) \setminus \mathcal{H}(CM_n(\mathbf{d}, p_1))$, conditional on $\mathbf{d}(0, 1) = \mathbf{d}_0(0, 1)$, $\mathbf{d}(1, 2) = \mathbf{d}_0(1, 2)$ is given by

$$\frac{1}{\binom{\ell_n/2}{k_2}\binom{k_2}{k_1}} \frac{(\ell_n - 2k_2 - 1)!!}{(\ell_n - 1)!!} \binom{\ell_n}{2k_1}\binom{\ell_n - 2k_1}{2k_2 - 2k_1} = \frac{1}{(2k_1 - 1)!!(2k_2 - 2k_1 - 1)!!},$$

and the proof is complete. $\qquad\square$

### 2.7.2   The dynamic construction

Let us now describe a dynamic construction of $CM_n(\mathbf{d})$ that turns out to be easier to work with. This dynamic construction was introduced in [29] to study the metric-space limits of the large components of the percolated configuration model. It will be shown that the graphs generated by this dynamic construction at a suitable range of time *approximate* the process $(CM_n(\mathbf{d}, p_n(\lambda)))_{\lambda \in \mathbb{R}}$.

**Algorithm 2.5.** At time $t = 0$, assume that there are $d_i$ *open* half-edges associated with vertex $i$, for all $i \in [n]$. Associate i.i.d unit rate exponential clocks to

each of the open half-edges. Each time an exponential clock rings, the corresponding half-edge selects another open half-edge uniformly at random and gets paired to it. The two paired half-edges are declared to be closed and the associated exponential clocks are removed. The process continues until the open half-edges are exhausted.

Let $\mathcal{G}_n(t)$ denote the graph generated upto time $t$. Notice that $\mathcal{G}_n(\infty)$ is distributed as $CM_n(\mathbf{d})$ since each half-edge chooses to pair with another uniformly chosen open half-edge. Denote the total number of open-half-edges remaining at time $t$ while implementing Algorithm 2.5 by $s_1(t)$. The graph process, given by Algorithm 2.5, can also be constructed as follows:

**Algorithm 2.6.** Let $\Xi_n$ be an inhomogeneous Poisson process with rate $s_1(t)$ at time $t$. Let $e_1 < e_2 < \ldots$ be the event times of $\Xi_n$.

(S1) At each event time, choose two unpaired half-edges uniformly at random and pair them. The graph $\mathcal{G}_n(t)$ is obtained by adding this edge to $\mathcal{G}_n(t-)$.

Notice the similarity between Algorithm 2.4 (S1) and Algorithm 2.6 (S1). Now, the idea is to compare the number of half-edges that have been paired by Algorithms 2.4 and 2.6. For that, we need the following lemma that describes the evolution of the count of the total number of open half-edges in Algorithm 2.6:

**Lemma 2.7.3** ([29, Lemma 8.2]). *Let* $s_1(t)$ *denote the total number of open half-edges at time* $t$. *Suppose that Assumption 2.2 holds. Then, for any* $T > 0$ *and some* $1/3 < \gamma < 1/2$,

$$\sup_{t \leqslant T} \left| \frac{1}{\ell_n} s_1(t) - e^{-2t} \right| = o_{\mathbb{P}}(n^{-\gamma}).$$

Notice that the proof of [29, Lemma 8.2] is stated only under some more stringent assumptions, however the identical argument can be carried out under Assumption 2.2. The next proposition ensures that the graphs generated by percolation in Algorithm 2.4 and the dynamic construction in Algorithm 2.5 are uniformly close in the critical window. Define

$$t_n(\lambda) = \frac{1}{2} \log \left( \frac{\nu_n}{\nu_n - 1} \right) + \frac{1}{2(\nu_n - 1)} \frac{\lambda}{n^{1/3}}. \tag{2.48}$$

**Proposition 2.7.4.** *Fix $-\infty < \lambda_\star < \lambda^\star < \infty$. There exists a coupling such that with high probability*

$$\mathcal{G}_n(t_n(\lambda) - \varepsilon_n) \subset CM_n(\mathbf{d}, p_n(\lambda)) \subset \mathcal{G}_n(t_n(\lambda) + \varepsilon_n), \quad \forall \lambda \in [\lambda_\star, \lambda^\star]$$

*where $\varepsilon_n = cn^{-\gamma_0}$, for some $1/3 < \gamma_0 < 1/2$ and the constant $c$ does not depend on $\lambda$.*

*Proof.* Notice the similarity between Algorithm 2.4 (S1) and Algorithm 2.6 (S1). Let $\#E(G)$ denote the number of edges in a graph $G$. Suppose with high probability the following holds: $\forall \lambda \in [\lambda_\star, \lambda^\star]$

$$\#E(\mathcal{G}_n(t_n(\lambda) - \varepsilon_n)) \leqslant \#E(CM_n(\mathbf{d}, p_n(\lambda))) \leqslant \#E(\mathcal{G}_n(t_n(\lambda) + \varepsilon_n)), \quad .(2.49)$$

On the event $\{\#E(CM_n(\mathbf{d}, p_n(\lambda))) \leqslant \#E(\mathcal{G}_n(t_n(\lambda) + \varepsilon_n)), \forall \lambda \in [\lambda_\star, \lambda^\star]\}$, the choice of the uniform pair of half-edges at the $k^{th}$ pairing in Algorithm 2.4 (S1) can be taken to be exactly the same as the $k^{th}$ pairing in Algorithm 2.6 (S1). Under the above coupling $CM_n(\mathbf{d}, p_n(\lambda_\star)) \subset \mathcal{G}_n(t_n(\lambda_\star) + \varepsilon_n)$. Moreover, since $\#E(CM_n(\mathbf{d}, p_n(\lambda)))$ is dominated by $\#E(\mathcal{G}_n(t_n(\lambda) + \varepsilon_n))$, uniformly over $\lambda \in [\lambda_\star, \lambda^\star]$, the above coupling also yields $CM_n(\mathbf{d}, p_n(\lambda)) \subset \mathcal{G}_n(t_n(\lambda) + \varepsilon_n)$ for all $\lambda \in [\lambda_\star, \lambda^\star]$. Further, on the event $\{\#E(\mathcal{G}_n(t_n(\lambda) - \varepsilon_n)) \leqslant \#E(CM_n(\mathbf{d}, p_n(\lambda))), \forall \lambda \in [\lambda_\star, \lambda^\star]\}$, under the same coupling, $\mathcal{G}_n(t_n(\lambda) - \varepsilon_n) \subset CM_n(\mathbf{d}, p_n(\lambda))$ for all $\lambda \in [\lambda_\star, \lambda^\star]$. Thus, it remains to show (2.49). An application of Lemma 2.7.3 along with (2.48) yields, for some $1/3 < \gamma_0 < \gamma < 1/2$, with high probability,

$$\left| \#E(\mathcal{G}_n(t_n(\lambda))) - \left( \frac{\ell_n}{2\nu_n} + \frac{\lambda \ell_n}{2\nu_n n^{1/3}} + \frac{n\varepsilon_n(\nu_n - 1)}{\nu_n} \right) \right| \leqslant n^{1-\gamma}, \quad \lambda \in [\lambda_\star, \lambda^\star].$$
$$(2.50)$$

Notice that the total number of half-edges in $CM_n(\mathbf{d}, p_n(\lambda))$ follows a binomial distribution with parameters $\ell_n/2$ and $p_n(\lambda)$. Thus, with high probability,

$$\left| \#E(CM_n(\mathbf{d}, p_n(\lambda))) - \left( \frac{\ell_n}{2\nu_n} + \frac{\lambda \ell_n}{2\nu_n n^{1/3}} \right) \right| \leqslant n^{1-\gamma}, \quad \lambda \in [\lambda_\star, \lambda^\star]. \quad (2.51)$$

The fact that the error can be chosen to be uniform over $\lambda \in [\lambda_\star, \lambda^\star]$ follows from the DKW inequality [132]. Thus, (2.50) and (2.51) together show that, with high probability,

$$\#E(CM_n(\mathbf{d}, p_n(\lambda))) \leqslant \#E(\mathcal{G}_n(t_n(\lambda) + \varepsilon_n)), \quad \forall \lambda \in [\lambda_\star, \lambda^\star].$$

The other part follows similarly and the proof is now complete. □

**Remark 11.** Notice that the proof of Proposition 2.7.4 can be directly modified to show that there exists a coupling such that, with high probability,

$$CM_n(\mathbf{d}, p_n(\lambda) - \varepsilon_n) \subset \mathcal{G}_n(t_n(\lambda)) \subset CM_n(\mathbf{d}, p_n(\lambda) + \varepsilon_n), \quad \forall \lambda \in [\lambda_\star, \lambda^\star],$$

where $\varepsilon_n = cn^{-\gamma_0}$, for some $1/3 < \gamma_0 < 1/2$ and the constant $c$ does not depend on $\lambda$. Therefore, the scaling limits of different functionals like rescaled component-sizes, surplus edges for $\mathcal{G}_n(t_n(\lambda))$ and $CM_n(\mathbf{d}, p_n(\lambda))$ are the same.

### 2.7.3    The modified process

From here onward, we often augment $\lambda$ to a predefined notation to emphasize the dependence on $\lambda$. We write $\mathscr{C}_{(i)}(\lambda)$ for the $i^{th}$ largest component of $\mathcal{G}_n(t_n(\lambda))$ and define

$$\mathcal{O}_i(\lambda) = \# \text{ open half-edges in } \mathscr{C}_{(i)}(\lambda).$$

Think of $\mathcal{O}_i(\lambda)$ as the *mass* of the component $\mathscr{C}_{(i)}(\lambda)$. Define the vector $\mathbf{C}_n(\lambda) = (n^{-2/3}|\mathscr{C}_{(i)}(\lambda)|)_{i \geqslant 1}$, and $\mathbf{O}_n(\lambda) = (n^{-2/3}\mathcal{O}_i(\lambda))_{i \geqslant 1}$. Let $\ell_n^o(\lambda) = \sum_{i \geqslant 1} \mathcal{O}_i(\lambda)$. By Lemma 2.7.3 and (2.48), $\ell_n^o(\lambda) \approx n\mu(\nu - 1)/\nu$. Now, observe that, during the evolution of the graph process generated by Algorithm 2.5, during the time interval $[t_n(\lambda), t_n(\lambda + d\lambda)]$, the $i^{th}$ and $j^{th}$ ($i > j$) largest components, merge at rate

$$2\mathcal{O}_i(\lambda)\mathcal{O}_j(\lambda) \times \frac{1}{\ell_n^o(\lambda) - 1} \times \frac{1}{2(\nu_n - 1)n^{1/3}}$$
$$\approx \frac{\nu}{\mu(\nu - 1)^2}\left(n^{-2/3}\mathcal{O}_i(\lambda)\right)\left(n^{-2/3}\mathcal{O}_j(\lambda)\right),$$

and create a component with $\mathcal{O}_i(\lambda) + \mathcal{O}_j(\lambda) - 2$ open half-edges. Thus the open half-edges $(\mathbf{O}_n(\lambda))_{\lambda \in \mathbb{R}}$ does *not* evolve as a multiplicative coalescent, but it is close. The fact that two half-edges are killed after pairing, makes the masses (the number of open half-edges) of the components deplete. If there were no such depletion of mass, then the vector of open half-edges would in fact merge as multiplicative coalescent. Let us formalize this idea below:

**Algorithm 2.7.** Initialize $\bar{\mathcal{G}}_n(t_n(\lambda_\star)) = \mathcal{G}_n(t_n(\lambda_\star))$. Let $\mathscr{O}$ denote the set of open half-edges in the graph $\mathcal{G}_n(t_n(\lambda_\star))$, $\bar{s}_1 = |\mathscr{O}|$ and $\bar{\Xi}_n$ denote a Poisson process with rate $\bar{s}_1$. At each event time of the Poisson process $\bar{\Xi}_n$, select two half-edges from $\mathscr{O}$ and create an edge between the corresponding vertices. However, the selected half-edges are kept alive, so that they can be selected again.

**Remark 12.** The only difference between Algorithm 2.6 and Algorithm 2.7, is that the *paired* half-edges are not discarded and thus more edges are created by Algorithm 2.7. Thus, there is a natural coupling between the graphs generated by Algorithms 2.6 and 2.7 such that $\mathcal{G}_n(t_n(\lambda)) \subset \bar{\mathcal{G}}_n(t_n(\lambda))$ for all $\lambda \in [\lambda_\star, \lambda^\star]$, with probability one. In the subsequent part of this section, we always work under this coupling. The extra edges that are created by Algorithm 2.7 will be called *bad* edges.

**Remark 13.** In the subsequent part of this chapter, we shall augment a predefined notation with a bar to denote the corresponding quantity for $\bar{\mathcal{G}}_n(t_n(\lambda))$. Denote $\beta_n = (\bar{s}_1(\nu_n - 1)n^{1/3})^{1/2}$ and $\bar{\mathbf{O}}'_n(\lambda)$ denote the vector $\mathrm{ord}((\beta_n^{-1}\bar{O}_i(\lambda))_{i\geqslant 1})$. By the description in Algorithm 2.7, $(\bar{\mathbf{O}}'_n(\lambda))_{\lambda\geqslant\lambda_\star}$ evolves as a standard multiplicative coalescent. Further, note that there exists a constant $c > 0$ such that $\beta_n = cn^{2/3}(1 + o_{\mathbb{P}}(1))$ which enables us to deduce the scaling limit results for $(\bar{\mathbf{O}}_n(\lambda))_{\lambda\geqslant\lambda_\star}$ from $(\bar{\mathbf{O}}'_n(\lambda))_{\lambda\geqslant\lambda_\star}$.

**Multiplicative coalescent with mass and weight**

The Feller property of the multiplicative coalescent [8, Proposition 5] ensures the joint convergence of the number of open half-edges in each component of $\bar{\mathcal{G}}_n(t_n(\lambda))$ at multiple values of $\lambda$ as we shall see below. To deduce the scaling limits involving the components sizes let us consider a dynamic process that is further augmented by a certain weight. Initially, the system consists of particles (possibly infinitely many) where particle $i$ has mass $x_i$, and weight $z_i$. Let $(X_i(t), Z_i(t))_{i\geqslant 1}$ denote the vector of masses, and weights at time $t$. The dynamics of the system is described as follows:

> At time $t$, particles $i$ and $j$ coalesce at rate $X_i(t)X_j(t)$ and create a particle with mass $X_i(t) + X_j(t)$, and weight $Z_i(t) + Z_j(t)$.

Denote by $\mathrm{MC}_2(\mathbf{x}, \mathbf{z}, t)$ the vector $(X_i(t), Z_i(t))_{i\geqslant 1}$ with initial mass $\mathbf{x}$, and weight $\mathbf{z}$. We shall need the following theorem:

**Theorem 2.7.5.** *Suppose that $(\mathbf{x}_n, \mathbf{z}_n) \to (\mathbf{x}, \mathbf{x})$ in $(\ell^2_\downarrow)^2$. Then, for any $t \geqslant 0$*

$$\mathrm{MC}_2(\mathbf{x}_n, \mathbf{z}_n, t) \xrightarrow{d} \mathrm{MC}_2(\mathbf{x}, \mathbf{x}, t).$$

*Proof.* For $\mathbf{x}_n = (x_i^n)_{i\geqslant 1}$ and $\mathbf{z}_n = (z_i^n)_{i\geqslant 1}$, let $\mathbf{w}_n^+ = \mathrm{ord}(x_i^n \vee z_i^n)$, $\mathbf{w}_n^- = \mathrm{ord}(x_i^n \wedge z_i^n)$, where ord denotes the decreasing ordering of the elements. Notice that $\mathbf{w}_n^+ \to \mathbf{x}$, and $\mathbf{w}_n^- \to \mathbf{x}$ in $\ell^2_\downarrow$. Using the Feller property of the multiplicative coalescent [8, Proposition 5], it follows that

$$\mathrm{MC}_2(\mathbf{w}_n^+, \mathbf{w}_n^+, t) \xrightarrow{d} \mathrm{MC}_2(\mathbf{x}, \mathbf{x}, t), \quad \mathrm{MC}_2(\mathbf{w}_n^-, \mathbf{w}_n^-, t) \xrightarrow{d} \mathrm{MC}_2(\mathbf{x}, \mathbf{x}, t), \quad (2.52)$$

with respect to the $(\ell_\downarrow^2)^2$ topology. Now suppose that $MC_2(\mathbf{w}_n^+, \mathbf{w}_n^+, t)$ and $MC_2(\mathbf{w}_n^-, \mathbf{w}_n^-, t)$ are coupled through the subgraph coupling (see [8, Page 838]). For $(\mathbf{x}, \mathbf{z}) \in (\ell_\downarrow^2)^2$, denote $\|(\mathbf{x}, \mathbf{z})\|_{22} = (\sum_{i \geqslant 1} x_i^2)^{1/2} + (\sum_{i \geqslant 1} z_i^2)^{1/2}$. Under the subgraph coupling, (2.52) yields

$$\|MC_2(\mathbf{w}_n^+, \mathbf{w}_n^+, t)\|_{22}^2 - \|MC_2(\mathbf{w}_n^-, \mathbf{w}_n^-, t)\|_{22}^2 \xrightarrow{\mathbb{P}} 0.$$

Moreover,

$$\|MC_2(\mathbf{w}_n^-, \mathbf{w}_n^-, t)\|_{22}^2 \leqslant \|MC_2(\mathbf{x}_n, \mathbf{z}_n, t)\|_{22}^2 \leqslant \|MC_2(\mathbf{w}_n^+, \mathbf{w}_n^+, t)\|_{22}^2.$$

Hence, using [8, Corollary 18 (a)], under the subgraph coupling,

$$\|MC_2(\mathbf{w}_n^+, \mathbf{w}_n^+, t) - MC_2(\mathbf{x}_n, \mathbf{z}_n, t)\|_{22}^2$$
$$\leqslant \|MC_2(\mathbf{w}_n^+, \mathbf{w}_n^+, t)\|_{22}^2 - \|MC_2(\mathbf{x}_n, \mathbf{z}_n, t)\|_{22}^2 \xrightarrow{\mathbb{P}} 0,$$

and the proof follows. □

### 2.7.4   Asymptotics for the open half-edges

In this section, we show that the open half-edges in the components of $\mathcal{G}_n(t_n(\lambda))$ are *approximately* proportional to the component sizes. This will enable us to apply Theorem 2.7.5 for deducing the scaling limits of the required quantities for the graph $\bar{\mathcal{G}}_n(t_n(\lambda))$.

**Lemma 2.7.6.** *There exists a constant $\kappa > 0$ such that, for any $\lambda \in \mathbb{R}$ and $i \geqslant 1$,*

$$\mathcal{O}_i(\lambda) = \kappa |\mathscr{C}_{(i)}(\lambda)| + o_{\mathbb{P}}(b_n). \tag{2.53}$$

*Further, $(\mathbf{O}_n(\lambda))_{n \geqslant 1}$ is tight in $\ell_\downarrow^2$ and consequently*

$$n^{-4/3} \sum_{i \geqslant 1} (\mathcal{O}_i(\lambda) - \kappa |\mathscr{C}_{(i)}(\lambda)|)^2 \xrightarrow{\mathbb{P}} 0.$$

*Proof.* Let $(d_k^\lambda)_{k \in [n]}$ denote the degree sequence of $CM_n(\mathbf{d}, p_n(\lambda))$ and define

$$\mathcal{O}_i^P(\lambda) = \sum_{k \in \mathscr{C}_{(i)}^P(\lambda)} (d_k - d_k^\lambda) = \sum_{k \in \mathscr{C}_{(i)}^P(\lambda)} d_k - 2(|\mathscr{C}_{(i)}^P(\lambda)| - 1 + SP(\mathscr{C}_{(i)}^P(\lambda))).$$

Using Remark 11 and the fact that the surplus edges in the large components is tight, it is enough to prove the lemma by replacing $\mathcal{O}_i(\lambda)$ by $\mathcal{O}_i^P(\lambda)$ and $\mathscr{C}_{(i)}(\lambda)$ by $\mathscr{C}_{(i)}^P(\lambda)$. For a component $\mathscr{C}$ of $CM_{\tilde{n}}(\tilde{\mathbf{d}})$, the corresponding

component $\mathscr{C}^p$ in the percolated graph is obtained by cleaning up $R(\mathscr{C})$ red degree-one vertices, see Algorithm 2.3. Thus, the number of open half-edges in $\mathscr{C}^p$ is given by

$$\sum_{k \in \mathscr{C} \cap [n]} d_k - \sum_{k \in \mathscr{C} \cap [n]} \tilde{d}_k + R(\tilde{\mathscr{C}}). \tag{2.54}$$

Now, all the three terms appearing in the right hand side of (2.54) can be estimated using Lemma 2.4.2. Indeed, we can consider weights $w_{i1} = d_i$, $w_{i2} = \tilde{d}_i$, and $w_{i3} =$ the number of red neighbors of vertex i in $CM_{\tilde{n}}(\tilde{\mathbf{d}})$. The conditions in (2.10) are satisfied by Lemma 2.6.1, and observing that

$$\max\{\max_i w_{i1}, \max_i w_{i2}, \max_i w_{i3}\} \leqslant d_{max} = o(n^{1/3}).$$

Note that, using an argument identical to Lemma 2.6.1, $(1/n) \sum_{i \in [\tilde{n}]} w_{ik} \tilde{d}_i$ converges $\mathbb{P}_p$ almost surely, for all $k = 1, 2, 3$. Now, (2.53) is a consequence of Lemma 2.4.11. Denote

$$D_i = \sum_{k \in \mathscr{C}_{(i)} \cap [n]} d_k, \quad \tilde{D}_i = \sum_{k \in \tilde{\mathscr{C}}_{(i)} \cap [n]} \tilde{d}_k,$$

$$\mathbf{D}_n = ord((D_i)_{i \geqslant 1}), \quad \text{and} \quad \tilde{\mathbf{D}}_n = ord((\tilde{D}_i)_{i \geqslant 1}).$$

Using (2.31), $(\tilde{\mathbf{D}}_n)_{n \geqslant 1}$ is tight in $\ell^2_{\downarrow}$. Further $w_{i3} \leqslant d_i$ for all i. Thus, for the $\ell^2_{\downarrow}$ tightness of $(\mathbf{O}_n(\lambda))_{n \geqslant 1}$, it is enough to show the $\ell^2_{\downarrow}$ tightness of $(\mathbf{D}_n)_{n \geqslant 1}$. Denote the conditional probability, conditioned on the uniform perfect matching in Algorithm 2.3 (S2), by $\tilde{\mathbb{P}}(\cdot)$. Notice that, since Algorithm 2.3 (S1), and (S2) are carried out independently, $\tilde{D}_i \sim Bin(D_i, \sqrt{p_n})$ under $\tilde{\mathbb{P}}$. Using standard concentration inequalities [115, (2.9)], it follows that

$$\tilde{\mathbb{P}}(\tilde{D}_i < D_i \sqrt{p_n}(1 - \sqrt{p_n})) \leqslant 2e^{-D_i p_n^{3/2}/3},$$

and thus for $\mathcal{I} = \{k : D_k > n^\varepsilon\}$, the union bound yields

$$\mathbb{P}(\exists i \in \mathcal{I} : D_i > a\tilde{D}_i) \to 0, \tag{2.55}$$

for some constant $a > 0$. Let $\mathcal{E}_n$ denote the corresponding event in (2.55). Thus, for any $\eta > 0$,

$$\mathbb{P}\left(n^{-4/3} \sum_{k > K, k \in \mathcal{I}} D_k^2 > \eta\right) \leqslant \mathbb{P}\left(n^{-4/3} \sum_{k > K} \tilde{D}_k^2 > \frac{\eta}{a}\right) + \mathbb{P}(\mathcal{E}_n) \to 0,$$

if we first take the limit as $n \to \infty$, and then $K \to \infty$, and use the $\ell^2_{\downarrow}$ tightness of $(\tilde{\mathbf{D}}_n)_{n \geqslant 1}$. Further, $\sum_{k \notin \mathcal{I}} D_k^2 \leqslant n^{1+2\varepsilon} = o(n^{4/3})$, if $\varepsilon < 1/6$. This completes the proof of the $\ell^2_{\downarrow}$ tightness of $(\mathbf{D}_n)_{n \geqslant 1}$ and consequently that of $(\mathbf{O}_n(\lambda))_{n \geqslant 1}$. $\qed$

### 2.7.5    Proof of Theorem 2.2.4

We will consider the case $k = 2$ only, since the case for general $k$ can be proved inductively. Fix $-\infty < \lambda_0 < \lambda_1 < \infty$. Suppose that the modified Algorithm 2.7 starts at time $\lambda_\star = \lambda_0$. By Lemma 2.7.6 and Theorem 2.2.3, $(\mathbf{O}_n(\lambda_0), \kappa \mathbf{C}_n(\lambda_0))$ converges in distribution to $\kappa \sqrt{\nu}(\tilde{\gamma}^{\lambda_0}, \tilde{\gamma}^{\lambda_0})$. Now, from Remark 13, an application of Theorem 2.7.5 gives

$$(\mathbf{C}_n(\lambda_0), \bar{\mathbf{C}}_n(\lambda_1)) \xrightarrow{\mathrm{d}} \sqrt{\nu}(\tilde{\gamma}^{\lambda_0}, \tilde{\gamma}^{\lambda_1}). \tag{2.56}$$

The fact that the limiting distribution corresponding to $\bar{\mathbf{C}}_n(\lambda_1)$ is equal to $\sqrt{\nu}\tilde{\gamma}^{\lambda_1}$ follows from the Feller property of multiplicative coalescent, [11, Theorem 2], and Theorem 2.7.5. For $\mathbf{x}, \mathbf{y} \in \ell^2_\downarrow$, denote $\mathbf{x} \preceq \mathbf{y}$ if $\mathbf{x}$ is the vector in decreasing order of elements $\{y_{ij} : i, j \geqslant 1\}$ such that $\sum_j y_{ij} \leqslant y_i$ for all $i \geqslant 1$. Thus if $\mathbf{y}$ is obtained by *coalescing* elements of $\mathbf{x}$, then $\mathbf{x} \preceq \mathbf{y}$. Under the coupling in Remark 12, it follows that $\mathbf{C}_n(\lambda) \preceq \bar{\mathbf{C}}_n(\lambda)$ almost surely, for each $\lambda \geqslant \lambda_0$. Using [8, Corollary 18 (a)], it follows that

$$\|\bar{\mathbf{C}}_n(\lambda_1) - \mathbf{C}_n(\lambda_1)\|_2^2 \leqslant \|\bar{\mathbf{C}}_n(\lambda_1)\|_2^2 - \|\mathbf{C}_n(\lambda_1)\|_2^2, \tag{2.57}$$

where $\|\cdot\|_2$ denotes the $\ell^2$-norm. The final ingredient is the following straightforward lemma:

**Lemma 2.7.7.** *Suppose $X_n$, $Y_n$ are non-negative random variables such that $X_n \leqslant Y_n$ a.s. and $X_n \xrightarrow{\mathcal{L}} X$, $Y_n \xrightarrow{\mathcal{L}} X$. Then,*

$$Y_n - X_n \xrightarrow{\mathbb{P}} 0.$$

*Proof.* Note that $((X_n, Y_n))_{n \geqslant 1}$ is tight in $\mathbb{R}^2$. Thus, for any $(n'_i)_{i \geqslant 1}$ there exists a subsequence $(n_i)_{i \geqslant 1} \subset (n'_i)_{i \geqslant 1}$ such that $(X_{n_i}, Y_{n_i}) \xrightarrow{\mathcal{L}} (Z_1, Z_2)$. Using the marginal distributional limits we get $Z_1 \overset{\mathcal{L}}{=} X$, $Z_2 \overset{\mathcal{L}}{=} X$. Also the joint distribution of $(Z_1, Z_2)$ is concentrated on the line $y = x$ in the $xy$ plane. Thus, $(X_{n_i}, Y_{n_i}) \xrightarrow{\mathcal{L}} (X, X)$. This limiting distribution does not depend on the subsequence $(n_i)_{i \geqslant 1}$. Thus the tightness of $((X_n, Y_n))_{n \geqslant 1}$ implies $(X_n, Y_n) \xrightarrow{\mathcal{L}} (X, X)$. The proof is now complete. $\qquad\square$

Now, observe that $\|\mathbf{C}_n(\lambda_1)\|_2^2 \leqslant \|\bar{\mathbf{C}}_n(\lambda_1)\|_2^2$ and $\|\mathbf{C}_n(\lambda_1)\|_2^2$, and $\|\bar{\mathbf{C}}_n(\lambda_1)\|_2^2$ have the same distributional limit by Theorem 2.2.2, and (2.56). Thus, using Lemma 2.7.7, it follows that $\|\bar{\mathbf{C}}_n(\lambda_1)\|_2^2 - \|\mathbf{C}_n(\lambda_1)\|_2^2 \xrightarrow{\mathbb{P}} 0$, and (2.56), (2.57) yield

$$(\mathbf{C}_n(\lambda_0), \mathbf{C}_n(\lambda_1)) \xrightarrow{\mathrm{d}} \sqrt{\nu}(\tilde{\gamma}^{\lambda_0}, \tilde{\gamma}^{\lambda_1}).$$

Finally, the proof of Theorem 2.2.4 is completed by applying Proposition 2.7.4.

$\qquad\square$

## 2.8    Conclusion

In this chapter, we have shown that whenever the third moment of the empirical degree distribution converges, the critical window for the configuration model is given by $\nu_n = 1 + \lambda n^{-1/3}$, and the largest component sizes have $\Theta(n^{2/3})$ vertices and $O(1)$ surplus edges. Theorem 2.2.2 identifies the precise limiting distribution of the rescaled component sizes and surplus edges, and the convergence results hold under $\mathbb{U}_\downarrow^0$ topology. We apply these results to percolated $CM_n(\mathbf{d})$. Analyzing the exploration process directly on percolated $CM_n(\mathbf{d})$ is difficult. This is because many paired half-edges are not retained by percolation during the exploration, which changes the number of available half-edges of an unexplored vertex. For this reason, one has to keep updating the degree distribution of unexplored vertices, which becomes difficult to track when the maximum degree is unbounded [135]. We circumvent this difficulty by using Janson's construction of the percolated configuration model. Further, for the joint convergence of the percolated clusters over the critical window, we give a construction of the percolation process in Algorithm 2.4, which allows us to approximate percolated graphs by a dynamically growing Markovian graph process. A further difficulty for proving Theorem 2.2.4 was that even the later Markovian graph process is only *approximately* multiplicative coalescent. The ideas for dealing with this approximate multiplicative coalescent are general, and we believe that these techniques are applicable to many other dynamic graph models.

# Chapter 3

# Heavy-tailed configuration models at criticality

**Abstract.** We study the critical behavior of the component sizes for the configuration model when the tail of the degree distribution of a randomly chosen vertex is a regularly-varying function with exponent $\tau - 1$ with $\tau \in (3, 4)$. The component sizes are shown to be of the order $n^{(\tau-2)/(\tau-1)} L(n)^{-1}$ for some slowly-varying function $L(\cdot)$. We show that the re-scaled ordered component sizes converge in distribution to the ordered excursions of a thinned Lévy process. This proves that the scaling limits for the component sizes for these heavy-tailed configuration models are in a different universality class compared to those for the Erdős-Rényi random graphs. Also the joint re-scaled vector of ordered component sizes and their surplus edges is shown to have a distributional limit under a strong topology. Our proof resolves a conjecture by Joseph, *Ann. Appl. Probab.* (2014) about the scaling limits of uniform simple graphs with i.i.d. degrees in the critical window, and sheds light on the relation between the scaling limits obtained by Joseph and in this chapter, which appear to be quite different. Further, we use percolation to study the evolution of the component sizes and the surplus edges within the critical scaling window, whose finite-dimensional distributions are shown to converge to the augmented multiplicative coalescent process introduced by Bhamidi et al., *Probab. Theory Related Fields* (2014). The main results of this chapter are proved under rather general assumptions on the vertex degrees. We also discuss how these assumptions are satisfied by some of the frameworks that have been studied previously.

In this chapter, we focus on the critical behavior of the configuration model, and critical percolation when the third moment of the empirical degree distribution tends to infinity. We include detailed proofs of the scaling limit results related to component sizes and surplus edges described in Chapter 1. As in Chapter 2, the scaling limit result are shown to hold for critical percolation on the configuration model. We also study the evolution of both component sizes and surplus edges over the critical window, and describe its asymptotic distribution by a version of the augmented multiplicative coalescent process. The scaling limits lie in the universality class identified in [33], and are fundamentally different than in Chapter 2. The results in this chapter provide a detailed understanding of the component sizes and surplus edges for heavy-tailed graphs in the critical window. Before stating the main results, we need to introduce some notation and concepts.

## 3.1   Definitions and notation

Recall the definitions from Chapter 1.2.1. Let $(\mathbb{U}_\downarrow^0)^k$ denote the k-fold product space of $\mathbb{U}_\downarrow^0$. For any $\mathbf{z} \in \mathbb{U}_\downarrow$, $\mathrm{ord}(\mathbf{z})$ will denote the element of $\mathbb{U}_\downarrow^0$ obtained by suitably ordering the coordinates of $\mathbf{z}$. We often use the boldface notation $\mathbf{X}$ for the process $(X(s))_{s \geqslant 0}$, unless stated otherwise. $\mathbb{D}[I, E]$ will denote the space of càdlàg functions from an interval I to the metric space $E = (E, d)$ equipped with Skorohod $J_1$-topology. Consider a decreasing sequence $\theta = (\theta_1, \theta_2, \dots) \in \ell_\downarrow^3 \setminus \ell_\downarrow^2$. Denote by $\mathcal{I}_i(s) := \mathbf{1}\{\xi_i \leqslant s\}$ where $\xi_i \sim \mathrm{Exp}(\theta_i/\mu)$ independently, and $\mathrm{Exp}(r)$ denotes the exponential distribution with rate r. Consider the process

$$\bar{S}_\infty^\lambda(t) = \sum_{i=1}^\infty \theta_i \left( \mathcal{I}_i(t) - (\theta_i/\mu)t \right) + \lambda t, \tag{3.1}$$

for some $\lambda \in \mathbb{R}, \mu > 0$ and define the reflected version of $\bar{S}_\infty^\lambda(t)$ by

$$\mathrm{refl}(\bar{S}_\infty^\lambda(t)) = \bar{S}_\infty^\lambda(t) - \min_{0 \leqslant u \leqslant t} \bar{S}_\infty^\lambda(u). \tag{3.2}$$

The process of the form (3.1) was termed *thinned* Lévy processes in [33] (see also [5, 102]), since the summands are thinned versions of Poisson processes. For any function $f \in \mathbb{D}[0, \infty)$, define $\underline{f}(x) = \inf_{y \leqslant x} f(y)$. $\mathbb{D}_+[0, \infty)$ is the subset of $\mathbb{D}[0, \infty)$ consisting of functions with positive jumps only. Note that $\underline{f}$ is continuous when $f \in \mathbb{D}_+[0, \infty)$. An *excursion* of a function $f \in \mathbb{D}_+[0, T]$

is an interval $(l, r)$ such that

$$\min\{f(l-), f(l)\} = \underline{f}(l) = \underline{f}(r) = \min\{f(r-), f(r)\}$$
$$\text{and} \quad f(x) > \underline{f}(r), \ \forall x \in (l, r) \subset [0, T].$$

Excursions of a function $f \in \mathbb{D}_+[0, \infty)$ are defined similarly. We will use $\gamma$ to denote an excursion, as well as the length of the excursion $\gamma$ to simplify notation.

Also, define the counting process $\mathbf{N}$ to be the Poisson process that has intensity $\mathrm{refl}(\bar{S}_\infty^\lambda(t))$ at time $t$ conditional on $(\bar{S}_\infty^\lambda(u))_{u \leqslant t}$. Formally, $\mathbf{N}$ is characterized as the counting process for which

$$N(t) - \int_0^t \mathrm{refl}(\bar{S}_\infty^\lambda(u)) du \tag{3.3}$$

is a martingale. We use the notation $N(\gamma)$ to denote the number of marks in the interval $\gamma$.

Finally, we define a Markov process $(\mathbf{Z}(s))_{s \in \mathbb{R}}$ on $\mathbb{D}(\mathbb{R}, \mathbb{U}_\downarrow^0)$, called the augmented multiplicative coalescent (AMC) process. Think of a collection of particles in a system with $\mathbf{X}(s)$ describing their masses and $\mathbf{Y}(s)$ describing an additional attribute at time $s$. Let $K_1, K_2 > 0$ be constants. The evolution of the system takes place according to the following rule at time $s$:

▷ For $i \neq j$, at rate $K_1 X_i(s) X_j(s)$, the $i^{th}$ and $j^{th}$ component merge and create a new component of mass $X_i(s) + X_j(s)$ and attribute $Y_i(s) + Y_j(s)$.

▷ For any $i \geqslant 1$, at rate $K_2 X_i^2(s)$, $Y_i(s)$ increases to $Y_i(s) + 1$.

Of course, at each event time, the indices are re-organized to give a proper element of $\mathbb{U}_\downarrow^0$. This process was first introduced in [30] to study the joint behavior of the component sizes and the surplus edges over the critical window. In [30], the authors extensively study the properties of the standard version of AMC, i.e., the case $K_1 = 1, K_2 = 1/2$ and showed in [30, Theorem 3.1] that this is a (nearly) Feller process, a property that will play a crucial rule in the final part of this chapter.

**Remark 14.** Notice that the summation term in (3.1), after replacing $\theta_i$ by $\mu\theta_i$, is of the form $V^\theta(s) = \mu^\alpha \sum_{i=1}^\infty \left(\theta_i \mathbf{1}\{\xi_i \leqslant s\} - \theta_i^2 s\right)$, where $\xi_i \sim \mathrm{Exp}(\theta_i)$ independently over $i$ and $\theta \in \ell_\downarrow^3 \setminus \ell_\downarrow^2$. Therefore, by [11, Lemma 1], the process $\mathrm{refl}(\bar{S}_\infty^\lambda)$ has no infinite excursions almost surely and only finitely many excursions with length at least $\delta$, for any $\delta > 0$.

## 3.2    Main results

### 3.2.1    Main results for critical configuration models

Throughout this chapter we will use the shorthand notation

$$\alpha = 1/(\tau - 1), \qquad \rho = (\tau - 2)/(\tau - 1), \qquad \eta = (\tau - 3)/(\tau - 1),$$
$$a_n = n^\alpha L(n), \qquad b_n = n^\rho (L(n))^{-1}, \qquad c_n = n^\eta (L(n))^{-2}, \tag{3.4}$$

where $\tau \in (3, 4)$ and $L(\cdot)$ is a slowly-varying function. We state our results under the following assumptions:

**Assumption 3.1.** Fix $\tau \in (3, 4)$. Let $\mathbf{d} = (d_1, \ldots, d_n)$ be a degree sequence such that the following conditions hold:

(i) (*High-degree vertices*) For any fixed $i \geqslant 1$,

$$\frac{d_i}{a_n} \to \theta_i,$$

where $\theta = (\theta_1, \theta_2, \ldots) \in \ell_\downarrow^3 \setminus \ell_\downarrow^2$.

(ii) (*Moment assumptions*) Let $D_n$ denote the degree of a vertex chosen uniformly at random from $[n]$, independently of $CM_n(\mathbf{d})$. Then, $D_n \xrightarrow{d} D$, for some integer-valued random variable $D$ and

$$\frac{1}{n} \sum_{i \in [n]} d_i \to \mu := \mathbb{E}[D], \quad \frac{1}{n} \sum_{i \in [n]} d_i^2 \to \mathbb{E}[D^2],$$
$$\lim_{K \to \infty} \limsup_{n \to \infty} \left( a_n^{-3} \sum_{i=K+1}^{n} d_i^3 \right) = 0.$$

(iii) (*Critical window*) For some $\lambda \in \mathbb{R}$,

$$\nu_n(\lambda) := \frac{\sum_{i \in [n]} d_i(d_i - 1)}{\sum_{i \in [n]} d_i} = 1 + \lambda c_n^{-1} + o(c_n^{-1}).$$

(iv) Let $n_1$ be the number of vertices of degree-one. Then $n_1 = \Theta(n)$, which is equivalent to assuming that $\mathbb{P}(D = 1) > 0$.

Note that Assumption 3.1 (i)-(ii) implies $\liminf_{n \to \infty} \mathbb{E}[D_n^3] = \infty$. The following three results hold for any $CM_n(\mathbf{d})$ satisfying Assumption 3.1:

**Theorem 3.2.1.** *Consider* $CM_n(\mathbf{d})$ *with the degrees satisfying Assumption 3.1. Denote the* $i^{th}$*-largest cluster of* $CM_n(\mathbf{d})$ *by* $\mathscr{C}_{(i)}$*. Then,*

$$\left(b_n^{-1}|\mathscr{C}_{(i)}|\right)_{i\geqslant 1} \xrightarrow{d} (\gamma_i(\lambda))_{i\geqslant 1},$$

*with respect to the* $\ell^2_\downarrow$*-topology where* $\gamma_i(\lambda)$ *is the length of the* $i^{th}$ *largest excursion of the process* $\bar{\mathbf{S}}^\lambda_\infty$*, while* $b_n$ *and the constants* $\lambda, \mu$ *are defined in (3.4) and Assumption 3.1.*

**Theorem 3.2.2.** *Consider* $CM_n(\mathbf{d})$ *with the degrees satisfying Assumption 3.1. Let* $SP(\mathscr{C}_{(i)})$ *denote the number of surplus edges in* $\mathscr{C}_{(i)}$ *and define the vectors* $\mathbf{Z}_n := \mathrm{ord}(b_n^{-1}|\mathscr{C}_{(i)}|, SP(\mathscr{C}_{(i)}))_{i\geqslant 1}$, $\mathbf{Z} := \mathrm{ord}(\gamma_i(\lambda), N(\gamma_i))_{i\geqslant 1}$. *Then, as* $n \to \infty$,

$$\mathbf{Z}_n \xrightarrow{d} \mathbf{Z} \tag{3.5}$$

*with respect to the* $\mathbb{U}^0_\downarrow$ *topology, where* $\mathbf{N}$ *is defined in (3.3).*

**Theorem 3.2.3.** *The results in Theorems 3.2.1 and 3.2.2 also hold for uniform random graphs with degree* $\mathbf{d}$*.*

**Remark 15.** The only previous work to understand the critical behavior of the configuration model with heavy-tailed degrees was by Joseph [121] where the degrees were assumed to be i.i.d an sample from an exact power-law distribution and the results were obtained for the component sizes of $CM_n(\mathbf{d})$ (Theorem 3.2.1). We will see that Assumption 3.1 is satisfied for i.i.d degrees in Section 3.3.2. Thus, a quenched version of [121, Theorem 8.3] follows from our results. Further, if the degrees are chosen approximately as the weights chosen in [33], then our results continue to hold. This sheds light on the relation between the scaling limits in [33] and [121] (see Remark 24). Moreover, Theorem 3.2.3 resolves [121, Conjecture 8.5].

**Remark 16.** The conclusions of Theorems 3.2.1, 3.2.2, and 3.2.3 hold for more general functionals of the components. Suppose that each vertex $i$ has a weight $w_i$ associated to it and let $\mathscr{W}_i$ denote the total weight of the component $\mathscr{C}_{(i)}$, i.e., $\mathscr{W}_i = \sum_{k \in \mathscr{C}_{(i)}} w_k$. Then, under some regularity conditions on the weight sequence $\mathbf{w} = (w_i)_{i \in [n]}$, in Section 3.8 we will show that the scaling limit for $\mathbf{Z}_n^w := \mathrm{ord}(b_n^{-1}\mathscr{W}_i, SP(\mathscr{C}_{(i)}))_{i\geqslant 1}$ is given by $\mathbf{Z} = \mathrm{ord}(\kappa\gamma_i(\lambda), N(\gamma_i))_{i\geqslant 1}$, where the constant $\kappa$ is given by

$$\kappa = \lim_{n\to\infty} \frac{\sum_{i\in[n]} d_i w_i}{\sum_{i\in[n]} d_i}.$$

Observe that, for $w_i = \mathbf{1}\{d_i = k\}$, $\mathscr{W}_i$ gives the asymptotic number of vertices of degree $k$ in the $i^{th}$ largest component.

**Remark 17.** It might not be immediate why we should work with Assumption 3.1. We will see in Section 3.3.1 that Assumption 3.1 is satisfied by the degree sequences in some important and natural cases. The reason to write the assumptions in this form is to make the properties of the degree distribution explicit (e.g. in terms of moment conditions and the asymptotics of the highest degrees) that jointly lead to this universal critical limiting behavior. We explain the significance of Assumption 3.1 in more detail in Section 3.4.

### 3.2.2   Percolation on heavy-tailed configuration models

Percolation refers to deleting each edge of a graph independently with probability $1 - p$. Consider percolation on a configuration model $CM_n(\mathbf{d})$ under the following assumptions:

**Assumption 3.2.**    (i) Assumption 3.1 (i), and (ii) hold for the degree sequence and $CM_n(\mathbf{d})$ is super-critical, i.e.,

$$\nu_n = \frac{\sum_{i \in [n]} d_i(d_i - 1)}{\sum_{i \in [n]} d_i} \to \nu = \frac{\mathbb{E}\,[D(D - 1)]}{\mathbb{E}\,[D]} > 1.$$

(ii) (*Critical window for percolation*) The percolation parameter $p_n$ satisfies

$$p_n = p_n(\lambda) := \frac{1}{\nu_n}\left(1 + \lambda c_n^{-1} + o(c_n^{-1})\right)$$

for some $\lambda \in \mathbb{R}$.

Let $CM_n(\mathbf{d}, p_n(\lambda))$ denote the graph obtained through percolation on $CM_n(\mathbf{d})$ with bond retention probability $p_n(\lambda)$. The following result gives the asymptotics for the ordered component sizes and the surplus edges for $CM_n(\mathbf{d}, p_n(\lambda))$:

**Theorem 3.2.4.** *Consider* $CM_n(\mathbf{d}, p_n(\lambda))$ *satisfying Assumption 3.2. Let* $\tilde{\mathbf{S}}_\infty^\lambda$ *denote the process in (3.1) with* $\theta_i$ *replaced by* $\theta_i/\sqrt{\nu}$, *and* $\mathscr{C}_{(i)}^p$ *denote the* $i^{th}$ *largest component of* $CM_n(\mathbf{d}, p_n)$ *and let* $\mathbf{Z}_n^p(\lambda) := \mathrm{ord}(b_n^{-1}|\mathscr{C}_{(i)}^p|, SP(\mathscr{C}_{(i)}^p))_{i \geqslant 1}$, $\mathbf{Z}^p(\lambda) := \mathrm{ord}((\nu^{1/2}\tilde{\gamma}_i(\lambda), N(\tilde{\gamma}_i(\lambda)))_{i \geqslant 1}$, *where* $\tilde{\gamma}_i(\lambda)$ *is the largest excursion of* $\tilde{\mathbf{S}}_\infty^\lambda$. *Then, for any* $\lambda \in \mathbb{R}$, *as* $n \to \infty$,

$$\mathbf{Z}_n^p(\lambda) \xrightarrow{d} \mathbf{Z}^p(\lambda) \tag{3.6}$$

*with respect to the* $\mathbb{U}_\downarrow^0$ *topology.*

Now, consider a graph $CM_n(\mathbf{d})$ satisfying Assumption 3.2 (i). To any edge $(ij)$ between vertices $i$ and $j$ (if any), associate an independent uniform

random variable $U_{(ij)}$. Note that the graph obtained by keeping only those edges satisfying $U_{(ij)} \leqslant p_n(\lambda)$ is distributed as $CM_n(\mathbf{d}, p_n(\lambda))$. This construction naturally couples the graphs $(CM_n(\mathbf{d}, p_n(\lambda)))_{\lambda \in \mathbb{R}}$ using the same set of uniform random variables. Our next result shows that the evolution of the component sizes and the surplus edges of $CM_n(\mathbf{d}, p_n(\lambda))$, as $\lambda$ varies, can be described by a version of the augmented multiplicative coalescent process described in Section 3.1:

**Theorem 3.2.5.** *Suppose that Assumption 3.2 holds, and $\ell_n/n = \mu + o(n^{-\zeta})$ for some $\eta < \zeta < 1/2$. Fix any $k \geqslant 1$, $-\infty < \lambda_1 < \cdots < \lambda_k < \infty$. Then, there exists a version $\mathbf{AMC} = (\mathbf{AMC}(\lambda))_{\lambda \in \mathbb{R}}$ of the augmented multiplicative coalescent such that, as $n \to \infty$,*

$$(\mathbf{Z}_n^p(\lambda_1), \ldots \mathbf{Z}_n^p(\lambda_k)) \xrightarrow{d} (\mathbf{AMC}(\lambda_1), \ldots, \mathbf{AMC}(\lambda_k))$$

*with respect to the $(\mathbb{U}_\downarrow^0)^k$ topology, where at each $\lambda$, $\mathbf{AMC}(\lambda)$ is distributed as the limiting object in* (3.6).

**Remark 18.** Theorem 3.2.5 also holds when $\mathbb{E}[D_n^3] \to \mathbb{E}[D^3] < \infty$ with $\alpha = \eta = 1/3$, $\rho = 2/3$ and $L(n) = 1$. This improves [70, Theorem 4], which was proved only for the cluster sizes.

**Remark 19.** Theorem 3.2.5, in fact, shows that there exists a version of the AMC process whose distribution at each fixed $\lambda$ can be described by the excursions of a thinned Lévy process and an associated Poisson process. This did not appear in [30, 57], since the scaling limits in their settings were described in terms of the excursions of a Brownian motion with parabolic drift.

**Remark 20.** The additional assumption in Theorem 3.2.5 about the asymtotics $\ell_n/n$ is required only in one place for Proposition 3.10.1 and the rest of the proof works under Assumption 3.2 only. That is why we have separated this assumption from the set of conditions in Assumption 3.2. It is worthwhile mentioning that the condition is not stringent at all, e.g., we will see that this condition is satisfied under the two widely studied set-ups in Section 3.3.1.

**Remark 21.** As we will see in Section 3.10, the proof of Theorem 3.2.5 can be extended to more general functionals of the components. For example, the evolution of the number of degree $k$ vertices along with the surplus edges can also be described by an AMC process. The key idea here is that these component functionals become approximately proportional to the component sizes

in the critical window and thus the scaling limit for the component functionals becomes a constant multiple of the scaling limit for the component sizes.

## 3.3    Important examples

### 3.3.1    Power-law degrees with small perturbation

As discussed in the introduction, our main goal is to obtain results for the critical configuration model with $\mathbb{P}(D_n \geqslant k) \sim L_0(k)k^{-(\tau-1)}$ for some $\tau \in (3,4)$. In this section, we consider such an example and show that the conditions of Assumption 3.1 are satisfied. Thus, the results in Section 3.2.1 hold for $CM_n(\mathbf{d})$ in the following set-up that is closely related to the model studied in [33] for rank-1 inhomogeneous random graphs.

Fix $\tau \in (3,4)$. Suppose that $F$ is the distribution function of a discrete non-negative random variable $D$ such that

$$G(x) = 1 - F(x) = \frac{C_F L_0(x)}{x^{\tau-1}}(1 + o(1)) \quad \text{as } x \to \infty, \tag{3.7}$$

where $L_0(\cdot)$ is a slowly-varying function so that the tail of the distribution is decaying like a regularly-varying function. Recall that the inverse of a locally bounded non-increasing function $f : \mathbb{R} \to \mathbb{R}$ is defined as $f^{-1}(x) := \inf\{y : f(y) \leqslant x\}$. Therefore, using [38, Theorem 1.5.12],

$$G^{-1}(x) = \frac{C_F^{1/(\tau-1)} L(1/x)}{x^{1/(\tau-1)}}(1 + o(1)) \quad \text{as } x \to 0, \tag{3.8}$$

where $L(\cdot)$ is another slowly-varying function. Note that [38, Theorem 1.5.12] is stated for positive exponents only. Since our exponent is negative, the asymptotics in (3.8) holds for $x \to 0$. Suppose that the random variable $D$ is such that

$$\nu = \frac{\mathbb{E}[D(D-1)]}{\mathbb{E}[D]} = 1. \tag{3.9}$$

Define the degree sequence $\mathbf{d}_\lambda$ by taking the degree of the $i^{th}$ vertex to be

$$d_i = d_i(\lambda) := G^{-1}(i/n) + \delta_{i,n}(\lambda), \tag{3.10}$$

where the $\delta_{i,n}(\lambda)$'s are non-negative integers satisfying the asymptotic equivalence

$$\delta_{i,n}(\lambda) \sim \lambda G^{-1}(i/n)c_n^{-1}, \quad \text{as } n \to \infty. \tag{3.11}$$

The $\delta_{i,n}(\lambda)$'s are chosen in such a way that Assumption 3.1 (iv) is satisfied. Fix any $K \geqslant 1$. Notice that (3.8) and (3.11) imply that, for all large enough $n$ (independently of K), the first K largest degrees $(d_i)_{i\in[K]}$ satisfy

$$d_i = \left(\frac{n^\alpha C_F^\alpha L(n/i)}{i^\alpha}\right)\left(1 + \lambda c_n^{-1} + o(c_n^{-1})\right). \qquad (3.12)$$

Therefore, $\mathbf{d}_\lambda$ satisfies Assumption 3.1 (i) with $\theta_i = (C_F/i)^\alpha$. The next two lemmas verify Assumption 3.1 (ii), (iii):

**Lemma 3.3.1.** *The degree sequence $\mathbf{d}_\lambda$ defined in (3.10) satisfies Assumption 3.1 (ii).*

*Proof.* Note that, by (3.12), $d_1^2 = o(n)$. Also, since $G^{-1}$ is non-increasing

$$\int_0^1 G^{-1}(x)dx - \frac{d_1}{n} \leqslant \frac{1}{n}\sum_{i\in[n]} G^{-1}(i/n) \leqslant \int_0^1 G^{-1}(x)dx.$$

Therefore,

$$\frac{1}{n}\sum_{i\in[n]} d_i = \frac{1}{n}\sum_{i\in[n]} G^{-1}(i/n)(1 + O(c_n^{-1}))$$
$$= \int_0^1 G^{-1}(x)dx + O(d_1/n) + O(c_n^{-1}) = \mathbb{E}[D] + O(b_n^{-1}). \qquad (3.13)$$

Similarly, $\sum_{i\in[n]} d_i^2 = n\mathbb{E}[D^2] + O(d_1^2) = n\mathbb{E}[D^2] + o(n)$. To prove the condition involving the third-moment, we use Potter's theorem [38, Theorem 1.5.6]. First note that $3\alpha - 1 = (4 - \tau)/(\tau - 1) > 0$ since $\tau \in (3,4)$. Fix $0 < \delta < \alpha - 1/3$ and $A > 1$ and choose $C = C(\delta, A)$ such that for all $i \leqslant nC^{-1}$, $L(n/i)/L(n) < Ai^\delta$. Therefore, (3.8) implies

$$a_n^{-3}\sum_{i>K} d_i^3 \leqslant A\sum_{i>K} i^{-3\alpha+3\delta} + \frac{\sup_{1\leqslant x\leqslant C} L(x)^3}{L(n)^3}\sum_{i>nC^{-1}} i^{-3\alpha}. \qquad (3.14)$$

From our choice of $\delta$, $-3\alpha + 3\delta < -1$ and therefore $\sum_{i\geqslant 1} i^{-3\alpha+3\delta} < \infty$. By [38, Lemma 1.3.2], $\sup_{1\leqslant x\leqslant C} L(x)^3 < \infty$. Moreover, $\sum_{i>nC^{-1}} i^{-3\alpha} = O(n^{1-3\alpha})$ and $1 - 3\alpha < 0$. Thus, the proof follows by first taking $n \to \infty$ and then $K \to \infty$. $\qquad \square$

**Lemma 3.3.2.** *The degree sequence $\mathbf{d}_\lambda$ defined in (3.10) satisfies Assumption 3.1 (iii), i.e., there exists $\lambda_0 \in \mathbb{R}$ such that*

$$\nu_n(\lambda) = 1 + (\lambda + \lambda_0)c_n^{-1} + o(c_n^{-1}). \qquad (3.15)$$

*Proof.* Firstly, Lemma 3.3.1 guarantees the convergence of the second moment of the degree sequence. However, (3.15) is more about obtaining sharper asymptotics for $\nu_n(\lambda)$. We use similar arguments as in [33, Lemma 2.2]. Denote $\nu_n := \nu_n(0)$. Note that $\nu_n(\lambda) = \nu_n(1 + \lambda c_n^{-1}) + o(c_n^{-1})$, so it is enough to verify that

$$\nu_n = 1 + \lambda_0 c_n^{-1} + o(c_n^{-1}).$$

Consider $d_i(0)$ as given in (3.10) with $\lambda = 0$. Lemma 3.3.1 implies

$$\nu_n = \frac{\sum_{i \in [n]} d_i(0)^2}{n \mathbb{E}[D]} - 1 + o(c_n^{-1}).$$

Fix any $K \geqslant 1$. We have

$$\int_{K/n}^1 G^{-1}(u)^2 du - \frac{d_K^2}{n} \leqslant \frac{1}{n} \sum_{i=K+1}^n d_i^2 \leqslant \int_{K/n}^1 G^{-1}(u)^2 du.$$

Now by (3.10), $d_K^2/n = \Theta(K^{-2\alpha} L(n/K)^2 n^{-\eta})$. Therefore,

$$\nu - \nu_n$$

$$= \frac{1}{\mathbb{E}[D]} \left( \sum_{i=1}^K \int_{(i-1)/n}^{i/n} G^{-1}(u)^2 du - \frac{1}{n} \sum_{i=1}^K d_i^2 \right) + O(K^{-2\alpha} L(n/K)^2 n^{-\eta}). \tag{3.16}$$

Again, using (3.10),

$$\frac{1}{n} \sum_{i=1}^K d_i^2 = n^{-\eta} \sum_{i=1}^K \left( \frac{C_F}{i} \right)^{2\alpha} L(n/i)^2 + o(c_n^{-1}) = c_n^{-1} \sum_{i=1}^K \left( \frac{C_F}{i} \right)^{2\alpha} + \varepsilon(c_n, K), \tag{3.17}$$

where the last equality follows using the fact that $L(\cdot)$ is a slowly-varying function. Note that the error term $\varepsilon(c_n, K)$ in (3.17) satisfies $\lim_{n \to \infty} c_n \varepsilon(c_n, K) = 0$ for each fixed $K \geqslant 1$. Again,

$$\sum_{i=1}^K \int_{(i-1)/n}^{i/n} G^{-1}(u)^2 du = n^{-\eta} \sum_{i=1}^K \int_{(i-1)}^i \left( \frac{C_F}{u} \right)^{2\alpha} L(n/u)^2 du + o(c_n^{-1})$$

$$= c_n^{-1} \sum_{i=1}^K \int_{(i-1)}^i \left( \frac{C_F}{u} \right)^{2\alpha} du + \varepsilon'(c_n, K), \tag{3.18}$$

where $\lim_{n \to \infty} c_n \varepsilon'(c_n, K) = 0$ for each fixed $K \geqslant 1$. Thus combining (3.16), (3.17), and (3.18) and first letting $n \to \infty$ and then $K \to \infty$, we get

$$\lim_{n \to \infty} c_n(\nu_n - \nu) = \lambda_0,$$

where

$$\lambda_0 = -\frac{C_F^{2\alpha}}{\mathbb{E}\,[D]} \sum_{i=1}^{\infty} \left( \int_{i-1}^{i} u^{-2\alpha} du - i^{-2\alpha} \right).$$

Using Euler-Maclaurin summation [88, Page 333] it can be seen that $\lambda_0$ is finite which completes the proof. $\qquad\square$

**Remark 22.** Note that if we add approximately $cn^{1-\eta}$ ($c > 0$ is a constant) ones in the degree sequence given in (3.10), then we end up with another configuration model for which $\lim_{n\to\infty} n^{\eta}(\nu_n - \nu) = \zeta'$ with $\zeta > \zeta'$. Similarly, deleting $cn^{1-\eta}$ ones from the degree sequence increases the new $\zeta$ value. This gives an obvious way to perturb the degree sequence in such a way that the configuration model is in different locations within the critical scaling window. In our proofs, we will only use the precise asymptotics of the *high*-degree vertices. Thus, a small (suitable) perturbation in the degrees of the *low*-degree vertices does not change the scaling behavior fundamentally, except for a change in the location inside the scaling window.

**Remark 23.** If $\nu$ in (3.9) is larger than one, then the degree sequence satisfies Assumption 3.2. Therefore, the results for critical percolation also hold in this setting. (3.13) implies that the additional assumption in Theorem 3.2.5 is also satisfied.

### 3.3.2    Random degrees sampled from a power-law distribution

We now consider the set-up discussed in [121]. Let $D_1, \ldots, D_n$ be i.i.d samples from a distribution $F$, where $F$ is defined in (3.7). Therefore, the asymptotic relation in (3.8) holds. Consider the random degree sequence $\mathbf{d}$ where $d_i = D_{(i)}$, $D_{(i)}$ being the $i^{\text{th}}$ order statistic of $(D_1, \ldots, D_n)$. We show that $\mathbf{d}$ satisfies Assumption 3.1 almost surely under a suitable coupling. We use a coupling from [53, Section 13.6]. Let $(E_1, E_2, \ldots)$ be an i.i.d sequence of unit rate exponential random variables and let $\Gamma_i := \sum_{j=1}^{i} E_j$. Let

$$\bar{d}_i = \bar{D}_{(i)} = G^{-1}(\Gamma_i/\Gamma_{n+1}).$$

It can be checked that $(d_1, \ldots, d_n) \stackrel{d}{=} (\bar{d}_1, \ldots, \bar{d}_n)$ and therefore, we will ignore the bar in the subsequent notation. Note that, by the strong law of large numbers, $\Gamma_{n+1}/n \stackrel{\text{a.s.}}{\longrightarrow} 1$. Thus, for each fixed $i \geqslant 1$, $\Gamma_{n+1}/(n\Gamma_i) \stackrel{\text{a.s.}}{\longrightarrow} 1/\Gamma_i$. Using (3.8), we see that $\mathbf{d}$ satisfies Assumption 3.1 (i) almost surely under this coupling with $\theta_i = (C_F/\Gamma_i)^{\alpha}$. The first two conditions of Assumption 3.1 (ii)

are trivially satisfied by $\mathbf{d}$ almost surely using the strong law of large numbers. Using the third condition, we first claim that

$$\mathbb{P}\left(\sum_{i=1}^{\infty} \Gamma_i^{-3\alpha} < \infty\right) = 1. \tag{3.19}$$

To see (3.19), note that $\Gamma_i$ has a Gamma distribution with shape parameter $i$ and scale parameter 1. Thus, for $i > 3\alpha$,

$$\mathbb{E}[\Gamma_i^{-3\alpha}] = \frac{\Gamma(i-3\alpha)}{\Gamma(i)} = i^{-3\alpha}(1+O(1/i)),$$

where $\Gamma(x)$ is the Gamma function and the last equality follows from Stirling's approximation. Therefore,

$$\mathbb{E}\left[\sum_{i=1}^{\infty} \Gamma_i^{-3\alpha}\right] = \sum_{i=1}^{\infty} \mathbb{E}\left[\Gamma_i^{-3\alpha}\right] < \infty$$

and (3.19) follows. Now, using the fact that $\Gamma_{n+1}/n \xrightarrow{\text{a.s.}} 1$, we can use arguments identical to (3.14) to show that $\lim_{K\to\infty} \limsup_{n\to\infty} a_n^{-3} \sum_{i>K} d_i^3 = 0$ on the event $\{\sum_{i=1}^{\infty} \Gamma_i^{-3\alpha} < \infty\} \cap \{\Gamma_{n+1}/n \to 1\}$. Thus, we have shown that the third condition of Assumption 3.1 (ii) holds almost surely.

To see Assumption 3.1 (iii), an argument similar to Lemma 3.3.2 can be carried out to prove that

$$\lim_{n\to\infty} c_n(\nu_n - \nu) \xrightarrow{\text{a.s.}} \Lambda_0,$$

where

$$\Lambda_0 := -\frac{C_F^{2\alpha}}{\mathbb{E}[D]} \sum_{i=1}^{\infty} \left(\int_{\Gamma_{i-1}}^{\Gamma_i} u^{-2\alpha} du - \Gamma_i^{-2\alpha}\right). \tag{3.20}$$

Therefore, the results in Section 3.2.1 hold conditionally on the degree sequence if we assume the degrees to be i.i.d samples from a distribution of the form (3.7). For the percolation results, notice that the additional condition in Theorem 3.2.5 is a direct consequence of the convergence rates of sums of i.i.d sequences of random variables [122, Corollary 3.22].

**Remark 24.** Let us recall the limiting object obtained in [121, Theorem 8.1] and compare this with the limiting object $\bar{\mathbf{S}}_\infty^{\Lambda_0}$, defined in (3.1) with $\Lambda_0$ given by (3.20). We will prove an analogue of [121, Theorem 8.1] in Theorem 3.5.1. Although we use a different exploration process from [121], the fact that the component sizes are *huge* compared to the number of cycles in a component, means that one can prove Theorem 3.5.1 for the exploration process in [121]

also. This will indirectly imply that Joseph's limiting exploration process in [121, Theorem 8.1] obeys the law of $\bar{\mathbf{S}}_\infty^{\wedge_0}$, averaged out over the $\Gamma$-values. This is counter intuitive, given the vastly different descriptions of the two processes; for example our process does not have independent increments. We do not have a direct way to prove the above mentioned claim.

## 3.4 Discussion

**Assumptions on the degree distribution.** Let us now briefly explain the significance of Assumption 3.1. Unlike the finite third-moment case [70], the high-degree vertices dictate the scaling limit in Theorem 3.2.1 and therefore it is essential to fix their asymptotics through Assumption 3.1 (i). Assumption 3.1 (iii) defines the critical window of the phase transition and Assumption 3.1 (iv) is reminiscent of the fact that a configuration model with negligibly small amount of degree-one vertices is always supercritical. Assumption 3.1 (ii) states the finiteness of the first two moments of the degree distribution and fixes the asymptotic order of the third moment. The order of the third moment is crucial in our case. The derivation of the scaling limits for the components sizes is based on the analysis of a walk which encodes the information about the component sizes in terms of the excursions above its past minima [8, 32, 33, 70, 144]. Now, the increment distribution turns out to be the size-biased distribution with the sizes being the degrees. Therefore, the third-moment assumption controls the variance of the increment distribution. Another viewpoint is that the components can be locally approximated by a branching process $\mathcal{X}_n$ with the variance of the same order as the third moment of the degree distribution. Thus Assumption 3.1 (ii) controls the order of the survival probability of $\mathcal{X}_n$, which is intimately related to the asymptotic size of the largest components.

**Connecting the barely subcritical and supercritical regimes.** The barely subcritical (supercritical) regime corresponds to the case when $\nu_n(\lambda_n) = 1 + \lambda_n c_n^{-1}$ for some $\lambda_n \to -\infty$ ($\lambda_n \to \infty$) and $\lambda_n = o(c_n^{-1})$. Janson [108] showed that the size of the $k^{\text{th}}$ largest cluster for a subcritical configuration model (i.e., the case $\nu_n \to \nu$ and $\nu < 1$) is $d_k/(1-\nu)$ (see [108, Remark 1.4]). In [31], we show that this is indeed the case for the entire barely subcritical regime, i.e., the size of the $k^{\text{th}}$ largest cluster is $d_k/(1 - \nu_n(\lambda_n)) = \Theta(b_n|\lambda_n|^{-1})$. In the barely supercritical case, the giant component can be *locally* approximated by a branching process $\mathcal{X}_n$ having variance of the order $a_n^3/n$ and the size of the giant component is of the order $n\rho_n$, where $\rho_n$ is the survival proba-

bility of $\mathcal{X}_n$ [101]. The asymptotic size of the giant component turns out to be $\Theta(b_n|\lambda_n|)$. Therefore, the fact that the sizes of the maximal components in the critical scaling window are $\Theta(b_n)$ for $\lambda_n = \Theta(1)$ proves a continuous phase transition property for the configuration model within the whole critical regime.

**Percolation.** The main reason to study percolation in this chapter is to understand the evolution of the component sizes and the surplus edges over the critical window in Theorem 3.2.5. It turns out that a precise characterization of the evolution of the percolation clusters is necessary for understanding the minimal spanning tree of the giant component with i.i.d. weights on each edge [4]. Also, since the percolated configuration model is again a configuration model [85, 109], the natural way to study the evolution of the clusters sizes of configuration models over the critical window is through percolation.

**Universality.** The limiting object in Theorem 3.2.1 is identical to that in [33, Theorem 1.1] for rank-1 inhomogeneous random graphs. Thus, $CM_n(\mathbf{d})$ with regularly-varying tails falls onto the domain of attraction of the new universality class studied in [33]. This is again confirming the predictions made by statistical physicists that the nature of the phase transition does not depend on the precise details of the model. Our scaling limit fits into the general class of limits predicted in [11]. In the notation of [11, (6)], the scaling limits $CM_n(\mathbf{d})$, under Assumption 3.1, give rise to the case $\kappa = 0$. To understand this, let us discuss some existing works. In [8, 15, 32, 121], and Chapter 2 the limiting component sizes are described by the excursions of a Brownian motion with a parabolic drift. All these models had a common property: if the component sizes in the barely subcritical regime are viewed as masses then (i) these masses merge as approximate multiplicative coalescents in the critical window, and (ii) each individual mass is negligible/"dust" compared to the sum of squares of the masses in the barely subcritical regime. Indeed, (ii) is observed in [8, (10)], [15, (4)]. In the case of [33] and this chapter, the barely-subcritical component sizes do not become negligible due to the existence of the high-degree vertices (see [33, Theorem 1.3]). As discussed in [11, Section 1.4], these *large* barely-subcritical clusters can be thought of as nuclei, not interacting with each other and "sweeping up the smaller clusters in such a way that the relative masses converge". It will be fascinating to find a class of random graphs, used to model real-life networks, that has both the nuclei and a good amount of dust in the barely-subcritical regime, so that the

scaling limits predicted by [11] can be observed in complete generality.

**Component sizes and the width of the critical window**. We have already discussed how the width of the scaling window and the order of the maximal degrees should lead the asymptotic size of the components to be of the order $b_n$. For the finite third-moment case, the size of the largest component is of the order $n^{2/3} \gg b_n$. We do not have a very intuitive explanation for the reduced sizes of the components except for the fact that a similar property is true for the survival probability of a slightly supercritical branching process. The width of the critical window decreases by a factor of $L(n)^{-2}$ as compared to [33] if the size of the high-degree vertices increases by a factor of $L(n)$ (see (3.4)). Indeed, an increase in the degrees of the high-degree vertices is expected to start the merging of the barely-subcritical nuclei earlier, resulting in an increase in the width of the critical window. The fact that the width decreases by a factor of $L(n)^{-2}$ comes out of our calculations.

**Overview of the proofs**. The proofs of Theorems 3.2.1 and 3.2.2 consist of two important steps. First, we define an exploration algorithm on the graph that explores one edge of the graph at each step. The algorithm produces a walk, termed exploration process, that encodes the information about the number of edges in the explored components in terms of the hitting times to its past minima. In Section 3.5, the exploration process, suitably rescaled, is shown to converge. The surplus edges in the components are asymptotically negligible compared to the component sizes; these two facts together give us the finite-dimensional scaling limit of the re-scaled component sizes. The proof of Theorem 3.2.1 follows from the asymptotics of the susceptibility function in Section 3.5.1. The joint convergence of the component sizes and surplus edges is proved by verifying a uniform tightness condition on the surplus edges in Section 3.6. Then, in Section 3.7, we exploit the idea that the large components are explored before any self-loops or multiple edges are created and conclude the proof of Theorem 3.2.3. The proof of Theorem 3.2.4 is completed by showing that the percolated degree sequence is again a configuration model satisfying Assumption 3.1. Section 3.10 is devoted to the proof of Theorem 3.2.5 which exploits different properties of the augmented multiplicative coalescent process.

## 3.5    Convergence of the exploration process

We start by describing how the connected components in the graph can be explored while generating the random graph simultaneously:

**Algorithm 3.1** (Exploring the graph)**.** The algorithm carries along vertices that can be alive, active, exploring and killed and half-edges that can be alive, active or killed. We sequentially explore the graph as follows:

(S0)  At stage $i = 0$, all the vertices and the half-edges are *alive* but none of them are *active*. Also, there are no *exploring* vertices.

(S1)  At each stage $i$, if there is no active half-edge at stage $i$, choose a vertex $\nu$ proportional to its degree among the alive (not yet killed) vertices and declare all its half-edges to be *active* and declare $\nu$ to be *exploring*. If there is an active vertex but no exploring vertex, then declare the *smallest* vertex to be exploring.

(S2)  At each stage $i$, take an active half-edge $e$ of an exploring vertex $\nu$ and pair it uniformly to another alive half-edge $f$. Kill $e, f$. If $f$ is incident to a vertex $\nu'$ that has not been discovered before, then declare all the half-edges incident to $\nu'$ active, except $f$ (if any). If degree$(\nu') = 1$ (i.e. the only half-edge incident to $\nu'$ is $f$) then kill $\nu'$. Otherwise, declare $\nu'$ to be active and larger than all other vertices that are alive. After killing $e$, if $\nu$ does not have another active half-edge, then kill $\nu$ also.

(S3)  Repeat from (S1) at stage $i + 1$ if not all half-edges are already killed.

Algorithm 3.1 gives a *breadth-first* exploration of the connected components of $CM_n(\mathbf{d})$. Define the exploration process by

$$S_n(0) = 0, \quad S_n(l) = S_n(l-1) + d_{(l)}J_l - 2, \tag{3.21}$$

where $J_l$ is the indicator that a new vertex is discovered at time $l$ and $d_{(l)}$ is the degree of the new vertex chosen at time $l$ when $J_l = 1$. Suppose $\mathscr{C}_k$ is the $k^{\text{th}}$ connected component explored by the above exploration process and define $\tau_k = \inf\{i : S_n(i) = -2k\}$. Then $\mathscr{C}_k$ is discovered between the times $\tau_{k-1} + 1$ and $\tau_k$, and $\tau_k - \tau_{k-1} - 1$ gives the total number of edges in $\mathscr{C}_k$. Call a vertex *discovered* if it is either active or killed. Let $\mathcal{V}_l$ denote the set of vertices discovered up to time $l$ and $\mathcal{I}_i^n(l) := \mathbf{1}\{i \in \mathcal{V}_l\}$. Note that

$$S_n(l) = \sum_{i \in [n]} d_i \mathcal{I}_i^n(l) - 2l = \sum_{i \in [n]} d_i \left( \mathcal{I}_i^n(l) - \frac{d_i}{\ell_n} l \right) + (\nu_n(\lambda) - 1) l.$$

Recall the notation in (3.4). Define the re-scaled version $\bar{\mathbf{S}}_n$ of $\mathbf{S}_n$ by $\bar{S}_n(t) = a_n^{-1} S_n(\lfloor b_n t \rfloor)$. Then, by Assumption 3.1 (iii),

$$\bar{S}_n(t) = a_n^{-1} \sum_{i \in [n]} d_i \left( \mathcal{I}_i^n(t b_n) - \frac{d_i}{\ell_n} t b_n \right) + \lambda t + o(1). \tag{3.22}$$

Note the similarity between the expressions in (3.1) and (3.22). We will prove the following:

**Theorem 3.5.1.** *Consider the process $\bar{\mathbf{S}}_n := (\bar{S}_n(t))_{t \geqslant 0}$ defined in (3.22) and recall the definition of $\bar{\mathbf{S}}_\infty := (\bar{S}_\infty(t))_{t \geqslant 0}$ from (3.1). Then,*

$$\bar{\mathbf{S}}_n \xrightarrow{d} \bar{\mathbf{S}}_\infty$$

*with respect to the Skorohod $J_1$ topology.*

The proof of Theorem 3.5.1 is completed by showing that the summation term in (3.22) is predominantly carried by the first *few* terms and the limit of the first few terms gives rise to the limiting process given in (3.1). Fix $K \geqslant 1$ to be large. Denote by $\mathscr{F}_l$ the sigma-field containing the information generated up to time $l$ by Algorithm 3.1. Also, let $\Upsilon_l$ denote the set of time points up to time $l$ when a component was discovered and $\upsilon_l = |\Upsilon_l|$. Note that we have lost $2(l - \upsilon_l)$ half-edges by time $l$. Thus, on the set $\{\mathcal{I}_i^n(l) = 0\}$,

$$\mathbb{P}\left( \mathcal{I}_i^n(l+1) = 1 \big| \mathscr{F}_l \right) = \begin{cases} \frac{d_i}{\ell_n - 2(l - \upsilon_l) - 1} & \text{if } l \notin \Upsilon_l, \\ \frac{d_i}{\ell_n - 2(l - \upsilon_l)} & \text{otherwise} \end{cases} \tag{3.23}$$

and, uniformly over $l \leqslant T b_n$,

$$\mathbb{P}\left( \mathcal{I}_i^n(l+1) = 1 \big| \mathscr{F}_l \right) \geqslant \frac{d_i}{\ell_n} \quad \text{on the set } \{\mathcal{I}_i^n(l) = 0\}. \tag{3.24}$$

Denote $M_n^K(l) = a_n^{-1} \sum_{i > K} d_i \left( \mathcal{I}_i^n(l) - \frac{d_i}{\ell_n} l \right)$. Then,

$$\mathbb{E}\left[ M_n^K(l+1) - M_n^K(l) \big| \mathscr{F}_l \right] = \mathbb{E}\left[ \sum_{i=K+1}^n a_n^{-1} d_i \left( \mathcal{I}_i^n(l+1) - \mathcal{I}_i^n(l) - \frac{d_i}{\ell_n} \right) \bigg| \mathscr{F}_l \right]$$

$$= \sum_{i=K+1}^n a_n^{-1} d_i \left( \mathbb{E}\left[ \mathcal{I}_i^n(l+1) \big| \mathscr{F}_l \right] \mathbf{1}\{\mathcal{I}_i^n(l) = 0\} - \frac{d_i}{\ell_n} \right) \geqslant 0.$$

Thus $(M_n^K(l))_{l=1}^{T b_n}$ is a sub-martingale. Further, (3.23) implies that, uniformly for all $l \leqslant T b_n$,

$$\mathbb{P}\left( \mathcal{I}_i^n(l) = 0 \right) \geqslant \left( 1 - \frac{d_i}{\ell_n'} \right)^l, \tag{3.25}$$

where $\ell_n' = \ell_n - 2Tb_n - 1$. Thus, Assumption 3.1 (ii) gives

$$\left| \mathbb{E}[M_n^K(\mathfrak{l})] \right| = a_n^{-1} \sum_{i=K+1}^{n} d_i \left( \mathbb{P}\left(\mathcal{I}_i^n(\mathfrak{l}) = 1\right) - \frac{d_i}{\ell_n}\mathfrak{l}\right)$$

$$\leqslant a_n^{-1} \sum_{i=K+1}^{n} d_i \left( 1 - \left(1 - \frac{d_i}{\ell_n'}\right)^{\mathfrak{l}} - \frac{d_i}{\ell_n'}\mathfrak{l}\right) + a_n^{-1}\mathfrak{l} \sum_{i\in[n]} d_i^2 \left(\frac{1}{\ell_n'} - \frac{1}{\ell_n}\right)$$

$$\leqslant \frac{\mathfrak{l}^2}{2\ell_n'^2 a_n} \sum_{i=K+1}^{n} d_i^3 + o(1)$$

$$\leqslant \frac{T^2 n^{2\rho} n^{3\alpha} L(n)^3}{2\ell_n'^2 L(n)^2 n^\alpha L(n)} \left( a_n^{-3} \sum_{i=K+1}^{n} d_i^3 \right) + o(1)$$

$$= C\left( a_n^{-3} \sum_{i=K+1}^{n} d_i^3 \right) + o(1),$$

for some constant $C > 0$, where we have used the fact that

$$a_n^{-1}\mathfrak{l} \sum_{i\in[n]} d_i^2 \left(\frac{1}{\ell_n'} - \frac{1}{\ell_n}\right) = O(n^{2\rho+1-\alpha-2}/L(n)^3)$$

$$= O(n^{(\tau-4)/(\tau-1)}/L(n)^3) = o(1),$$

uniformly for $\mathfrak{l} \leqslant Tb_n$. Therefore, uniformly over $\mathfrak{l} \leqslant Tb_n$,

$$\lim_{K\to\infty} \limsup_{n\to\infty} \left|\mathbb{E}[M_n^K(\mathfrak{l})]\right| = 0. \qquad (3.26)$$

Now, note that for any $(x_1, x_2, \dots)$, $0 \leqslant a + b \leqslant x_i$ and $a, b > 0$ one has $\prod_{i=1}^{R}(1 - a/x_i)(1 - b/x_i) \geqslant \prod_{i=1}^{R}(1 - (a+b)/x_i)$. Thus, by (3.23), for all $\mathfrak{l} \geqslant 1$ and $i \neq j$,

$$\mathbb{P}\left(\mathcal{I}_i^n(\mathfrak{l}) = 0, \mathcal{I}_j^n(\mathfrak{l}) = 0\right) \leqslant \mathbb{P}\left(\mathcal{I}_i^n(\mathfrak{l}) = 0\right)\mathbb{P}\left(\mathcal{I}_j^n(\mathfrak{l}) = 0\right) \qquad (3.27)$$

and therefore $\mathcal{I}_i^n(\mathfrak{l})$ and $\mathcal{I}_j^n(\mathfrak{l})$ are negatively correlated. Observe also that, uniformly over $\mathfrak{l} \leqslant Tb_n$,

$$\text{Var}\left(\mathcal{I}_i^n(\mathfrak{l})\right) \leqslant \mathbb{P}\left(\mathcal{I}_i^n(\mathfrak{l}) = 1\right)$$

$$\leqslant \sum_{\mathfrak{l}_1=1}^{\mathfrak{l}} \mathbb{P}\left(\text{vertex } i \text{ is first discovered at stage } \mathfrak{l}_1\right) \leqslant \frac{\mathfrak{l}d_i}{\ell_n'}. \qquad (3.28)$$

Therefore, using the negative correlation in (3.27), uniformly over $\mathfrak{l} \leqslant Tb_n$,

$$\text{Var}\left(M_n^K(\mathfrak{l})\right) \leqslant a_n^{-2} \sum_{i=K+1}^{n} d_i^2 \text{Var}\left(\mathcal{I}_i^n(\mathfrak{l})\right) \leqslant \frac{\mathfrak{l}}{\ell_n' a_n^2} \sum_{i=K+1}^{n} d_i^3 \leqslant C a_n^{-3} \sum_{i=K+1}^{n} d_i^3,$$

$$(3.29)$$

for some constant $C > 0$ and by using Assumption 3.1 (ii) again,

$$\lim_{K \to \infty} \limsup_{n \to \infty} \mathrm{Var}\left(M_n^K(l)\right) = 0,$$

uniformly for $l \leqslant Tb_n$. Now we can use the super-martingale inequality [147, Lemma 2.54.5] stating that for any super-martingale $(M(t))_{t \geqslant 0}$, satisfying $M(0) = 0$,

$$\varepsilon \mathbb{P}\left(\sup_{s \leqslant t} |M(s)| > 3\varepsilon\right) \leqslant 3\mathbb{E}\left[|M(t)|\right] \leqslant 3\left(|\mathbb{E}\left[M(t)\right]| + \sqrt{\mathrm{Var}\left(M(t)\right)}\right). \quad (3.30)$$

Using (3.26), (3.29), and (3.30), together with the fact that $(-M_n^K(l))_{l=1}^{Tb_n}$ is a super-martingale, we get, for any $\varepsilon > 0$,

$$\lim_{K \to \infty} \limsup_{n \to \infty} \mathbb{P}\left(\sup_{l \leqslant Tb_n} |M_n^K(l)| > \varepsilon\right) = 0. \quad (3.31)$$

Define the truncated exploration process

$$\bar{S}_n^K(t) = a_n^{-1} \sum_{i=1}^{K} d_i \left(\mathcal{I}_i^n(tb_n) - \frac{d_i}{\ell_n} tb_n\right) + \lambda t.$$

Define $\mathcal{I}_i^n(tb_n) = \mathcal{I}_i^n(\lfloor tb_n \rfloor)$ and recall that $\mathcal{I}_i(s) := \mathbf{1}\{\xi_i \leqslant s\}$ where $\xi_i \sim \mathrm{Exp}(\theta_i/\mu)$.

**Lemma 3.5.2.** *Fix any $K \geqslant 1$. As $n \to \infty$,*

$$(\mathcal{I}_i^n(tb_n))_{i \in [K], t \geqslant 0} \xrightarrow{d} (\mathcal{I}_i(t))_{i \in [K], t \geqslant 0}.$$

*Proof.* By noting that $(\mathcal{I}_i^n(tb_n))_{t \geqslant 0}$ are indicator processes, it is enough to show that

$$\mathbb{P}\left(\mathcal{I}_i^n(t_i b_n) = 0, \forall i \in [K]\right) \to \mathbb{P}\left(\mathcal{I}_i(t_i) = 0, \forall i \in [K]\right) = \exp\left(-\mu^{-1} \sum_{i=1}^{K} \theta_i t_i\right).$$

for any $t_1, \ldots, t_K \in \mathbb{R}$. Now,

$$\mathbb{P}\left(\mathcal{I}_i^n(m_i) = 0, \forall i \in [K]\right) = \prod_{l=1}^{\infty} \left(1 - \sum_{i \leqslant K : l \leqslant m_i} \frac{d_i}{\ell_n - \Theta(l)}\right), \quad (3.32)$$

where the $\Theta(l)$ term arises from the expression in (3.23) and noting that $\upsilon_l \leqslant l$. Taking logarithms on both sides of (3.32) and using the fact that

$l \leqslant \max m_i = \Theta(b_n)$ we get

$$\mathbb{P}\left(\mathcal{I}_i^n(m_i) = 0 \, \forall i \in [K]\right) = \exp\left(-\sum_{l=1}^{\infty} \sum_{i \leqslant K : l \leqslant m_i} \frac{d_i}{\ell_n} + o(1)\right)$$

$$= \exp\left(-\sum_{i \in [K]} \frac{d_i m_i}{\ell_n} + o(1)\right). \tag{3.33}$$

Putting $m_i = t_i b_n$, Assumption 3.1 (i), (ii) gives

$$\frac{m_i d_i}{\ell_n} = \frac{\theta_i t_i}{\mu}(1 + o(1)). \tag{3.34}$$

Hence (3.33) and (3.34) complete the proof of Lemma 3.5.2. $\qquad\square$

*Proof of Theorem 3.5.1.* The proof of Theorem 3.5.1 now follows from (3.22), (3.31) and Lemma 3.5.2 by first taking the limit as $n \to \infty$ and then taking the limit as $K \to \infty$. $\qquad\square$

**Theorem 3.5.3.** *Recall the definition of* $\mathrm{refl}(\bar{\mathbf{S}}_\infty)$ *from* (3.2). *As* $n \to \infty$,

$$\mathrm{refl}(\bar{\mathbf{S}}_n) \xrightarrow{d} \mathrm{refl}(\bar{\mathbf{S}}_\infty).$$

*Proof.* This follows from Theorem 3.5.1 and the fact that the reflection is Lipschitz continuous with respect to the Skorohod $J_1$ topology (see [150, Theorem 13.5.1]). $\qquad\square$

This also implies that Algorithm 3.1 explores the *large* components before time $Tb_n$ for large $T$. Next, we show that the function mapping an element of $\mathbb{D}[0, \infty)$ to its largest excursions, is continuous on a special subset $A$ of $\mathbb{D}[0, \infty)$ and the process $\mathrm{refl}(\bar{\mathbf{S}}_\infty)$ has sample paths in $A$ almost surely. Therefore, Theorem 3.5.1 gives the scaling limit of the number of edges in the components ordered as a non-increasing sequence. Finally, we show that the number of surplus edges discovered up to time $Tb_n$ are negligible and thus the convergence of the component sizes in Theorem 3.2.1 follows.

### 3.5.1  Tightness of the component sizes

The following proposition establishes a uniform tail summability condition that is required for the tightness of the (scaled) ordered vector of component sizes with respect to the $\ell_\downarrow^2$ topology:

**Proposition 3.5.4.** *For any* $\varepsilon > 0$,

$$\lim_{K \to \infty} \limsup_{n \to \infty} \mathbb{P}\left(\sum_{i > K} |\mathscr{C}_{(i)}|^2 > \varepsilon b_n^2\right) = 0.$$

Roughly speaking, the proof is based on the fact that the graph, obtained by removing a large number of high-degree vertices, yields a graph that approaches subcriticality. More precisely, we prove Lemma 3.5.5 below to complete the proof of Proposition 3.5.4. This fact is not true for the finite third-moment setting [70]. However, since the large-degree vertices guide the scaling behavior in the infinite third-moment case, the observation in Lemma 3.5.5 saves some computational complexity, and gives a different proof of the $\ell_\downarrow^2$ tightness than the arguments with size-biased point processes originally described in [8].

**Lemma 3.5.5.** *Consider* $\mathrm{CM}_n(\mathbf{d})$ *satisfying Assumption 3.1. Let* $\mathcal{G}^{[K]}$ *be the random graph obtained by removing all edges attached to vertices* $1, \ldots, K$ *and let* $\mathbf{d}'$ *be the obtained degree sequence. Suppose* $V_n$ *is a random vertex of* $\mathcal{G}^{[K]}$ *chosen independently of the graph and let* $\mathscr{C}^{[K]}(V_n)$ *be the corresponding component. Let* $\{\mathscr{C}_{(i)}^{[K]} : i \geqslant 1\}$ *be the components of* $\mathcal{G}^{[K]}$, *ordered according to their sizes. Then,*

$$\lim_{K \to \infty} \limsup_{n \to \infty} c_n^{-1} \mathbb{E}\left[|\mathscr{C}^{[K]}(V_n)|\right] = 0. \tag{3.35}$$

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

$$\lim_{K \to \infty} \limsup_{n \to \infty} \mathbb{P}\left(\sum_{i \geqslant 1} |\mathscr{C}_{(i)}^{[K]}|^2 > \varepsilon b_n^2\right) = 0. \tag{3.36}$$

*Proof.* We make use of a result due to Janson [111] regarding bounds on the susceptibility functions for the configuration model. In fact, [111, Lemma 5.2] shows that, for any configuration model $\mathrm{CM}_n(\mathbf{d})$ with $\nu_n < 1$,

$$\mathbb{E}\left[|\mathscr{C}(V_n)|\right] \leqslant 1 + \frac{\mathbb{E}\left[D_n\right]}{1 - \nu_n}. \tag{3.37}$$

Now, conditional on the set of removed half-edges, $\mathcal{G}^{[K]}$ is still a configuration model with some degree sequence $\mathbf{d}'$ with $d_i' \leqslant d_i$ for all $i \in [n] \setminus [K]$ and $d_i' = 0$ for $i \in [K]$. Further, the criticality parameter of $\mathcal{G}^{[K]}$ satisfies

$$\nu_n^{[K]} = \frac{\sum_{i \in [n]} d_i'(d_i' - 1)}{\sum_{i \in [n]} d_i'} \leqslant \frac{\sum_{i \in [n]} d_i(d_i - 1) - \sum_{i=1}^K d_i(d_i - 1)}{\ell_n - 2\sum_{i=1}^K d_i}$$

$$= \nu_n - C_1 n^{2\alpha - 1} L(n)^2 \sum_{i \leqslant K} \theta_i^2 = \nu_n - C_1 c_n^{-1} \sum_{i \leqslant K} \theta_i^2$$

for some constant $C_1 > 0$. Since $\theta \notin \ell_\downarrow^2$, $K$ can be chosen large enough such that $\nu_n^{[K]} < 1$ uniformly for all $n$. Also $\sum_{i \in [n]} d_i' = \ell_n + o(n)$ for each fixed

K. Let $\mathbb{E}_K[\cdot]$ denote the conditional expectation, conditioned on the set of removed half-edges. Using (3.37) on $\mathcal{G}^{[K]}$, we get

$$\mathbb{E}_K\left[|\mathscr{C}^{[K]}(V_n)|\right] \leqslant \frac{C_2}{1 - \nu_n^{[K]}} \leqslant \frac{C_2}{1 - \nu_n + C_1 c_n^{-1} \sum_{i \leqslant K} \theta_i^2} \leqslant \frac{C_2 c_n}{-\lambda + C_1 \sum_{i \leqslant K} \theta_i^2},$$

for some constant $C_2 > 0$. Using the fact that $\theta \notin \ell_{\downarrow}^2$, this concludes the proof of (3.35). The proof of (3.36) follows from (3.35) by using the Markov inequality and the observation that

$$\mathbb{E}\left[\sum_{i \geqslant 1} |\mathscr{C}_{(i)}^{[K]}|^2\right] = n\mathbb{E}\left[|\mathscr{C}^{[K]}(V_n)|\right]. \tag{3.38}$$

$\square$

*Proof of Proposition 3.5.4.* Denote the sum of squares of the component sizes excluding the components containing vertices $1, 2, \ldots, K$ by $\mathscr{S}_K$. Note that

$$\sum_{i > K} |\mathscr{C}_{(i)}|^2 \leqslant \mathscr{S}_K \leqslant \sum_{i \geqslant 1} |\mathscr{C}_{(i)}^{[K]}|^2.$$

Thus, Proposition 3.5.4 follows from Lemma 3.5.5. $\square$

### 3.5.2  Large components are explored early

An important consequence of Proposition 3.5.4 is that after time $\Theta(b_n)$, Algorithm 3.1 does not explore large components. The precise statement needed to complete our proof is given below. This is an essential ingredient to conclude the convergence of the component sizes from the convergence of the exploration process since Theorem 3.5.1 only gives information about the components explored on the time scale of the order $b_n$.

**Lemma 3.5.6.** *Let $\mathscr{C}_{\max}^{\geqslant T}$ be the largest among those components which are started exploring after time $Tb_n$ by Algorithm 3.1. Then, for any $\varepsilon > 0$,*

$$\lim_{T \to \infty} \limsup_{n \to \infty} \mathbb{P}\left(|\mathscr{C}_{\max}^{\geqslant T}| > \varepsilon b_n\right) = 0.$$

*Proof.* Define the event

$$\mathscr{A}_{K,T}^n := \{\text{all the vertices of } [K] \text{ are explored before time } Tb_n\}.$$

Recall the definition of $\mathscr{C}_{(i)}^{[K]}$ from Lemma 3.5.5. Firstly, note that

$$\mathbb{P}\left(|\mathscr{C}_{\max}^{\geqslant T}| > \varepsilon b_n, \mathscr{A}_{K,T}^n\right) \leqslant \mathbb{P}\left(\sum_{i \geqslant 1} |\mathscr{C}_{(i)}^{[K]}|^2 > \varepsilon^2 b_n^2\right). \tag{3.39}$$

Moreover, using (3.23) and the fact that $d_j b_n = \Theta(n)$, we get

$$
\mathbb{P}\left((\mathscr{A}_{K,T}^n)^c\right) = \mathbb{P}\left(\exists j \in [K] : j \text{ is not explored before } Tb_n\right)
$$

$$
\leqslant \sum_{j=1}^{K} \mathbb{P}\left(j \text{ is not explored before } Tb_n\right) \tag{3.40}
$$

$$
\leqslant \sum_{j=1}^{K} \left(1 - \frac{d_j}{\ell_n - \Theta(Tb_n)}\right)^{Tb_n} \leqslant \sum_{j=1}^{K} e^{-CT},
$$

where $C > 0$ is a constant that may depend on $K$. Now, by (3.39),

$$
\mathbb{P}\left(|\mathscr{C}_{\max}^{\geqslant T}| > \varepsilon b_n\right) \leqslant \mathbb{P}\left(\sum_{i \geqslant 1} |\mathscr{C}_{(i)}^{[K]}|^2 > \varepsilon^2 b_n^2\right) + \mathbb{P}\left((\mathscr{A}_{K,T}^n)^c\right).
$$

The proof follows by taking $\limsup_{n\to\infty}$, $\lim_{T\to\infty}$, $\lim_{K\to\infty}$ respectively and using (3.36), (3.40). $\qquad\square$

Define the set of excursions of a function $f$ by

$$
\mathcal{E} := \{(l, r) : (l, r) \text{ is an excursion of } f\}.
$$

We also denote the set of excursion end-points by $\mathcal{Y}$, i.e.,

$$
\mathcal{Y} := \{r > 0 : (l, r) \in \mathcal{E}\}.
$$

**Definition 1.** A function $f \in \mathbb{D}_+[0, T]$ is said to be *good* if the following holds:

(a) $\mathcal{Y}$ does not have an isolated point and the complement of $\cup_{(l,r)\in\mathcal{E}}(l, r)$ has Lebesgue measure zero;

(b) $f$ does not attain a local minimum at any point of $\mathcal{Y}$.

**Remark 25.** We claim that if a function $f \in \mathbb{D}_+[0, T]$ is good, then $f$ is continuous on $\mathcal{Y}$. To see this, fix any $\delta > 0$ and denote the set of excursions of length at least $\delta$ by $\mathcal{E}_\delta$. Let $r$ be the excursion endpoint of an excursion in $\mathcal{E}_\delta$ and suppose that $f(r) > f(r-)$. Thus, there is no excursion endpoint in $(r - \delta, r)$. Moreover, since $f$ is right-continuous, there exists $\delta' > 0$ such that $f(x) > f(r-) + \varepsilon$ for all $x \in (r, r + \delta')$, where $\varepsilon = (f(r) - f(r-))/2 > 0$. Thus there is no excursion endpoint on $(r - \delta, r + \delta')$ and thus $r$ is an isolated point contradicting Definition 1. We conclude that $f$ is continuous at excursion endpoints of the excursions in $\mathcal{E}_\delta$, and since $\delta > 0$ is arbitrary the claim is established.

Let $\mathcal{L}_i(f)$ be the length of the $i^{\text{th}}$ largest excursion of $f$ and define $\Phi_m :$ $\mathbb{D}_+[0, T] \to \mathbb{R}^m$ by

$$\Phi_m(f) = (\mathcal{L}_1(f), \mathcal{L}_2(f), \ldots, \mathcal{L}_m(f)).$$

Note that $\Phi_m(\cdot)$ is well-defined for any good function defined in Definition 2.

**Lemma 3.5.7.** *Suppose that $f \in \mathbb{D}_+[0, T]$ is good. Then, $\Phi_m$ is continuous at $f$ with respect to the subspace topology on $\mathbb{D}_+[0, T]$ induced by the Skorohod $J_1$ topology.*

*Proof.* We extend the arguments of [135, Proposition 22]. The proof here is for $m = 1$ and similar arguments hold for $m > 1$. Let $\mathfrak{L}$ denote the set of continuous functions $\Lambda : \mathbb{R}_+ \to \mathbb{R}_+$ that are strictly increasing and $\Lambda(0) = 0, \Lambda(T) = T$. Suppose $E_1 = (l, r)$ is the longest excursion of $f$ on $[0, T]$, thus $\Phi_1(f) = r - l$. For any $\varepsilon > 0$ (small), choose $\delta > 0$ such that

$$f(x) > \min\{f(r-), f(r)\} + \delta \quad \forall x \in (l + \varepsilon, r - \varepsilon). \tag{3.41}$$

Let $\|\cdot\|$ denote the sup-norm on $[0, T]$. Take any sequence of functions $f_n \in \mathbb{D}_+[0, T]$ such that $f_n \to f$, i.e., there exists $\{\Lambda_n\}_{n \geqslant 1} \subset \mathfrak{L}$ such that for all large enough $n$,

$$\|f_n \circ \Lambda_n - f\| < \frac{\delta}{6} \text{ and } \|\Lambda_n - I\| < \varepsilon, \tag{3.42}$$

where $I$ is the identity function. Now, by Remark 25, $f$ is continuous at $r$. This implies that $f(r-) = f(r)$, and using (3.41) and (3.42), for all large enough $n$,

$$f_n(y) > f_n \circ \Lambda_n(r) + \frac{2\delta}{3} \quad \forall y \in (l + 2\varepsilon, r - 2\varepsilon). \tag{3.43}$$

Further, using the continuity of $f$ at $r$, $f_n(r) \to f(r)$ and thus, for all sufficiently large $n$,

$$|f_n \circ \Lambda_n(r) - f_n(r)| \leqslant |f_n \circ \Lambda_n(r) - f(r)| + |f_n(r) - f(r)| < \frac{\delta}{3}.$$

Hence, (3.43) implies that, for all sufficiently large $n$,

$$f_n(y) > f_n(r) + \frac{\delta}{3} \quad \forall y \in (l + 2\varepsilon, r - 2\varepsilon).$$

Thus, for any $\varepsilon > 0$, we have

$$\liminf_{n \to \infty} \Phi_1(f_n) \geqslant r - l - 4\varepsilon = \Phi_1(f) - 4\varepsilon. \tag{3.44}$$

Now we turn to a suitable upper bound on $\limsup_{n\to\infty} \Phi_1(f_n)$. First, we claim that one can find $r_1, \ldots, r_k \in \mathcal{Y}$ such that $r_1 \leqslant \Phi_1(f) + \varepsilon, T - r_k < \Phi_1(f) + \varepsilon$, and $r_i - r_{i-1} \leqslant \Phi_1(f) + \varepsilon, \forall i = 2, \ldots, k$. The claim is a consequence of Definition 1 (a). Now, Definition 1 (b) implies that for any small $\varepsilon > 0$, there exists $\delta > 0$ and $x_i \in (r_i, r_i + \varepsilon)$ such that $f(r_i) - f(x_i) > \delta \ \forall i$. Again, since $r_i$ is a continuity point of $f$, $f_n(r_i) \to f(r_i)$. Thus, using (3.42), for all large enough $n$,

$$f_n(r_i) - f_n(\Lambda_n(x_i)) > \frac{\delta}{2}.$$

Now, $\Lambda_n(x_i) \in (r_i, r_i + \varepsilon)$ for all sufficiently large $n$, since $x_i \in (r_i, r_i + \varepsilon)$. Thus, for all large enough $n$, there exists a point $z_i^n \in (r_i, r_i + \varepsilon)$ such that

$$f_n(r_i) - f_n(z_i^n) > \frac{\delta}{2}.$$

Also the function $f_n$ only has positive jumps and $\underline{f_n}(r_i) \to \underline{f}(r_i)$, as $\underline{f_n}$ is continuous, where we recall that $\underline{f}(x) = \inf_{y \leqslant x} f(y)$. Therefore, $f_n$ must have an excursion end point on $(r_i, r_i + \varepsilon)$ for all large enough $n$. Also, using the fact that the complement of $\cup_{(l,r) \in \mathcal{E}}(l, r)$ has Lebesgue measure zero, $f$ has an excursion endpoint $r_i^0 \in (l_i - \varepsilon, l_i)$. The previous argument shows that $f_n$ has to have an excursion endpoint in $(r_i^0, r_i^0 + \varepsilon)$ and thus in $(l_i - \varepsilon, l_i + \varepsilon)$, for all large $n$. Therefore, for any $\varepsilon > 0$,

$$\limsup_{n\to\infty} \Phi_1(f_n) \leqslant \Phi_1(f) + 3\varepsilon. \tag{3.45}$$

Hence the proof follows from (3.44) and (3.45).   □

**Remark 26.** For $f \in \mathbb{D}_+[0, T]$, let $\mathcal{A}_i(f)$ denote the area under the excursion $\mathcal{L}_i(f)$. Let $(f_n)_{n\geqslant 1}$ be a sequence of functions on $f \in \mathbb{D}_+[0, \infty)$ such that $f_n \to f$, with respect to the Skorohod $J_1$ topology, where $f$ is good. Then, (3.42), (3.44) and (3.45) also implies that $(\mathcal{A}_1(f_n), \ldots, \mathcal{A}_m(f_n))$ converges to $(\mathcal{A}_1(f), \ldots, \mathcal{A}_m(f))$, for any $m \geqslant 1$.

**Definition 2.** A stochastic process $\mathbf{X} \in \mathbb{D}_+[0, \infty)$ is said to be good if

(a) The sample paths are good almost surely when restricted to $[0, T]$, for every fixed $T > 0$;

(b) $\mathbf{X}$ does not have an infinite excursion almost surely;

(c) For any $\varepsilon > 0$, $\mathbf{X}$ has only finitely many excursions of length more than $\varepsilon$ almost surely.

**Lemma 3.5.8.** *The thinned Lévy process* $\mathbf{S}^\lambda_\infty$ *defined in* (3.1) *is good.*

*Proof.* Let us make use of the properties of the process $\mathbf{S}^\lambda_\infty$ that were established in [11]. $\mathbf{S}^\lambda_\infty$ satisfies Definition 2 (b),(c) by [11, (8)]. The fact that the excursion endpoints of $\mathbf{S}^\lambda_\infty$ do not have any isolated points almost surely follows directly from [11, Proposition 14 (d)]. Further, [11, Proposition 14 (b)] implies that, for any $u > 0$, $\mathbb{P}\left(S^\lambda_\infty(u) = \inf_{u' \leqslant u} S^\lambda_\infty(u')\right) = 0$. Taking the integral with respect to the Lebesgue measure and interchanging the limit by using Fubini's theorem, we conclude that almost surely

$$\int_0^T \mathbf{1}\left\{ S^\lambda_\infty(u) = \inf_{u' \leqslant u} S^\lambda_\infty(u') \right\} du = 0,$$

which verifies Definition 1 (a). Now, let $\mathbf{L}$ be the Lévy process defined as

$$L(t) = \sum_{i=1}^\infty \theta_i \left(\mathcal{N}_i(t) - (\theta_i/\mu)t\right) + \lambda t,$$

where $(\mathcal{N}_i(t))_{t \geqslant 0}$ is a Poisson process with rate $\theta_i$ which are independent for different $i$. Via the natural coupling that states $\mathcal{I}_i(t) \leqslant \mathcal{N}_i(t)$, we can assume that $S^\lambda_\infty(t) \leqslant L(t)$ for all $t > 0$. Using [28, Theorem VII.1],

$$\inf\{t > 0 : L(t) < 0\} = 0, \quad \text{almost surely.} \tag{3.46}$$

Moreover, for any stopping time $\mathcal{T} > 0$, $(S^\lambda_\infty(\mathcal{T}+t) - S^\lambda_\infty(\mathcal{T}))_{t \geqslant 0}$, conditioned on the sigma-field $\sigma(S^\lambda_\infty(s) : s \leqslant \mathcal{T})$, is distributed as a process defined in (3.1) for some random $\theta$ and $\Lambda$. Now we can take $\mathcal{T}$ to be an excursion endpoint and an application of (3.46) verifies Definition 1 (b).                          $\square$

As described in Section 3.5, the excursion lengths of the exploration process $\bar{\mathbf{S}}_n$ gives the total number of edges in the explored components. Lemma 3.5.9 below estimates the number of surplus edges in the components explored upto time $\Theta(b_n)$. This enables us to compute the scaling limits for the component sizes using the results from the previous section and complete the proof of Theorem 3.2.1.

**Lemma 3.5.9.** *Let* $N^\lambda_n(k)$ *be the number of surplus edges discovered up to time* $k$ *and* $\bar{N}^\lambda_n(u) = N^\lambda_n(\lfloor ub_n \rfloor)$. *Then, as* $n \to \infty$,

$$(\bar{\mathbf{S}}_n, \bar{\mathbf{N}}^\lambda_n) \xrightarrow{d} (\mathbf{S}^\lambda_\infty, \mathbf{N}^\lambda),$$

*where* $\mathbf{N}^\lambda$ *is defined in* (3.3).

*Proof.* We write $N_n^\lambda(l) = \sum_{i=2}^l \xi_i$, where $\xi_i = \mathbf{1}\{\mathcal{V}_i = \mathcal{V}_{i-1}\}$. Let $A_i$ denote the number of active half-edges after stage $i$ while implementing Algorithm 3.1. Note that

$$\mathbb{P}\left(\xi_i = 1|\mathscr{F}_{i-1}\right) = \frac{A_{i-1}-1}{\ell_n - 2i - 1} = \frac{A_{i-1}}{\ell_n}(1 + O(i/n)) + O(n^{-1}),$$

uniformly for $i \leqslant Tb_n$ for any $T > 0$. Therefore, the instantaneous rate of change of the re-scaled process $\bar{N}^\lambda$ at time $t$, conditional on the past, is

$$b_n \frac{A_{\lfloor tb_n \rfloor}}{n\mu}(1 + o(1)) + o(1) = \frac{1}{\mu}\mathrm{refl}(\bar{S}_n(t))(1 + o(1)) + o(1).$$

Recall from Theorem 3.5.3 that $\mathrm{refl}(\bar{\mathbf{S}}_n) \xrightarrow{d} \mathrm{refl}(\bar{\mathbf{S}}_\infty)$. Then, by the Skorohod representation theorem, we can assume that $\mathrm{refl}(\bar{\mathbf{S}}_n) \to \mathrm{refl}(\bar{\mathbf{S}}_\infty)$ almost surely on some probability space. Observe that $(\int_0^t \mathrm{refl}(\bar{S}_\infty(u))du)_{t \geqslant 0}$ has continuous sample paths. Therefore, the conditions of [128, Corollary 1, Page 388] are satisfied and the proof is complete. $\qquad\square$

**Theorem 3.5.10.** *For any $m \geqslant 1$, as $n \to \infty$*

$$b_n^{-1}\left(|\mathscr{C}_{(1)}|, |\mathscr{C}_{(2)}|, \ldots, |\mathscr{C}_{(m)}|\right) \xrightarrow{d} (\gamma_1(\lambda), \gamma_2(\lambda), \ldots, \gamma_m(\lambda))$$

*with respect to the product topology, where $\gamma_i(\lambda)$ is the $i^{th}$ largest excursion of $\bar{\mathbf{S}}_\infty$ defined in (3.1).*

*Proof.* Fix any $m \geqslant 1$. Let $\mathscr{C}_{(i)}^T$ be the $i^{th}$ largest component explored by Algorithm 3.1 up to time $Tb_n$. Denote by $\mathscr{D}_{(i)}^{\mathrm{ord},T}$ the $i^{th}$ largest value of $(\sum_{k \in \mathscr{C}_{(i)}^T} d_k)_{i \geqslant 1}$. Let $g : \mathbb{R}^m \mapsto \mathbb{R}$ be a bounded continuous function. By Lemma 3.5.8 the sample paths of $\bar{\mathbf{S}}_\infty$ are almost surely good. Thus, using Theorem 3.5.1, Lemma 3.5.7 gives

$$\lim_{n\to\infty} \mathbb{E}\left[g\left((2b_n)^{-1}\left(\mathscr{D}_{(1)}^{\mathrm{ord},T}, \mathscr{D}_{(2)}^{\mathrm{ord},T}, \ldots, \mathscr{D}_{(m)}^{\mathrm{ord},T}\right)\right)\right]$$
$$= \mathbb{E}\left[g\left(\gamma_1^T(\lambda), \gamma_2^T(\lambda), \ldots, \gamma_m^T(\lambda)\right)\right],$$

where $\gamma_i^T(\lambda)$ is the $i^{th}$ largest excursion of $\bar{\mathbf{S}}_\infty$ restricted to $[0, T]$. Now the support of the joint distribution of $(\gamma_i^T(\lambda))_{i \geqslant 1}$ is concentrated on $\{(x_1, x_2, \ldots) : x_1 > x_2 > \ldots\}$. Thus, using Lemma 3.5.9, it follows that

$$\lim_{n\to\infty} \mathbb{E}\left[g\left(b_n^{-1}\left(|\mathscr{C}_{(1)}^T|, |\mathscr{C}_{(2)}^T|, \ldots, |\mathscr{C}_{(m)}^T|\right)\right)\right] = \mathbb{E}\left[g\left(\gamma_1^T(\lambda), \gamma_2^T(\lambda), \ldots, \gamma_m^T(\lambda)\right)\right]. \tag{3.47}$$

Since $\mathbf{S}_\infty^\lambda$ satisfies Definition 2 (b), (c), it follows that

$$\lim_{T\to\infty} \mathbb{E}\left[g\left(\gamma_1^T(\lambda), \gamma_2^T(\lambda), \ldots, \gamma_m^T(\lambda)\right)\right] = \mathbb{E}\left[g\left(\gamma_1(\lambda), \gamma_2(\lambda), \ldots, \gamma_m(\lambda)\right)\right] \tag{3.48}$$

Finally, using Lemma 3.5.6, the proof of Theorem 3.5.10 is completed by (3.47) and (3.48).

<div align="right">□</div>

*Proof of Theorem 3.2.1.* The proof of Theorem 3.2.1 now follows directly from Theorem 3.5.10 and Proposition 3.5.4.    □

## 3.6    Proof of Theorem 3.2.2

The goal of this section is to prove the joint convergence of the component sizes and the surplus edges as described in Theorem 3.2.2. We start with a preparatory lemma:

**Lemma 3.6.1.** *The convergence in* (3.5) *holds with respect to the* $\ell_{\downarrow}^2 \times \mathbb{N}^\infty$ *topology.*

*Proof.* Note that Lemma 3.5.6 already states that we do not see large components being explored after the time $Tb_n$ for large $T > 0$. Thus the proof is a consequence of Lemmas 3.5.7, 3.5.9, Remark 26 and Theorem 3.2.1.    □

Recall the definition of the metric $d_U$ from Chapter 2.1. Using Lemma 3.6.1, it now remains to obtain a uniform summability condition on the tail of the sum of products of the scaled component sizes and surplus edges. This is formally stated in Proposition 3.6.2 below. The proof is completed in the similar spirit as the finite third-moment case [70].

**Proposition 3.6.2.** *For any* $\varepsilon > 0$,

$$\lim_{\delta \to 0} \limsup_{n \to \infty} \mathbb{P}\left( \sum_{i:|\mathscr{C}_{(i)}| \leqslant \delta b_n} |\mathscr{C}_{(i)}| \times \mathrm{SP}(\mathscr{C}_{(i)}) > \varepsilon b_n \right) = 0.$$

The following estimate will be the crucial ingredient to complete the proof of Proposition 3.6.2. The proof of Lemma 3.6.3 is postponed to Appendix 3.B since this uses similar ideas as [70].

**Lemma 3.6.3.** *Assume that* $\limsup_{n \to \infty} c_n(\nu_n - 1) < 0$. *Let* $V_n$ *denote a vertex chosen uniformly at random, independently of the graph* $\mathrm{CM}_n(\mathbf{d})$ *and let* $\mathscr{C}(V_n)$ *denote the component containing* $V_n$. *Let* $\delta_k = \delta k^{-0.12}$. *Then, for* $\delta > 0$ *sufficiently small,*

$$\mathbb{P}\left(\mathrm{SP}(\mathscr{C}(V_n)) \geqslant K, |\mathscr{C}(V_n)| \in (\delta_K b_n, 2\delta_K b_n)\right) \leqslant \frac{C\sqrt{\delta}}{a_n K^{1.1}}$$

*where* $C$ *is a fixed constant independent of* $n, \delta, K$.

<div align="center">136</div>

*Proof of Proposition 3.6.2 using Lemma 3.6.3.* First consider the case $\lambda < 0$. Fix any $\varepsilon, \eta > 0$. Note that

$$\mathbb{P}\left(\sum_{|\mathscr{C}_{(i)}| \leqslant \varepsilon b_n} |\mathscr{C}_{(i)}| \mathrm{SP}(\mathscr{C}_{(i)}) > \eta b_n\right) \leqslant \frac{1}{\eta b_n} \mathbb{E}\left[\sum_{i=1}^{\infty} |\mathscr{C}_{(i)}| \mathrm{SP}(\mathscr{C}_{(i)}) \mathbf{1}_{\{|\mathscr{C}_{(i)}| \leqslant \varepsilon b_n\}}\right]$$

$$= \frac{a_n}{\eta} \mathbb{E}\left[\mathrm{SP}(\mathscr{C}(V_n)) \mathbf{1}_{\{|\mathscr{C}(V_n)| \leqslant \varepsilon b_n\}}\right]$$

$$= \frac{a_n}{\eta} \sum_{k=1}^{\infty} \sum_{i \geqslant \log_2(1/(k^{0.12}\varepsilon))} \mathbb{P}\left(\mathrm{SP}(\mathscr{C}(V_n)) \geqslant k, |\mathscr{C}(V_n)| \in \left(\frac{b_n}{2^{i+1}k^{0.12}}, \frac{b_n}{2^i k^{0.12}}\right]\right)$$

$$\leqslant \frac{C}{\eta} \sum_{k=1}^{\infty} \frac{1}{k^{1.1}} \sum_{i \geqslant \log_2(1/(k^{0.12}\varepsilon))} 2^{-i/2} \leqslant \frac{C}{\eta} \sum_{k=1}^{\infty} \frac{\sqrt{\varepsilon}}{k^{1.04}} = O(\sqrt{\varepsilon}),$$

where the last-but-two step follows from Lemma 3.6.3. The proof of Proposition 3.6.2 now follows for $\lambda < 0$.

Now consider the case $\lambda > 0$. Fix a large integer $R \geqslant 1$ such that $\lambda - \sum_{i=1}^{R} \theta_i^2 < 0$. This can be done because $\theta \notin \ell_{\downarrow}^2$. Using (3.39), for any $\eta > 0$, it is possible to choose $T > 0$ such that for all sufficiently large $n$,

$$\mathbb{P}\left(\text{all the vertices } 1, \dots, R \text{ are explored within time } Tb_n\right) > 1 - \eta. \quad (3.49)$$

Let $T_e$ denote the first time after $Tb_n$ when we finish exploring a component. By Theorem 3.5.1, $(b_n^{-1} T_e)_{n \geqslant 1}$ is a tight sequence. Let $\mathscr{G}_T^*$ denote the graph obtained by removing the components explored up to time $T_e$. Then, $\mathscr{G}_T^*$ is again a configuration model conditioned on its degrees. Let $\nu_n^*$ denote the value of the criticality parameter for $\mathscr{G}^*$. Note that

$$\sum_{i \notin \mathscr{V}_{T_e}} d_i \geqslant \ell_n - 2Tb_n \implies \sum_{i \notin \mathscr{V}_{T_e}} d_i = \ell_n + o_{\mathbb{P}}(n),$$

and thus conditionally on $\mathscr{F}_{T_e}$ and the fact that $(1, \dots, R)$ are explored within time $Tb_n$,

$$\nu_n^* \leqslant \frac{\sum_{i \in [n]} d_i^2 - \sum_{i=1}^{R} d_i^2}{\sum_{i \notin \mathscr{V}_{T_e}} d_i} - 1 = 1 + c_n^{-1}\left(\lambda - \sum_{i=1}^{R} \theta_i^2\right) + o(c_n^{-1}). \quad (3.50)$$

Therefore, combining (3.49), (3.50), we can use Lemma 3.6.3 on $\mathscr{G}_T^*$ since $c_n(\nu_n^* - 1) < 0$. Thus, if $\mathscr{C}_{(i)}^*$ denotes the $i^{\text{th}}$ largest component of $\mathscr{G}_T^*$, then

$$\lim_{T \to \infty} \lim_{\delta \to 0} \limsup_{n \to \infty} \mathbb{P}\left(\sum_{i:|\mathscr{C}_{(i)}^*| \leqslant \delta b_n} |\mathscr{C}_{(i)}^*| \times \mathrm{SP}(\mathscr{C}_{(i)}^*) > \varepsilon b_n\right) = 0. \quad (3.51)$$

To conclude the proof for the whole graph $CM_n(\mathbf{d})$ (with $\lambda > 0$), let

$$\mathcal{K}_n^T := \{i : |\mathscr{C}_{(i)}| \leqslant \delta b_n, |\mathscr{C}_{(i)}| \text{ is explored before the time } T_e\}.$$

Note that

$$\sum_{i \in \mathcal{K}_n^T} |\mathscr{C}_{(i)}| \cdot SP(\mathscr{C}_{(i)}) \leqslant \left( \sum_{i \in \mathcal{K}_n} |\mathscr{C}_{(i)}|^2 \right)^{1/2} \times \left( \sum_{i \in \mathcal{K}_n} SP(\mathscr{C}_{(i)})^2 \right)^{1/2}$$

$$\leqslant \left( \sum_{|\mathscr{C}_{(i)}| \leqslant \delta b_n} |\mathscr{C}_{(i)}|^2 \right)^{1/2} \times SP(T_e),$$

where $SP(t)$ is the number of surplus edges explored up to time $t b_n$ and we have used the fact that $\sum_{i \in \mathcal{K}_n} SP(\mathscr{C}_{(i)})^2 \leqslant (\sum_{i \in \mathcal{K}_n} SP(\mathscr{C}_{(i)}))^2 \leqslant SP(T_e)^2$. From Lemma 3.5.9 and Proposition 3.5.4 we can conclude that for any $T > 0$,

$$\lim_{\delta \to 0} \limsup_{n \to \infty} \mathbb{P}\left( \sum_{i \in \mathcal{K}_n^T} |\mathscr{C}_{(i)}| \cdot SP(\mathscr{C}_{(i)}) > \varepsilon b_n \right) = 0. \tag{3.52}$$

The proof is now complete for the case $\lambda > 0$ by combining (3.51) and (3.52).

$\square$

## 3.7   Proof for simple graphs

In this section, we give a proof of Theorem 3.2.3. Let $\mathbb{P}_s(\cdot)$ (respectively $\mathbb{E}_s[\cdot]$) denote the probability measure (respectively the expectation) conditionally on the graph $CM_n(\mathbf{d})$ being simple. For any process $\mathbf{X}$ on $\mathbb{D}([0, \infty), \mathbb{R})$, we define $\mathbf{X}^T := (X(t))_{t \leqslant T}$. Thus the truncated process $\mathbf{X}^T$ is $\mathbb{D}([0, T], \mathbb{R})$-valued. Now, by [110, Theorem 1.1], $\liminf_{n \to \infty} \mathbb{P}(CM_n(\mathbf{d}) \text{ is simple}) > 0$. This fact ensures that, under the conditional measure $\mathbb{P}_s$, $(b_n^{-1}|\mathscr{C}_{(i)}|)_{i \geqslant 1}$ is tight with respect to the $\ell^2_\downarrow$ topology. Therefore, to conclude Theorem 3.2.3, it suffices to show that the exploration process $\bar{\mathbf{S}}_n$, defined in (3.22), has the same limit (in distribution) under $\mathbb{P}_s$ as obtained in Theorem 3.5.1 so that the finite-dimensional limit of $(b_n^{-1}|\mathscr{C}_{(i)}|)_{i \geqslant 1}$ remains unchanged under $\mathbb{P}_s$. Thus, it is enough to show that for any bounded continuous function $f : \mathbb{D}([0, T], \mathbb{R}) \mapsto \mathbb{R}$,

$$\left| \mathbb{E}[f(\bar{\mathbf{S}}_n^T)] - \mathbb{E}_s[f(\bar{\mathbf{S}}_n^T)] \right| \to 0. \tag{3.53}$$

Let $\ell_n' := \ell_n - 2Tb_n$. We first estimate the number of multiple edges or self-loops discovered in the graph up to time $Tb_n$. Let $v_l$ denote the *exploring* vertex in the breadth-first exploration given by Algorithm 3.1, $d_{v_l}$ the degree of $v_l$ and $(e_1, \ldots, e_r)$ the ordered set of active half-edges of $v_l$ when $v_l$ is

declared to be exploring. Note that, for $l \leqslant Tb_n$, $e_i$ creates a self-loop with probability at most $(d_{v_l} - i)/\ell'_n$ and creates a multiple edge with probability at most $(i-1)/\ell'_n$. Therefore,

$$\mathbb{E}\left[\#\{\text{self-loops/multiple edges discovered while exploring} v_l\}|\mathscr{F}_{l-1}\right] \leqslant \frac{2d^2_{v_l}}{\ell'_n}.$$

Thus, for any $T > 0$,

$$\mathbb{E}\left[\#\{\text{self-loops or multiple edges discovered up to time } Tb_n\right]$$

$$\leqslant \frac{2}{\ell'_n}\mathbb{E}\left[\sum_{i \in [n]} d^2_i \mathcal{I}^n_i(Tb_n)\right]$$

$$= \frac{2}{\ell'_n}\mathbb{E}\left[\sum_{i=1}^{K} d^2_i \mathcal{I}^n_i(Tb_n)\right] + \frac{2}{\ell'_n}\mathbb{E}\left[\sum_{i=K+1}^{n} d^2_i \mathcal{I}^n_i(Tb_n)\right],$$

where $\mathcal{I}^n_i(l) = \mathbf{1}\{i \in \mathscr{V}_l\}$. Now, using Assumption 3.1 (i), for every fixed $K \geqslant 1$,

$$\frac{2}{\ell'_n}\mathbb{E}\left[\sum_{i=1}^{K} d^2_i \mathcal{I}^n_i(Tb_n)\right] \leqslant \frac{2}{\ell'_n}\sum_{i=1}^{K} d^2_i \to 0,$$

since $2\alpha - 1 < 0$. Moreover, recall from (3.24) that $\mathbb{P}\left(\mathcal{I}^n_i(Tb_n) = 1\right) \leqslant Tb_n d_i/\ell'_n$. Therefore, for some constant $C > 0$,

$$\frac{2}{\ell'_n}\mathbb{E}\left[\sum_{i=K+1}^{n} d^2_i \mathcal{I}^n_i(Tb_n)\right] \leqslant \frac{Tb_n}{\ell'^2_n}\sum_{i=K+1}^{n} d^3_i \leqslant C\left(a_n^{-3}\sum_{i=K+1}^{n} d^3_i\right),$$

which, by Assumption 3.1 (ii), tends to zero if we first take $\limsup_{n\to\infty}$ and then take $\lim_{K\to\infty}$. Consequently, for any fixed $T > 0$, as $n \to \infty$,

$$\mathbb{P}\left(\text{at least one self-loop or multiple edge is discovered before time } Tb_n\right) \to 0.$$

Now,

$$\mathbb{E}\left[f(\bar{\mathbf{S}}^T_n)\mathbf{1}\{CM_n(\mathbf{d}) \text{ is simple}\}\right]$$

$$= \mathbb{E}\left[f(\bar{\mathbf{S}}^T_n)\mathbf{1}\{\text{no self-loops or multiple edges found after } Tb_n\}\right] + o(1)$$

$$= \mathbb{E}\left[f(\bar{\mathbf{S}}^T_n)\mathbb{P}\left(\text{no self-loops or multiple edges found after } Tb_n|\mathscr{F}_{Tb_n}\right)\right] + o(1).$$

$$(3.54)$$

Define, $T_e = \inf\{l \geqslant Tb_n : \text{a component is finished exploring at time } l\}$. Using the fact that $(b_n^{-1}T_e)_{n\geqslant 1}$ is a tight sequence, the limit of the expected

number of loops or multiple edges discovered between time $Tb_n$ and $T_e$ is again zero. As in the proof of Proposition 3.6.2, consider the graph $\mathcal{G}^*$, obtained by removing the components obtained up to time $T_e$. Thus, $\mathcal{G}^*$ is a configuration model, conditioned on its degree sequence. Let $\nu_n^*$ be the criticality parameter. Then, we claim that $\nu_n^* \xrightarrow{\mathbb{P}} 1$. To see this note that $\sum_{i \notin \mathscr{V}_{T_e}} d_i = \ell_n + o_{\mathbb{P}}(n)$. Further, note that by Assumption 3.1 (ii) (3.23), for any $t > 0$,

$$\limsup_{n \to \infty} \mathbb{E}\left[a_n^{-2} \sum_{i \in [n]} d_i^2 \mathcal{I}_i(tb_n)\right] \leqslant \limsup_{n \to \infty} a_n^{-2} tb_n \frac{\sum_{i \in [n]} d_i^3}{\ell_n - 2tb_n} < \infty,$$

which implies that $\sum_{i \notin \mathscr{V}_{T_e}} d_i^2 = \sum_{i \in [n]} d_i^2 + o_{\mathbb{P}}(n)$ and thus the claim is proved. Since the degree distribution has finite second moment, using [98, Theorem 7.11] we get

$$\mathbb{P}\left(\mathcal{G}^* \text{ is simple} \middle| \mathscr{F}_{T_e}\right) \xrightarrow{\mathbb{P}} e^{-3/4}. \tag{3.55}$$

Now using (3.54), (3.55) and the dominated convergence theorem, we conclude that

$$\mathbb{E}\left[f(\bar{\mathbf{S}}_n^\mathsf{T})\mathbf{1}\{CM_n(\mathbf{d}) \text{ is simple}\}\right] = \mathbb{E}\left[f(\bar{\mathbf{S}}_n^\mathsf{T})\right]\mathbb{P}\left(CM_n(\mathbf{d}) \text{ is simple}\right) + o(1).$$

Therefore, (3.53) follows and the proof of Theorem 3.2.3 is complete. □

## 3.8    Scaling limits for component functionals

Suppose that vertex $i$ has an associated weight $w_i$. The total weight of the component $\mathscr{C}_{(i)}$ is denoted by $\mathscr{W}_i = \sum_{k \in \mathscr{C}_{(i)}} w_k$. The goal of this section is to derive the scaling limits for $(\mathscr{W}_i)_{i \geqslant 1}$ when the weight sequence satisfies some regularity conditions given below:

**Assumption 3.3.** The weight sequences $w = (w_i)_{i \in [n]}$ satisfies

(i) $\sum_{i \in [n]} w_i = O(n)$, and $\lim_{n \to \infty} \frac{1}{\ell_n} \sum_{i \in [n]} d_i w_i = \mu_w$.

(ii) $\max\{\sum_{i \in [n]} d_i w_i^2, \sum_{i \in [n]} d_i^2 w_i\} = O(a_n^3)$.

**Theorem 3.8.1.** *Consider* $CM_n(\mathbf{d})$ *satisfying Assumption 3.1 and a weight sequence* $w$ *satisfying Assumption 3.3. Denote* $\mathbf{Z}_n^w = \mathrm{ord}(b_n^{-1}\mathscr{W}_i, \mathrm{SP}(\mathscr{C}_{(i)}))_{i \geqslant 1}$ *and* $\mathbf{Z}^w := \mathrm{ord}(\mu_w \gamma_i(\lambda), N(\gamma_i))_{i \geqslant 1}$, *where* $\gamma_i(\lambda)$, *and* $N(\gamma_i)$ *are defined in Theorem 3.2.2. As* $n \to \infty$,

$$\mathbf{Z}_n^w \xrightarrow{d} \mathbf{Z}^w, \tag{3.56}$$

*with respect to the* $\mathbb{U}_\downarrow^0$ *topology.*

The proof of Theorem 3.8.1 can be decomposed in two main steps: the first one is to obtain the finite-dimensional limits of $\mathbf{Z}_n^w$ and then prove the $\mathbb{U}_\downarrow^0$ convergence. The finite-dimensional limit is a consequence of the fact that the total weight of the clusters is approximately equal to the cluster sizes. The argument for the tightness with respect to the $\mathbb{U}_\downarrow^0$ topology is similar to Propositions 3.5.4 and 3.6.2 and therefore we only provide a sketch with pointers to all the necessary ingredients. Recall that $\mathcal{I}_i^n(l) = \mathbf{1}\{i \in \mathcal{V}_l\}$, where $\mathcal{V}_l$ is the set of discovered vertices upto time $l$ by Algorithm 3.1.

**Lemma 3.8.2.** *Under Assumptions 3.1, 3.3, for any* $T > 0$,

$$\sup_{u \leqslant T} \left| \sum_{i \in [n]} w_i \mathcal{I}_i^n(ub_n) - \frac{\sum_{i \in [n]} d_i w_i}{\ell_n} ub_n \right| = O_{\mathbb{P}}(a_n). \qquad (3.57)$$

*Consequently, for each fixed* $i \geqslant 1$,

$$\mathscr{W}_i = \mu_w \left| \mathscr{C}_{(i)} \right| + o_{\mathbb{P}}(b_n). \qquad (3.58)$$

*Proof.* Fix any $T > 0$. Define ,

$$W_n(l) = \sum_{i \in [n]} w_i \mathcal{I}_i^n(l) - \frac{\sum_{i \in [n]} d_i w_i}{\ell_n} l.$$

The goal is to use the supermartingale inequality (3.30) in the same spirit as in the proof of (3.31). Firstly, observe from (3.24) that

$$\mathbb{E}[W_n(l+1) - W_n(l)|\mathscr{F}_l]$$

$$= \mathbb{E}\left[ \sum_{i \in [n]} w_i \left( \mathcal{I}_i^n(l+1) - \mathcal{I}_i^n(l) \right) \Big| \mathscr{F}_l \right] - \frac{\sum_{i \in [n]} d_i w_i}{\ell_n}$$

$$= \sum_{i \in [n]} w_i \mathbb{E}\left[ \mathcal{I}_i^n(l+1) \big| \mathscr{F}_l \right] \mathbf{1}\{\mathcal{I}_i^n(l) = 0\} - \frac{\sum_{i \in [n]} d_i w_i}{\ell_n} \geqslant 0,$$

uniformly over $l \leqslant Tb_n$ and therefore, $(W_n(l))_{l=1}^{Tb_n}$ is a sub-martingale. Let $\ell_n' = \ell_n - 2Tb_n - 1$. Using (3.25), we compute

$$\left| \mathbb{E}[W_n(l)] \right| = \sum_{i \in [n]} w_i \left( \mathbb{P}\left( \mathcal{I}_i^n(l) = 1 \right) - \frac{d_i}{\ell_n} l \right)$$

$$\leqslant \sum_{i \in [n]} w_i \left( 1 - \left( 1 - \frac{d_i}{\ell_n'} \right)^l - \frac{d_i}{\ell_n'} l \right) + l \sum_{i \in [n]} w_i \left( \frac{d_i}{\ell_n'} - \frac{d_i}{\ell_n} \right)$$

$$\leqslant 2(2Tb_n)^2 \frac{\sum_{i \in [n]} d_i^2 w_i}{\ell_n'^2} = O(b_n^2 a_n^3/n^2) = O(a_n),$$

$$(3.59)$$

uniformly over $l \leqslant Tb_n$. Also, using (3.27), (3.28), and Assumption 3.3 (ii),

$$\text{Var}(W_n(l)) \leqslant \sum_{i \in [n]} w_i^2 \text{var}(\mathfrak{I}_i^n(l)) \leqslant Tb_n \frac{\sum_{i \in [n]} d_i w_i^2}{\ell_n'} = O(a_n^2), \qquad (3.60)$$

uniformly over $l \leqslant Tb_n$. Using (3.30), (3.59) and (3.60), we conclude the proof of (3.57). The proof of (3.58) follows using Lemma 3.5.6 and simply observing that $a_n = o(b_n)$. $\qquad\qquad\square$

*Proof of Theorem 3.8.1.* Lemma 3.8.2 ensures the finite-dimensional convergence in (3.56). Thus, the proof is complete if we can show that, for any $\varepsilon > 0$

$$\lim_{K \to \infty} \limsup_{n \to \infty} \mathbb{P}\left( \sum_{i > K} \mathscr{W}_i^2 > \varepsilon b_n^2 \right) = 0, \qquad (3.61a)$$

and

$$\lim_{\delta \to 0} \limsup_{n \to \infty} \mathbb{P}\left( \sum_{|\mathscr{C}_{(i)}| \leqslant \delta b_n} \mathscr{W}_i \times \text{SP}(\mathscr{C}_{(i)}) > \varepsilon b_n \right) = 0. \qquad (3.61b)$$

The arguments for proving (3.61a), and (3.61b) are similar to those for ropositions 3.5.4, and 3.6.2 and thus we only sketch a brief outline. Denote $\ell_n^w = \sum_{i \in [n]} w_i$. The main ingredient to the proof of Proposition 3.5.4 is Lemma 3.5.5, and the proof of Lemma 3.5.5 uses the fact that the expected sum of squares of the cluster sizes can be written in terms of susceptibility functions in (3.38) and then we made use of the estimate for the susceptibility function in (3.37). Let $V_n'$ denote a vertex chosen according to the distribution $(w_i/\ell_n^w)_{i \in [n]}$, independently of the graph. Notice that for any $CM_n(\mathbf{d})$,

$$\mathbb{E}\left[ \sum_{i \geqslant 1} \mathscr{W}_i^2 \right] = \ell_n^w \mathbb{E}\left[ \mathscr{W}(V_n') \right]. \qquad (3.62)$$

Now, [111, Lemma 5.2] can be extended using an identical argument to compute the weight-based susceptibility function in the right hand side of (3.62). See Lemma 3.A.1 given in Appendix 3.A. The proof of (3.61b) can also be completed using an identical argument as Proposition 3.6.2 by observing that

$$\mathbb{P}\left( \sum_{|\mathscr{C}_{(i)}| \leqslant \delta b_n} \mathscr{W}_i \times \text{SP}(\mathscr{C}_{(i)}) > \varepsilon b_n \right) \leqslant \frac{\ell_n^w}{\varepsilon b_n} \mathbb{E}\left[ \text{SP}(\mathscr{C}(V_n')) \mathbf{1}_{\{|\mathscr{C}(V_n')| \leqslant \delta b_n\}} \right].$$

Moreover, an analogue of Lemma 3.6.3 also holds for $V_n'$ (see Appendix 3.B), and the proof of (3.61b) can now be completed in an identical manner as the proof of Proposition 3.6.2. $\qquad\qquad\square$

While studying percolation in the next section, we will need an estimate for the proportion of degree-one vertices in the large components. In fact, an application of Theorem 3.8.1, yields the following result about the degree composition of the largest clusters:

**Corollary 3.8.3.** *Consider* $\mathrm{CM}_n(\mathbf{d})$ *satisfying Assumption 3.1. Let* $\nu_k(G)$ *denote the number of vertices of degree* $k$ *in the graph* $G$. *Then, for any fixed* $i \geqslant 1$,

$$\nu_k(\mathscr{C}_{(i)}) = \frac{kr_k}{\mu}|\mathscr{C}_{(i)}| + o_{\mathbb{P}}(b_n), \tag{3.63}$$

*where* $r_k = \mathbb{P}(D = k)$. *Denote* $\mathbf{Z}_n^k = \mathrm{ord}(b_n^{-1}\nu_k(\mathscr{C}_{(i)}), \mathrm{SP}(\mathscr{C}_{(i)}))_{i\geqslant 1}$, $\mathbf{Z}^k := \mathrm{ord}(\frac{kr_k}{\mu}\gamma_i(\lambda), N(\gamma_i))_{i\geqslant 1}$, *where* $\gamma_i(\lambda)$, *and* $N(\gamma_i)$ *are defined in Theorem 3.2.2. As* $n \to \infty$,

$$\mathbf{Z}_n^k \xrightarrow{d} \mathbf{Z}^k,$$

*with respect to the* $\mathbb{U}_\downarrow^0$ *topology.*

*Proof.* The proof follows directly from Theorem 3.8.1 by putting $w_i = \mathbf{1}\{d_i = k\}$. The fact that this weight sequence satisfies Assumption 3.3 is a consequence of Assumption 3.1. □

## 3.9    Percolation

In this section, we study critical percolation on the configuration model for fixed $\lambda \in \mathbb{R}$ and complete the proof of Theorem 3.2.4. As discussed earlier, $\mathrm{CM}_n(\mathbf{d}, p)$ is obtained by first constructing $\mathrm{CM}_n(\mathbf{d})$ and then deleting each edge with probability $1 - p$, independently of each other, and the graph $\mathrm{CM}_n(\mathbf{d})$. An interesting property of the configuration model is that $\mathrm{CM}_n(\mathbf{d}, p)$ is also distributed as a configuration model conditional on the degrees [85]. The rough idea here is to show that the degree distribution of $\mathrm{CM}_n(\mathbf{d}, p_n(\lambda))$ satisfies Assumption 3.1, where $p_n(\lambda)$ is given by Assumption 3.2. This allows us to invoke Theorem 3.2.2 and complete the proof of Theorem 3.2.4. Recall from Assumption 3.2 that $\nu = \lim_{n\to\infty} \nu_n > 1$, and $p_n = p_n(\lambda) = \nu_n^{-1}(1 + \lambda c_n^{-1})$. We start by describing an algorithm due to Janson [109] that is easier to work with.

**Algorithm 3.2 (Construction of $\mathbf{CM_n(d, p_n)}$).** Initially, vertex $i$ has $d_i$ half-edges incident to it. For each half-edge $e$, let $v_e$ be the vertex to which $e$ is incident.

(S1) With probability $1 - \sqrt{p_n}$, one detaches $e$ from $v_e$ and associates $e$ to a new vertex $v'$ of degree-one. Color the new vertex red. This is done independently for every existing half-edge and we call this whole process explosion. Let $n_+$ be the number of red vertices created by explosion and $\tilde{n} = n + n_+$. Denote the degree sequence obtained from the above procedure by $\tilde{d} = (\tilde{d}_i)_{i \in [\tilde{n}]}$, i.e., $\tilde{d}_i \sim \text{Bin}(d_i, \sqrt{p_n})$ for $i \in [n]$ and $\tilde{d}_i = 1$ for $i \in [\tilde{n}] \setminus [n]$;

(S2) Construct $\text{CM}_{\tilde{n}}(\tilde{d})$ independently of (S1);

(S3) Delete all the red vertices and the edges attached to them.

It was also shown in [109] that the obtained multigraph has the same distribution as $\text{CM}_n(d, p)$ if we replace (S3) by

(S3′) Instead of deleting red vertices, choose $n_+$ degree-one vertices uniformly at random without replacement, independently of (S1), and (S2) and delete them.

**Remark 27.** Notice that Algorithm 3.2 (S1) induces a probability measure $\mathbb{P}_p^n$ on $\mathbb{N}^\infty$. Denote their product measure by $\mathbb{P}_p$. In words, for different $n$, (S1) is carried out independently. All the almost sure statements about the degrees in this section will be with respect to the probability measure $\mathbb{P}_p$.

Let us first show that $\tilde{d}$ also satisfies Assumption 3.1 (ii). Note that the total number of half-edges remains unchanged during the explosion in Algorithm 3.2 (S1) and therefore, $\sum_{i \in [\tilde{n}]} \tilde{d}_i = \sum_{i \in [n]} d_i$ and by Assumption 3.2 (i),

$$\frac{1}{n} \sum_{i \in [\tilde{n}]} \tilde{d}_i \to \mu \quad \mathbb{P}_p \text{ a.s.} \tag{3.64}$$

This verifies the first moment condition in Assumption 3.1 (ii) for the percolated degree sequence $\mathbb{P}_p$ a.s. Let $I_{ij} :=$ the indicator of the $j^{\text{th}}$ half-edge corresponding to vertex $i$ being kept after the explosion. Then $I_{ij} \sim \text{Ber}(\sqrt{p_n})$ independently for $i \in [n], j \in [d_i]$. Let

$$\mathbf{I} := (I_{ij})_{j \in [d_i], i \in [n]} \quad \text{and} \quad f_1(\mathbf{I}) := \sum_{i \in [n]} \tilde{d}_i(\tilde{d}_i - 1).$$

Note that $f_1(\mathbf{I}) = \sum_{i \in [\tilde{n}]} \tilde{d}_i(\tilde{d}_i - 1)$ since the degree-one vertices do not contribute to the sum. One can check that by changing the status of one half-edge

corresponding to vertex k we can change $f_1$ by at most $2(d_k + 1)$. Therefore an application of [115, Corollary 2.27] yields

$$
\mathbb{P}_p\left(\left| \sum_{i\in[n]} \tilde{d}_i(\tilde{d}_i - 1) - p_n \sum_{i\in[n]} d_i(d_i - 1)\right| > t\right)
$$
$$
\leqslant 2\exp\left(-\frac{t^2}{2\sum_{i\in[n]} d_i(d_i + 1)^2}\right). \tag{3.65}
$$

Now by Assumption 3.2 (i), $\sum_{i\in[n]} d_i^3 = O(a_n^3)$. If we set $t = n^{1-\varepsilon}c_n^{-1}$, then $t^2/(\sum_{i\in[n]} d_i^3)$ is of the order $n^{\alpha-2\varepsilon}/L(n)$. Thus, choosing $\varepsilon < \alpha/2$, using (3.65) and the Borel-Cantelli lemma we conclude that

$$
\sum_{i\in[n]} \tilde{d}_i(\tilde{d}_i - 1) = p_n \sum_{i\in[n]} d_i(d_i - 1) + o(nc_n^{-1}) \quad \mathbb{P}_p \text{ a.s.} \tag{3.66}
$$

Thus, using Assumption 3.2, the second moment condition in Assumption 3.1 (ii) is verified for the percolated degree sequence $\mathbb{P}_p$ a.s. Let $\tilde{d}_{(i)}$ denote the $i^{\text{th}}$ largest value of $(\tilde{d}_i)_{i\in[\tilde{n}]}$. The third-moment condition in Assumption 3.1 (ii) is obtained by noting that $\tilde{d}_i \leqslant d_i$ for all $i \in [n]$ and

$$
\lim_{K\to\infty} \limsup_{n\to\infty} a_n^{-3} \sum_{i=K+1}^{\tilde{n}} \tilde{d}_{(i)}^3 \leqslant \lim_{K\to\infty} \limsup_{n\to\infty} a_n^{-3} \sum_{i=K+1}^{\tilde{n}} \tilde{d}_i^3
$$
$$
\leqslant \lim_{K\to\infty} \limsup_{n\to\infty} a_n^{-3}\left(\sum_{i=K+1}^{n} \tilde{d}_i^3 + n_+\right) \leqslant \lim_{K\to\infty} \limsup_{n\to\infty} a_n^{-3}\left(\sum_{i=K+1}^{n} d_i^3 + n_+\right), \tag{3.67}
$$

which tends to zero $\mathbb{P}_p$ a.s., where we have used Assumption 3.2 (i) and the fact that $a_n^{-3}n_+ \to 0$, $\mathbb{P}_p$ a.s., which follows by observing that $n_+ \sim \text{Bin}(\ell_n, 1 - \sqrt{p_n})$. To see that $\tilde{d}$ satisfies Assumption 3.1 (iii) note that by (3.66),

$$
\frac{\sum_{i\in[\tilde{n}]} \tilde{d}_i(\tilde{d}_i - 1)}{\sum_{i\in[\tilde{n}]} \tilde{d}_i} = p_n \frac{\sum_{i\in[n]} d_i(d_i - 1)}{\sum_{i\in[n]} d_i} + o(c_n^{-1}) = 1 + \lambda c_n^{-1} + o(c_n^{-1})
$$

$\mathbb{P}_p$ a.s., where the last step follows from Assumption 3.2 (ii). Assumption 3.1 (iv) is trivially satisfied by $\tilde{d}$. Finally, in order to verify Assumption 3.1 (i), it suffices to show that

$$
\frac{\tilde{d}_{(i)}}{a_n} \to \theta_i\sqrt{p}, \quad \mathbb{P}_p \text{ a.s.}, \tag{3.68}
$$

where $p = 1/v$. Recall that $\tilde{d}_i \sim \text{Bin}(d_i, \sqrt{p_n})$. A standard concentration inequality for the binomial distribution [115, (2.9)] yields that, for any $0 <$

$\varepsilon \leqslant 3/2$,

$$\mathbb{P}(|\tilde{d}_i - d_i\sqrt{p_n}| > \varepsilon d_i\sqrt{p_n}) \leqslant 2\exp(-\varepsilon^2 d_i\sqrt{p_n}/3),$$

and using the Borel-Cantelli lemma it follows that $\mathbb{P}_p$ almost surely, $\tilde{d}_i = d_i\sqrt{p_n}(1 + o(1))$ for all fixed $i$. Moreover, an application of (3.67) yields that

$$\lim_{K\to\infty} \limsup_{n\to\infty} a_n^{-3} \max_{i>K} \tilde{d}_i^3 = 0.$$

Now, since $\theta$ is an ordered vector, the proof of (3.68) follows.

To summarize, the above discussion in (3.64), (3.66), (3.67), and (3.68) yields that the degree sequence $\tilde{d}$ satisfies all the conditions in Assumption 3.1. Therefore, Theorem 3.2.2 can be applied to $\mathrm{CM}_{\tilde{n}}(\tilde{d})$. Denote by $\mathscr{C}_{(i)}$ the $i^{\text{th}}$ largest component of $\mathrm{CM}_{\tilde{n}}(\tilde{d})$. Let $\tilde{Z}_n = \mathrm{ord}(b_n^{-1}|\mathscr{C}_{(i)}|, \mathrm{SP}(\mathscr{C}_{(i)})_{i\geqslant 1}$ and $\tilde{Z} := \mathrm{ord}(\tilde{\gamma}_i(\lambda), N(\tilde{\gamma}_i))_{i\geqslant 1}$, where $\gamma_i(\lambda)$, and $N(\gamma_i)$ are defined in Theorem 3.2.4. Now, Theorem 3.2.2 implies

$$\tilde{Z}_n \xrightarrow{\mathrm{d}} \tilde{Z},$$

with respect to the $\mathbb{U}_{\downarrow}^0$ topology.

Since the percolated degree sequence satisfies Assumption 3.1 $\mathbb{P}_p$ a.s., (3.63) holds for $\mathscr{C}_{(i)}$ also. Let $v_1^d(\mathscr{C}_{(i)})$ be the number of degree-one vertices of $\mathscr{C}_{(i)}$ which are deleted while creating the graph $\mathrm{CM}_n(d, p_n)$ from $\mathrm{CM}_{\tilde{n}}(\tilde{d})$. Since the vertices are to be chosen uniformly from all degree-one vertices as described in (S3′),

$$v_1^d(\mathscr{C}_{(i)}) = \frac{n_+}{\tilde{n}_1}v_1(\mathscr{C}_{(i)}) + o_{\mathbb{P}}(b_n) = \frac{n_+}{\tilde{n}_1}\frac{\tilde{n}_1}{\ell_n}|\mathscr{C}_{(i)}| + o_{\mathbb{P}}(b_n) = \frac{n_+}{\ell_n}|\mathscr{C}_{(i)}| + o_{\mathbb{P}}(b_n)$$

$$= \frac{\mu(1 - \sqrt{p_n}) + o(1)}{\mu + o(1)}|\mathscr{C}_{(i)}| + o_{\mathbb{P}}(b_n) = (1 - \sqrt{p_n})|\mathscr{C}_{(i)}| + o_{\mathbb{P}}(b_n),$$

where the last-but-one equality follows by observing that $n_+ \sim \mathrm{Bin}(\ell_n, 1 - \sqrt{p_n})$. Now, notice that by removing degree-one vertices, the components are not broken up, so the vector of component sizes for percolation can be obtained by just subtracting the number of red vertices from the component sizes of $\mathrm{CM}_{\tilde{n}}(\tilde{d})$. Moreover, the removal of degree-one vertices does not effect the count of surplus edges. Therefore, the proof of Theorem 3.2.4 is complete by using Corollary 3.8.3.

## 3.10    Convergence to augmented multiplicative coalescent

Let us give an overview of the organization of this section: In Section 3.10.1, we discuss an alternative dynamic construction that approximates the percolated graph process, coupled in a natural way. This construction enables

us to compare the coupled percolated graphs with a dynamic construction. Then, we describe a modified system that evolves as an exact augmented multiplicative coalescent and the rest of the section is devoted to comparing the exact augmented multiplicative coalescent and the corresponding quantities for the graphs generated by the dynamic construction. The ideas are similar to [70, Section 8], and we only give the overall idea and the necessary details specific to this chapter.

### 3.10.1   The dynamic construction and the coupling

Let us consider graphs generated dynamically as follows:

**Algorithm 3.3.** Let $s_1(t)$ be the total number of unpaired or *open* half-edges at time t, and $\Xi_n$ be an inhomogeneous Poisson process with rate $s_1(t)$ at time t.

(S0)  Initially, $s_1(0) = \ell_n$, and $\mathcal{G}_n(0)$ is the empty graph on vertex set $[n]$.

(S1)  At each event time of $\Xi_n$, choose two open half-edges uniformly at random and pair them. The graph $\mathcal{G}_n(t)$ is obtained by adding this edge to $\mathcal{G}_n(t-)$. Decrease $s_1(t)$ by two. Continue until $s_1(t)$ becomes zero.

Notice that $\mathcal{G}_n(\infty)$ is distributed as $CM_n(\mathbf{d})$ since an open half-edge is paired with another uniformly chosen open half-edge. The next proposition ensures that the graph process generated by Algorithm 3.3 *sandwiches* the graph process $(CM_n(\mathbf{d}, p_n(\lambda)))_{\lambda \in \mathbb{R}}$. This result was proved in [70, Proposition 28]. The proof is identical under Assumption 3.2 and therefore is omitted here. Define,

$$t_n(\lambda) = \frac{1}{2} \log \left( \frac{\nu_n}{\nu_n - 1} \right) + \frac{1}{2(\nu_n - 1)} \frac{\lambda}{c_n}.$$

**Proposition 3.10.1.** *Fix* $-\infty < \lambda_\star < \lambda^\star < \infty$. *There exists a coupling such that with high probability*

$$\mathcal{G}_n(t_n(\lambda) - \varepsilon_n) \subset CM_n(\mathbf{d}, p_n(\lambda)) \subset \mathcal{G}_n(t_n(\lambda) + \varepsilon_n), \quad \forall \lambda \in [\lambda_\star, \lambda^\star]$$

*and*

$$CM_n(\mathbf{d}, p_n(\lambda) - \varepsilon_n) \subset \mathcal{G}_n(t_n(\lambda)) \subset CM_n(\mathbf{d}, p_n(\lambda) + \varepsilon_n), \quad \forall \lambda \in [\lambda_\star, \lambda^\star] \tag{3.69a}$$

*where* $\varepsilon_n = cn^{-\gamma_0}$, *for some* $\eta < \gamma_0 < 1/2$ *and the constant* $c$ *does not depend on* $\lambda$.

From here onward, we augment $\lambda$ to a predefined notation to emphasize the dependence on $\lambda$. We write $\mathscr{C}_{(i)}(\lambda)$ for the $i^{\text{th}}$ largest component of $\mathcal{G}_n(t_n(\lambda))$ and define

$$\mathcal{O}_i(\lambda) = \# \text{ open half-edges in } \mathscr{C}_{(i)}(\lambda).$$

Think of $\mathcal{O}_i(\lambda)$ as the *mass* of the component $\mathscr{C}_{(i)}(\lambda)$. Let $\mathbf{Z}_n^o(\lambda)$ denote the vector of the number of open half-edges (re-scaled by $b_n$) and surplus edges of $\mathcal{G}_n(t_n(\lambda))$, ordered as an element of $\mathbb{U}_\downarrow^0$. For a process $\mathbf{X}$, we will write $\mathbf{X}[\lambda_\star, \lambda^\star]$ to denote the restricted process $(X(\lambda))_{\lambda \in [\lambda_\star, \lambda^\star]}$. Let $\ell_n^o(\lambda) = \sum_{i \geqslant 1} \mathcal{O}_i(\lambda)$. Note that

$$\ell_n^o(\lambda) = \frac{n\mu(\nu - 1)}{\nu}(1 + o_{\mathbb{P}}(1)). \tag{3.70}$$

(3.70) is a consequence of [29, Lemma 8.2] since the proof only uses the facts that $|\ell_n/n - \mu| = o(n^{-\gamma})$ for all $\gamma < 1/2$, and $\sum_{i \in [n]} d_i(d_i - 1)/\ell_n \to \nu$. Now, observe that, during the evolution of the graph process generated by Algorithm 3.3, during the time interval $[t_n(\lambda), t_n(\lambda + d\lambda)]$, the $i^{\text{th}}$ and $j^{\text{th}}$ $(i > j)$ largest components, merge at rate

$$2\mathcal{O}_i(\lambda)\mathcal{O}_j(\lambda) \times \frac{1}{\ell_n^o(\lambda) - 1} \times \frac{1}{2(\nu_n - 1)c_n} \approx \frac{\nu}{\mu(\nu - 1)^2}\left(b_n^{-1}\mathcal{O}_i(\lambda)\right)\left(b_n^{-1}\mathcal{O}_j(\lambda)\right), \tag{3.71}$$

and create a component with $\mathcal{O}_i(\lambda) + \mathcal{O}_j(\lambda) - 2$ open half-edges and $SP(\mathscr{C}_{(i)}(\lambda)) + SP(\mathscr{C}_{(j)}(\lambda))$ surplus edges. Also, a surplus edge is created in $\mathscr{C}_{(i)}(\lambda)$ at rate

$$\mathcal{O}_i(\lambda)(\mathcal{O}_i(\lambda) - 1) \times \frac{1}{\ell_n^o(\lambda) - 1} \times \frac{1}{2(\nu_n - 1)c_n} \approx \frac{\nu}{2\mu(\nu - 1)^2}\left(b_n^{-1}\mathcal{O}_i(\lambda)\right)^2, \tag{3.72}$$

and $\mathscr{C}_{(i)}(\lambda)$ becomes a component with surplus edges $SP(\mathscr{C}_{(i)}(\lambda)) + 1$ and open half-edges $\mathcal{O}_i(\lambda) - 2$. Thus $\mathbf{Z}_n^o[\lambda_\star, \lambda^\star]$ does *not* evolve as an AMC process but it is close. The fact that two half-edges are killed after pairing, makes the masses (the number of open half-edges) of the components deplete. If there were no such depletion of mass, then the vector of open half-edges, along with the surplus edges, would in fact merge as an augmented multiplicative coalescent. Let us define the modified process [70, Algorithm 7] that in fact evolves as augmented multiplicative coalescent:

**Algorithm 3.4.** Initialize $\bar{\mathcal{G}}_n(t_n(\lambda_\star)) = \mathcal{G}_n(t_n(\lambda_\star))$. Let $\mathscr{O}$ denote the set of open half-edges in the graph $\mathcal{G}_n(t_n(\lambda_\star))$, $\bar{s}_1 = |\mathscr{O}|$ and $\bar{\Xi}_n$ denote a Poisson process with rate $\bar{s}_1$. At each event time of the Poisson process $\bar{\Xi}_n$, select two

half-edges from $\mathcal{O}$ and create an edge between the corresponding vertices. However, the selected half-edges are kept alive, so that they can be selected again.

**Remark 28.** The only difference between Algorithms 3.3 and 3.4, is that the *paired* half-edges are not discarded and thus more edges are created by Algorithm 3.4. Thus, there is a natural coupling between the graphs generated by Algorithms 3.3 and 3.4 such that $\mathcal{G}_n(t_n(\lambda)) \subset \bar{\mathcal{G}}_n(t_n(\lambda))$ for all $\lambda \in [\lambda_\star, \lambda^\star]$, with probability one. In the subsequent part of this section, we will always work under this coupling. The extra edges that are created by Algorithm 3.4 will be called *bad* edges.

In the subsequent part of this chapter, we will augment a predefined notation with a bar to denote the corresponding quantity for $\bar{\mathcal{G}}_n(t_n(\lambda))$. Denote $\beta_n = (\bar{s}_1(\nu_n - 1)c_n)^{1/2}$ and $\bar{\mathbf{Z}}_n^{o,\mathrm{scl}}(\lambda)$ denote the vector $\mathrm{ord}(\beta_n^{-1}\bar{O}_i(\lambda), \mathrm{SP}(\bar{\mathscr{C}}_{(i)}(\lambda)))_{i\geq 1}$. Using an argument identical to (3.71), and (3.72), it follows that $\bar{\mathbf{Z}}_n^{o,\mathrm{scl}}[\lambda_\star, \lambda^\star]$ evolves as a standard augmented multiplicative coalescent. Note that there exists a constant $c > 0$ such that $\beta_n = cb_n(1 + o_\mathbb{P}(1))$, and therefore the scaling limit of any finite-dimensional distributions of $\bar{\mathbf{Z}}_n^o[\lambda_\star, \lambda^\star]$ can be obtained from $\bar{\mathbf{Z}}_n^{o,\mathrm{scl}}[\lambda_\star, \lambda^\star]$.

**Augmented multiplicative coalescent with mass and weight**

The near Feller property of the augmented multiplicative coalescent [30, Theorem 3.1] ensures the joint convergence of the number of open half-edges in each component together with the surplus edges of $\bar{\mathcal{G}}_n(t_n(\lambda))$. To deduce the scaling limits involving the components sizes let us consider a dynamic process that is further augmented by weight. Initially, the system consists of particles (possibly infinitely many) where particle $i$ has mass $x_i$, weight $z_i$ and an attribute $y_i$. Let $(X_i(t), Z_i(t), Y_i(t))_{i\geq 1}$ denote masses, weights, and attribute values at time $t$. The dynamics of the system is described as follows: At time $t$,

$\triangleright$ particles $i$ and $j$ coalesce at rate $X_i(t)X_j(t)$ and create a particle with mass $X_i(t) + X_j(t)$, weight $Z_i(t) + Z_j(t)$ and attribute $Y_i(t) + Y_j(t)$.

$\triangleright$ for each $i$, attribute $Y_i(t)$ increases by 1 at rate $Y_i^2(t)/2$.

For $(x, y), (z, y) \in \mathbb{U}_\downarrow^0$, we write $(x, z, y)$ for $((x, y), (z, y)) \in (\mathbb{U}_\downarrow^0)^2$. Denote by $\mathrm{MC}_2(x, z, t)$ and $\mathrm{AMC}_2(x, z, y, t)$ respectively the vector $(X_i(t), Z_i(t))_{i\geq 1}$ and $(X_i(t), Z_i(t), Y_i(t))_{i\geq 1}$ with initial mass $x$, weight $z$ and attribute value $y$. We will need the following theorem:

**Theorem 3.10.2.** *Suppose that* $(\boldsymbol{x_n}, \boldsymbol{z_n}, \boldsymbol{y_n}) \to (\boldsymbol{x}, \boldsymbol{x}, \boldsymbol{y})$ *in* $(\mathbb{U}_\downarrow^0)^2$ *and* $\sum_i x_i = \infty$. *Then, for any* $t \geqslant 0$

$$AMC_2(\boldsymbol{x_n}, \boldsymbol{z_n}, \boldsymbol{y_n}) \xrightarrow{d} AMC_2(\boldsymbol{x}, \boldsymbol{x}, \boldsymbol{y}).$$

*Proof.* By [70, Theorem 29],

$$MC_2(\boldsymbol{x_n}, \boldsymbol{z_n}, t) \xrightarrow{d} MC_2(\boldsymbol{x}, \boldsymbol{x}, t). \tag{3.73}$$

For $\boldsymbol{x_n} = (x_i^n)_{i \geqslant 1}$, and $\boldsymbol{z_n} = (z_i^n)_{i \geqslant 1}$ let $\boldsymbol{w_n^+} = \mathrm{sort}(x_i^n \vee z_i^n)$, $\boldsymbol{w_n^-} = \mathrm{sort}(x_i^n \wedge z_i^n)$, where sort denotes the decreasing ordering of the elements. Notice that $\boldsymbol{w_n^+} \to \boldsymbol{x}$, and $\boldsymbol{w_n^-} \to \boldsymbol{x}$ in $\ell_\downarrow^2$. Let us denote by $AMC_1(\boldsymbol{x}, \boldsymbol{y}, t)$ the usual augmented multiplicative coalescent process at time $t$ with starting state $(\boldsymbol{x}, \boldsymbol{y})$. Now, since $\sum_i x_i = \infty$, we can use the near Feller property [30, Theorem 3.1] to conclude that $AMC_1(\boldsymbol{x_n}, \boldsymbol{y_n}, t) \xrightarrow{d} AMC_1(\boldsymbol{x}, \boldsymbol{y}, t)$. Moreover, $AMC_2(\boldsymbol{w_n^+}, \boldsymbol{w_n^+}, \boldsymbol{y_n}, t)$ and $AMC_2(\boldsymbol{w_n^-}, \boldsymbol{w_n^-}, \boldsymbol{y_n}, t)$ converges to the same limit. For $(\boldsymbol{x}, \boldsymbol{z}, \boldsymbol{y}) \in (\mathbb{U}_\downarrow^0)^2$, if $S_{pr}(\boldsymbol{x}, \boldsymbol{z}, \boldsymbol{y}) = \sum_i z_i y_i$, then under the subgraph coupling

$$S_{pr}(AMC_2(\boldsymbol{w_n^+}, \boldsymbol{w_n^+}, \boldsymbol{y_n}, t)) - S_{pr}(AMC_2(\boldsymbol{w_n^-}, \boldsymbol{w_n^-}, \boldsymbol{y_n}, t)) \xrightarrow{\mathbb{P}} 0,$$

which implies that

$$\begin{aligned}
(AMC_1(\boldsymbol{x_n}, \boldsymbol{y_n}, t), S_{pr}(AMC_2(\boldsymbol{x_n}, \boldsymbol{z_n}, \boldsymbol{y_n}, t))) \\
\xrightarrow{d} (AMC_1(\boldsymbol{x}, \boldsymbol{y}, t), S_{pr}(AMC_2(\boldsymbol{x}, \boldsymbol{x}, \boldsymbol{y}, t))).
\end{aligned} \tag{3.74}$$

Now, using (3.73), (3.74), an application of [30, Lemma 4.11] concludes the proof. ∎

### 3.10.2   Asymptotics for the open half-edges

The following lemma shows that the number of open half-edges in $\mathcal{G}_n(t_n(\lambda))$ is *approximately* proportional to the component sizes. This will enable us to apply Theorem 3.10.2 for deducing the scaling limits of the required quantities for the graph $\bar{\mathcal{G}}_n(t_n(\lambda))$.

**Lemma 3.10.3.** *There exists a constant* $\kappa > 0$ *such that, for any* $i \geqslant 1$,

$$\mathcal{O}_i(\lambda) = \kappa |\mathscr{C}_{(i)}(\lambda)| + o_{\mathbb{P}}(b_n). \tag{3.75}$$

*Further,* $(\mathbf{Z}_n^o(\lambda))_{n \geqslant 1}$ *is tight in* $\mathbb{U}_\downarrow^0$.

*Proof.* Let $(d_k^\lambda)_{k\in[n]}$ denote the degree sequence of $\mathrm{CM}_n(\mathbf{d}, p_n(\lambda))$ and define

$$\mathcal{O}_i^p(\lambda) = \sum_{k\in\mathscr{C}_{(i)}^p(\lambda)} (d_k - d_k^\lambda) = \sum_{k\in\mathscr{C}_{(i)}^p(\lambda)} d_k - 2(|\mathscr{C}_{(i)}^p(\lambda)| - 1 + \mathrm{SP}(\mathscr{C}_{(i)}^p(\lambda))).$$

Using (3.69a) and the fact that the number of surplus edges in the large components are tight, it is enough to prove the lemma by replacing $\mathcal{O}_i(\lambda)$ by $\mathcal{O}_i^p(\lambda)$ and $\mathscr{C}_{(i)}'(\lambda)$ by $\mathscr{C}_{(i)}^p(\lambda)$. For a component $\mathscr{C}$ of $\mathrm{CM}_{\tilde{n}}(\tilde{\mathbf{d}})$, the corresponding component in the percolated graph is obtained by cleaning up $\mathrm{R}(\tilde{\mathscr{C}})$ red degree-one vertices. Thus, the degree deficiency of that percolated cluster is given by

$$\sum_{k\in\mathscr{C}\cap[n]} d_k - \sum_{k\in\mathscr{C}\cap[n]} \tilde{d}_k + 2\mathrm{R}(\mathscr{C}). \tag{3.76}$$

Now, all the three terms appearing in the right hand side of (3.76) can be estimated using Theorem 3.8.1, where we recall from Section 3.9 that $\tilde{\mathbf{d}}$ satisfies Assumption 3.1. The proof is now complete. $\qquad\square$

For an element $\mathbf{z} = (x_i, y_i)_{i\geqslant 1} \in \mathbb{U}_{\downarrow}^0$ and a constant $c > 0$, denote $c\mathbf{z} = (cx_i, y_i)_{i\geqslant 1}$. Thus, Lemma 3.10.3 states that, for each fixed $\lambda$, $\mathbf{Z}_n^o(\lambda)$ is close to $\kappa\mathbb{Z}_n(\lambda)$. The following lemma states that formally:

**Corollary 3.10.4.** *For each fixed $\lambda$, as $n \to \infty$, $d_U(\mathbf{Z}_n^o(\lambda), \kappa\mathbb{Z}_n(\lambda)) \xrightarrow{\mathbb{P}} 0$.*

*Proof.* Let $\pi_k, T_k : \mathbb{U}_{\downarrow}^0 \mapsto \mathbb{U}_{\downarrow}^0$ be the functions such that for $\mathbf{z} = ((x_i, y_i))_{i\geqslant 1}$, $\pi_k(\mathbf{z})$ consists of only $(x_i, y_i)$ for $i \leqslant k$ and zeroes in other coordinates, and $T_k(\mathbf{z})$ consists only of $(x_i, y_i)$ for $i > k$. Thus,

$$\begin{aligned} d_U\left(\mathbf{Z}_n^o(\lambda), \kappa\mathbb{Z}_n(\lambda)\right) \leqslant &\; d_U\left(\pi_K(\mathbf{Z}_n^o(\lambda)), \pi_K(\kappa\mathbb{Z}_n(\lambda))\right) \\ &+ \|T_K(\mathbf{Z}_n^o(\lambda))\|_U + \|T_K(\kappa\mathbb{Z}_n(\lambda))\|_U. \end{aligned} \tag{3.77}$$

Now, for each fixed $K \geqslant 1$ the first term in the right hand side of (3.77) converges in probability to zero, by (3.75). Also, using the tightness of both $(\mathbb{Z}_n(\lambda))_{n\geqslant 1}$ and $(\mathbb{Z}_n^o(\lambda))_{n\geqslant 1}$ with respect to the $\mathbb{U}_{\downarrow}^0$ topology, it follows that for any $\varepsilon > 0$,

$$\lim_{K\to\infty} \lim_{n\to\infty} \mathbb{P}\left(\|T_K(\mathbf{Z}_n(\lambda))\|_U > \varepsilon\right) = \lim_{K\to\infty} \lim_{n\to\infty} \mathbb{P}\left(\|T_K(\mathbf{Z}_n^o(\lambda))\|_U > \varepsilon\right) = 0,$$

and the proof is now complete. $\qquad\square$

### 3.10.3    Comparison between the dynamic construction and the modified process

Suppose that, at time $\lambda_\star$, we have colored the components $(\mathscr{C}_{(i)}(\lambda_\star))_{i \in [M]}$ blue, say, and then let Algorithms 3.3 and 3.4 evolve. Additionally, we color all the components blue that get connected to one of the blue components during the evolution. Let $\mathcal{C}_M(\lambda)$, $\bar{\mathcal{C}}_M(\lambda)$ denote the union of all such blue components in $\mathcal{G}_n(t_n(\lambda))$ and $\bar{\mathcal{G}}_n(t_n(\lambda))$. In this section, we show that (i) no bad edges are created that are surplus edge of some component, (ii) $|\bar{\mathcal{C}}_M(\lambda)| - |\mathcal{C}_M(\lambda)|$ is asymptotically negligible, (iii) no bad edge is created between the large components, and (iv) with sufficiently large probability, the largest components of $\bar{\mathcal{G}}_n(t_n(\lambda))$ are contained within $\bar{\mathcal{C}}_M(\lambda)$, where $M$ is large. These facts together ensure that the scaling limit for the largest connected components and surplus edges of $\mathcal{G}_n(t_n(\lambda))$ and $\bar{\mathcal{G}}_n(t_n(\lambda))$ are identical. Consider the coupled evolution of Algorithms 3.3 and 3.4. Thus, in the modified set-up, more components get merged due to the creation of bad-edges. Denote $\mathcal{B}_M(\lambda) = |\bar{\mathcal{C}}_M(\lambda)| - |\mathcal{C}_M(\lambda)|$ and $B_{SP}(\lambda)$ the number of bad-edges that are created as surplus edge of some component.

**Lemma 3.10.5.** *For any* $\lambda \geqslant \lambda_\star$, $B_{SP}(\lambda) \xrightarrow{\mathbb{P}} 0$ *and for all* $M \geqslant 1$, $b_n^{-1} \mathcal{B}_M(\lambda) \xrightarrow{\mathbb{P}} 0$.

*Proof.* Before going into the proof, recall Algorithm 3.4, and all the definitions. A bad edge is created if, during some event time of $\bar{\Xi}_n$, a half-edge from $\mathscr{O}$ is selected that was already selected before. Now, for some given pair $(e_0, f_0)$, $e_0 \neq f_0$, the number of ways in which one can choose a pair $(e, f)$, $e \neq f$ such that $e = e_0$, or $f = f_0$, is given by $2\bar{s}_1 - 3$. Thus, the bad edges are created between times $[t_n(\lambda), t_n(\lambda + d\lambda)]$ at rate $(2(\nu_n - 1)\bar{s}_1 c_n)^2/(2\bar{s}_1 - 3)$. Denote $\mathcal{I}_M = \mathcal{I}_M(\lambda) = \{i : \mathscr{C}_{(i)}(\lambda) \subset \bar{\mathcal{C}}_M(\lambda)\}$. The created bad edge adds an additional mass of $|\bar{\mathscr{C}}_{(i)}(\lambda)|$ to $\bar{\mathcal{C}}_M(\lambda)$ if one end is from $\bar{\mathcal{C}}_M(\lambda)$ (for which there are $\sum_{i \in \mathcal{I}_M} \mathcal{O}_i(\lambda)$ possibilities) and the other half-edge is in $\bar{\mathscr{C}}_{(i)}(\lambda)$. The created bad edge is a surplus edge if both of its endpoints come from the same component. For any semi-martingale $(Y_t)_{t \geqslant 0}$, we write $D(Y)(t)$ and $QV(Y)(t)$, respectively to denote the compensator and the quadratic variation, i.e.,

$$Y_t - D(Y)(t), \quad \text{and} \quad (Y_t - D(Y)(t))^2 - QV(Y)(t)$$

are both martingales. Now, $D(B_{SP}(\lambda)) \geqslant 0$, $D(b_n^{-1}\mathcal{B}_1(\lambda)) \geqslant 0$, and for some

constants $C_1, C_2 > 0$

$$D(B_{SP})(\lambda) = \int_{\lambda_\star}^\lambda \frac{2\bar{s}_1 - 3}{4(v_n - 1)^2 \bar{s}_1^2 c_n^2} \sum_{i \geqslant 1} \binom{\bar{\mathcal{O}}_i(\lambda')}{2} d\lambda'$$

$$\leqslant \frac{C_1 n}{b_n^2} \int_{\lambda_\star}^\lambda \|\bar{\mathbf{O}}_n(\lambda')\|_2^2 d\lambda' + o_{\mathbb{P}}(1)$$

$$\leqslant \frac{C_1 n}{b_n^2} (\lambda^\star - \lambda_\star) \|\bar{\mathbf{O}}_n(\lambda^\star)\|_2^2 + o_{\mathbb{P}}(1),$$

$$D(b_n^{-1}\mathcal{B}_1)(\lambda) \leqslant b_n^{-1} \int_{\lambda_\star}^\lambda \frac{2\bar{s}_1 - 3}{4(v_n - 1)^2 \bar{s}_1^2 c_n^2} \sum_{i \in \mathcal{I}_M(\lambda)} \bar{\mathcal{O}}_i(\lambda') \sum_{i \geqslant 1} \bar{\mathcal{O}}_i(\lambda') |\bar{\mathscr{C}}_{(i)}(\lambda')| d\lambda'$$

$$\leqslant \frac{C_2 n}{b_n^2} \int_{\lambda_\star}^\lambda \left( b_n^{-1} \sum_{i=1}^M \bar{\mathcal{O}}_{(i)}(\lambda') \right) \|\bar{\mathbf{O}}_n(\lambda')\|_2 \|\bar{\mathbf{C}}_n(\lambda')\|_2 d\lambda' + o_{\mathbb{P}}(1)$$

$$\leqslant \frac{C_2 n}{b_n^2} (\lambda^\star - \lambda_\star) \left( b_n^{-1} \sum_{i=1}^M \bar{\mathcal{O}}_{(i)}(\lambda^\star) \right) \|\bar{\mathbf{O}}_n(\lambda^\star)\|_2 \|\bar{\mathbf{C}}_n(\lambda^\star)\|_2 + o_{\mathbb{P}}(1),$$

where $\bar{\mathcal{O}}_{(i)}$ denotes the $i^{\text{th}}$ largest value of $(\bar{\mathcal{O}}_i)_{i \geqslant 1}$. Further,

$$QV(B_{SP})(\lambda) \leqslant \frac{C_1 n}{b_n^2} (\lambda^\star - \lambda_\star) \|\bar{\mathbf{O}}_n(\lambda^\star)\|_2^2 + o_{\mathbb{P}}(1),$$

and

$$QV(b_n^{-1}\mathcal{B}_1)(\lambda) \leqslant b_n^{-2} \int_{\lambda_\star}^\lambda \frac{2\bar{s}_1 - 3}{4(v_n - 1)^2 \bar{s}_1^2 c_n^2} \sum_{i \in \mathcal{I}_M(\lambda)} \bar{\mathcal{O}}_i(\lambda') \sum_{i \geqslant 1} \bar{\mathcal{O}}_i(\lambda') |\bar{\mathscr{C}}_{(i)}(\lambda')|^2 d\lambda'$$

$$\leqslant \frac{C_2 n}{b_n^2} \int_{\lambda_\star}^\lambda \left( b_n^{-1} \sum_{i=1}^M \bar{\mathcal{O}}_{(i)}(\lambda') \right) \left( b_n^{-1} |\bar{\mathscr{C}}_{(1)}(\lambda')| \right) \|\bar{\mathbf{O}}_n(\lambda')\|_2 \|\bar{\mathbf{C}}_n(\lambda')\|_2 d\lambda' + o_{\mathbb{P}}(1)$$

$$\leqslant \frac{C_2 n}{b_n^2} (\lambda^\star - \lambda_\star) \left( b_n^{-1} \sum_{i=1}^M \bar{\mathcal{O}}_{(i)}(\lambda^\star) \right) \left( b_n^{-1} |\bar{\mathscr{C}}_{(1)}(\lambda^\star)| \right) \|\bar{\mathbf{O}}_n(\lambda^\star)\|_2 \|\bar{\mathbf{C}}_n(\lambda^\star)\|_2 + o_{\mathbb{P}}(1).$$

Recall that using Lemma 3.10.3, an application of Theorem 3.10.2 yields $(\bar{\mathbf{Z}}_n(\lambda))_{n \geqslant 1}$ is tight in $\mathbb{U}_\downarrow^0$. The proof now follows using the fact that $n/b_n^2 \to 0$. $\qquad\square$

Suppose that a bad edge is being created at time $\lambda'$. Now, this bad edge may be created by choosing the open half-edges from $\mathscr{C}_{(i)}(\lambda')$ and $\mathscr{C}_{(j)}(\lambda')$ for $1 \leqslant i, j \leqslant M$. For fixed $M$, let $F_M(\lambda)$ denote the number of such bad-edges created upto time $\lambda$. Using an argument identical to Lemma 3.10.5 one can show the following:

**Lemma 3.10.6.** *For any* $\lambda \geqslant \lambda_\star$ *and* $M \geqslant 1$, $F_M(\lambda) \xrightarrow{\mathbb{P}} 0$.

The following is the last ingredient that will be needed in the proof:

**Lemma 3.10.7.** *Fix any* $\lambda \in [\lambda_\star, \lambda^\star]$. *For any* $\varepsilon > 0$, *and* $K \geqslant 1$, *there exists* $M = M(\varepsilon, K)$ *such that*

$$\limsup_{n \to \infty} \mathbb{P}\left(\bar{\mathscr{C}}_{(1)}(\lambda), \dots, \bar{\mathscr{C}}_{(K)}(\lambda) \text{ are not contained in } \bar{\mathbb{C}}_M(\lambda)\right) \leqslant \varepsilon.$$

*Proof.* Let $\mathcal{I}_M := \{i : \bar{\mathscr{C}}_{(i)}(\lambda) \subset \bar{\mathbb{C}}_M(\lambda)\}$. It is enough to show that, for any $\varepsilon > 0$, there exists $M$ such that

$$\limsup_{n \to \infty} \mathbb{P}\left(\sum_{i \notin \mathcal{I}_M} |\bar{\mathscr{C}}_{(i)}(\lambda)|^2 > \varepsilon b_n^2\right) \leqslant \varepsilon.$$

For any $M \geqslant 1$, consider the merging dynamics of Algorithm 3.4, where at time $\lambda_\star$, all the components $(\bar{\mathscr{C}}_{(i)}(\lambda_\star))_{i \in [M]}$ are removed. We refer to the above evolution as M-truncated system. We augment a previously defined notation with a superscript $> M$ to denote the corresponding quantity for the M-truncated system. We assume that the M-truncated system and the modified system are coupled in a natural way that at each event time of the modified truncated system, an edge is created in the M-truncated system if both the half-edges are selected from the outside of $\cup_{i=1}^{M} \bar{\mathscr{C}}_{(i)}(\lambda_\star)$. Under this coupling,

$$\sum_{i \notin \mathcal{I}_M} |\bar{\mathscr{C}}_{(i)}(\lambda)|^2 \leqslant \sum_{i \geqslant 1} |\bar{\mathscr{C}}_{(i)}^{>M}(\lambda)|^2.$$

Now, using Lemma 3.10.3, an application of Theorem 3.10.2 yields that $(\bar{\mathbf{Z}}_n(\lambda))_{n \geqslant 1}$ is tight in $\mathbb{U}_\downarrow^0$. Thus the proof now follows. $\qquad\square$

### 3.10.4   Proof of Theorem 3.2.5

We now have all the ingredients to complete the proof of Theorem 3.2.5. For simplicity in writing, we only give a proof for the case $k = 2$ since the proof for general $k$ is identical. Take $\lambda_\star = \lambda_1$. Using Lemma 3.10.3, Theorem 3.10.2 implies

$$(\bar{\mathbf{Z}}_n(\lambda_1), \bar{\mathbf{Z}}_n(\lambda_2)) \xrightarrow{d} (\bar{\mathbf{Z}}(\lambda_1), \bar{\mathbf{Z}}(\lambda_1, \lambda_2)), \tag{3.80}$$

for some random elements $\mathbf{Z}(\lambda_1), \mathbf{Z}(\lambda_1, \lambda_2)$ of $\mathbb{U}_\downarrow^0$. Now, $\bar{\mathbf{Z}}_n(\lambda_1) = \mathbf{Z}_n(\lambda_1)$. Moreover, using Lemmas 3.10.5, 3.10.6, and 3.10.7 and the facts that both $(\bar{\mathbf{Z}}_n(\lambda_2))_{n \geqslant 1}$ and $(\mathbf{Z}_n(\lambda_2))_{n \geqslant 1}$ converge, it follows that (see the argument in Corollary 3.10.4)

$$d_{\mathbb{U}}(\bar{\mathbf{Z}}_n(\lambda_2), \mathbf{Z}_n(\lambda_2)) \xrightarrow{\mathbb{P}} 0.$$

Thus, $(\mathbf{Z}_n(\lambda_1), \mathbf{Z}_n(\lambda_2))$ converge jointly. Moreover, the limiting object $\mathbf{Z}(\lambda_1, \lambda_2)$ appearing in (3.80) does not depend on $\lambda_1$ by Theorem 3.2.2. Now, using induction, there exists a version of the augmented multiplicative coalescent $\mathbf{AMC} = (\mathrm{AMC}(\lambda))_{\lambda \in \mathbb{R}}$ such that for any $k \geqslant 1$

$$(\mathbf{Z}_n(\lambda_1), \ldots, \mathbf{Z}_n(\lambda_k)) \xrightarrow{\mathrm{d}} (\mathrm{AMC}(\lambda_1), \ldots, \mathrm{AMC}(\lambda_k)).$$

Finally, the proof of Theorem 3.2.2 is completed by using Proposition 3.10.1.
□

## 3.11    Conclusion

In this chapter, we have shown that, when the third moment of the empirical degree distribution tends to infinity, the critical window for the configuration model is primarily dictated by the vertices of highest degree or hubs. In fact, the asymptotics of hubs completely specify the scaling limits for the component sizes and the surplus edges. The proof techniques in this case is completely different than in Chapter 2. Since the increment distribution of the exploration process has infinite third moment, the analysis does not fall under the framework of invariance principles such as Martingale FCLT. Moreover, since the limiting process does not have independent increments, general methods for stochastic process convergence are not applicable. The core of the analysis rests on the fact that the hubs cause jumps in the exploration process, and the contribution due to the low-degree vertices turns out to be asymptotically negligible. The proof for $\ell^2_\downarrow$ convergence is also more direct in this case, as compared to the size-biased point processes in Chapter 2.4.6. The fact that the graph becomes more subcritical only after the removal of the hubs plays a crucial role in the analysis. As in Chapter 2, we use Janson's construction to study critical percolation. Further, the evolution over the critical window is studied for both the component sizes and the surplus edges. In fact, the scaling limit in Theorem 3.2.5 establishes that there exists a version of augmented multiplicative coalescent with finite-dimensional distributions being described by excursions of a thinned Lévy process.

## 3.A    Appendix: Path counting

In this section, we derive a generalization of [111, Lemma 5.1] by extending their argument. Let $V'_n$ denote the vertex chosen according to the distribution

$F_n$ on $[n]$, independently of the graph. Also, let $D'_n$ denote the degree of $V'_n$, $D_n$ denote the degree of a uniformly chosen vertex (independently of the graph) and $\mathscr{C}(v)$ denote the connected component containing $v$.

**Lemma 3.A.1.** *Let $w = (w_i)_{i\in[n]}$ be a weight sequence and consider $\mathrm{CM}_n(\mathbf{d})$ such that $\nu_n < 1$. Then,*

$$\mathbb{E}\left[\sum_{i\in\mathscr{C}(V'_n)} w_i\right] \leqslant \mathbb{E}\left[w_{V'_n}\right] + \frac{\mathbb{E}\left[D'_n\right]\mathbb{E}\left[D_n w_{V_n}\right]}{\mathbb{E}\left[D_n\right](1-\nu_n)}.$$

*Proof.* Consider all possible paths of length $l$ starting from $V'_n$ and the $w$-value at the end of those paths. If we sum over all such paths together with a sum over all possible $l$, then we obtain an upper bound on $\sum_{i\in\mathscr{C}(V'_n)} w_i$. Write $\mathbb{E}_v[\cdot]$ for the expectation conditional on $V'_n = v$. Thus,

$$\mathbb{E}_v\left[\sum_{i\in\mathscr{C}(V'_n)} w_i\right] \leqslant w_v + d_v \sum_{\substack{l\geqslant 1}} \sum_{\substack{x_1,\dots,x_l \\ x_i\neq x_j, \forall i\neq j}} \frac{\prod_{i=1}^{l-1} d_{x_i}(d_{x_i}-1)d_{x_l}w_{x_l}}{(\ell_n-1)\dots(\ell_n-2l+1)}.$$

Now, using the exactly same arguments as [111, Lemma 5.1], it follows that

$$\mathbb{E}\left[\sum_{i\in\mathscr{C}(V'_n)} w_i\right] \leqslant \mathbb{E}\left[w_{V'_n}\right] + \frac{\mathbb{E}\left[D'_n\right]\mathbb{E}\left[D_n w_{V_n}\right]}{\mathbb{E}\left[D_n\right]} \sum_{l\geqslant 1} \nu_n^{l-1},$$

and this completes the proof.  □

## 3.B   Appendix: Proof of Lemma 3.6.3

The proof is an adaptation of the proof of [70, Lemma 20]. Let $V'_n$ denote the vertex chosen according to the distribution $F_n$ on $[n]$, independently of the graph and let $D'_n$ denote the degree of $V'_n$. Suppose that $\limsup_{n\to\infty} \mathbb{E}[D'_n] < \infty$. We use a generic constant $C$ to denote a positive constant independent of $n, \delta, K$. Consider the graph exploration described in Algorithm 3.1, but now we start by choosing vertex $V'_n$ at Stage 0 and declaring all its half-edges active. The exploration process is still given by (3.21) with $S_n(0) = D'_n$. Note that $\mathscr{C}(V'_n)$ is explored when $\mathbf{S}_n$ hits zero. For $H > 0$, let

$$\gamma := \inf\{l \geqslant 1 : S_n(l) \geqslant H \text{ or } S_n(l) = 0\} \wedge 2\delta_K b_n.$$

Note that

$$
\mathbb{E}\left[S_n(l+1) - S_n(l) \mid (\mathcal{J}_i^n(l))_{i=1}^n\right] = \sum_{i \in [n]} d_i \mathbb{P}\left(i \notin \mathcal{V}_l, i \in \mathcal{V}_{l+1} \mid (\mathcal{J}_i^n(l))_{i=1}^n\right) - 2
$$

$$
= \frac{\sum_{i \notin \mathcal{V}_l} d_i^2}{\ell_n - 2l - 1} - 2 \leqslant \frac{\sum_{i \in [n]} d_i^2}{\ell_n - 2l - 1} - 2
$$

$$
:= \lambda c_n^{-1} + o(c_n^{-1}) + \frac{2l+1}{\ell_n - 2l - 1} \times \frac{\sum_{i \in [n]} d_i^2}{\ell_n} \leqslant 0
$$

uniformly over $l \leqslant 2\delta_K b_n$ for all small $\delta > 0$ and large $n$, where the last step follows from the fact that $\lambda < 0$. Therefore, $\{S_n(l)\}_{l=1}^{2\delta_K b_n}$ is a super-martingale. The optional stopping theorem now implies

$$
\mathbb{E}\left[D_n'\right] \geqslant \mathbb{E}\left[S_n(\gamma)\right] \geqslant H\mathbb{P}\left(S_n(\gamma) \geqslant H\right).
$$

Thus,

$$
\mathbb{P}\left(S_n(\gamma) \geqslant H\right) \leqslant \frac{\mathbb{E}\left[D_n'\right]}{H}.
$$

Put $H = a_n K^{1.1}/\sqrt{\delta}$. To simplify the writing, we write $S_n[0, t] \in A$ to denote that $S_n(l) \in A$, for all $l \in [0, t]$. Notice that

$$
\mathbb{P}\left(\mathrm{SP}(\mathscr{C}(V_n')) \geqslant K, |\mathscr{C}(V_n')| \in (\delta_K b_n, 2\delta_K b_n)\right)
$$
$$
\leqslant \mathbb{P}\left(S_n(\gamma) \geqslant H\right) + \mathbb{P}\left(\mathrm{SP}(\mathscr{C}(V_n')) \geqslant K, S_n[0, 2\delta_K b_n] < H, S_n[0, \delta_K b_n] > 0\right).
$$
$$
(3.81)
$$

Define $A(l_1, \ldots, l_K)$ to be the event that $\{$surpluses occur at times $l_1, \ldots, l_K$, and $S_n[0, 2\delta_K b_n] < H, S_n[0, \delta_K b_n] > 0\}$. Now,

$$
\mathbb{P}\left(\mathrm{SP}(\mathscr{C}(V_n')) \geqslant K, S_n[0, 2\delta_K b_n] < H, S_n[0, \delta_K b_n] > 0\right)
$$
$$
\leqslant \sum_{1 \leqslant l_1 < \cdots < l_K \leqslant 2\delta_K b_n} \mathbb{P}\left(A(l_1, \ldots, l_K)\right)
$$
$$
= \sum_{1 \leqslant l_1 < \cdots < l_K \leqslant 2\delta_K b_n} \mathbb{E}\left[\mathbf{1}\{0 < S_n[0, l_K - 1] < H, \mathrm{SP}(l_K - 1) = K - 1\}Y\right],
$$

where

$$
Y = \mathbb{P}\left(K^{\text{th}} \text{ surplus occurs at time } l_K, S_n[l_K, 2\delta_K b_n] < H, S_n[l_K, \gamma] > 0 \mid \mathscr{F}_{l_K - 1}\right)
$$
$$
\leqslant \frac{C K^{1.1} a_n}{\ell_n \sqrt{\delta}} \leqslant \frac{C K^{1.1}}{b_n \sqrt{\delta}}.
$$

Therefore, using induction, (3.81) yields

$$\mathbb{P}\left(\text{SP}(\mathscr{C}(V_n')) \geqslant K, S_n[0, 2\delta_K b_n] < H, S_n[0, \delta_K b_n] > 0\right)$$

$$\leqslant C\left(\frac{K^{1.1}}{\sqrt{\delta b_n}}\right)^K \frac{(2\delta b_n)^{K-1}}{K^{0.12(K-1)}(K-1)!} \sum_{l_1=1}^{2\delta_K b_n} \mathbb{P}\left(|\mathscr{C}(V_n')| \geqslant l_1\right)$$

$$\leqslant C\frac{\delta^{K/2}}{K^{1.1}b_n}\mathbb{E}\left[|\mathscr{C}(V_n')|\right],$$

where we have used the fact that $\#\{1 \leqslant l_2, \ldots, l_K \leqslant 2\delta b_n\} = (2\delta b_n)^{K-1}/(K-1)!$ and Stirling's approximation for $(K-1)!$ in the last step. Since $\lambda < 0$, we can use Lemma 3.A.1 to conclude that for all sufficiently large $n$

$$\mathbb{E}\left[|\mathscr{C}(V_n)|\right] \leqslant Cc_n,$$

for some constant $C > 0$ and we get the desired bound for (3.81). The proof of Lemma 3.6.3 is now complete. $\qquad\square$

# Chapter 4

# Metric space limit for critical components in the infinite third-moment regime

**Abstract.** This chapter establishes general universality principles for random network models whose component sizes in the critical regime lie in the multiplicative coalescent universality class with heavy-tailed degrees resulting in hubs. For graphs whose components evolve exactly as a multiplicative coalescent in this regime, scaling limits for the metric structure of maximal components were derived in [36]. In this chapter, we derive sufficient uniform asymptotic negligibility conditions for general network models to satisfy in the barely subcritical regime such that, if the evolution of the components can be *approximated* by a multiplicative coalescent as one transitions from the barely subcritical regime through the critical regime, then the maximal components belong to the same universality class as in [36]. As a canonical example, we study critical percolation on configuration models with heavy-tailed degrees. Of independent interest, we derive refined asymptotics for various susceptibility functions and the maximal diameter in the barely subcritical regime. These estimates, coupled with the universality result, allow us to derive the asymptotic metric structure of the large components through the critical scaling window for percolation.

The aim of this chapter is to understand universality principles for the metric structure of the critical components when the degree distribution satisfies an infinite third-moment condition. To describe our results, we start with an analogy. In classical limit theorems for sums of independent random variables, there are two major steps: (1) Identifying possible limit laws (e.g. normal distribution, stable laws, etc.), and (2) understanding uniform asymptotic negligibility conditions under which sums of random variables (appropriately re-scaled) converge to the appropriate limit. In the context of critical random graphs with degree-exponent $\tau \in (3, 4)$, candidate limit law of maximal components with each edge rescaled to have length $n^{-(\tau-3)/(\tau-1)}$ was established in [36]. In this chapter we establish sufficient uniform asymptotic negligibility (UAN) conditions for a random graph model in the barely subcritical regime which, coupled with appropriate merging dynamics of components as one increases edge density through the critical regime, implies convergence to limits established in [36]. This is described in Theorem 4.4.2. As a canonical example, we analyze the critical regime for percolation on the uniform random graph model and the configuration model with a prescribed heavy-tailed degree distribution (see Theorems 4.1.1 and 4.1.2). Of independent interest, we obtain refined estimates for various susceptibility functionals and bounds on the diameter of largest connected components in the barely-subcritical regime for the configuration model; these are described in Theorems 4.1.3 and 4.1.4.

**Organization of the chapter.**    Section 4.1 describes the canonical random graph model motivating this work and describes associated results. A full description of the limit objects and notions of convergence of metric space valued random variables are deferred to Section 4.3. Section 4.2 has a detailed discussion of related work and relevance of this work. Section 4.4 describes and proves the general universality result. Sections 4.4.1 and 4.5 prove results about the configuration model.

## 4.1    Main results

Owing to technical overhead, the statement of our main universality result is deferred to Section 4.4. In this section, we present the results about the largest connected components obtained via percolation on the uniform random graph model and the configuration model. We defer definitions of the

limiting objects as well as notions of convergence of measured metric spaces to Section 4.3.

### 4.1.1    Critical percolation on the configuration model: the metric structure

For $p > 0$, define metric space $\ell_\downarrow^p = \{(x_i)_{i \geqslant 1} : x_i > x_{i+1}, \sum_i x_i^p < \infty\}$, with the metric $d(x, y) = (\sum_i |x_i - y_i|^p)^{1/p}$. Fix $\tau \in (3, 4)$. Throughout this chapter we set:

$$\alpha = 1/(\tau - 1), \qquad \rho = (\tau - 2)/(\tau - 1), \qquad \eta = (\tau - 3)/(\tau - 1).$$

**Assumption 4.1** (Degree sequence). For each $n \geqslant 1$, let $\mathbf{d} = \mathbf{d}_n = (d_i)_{i \in [n]}$ be a degree sequence. We assume the following about $(\mathbf{d}_n)_{n \geqslant 1}$ as $n \to \infty$:

(i) (*High degree vertices*) For $i \geqslant 1$, $n^{-\alpha} d_i \to \theta_i$, where $\theta = (\theta_i)_{i \geqslant 1} \in \ell_\downarrow^3 \setminus \ell_\downarrow^2$.

(ii) (*Moment assumptions*) Let $D_n$ denote the degree of a vertex chosen uniformly at random, independently of $CM_n(\mathbf{d})$. Then, $D_n$ converges in distribution to some discrete random variable $D$, and

$$\frac{1}{n} \sum_{i \in [n]} d_i \to \mu := \mathbb{E}[D], \quad \frac{1}{n} \sum_{i \in [n]} d_i^2 \to \mu_2 := \mathbb{E}[D^2],$$

$$\lim_{K \to \infty} \limsup_{n \to \infty} n^{-3\alpha} \sum_{i=K+1}^{n} d_i^3 = 0.$$

As discussed in Chapter 1, the component sizes of $CM_n(\mathbf{d})$ undergo a phase transition [114, 133] depending on the parameter

$$\nu_n = \frac{\sum_{i \in [n]} d_i(d_i - 1)}{\sum_{i \in [n]} d_i} \to \nu = \frac{\mathbb{E}[D(D-1)]}{\mathbb{E}[D]}.$$

Precisely, when $\nu > 1$, $CM_n(\mathbf{d})$ is super-critical in the sense that there exists a unique *giant* component whp, and when $\nu < 1$, all the components have size $o_\mathbb{P}(n)$. In this chapter, we will always assume that

$$\nu > 1, \quad \text{i.e.} \quad CM_n(\mathbf{d}) \text{ is supercritical.} \tag{4.1}$$

The focus of this chapter is to study *critical percolation* on $CM_n(\mathbf{d})$. Percolation refers to deleting each edge of a graph independently with probability $1 - p$. Let $CM_n(\mathbf{d}, p_n)$, and $UM_n(\mathbf{d}, p_n)$ denote the graphs obtained from

percolation with probability $p_n$ on graphs $CM_n(\mathbf{d})$ and $UM_n(\mathbf{d})$, respectively. For $p_n \to p$, it was shown in [109] that the critical point for the phase transition of the component sizes is $p = 1/\nu$. The critical window for percolation was studied in Chapters 2 and 3 to obtain the asymptotics of the largest component sizes and corresponding surplus edges. In this chapter, we will assume that $CM_n(\mathbf{d}, p_c(\lambda))2$ is in the critical window, i.e.,

$$p_n = p_n(\lambda) = \frac{1}{\nu_n} + \frac{\lambda}{n^\eta} + o(n^{-\eta}). \tag{4.2}$$

Let $\mathscr{C}^p_{(i)}(\lambda)$ denote the $i$-th largest component of $CM(\mathbf{d}, p_n(\lambda))$. Each component $\mathscr{C}$ can be viewed as a measured metric space with (i) the metric being the graph distance where each edge has length one, (ii) the measure being proportional to the counting measure, i.e., for any $A \subset \mathscr{C}$, the measure of $A$ is given by $\mu_{ct,i}(A) = |A|/|\mathscr{C}|$. For a generic measured metric space $M = (M, d, \mu)$ and $a > 0$, write $aM$ to denote the measured metric space $(M, ad, \mu)$. Write $\mathscr{S}_*$ for the space of all measured metric spaces equipped with the Gromov weak topology (see Section 4.3.1) and let $\mathscr{S}_*^{\mathbb{N}}$ denote the corresponding product space with the accompanying product topology. For each $n \geqslant 1$, view $\left(n^{-\eta}\mathscr{C}^p_{(i)}(\lambda)\right)_{i \geqslant 1}$ as an object in $\mathscr{S}_*^{\mathbb{N}}$ by appending an infinite sequence of empty metric spaces after enumerating the components in $CM_n(\mathbf{d}, p_n(\lambda))$. The main results for the configuration model are as follows:

**Theorem 4.1.1.** *Consider* $CM_n(\mathbf{d}, p_n(\lambda))$ *satisfying Assumption 4.1, (4.1) and (4.2) for some* $\lambda \in \mathbb{R}$. *There exists a sequence of random measured metric spaces* $(\mathscr{M}_i(\lambda))_{i \geqslant 1}$ *such that on* $\mathscr{S}_*^{\mathbb{N}}$ *as* $n \to \infty$

$$\left(n^{-\eta}\mathscr{C}^p_{(i)}(\lambda)\right)_{i \geqslant 1} \xrightarrow{d} \left(\mathscr{M}_i(\lambda)\right)_{i \geqslant 1}. \tag{4.3}$$

**Theorem 4.1.2.** *Under Assumption 4.1, (4.1) and (4.2) for some* $\lambda \in \mathbb{R}$, *the convergence in (4.3) also holds for the components of* $UM_n(\mathbf{d}, p_n(\lambda))$, *with the identical limiting object.*

**Remark 29.** The limiting objects are precisely described in Section 4.3.5. The conclusion of Theorem 4.1.1 holds if the measure $\mu_{ct,i}$ on $\mathscr{C}^p_{(i)}$ is replaced by more general measures, see Remark 37.

## 4.1.2   Mesoscopic properties of the critical clusters: barely subcritical regime

One of the main ingredients in the proof of Theorem 4.1.1 is a refined analysis of various susceptibility functions in the *barely subcritical* regime which are

of independent interest. In this context we prove general statements about the susceptibility functions applicable not just to percolation on the supercritical configuration model, rather to any barely subcritical configuration model. Since percolation on a configuration model yields a configuration model [85, 109], the above yields susceptibility functions for percolation on a configuration model as a special case.

**Assumption 4.2** (Barely subcritical degree sequence). Let $\mathbf{d}' = (d_1', \dots, d_n')$ be a degree sequence and let $w_{(\cdot)} : [n] \mapsto \mathbb{R}_+$ be such that

(i)  Assumption 4.1 holds for $\mathbf{d}'$ with some $\mathbf{c} \in \ell_\downarrow^3 \setminus \ell_\downarrow^2$, and

$$\lim_{n \to \infty} \frac{1}{n} \sum_{i \in [n]} d_i' = \mu_d, \quad \lim_{n \to \infty} \frac{1}{n} \sum_{i \in [n]} w_i = \mu_w,$$

$$\lim_{n \to \infty} \frac{1}{n} \sum_{i \in [n]} d_i' w_i = \mu_{d,w}.$$

(ii)  $\max\{\sum_{i \in [n]} w_i^3, \sum_{i \in [n]} d_i'^2 w_i, \sum_{i \in [n]} d_i' w_i^2\} = O(n^{3\alpha})$.

(iii)  (*Barely subcritical regime*:) There exists $0 < \delta < \eta$ and $\lambda_0 > 0$ such that

$$\nu_n' = \frac{\sum_{i \in [n]} d_i'(d_i' - 1)}{\sum_{i \in [n]} d_i'} = 1 - \lambda_0 n^{-\delta} + o(n^{-\delta}). \tag{4.4}$$

We will consider a configuration model, where vertex $i$ has degree $d_i'$ and weight $w_i$. Let $\mathscr{C}_{(i)}'$ denote the $i$-th largest component of $\mathrm{CM}_n(\mathbf{d}')$, $\mathscr{W}_{(i)} = \sum_{k \in \mathscr{C}_{(i)}'} w_k$ and define the weight-based susceptibility functions as

$$s_r^\star = \frac{1}{n} \sum_{i \geqslant 1} \mathscr{W}_{(i)}^r \; ; \; r \geqslant 1, \qquad s_{pr}^\star = \frac{1}{n} \sum_{i \geqslant 1} \mathscr{W}_{(i)} |\mathscr{C}_{(i)}'|. \tag{4.5}$$

Also, define the weighted distance-based susceptibility as

$$\mathcal{D}_n^\star = \frac{1}{n} \sum_{i,j \in [n]} w_i w_j d(i,j) \mathbf{1}\{i, j \text{ are in the same connected component}\}, \tag{4.6}$$

where $d$ denotes the graph distance in the component $\mathscr{C}_{(k)}'$ for which $i, j \in \mathscr{C}_{(k)}'$. The goal is to show that the quantities defined in (4.5) and (4.6) satisfy asymptotic regularity conditions. These are summarized in the following theorem:

**Theorem 4.1.3** (Susceptibility functions)**.** *Under Assumption 4.2, as* $n \to \infty$,

$$n^{-\delta} s_2^{\star} \xrightarrow{\mathbb{P}} \frac{\mu_{d,w}^2}{\mu_d \lambda_0}, \quad n^{-\delta} s_{pr}^{\star} \xrightarrow{\mathbb{P}} \frac{\mu_{d,w}}{\lambda_0}, \quad n^{-(\alpha+\delta)} \mathscr{W}_{(j)} \xrightarrow{\mathbb{P}} \frac{\mu_{d,w}}{\mu_d \lambda_0} c_j,$$

$$n^{-3\alpha-3\delta+1} s_3^{\star} \xrightarrow{\mathbb{P}} \left( \frac{\mu_{d,w}}{\mu_d \lambda_0} \right)^3 \sum_{i=1}^{\infty} c_i^3, \quad n^{-2\delta} \mathcal{D}_n^{\star} \xrightarrow{\mathbb{P}} \frac{\mu_{d,w}^2}{\mu_d \lambda_0^2}.$$

For a connected graph $G$, $\Delta(G)$ denotes the diameter of the graph, and for any arbitrary graph $G$, $\Delta_{max}(G) := \max \Delta(\mathscr{C})$, where the maximum is taken over all connected components $\mathscr{C} \subset G$. We simply write $\Delta_{max}$ for $\Delta_{max}(CM_n(\mathbf{d}'))$.

**Theorem 4.1.4** (Maximum diameter)**.** *Under Assumption 4.2, as* $n \to \infty$,

$$\mathbb{P}(\Delta_{max} > 6n^{\delta} \log(n)) \to 0.$$

**Remark 30.** Note that $w_i = 1$ for all $i \in [n]$ implies that $\mathscr{W}_{(i)} = |\mathscr{C}_{(i)}|$, and thus Theorem 4.1.3 holds for the usual susceptibility functions defined in terms of the component sizes (cf. [111]). In the proof of Theorem 4.1.1, we will require a more general weight function.

## 4.2  Discussion

In this section, we describe related work and discuss the relevance of the results in this chapter.

**Related work.**   A wide array of universality conjectures have been postulated about functionals of network models. Of particular relevance here is [51, 52], where via simulations the so-called strong disorder regime (which in the extremal case is the minimal spanning tree where edges have i.i.d. positive random edge weights) was studied. From the probabilisitic combinatorics community both for the universality principle and the results for the configuration model, a major role is played by the *multiplicative coalescent*. The process was rigorously constructed in [8] whilst a complete description of the entrance boundary of this Markov process was laid out in [11]. In the context of the critical regime for random graphs especially with heavy tails, *component sizes* for the closely related rank-one random graph model were derived in [33, 97]. These were then extended to the configuration model in [121] culminating in a complete description of component sizes and surplus edges in [69]. Rigorous results for the metric-space structure of components

in the heavy-tailed regime was first derived in [36]; the limiting objects are described in Section 4.3.5. In this chapter, we develop general conditions under which the metric structure of the critical components are identical to those for critical rank-one inhomogeneous random graphs as derived in [36], and hence derive Theorem 4.1.1. Similar universality principles for $\tau > 4$ were derived in [29].

**Proof techniques in the barely subcritical regime.**    A key part of the contributions of this chapter is a refined analysis of the barely subcritical regime for $\mathrm{CM}_n(\mathbf{d})$ in the heavy-tailed regime; for related results see e.g. [108, 111, 113, 116] and the references therein. The bounds in this chapter, in particular the extension to the barely subcritical regime are new. The proof techniques are also novel, and involve a combination of generalizing path-counting techniques [111], formalizing branching process heuristics, as well as leveraging the differential equation method [152] to analyze various susceptibility functions in the barely subcritical regime.

**Remark 31.** In an ongoing work, [65] derives the scaling limit of the maximal components at criticality for $\mathrm{CM}_n(\mathbf{d})$ when the degrees are i.i.d samples from a power-law distribution with $\tau \in (3, 4)$, and [90] investigates the properties of these limiting objects, obtained via appropriate tilts of Levy trees [76]. Interestingly, the description of the limiting objects in the i.i.d setting turns out to be quite different than Theorem 4.1.1. It will be interesting to explore the connections between the results in the above papers and the current work.

## 4.3    Convergence of metric spaces, discrete structures and limit objects

The aim of this section is to define the proper notion of convergence relevant to this chapter (Section 4.3.1), set-up discrete structures required in the statement and proof of the universality result in Theorem 4.4.2 (Sections 4.3.2, 4.3.3, 4.3.4), and describe limit objects that arise in Theorem 4.1.1 (Sections 4.3.4 and 4.3.5).

### 4.3.1    Gromov-weak topology

A complete separable measured metric space (denoted by $(X, d, \mu)$) is a complete, separable metric space $(X, d)$ with an associated probability measure $\mu$ on the Borel sigma algebra $\mathcal{B}(X)$. The Gromov-weak topology is defined

on $\mathscr{S}_0$, the space of all complete and separable measured metric spaces (see [36, Section 2.1.2], [91, 93]). The notion is formulated based on the philosophy of finite-dimensional convergence. Two measured metric spaces $(X_1, d_1, \mu_1), (X_2, d_2, \mu_2)$ are considered to be equivalent if there exists an isometry $\psi : \text{support}(\mu_1) \mapsto \text{support}(\mu_2)$ such that $\mu_2 = \mu_1 \circ \psi^{-1}$. Let $\mathscr{S}_*$ be the space of all equivalence classes of $\mathscr{S}_0$. We abuse the notation by not distinguishing between a metric space and its corresponding equivalence class. Fix $l \geqslant 2$, $(X, d, \mu) \in \mathscr{S}_*$. Given any collection of points $\mathbf{x} = (x_1, \ldots, x_l) \in X^l$, define $D(\mathbf{x}) := (d(x_i, x_j))_{i,j \in [l]}$ to be the matrix of pairwise distances of the points in $\mathbf{x}$. A function $\Phi : \mathscr{S}_* \mapsto \mathbb{R}$ is called a polynomial if there exists a bounded continuous function $\phi : \mathbb{R}^{l^2} \mapsto \mathbb{R}$ such that

$$\Phi((X, d, \mu)) = \int \phi(D(\mathbf{x})) d\mu^{\otimes l},$$

where $\mu^{\otimes l}$ denotes the $l$-fold product measure. A sequence $\{(X_n, d_n, \mu_n)\}_{n \geqslant 1} \subset \mathscr{S}_*$ is said to converge to $(X, d, \mu) \in \mathscr{S}_*$ if and only if $\Phi((X_n, d_n, \mu_n)) \to \Phi((X, d, \mu))$ for all polynomials $\Phi$ on $\mathscr{S}_*$. By [91, Theorem 1], $\mathscr{S}_*$ is a Polish space under the Gromov-weak topology.

### 4.3.2    Super graphs

Our *super graphs* consist of three main ingredients: 1) a collection of metric spaces called *blobs*, 2) a graphical super-structure determining the connections between the blobs, 3) connection points or junction points at each blob. In more detail, super graphs contain the following structures:

(a) **Blobs**: A collection $\{(M_i, d_i, \mu_i)\}_{i \in [m]}$ of connected, compact measured metric spaces.

(b) **Superstructure**: A (random) graph $\mathcal{G}$ with vertex set $[m]$. The graph has a weight sequence $\mathbf{p} = (p_i)_{i \in [m]}$ associated to the vertex set $[m]$. We regard $M_i$ as the $i$-th vertex of $\mathcal{G}$.

(c) **Junction points**: An independent collection of random points $\mathbf{X} := (X_{i,j} : i, j \in [m])$ such that $X_{i,j} \sim \mu_i$ for all $i, j$. Further, $\mathbf{X}$ is independent of $\mathcal{G}$.

Using these three ingredients, define a metric space $(\bar{M}, \bar{d}, \bar{\mu}) = \Gamma(\mathcal{G}, \mathbf{p}, \mathbf{M}, \mathbf{X})$, with $\bar{M} = \sqcup_{i \in [m]} M_i$, by putting an edge of *length* one between the pair of points $\{(X_{i,j}, X_{j,i}) : (i, j)$ is an edge of $\mathcal{G}\}$. The distance metric $\bar{d}$ is the natural metric obtained from the graph distance and the inter-blob distance on a

path. More precisely, for any $x, y \in \bar{M}$ with $x \in M_{j_1}$ and $y \in M_{j_2}$,

$$\bar{d}(x,y) = \inf \left\{ k + d_{j_1}(x, X_{j_1, i_1}) + \sum_{l=1}^{k-1} d_{i_l}(X_{i_l, i_{l-1}}, X_{i_{l+1}, i_l}) + d_{j_2}(X_{j_2, i_{k-1}}, y) \right\},$$

(4.7)

where the infimum is taken over all paths $(i_1, \ldots, i_{k-1})$ in $\mathcal{G}$ and all $k \geqslant 1$ and we interpret $i_0$ and $i_k$ as $j_1$ and $j_2$ respectively. The measure $\bar{\mu}$ is given by $\bar{\mu}(A) := \sum_{i \in [m]} p_i \mu_i (A \cap M_i)$, for any measurable subset $A$ of $\bar{M}$. Note that there is a one-to-one correspondence between the components of $\mathcal{G}$ and $\Gamma(\mathcal{G}, \mathbf{p}, \mathbf{M}, \mathbf{X})$ as the blobs are connected.

### 4.3.3   Space of trees with edge lengths, leaf weights, root-to-leaf measures, and blobs

In the proof of the main results we need the following spaces built on top of the space of discrete trees. The first space $\mathbf{T}_{IJ}$ was formulated in [13, 14] where it was used to study trees spanning a finite number of random points sampled from an inhomogeneous continuum random tree (as described in the next section).

**The space $\mathbf{T}_{IJ}$.**

Fix $I \geqslant 0$ and $J \geqslant 1$. Let $\mathbf{T}_{IJ}$ be the space of trees with each element $\mathbf{t} \in \mathbf{T}_{IJ}$ having the following properties:

1. There are exactly $J$ leaves labeled $1+, \ldots, J+$, and the tree is rooted at the labeled vertex $0+$.

2. There may be extra labeled vertices (called hubs) with labels in $\{1, \ldots, I\}$. (It is possible that only some, and not all, labels in $\{1, \ldots, I\}$ are used.)

3. Every edge $e$ has a strictly positive edge length $l_e$.

A tree $\mathbf{t} \in \mathbf{T}_{IJ}$ can be viewed as being composed of two parts: (1) shape($\mathbf{t}$) describing the shape of the tree (including the labels of leaves and hubs) but ignoring edge lengths. The set of all possible shapes $\mathbf{T}_{IJ}^{\text{shape}}$ is obviously finite for fixed $I, J$. (2) The edge lengths $\mathbf{l}(\mathbf{t}) := (l_e : e \in \mathbf{t})$. We will consider the product topology on $\mathbf{T}_{IJ}$ consisting of the discrete topology on $\mathbf{T}_{IJ}^{\text{shape}}$ and the product topology on $\mathbb{R}^E$, where $E$ is the number of edges of $\mathbf{t}$.

**The space $T^*_{IJ}$.**

Along with the three attributes above in $\mathbf{T}_{IJ}$, the trees in this space have the following two additional properties. Let $\mathcal{L}(\mathbf{t}) := \{1+, \ldots, J+\}$ denote the collection of leaves in $\mathbf{t}$. Then every leaf $v \in \mathcal{L}(\mathbf{t})$ has the following attributes:

(d) **Leaf weights:** A strictly positive number $A(v)$.

(e) **Root-to-leaf measures:** A probability measure $v_{\mathbf{t},v}$ on the path $[0+, v]$ connecting the root and the leaf $v$.

The path $[0+, v]$ for each $v \in \mathcal{L}(\mathbf{t})$, can be viewed as a compact measured metric space with the measure being $v_{\mathbf{t},v}$. Let $\mathcal{X}$ denote the space of compact measured metric spaces endowed with the Gromov-Hausdorff-Prokhorov topology (see [36, Section 2.1.1]). In addition to the topology on $\mathbf{T}_{IJ}$, the space $\mathbf{T}^*_{IJ}$ with the additional two attributes inherits the product topology on $\mathbb{R}^J$ due to leaf weights and $\mathcal{X}^J$ due to the paths $[0+, v]$ endowed with $v_{\mathbf{t},v}$ for each $v \in \mathcal{L}(\mathbf{t})$. For consistency, we add a conventional state $\partial$ to the spaces $\mathbf{T}_{IJ}$ and $\mathbf{T}^*_{IJ}$. Its use will be made clear in Section 4.4.

For all instances in this chapter, the shape of a tree shape$(\mathbf{t})$ will be viewed as a subgraph of a graph with $m$ vertices. In that case, the tree will be assumed to inherit the vertex labels from the original graph. We will often write $\mathbf{t} \in T^{*m}_{IJ}$ to emphasize the fact that the vertices of $\mathbf{t}$ are labeled from a subset of $[m]$.

**The space $\overline{T}^{*m}_{IJ}$.**

We enrich the space $T^{*m}_{IJ}$ with some additional elements to accommodate the blobs. Consider $\mathbf{t} \in T^{*m}_{IJ}$ and construct $\bar{\mathbf{t}}$ as follows: Let $(M_i, d_i, \mu_i)_{i \in [m]}$ be a collection of blobs and $\mathbf{X} = (X_{ij} : i, j \in [m])$ be the collection of junction points as defined in Section 4.3.2. Construct the metric space $\bar{\mathbf{t}}$ with elements in $\bar{M}(\mathbf{t}) = \sqcup_{i \in \mathbf{t}} M_i$, by putting an edge of 'length' one between the pair of vertices $\{(X_{i,j}, X_{j,i}) : (i, j) \text{ is an edge of } \mathbf{t}\}$. The distance metric is given by (4.7). The path from the leaf $v$ to the root $0+$ now contain blobs. Replace the root-to-leaf measure by $\bar{v}_{\mathbf{t},v}(A) := \sum_{i \in [0+,v]} v_{\mathbf{t},v}(i) \mu(M_i \cap A)$ for $A \subset \sqcup_{i \in [0+,v]} M_i$, where $v_{\mathbf{t},v}$ is the root-to-leaf measure on $[0+, v]$ for $\mathbf{t}$. Notice that $T^{*m}_{IJ}$ can be viewed as a subset of $\overline{T}^{*m}_{IJ}$. In the proof of the universality theorem in Section 4.4, the blobs will be a fixed collection and, therefore, any $\mathbf{t} \in T^{*m}_{IJ}$ corresponds to a unique $\bar{\mathbf{t}} \in \overline{T}^{*m}_{IJ}$.

### 4.3.4   p-trees

For fixed $m \geqslant 1$, write $\mathbb{T}_m$ and $\mathbb{T}_m^{\text{ord}}$ for the collection of all rooted trees with vertex set $[m]$ and rooted ordered trees with vertex set $[m]$ respectively. An ordered rooted tree is a rooted tree where children of each individual are assigned an order. We define a random tree model called **p**-trees [59, 142], and their corresponding limits, the so-called inhomogeneous continuum random trees, which play a key role in describing the limiting metric spaces. Fix $m \geqslant 1$, and a probability mass function $\mathbf{p} = (p_i)_{i \in [m]}$ with $p_i > 0$ for all $i \in [m]$. A **p**-tree is a random tree in $\mathbb{T}_m$, with law as follows: For any fixed $\mathbf{t} \in \mathbb{T}_m$ and $v \in \mathbf{t}$, write $d_v(\mathbf{t})$ for the number of children of $v$ in the tree $\mathbf{t}$. Then the law of the **p**-tree, denoted by $\mathbb{P}_{\text{tree}}$, is defined as

$$\mathbb{P}_{\text{tree}}(\mathbf{t}) = \mathbb{P}_{\text{tree}}(\mathbf{t}; \mathbf{p}) = \prod_{v \in [m]} p_v^{d_v(\mathbf{t})}, \quad \mathbf{t} \in \mathbb{T}_m. \tag{4.8}$$

Generating a random **p**-tree $\mathcal{T} \sim \mathbb{P}_{\text{tree}}$ and then assigning a uniform random order on the children of every vertex $v \in \mathcal{T}$ gives a random element with law $\mathbb{P}_{\text{ord}}(\cdot; \mathbf{p})$ given by

$$\mathbb{P}_{\text{ord}}(\mathbf{t}) = \mathbb{P}_{\text{ord}}(\mathbf{t}; \mathbf{p}) = \prod_{v \in [m]} \frac{p_v^{d_v(\mathbf{t})}}{(d_v(\mathbf{t}))!}, \quad \mathbf{t} \in \mathbb{T}_m^{\text{ord}}. \tag{4.9}$$

**The birthday construction of p-trees.**

We now describe a construction of **p**-trees, formulated in [59], that is relevant to this work. Let $\mathbf{Y} := (Y_0, Y_1, \ldots)$ be a sequence of i.i.d. random variables with distribution $\mathbf{p}$. Let $R_0 = 0$ and for $l \geqslant 1$, let $R_l$ denote the $l$-th repeat time, i.e., $R_l = \min \{k > R_{l-1} : Y_k \in \{Y_0, \ldots, Y_{k-1}\}\}$. Now consider the directed graph formed via the edges $\mathcal{T}(\mathbf{Y}) := \{(Y_{j-1}, Y_j) : Y_j \notin \{Y_0, \ldots, Y_{j-1}\}, j \geqslant 1\}$. This gives a tree which we view as rooted at $Y_0$. The following striking result was shown in [59]:

**Theorem 4.3.1** ([59, Lemma 1 and Theorem 2]). *The random tree $\mathcal{T}(\mathbf{Y})$, viewed as an object in $\mathbb{T}_m$, is distributed as a **p**-tree with distribution (4.8) independently of $Y_{R_1-1}, Y_{R_2-1}, \ldots$ which are i.i.d with distribution $\mathbf{p}$.*

**Remark 32.** The independence between the sequence $Y_{R_1-1}, Y_{R_2-1}, \ldots$ and the constructed **p**-tree $\mathcal{T}(\mathbf{Y})$ is truly remarkable. In particular, let $\mathcal{T}_r \subset \mathcal{T}(\mathbf{Y})$ denote the subtree with vertex set $\{Y_0, Y_1, \ldots, Y_{R_r-1}\}$, namely the tree constructed in the first $R_r$ steps. Further take $\tilde{\mathbf{Y}} = (\tilde{Y}_1, \ldots \tilde{Y}_r)$ an i.i.d. sample

from $\mathbf{p}$ and then construct the subtree $\mathcal{S}_r$ spanned by $\tilde{\mathbf{Y}}$. Then the above result (formalized as [59, Corollary 3]) implies that

$$(\tilde{Y}_1, \tilde{Y}_2, \ldots, \tilde{Y}_r; \mathcal{S}_r) \overset{d}{=} (Y_{R_1-1}, Y_{R_2-1}, \ldots Y_{R_r-1}; \mathcal{T}_r).$$

We will use this fact in Section 4.4 to complete the proof of the universality theorem.

**Tilted p-trees and connected components of $NR_n(x,t)$.**

Consider the vertex set $[n]$ and assign weight $x_i$ to vertex $i$. Now, connect each pair of vertices $i, j$ ($i \neq j$) independently with probability $q_{ij} := 1 - \exp(-t x_i x_j)$. The resulting random graph, denoted by $NR_n(\mathbf{x}, t)$, is known as the Norros-Reittu model or the Poisson graph process [98]. For a connected component $\mathcal{C} \subseteq NR_n(\mathbf{x}, t)$, let $\mathrm{mass}(\mathcal{C}) := \sum_{i \in \mathcal{C}} x_i$ and, for any $t \geqslant 0$, $(\mathcal{C}_i(t))_{i \geqslant 1}$ denotes the components in decreasing order of their mass sizes. In this section, we describe results from [35] that gave a method of constructing connected components of $NR_n(\mathbf{x}, t)$, conditionally on the vertices of the components. This construction involves tilted versions of $\mathbf{p}$-trees introduced in Section 4.3.4. Since these trees are parametrized via a driving probability mass function (pmf) $\mathbf{p}$, it will be easy to parametrize various random graph constructions in terms of pmfs as opposed to vertex weights $\mathbf{x}$. Proposition 4.3.2 will relate vertex weights to pmfs.

Fix $n \geqslant 1$ and $\mathcal{V} \subset [n]$, and write $\mathbb{G}_{\mathcal{V}}^{\mathrm{con}}$ for the space of all simple connected graphs with vertex set $\mathcal{V}$. For fixed $a > 0$, and probability mass function $\mathbf{p} = (p_v)_{v \in \mathcal{V}}$, define probability distributions $\mathbb{P}_{\mathrm{con}}(\cdot; \mathbf{p}, a, \mathcal{V})$ on $\mathbb{G}_{\mathcal{V}}^{\mathrm{con}}$ as follows: For $i, j \in \mathcal{V}$, denote

$$q_{ij} := 1 - \exp(-a p_i p_j). \tag{4.10}$$

Then, for $G \in \mathbb{G}_{\mathcal{V}}^{\mathrm{con}}$,

$$\mathbb{P}_{\mathrm{con}}(G; \mathbf{p}, a, \mathcal{V}) := \frac{1}{Z(\mathbf{p}, a)} \prod_{(i,j) \in E(G)} q_{ij} \prod_{(i,j) \notin E(G)} (1 - q_{ij}), \tag{4.11}$$

where $Z(\mathbf{p}, a)$ is the normalizing constant. Now let $\mathcal{V}^{(i)}$ be the vertex set of $\mathcal{C}_i(t)$ for $i \geqslant 1$, and note that $(\mathcal{V}^{(i)})_{i \geqslant 1}$ denotes a random finite partition of the vertex set $[n]$. The next proposition yields a construction of the random (connected) graphs $(\mathcal{C}_i(t))_{i \geqslant 1}$:

**Proposition 4.3.2** ([35, Proposition 6.1]). *Given the partition* $(\mathcal{V}^{(i)})_{i \geqslant 1}$*, define, for* $i \geqslant 1$,

$$\mathbf{p}_n^{(i)} := \left( \frac{x_v}{\sum_{v \in \mathcal{V}^{(i)}} x_v} : v \in \mathcal{V}^{(i)} \right), \quad a_n^{(i)} := t \left( \sum_{v \in \mathcal{V}_{(i)}} x_v \right)^2. \tag{4.12}$$

*For each fixed* $i \geqslant 1$*, let* $G_i \in \mathbb{G}_{\mathcal{V}^{(i)}}^{con}$ *be a connected simple graph with vertex set* $\mathcal{V}^{(i)}$*. Then*

$$\mathbb{P}\left( \mathcal{C}_i(t) = G_i, \ \forall i \geqslant 1 \ \middle| \ (\mathcal{V}^{(i)})_{i \geqslant 1} \right) = \prod_{i \geqslant 1} \mathbb{P}_{con}(G_i; \mathbf{p}_n^{(i)}, a_n^{(i)}, \mathcal{V}^{(i)}).$$

**Algorithm 4.1.** The random graph $NR_n(\mathbf{x}, t)$ can be generated in two stages:

(S0) Generate the random partition $(\mathcal{V}^{(i)})_{i \geqslant 1}$ of the vertices into different components.

(S1) Conditional on the partition, generate the internal structure of each component following the law of $\mathbb{P}_{con}(\cdot; \mathbf{p}^{(i)}, a^{(i)}, \mathcal{V}^{(i)})$, independently across different components.

Let us now describe an algorithm to generate such connected components using the distribution in (4.11). To ease notation, let $\mathcal{V} = [m]$ for some $m \geqslant 1$ and fix a probability mass function $\mathbf{p}$ on $[m]$ and a constant $a > 0$ and write $\mathbb{P}_{con}(\cdot) := \mathbb{P}_{con}(\cdot; \mathbf{p}, a, [m])$ on $\mathbb{G}_m^{con} := \mathbb{G}_{[m]}^{con}$. As a matter of convention, we view ordered rooted trees via their planar embedding using the associated ordering to determine the relative locations of siblings of an individual. We think of the left-most sibling as the "oldest". Further, in a depth-first exploration, we explore the tree from left to right. Now given a planar rooted tree $\mathbf{t} \in \mathbb{T}_m$, let $\rho$ denote the root and for every vertex $v \in [m]$, let $[\rho, v]$ denote the path connecting $\rho$ to $v$ in the tree. Given this path and a vertex $i \in [\rho, v]$, write $RC(i, [\rho, v])$ for the set of all children of $i$ that fall to the right of $[\rho, v]$. Define $\mathfrak{P}(v, \mathbf{t}) := \cup_{i \in [m]} RC(i, [\rho, v])$. In the terminology of [3, 36], $\mathfrak{P}(v, \mathbf{t})$ denotes the set of endpoints of all *permitted edges* emanating from $v$. The surplus edges of the graph $G$, sampled from $\mathbb{P}_{con}(\cdot)$, are formed only between $v$ and $\mathfrak{P}(v, \mathbf{t})$, as $v$ varies. Define

$$\mathfrak{G}_{(m)}(v) := \sum_{i \in [\rho, v]} \sum_{j \in [m]} p_j \mathbf{1}\{j \in RC(i, [\rho, v])\}. \tag{4.13}$$

Let $(v(1), v(2), \ldots, v(m))$ denote the order of the vertices in the depth-first exploration of the tree $\mathbf{t}$. Let $y^*(0) = 0$ and $y^*(i) = y^*(i-1) + p_{v(i)}$ and

define

$$A_{(m)}(u) = \mathfrak{G}_{(m)}(u), \text{ for } u \in (y^*(i-1), y^*(i)], \quad \text{and} \quad \bar{A}_{(m)}(\cdot) := a A_{(m)}(\cdot),$$
(4.14)

where $a$ is defined in (4.10). Define the function

$$\Lambda_{(m)}(\mathbf{t}) := a \sum_{v \in [m]} p_v \mathfrak{G}_{(m)}(v).$$
(4.15)

Finally, let $E(\mathbf{t})$ denote the set of edges of $\mathbf{t}$, $\mathscr{T}_m^{\mathbf{p}}$ the $\mathbf{p}$-tree defined in (4.9), $\mathfrak{P}(\mathbf{t}) = \cup_{v \in [m]} \mathfrak{P}(v, \mathbf{t})$, and define the tilt function $L : \mathbb{T}_m^{\mathrm{ord}} \to \mathbb{R}_+$ by

$$L(\mathbf{t}) = L_{(m)}(\mathbf{t}) := \prod_{(k,\ell) \in E(\mathbf{t})} \left[ \frac{\exp(a p_k p_\ell) - 1}{a p_k p_\ell} \right] \exp \left( \sum_{(k,\ell) \in \mathfrak{P}(\mathbf{t})} a p_k p_\ell \right), \quad (4.16)$$

for $\mathbf{t} \in \mathbb{T}_m^{\mathrm{ord}}$. Recall the (ordered) $\mathbf{p}$-tree distribution from (4.9). Using $L(\cdot)$ to tilt this distribution results in the distribution

$$\mathbb{P}_{\mathrm{ord}}^\star(\mathbf{t}) := \mathbb{P}_{\mathrm{ord}}(\mathbf{t}) \cdot \frac{L(\mathbf{t})}{\mathbb{E}_{\mathrm{ord}}[L(\mathscr{T}_m^{\mathbf{p}})]}, \qquad \mathbf{t} \in \mathbb{T}_m^{\mathrm{ord}}. \quad (4.17)$$

While all of these objects depend on the tree $\mathbf{t}$, we suppress this dependence to ease notation.

**Algorithm 4.2.** Let $\tilde{\mathcal{G}}_m(\mathbf{p}, a)$ denote a random graph sampled from $\mathbb{P}_{\mathrm{con}}(\cdot)$. This algorithm gives a construction of $\tilde{\mathcal{G}}_m(\mathbf{p}, a)$, proved in [36].

(S1) **Tilted $\mathbf{p}$-tree:** Generate a tilted ordered $\mathbf{p}$-tree $\mathscr{T}_m^{\mathbf{p}, \star}$ with distribution (4.17). Now consider the (random) objects $\mathfrak{P}(v, \mathscr{T}_m^{\mathbf{p}, \star})$ for $v \in [m]$ and the corresponding (random) functions $\mathfrak{G}_{(m)}(\cdot)$ on $[m]$ and $A_{(m)}(\cdot)$ on $[0, 1]$.

(S2) **Poisson number of possible surplus edges:** Let $\mathcal{P}$ denote a rate-one Poisson process on $\mathbb{R}_+^2$ that is independent of all other randomness and define

$$\bar{A}_{(m)} \cap \mathcal{P} := \left\{ (s, t) \in \mathcal{P} : s \in [0, 1], t \leqslant \bar{A}_{(m)}(s) \right\}.$$

Write $\bar{A}_{(m)} \cap \mathcal{P} := \{(s_j, t_j) : 1 \leqslant j \leqslant N_{(m)}^\star\}$ where $N_{(m)}^\star = |\bar{A}_{(m)} \cap \mathcal{P}|$. We next use the set $\{(s_j, t_j) : 1 \leqslant j \leqslant N_{(m)}^\star\}$ to generate pairs of points $\{(\mathcal{L}_j, \mathcal{R}_j) : 1 \leqslant j \leqslant N_{(m)}^\star\}$ in the tree that will be joined to form the surplus edges.

(S3) **"First" endpoints:** Fix $j$ and suppose $s_j \in (y^*(i-1), y^*(i)]$ for some $i \geqslant 1$, where $y^*(i)$ is as given right above (4.14). Then the *first endpoint* of the surplus edge corresponding to $(s_j, t_j)$ is $\mathcal{L}_j := v(i)$, where $v(i)$ is defined right below (4.13).

(S4) **"Second" endpoints:** Note that in the interval $(y^*(i-1), y^*(i)]$, the function $\bar{A}_{(m)}$ is of constant height $a\mathfrak{G}_{(m)}(v(i))$. We will view this height as being partitioned into sub-intervals of length $ap_u$ for each element $u \in \mathfrak{P}(v(i), \mathscr{T}_m^{\mathbf{p}, \star})$, the collection of endpoints of permitted edges emanating from $\mathcal{L}_k$. (Assume that this partitioning is done according to some preassigned rule, e.g., using the order of the vertices in $\mathfrak{P}(v(i), \mathscr{T}_m^{\mathbf{p}, \star})$.) Suppose $t_j$ belongs to the interval corresponding to $u$. Then the *second endpoint* is $\mathcal{R}_j = u$. Form an edge between $(\mathcal{L}_j, \mathcal{R}_j) = (v(i), u)$.

(S5) In this construction, it is possible that one creates more than one surplus edge between two vertices. Remove any multiple surplus edges. This has vanishing probability in our applications.

**Definition 3.** *Consider the connected random graph $\tilde{\mathcal{G}}_m(\mathbf{p}, a)$, given by Algorithm 4.2, viewed as a measured metric space via the graph distance and each vertex $v$ is assigned measure $p_v$.*

The following lemma describes the law of $\tilde{\mathcal{G}}_m(\mathbf{p}, a)$:

**Lemma 4.3.3** ([36, Lemma 4.10]). *The random graph $\tilde{\mathcal{G}}_m(\mathbf{p}, a)$ generated by Algorithm 4.2 has the same law as $\mathbb{P}_{\mathrm{con}}(\cdot)$. Further, conditionally on $\mathscr{T}_m^{\mathbf{p}, \star}$,*

1. *$N_{(m)}^\star$ has Poisson distribution with mean $\Lambda_{(m)}(\mathscr{T}_m^{\mathbf{p}, \star})$ where $\Lambda_{(m)}$ is as in (4.15);*

2. *conditionally further on $N_{(m)}^\star = k$, the first endpoints $(\mathcal{L}_j)_{j \in [k]}$ can be generated in an i.i.d fashion by sampling from the vertex set $[m]$ with probability distribution $\mathcal{J}^{(m)}(v) \propto p_v \mathfrak{G}_{(m)}(v), v \in [m]$.;*

3. *conditionally further on $N_{(m)}^\star = k$ and the first endpoints $(\mathcal{L}_j)_{j \in [k]}$, generate the second endpoints in an i.i.d. fashion where conditionally on $\mathcal{L}_j = v$, the probability distribution of $\mathcal{R}_j$ is given by*

$$Q_v^{(m)}(y) := \begin{cases} \sum_u p_u \mathbf{1}\{u \in \mathrm{RC}(y, [\rho, v])\}/\mathfrak{G}_{(m)}(v) & \text{if } y \in [\rho, v], \\ 0 & \text{otherwise}, \end{cases}$$

*and create an edge between $\mathcal{L}_j$ and $\mathcal{R}_j$ for $1 \leqslant j \leqslant k$.*

In a series of papers [12–14] it was shown that **p**-trees, under various assumptions, converge to inhomogeneous continuum random trees that we now describe. Recall from [81, 127] that a real tree is a metric space $(\mathcal{T}, d)$ that satisfies the following for every pair $a, b \in \mathcal{T}$:

1. There is a **unique** isometric map $f_{a,b} \colon [0, d(a, b)] \to \mathcal{T}$ such that $f_{a,b}(0) = a$, $f_{a,b}(d(a, b)) = b$.

2. For any continuous one-to-one map $g \colon [0, 1] \to \mathcal{T}$ with $g(0) = a$ and $g(1) = b$, we have $g([0, 1]) = f_{a,b}([0, d(a, b)])$.

**Construction of the ICRT:** Given $\beta \in \ell^2_\downarrow \setminus \ell^1_\downarrow$ with $\sum_i \beta_i^2 = 1$, we will now define the inhomogeneous continuum random tree $\mathcal{T}^\beta$. We mainly follow the notation in [14]. Assume that we are working on a probability space $(\Omega, \mathcal{F}, \mathbb{P}_\beta)$ rich enough to support the following:

1. For each $i \geqslant 1$, let $\mathcal{P}_i := (\xi_{i,1}, \xi_{i,2}, \ldots)$ be rate $\beta_i$ Poisson processes that are independent for different $i$. The first point of each process $\xi_{i,1}$ is special and is called a *joinpoint*, while the remaining points $\xi_{i,j}$ with $j \geqslant 2$ will be called $i$-*cutpoints* [14].

2. Independently of the above, let $\mathbf{U} = (U_j^{(i)})_{i,j \geqslant 1}$ be a collection of i.i.d. uniform $(0, 1)$ random variables. These are not required to construct the tree but will be used to define a certain function on the tree.

The **random** real tree (with marked vertices) $\mathcal{T}^\beta_{(\infty)}$ is then constructed as follows:

1. Arrange the cutpoints $\{\xi_{i,j} : i \geqslant 1, j \geqslant 2\}$ in increasing order as $0 < \eta_1 < \eta_2 < \cdots$. The assumption that $\sum_i \beta_i^2 < \infty$ implies that this is possible. For every cutpoint $\eta_k = \xi_{i,j}$, let $\eta_k^* := \xi_{i,1}$ be the corresponding joinpoint.

2. Next, build the tree inductively. Start with the branch $[0, \eta_1]$. Inductively assuming that we have completed step $k$, attach the branch $(\eta_k, \eta_{k+1}]$ to the joinpoint $\eta_k^*$ corresponding to $\eta_k$.

Write $\mathcal{T}^\beta_0$ for the corresponding tree after one has used up all the branches $[0, \eta_1], \{(\eta_k, \eta_{k+1}] : k \geqslant 1\}$. Note that for every $i \geqslant 1$, the joinpoint $\xi_{i,1}$ corresponds to a vertex with infinite degree. Label this vertex $i$. The ICRT $\mathcal{T}^\beta_{(\infty)}$ is the completion of the marked metric tree $\mathcal{T}^\beta_0$. As argued in [14, Section 2], this is a real-tree as defined above which can be viewed as rooted at the vertex corresponding to zero. We call the vertex corresponding to joinpoint $\xi_{i,1}$

**hub** i. Since $\sum_i \beta_i = \infty$, one can check that hubs are almost everywhere dense on $\mathscr{T}_{(\infty)}^{\beta}$.

The uniform random variables $(U_j^{(i)})_{i,j \geqslant 1}$ give rise to a natural ordering on $\mathscr{T}_{(\infty)}^{\beta}$ (or a planar embedding of $\mathscr{T}_{(\infty)}^{\beta}$) as follows: For $i \geqslant 1$, let $(\mathscr{T}_j^{(i)})_{j \geqslant 1}$ be the collection of subtrees hanging off the $i$-th hub. Associate $U_j^{(i)}$ with the subtree $\mathscr{T}_j^{(i)}$, and think of $\mathscr{T}_{j_1}^{(i)}$ appearing "to the right of" $\mathscr{T}_{j_2}^{(i)}$ if $U_{j_1}^{(i)} < U_{j_2}^{(i)}$. This is the natural ordering on $\mathscr{T}_{(\infty)}^{\beta}$ when it is being viewed as a limit of ordered **p**-trees. We can think of the pair $(\mathscr{T}_{(\infty)}^{\beta}, \mathbf{U})$ as the **ordered ICRT**.

### 4.3.5   Continuum limits of components

The aim of this section is to give an explicit description of the limiting (random) metric spaces in Theorem 4.1.1. We start by constructing a specific metric space using the tilted version of the ICRT in Section 4.3.5. Then we describe the limits of maximal components in Section 4.3.5.

**Tilted ICRTs and vertex identification**

Let $(\Omega, \mathcal{F}, \mathbb{P}_\beta)$ and $\mathscr{T}_{(\infty)}^{\beta}$ be as in Section 4.3.4. In [14], it was shown that one can associate a natural probability measure $\mu$, called the **mass measure**, to $\mathscr{T}_{(\infty)}^{\beta}$, satisfying $\mu(\mathfrak{L}(\mathscr{T}_{(\infty)}^{\beta})) = 1$. Here we recall that $\mathfrak{L}(\cdot)$ denotes the set of leaves. Before moving to the desired construction of the random metric space, we will need to define some more quantities that describes the asymptotic analogues of the quantities appearing in Algorithm 4.2. Similarly to (4.13), define

$$\mathfrak{G}_{(\infty)}(y) = \sum_{i \geqslant 1} \beta_i \left( \sum_{j \geqslant 1} U_j^{(i)} \times \mathbf{1}\{y \in \mathscr{T}_j^{(i)}\} \right).$$

It was shown in [36] that $\mathfrak{G}_{(\infty)}(y)$ is finite for almost every realization of $\mathscr{T}_{(\infty)}^{\beta}$ and for $\mu$-almost every $y \in \mathscr{T}_{(\infty)}^{\beta}$. For $y \in \mathscr{T}_{(\infty)}^{\beta}$, let $[\rho, y]$ denote the path from the root $\rho$ to $y$. For every $y$, define a probability measure on $[\rho, y]$ as

$$Q_y^{(\infty)}(v) := \frac{\beta_i U_j^{(i)}}{\mathfrak{G}_{(\infty)}(y)}, \quad \text{if } v \text{ is the } i\text{-th hub and } y \in \mathscr{T}_j^{(i)} \text{ for some } j. \quad (4.18)$$

Thus, this probability measure is concentrated on the hubs on the path from $y$ to the root. Let $\gamma > 0$ be a constant. The choice of the function $\gamma$ is indicated in Assumption 4.3. Informally, the construction goes as follows: We will first tilt the distribution of the original ICRT $\mathscr{T}_{(\infty)}^{\beta}$ using the exponential functional

$$L_{(\infty)}(\mathscr{T}_{(\infty)}^{\beta}, \mathbf{U}) := \exp \left( \gamma \int_{y \in \mathscr{T}_{(\infty)}^{\beta}} \mathfrak{G}_{(\infty)}(y) \mu(dy) \right)$$

to get a tilted tree $\mathscr{T}_{(\infty)}^{\beta,\star}$. We then generate a random but finite number $N_{(\infty)}^{\star}$ of pairs of points $\{(x_k, y_k) : 1 \leqslant k \leqslant N_{(\infty)}^{\star}\}$ that will provide the surplus edges. The final metric space is obtained by creating "shortcuts" by identifying the points $x_k$ and $y_k$. Formally the construction proceeds in four steps:

1. **Tilted ICRT:** Define $\mathbb{P}_\beta^\star$ on $\Omega$ by

$$\frac{d\mathbb{P}_\beta^\star}{d\mathbb{P}_\beta} = \frac{\exp\left(\gamma \int_{y \in \mathscr{T}_{(\infty)}^\beta} \mathfrak{G}_{(\infty)}(y)\mu(dy)\right)}{\mathbb{E}\left[\exp\left(\gamma \int_{x \in \mathscr{T}_{(\infty)}^\beta} \mathfrak{G}_{(\infty)}(x)\mu(dx)\right)\right]}.$$

   The expectation in the denominator is with respect to the original measure $\mathbb{P}_\beta$. Write $(\mathscr{T}_{(\infty)}^{\beta,\star}, \mu^\star)$ and $\mathbf{U}^\star = (U_j^{(i),\star})_{i,j \geqslant 1}$ for the tree and the mass measure on it, and the associated random variables under this change of measure.

2. **Poisson number of identification points:** Conditionally on the object $((\mathscr{T}_{(\infty)}^{\beta,\star}, \mu^\star), \mathbf{U}^\star)$, generate $N_{(\infty)}^\star$ having a Poisson$(\Lambda_{(\infty)}^\star)$ distribution, where

$$\Lambda_{(\infty)}^\star := \gamma \int_{y \in \mathscr{T}_{(\infty)}^{\beta,\star}} \mathfrak{G}_{(\infty)}(y)\mu^\star(dy) = \gamma \sum_{i \geqslant 1} \beta_i \left[\sum_{j \geqslant 1} U_j^{(i),\star}\mu^\star(\mathscr{T}_j^{(i),\star})\right].$$

   Here, $(\mathscr{T}_j^{(i),\star})_{j \geqslant 1}$ denotes the collection of subtrees of hub $i$ in $\mathscr{T}_{(\infty)}^{\beta,\star}$.

3. **"First" endpoints (of shortcuts):** Conditionally on (a) and (b), sample $x_k$ from $\mathscr{T}_{(\infty)}^{\beta,\star}$ with density proportional to $\mathfrak{G}_{(\infty)}(x)\mu^\star(dx)$ for $1 \leqslant k \leqslant N_{(\infty)}^\star$.

4. **"Second" endpoints (of shortcuts) and identification:** Having chosen $x_k$, choose $y_k$ from the path $[\rho, x_k]$ joining the root $\rho$ and $x_k$ according to the probability measure $Q_{x_k}^{(\infty)}$ as in (4.18) but with $U_j^{(i),\star}$ replacing $U_j^{(i)}$. (Note that $y_k$ is always a **hub** on $[\rho, x_k]$.) Identify $x_k$ and $y_k$, i.e., form the quotient space by introducing the equivalence relation $x_k \sim y_k$ for $1 \leqslant k \leqslant N_{(\infty)}^\star$.

**Definition 4.** *Fix $\gamma \geqslant 0$ and $\beta \in \ell_\downarrow^2 \setminus \ell_\downarrow^1$ with $\sum_i \beta_i^2 = 1$. Let $\mathcal{G}_{(\infty)}(\beta, \gamma)$ be the metric measure space constructed via the four steps above equipped with the measure inherited from the mass measure on $\mathscr{T}_{(\infty)}^{\beta,\star}$.*

### Scaling limit for the component sizes and surplus edges

Let us describe the scaling limit results for the component sizes and the surplus edges (#edges $-$ #vertces $+$ 1) for the largest components of $CM_n(\mathbf{d}, p_n(\lambda))$

from Chapter 3. Although we need to define the limiting object only for describing the limiting metric space, the convergence result will turn out to be crucial in Section 4.5 in the proof of Theorem 4.1.1, and therefore we state it here as well. Consider a decreasing sequence $\theta \in \ell^3_{\downarrow} \setminus \ell^2_{\downarrow}$. Denote by $\mathcal{I}_i(s) := \mathbf{1}\{\zeta_i \leqslant s\}$ where $\zeta_i \sim \text{Exp}(\theta_i)$ independently, and $\text{Exp}(r)$ denotes the exponential distribution with rate $r$. Consider the process

$$\bar{S}^\lambda_\infty(t) = \sum_{i=1}^{\infty} \theta_i \left( \mathcal{I}_i(t) - \theta_i t \right) + \lambda t, \tag{4.19}$$

for some $\lambda \in \mathbb{R}$. Define the reflected version of $\bar{S}^\lambda_\infty(t)$ by $\text{refl}(\bar{S}^\lambda_\infty(t)) = \bar{S}^\lambda_\infty(t) - \inf_{0 \leqslant u \leqslant t} \bar{S}^\lambda_\infty(u)$. The processes of the form (4.19) were termed *thinned* Lévy processes in [33] since the summands are thinned versions of Poisson processes. Let $(\Xi_i(\theta, \lambda))_{i \geqslant 1}$, $(\xi_i(\theta, \lambda))_{i \geqslant 1}$, respectively, denote the vector of excursions and excursion-lengths, ordered according to the excursion lengths in a decreasing manner. Denote the vector $(\xi_i(\theta, \lambda))_{i \geqslant 1}$ by $\xi(\theta, \lambda)$. The fact that $\xi(\theta, \lambda)$ is always well defined follows from [11, Lemma 1]. Also, define the counting process of marks $N$ to be a Poisson process that has intensity $\text{refl}(\bar{S}^\lambda_\infty(t))$ at time $t$ conditional on $(\text{refl}(\bar{S}^\lambda_\infty(u)))_{u \leqslant t}$. We use the notation $\mathcal{N}_i(\theta, \lambda)$ to denote the number of marks within the $i$-th largest excursion $\Xi_i(\theta, \lambda)$.

For a connected graph $G$, let $\text{SP}(G) = \#\text{edges} - \#\text{vertices} + 1$ denote its surplus edges. In the context of this chapter, we simply write $\xi_i$, $\xi$ and $\mathcal{N}_i$ respectively for $\xi_i(\theta/(\mu\nu), \lambda/\mu)$, $\xi(\theta/(\mu\nu), \lambda/\mu)$ and $\mathcal{N}_i(\theta/(\mu\nu), \lambda/\mu)$.

**Proposition 4.3.4.** *Under Assumption 4.1, as $n \to \infty$,*

$$\left( n^{-\rho} |\mathscr{C}^{\text{p}}_{(i)}(\lambda)|, \text{SP}(\mathscr{C}^{\text{p}}_{(i)}(\lambda)) \right)_{i \geqslant 1} \xrightarrow{d} \left( \frac{1}{\nu} \xi_i, \mathcal{N}_i \right)_{i \geqslant 1}, \tag{4.20}$$

*with respect to the topology on the product space $\ell^2_{\downarrow} \times \mathbb{N}^{\mathbb{N}}$.*

Proposition 4.3.4 was proved in Chapter 3. The limiting object in Theorem 3.2.1 is stated in a slightly different form compared to the right-hand side of (4.20). However, the limiting objects are identical in distribution with suitable rescaling of time and space, and by observing that $r\text{Exp}(r) \stackrel{d}{=} \text{Exp}(1)$, where $\text{Exp}(r)$ denotes an exponential random variable with rate $r$ (See Appendix 4.A). In fact, the arguments in Appendix 4.A establish the following lemma that will be used extensively in Section 4.5.

**Lemma 4.3.5.** *For $\eta_1, \eta_2 > 0$, $\theta \in \ell^3_{\downarrow} \setminus \ell^2_{\downarrow}$ and $\lambda \in \mathbb{R}$, $\xi(\eta_1 \theta, \eta_2 \lambda) \stackrel{d}{=} \frac{1}{\eta_1} \xi\left( \theta, \frac{\eta_2}{\eta_1^2} \lambda \right)$.*

**Limiting component structures**

We are now all set to describe the metric space $M_i$ appearing in Theorem 4.1.1. Recall the graph $\mathcal{G}_\infty(\boldsymbol{\beta}, \boldsymbol{\gamma})$ from Definition 4. Using the notation of Section 4.3.5, write $\xi_i^*$ for $\xi_i((\mu(\nu-1))^{-1}\boldsymbol{\theta}, (\mu(\nu-1)^2)^{-1}\nu^2\lambda)$ and $\Xi_i^*$ for the excursion corresponding to $\xi_i^*$. Note that $\xi_i^*$ has the same distribution as $(\nu-1)\xi_i/\nu$, where $\xi_i$ is as in Proposition 4.3.4. Then the limiting space $M_i$ is distributed as

$$M_i \overset{d}{=} \frac{\nu}{\nu-1} \frac{\xi_i^*}{\left(\sum_{v \in \Xi_i^*} \theta_v^2\right)^{1/2}} \mathcal{G}_\infty(\boldsymbol{\theta}^{(i)}, \boldsymbol{\gamma}^{(i)}),$$

where $\boldsymbol{\theta}^{(i)} = \left(\frac{\theta_j}{\sum_{v \in \Xi_i^*} \theta_v^2} : j \in \Xi_i^*\right)$ and $\boldsymbol{\gamma}^{(i)} = \frac{\xi_i^*}{\mu(\nu-1)}\left(\sum_{v \in \Xi_i^*} \theta_v^2\right)^{1/2}$.

## 4.4   Universality theorem

In this section, we develop universality principles that enable us to derive the scaling limits of the components for graphs that can be compared with the critical rank-one inhomogeneous random graph in a suitable sense. Our universality theorem closely resembles that in [29, Theorem 6.4] which was developed in a different context to derive the scaling limits of the components for general inhomogeneous random graphs with a finite number of types and the configuration model with an exponential moment condition on the degrees. We first state the relevant result from [36] that was used in the context of rank-one inhomogeneous random graphs and then state our main result below. The convergence of metric spaces is with respect to the Gromov-weak topology, unless stated otherwise. Recall the measured metric spaces $\tilde{\mathcal{G}}_m(\mathbf{p}, a)$ and $\mathcal{G}_\infty(\boldsymbol{\beta}, \boldsymbol{\gamma})$ defined in Definitions 3 and 4.

**Assumption 4.3.**   (i) Let $\sigma(\mathbf{p}) := \left(\sum_i p_i^2\right)^{1/2}$. As $m \to \infty$, $\sigma(\mathbf{p}) \to 0$, and for each fixed $i \geqslant 1$, $p_i/\sigma(\mathbf{p}) \to \beta_i$, where $\boldsymbol{\beta} = (\beta_i)_{i \geqslant 1} \in \ell^2_\downarrow \setminus \ell^1_\downarrow$, $\sum_i \beta_i^2 = 1$.

(ii) Recall $a$ from (4.10). There exists a constant $\gamma > 0$ such that $a\sigma(\mathbf{p}) \to \gamma$.

Assumption 4.3 (i) is a sufficient condition for the convergence of $\mathbf{p}$-trees [59] when the edges are assigned edge-length $\sigma(\mathbf{p})$. Assumption 4.3 (ii) is required for the tilting function $L(\cdot)$ to converge. This suggests that the tilted $\mathbf{p}$-tree in Algorithm 4.2 (S1) converges to tilted ICRT. In [36], the above fact was proved, and it was further shown that the scaling limit holds after the shortcuts are created during Algorithm 4.2 (S2) onwards.

**Theorem 4.4.1** ([36, Theorem 4.5]). *Under Assumption 4.3,* $\sigma(\mathbf{p})\tilde{\mathcal{G}}_m(\mathbf{p}, a) \xrightarrow{d} \mathcal{G}_{(\infty)}(\boldsymbol{\beta}, \gamma)$, *as* $m \to \infty$.

For each $m \geqslant 1$, fix a collection of blobs $\mathbf{M}_m := \{(M_i, d_i, \mu_i) : i \in [m]\}$. Recall the definition of super graphs from Section 4.3.2 and denote

$$\tilde{\mathcal{G}}_m^{bl}(\mathbf{p}, a) = \Gamma(\tilde{\mathcal{G}}_m(\mathbf{p}, a), \mathbf{p}, \mathbf{M}_m, \mathbf{X}),$$

where $\mathbf{X} = (X_{ij})_{i,j \in [m]}$, $X_{ij} \sim \mu_i$ independently for each $i$. Moreover, $\mathbf{X}$ is independent of the graph $\tilde{\mathcal{G}}_m(\mathbf{p}, a)$. Let $u_i := \mathbb{E}[d_i(X_i, X_i')]$ where $X_i, X_i' \sim \mu_i$ independently and $B_m := \sum_{i \in [m]} p_i u_i$. Let $\Delta_i := \mathrm{diam}(M_i)$, $\Delta_{\max} := \max_{i \in [m]} \Delta_i$.

**Assumption 4.4** (Maximum inter-blob-distance). $\lim_{m \to \infty} \frac{\sigma(\mathbf{p})\Delta_{\max}}{B_m + 1} = 0$.

Assumption 4.4 basically says that the blobs have negligible diameter compared to the average distances in the metric space. The next theorem is the universality theorem, which basically says that the introduction of the blobs does not change the scaling limits in [36, Theorem 4.5] if the distances are normalized accordingly.

**Theorem 4.4.2** (Universality theorem). *Under Assumptions 4.3, 4.4, as* $m \to \infty$,

$$\frac{\sigma(\mathbf{p})}{B_m + 1} \tilde{\mathcal{G}}_m^{bl}(\mathbf{p}, a) \xrightarrow{d} \mathcal{G}_{(\infty)}(\boldsymbol{\beta}, \gamma).$$

### 4.4.1    Completing the proof of Theorem 4.4.2

This section is devoted to the proof of Theorem 4.4.2. To simplify notation, we write $\tilde{\mathcal{G}}_m$, $\tilde{\mathcal{G}}_m^{bl}$ respectively instead of $\tilde{\mathcal{G}}_m(\mathbf{p}, a)$ and $\tilde{\mathcal{G}}_m^{bl}(\mathbf{p}, a)$.

**Lemma 4.4.3** ([36, Lemma 4.11]). *Recall the definition of* $N_{(m)}^\star$ *from Algorithm 4.2. The sequence of random variables* $(N_{(m)}^\star)_{m \geqslant 1}$ *is tight.*

Recall the definition of Gromov-weak topology from Section 4.3.1. Fix some $l \geqslant 1$ and take any bounded continuous function $\phi : \mathbb{R}^{l^2} \mapsto \mathbb{R}$. We simply write $\Phi(X)$ for $\Phi((X, d, \mu))$.

**Key step 1.** Let us write the scaled metric spaces as $\tilde{\mathcal{G}}_m^s = \sigma(\mathbf{p})\tilde{\mathcal{G}}_m$ and $\tilde{\mathcal{G}}_m^{bl,s} = \frac{\sigma(\mathbf{p})}{B_m+1}\tilde{\mathcal{G}}_m^{bl}$. Using Theorem 4.4.1 it is enough to show that

$$\lim_{m \to \infty} \left| \mathbb{E}\left[\Phi\left(\tilde{\mathcal{G}}_m^{bl,s}\right)\right] - \mathbb{E}\left[\Phi\left(\tilde{\mathcal{G}}_m^s\right)\right] \right| = 0. \tag{4.21}$$

The above is the main step in the proof of Theorem 4.4.2. By (4.21), we need only to compare the structure of $\tilde{\mathcal{G}}_m^{bl}$ to that of $\tilde{\mathcal{G}}_m$.

**Key step 2.**   For any $K \geqslant 1$,

$$\left| \mathbb{E}\left[ \Phi(\tilde{\mathcal{G}}_m^s) \right] - \sum_{k=0}^{K} \mathbb{E}\left[ \Phi(\tilde{\mathcal{G}}_m^s) \mathbf{1}\{ N_{(m)}^\star = k \} \right] \right| \leqslant \| \phi \|_\infty \mathbb{P}\left( N_{(m)}^\star \geqslant K+1 \right),$$

and the same inequality also holds for $\tilde{\mathcal{G}}_m^{\text{bl},s}$. Thus, using Lemma 4.4.3, the proof of (4.21) reduces to showing that, for each fixed $k \geqslant 1$,

$$\lim_{m \to \infty} \left| \mathbb{E}\left[ \Phi(\tilde{\mathcal{G}}_m^{\text{bl},s}) \mathbf{1}\{ N_{(m)}^\star = k \} \right] - \mathbb{E}\left[ \Phi(\tilde{\mathcal{G}}_m^s) \mathbf{1}\{ N_{(m)}^\star = k \} \right] \right| = 0. \tag{4.22}$$

**Main aim of this section.**   Below, we define a function $g_\phi^k(\cdot)$ on the space $\overline{T}_{IJ}^*$ which captures the behavior of pairwise distances after creating $k$ surplus edges. Under Assumption 4.4, we show that the introduction of blobs changes the distances within the tilted $\mathbf{p}$-trees and the $g_\phi^k$ values negligibly. This completes the proof of (4.22).

For any fixed $k \geqslant 0$, consider $\mathbf{t} \in T_{I,(k+l)}^*$ with root $0+$, leaves $\mathbf{i} = (1+, \ldots, (k+l)+)$ and root-to-leaf measures $\nu_{\mathbf{t},i}$ on the path $[0+, i+]$ for all $1 \leqslant i \leqslant k+l$. We create a graph $G(\mathbf{t})$ by sampling, for each $1 \leqslant i \leqslant k$, points $i(s)$ on $[0+, i+]$ according $\nu_{\mathbf{t},i}$ and connecting $i+$ with $i(s)$. Let $d_{G(\mathbf{t})}$ denote the distance on $G(\mathbf{t})$ given by the sum of edge lengths in the shortest path. Then, the function $g_\phi^k : T_{IJ}^* \mapsto \mathbb{R}$ is defined as

$$g_\phi^k(\mathbf{t}) = \mathbb{E}\left[ \phi\big( d_{G(\mathbf{t})}(i+, j+) : k+1 \leqslant i, j \leqslant k+l \big) \right] \mathbf{1}\{ \mathbf{t} \neq \partial \}, \tag{4.23a}$$

where $\partial$ is a forbidden state defined as follows: Given any $\mathbf{t} \in T_{IJ}^*$, and a set of vertices $\mathbf{v} = (v_1, \ldots, v_r)$, we denote the subtree of $\mathbf{t}$ spanned by $\mathbf{v}$ with $\mathbf{t}(\mathbf{v})$. We declare $\mathbf{t}(\mathbf{v}) = \partial$ if either two vertices in $\mathbf{v}$ are the same or one of them is an ancestor of another vertex in $\mathbf{v}$. Thus, if $\mathbf{t}(\mathbf{v}) \neq \partial$, the tree $\mathbf{t}(\mathbf{v})$ necessarily has $r$ leaves. Notice that the expectation in (4.23a) is over the choices of $i(s)$-values only. In our context, $\mathbf{t}$ is always considered as a subgraph of the graph on vertex set $[m]$ and thus we assume that $\mathbf{t}$ has inherited the labels from the corresponding graph. Thus $\mathbf{t} \in T_{I,(k+l)}^{*m}$. There is a natural way to extend $g_\phi^k(\cdot)$ to $\overline{T}_{IJ}^{*m}$ as follows: Consider $\overline{\mathbf{t}} \in \overline{T}_{IJ}^{*m}$ and the corresponding $\mathbf{t} \in T_{IJ}^{*m}$ (see Section 4.3.3). Let $0+$, $\mathbf{i}$, $(\nu_{\mathbf{t},i})_{i \in [k+l]}$ and $(i(s))_{i \in [k+l]}$ be as defined above. Let $\bar{G}(\bar{\mathbf{t}})$ denote the metric space by introducing an edge of length one between $X_{i+i(s)}$ and $X_{i(s)i+}$, where $X_{ij}$ has distribution $\mu_i$ for all $j \geqslant 1$, independently of each other and other shortcuts. For $k+1 \leqslant i \leqslant k+l$, $X_i \in M_{x_i}$ have distribution $\mu_{x_i}$ independently for all $i \geqslant 1$. Let $\bar{d}_{\bar{G}(\bar{\mathbf{t}})}$ denote

the distance on $\bar{G}(\bar{\mathbf{t}})$. Then, let

$$g_\phi^k(\bar{\mathbf{t}}) = \mathbb{E}\big[\phi\big(\bar{d}_{\bar{G}(\bar{\mathbf{t}})}(X_i, X_j) : k+1 \leqslant i, j \leqslant k+l\big)\big]\mathbf{1}\{\mathbf{t} \neq \partial\}, \qquad (4.23b)$$

where the expectation is taken over the collection of random variables $X_{i+i(s)}$ and $X_{i(s)i+}$. At this moment, we urge the reader to recall the construction in Algorithm 4.2, Lemma 4.3.3 and all the associated notations. Now, conditional on $\mathscr{T}_m^{\mathbf{p},\star}$, we can construct the tree $\mathscr{T}_m^{\mathbf{p},\star}(\tilde{\mathbf{V}}_m^{k,k+l})$ where

(a) $\tilde{\mathbf{V}}_m^{k,k+l} = (\tilde{V}_1^m, \ldots, \tilde{V}_k^m, V_{k+1}^m, \ldots, V_{k+l}^m)$ is an independent collection of vertices;

(b) $\tilde{V}_i^m$ is distributed as $\mathcal{J}^{(m)}(\cdot)$, for $1 \leqslant i \leqslant k$ and $V_i^m$ is distributed as $\mathbf{p}$, for $k+1 \leqslant i \leqslant k+l$.

Note that, by [36, (4.30)], $\lim_{m\to\infty} \mathbb{P}(\mathscr{T}_m^{\mathbf{p},\star}(\tilde{\mathbf{V}}_m^{k,k+l}) = \partial) = 0$. Now, whenever $\mathscr{T}_m^{\mathbf{p},\star}(\tilde{\mathbf{V}}_m^{k,k+l}) \neq \partial$, $\mathscr{T}_m^{\mathbf{p},\star}(\tilde{\mathbf{V}}_m^{k,k+l})$ can be considered as an element of $T_{I,k+l}^{*m}$ using the leaf-weights $(\mathfrak{G}_{(m)}(\tilde{V}_i))_{i=1}^k$, $(\mathfrak{G}_{(m)}(V_i))_{i=k+1}^{k+l}$ and root-to-leaf measures given by $(Q_{\tilde{V}_i}^m(\cdot))_{i=1}^k$, $(Q_{V_i}^m(\cdot))_{i=k+1}^{k+l}$. Let $\bar{\mathscr{T}}_m^{\mathbf{p},\star}(\tilde{\mathbf{V}}_m^{k,k+l})$ denote the element corresponding to $\mathscr{T}_m^{\mathbf{p},\star}(\tilde{\mathbf{V}}_m^{k,k+l})$ with blobs. Thus, $\bar{\mathscr{T}}_m^{\mathbf{p},\star}(\tilde{\mathbf{V}}_m^{k,k+l})$ is viewed as an element of $\overline{T}_{IJ}^{*m}$. Let $\mathbf{V}_m = (V_1, \ldots, V_{k+l})$ be an i.i.d. collection of random variables with distribution $\mathbf{p}$. Let $\mathbb{E}_{\mathbf{p},\star}$ denote the expecation conditionally on $\mathscr{T}_m^{\mathbf{p},\star}$ and $N_{(m)}^\star$. The proof of (4.22) now reduces to

$$\left| \mathbb{E}\big[\Phi\big(\tilde{\mathcal{G}}_m^{bl,s}\big)\mathbf{1}\{N_{(m)}^\star = k\}\big] - \mathbb{E}\big[\Phi(\tilde{\mathcal{G}}_m^s)\mathbf{1}\{N_{(m)}^\star = k\}\big] \right|$$

$$= \left| \mathbb{E}\left[\mathbb{E}_{\mathbf{p},\star}\left[g_\phi^k\left(\frac{\sigma(\mathbf{p})}{B_m+1}\bar{\mathscr{T}}_m^{\mathbf{p},\star}(\tilde{\mathbf{V}}_m^{k,k+l})\right)\right]\mathbf{1}\{N_{(m)}^\star = k\}\right] \right. \qquad (4.24)$$

$$\left. - \mathbb{E}\left[\mathbb{E}_{\mathbf{p},\star}\big[g_\phi^k\big(\sigma(\mathbf{p})\mathscr{T}_m^{\mathbf{p},\star}(\tilde{\mathbf{V}}_m^{k,k+l})\big)\big]\mathbf{1}\{N_{(m)}^\star = k\}\right] \right| + o(1).$$

Notice that the tilting does not affect the blobs themselves but only the superstructure. Recall also the definition of the tilting function $L(\cdot)$ from (4.16). Using the fact that $\mathcal{J}^{(m)}(v) \propto p_v \mathfrak{G}_{(m)}(v)$,

$$\mathbb{E}_{\mathbf{p},\star}\big[g_\phi^k\big(\sigma(\mathbf{p})\mathscr{T}_m^{\mathbf{p},\star}(\tilde{\mathbf{V}}_m^{k,k+l})\big)\big] = \frac{\mathbb{E}_{\mathbf{p},\star}\big[\prod_{i=1}^k \mathfrak{G}_{(m)}(V_i)g_\phi^k\big(\sigma(\mathbf{p})\mathscr{T}_m^{\mathbf{p},\star}(\mathbf{V}_m)\big)\big]}{\big(\mathbb{E}_{\mathbf{p},\star}[\mathfrak{G}_{(m)}(V_1)]\big)^k},$$

and an identical expression holds by replacing $\sigma(\mathbf{p})\mathscr{T}_m^{\mathbf{p},\star}$ by $\frac{\sigma(\mathbf{p})}{B_m+1}\bar{\mathscr{T}}_m^{\mathbf{p}}$. Denote the expectation conditionally on $\mathscr{T}_m^{\mathbf{p}}$ and $N_{(m)}$ by $\mathbb{E}_{\mathbf{p}}$ and simply write

$\bar{\mathscr{T}}_m^{\mathbf{p},s}$, $\mathscr{T}_m^{\mathbf{p},s}$ for $\frac{\sigma(\mathbf{p})}{B_m+1}\bar{\mathscr{T}}_m^{\mathbf{p}}(\mathbf{V}_m)$, $\sigma(\mathbf{p})\mathscr{T}_m^{\mathbf{p}}(\mathbf{V}_m)$ respectively. Now, (4.24) simplifies to

$$\left| \mathbb{E}\big[\Phi\big(\tilde{\mathcal{G}}_m^{\mathrm{bl},s}\big)\mathbf{1}\{N_{(m)}^\star = k\}\big] - \mathbb{E}\big[\Phi\big(\tilde{\mathcal{G}}_m^s\big)\mathbf{1}\{N_{(m)}^\star = k\}\big] \right|$$

$$\leqslant \frac{1}{\mathbb{E}\left[L(\mathscr{T}_m^{\mathbf{p}})\right]}\left| \mathbb{E}\left[ \frac{\mathbb{E}_{\mathbf{p}}\big[\prod_{i=1}^k \mathfrak{G}_{(m)}(V_i)g_\phi^k\big(\bar{\mathscr{T}}_m^{\mathbf{p},s}\big)\big]}{\big(\mathbb{E}_{\mathbf{p}}[\mathfrak{G}_{(m)}(V_1)]\big)^k}L(\mathscr{T}_m^{\mathbf{p}})\mathbf{1}\big\{N_{(m)} = k\big\} \right] \right.$$

$$\left. - \mathbb{E}\left[ \frac{\mathbb{E}_{\mathbf{p}}\big[\prod_{i=1}^k \mathfrak{G}_{(m)}(V_i)g_\phi^k\big(\mathscr{T}_m^{\mathbf{p},s}\big)\big]}{\big(\mathbb{E}_{\mathbf{p}}[\mathfrak{G}_{(m)}(V_1)]\big)^k}L(\mathscr{T}_m^{\mathbf{p}})\mathbf{1}\big\{N_{(m)} = k\big\} \right] \right|. \qquad (4.25)$$

**Proposition 4.4.4.** *As $m \to \infty$, $\big|g_\phi^k\big(\bar{\mathscr{T}}_m^{\mathbf{p},s}\big) - g_\phi^k\big(\mathscr{T}_m^{\mathbf{p},s}\big)\big| \xrightarrow{\mathbb{P}} 0$.*

We first show that it is enough to prove Proposition 4.4.4 to complete the proof of (4.25), but before that we first need to state some results. The proof of Proposition 4.4.4 is deferred till the end of this section.

**Lemma 4.4.5** ([36, Proposition 4.8, Theorem 4.15]). *$(L(\mathscr{T}_m^{\mathbf{p}}))_{m\geqslant 1}$ is uniformly integrable. Also, for each $k \geqslant 0$, the quantity*

$$\left( \mathbb{E}_{\mathbf{p}}\left[ \frac{\mathfrak{G}_{(m)}(V_1^{(m)})}{\sigma(\mathbf{p})} \right], \ \mathbb{E}_{\mathbf{p}}\left[ \left(\prod_{i=1}^k \frac{\mathfrak{G}_{(m)}(V_i^{(m)})}{\sigma(\mathbf{p})}\right)g_\phi^k\big(\mathscr{T}_m^{\mathbf{p},s}\big) \right] \right)$$

*converges in distribution to some random variable.*

**Fact 1.** *Consider three sequences of random variables $(X_m)_{m\geqslant 1}$, $(Y_m)_{m\geqslant 1}$ and $(Y'_m)_{m\geqslant 1}$ with (i) $(X_m)_{m\geqslant 1}$ is uniformly integrable, (ii) $(Y_m)_{m\geqslant 1}$ and $(Y'_m)_{m\geqslant 1}$ are almost surely bounded and (iii) $Y_m - Y'_m \xrightarrow{\mathbb{P}} 0$. Then, as $m \to \infty$,*

$$\mathbb{E}\left[|X_m Y_m - X_m Y'_m|\right] \to 0.$$

**Fact 2.** *Suppose that $(X_m)_{m\geqslant 1}$ is a sequence of random variables such that for every $m \geqslant 1$, there exists a further sequence $(X_{m,r})_{r\geqslant 1}$ satisfying (i) for each fixed $m \geqslant 1$, $X_{m,r} \xrightarrow{\mathbb{P}} 0$ as $r \to \infty$, and (ii) $\lim_{r\to\infty}\limsup_{m\to\infty}\mathbb{P}(|X_m - X_{m,r}| > \varepsilon) = 0$ for any $\varepsilon > 0$. Then $X_m \xrightarrow{\mathbb{P}} 0$ as $m \to \infty$.*

*Proof of (4.25) from Proposition 4.4.4.* By Lemma 4.4.5 and Fact 1, the proof of (4.25) reduces to showing

$$\mathbb{E}_{\mathbf{p}}\left[ \left(\prod_{i=1}^k \frac{\mathfrak{G}_{(m)}(V_i^{(m)})}{\sigma(\mathbf{p})}\right)\big(g_\phi^k\big(\bar{\mathscr{T}}_m^{\mathbf{p},s}\big) - g_\phi^k\big(\mathscr{T}_m^{\mathbf{p},s}\big)\big) \right] \xrightarrow{\mathbb{P}} 0. \qquad (4.26)$$

Let $X_m$ denote the term inside the expectation in (4.26). Further, sample the set of leaves $\mathbf{V}_m$ independently $r$ times on the same tree $\mathscr{T}_m^{\mathbf{p}}$ and let $X_m^i$

denote the observed value of $X_m$ in the $i$-th sample. Further, let $X_{m,r} = r^{-1} \sum_{i=1}^{r} X_m^i$. Obviously, condition (i) in Fact 2 is satisfied due to Proposition 4.4.4. To verify condition (ii), note that $\mathbb{E}_{\mathbf{p}}(X_{m,r}) = X_m$ and therefore Chebyshev's inequality yields

$$\mathbb{P}\left(|X_m - X_{m,r}| > \varepsilon\right) \leqslant \frac{\mathbb{E}[X_m^2]}{\varepsilon^2 r} \leqslant \frac{4\|\phi\|_\infty^2}{\varepsilon^2 r} \frac{\mathbb{E}\left[\mathfrak{G}_{(m)}(V_1^m)^{2k}\right]}{\sigma(\mathbf{p})^{2k}}.$$

The final term is uniformly bounded over $m \geqslant 1$ and vanishes as $r \to \infty$. For an interested reader, using the notation of [36], the above is a consequence of the fact that $\|\mathfrak{G}_{(m)}\|_\infty \leqslant \|F^{\mathrm{exc},\mathbf{p}}\|_\infty$ and [36, Lemma 4.9]. $\qquad \square$

In this section, we will use the notion of Gromov-Hausdorff-Prokhorov topology on the collection of $(X, d, \mu)$, where $(X, d)$ is a compact metric space and $\mu$ is a probability measure on the corresponding Borel sigma algebra. Without re-defining all the required notions, we refer the reader to [36, Section 2.1.1]. We further recall the notation dis for distortion and $D(\mu; \mu_1, \mu_2)$ for discrepancy of measures as defined in [36, Section 2.1.1]. Denote the root of $\mathscr{T}_m^{\mathbf{P}}(\mathbf{V}_m)$ by $0+$ and the $j$th leaf by $j+$. Let $\mathscr{M}_j^m := \{[0+, j+], d, \nu_j\}$ be the random measured metric space with the corresponding root-to-leaf measure $\nu_j$. Let $\bar{\mathscr{M}}_j^m := \{\bar{M}_j, \bar{d}, \bar{\nu}_j\}$ be the measured metric space with $\bar{M}_j := \sqcup_{i \in [0+, j+]} M_i$ and the induced root-to-leaf measure $\bar{\nu}_j(A) = \sum_{i \in [0+, j+]} \nu_j(\{i\}) \mu_i(A \cap M_i)$. For convenience, we have suppressed the dependence on $\mathscr{T}_m^{\mathbf{P}}(\mathbf{V}_m)$ in the notation. $\mathscr{M}_j^m$ is coupled to $\bar{\mathscr{M}}_j^m$ in the obvious way that the superstructure of $\bar{\mathscr{M}}_j^m$ is given by $\mathscr{M}_j^m$. We need the following lemma to prove Proposition 4.4.4:

**Lemma 4.4.6.** *For $j \geqslant 1$, as $m \to \infty$, $d_{\mathrm{GHP}}\left(\sigma(\mathbf{p}) \mathscr{M}_j^m, \frac{\sigma(\mathbf{p})}{B_m + 1} \bar{\mathscr{M}}_j^m\right) \xrightarrow{\mathbb{P}} 0$.*

*Proof.* We prove this for $j = 1$ only. The proof for $j \geqslant 2$ is identical. For $x \in \bar{M}_1$, we denote its corresponding vertex label by $i(x)$, i.e., $i(x) = k$ iff $x \in M_k$. Consider the correspondence $C_m$ and the measure $\mathfrak{m}$ on the product space $[0+, 1+] \times \bar{M}_1$ defined as

$$C_m := \{(i, x) : i \in [0+, 1+], x \in M_i\}, \quad \mathfrak{m}(\{i\} \times A) = \nu_1(\{i\}) \bar{\nu}_1(A \cap M_i). \quad (4.27)$$

Note that the discrepancy of $\mathfrak{m}$ satisfies $D(\mathfrak{m}; \nu_1, \bar{\nu}_1) = 0$. Further, $\mathfrak{m}(C_m^c) = 0$. Therefore, Lemma 4.4.6 follows if we can prove that

$$\mathrm{dis}(C_m) := \sup_{x,y \in \bar{M}_1} \left\{\sigma(\mathbf{p}) d(i(x), i(y)) - \frac{\sigma(\mathbf{p})}{B_m + 1} \bar{d}(x, y)\right\} \xrightarrow{\mathbb{P}} 0. \quad (4.28)$$

To simplify the expression for $\operatorname{dis}(C_m)$, suppose that $i(x)$ is an ancestor of $i(y)$ on the path from $0+$ to $1+$. Then,

$$d(i(x), i(y)) = d(0+, i(y)) - d(0+, i(x)),$$

$$\bar{d}(x_0, y) - \bar{d}(x_0, x) \leqslant \bar{d}(x, y) \leqslant \bar{d}(x_0, y) - \bar{d}(x_0, x) + 2\Delta_{\max},$$

for any $x_0 \in M_{0+}$. This implies that

$$
\begin{aligned}
\sup_{x,y \in \bar{M}_1} & \left\{ \sigma(\mathbf{p}) d(i(x), i(y)) - \frac{\sigma(\mathbf{p})}{B_m + 1} \bar{d}(x, y) \right\} \\
& \leqslant 2 \sup_{y \in \bar{M}_1} \left\{ \sigma(\mathbf{p}) d(0+, i(y)) - \frac{\sigma(\mathbf{p})}{B_m + 1} \bar{d}(x_0, y) \right\} + \frac{2\sigma(\mathbf{p})\Delta_{\max}}{B_m + 1}.
\end{aligned}
\tag{4.29}
$$

Further, replacing $y$ by any other point $y'$ in the right hand side in (4.29) incurs an error of at most $\Delta_{\max}$. Now, write the path $[0+, 1+]$ as $0+ = i_0 \to i_1 \to \cdots \to i_{R^*-2} \to i_{R^*-1} = 1+$. Then

$$
\operatorname{dis}(C_m) \leqslant 2 \sup_{k \leqslant R^*-1} \left| \sigma(\mathbf{p}) d(i_0, i_k) - \frac{\sigma(\mathbf{p})}{B_m + 1} \bar{d}(X_{i_0, i_1}, X_{i_k, i_{k+1}}) \right| + \frac{6\sigma(\mathbf{p})\Delta_{\max}}{B_m + 1},
\tag{4.30}
$$

where $(X_{i,j})_{i,j \in [m]}$ are the junction-points. Using Assumption 4.4 and (4.30), it is now enough to show that for any $\varepsilon > 0$,

$$
\lim_{m \to \infty} \mathbb{P}\left( \sup_{k \leqslant R^*-1} \left| \sigma(\mathbf{p}) d(i_0, i_k) - \frac{\sigma(\mathbf{p})}{B_m + 1} d(X_{i_0, i_1}, X_{i_k, i_{k+1}}) \right| > \varepsilon \right) = 0.
$$

Denote the term inside the above supremum by $Q_k$. Then,

$$
\begin{aligned}
Q_k & := \left[ \sigma(\mathbf{p}) d(i_0, i_k) - \frac{\sigma(\mathbf{p})}{B_m + 1} \bar{d}(X_{i_0, i_1}, X_{i_k, i_{k+1}}) \right] \\
& = \left[ \sigma(\mathbf{p}) k - \frac{\sigma(\mathbf{p})}{B_m + 1} \left( k + \sum_{j=1}^{k} d_{i_j}(X_{i_j, i_{j-1}}, X_{i_j, i_{j+1}}) \right) \right] \\
& = \frac{\sigma(\mathbf{p})}{B_m + 1} \left[ \sum_{j=1}^{k} \left( B_m - d_{i_j}(X_{i_j, i_{j-1}}, X_{i_j, i_{j+1}}) \right) \right].
\end{aligned}
\tag{4.31}
$$

Recall the construction of the path $[0+, 1+]$ via the birthday problem from Section 4.3.4. Take $\mathbf{J} := (J_i)_{i \geqslant 1}$ such that $J_i$ are i.i.d. samples from $\mathbf{p}$. Further let $\xi := (\xi_i)_{i \in [m]}$ be an independent sequence such that $\xi_i$ is the distance between two points, chosen randomly from $M_i$ according to $\mu_i$. Further, let $\mathbf{J}$ and $\xi$ be independent. Then $R^*$ can be thought of as the first repeat time of the sequence $\mathbf{J}$. Thus, $Q_k$ in (4.31) has the same distribution as

$$
\hat{Q}_k := \frac{\sigma(\mathbf{p})}{B_m + 1} \sum_{i=1}^{k} (B_m - \xi_{J_i}).
$$

From the birthday construction $\mathbb{E}\left[\xi_{J_i}\right] = \sum_{i\in[m]} p_i u_i = B_m$ and $(\xi_{J_i})_{i\geqslant 1}$ is an independent sequence. Therefore, $(\hat{Q}_k)_{k\geqslant 0}$ is a martingale. Further,

$$\text{Var}(\hat{Q}_k) \leqslant \left(\frac{\sigma(\mathbf{p})}{B_m+1}\right)^2 k\Delta_{\max} \sum_{i\in[m]} p_i u_i = \frac{\sigma(\mathbf{p})^2 k\Delta_{\max} B_m}{(B_m+1)^2}.$$

Thus, by Doob's martingale inequality [147, Lemma 2.54.5], it follows that, for any $\varepsilon > 0$ and $T > 0$,

$$\mathbb{P}\left(\sup_{k\leqslant T} |\hat{Q}_k| > \varepsilon\right) \leqslant \frac{T\sigma(\mathbf{p})^2\Delta_{\max} B_m}{(B_m+1)^2\varepsilon^2}.$$

Recall from [59, Theorem 4] that $(\sigma(\mathbf{p})R^*)_{m\geqslant 1}$ is a tight sequence of random variables. The proof now follows using Assumption 4.4. □

*Proof of Proposition 4.4.4 using Lemma 4.4.6.* We use objects defined in (4.27), (4.28) in the proof of Lemma 4.4.6 for all the path metric spaces with $j \leqslant k$. We assume that we are working on a probability space such that the convergence (4.28) holds almost surely for all $j \leqslant k$. To summarize, for fixed $\varepsilon > 0$ and for each $j \leqslant k$, we can choose the correspondence $C_m^j$ and a measure $\mathfrak{m}_j$ of $[0+,j+] \times \bar{M}_j$ satisfying (i) $(i, X_{ik}) \in C_m^j$, for all $i, k \in [0+,j+]$, (ii) $\text{dis}(C_m^j) < \varepsilon/2k$ almost surely, and (iii) $D(\mathfrak{m}_j; \nu_j, \bar{\nu}_j) = 0$ and $\mathfrak{m}_j((C_m^j)^c) = 0$. Recall the definitions of the function $g_\phi^k$ from (4.23a), (4.23b) and the associated graphs $G(\cdot)$, $\bar{G}(\cdot)$. We simply write $G$ and $\bar{G}$ for $G(\sigma(\mathbf{p})\mathscr{T}_m^\mathbf{P}(\mathbf{V}_m))$ and $\bar{G}(\frac{\sigma(\mathbf{p})}{B_m+1}\bar{\mathscr{T}}_m^\mathbf{P}(\mathbf{V}_m))$, respectively. Let $\mathfrak{m}^{\otimes k}$ denote the k-fold product measure of $\mathfrak{m}_j$ for $j \leqslant k$. We denote the graph distance on a graph $H$ by $d_H$. Note that

$$\left|g_\phi^k\left(\sigma(\mathbf{p})\mathscr{T}_m^\mathbf{P}(\mathbf{V}_m)\right) - g_\phi^k\left(\frac{\sigma(\mathbf{p})}{B_m+1}\bar{\mathscr{T}}_m^\mathbf{P}(\mathbf{V}_m)\right)\right|$$
$$\leqslant \mathbb{E}\left[\left|\phi\left((d_G(i+,j+))_{i,j=k+1}^{k+l}\right) - \phi\left((d_G(X_i,X_j))_{i,j=k+1}^{k+l}\right)\right|\right], \tag{4.32}$$

where $X_i \sim \mu_i$ independently for $i \in [m]$, and the above expectation is with respect to the measure $\mathfrak{m}^{\otimes k}$. Recall the notation while defining $g_\phi^k(\cdot)$ in (4.23a), (4.23b). Notice that for any point $k \in [0+,i+]$ and $x_k \in M_k$ and $x_{i(s)} \in M_{i(s)}$,

$$|d_\mathfrak{t}(k, i(s)) - d_\mathfrak{t}(x_k, x_{i(s)})| \leqslant \frac{\varepsilon}{2k}. \tag{4.33}$$

Now, for any path $i+$ to $j+$ in $G$, we can essentially take the same path from $X_i$ to $X_j$ in $\bar{G}$ and take the corresponding inter-blob paths on the way. The distance traversed in $\bar{G}$ in this way gives an upper bound on $d_{\bar{G}}(X_i, X_j)$. Notice that, by (4.33), taking a shortcut contributes at most $\varepsilon/2k$ to the difference of the distance traveled in $G$ and $\bar{G}$. Also, traversing a shortcut edge contributes

$\sigma(\mathbf{p})B_m/(B_m + 1)$ and there are at most $k$ shortcuts on the path. Furthermore, it may be required to reach the relevant junction points from $X_i$ and $X_j$ and that contributes at most $2\sigma(\mathbf{p})\Delta_{\max}/(B_m + 1)$. Thus, for $k + 1 \leqslant i, j \leqslant k + l$, and sufficiently large $m$,

$$d_{\tilde{G}}(X_i, X_j) \leqslant d_G(i+, j+) + \frac{\varepsilon}{2} + \frac{k\sigma(\mathbf{p})B_m}{B_m + 1} + \frac{2\sigma(\mathbf{p})\Delta_{\max}}{B_m + 1} \leqslant d_G(i+, j+) + \varepsilon.$$

By symmetry we can conclude the lower bound also and the continuity of $\phi(\cdot)$ (see [36, Theorem 4.18]) along with (4.32) completes the proof of Proposition 4.4.4. $\qquad\square$

At this moment, we urge the reader to recall the definitions from (4.4), (4.5) and (4.6). The configuration model graphs considered in this section will be assumed to have degree sequence $\mathbf{d}'$ and the vertices have an associated weight sequence $w$ such that Assumption 4.2 is satisfied. We treat the different terms arising in Theorem 4.1.3 in different subsections.

## 4.4.2  Analysis of $s_2^\star$

The asymptotics of $s_2^\star$ is a consequence of the Chebyshev inequality. In the following lemma, we compute its mean and variance. Consider the size-biased distribution on the vertex set $[n]$ with sizes $(w_i)_{i\in[n]}$. Let $V_n$ and $V_n^*$, respectively, denote a vertex chosen uniformly at random and according to the size-biased distribution, independently of the underlying graph $CM_n(\mathbf{d}')$. Let $D_n'$, $W_n$ (respectively $D_n^*$, $W_n^*$) denote the degree and weight of $V_n$ (respectively $V_n^*$). For a vertex $v \in [n]$, let $\mathscr{W}(v) := \sum_{k \in \mathscr{C}'(v)} w_k$, where $\mathscr{C}'(v)$ denotes the component of $CM_n(\mathbf{d}')$ containing $v$.

**Lemma 4.4.7.** *Under Assumption 4.2, (i)* $\mathbb{E}\left[\mathscr{W}(V_n^*)\right] = \frac{\mathbb{E}[D_n^*]\mathbb{E}[D_n'W_n]}{\mathbb{E}[D_n'](1-\nu_n')}(1 + o(1))$, *(ii)* $\mathbb{E}\left[(\mathscr{W}(V_n^*))^2\right] = O\left(\frac{n^{3\alpha-1}}{(1-\nu_n')^3}\right)$, *and* $\mathbb{E}\left[(\mathscr{W}(V_n^*))^3\right] = o(n^{1+2\delta})$.

*Asymptotics of $s_2^\star$.* Denote $\ell_n^w = \sum_{i\in[n]} w_i$. Firstly, if $\mathbb{E}_{\mathbf{d}'}$ denotes the conditional expectation given $CM_n(\mathbf{d}')$, then for any $r \geqslant 1$,

$$\mathbb{E}_{\mathbf{d}'}\left[(\mathscr{W}(V_n^*))^{r-1}\right] = \sum_{i\geqslant 1}\sum_{k\in\mathscr{C}'_{(i)}} \frac{w_k}{\ell_n^w}\left(\sum_{k\in\mathscr{C}'_{(i)}} w_k\right)^{r-1} = \frac{1}{\ell_n^w}\sum_{i\geqslant 1}(\mathscr{W}_{(i)})^r. \quad (4.34)$$

Therefore, using Lemma 4.4.7 and (4.4), it follows from Assumption 4.2 that

$$n^{-\delta}\mathbb{E}\left[s_2^\star\right] = \frac{\ell_n^w}{n}n^{-\delta}\mathbb{E}\left[\mathscr{W}(V_n^*)\right] \to \frac{\mu_{d,w}^2}{\mu_d\lambda_0}, \quad (4.35)$$

where we have used the fact that $\mathbb{E}\left[D_n^*\right] \to \mu_{d,w}/\mu_w$. It remains to compute the variance. Let $U_n^*$ denote another vertex chosen in a size-biased way with the sizes being $(w_i)_{i \in [n]}$, independently of the graph $CM_n(\mathbf{d}')$ and $V_n^*$. Then (4.34) yields

$$
\begin{aligned}
\mathbb{E}\left[(s_2^\star)^2\right] &= \frac{1}{n^2}\mathbb{E}\left[\sum_{i,j \geqslant 1} \mathscr{W}_{(i)}^2 \mathscr{W}_{(j)}^2\right] = \frac{1}{n^2}\mathbb{E}\left[\sum_{i \geqslant 1} \mathscr{W}_{(i)}^4\right] + \frac{1}{n^2}\mathbb{E}\left[\sum_{i \neq j} \mathscr{W}_{(i)}^2 \mathscr{W}_{(j)}^2\right] \\
&= \frac{\ell_n^w}{n}\frac{1}{n}\mathbb{E}\left[\left(\mathscr{W}(V_n^*)\right)^3\right] + \left(\frac{\ell_n^w}{n}\right)^2 \mathbb{E}\left[\mathscr{W}(U_n^*)\mathscr{W}(V_n^*)\mathbf{1}\{U_n^* \notin \mathscr{C}'(V_n^*)\}\right],
\end{aligned}
$$

$$(4.36)$$

where the second term in the third equality follows using similar arguments as in (4.34). Denote the last two terms of (4.36) by $(\mathbf{I})$ and $(\mathbf{II})$ respectively. To estimate $(\mathbf{II})$, observe that, conditionally on the graph $\mathscr{C}'(V_n^*)$, the graph obtained by removing $\mathscr{C}'(V_n^*)$ from $CM_n(\mathbf{d}')$ is again a configuration model with the induced degree sequence $\tilde{\mathbf{d}}$ and number of vertices $\tilde{n}$. Let $\tilde{v}_n$ denote the corresponding criticality parameter. In the proof of Lemma 4.4.7 (i), we will see that the upper bound holds whenever $\tilde{v}_n < 1$ (see Remark 33). Thus, let us first show that

for all sufficiently large $n$, $\mathbb{P}\left(\tilde{v}_n < 1 \mid \mathscr{C}'(V_n^*)\right) = 1$, almost surely. (4.37)

Denote $\ell_n' = \sum_{i \in [n]} d_i'$. To see (4.37), first notice that

$$
\tilde{v}_n - 1 = v_n' - 1 - \frac{\sum_{j \in \mathscr{C}'(V_n^*)} d_j'(d_j' - 2)}{\sum_{j \in [n]} d_j'} + (\tilde{v}_n - 1)\frac{\sum_{j \in \mathscr{C}'(V_n^*)} d_j'}{\ell_n'}.
$$

Moreover, for any connected graph $\mathcal{G}$, $\sum_{i \in \mathcal{G}} d_i'(d_i' - 2) \geqslant -2$ (this can be proved by induction) so that $(\tilde{v}_n - 1)\left(1 - \sum_{j \in \mathscr{C}'(V_n^*)} d_j'/\ell_n'\right) \leqslant v_n' - 1 + \frac{2}{\ell_n'}$. The proof of (4.37) now follows. As mentioned above, now we can apply the upper bound from Lemma 4.4.7. Therefore,

$$
\begin{aligned}
&\mathbb{E}\left[\mathscr{W}(U_n^*)\mathbf{1}\{U_n^* \notin \mathscr{C}'(V_n^*)\}\big|\mathscr{C}'(V_n^*)\right] \\
&= \frac{\sum_{i \notin \mathscr{C}'(V_n^*)} w_i}{\ell_n^w}\mathbb{E}\left[\mathscr{W}(U_n^*)\big|\mathscr{C}'(V_n^*), U_n^* \notin \mathscr{C}'(V_n^*)\right] \\
&\leqslant \frac{\left(\sum_{i \in [n]} d_i' w_i\right)^2}{\ell_n^w \ell_n'(1 - v_n' - 2/\ell_n')} = \mathbb{E}\left[\mathscr{W}(V_n^*)\right]\left(1 + O(1/n)\right).
\end{aligned}
$$

$$(4.38)$$

Thus,

$$
\mathbb{E}\left[\mathscr{W}(U_n^*)\mathscr{W}(V_n^*)\mathbf{1}\{U_n^* \notin \mathscr{C}'(V_n^*)\}\right] \leqslant \left(\mathbb{E}\left[\mathscr{W}(V_n^*)\right]\right)^2\left(1 + O(1/n)\right). \quad (4.39)
$$

We conclude that (4.36), (4.39) together with Lemma 4.4.7 implies that $\mathrm{Var}\left(s_2^\star\right) = o(n^{2\delta})$. Thus, we can use the Chebyshev inequality and (4.35) to conclude that

$$n^{-\delta}s_2^\star \xrightarrow{\mathbb{P}} \mu_{d,w}^2/(\mu_d\lambda_0).$$

$\square$

*Proof of Lemma 4.4.7* (i). We use path-counting techniques for configuration models from [111, Lemma 5.1]. Let $\mathcal{A}(k,l)$ denote the event that there exists a path of length $l$ from $V_n^*$ to $k$ and $\mathcal{A}'(k,l)$ the event that there exist two different paths, one of length $l$ and another one of length at most $l$, from $V_n^*$ to $k$. Notice that

$$\mathbb{E}\left[\mathcal{W}(V_n^*)\right] \leqslant \mathbb{E}\left[\sum_{k\in[n]} w_k \mathbf{1}\{V_n^* \rightsquigarrow k\}\right] = \mathbb{E}\left[W_n^*\right] + \sum_{l\geqslant 1}\sum_{k\in[n]} w_k \mathbb{P}\left(\mathcal{A}(k,l)\right),$$

$$(4.40a)$$

$$\mathbb{E}\left[\mathcal{W}(V_n^*)\right] \geqslant \sum_{l\geqslant 1}\sum_{k\in[n]} w_k \mathbb{P}\left(\mathcal{A}(k,l)\right) - \sum_{l\geqslant 1}\sum_{k\in[n]} w_k \mathbb{P}\left(\mathcal{A}'(k,l)\right). \qquad (4.40b)$$

Now, by Assumption 4.2, (4.40a) yields

$$\mathbb{E}\left[\mathcal{W}(V_n^*)\right] \leqslant \mathbb{E}\left[W_n^*\right] + \mathbb{E}\left[D_n^*\right]\sum_{l=1}^{\infty}\sum_{k\in[n]} w_k \sum_{x_i\neq x_j,\forall i\neq j} \frac{\prod_{i=1}^{l-1} d_{x_i}'(d_{x_i}'-1)d_k'}{(\ell_n'-1)\cdots(\ell_n'-2l+1)}$$

$$\leqslant \mathbb{E}\left[W_n^*\right] + \frac{\mathbb{E}\left[D_n^*\right]\mathbb{E}\left[D_n'W_n\right]}{\mathbb{E}\left[D_n'\right]-1/n}\sum_{l=1}^{\infty} v_n'^{l-1} = \frac{\mathbb{E}\left[D_n^*\right]\mathbb{E}\left[D_n'W_n\right]}{\mathbb{E}\left[D_n'\right](1-v_n')}(1+o(1)),$$

$$(4.41)$$

where in the third step, we have used the fact that

$$\sum_{x_i\neq x_j,\forall i\neq j} \frac{\prod_{i=1}^{l-1} d_{x_i}'(d_{x_i}'-1)}{(\ell_n'-1)\cdots(\ell_n'-2l+1)} \leqslant v_n'^{l-1}$$
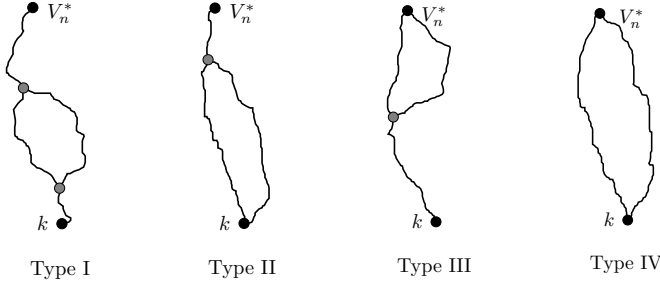
Figure 4.1: Possible paths corresponding to $\mathcal{A}'(k, l)$.

from [109, Lemma 5.1]. For the computation of the lower bound, observe that

$$
\sum_{l=1}^{n^{\eta}/\log(n)} \sum_{k \in [n]} w_k \mathbb{P}\left(\mathcal{A}(k, l)\right)
$$

$$
\geqslant \mathbb{E}[D_n^*] \sum_{l=1}^{n^{\eta}/\log(n)} \sum_{k \in [n]} w_k \sum_{x_i \neq x_j, \forall i \neq j} \frac{1}{\ell_n^{l-1}} \prod_{i=1}^{l-1} d'_{x_i}(d'_{x_i} - 1) d'_k
$$

$$
\geqslant \frac{\mathbb{E}[D_n^*]\mathbb{E}[D'_n W_n]}{\mathbb{E}[D'_n]} \sum_{l=1}^{n^{\eta}/\log(n)} \left( v_n'^{l-1} - \frac{d'_1 n^{\eta} \left( \sum_{i \in [n]} d'_i (d'_i - 1) \right)^{l-2}}{\log(n) \ell_n'^{l-1}} \right)
$$

$$
= \frac{\mathbb{E}[D_n^*] \mathbb{E}[D'_n W_n] (1 - (v_n')^{n^{\eta}/\log(n)})}{\mathbb{E}[D'_n](1 - v_n')}(1 + o(1))
$$

$$
= \frac{\mathbb{E}[D_n^*] \mathbb{E}[D'_n W_n]}{\mathbb{E}[D'_n](1 - v_n')}(1 + o(1)),
$$

$$(4.42)$$

where we have used the fact that $d'_1 l \leqslant d'_1 n^{\eta}/\log(n)$ and inclusion-exclusion to obtain the third step, and (4.4), $d'_1 n^{\eta}/\ell'_n = c_1/\mu_d(1 + o(1))$ and the fact that $\delta < \eta$ in the one-but-last step. To complete the proof of Lemma 4.4.7, we need to have an upper bound on the last term of (4.40b). Observe that if $\mathcal{A}'(k, l)$ happens, then one of the structures in Figure 4.1 occurs.

Denote by $\mathcal{A}'(k, l, i)$ the event that the structure of type $i$ ($i$=I, II, III, IV) in Figure 4.1 appears. We use the notation C to denote a generic constant. Using
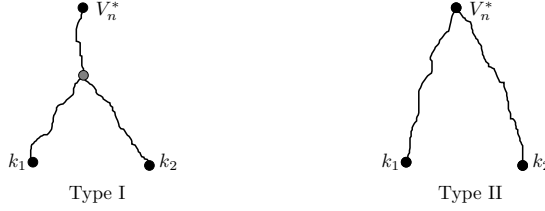
Figure 4.2: Possible paths when $V_n^* \rightsquigarrow k_1$, and $V_n^* \rightsquigarrow k_2$.

an argument identical to (4.41), and applying Assumption 4.2, it follows that

$$\sum_{l \geqslant 1} \sum_{k \in [n]} w_k \mathbb{P}\left(\mathcal{A}'(k,l,I)\right)$$

$$\leqslant C\mathbb{E}\left[D_n^*\right]\mathbb{E}\left[D_n' W_n\right] \sum_{l \geqslant 1} \sum_{r \geqslant 1} \frac{\sigma_3(n)^2}{\ell_n'}(l-1)(l-2)\nu_n'^{l+r-4}$$

$$\leqslant C\frac{n^{6\alpha-3}}{1-\nu_n'} \sum_{l \geqslant 3}(l-1)(l-2)\nu_n'^{l-3} \leqslant C\frac{n^{6\alpha-3}}{(1-\nu_n')^4} = o(n^\delta),$$

where we have used the fact $\delta < \eta$ in the last step. Identical arguments can be carried out to conclude that $\sum_{l \geqslant 1} \sum_{k \in [n]} w_k \mathbb{P}\left(\mathcal{A}'(k,l,i)\right) = o(n^\delta)$, $i = II, III, IV$. Combining this with (4.42) and applying them to (4.40b), it follows that

$$\mathbb{E}\left[\mathscr{W}(V_n^*)\right] \geqslant \frac{\mathbb{E}\left[D_n^*\right]\mathbb{E}\left[D_n' W_n\right]}{\mathbb{E}\left[D_n'\right](1-\nu_n')}(1+o(1)), \tag{4.43}$$

and the proof of Lemma 4.4.7 is now complete using (4.41). □

*Proof of Lemma 4.4.7* (ii). Notice that

$$\left(\mathscr{W}(V_n^*)\right)^r = \sum_{k_1,\dots,k_r \in [n]} w_{k_1} \cdots w_{k_r} \mathbf{1}\{V_n^* \rightsquigarrow k_1, \dots, V_n^* \rightsquigarrow k_r\}.$$

We can again count the contribution due to the different types of paths in Figure 4.2 by using similar argument as in (4.40a) to compute the second moment. Ignoring the re-computation, it follows that

$$\mathbb{E}\left[\left(\mathscr{W}(V_n^*)\right)^2\right] \leqslant \mathbb{E}\left[(W_n^*)^2\right] + \frac{\mathbb{E}\left[D_n^*\right]\left(\mathbb{E}\left[D_n' W_n\right]\right)^2 \sigma_3(n)}{(\mathbb{E}\left[D_n'\right])^3(1-\nu_n')^3}$$

$$+ \frac{\mathbb{E}\left[D_n^*(D_n^*-1)\right]\left(\mathbb{E}\left[D_n' W_n\right]\right)^2}{(\mathbb{E}\left[D_n'\right])^2(1-\nu_n')^2}, \tag{4.44}$$

which gives rise to the desired $O(\cdot)$ term. For the third moment, the leading contributions arise from the structures given in Figure 4.3. See Appendix 4.B for a detailed computation. □
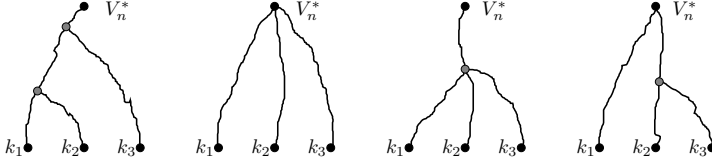
Figure 4.3: Possible paths when $V_n^* \rightsquigarrow k_1$, $V_n^* \rightsquigarrow k_2$, and $V_n^* \rightsquigarrow k_3$.

**Remark 33.** The upper bounds of $\mathbb{E}[(\mathscr{W}(V_n^*))^r]$ in (4.41) and (4.44) hold for any configuration model for which the value of the criticality parameter is less than one. The precise assumptions were needed to estimate the orders of these terms.

**Remark 34.** The method used to obtain the asymptotics of $s_2^\star$ can be followed verbatim to obtain the asymptotics of $s_{pr}^\star$. Indeed, notice that

$$\mathbb{E}\left[s_{pr}^\star\right] = \frac{1}{n}\mathbb{E}\left[\sum_{i \geqslant 1} \mathscr{W}_{(i)}|\mathscr{C}_{(i)}'|\right] = \mathbb{E}\left[\mathscr{W}(V_n)\right].$$

A similar identity for the second moment of $s_{pr}^\star$ also holds.

The main aim of this section is to prove the following proposition which will be required to obtain the asymptotics of $s_3^\star$, as well as $\mathscr{W}_{(i)}$:

**Proposition 4.4.8.** *Suppose that Assumption 4.2 holds. For any $\varepsilon > 0$,*

$$\lim_{K \to \infty} \limsup_{n \to \infty} \mathbb{P}\left(\sum_{i > K} \left(\mathscr{W}_{(i)}\right)^3 > \varepsilon n^{3(\alpha+\delta)}\right) = 0.$$

*Proof.* Let $\mathcal{G}^K$ denote the graph obtained by deleting all the edges incident to vertices $\{1, \ldots, K\}$. In this proof, a superscript $K$ to any previously defined object will correspond to the object in $\mathcal{G}^K$. Note that $\mathcal{G}^K$ is again distributed as a configuration model conditioned on the new degree sequence $\mathbf{d}^K$. Firstly, for each fixed $K$, there exists a constant $C_1 > 0$ such that

$$\nu_n^K = \frac{\sum_{i \in [n]} d_i^K(d_i^K - 1)}{\sum_{i \in [n]} d_i^K} \leqslant \frac{\sum_{i \in [n]} d_i'(d_i' - 1) - \sum_{i=1}^K d_i'(d_i' - 1)}{\ell_n' - 2\sum_{i=1}^K d_i'}$$

$$= \nu_n' - C_1 n^{2\alpha-1}\sum_{i \leqslant K} c_i^2 \leqslant \nu_n' - C_1 n^{-\delta}\sum_{i \leqslant K} c_i^2,$$

where we have used the fact that $\delta < \eta$ in the last step. Since $\nu_n^K < 1$, we can apply the upper bound in (4.44) (see Remark 33) and it follows that

$$\frac{1}{n}\mathbb{E}\left[\sum_{i \geqslant 1} \left(\mathscr{W}_{(i)}^K\right)^3\right] = \frac{\ell_n^w}{n}\mathbb{E}\left[\left(\mathscr{W}^K(V_n^*)\right)^2\right] \leqslant C\frac{n^{3(\alpha+\delta)-1}}{1 + C_1\sum_{i=1}^K c_i^2}, \qquad (4.45)$$

for some constant $C > 0$, and therefore, using the Markov inequality and the fact that $\mathbf{c} \in \ell_{\downarrow}^3 \setminus \ell_{\downarrow}^2$, it follows that, for any $\varepsilon > 0$,

$$\lim_{K \to \infty} \limsup_{n \to \infty} \mathbb{P}\left( \sum_{i \geqslant 1} \left( \mathscr{W}_{(i)}^K \right)^3 > \varepsilon n^{3(\alpha+\delta)} \right) = 0. \tag{4.46}$$

Now, the proof is complete by observing that $\sum_{i > K} (\mathscr{W}_{(i)})^3 \leqslant \sum_{i \geqslant 1} (\mathscr{W}_{(i)}^K)^3$.

$\square$

**Remark 35.** Notice that the proof of Proposition 4.4.8 can be modified to conclude the similar results for $\sum_{i > K} \left( \mathscr{W}_{(i)} \right)^2 |\mathscr{C}_{(i)}'|$ and $\sum_{i > K} \left( \mathscr{W}_{(i)} \right)^2 |\mathscr{C}_{(i)}'|$. Indeed, (4.45) can be replaced by observing that the following identities hold:

$$\mathbb{E}[\sum_{i \geqslant 1} (\mathscr{W}_{(i)}^K)^2 |\mathscr{C}_{(i)}'^K|] = n\mathbb{E}[\mathscr{W}^K(V_n)], \quad \mathbb{E}[\sum_{i \geqslant 1} \mathscr{W}_{(i)}^K (|\mathscr{C}_{(i)}'^K|)^2] = \ell_n^w \mathbb{E}[|\mathscr{C}'^K(V_n^*)|^2].$$

### 4.4.3   Barely sub-critical masses

We only prove the asymptotics of $\mathscr{W}_{(i)}$ in Theorem 4.1.3. Then the asymptotics of $s_3^\star$ follow by a direct application of Proposition 4.4.8. The idea is to obtain the asymptotics for $\mathscr{W}(j)$ for each fixed $j$. We will see that Proposition 4.4.8 implies that $\mathscr{W}_{(j)} = \mathscr{W}(j)$ with high probability. Consider the breadth-first exploration of the graph starting from vertex $j$ as follows:

**Algorithm 4.3.** The algorithm carries along three disjoint sets of half-edges: *active, neutral, dead.*

(S0) At stage $i = 0$, the half-edges incident to $j$ are active and all the other half-edges are neutral. Order the initially active half-edges arbitrarily.

(S1) At each stage, take the *largest* half-edge $e$ and pair it with another half-edge $f$, chosen uniformly at random from the set of half-edges that are either active or neutral. If $f$ is neutral, then the vertex $v$ to which $f$ is incident, is not discovered yet. Declare the half-edges incident to $v$ to be active and *larger* than all other active vertices (choose any order between the half-edges of $v$). Declare $e, f$ to be dead.

(S2) Repeat from (S1) until the set of active half-edges is empty.

Define the process $\mathbf{S}_n^j$ by $S_n^j(l) = S_n^j(l-1) + d_{(l)}' J_l - 2$, and $S_n^j(0) = d_j'$, where $J_l$ is the indicator that a new vertex is discovered at time $l$ and $d_{(l)}'$ is the degree of the discovered vertex, if any. Let $L := \inf\{l \geqslant 1 : S_n^j(l) = 0\}$. By convention, we assume that $S_n^j(l) = 0$ for $l > L$. Let $\mathscr{V}_l$ denote the vertex set

discovered upto time $l$ excluding $j$ and $\mathcal{I}_i^n(l) := \mathbf{1}\{i \in \mathcal{V}_l\}$. Define $\mathcal{I}_j^n(l) \equiv 0$. Also, let $\mathcal{F}_l$ denote the sigma-field containing all the information upto time $l$ in Algorithm 4.3. Note that

$$S_n^j(l) = d_j' + \sum_{i \in [n]} d_i' \mathcal{I}_i^n(l) - 2l = d_j' + \sum_{i \in [n]} d_i' \left( \mathcal{I}_i^n(l) - \frac{d_i'}{\ell_n'} l \right) + (\nu_n' - 1) l.$$

Consider the re-scaled process $\bar{S}_n^j$ defined as $\bar{S}_n^j(t) = n^{-\alpha} S_n^j(\lfloor tn^{\alpha+\delta} \rfloor)$. Then, using Assumption 4.2,

$$\bar{S}_n^j(t) = c_j + n^{-\alpha} \sum_{i \in [n]} d_i' \left( \mathcal{I}_i^n(tn^{\alpha+\delta}) - \frac{d_i'}{\ell_n'} tn^{\alpha+\delta} \right) - \lambda_0 t + o(1). \quad (4.47)$$

The following three lemmas determine the asymptotics of $\mathscr{W}_{(j)}$ and $s_3^\star$:

**Lemma 4.4.9.** *Let $L_j$ be the function with $L_j(t) = c_j - \lambda_0 t$ for $t \in [0, c_j \lambda_0^{-1}]$ and $L_j(t) = 0$ for $t > c_j \lambda_0^{-1}$. Then, under Assumption 4.2, as $n \to \infty$, $\bar{S}_n^j \xrightarrow{P} L_j$ with respect to the Skorohod $J_1$ topology.*

**Lemma 4.4.10.** *For any $T > 0$, $\sup_{l \leqslant Tn^{\alpha+\delta}} \left| \sum_{i \in [n]} w_i \mathcal{I}_i^n(l) - \frac{\sum_{i \in [n]} d_i' w_i}{\sum_{i \in [n]} d_i'} l \right| = o_P(n^{\alpha+\delta})$.*

**Lemma 4.4.11.** *Fix any $j \geqslant 1$. Then with high probability $\mathscr{W}(j) = \mathscr{W}_{(j)}$.*

*Asymptotics of $\mathscr{W}_{(j)}$.* Note that, since the exploration process explores one edge at each time, Lemma 4.4.9 implies that (see e.g. [150, Theorem 13.6.4])

$$\frac{1}{2n^{\alpha+\delta}} \sum_{k \in \mathscr{C}'(j)} d_k' \xrightarrow{\mathbb{P}} \frac{c_j}{\lambda_0}. \quad (4.48)$$

Moreover, Lemma 4.4.10 yields that

$$\begin{aligned}
\frac{1}{n^{\alpha+\delta}} \mathscr{W}(j) &= \frac{1}{n^{\alpha+\delta}} \sum_{k \in \mathscr{C}'(j)} w_k \\
&= \frac{\sum_{i \in [n]} d_i' w_i}{\ell_n' n^{\alpha+\delta}} \frac{1}{2} \sum_{k \in \mathscr{C}'(j)} d_k' + o_P(1) \xrightarrow{\mathbb{P}} \frac{\mu_{d,w}}{\mu_d \lambda_0} c_j.
\end{aligned} \quad (4.49)$$

Now the asymptotics of $\mathscr{W}_{(j)}$ in Theorem 4.1.3 follows by an application of Lemma 4.4.11 under Assumption 4.2.  □

Next we provide a proof for Lemma 4.4.11 subject to Lemmas 4.4.9, 4.4.10. The proofs of Lemmas 4.4.9 and 4.4.10 are similar to [69, Section 4] and are provided in Appendix 4.C.

*Proof of Lemma 4.4.11.* To simplify the writing, we only give a proof for $j = 1$, the general case follows similarly. Define the event $A_i := \{\mathscr{W}(i) > \mathscr{W}(1), i \notin \cup_{k \leqslant i-1}\mathscr{C}'(k)\}$ and let $A_{>K} = \cup_{i>K} A_i$. Fix $r$ such that $c_2 < r\lambda_0\mu_d/\mu_{d,w} < c_1$ and define the event $B = \{\mathscr{W}(1) > rn^{\alpha+\delta}\}$. Then, for any $K \geqslant 2$,

$$\mathbb{P}\left(\mathscr{W}(1) \neq \mathscr{W}_{(1)}\right) \leqslant \sum_{i=2}^{K} \mathbb{P}\left(A_i \cap B\right) + \mathbb{P}\left(A_{>K} \cap B\right) + \mathbb{P}\left(B^c\right). \tag{4.50}$$

Firstly, notice that due to the choice of $r$, (4.48) and (4.49) implies that $\mathbb{P}(B^c) \to 0$. Moreover, for each fixed $i \geqslant 2$, $\mathbb{P}\left(A_i \cap B\right) \leqslant \mathbb{P}(\mathscr{W}(i) > rn^{\alpha+\delta}) \to 0$. Further, recall (4.46) and the relevant notation. Note that

$$\mathbb{P}\left(A_{>K} \cap B\right) \leqslant \mathbb{P}\left(\sum_{i \geqslant 1}\left(\mathscr{W}_{(i)}^K\right)^3 > r^3 n^{3(\alpha+\delta)}\right).$$

Thus, the proof follows from (4.50) by taking first the limit as $n \to \infty$, and then as $K \to \infty$ and using (4.46). $\qquad\square$

### 4.4.4    Mesoscopic typical distances

Recall the definition of $\mathcal{D}_n^\star$ from (4.6). In this section, we obtain the asymptotics of $\mathcal{D}_n^\star$ in Theorem 4.1.3 using a similar analysis as in Section 4.4.2. Again the proof involves the Chebyshev inequality where the moments are estimated using path counting. We sketch the computation of $\mathbb{E}[\mathcal{D}_n^\star]$. Recall the notations $U_n^*, V_n^*, \mathcal{A}(k,l)$ and $\mathcal{A}'(k,l)$ from Section 4.4.2. Note that

$$\mathbb{E}\left[\mathcal{D}_n^\star\right] = \frac{1}{n}\mathbb{E}\left[\sum_{i,k\in[n]} w_i w_k d(i,k)\mathbf{1}\{k \in \mathscr{C}'(i)\}\right]$$

$$\leqslant \frac{\ell_n^w}{n}\sum_{k\in[n]} w_k \sum_{l\geqslant 1} l\,\mathbb{P}\left(\mathcal{A}(k,l)\right) = \frac{\ell_n^w}{n}\sum_{l\geqslant 1} l \sum_{k\in[n]} w_k\mathbb{P}\left(\mathcal{A}(k,l)\right),$$

and

$$\mathbb{E}\left[\mathcal{D}_n^\star\right] \geqslant \frac{\ell_n^w}{n}\sum_{l\geqslant 1} l\left(\sum_{k\in[n]} w_k\left(\mathbb{P}\left(\mathcal{A}(k,l)\right) - \mathbb{P}\left(\mathcal{A}'(k,l)\right)\right)\right).$$

Now compare the terms above to (4.40a), (4.40b). The only difference is that there is an extra multiplicative $l$ here which amounts to differentiating with respect to $v_n'$ in the obtained bounds. Thus, we can repeat an argument identical to (4.41), (4.43) to obtain that

$$\mathbb{E}\left[\mathcal{D}_n^\star\right] = \frac{\mathbb{E}\left[W_n\right]\mathbb{E}\left[D_n^*\right]\mathbb{E}\left[D_n' W_n\right]}{\mathbb{E}\left[D_n'\right](1-v_n')^2}(1+o(1)) = \frac{(\mathbb{E}\left[D_n' W_n\right])^2}{\mathbb{E}\left[D_n'\right](1-v_n')^2}(1+o(1)).$$

The variance terms can also be computed similarly. Due to the presence of a factor $l^2$ in the second moment, we have to differentiate the upper-bounds twice with respect to $\nu'_n$. Again, the identical arguments as (4.38) can be applied to show that $\text{Var}\left(\mathcal{D}^\star_n\right) = o(n^{4\delta})$. This completes the proof of the asymptotics of $\mathcal{D}^\star_n$.

### 4.4.5   Maximum diameter: Proof of Theorem 4.1.4

Firstly, let us investigate the diameter of $\mathscr{C}'(i)$. Notice that, if $\Delta(\mathscr{C}'(i)) > 6n^\delta \log(n)$, then there exists at least one path of length at least $3n^\delta \log(n)$ starting from $i$. Now, the expected number of such paths is at most $\sum_{l=3n^\delta \log(n)}^n \mathbb{E}\left[P_l\right]$, where $P_l$ denotes the number of paths of length $l$, starting from vertex $i$ and we have used the fact that a vertex disjoint path can be of size at most $n$. Again, the path-counting technique yields $\mathbb{E}\left[P_l\right] \leqslant d'_i \nu'^{l-1}_n$. Thus, for some constant $C > 0$,

$$\mathbb{P}\left(\Delta(\mathscr{C}'(i)) > 6n^\delta \log(n)\right) \leqslant \sum_{l=3n^\delta \log(n)}^n \mathbb{E}\left[P_l\right]$$

$$\leqslant C d'_i n (\nu'_n)^{3n^\delta \log(n)} \leqslant C \frac{d'_i}{n^2},$$

where in the last step we have used (4.4). Thus, the proof of Theorem 4.1.4 follows using the union bound.

## 4.5    Metric space limit for percolation clusters

Finally, the aim of this section is to complete the proof of Theorem 4.1.1. We start by defining the multiplicative coalescent process [8, 11] that will play a pivotal role in this section:

**Definition 5** (Multiplicative coalescent). Consider a (possibly infinite) collection of particles and let $\mathbf{X}(s) = (X_i(s))_{i \geqslant 1}$ denote the collection of masses of those particles at time $s$. Thus the $i$-th particle has mass $X_i(s)$ at time $s$. The evolution of the system takes place according to the following rule at time $s$: At rate $X_i(s)X_j(s)$, particles $i$ and $j$ merge into a new particle of mass $X_i(s) + X_j(s)$.

Before going into the details, let us describe the general idea and the organization of this section. Extending the approach of [29], we consider a dynamically growing process of graphs that approximates the percolation clusters in the critical window (see Chapter 2). Now, the graphs generated

by this dynamic evolution satisfy: (i) In the critical window, the components merge *approximately* as the multiplicative coalescent where the *mass* of each component is approximately proportional to the component size; (ii) the masses of the barely sub-critical clusters satisfy *nice* properties due to Theorem 4.1.3. In Section 4.5.1, we derive the required properties in the barely subcritical regime for the dynamically growing graph process using Theorems 4.1.3 and 4.1.4. In Section 4.5.2, we modify the dynamic process such that the components merge exactly as multiplicative coalescent. Since the exact multiplicative coalescent corresponds to the rank-one inhomogeneous case, thinking of these barely subcritical clusters as blobs, we use the universality theorem (Theorem 4.4.2) in Section 4.5.3 to determine the metric space limits of the largest components of the modified graph (Theorem 4.5.8). Section 4.5.4 is devoted to the structural comparison of the modified graph and the original graph, and we finally complete the proof of Theorems 4.1.1. in Section 4.5.5. The proof of Theorem 4.1.2 is given in Section 4.5.6.

**Algorithm 4.4** (The dynamic construction). Let $\mathcal{G}_n(t)$ be the graph obtained up to time $t$ by the following dynamic construction:

(S0) A half-edge can either be alive or dead. Initially, all the half-edges are alive. All the half-edges have an independent unit rate exponential clock attached to them.

(S1) Whenever a clock rings, we take the corresponding half-edge, kill it and pair it with a half-edge chosen uniformly at random among the alive half-edges. The paired half-edge is also killed and the exponential clocks associated with killed half-edges are discarded.

Since a half-edge is paired with another unpaired half-edge, chosen uniformly at random from the set of all unpaired half-edges, the final graph $\mathcal{G}_n(\infty)$ is distributed as $\mathrm{CM}_n(\mathbf{d})$. Define

$$t_c(\lambda) = \frac{1}{2} \log \left( \frac{\nu_n}{\nu_n - 1} \right) + \frac{\nu_n}{2(\nu_n - 1)} \frac{\lambda}{n^{\eta}}. \tag{4.51}$$

We denote the $i$-th largest component of $\mathcal{G}_n(t)$ by $\mathscr{C}_{(i)}(t)$. In the subsequent part of this chapter, we will derive the metric space limit of $(\mathscr{C}_{(i)}(t_c(\lambda)))_{i \geqslant 1}$. The following lemma (see Proposition 3.10.1) enables us to switch to the conclusions for the largest clusters of $\mathrm{CM}_n(\mathbf{d}, p_n(\lambda))$:

**Lemma 4.5.1** ([69, Proposition 24]). *There exists $\varepsilon_n = o(n^{-\eta})$ and a coupling such that, with high probability,*

$$\mathcal{G}_n(t_c(\lambda) - \varepsilon_n) \subset CM_n(\mathbf{d}, p_n(\lambda)) \subset \mathcal{G}_n(t_c(\lambda) + \varepsilon_n),$$
$$CM_n(\mathbf{d}, p_n(\lambda) - \varepsilon_n) \subset \mathcal{G}_n(t_c(\lambda) \subset CM_n(\mathbf{d}, p_n(\lambda) + \varepsilon_n).$$

Let $\omega_i(t)$ denote the number of unpaired/open half-edges incident to vertex $i$ at time $t$ in Algorithm 4.4. We end this section by understanding the evolution of some functionals of the degrees and the open half-edges in the graph $\mathcal{G}_n(t)$. Let $s_1(t)$ denote the total number of unpaired half-edges at time $t$. Denote also $s_2(t) = \sum_{i \in [n]} \omega_i(t)^2$, $s_{d,\omega}(t) = \sum_{i \in [n]} d_i \omega_i(t)$. Further, we write $\mu_n = \ell_n/n$.

**Lemma 4.5.2.** *Under Assumption 4.1, the quantities* $\sup_{t \leqslant T} |\frac{1}{n} s_1(t) - \mu_n e^{-2t}|$, $\sup_{t \leqslant T} |\frac{1}{n} s_2(t) - \mu_n e^{-4t}(\nu_n + e^{2t})|$, $\sup_{t \leqslant T} |\frac{1}{n} s_{d,\omega}(t) - \mu_n(1 + \nu_n)e^{-2t}|\}$ *are* $O_{\mathbb{P}}(n^{-1/2})$, *for any* $T > 0$.

*Proof.* The proof uses the differential equation method [152]. Notice that, after each exponential clock rings in Algorithm 4.4, $s_1(t)$ decreases by two. Let $Y$ denote a unit rate Poisson process. Using the random time change representation [80],

$$s_1(t) = \ell_n - 2Y\left(\int_0^t s_1(u)du\right) = \ell_n + M_n(t) - 2\int_0^t s_1(u)du,$$

where $M_n$ is a martingale. Now, the quadratic variation of $M_n$ satisfies $\langle M_n \rangle(t) \leqslant 4t\ell_n = O(n)$, which implies that $\sup_{t \leqslant T} |M_n(t)| = O_{\mathbb{P}}(\sqrt{n})$. Moreover, notice that the function $f(t) = \mu_n e^{-2t}$ satisfies $f(t) = \mu_n - 2\int_0^t f(u)du$. Therefore,

$$\sup_{t \leqslant T}\left|\frac{1}{n} s_1(t) - \mu_n e^{-2t}\right| \leqslant \sup_{t \leqslant T} \frac{|M_n(t)|}{n} + 2\int_0^T \sup_{t \leqslant u}\left|\frac{1}{n} s_1(t) - \mu_n e^{-2t}\right|du.$$

Using Grőnwall's inequality [131, Proposition 1.4], it follows that

$$\sup_{t \leqslant T}\left|\frac{1}{n} s_1(t) - \mu_n e^{-2t}\right| \leqslant e^{2T} \sup_{t \leqslant T} \frac{|M_n(t)|}{n} = O_{\mathbb{P}}(n^{-1/2}), \tag{4.52}$$

as required. For $s_2(t)$, note that if half-edges corresponding to vertices $i$ and $j$ are paired, then $s_2$ changes by $-2\omega_i - 2\omega_j + 2$ and if two half-edges

corresponding to $i$ are paired, $s_2$ then changes by $-4\omega_i + 4$. Thus,

$$\sum_{i \in [n]} \omega_i(t)^2$$

$$= \sum_{i \in [n]} d_i^2 + M_n'(t) + \int_0^t \sum_{i \neq j} \frac{\omega_i(u)\omega_j(u)(-2\omega_i(u) - 2\omega_j(u) + 2)}{s_1(u) - 1}$$

$$+ \int_0^t \sum_{i \in [n]} \frac{\omega_i(u)(\omega_i(u) - 1)(-4\omega_i(u) + 4)}{s_1(u) - 1}$$

$$= n\mu_n(1 + \nu_n) + M_n'(t) + \int_0^t (-4s_2(u) + 2s_1(u))du + O_{\mathbb{P}}(1),$$

where $M_n'$ is a martingale with quadratic variation given by $\langle M_n' \rangle(t) = O(n)$. Again, an estimate equivalent to (4.52) follows using Grönwall's inequality. Notice also that when a clock corresponding to vertex $i$ rings and it is paired to vertex $j$, then $s_{d,\omega}$ decreases by $d_i + d_j$. Thus,

$$s_{d,\omega}(t) = \sum_{i \in [n]} d_i^2 + M_n''(t) - \int_0^t \sum_{i \neq j} \frac{\omega_i(u)\omega_j(u)(d_i + d_j)}{s_1(u) - 1}du$$

$$- \int_0^t \sum_{i \in [n]} \frac{\omega_i(u)(\omega_i(u) - 1)2d_i}{s_1(u) - 1}du$$

$$= n\mu_n(1 + \nu_n) + M_n''(t) - 2\int_0^t s_{d,\omega}(u)du,$$

where $M_n''$ is a martingale with quadratic variation given by $\langle M_n'' \rangle(t) \leqslant 2t \sum_{i \in [n]} d_i^2 = O(n)$. The proof of Lemma 4.5.2 is now complete. $\square$

### 4.5.1   Entrance boundary for open half-edges

Define

$$t_n = \frac{1}{2} \log\left(\frac{\nu_n}{\nu_n - 1}\right) - \frac{\nu_n}{2(\nu_n - 1)} \frac{1}{n^\delta}, \quad 0 < \delta < \eta. \tag{4.53}$$

The goal is to show that the open half-edges satisfy the entrance boundary conditions. Let $\mathbf{d}(t) = (d_i(t))_{i \in [n]}$ denote the degree sequence of $\mathcal{G}_n(t)$ constructed by Algorithm 4.4. Recall that $\mathcal{G}_n(t)$ is a configuration model conditionally on $\mathbf{d}(t)$. Let us first derive the asymptotics of $\nu_n(t_n)$. Recall that $\omega_i(t)$ denotes the number of open half-edges adjacent to vertex $i$ in $\mathcal{G}_n(t)$. Notice that

$$\nu_n(t_n) = \frac{\sum_{i \in [n]} (d_i - \omega_i(t_n))^2}{\ell_n - s_1(t_n)} - 1 = \frac{\sum_{i \in [n]} d_i^2 - 2s_{d,\omega}(t_n) + s_2(t_n)}{\ell_n - s_1(t_n)} - 1.$$

Using Lemma 4.5.2 and Assumption 4.1,

$$\frac{1}{n}(\ell_n - s_1(t_n)) = \mu_n(1 - e^{-2t_n}) + o_{\mathbb{P}}(n^{-\delta}) = \frac{\mu_n}{\nu_n}\left(1 - \frac{\nu_n}{n^\delta}\right) + o_{\mathbb{P}}(n^{-\delta}),$$

$$(4.54)$$

$$\frac{1}{n}\left(\sum_{i\in[n]} d_i^2 - 2s_{d,\omega}(t_n) + s_2(t_n)\right) = \frac{\mu_n}{\nu_n}\left(2 - \frac{3\nu_n}{n^\delta}\right) + o_{\mathbb{P}}(n^{-\delta}). \quad (4.55)$$

Thus, (4.54) and (4.55) yield that $\nu_n(t_n) = 1 - \nu_n n^{-\delta} + o_{\mathbb{P}}(n^{-\delta})$. We aim to apply the results for the barely sub-critical regime in Theorem 4.1.3 to the number of open half-edges $\omega(t_n) = (\omega_i(t_n))_{i\in[n]}$. Notice that, by Lemma 4.5.2 and Assumption 4.1, $\omega(t_n)$ and $d(t_n)$ satisfy Assumption 4.2 with

$$\mu_\omega = \frac{\mu(\nu-1)}{\nu}, \quad \mu_d = \frac{\mu}{\nu}, \quad \mu_{d,\omega} = \frac{\mu(\nu-1)}{\nu}, \quad c_i = \frac{\theta_i}{\nu}.$$

Consider the quantities $s_2^\star$, $s_3^\star$, $\mathcal{D}_n^\star$ with the weights being the number of open half-edges and denote them by $s_2^\omega$, $s_3^\omega$, $\mathcal{D}_n^\omega$ respectively. Denote $f_i(t) = \sum_{k\in\mathscr{C}_{(i)}(t)} \omega_k(t)$ and $f(t) = (f_i(t))_{i\geqslant 1}$. The following theorem summarizes the entrance boundary conditions for $f(t)$:

**Theorem 4.5.3.** *Under Assumption 4.1, as $n \to \infty$,*

$$n^{-\delta}s_2^\omega \xrightarrow{\mathbb{P}} \frac{\mu(\nu-1)^2}{\nu^2}, \quad n^{-\delta}s_{pr}^\omega \xrightarrow{\mathbb{P}} \frac{\mu(\nu-1)}{\nu^2}, \quad n^{-(\alpha+\delta)}f_i(t_n) \xrightarrow{\mathbb{P}} \left(\frac{\nu-1}{\nu^2}\right)\theta_i,$$

$$n^{-3\alpha-3\delta+1}s_3^\omega \xrightarrow{\mathbb{P}} \left(\frac{\nu-1}{\nu^2}\right)^3 \sum_{i=1}^{\infty} \theta_i^3, \quad n^{-2\delta}\mathcal{D}_n^\omega \xrightarrow{\mathbb{P}} \frac{\mu(\nu-1)^2}{\nu^3}.$$

**Remark 36.** Setting $w_i = 1$ for all $i$, we get the entrance boundary conditions for the component sizes also. In this case $\mu_d = \mu_{d,w} = \mu/\nu$. Augmenting a predefined notation with $c$ in the superscript to denote the component susceptibilities, it follows that

$$n^{-\delta}s_2^c \xrightarrow{\mathbb{P}} \frac{\mu}{\nu^2}, \quad n^{-(\alpha+\delta)}|\mathscr{C}_{(i)}(t_n)| \xrightarrow{\mathbb{P}} \frac{\theta_i}{\nu^2}, \quad n^{-3\alpha-3\delta+1}s_3^c \xrightarrow{\mathbb{P}} \frac{1}{\nu^6}\sum_{i=1}^{\infty} \theta_i^3.$$

### 4.5.2   Coupling with the multiplicative coalescent

Recall the definitions of $t_c(\lambda)$ and $t_n$ from (4.51) and (4.53). Now, let us investigate the dynamics of $f(t)$ starting from time $t_n$. Notice that, in the time interval $[t_n, t_c(\lambda)]$, components with masses $f_i(t)$ and $f_j(t)$ merge at rate

$$f_i(t)\frac{f_j(t)}{s_1(t)-1} + f_j(t)\frac{f_i(t)}{s_1(t)-1} = \frac{2f_i(t)f_j(t)}{s_1(t)-1} \approx \frac{2\nu f_i(t)f_j(t)}{\mu(\nu-1)n},$$

and create a component with $f_i(t) + f_j(t) - 2$ open half-edges. Thus $f(t)$ does not *exactly* evolve as a multiplicative coalescent, but it is close. Now, we define an exact multiplicative coalescent that approximates the above process:

**Algorithm 4.5** (Modified process). Conditionally on $\mathcal{G}_n(t_n)$, associate a rate $2/(s_1(t_n) - 1)$ Poisson process $\mathcal{P}(e, f)$ to each of pair of unpaired-half-edges $(e, f)$. An edge $(e, f)$ is created between the vertices incident to $e$ and $f$ at the instance when $\mathcal{P}(e, f)$ rings. We denote the graph obtained at time $t$ by $\bar{\mathcal{G}}_n(t)$.

**Proposition 4.5.4.** *There exists a coupling such that $\mathcal{G}_n(t) \subset \bar{\mathcal{G}}_n(t)$ for all $t > t_n$ with probability one.*

*Proof.* Recall the construction of $\mathcal{G}_n(t)$ from Algorithm 4.4. We modify (S1) as follows: whenever two half-edges are paired, we do not kill the corresponding half-edges and do not discard the associated exponential clocks. Instead we reset the corresponding exponential clocks. The graphs generated by this modification of Algorithm 4.4 has the same distribution as $\bar{\mathcal{G}}_n(t)$, conditionally on $\mathcal{G}_n(t_n)$. Moreover, the above also gives a natural coupling such that $\mathcal{G}_n(t) \subset \bar{\mathcal{G}}_n(t)$, by viewing the event times of Algorithm 4.4 as a thinning of the event times of the modified process. $\square$

Henceforth, we will always assume that we are working on a probability space such that Proposition 4.5.4 holds. The connected components at time $t_n$, $(\mathscr{C}_{(i)}(t_n))_{i \geqslant 1}$ are regarded as blobs. Thus, for $t \geqslant t_n$, the graph $\bar{\mathcal{G}}_n(t)$ should be viewed as a super-graph with the superstructure being determined by the edges appearing after time $t_n$ in Algorithm 4.5. Let us denote the ordered connected components of $\bar{\mathcal{G}}_n(t)$ by $(\bar{\mathscr{C}}_{(i)}(t))_{i \geqslant 1}$. The components of $\bar{\mathcal{G}}_n(t)$ can be regarded as a union of the blobs. For a component $\mathscr{C}$, we use the notation $\mathfrak{B}(\mathscr{C})$ to denote the collection of indices corresponding to the blobs within $\mathscr{C}$ given by $\{b : \mathscr{C}_{(b)}(t_n) \subset \mathscr{C}\}$. Denote

$$\bar{\mathcal{F}}_i(t) = \sum_{b \in \mathfrak{B}(\bar{\mathscr{C}}_{(i)}(t))} f_b(t_n).$$

The $\bar{\mathcal{F}}$-value is regarded as the mass of component $\bar{\mathscr{C}}_{(i)}(t)$ at time $t$. Note that for the modified process in Algorithm 4.5, conditionally on $\mathcal{G}_n(t_n)$, at time $t \in [t_n, t_c(\lambda)]$, $\bar{\mathscr{C}}_{(i)}(t)$ and $\bar{\mathscr{C}}_{(j)}(t)$ merge at exact rate $2\bar{\mathcal{F}}_i(t)\bar{\mathcal{F}}_j(t)/(s_1(t_n) - 1)$ and the new component has mass $\bar{\mathcal{F}}_i(t) + \bar{\mathcal{F}}_j(t)$. Thus, the vector of masses $(\bar{\mathcal{F}}_i(t))_{i \geqslant 1}$ merge as an exact multiplicative coalescent.

### 4.5.3    Properties of the modified process

Notice that, conditionally on $\mathcal{G}_n(t_n)$, blobs $b_i$ and $b_j$ are connected in $\bar{\mathcal{G}}_n(t_c(\lambda))$ with probability $p_{ij}$ equal to

$$1 - \exp\left(-f_{b_i}(t_n)f_{b_j}(t_n)\left[\frac{1}{n^{1+\delta}}\frac{\nu^2}{\mu(\nu-1)^2} + \frac{1}{n^{1+\eta}}\frac{\nu^2}{\mu(\nu-1)^2}\lambda\right](1+o_{\mathbb{P}}(1))\right),$$
$$(4.56)$$

where the $o_{\mathbb{P}}(\cdot)$ term appearing above is uniform in $i, j$. Thus, using Theorem 4.5.3, (4.56) is of the form $1 - e^{-q x_i^n x_j^n (1+o_{\mathbb{P}}(1))}$ with

$$x_i^n = n^{-\rho}f_{b_i}(t_n), \quad q = \frac{1}{\sigma_2(x^n)} + \frac{\nu^2}{\mu(\nu-1)^2}\lambda, \qquad (4.57)$$

where $\sigma_r(x^n) = \sum_i (x_i^n)^r$. By Theorem 4.1.3, the sequence $x^n$ satisfies the entrance boundary conditions of [11], i.e.,

$$\frac{\sigma_3(x^n)}{(\sigma_2(x^n))^3} \xrightarrow{\mathbb{P}} \frac{1}{\mu^3(\nu-1)^3}\sum_{i=1}^{\infty}\theta_i^3, \quad \frac{x_i^n}{\sigma_2(x^n)} \xrightarrow{\mathbb{P}} \frac{1}{\mu(\nu-1)}\theta_i, \quad \sigma_2(x^n) \xrightarrow{\mathbb{P}} 0.$$

To simplify the notation, we write $\bar{\mathcal{F}}_i(\lambda)$ for $\bar{\mathcal{F}}_i(t_c(\lambda))$ and $\mathscr{C}_{(i)}(\lambda)$ for $\mathscr{C}_{(i)}(t_c(\lambda))$. The following result is a consequence of [11, Proposition 7] and Lemma 4.3.5:

**Proposition 4.5.5.** *As $n \to \infty$, $\left(n^{-\rho}\bar{\mathcal{F}}_i(\lambda)\right)_{i\geqslant 1} \xrightarrow{d} \frac{\nu-1}{\nu}\xi$ with respect to the $\ell_{\downarrow}^2$ topology, where $\xi$ is defined in Proposition 4.3.4.*

We next relate $(\bar{\mathcal{F}}_i(\lambda))_{i\geqslant 1}$ to $(\mathscr{C}_{(i)}(\lambda))_{i\geqslant 1}$, for each fixed $i$:

**Proposition 4.5.6.** *As $n \to \infty$, $\bar{\mathcal{F}}_i(\lambda) = (\nu-1)|\mathscr{C}_{(i)}(\lambda)| + o_{\mathbb{P}}(n^\rho)$. Consequently, $\left(n^{-\rho}|\mathscr{C}_{(i)}(\lambda)|\right)_{i\geqslant 1} \xrightarrow{d} \frac{1}{\nu}\xi$ with respect to the product topology.*

We will need the following lemma, the proof of which is same as [35, Lemma 8.2].

**Lemma 4.5.7** ([35, Lemma 8.2]). *Consider two ordered weight sequences $x = (x_i)_{i\in[m]}$ and $y = (y_i)_{i\in[m]}$. Consider the size-biased reordering $(v(1), v(2), \dots)$ of $[m]$ with respect to the weights $x$ and let $V(i) := \{v(1), \dots, v(i)\}$. Denote $m_{rs} = \sum_i x_i^r y_i^s$, define $c_n = m_{11}/m_{10}$ and assume that $c_n > 0$ for each $n$. Suppose that the following conditions hold:*

$$\frac{lm_{21}}{m_{10}m_{11}} \to 0, \quad \frac{m_{12}m_{10}}{lm_{11}^2} \to 0, \quad \frac{lm_{20}}{m_{10}^2} \to 0, \quad \text{as } n \to \infty. \qquad (4.58)$$

*Then, as $n \to \infty$, $\sup_{k\leqslant l}\left|\frac{1}{lc_n}\sum_i y_i\mathbf{1}\{i \in V(k)\} - \frac{k}{l}\right| \xrightarrow{\mathbb{P}} 0$.*

*Proof of Proposition 4.5.6.* We only prove the asymptotic relation of $\bar{\mathcal{F}}_1(\lambda)$ and $|\mathscr{C}_{(1)}(\lambda)|$. Consider the breadth-first exploration of the *supestructure* of graph $\bar{\mathcal{G}}_n(t_c(\lambda))$ (which is also a rank-one inhomogeneous random graph) using the Aldous-Limic construction from [11, Section 2.3]. Notice that the vertices are explored in a size-biased manner with the sizes being $x = (x_i)_{i \geqslant 1}$, where $x_i = n^{-\rho} f_i(t_n)$. Let $\nu(i)$ be the i-th vertex explored. Further, let $\bar{\mathscr{C}}_{(i)}^{st}(\lambda)$ denote the component $\bar{\mathscr{C}}_{(i)}(\lambda)$, where the blobs have been shrunk to single vertices. Then, from [11], one has the following:

(i) there exists random variables $m_L, m_R$ such that $\bar{\mathscr{C}}_{(i)}^{st}(\lambda)$ is explored between $m_L + 1$ and $m_R$;

(ii) $\sum_{i \leqslant m_R} x_{\nu(i)}$ is tight;

(iii) $\sum_{i=m_L+1}^{m_R} x_{\nu(i)} \xrightarrow{d} \gamma$, where $\gamma$ is some non-degenerate, positive random variable.

Let $y_i = n^{-\rho} |\mathscr{C}_{(i)}(t_n)|$. Using Theorem 4.5.3, Remark 35 and Remark 36, it follows that $\sum_i x_i^r y_i^s = O_{\mathbb{P}}(n^{3\delta - 3\eta})$; for $r + s = 3$, $\sum_i x_i = O_{\mathbb{P}}(n^{1-\rho})$, and $\sum_i x_i^r y_i^s = O_{\mathbb{P}}(n^{-2\rho+1+\delta})$; for $r + s = 2$. Below, we show that

$$\frac{\sum_{i=m_L+1}^{m_R} y_{\nu(i)}}{\sum_{i=m_L+1}^{m_R} x_{\nu(i)}} \times \frac{\sum_i x_i^2}{\sum_i x_i y_i} \xrightarrow{\mathbb{P}} 1. \tag{4.59}$$

The proof of Proposition 4.5.6 follows from (4.59) by observing that $\frac{\sum_i x_i^2}{\sum_i x_i y_i} = \frac{s_2^\omega(t_n)}{s_{pr}^\omega(t_n)} \xrightarrow{\mathbb{P}} \nu - 1$, and using Theorem 4.5.3. To prove (4.59), we will now apply Lemma 4.5.7. Denote $m_0 = \sum_i x_i / \sum_i x_i^2$ and consider $l = 2Tm_0$ for some fixed $T > 0$. Using Theorem 4.5.3, an application of Lemma 4.5.7 yields

$$\sup_{k \leqslant 2Tm_0} \left| \sum_{i=1}^k x_{\nu(i)} - \frac{k}{m_0} \right| \xrightarrow{\mathbb{P}} 0.$$

Now, for any $\varepsilon > 0$, $T > 0$ can be chosen large enough such that $\sum_{i=1}^{m_R} x_{\nu(i)} > T$ has probability at most $\varepsilon$ and on the event $\left\{ \sup_{k \leqslant 2Tm_0} \left| \sum_{i=1}^k x_{\nu(i)} - \frac{k}{m_0} \right| \leqslant \varepsilon \right\} \cap \left\{ \sum_{i=1}^{m_R} x_{\nu(i)} \leqslant T \right\}$, one has $m_L < m_R < 2Tm_0$. Thus, it follows that

$$\left| \sum_{i=m_L+1}^{m_R} x_{\nu(i)} - \frac{m_R - m_L}{m_0} \right| \xrightarrow{\mathbb{P}} 0. \tag{4.60}$$

An identical argument as above shows that

$$\left| \sum_{i=m_L+1}^{m_R} y_{\nu(i)} - \frac{m_R - m_L}{m_0'} \right| \xrightarrow{\mathbb{P}} 0, \tag{4.61}$$

where $m_0' = \sum_i x_i / \sum_i x_i y_i$. The proof of (4.59) now follows from (4.60) and (4.61). The asymptotic distribution for $(n^{-\rho} |\bar{\mathscr{C}}_{(i)}|)$ can be obtained using Proposition 4.5.5 and Lemma 4.3.5. $\hfill\square$

Recall that $\omega_i(t_n)$ denotes the number of open-half edges attached to vertex $i$ in the graph $\mathcal{G}_n(t_n)$. We now equip $\bar{\mathscr{C}}_{(i)}(\lambda)$ with the probability measure $\mu_{fr}^i$ given by $\mu_{fr}^i(A) = \sum_{k \in A} \omega_k(t_n) / \mathcal{F}_i(\lambda)$ for $A \subset \bar{\mathscr{C}}_{(i)}(\lambda)$, and denote the corresponding measured metric space by $\bar{\mathscr{C}}_{(i)}^{fr}(\lambda)$.

**Theorem 4.5.8.** *Under Assumption 4.1, as $n \to \infty$,*

$$\left(n^{-\eta} \bar{\mathscr{C}}_{(i)}^{fr}(\lambda)\right)_{i \geqslant 1} \xrightarrow{d} (M_i)_{i \geqslant 1}, \tag{4.62}$$

*with respect to the $\mathscr{S}_*^{\mathbb{N}}$ topology, where $M_i$ is defined in Section 4.3.5.*

*Proof.* We just consider the metric space limit of $\bar{\mathscr{C}}_{(i)}^{fr}(\lambda)$ for each fixed $i \geqslant 1$ and the joint convergence in (4.62) follows using the joint convergence of different functionals used throughout the proof. Recall the notation $\mathfrak{B}(\mathscr{C}) := \{b : \mathscr{C}_{(b)}(t_n) \subset \mathscr{C}\}$ for a component $\mathscr{C}$. Now, $\bar{\mathscr{C}}_{(i)}^{fr}(\lambda)$ can be seen as a super-graph as defined in Section 4.3.2 with

1. the collection of blobs $\{\mathscr{C}_{(b)}(t_n) : b \in \mathfrak{B}(\bar{\mathscr{C}}_{(i)}(\lambda))\}$ and within-blob measure $\mu_b$ given by $\mu_b(A) = \sum_{k \in A} \omega_k(t_n) / f_b(t_n)$, $A \subset \mathscr{C}_{(b)}(t_n)$, $b \in \mathfrak{B}(\bar{\mathscr{C}}_{(i)}(\lambda))$;

2. the superstructure consisting of the edges appearing during $[t_n, t_c(\lambda)]$ in Algorithm 4.5 and weight sequence $(f_b(t_n) / \mathcal{F}_i(\lambda) : b \in \mathfrak{B}(\bar{\mathscr{C}}_{(i)}(\lambda)))$.

Let $d(\cdot, \cdot)$ denote the graph distance on $\bar{\mathscr{C}}_{(i)}(\lambda)$ and define

$$u_b = \sum_{i,j \in \mathscr{C}_{(b)}(t_n)} \frac{\omega_i \omega_j}{f_b^2(t_n)} d(i,j), \quad B_n^{(i)} = \frac{\sum_{b \in \mathfrak{B}(\bar{\mathscr{C}}_{(i)}(\lambda))} x_b u_b}{\sum_{b \in \mathfrak{B}(\bar{\mathscr{C}}_{(i)}(\lambda))} x_b}.$$

Here $u_b$ gives the average distance within blob $\mathscr{C}_{(b)}(t_n)$. Using Lemma 4.5.7, we will show

$$B_n^{(i)} \times \frac{\sum_i x_i^2}{\sum_i x_i^2 u_i} \xrightarrow{\mathbb{P}} 1. \tag{4.63}$$

The argument is the same as the proof of (4.59). We only have to ensure that (4.58) holds with $y_i = x_i u_i$. Thus, we need to show that

$$\frac{n^{\rho - \delta} \sum_i x_i^3 u_i}{\sum_i x_i \sum_i x_i^2 u_i} \xrightarrow{\mathbb{P}} 0, \quad \text{and} \quad \frac{\sum_i x_i^3 u_i^2 \sum_i x_i}{n^{\rho - \delta} \left(\sum_i x_i^2 u_i\right)^2} \xrightarrow{\mathbb{P}} 0. \tag{4.64}$$

First of all, notice that, by Lemma 4.5.2 and Theorem 4.5.3,

$$
\begin{aligned}
c_n &= \frac{\sum_i x_i^2 u_i}{\sum_i x_i} = (1 + o_{\mathbb{P}}(1)) \frac{\nu n^{-1+\rho}}{\mu(\nu-1)} n^{-2\rho} \sum_b f_b^2(t_n) \sum_{i,j \in b} \frac{\omega_i \omega_j}{f_b^2(t_n)} d(i,j) \\
&= \frac{\nu n^{-1+\rho}}{\mu(\nu-1)} n^{1-2\rho} \mathcal{D}_n^\omega = n^{2\delta-\rho} \frac{\nu-1}{\nu^2} (1 + o_{\mathbb{P}}(1)).
\end{aligned}
$$

(4.65)

Also, recall from Theorem 4.1.4 that $u_{\max} = \max_b u_b = O_{\mathbb{P}}(n^\delta \log(n))$. Now,

$$
\begin{aligned}
\frac{n^{\rho-\delta} \sum_i x_i^3 u_i}{\sum_i x_i \sum_i x_i^2 u_i} &\leqslant \frac{n^{\rho-\delta} u_{\max} \sum_i x_i^3}{\sum_i x_i \sum_i x_i^2 u_i} = O_{\mathbb{P}}\left( \frac{n^{\rho-\delta} n^\delta \log(n) n^{-3\rho} n^{3\alpha+3\delta}}{n^{1-\rho} n^{2\delta-\rho} n^{1-\rho}} \right) \\
&= O_{\mathbb{P}}(n^{\delta-\eta} \log(n)) = o_{\mathbb{P}}(1), \\
\frac{\sum_i x_i^3 u_i^2 \sum_i x_i}{n^{\rho-\delta} \left( \sum_i x_i^2 u_i \right)^2} &\leqslant \frac{x_{\max} u_{\max} \sum_i x_i}{n^{\rho-\delta} \sum_i x_i^2 u_i} = O_{\mathbb{P}}\left( \frac{n^{-\rho} n^{\alpha+\delta} n^\delta \log(n)}{n^{\rho-\delta} n^{2\delta-\rho}} \right) \\
&= O_{\mathbb{P}}(n^{\delta-\eta} \log(n)) = o_{\mathbb{P}}(1),
\end{aligned}
$$

and (4.64) follows, and hence the proof of (4.63) also follows. Recall that the superstructure of $\bar{\mathcal{G}}_n(t_c(\lambda))$ has the same distribution as a $NR_n(x, q)$ random graph with the parameters given by (4.57). Thus, using Proposition 4.3.2, we now aim to use Theorem 4.4.2 on $\mathscr{C}_{(i)}^{fr}(\lambda)$ with the blobs being $(\mathscr{C}_{(i)}(t_n))_{i \geqslant 1}$, and $p_n^{(i)}$, $a_n^{(i)}$ given by (4.12). Define $\Upsilon_n^{(i)} = \left( p_b / \sigma(p_n^{(i)}) : b \in \mathfrak{B}(\bar{\mathscr{C}}_{(i)}(\lambda)) \right)$. Let $\mathcal{N}(\mathbb{R}_+)$ denote the space of all counting measures equipped with the vague topology and denote the product space $\mathbb{S} = \mathbb{R}_+^3 \times \mathcal{N}(\mathbb{R}_+)$. Define

$$
\mathscr{P}_n = \left( a_n^{(i)} \sigma(p_n^{(i)}), \sum_{b \in \mathfrak{B}(\mathscr{C}_{(i)}(\lambda))} x_b, \frac{1}{\sigma_2^2(x^n)} \sum_{b \in \mathfrak{B}(\mathscr{C}_{(i)}(\lambda))} x_b^2, \Upsilon_n^{(i)} \right)_{i \geqslant 1},
$$

viewed as an element of $\mathbb{S}^{\mathbb{N}}$. Recall the definition of $\xi_i^*$ and $\Xi_i^*$ from Section 4.3.5. Define

$$
\mathscr{P}^\infty = \left( \frac{\xi_i^*}{\mu(\nu-1)} \left( \sum_{\nu \in \Xi_i^*} \theta_\nu^2 \right)^{1/2}, \xi_i^*, \frac{1}{\mu^2(\nu-1)^2} \sum_{\nu \in \Xi_i^*} \theta_\nu^2, \left( \frac{\theta_j}{\sum_{\nu \in \Xi_i^*} \theta_\nu^2} : j \in \Xi_i^* \right) \right)_{i \geqslant 1}
$$

The following is a consequence of [36, Proposition 5.1, Lemma 5.4]:

$$
\sigma(p_n^{(i)}) \xrightarrow{\mathbb{P}} 0, \quad \text{and} \quad \mathscr{P}_n \xrightarrow{d} \mathscr{P}_\infty \text{ on } \mathbb{S}^{\mathbb{N}}.
$$

(4.66)

Without loss of generality, we assume that the convergence in (4.66) holds

almost surely. Now, using (4.63), it follows that

$$
\frac{\sigma(\mathbf{p}_n^{(i)})}{1 + B_n^{(i)}} = \frac{\sigma_2(\mathbf{x}^n)\big(\sum_{\nu\in\Xi_i^*}\theta_\nu^2\big)^{1/2}}{\mu(\nu-1)\xi_i^*} \times \frac{\sum_i x_i^2}{\sum_i x_i^2 u_i}(1 + o(1))
$$

$$
= \frac{\sigma_2^2(\mathbf{x}^n)\big(\sum_{\nu\in\Xi_i^*}\theta_\nu^2\big)^{1/2}}{\mu(\nu-1)\xi_i^*\sum_i x_i^2 u_i}(1 + o(1)) = n^{-\eta}\frac{\nu-1}{\nu}\frac{1}{\xi_i^*}\Big(\sum_{\nu\in\Xi_i^*}\theta_\nu^2\Big)^{1/2}(1 + o(1)),
$$

where the last step follows from Theorem 4.5.3, (4.65) and (4.66). The proof
of Theorem 4.5.8 is now complete using Theorem 4.4.2. $\qquad\square$

### 4.5.4    Properties of the original process

Let us denote the ordered components of $\mathcal{G}_n(t_c(\lambda))$ simply by $(\mathscr{C}_{(i)}(\lambda))_{i\geqslant 1}$.
To prove Theorem 4.1.1, we need to compare functionals of $\mathscr{C}_{(i)}(\lambda)$ and $\bar{\mathscr{C}}_{(i)}(\lambda)$
that describe the structures of these graphs. Firstly, the following is a direct
consequence of Lemma 4.5.1 and Proposition 4.3.4:

**Proposition 4.5.9.** *Let* $(\mathscr{C}_{(i)}(\lambda))_{i\geqslant 1}$ *denote the ordered vector of components sizes
of the graph* $\mathcal{G}_n(t_c(\lambda))$. *Then,* $\big(n^{-\rho}|\mathscr{C}_{(i)}(\lambda)|, SP(\mathscr{C}_{(i)}(\lambda))\big)_{i\geqslant 1} \xrightarrow{d} (\frac{1}{\nu}\xi_i, \mathscr{N}_i)_{i\geqslant 1}$ *as*
$n \to \infty$, *with respect to the topology on* $\ell_{\downarrow}^2 \times \mathbb{N}^{\mathbb{N}}$, *where the limiting objects are
defined in Proposition 4.3.4.*

Now, conditionally on $\mathcal{G}_n(t_n)$, $\mathscr{C}_{(i)}(\lambda)$ can also be viewed as consisting of
blobs $(\mathscr{C}_{(i)}(t_n))_{i\geqslant 1}$ and a superstructure connecting the blobs. Denote

$$
\mathcal{F}_i(\lambda) = \sum_{b\in\mathfrak{B}(\mathscr{C}_{(i)}(\lambda))} f_b(t_n).
$$

The components consist of surplus edges within the blobs and the surplus
edges in the superstructure. Let $SP'(\mathscr{C}_{(i)}(\lambda))$ denote the number of surplus
edges in the superstructure of $\mathscr{C}_{(i)}(\lambda)$. The following proposition relates the
superstructure and components:

**Proposition 4.5.10.** *Assume that* $\eta/2 < \delta < \eta$. *Then, for each* $1 \leqslant i \leqslant K$, *the
following hold:*

(a) *With high probability,* $SP'(\mathscr{C}_{(i)}(\lambda)) = SP(\mathscr{C}_{(i)}(\lambda))$. *Consequently, there are no
surplus edges within blobs in* $\mathscr{C}_{(i)}(\lambda)$ *with high probability;*

(b) $\mathcal{F}_i(\lambda)/|\mathscr{C}_{(i)}(\lambda)| \xrightarrow{\mathbb{P}} \nu - 1$. *Consequently,* $(n^{-\rho}\mathcal{F}_i(\lambda))_{i\geqslant 1}$ *and* $(n^{-\rho}\bar{\mathcal{F}}_i(\lambda))_{i\geqslant 1}$
*have the same distributional limit as Proposition 4.5.5.*

We start by explaining the idea of the proof. Since $SP'(\mathscr{C}_{(i)}(\lambda)) \leqslant SP(\mathscr{C}_{(i)}(\lambda))$ almost surely, for Part (a) it suffices to show that

$$SP'(\mathscr{C}_{(i)}(\lambda)) \text{ and } SP(\mathscr{C}_{(i)}(\lambda)) \text{ have the same distributional limit.} \qquad (4.67)$$

Let $\mathcal{G}'_n$ denote the graph obtained from $\mathcal{G}_n(t_c(\lambda))$ by shrinking each blob to a single node. Then, $SP'(\cdot)$ can be viewed as the surplus edges in the components of $\mathcal{G}'_n$. The graph $\mathcal{G}'_n$ can also be viewed to be constructed dynamically as in Algorithm 4.4 with the degree sequence being $(f_i(t_n))_{i \geqslant 1}$. In the following, we investigate the relations between $\mathcal{G}_n(t_n)$ and $\mathcal{G}'_n$ carefully. Lemma 4.5.2 implies that the number of unpaired half-edges in $\mathcal{G}_n(t_n)$ that are paired in $\mathcal{G}_n(t_c(\lambda))$ is given by

$$s_1(t_n) - s_1(t_c(\lambda)) = n\mu_n(n^{-\delta} + \lambda n^{-\eta}) + o_{\mathbb{P}}(n^{1-\gamma}), \quad \text{for some } \eta < \gamma. \ (4.68)$$

**Algorithm 4.6.** Define $\pi_n = \frac{v_n}{v_n - 1}(n^{-\delta} + \lambda n^{-\eta})$ and associate $f_i(t_n)$ half-edges to the vertex $i$ of $\mathcal{G}'_n$. Construct the graph $\mathcal{G}'_n(\pi_n)$ as follows:

(S1) Retain each half-edge independently with probability $\pi_n$.

(S2) Create a uniform perfect matching between the retained half-edges and obtain $\mathcal{G}'_n(\pi_n)$ by creating edges corresponding to any two pair of matched half-edges.

In (S1), if the total number of retained half-edges is odd, then add an extra half-edge to vertex 1. However, this will be ignored in the computations since it does not make any difference in the asymptotic computations. Notice that $a_i$, the number of half-edges attached to $i$ that are retained by Algorithm 4.6 (S1), is distributed as $\text{Bin}(f_i(t_n), \pi_n)$, independently for each $i$. Thus the number of half-edges in the graph $\mathcal{G}'_n(\pi_n)$ is distributed as a $\text{Bin}(s_1(t_n), \pi_n)$ random variable. We claim that there exists $\varepsilon_n = o(n^{-\eta})$ and a coupling such that, with high probability

$$\mathcal{G}'_n(\pi_n - \varepsilon_n) \subset \mathcal{G}'_n \subset \mathcal{G}'_n(\pi_n + \varepsilon_n). \qquad (4.69)$$

The proof follows from an identical argument as Lemma 4.5.1 using the estimate (4.68) and standard concentration inequalities for binomial random variables. We skip the proof here and refer the reader to Chapter 2.7. We now continue to analyze $\mathcal{G}'_n(\pi_n)$, keeping in mind that the relation (4.69) allows us make conclusions for $\mathcal{G}'_n$. To analyze the component sizes and the surplus edges of the components of $\mathcal{G}'(\pi_n)$ we first need some regularity conditions on $\mathbf{a}$, the degree sequence of $\mathcal{G}'_n(\pi_n)$, as summarized in the following lemma:

**Lemma 4.5.11.** *For some $\eta/2 < \delta < \eta$, as $n \to \infty$,*

$$n^{-\alpha} a_i \xrightarrow{\mathbb{P}} \frac{\theta_i}{\nu}, \qquad \frac{a_i}{\sum_i a_i} n^{\rho - \delta} \xrightarrow{\mathbb{P}} \frac{\theta_i}{\mu \nu},$$

$$\nu_n(a) = \frac{\sum_i a_i(a_i - 1)}{\sum_i a_i} = 1 + \lambda n^{-\eta + \delta} + o_\mathbb{P}(n^{-\eta + \delta}),$$

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

$$\lim_{K \to \infty} \limsup_{n \to \infty} \mathbb{P}\left( \sum_{i > K} a_i^3 > \varepsilon n^{3\alpha} \right) = 0. \tag{4.70}$$

*Proof.* Using Theorem 4.5.3 and the fact that $a_i \sim \text{Bin}(f_i(t_n), \pi_n)$, one gets $n^{-\alpha} a_i = (1 + o_\mathbb{P}(1))\theta_i/\nu$. Moreover, $\sum_i a_i \sim \text{Bin}(\sum_i f_i(t_n), \pi_n)$ and $\sum_i a_i = (1 + o_\mathbb{P}(1))\pi_n \sum_i f_i(t_n)$ yield the required asymptotics for $a_i / \sum_j a_j$. Let $I_{ij}$ be the indicator of the $j$-th half-edge corresponding to vertex $i$ is kept in Algorithm 4.6 (S1). Then $I_{ij} \sim \text{Ber}(\pi_n)$ independently for $j \in [f_i(t_n)]$, $i \geqslant 1$. Note that, by changing the status of one half-edge corresponding to vertex $k$, we can change $\sum_i a_i(a_i - 1)$ by at most $2(f_k(t_n) + 1)$. Therefore we can apply [115, Corollary 2.27] to conclude that

$$\mathbb{P}\left( \left| \sum_i a_i(a_i - 1) - \pi_n^2 \sum_i f_i(t_n)(f_i(t_n) - 1) \right| > t \,\middle|\, (f_i(t_n))_{i \geqslant 1} \right)$$

$$\leqslant 2 \exp\left( \frac{-t^2}{2 \sum_i f_i(t_n)(f_i(t_n) + 1)^2} \right).$$

Observe that $\pi_n^2 \sum_i f_i(t_n)(f_i(t_n) - 1) = \Theta_\mathbb{P}(n^{1-\delta})$. Take $t = \varepsilon n^{1-\delta}$ and recall that $\sum_i f_i^3(t_n) = \Theta_\mathbb{P}(n^{3\alpha + 3\delta})$. It is easy to check that $(2 - 3\alpha)/5 > \eta/2$, and therefore one can choose $\eta/2 < \delta < \eta$ such that $t^2 / \sum_i f_i^3(t_n) \to \infty$. Thus,

$$\sum_i a_i(a_i - 1) = (1 + o_\mathbb{P}(1))\pi_n^2 \sum_i f_i(t_n)(f_i(t_n) - 1).$$

Therefore, Theorem 4.5.3 yields the required asymptotics for $\nu_n(a)$. To see (4.70), note that $\mathbb{E}\left[ \sum_{i > K} a_i(a_i - 1)(a_i - 2) \mid (f_i(t_n))_{i \geqslant 1} \right] = \pi_n^3 \sum_{i > K} f_i(t_n)^3$, and the proof follows again by using the condition on $s_3^\omega$ in Theorem 4.5.3. $\qquad\square$

From here onward, we assume that $\delta > 0$ is such that Lemma 4.5.11 holds. Consider the exploration of the graph $\mathcal{G}_n'(\pi_n)$ via Algorithm 4.3, but now the first vertex is chosen proportional to its degree. Define the exploration process by $S_n$ similarly as the process $S_n^j(l)$ in Section 4.4.3. Call a vertex

*discovered* if it is either active or killed. Let $\mathscr{V}_l$ denote the set of vertices discovered up to time $l$ and $\mathfrak{I}_i^n(l) := \mathbf{1}\{i \in \mathscr{V}_l\}$. Note that

$$S_n(l) = \sum_i a_i \mathfrak{I}_i^n(l) - 2l = \sum_i a_i \left( \mathfrak{I}_i^n(l) - \frac{a_i}{\ell_n^a} l \right) + (\nu_n(\mathbf{a}) - 1) l, \quad (4.71)$$

where $\ell_n^a = \sum_i a_i$. Consider the re-scaled version $\bar{\mathbf{S}}_n$ of $\mathbf{S}_n$ defined as $\bar{S}_n(t) = n^{-\alpha} S_n(\lfloor t n^{\rho-\delta} \rfloor)$. Define the limiting process

$$S(t) = \sum_{i=1}^{\infty} \frac{\theta_i}{\nu} \left( \mathbf{1}\{E_i \leqslant t\} - \frac{\theta_i}{\mu\nu} t \right) + \lambda t, \quad (4.72)$$

where $E_i \sim \text{Exp}(\theta_i/(\mu\nu))$ independently for $i \geqslant 1$. The following proposition describes the scaling limit of $\bar{\mathbf{S}}_n$:

**Proposition 4.5.12.** *As* $n \to \infty$, $\bar{\mathbf{S}}_n \xrightarrow{d} \mathbf{S}$ *with respect to the Skorohod* $J_1$ *topology.*

The proof of Proposition 4.5.12 can be carried out using similar ideas as Chapter 3 Theorem 3.5.1. A sketch of the proof is given in Appendix 4.D. The excursion lengths of the exploration process give the number of edges in the explored components. Now, at each step $l$, the probability of discovering a surplus edge, conditioned on the past, is approximately the proportion of half-edges that are active. Note that the number of active half-edges is the reflected version of $\mathbf{S}_n$ given by $\text{refl}(S_n(t)) = S_n(t) - \inf_{u \leqslant t} S_n(u)$. Thus, conditionally on $(S_n(l))_{l \leqslant t n^{\rho-\delta}}$, the rate at which a surplus edge appears at time $t n^{\rho-\delta}$ is approximately $n^{\rho-\delta} \frac{\text{refl}(S_n(t n^{\rho-\delta}))}{\sum_i a_i} = \frac{1}{\mu} \text{refl}(\bar{S}_n(t))(1 + o_{\mathbb{P}}(1))$. Therefore, Proposition 4.5.12 implies that for each $K \geqslant 1$, there exists components $C_1, \ldots, C_K \subset \mathcal{G}_n'(\pi_n)$ such that

$$\left( n^{-\rho+\delta} |C_i|, \text{SP}(C_i) \right)_{i \in [K]} \xrightarrow{d} \left( \xi_i, \mathscr{N}_i \right)_{i \in [K]}, \quad (4.73)$$

where $\xi_i$ and $\mathscr{N}_i$ are defined in Proposition 4.3.4. Here we have also used the fact that the ordered excursion lengths of the process $(S(t))_{t \geqslant 0}$, defined in (4.72), are identically distributed as the ordered excursion lengths of $(S(t)/\mu)_{t \geqslant 0}$. Note that $C_i$ in (4.73) may not be the $i$-th largest component of $\mathcal{G}_n'(\pi_n)$ as we have not established that the $i$-th largest component is explored by time $\Theta(n^{\rho-\delta})$. However, that is not required for our purposes. We can now combine (4.69) and (4.73) to obtain the asymptotics for the number of blobs in the largest connected components and $\text{SP}'(\cdot)$. Denote $\mathscr{B}(\mathscr{C}) = |\mathfrak{B}(\mathscr{C})|$ for a component $\mathscr{C} \subset \mathcal{G}_n(t_c(\lambda))$. The following is a direct consequence of (4.69) and (4.73):

**Lemma 4.5.13.** *For* $K \geqslant 1$, *there exist components* $\mathscr{C}_1, \ldots, \mathscr{C}_K \subset \mathcal{G}_n(t_c(\lambda))$ *such that the following convergence holds:*

$$(n^{-\rho+\delta}\mathscr{B}(\mathscr{C}_i), \mathrm{SP}'(\mathscr{C}_i))_{i \in [K]} \xrightarrow{d} (\xi_i, \mathcal{N}_i)_{i \in [K]}.$$

Next we show that the components $\mathscr{C}_i$ in Lemma 4.5.13 indeed correspond to the $i$-th largest component of $\mathcal{G}_n(t_c(\lambda))$:

**Lemma 4.5.14.** *For any* $K \geqslant 1$, $\mathscr{C}_i = \mathscr{C}_{(i)}(\lambda)$, $\forall i \in [K]$ *with high probability.*

*Proof.* Notice that $\sum_{j \leqslant i} |\mathscr{C}_j| \leqslant \sum_{j \leqslant i} |\mathscr{C}_{(j)}(\lambda)|$ for all $i \in [K]$, almost surely. Thus, it is enough to prove that $|\mathscr{C}_i|$ and $|\mathscr{C}_{(i)}(\lambda)|$ involve the same re-scaling factor and have the same scaling limit. We again make use of the inclusions in graphs in (4.69). Algorithm 4.3 explores the components of $\mathcal{G}'_n(\pi_n)$ in a size-biased manner with the sizes being $(a_i)_{i \geqslant 1}$. An application of Lemma 4.5.7 with $y_i = \mathscr{C}_{(i)}(t_n)$ yields that, for any $t > 0$, uniformly for $l \leqslant tn^{\rho-\delta}$,

$$\sum_i |\mathscr{C}_{(i)}(t_n)| \mathcal{I}_i^n(l) = \sum_i |\mathscr{C}_{(i)}(t_n)| \frac{a_i}{\sum_i a_i} l + o_{\mathbb{P}}(n^\rho). \tag{4.74}$$

Since $a_i \sim \mathrm{Bin}(f_i(t_n), \pi_n)$, we can apply concentration inequalities like [115, Corollary 2.27] and use the asymptotics from Theorem 4.5.3 to conclude that

$$n^{-\delta} \frac{\sum_i a_i |\mathscr{C}_{(i)}(t_n)|}{\sum_i a_i} = \frac{\frac{\mu(\nu-1)}{\nu^2}}{\frac{\mu(\nu-1)}{\nu}}(1 + o_{\mathbb{P}}(1)) = \frac{1}{\nu}(1 + o_{\mathbb{P}}(1)). \tag{4.75}$$

Thus, (4.74) and (4.75), together with (4.69), imply that $\frac{\nu |\mathscr{C}_i|}{n^\delta \mathscr{B}(\mathscr{C}_i)} \xrightarrow{\mathbb{P}} 1$, and it follows from Lemma 4.5.13 and Lemma 4.3.5 that $(n^{-\rho}|\mathscr{C}_i|)_{i \in [K]} \xrightarrow{d} (\frac{1}{\nu}\xi_i)_{i \in [K]}$. $\square$

*Proof of Proposition 4.5.10.* We are now finally in the position to prove Proposition 4.5.10. Using Lemmas 4.5.13, 4.5.14, and Proposition 4.5.9 together with (4.69), we directly conclude Part (a) from (4.67). For Part (b), we can follow the same arguments as (4.74) to conclude that, uniformly for $l \leqslant tn^{\rho-\delta}$,

$$\sum_i f_i(t_n) \mathcal{I}_i^n(l) = \sum_i f_i(t_n) \frac{a_i}{\sum_i a_i} l + o_{\mathbb{P}}(n^\rho), \tag{4.76}$$

where $n^{-\delta} \frac{\sum_i a_i f_i(t_n)}{\sum_i a_i} = \frac{\nu-1}{\nu}(1 + o_{\mathbb{P}}(1))$. Now, (4.74) and (4.76) together with (4.69) prove Part (b). $\square$

In the final part of the proof, we will also need an estimate of the surplus edges in the components $\bar{\mathscr{C}}_{(i)}(\lambda)$, that can be obtained by following the

exact same argument as the proof outline of Lemma 4.5.13. Recall that the superstructure on the graph $\bar{\mathcal{G}}_n(t_c(\lambda))$ is a rank-one inhomogeneous random graph $NR_n(\boldsymbol{x}, \boldsymbol{q})$. The connection probabilities given by (4.56) can be written as $1 - \exp(-z_i(\lambda)z_j(\lambda)/\sum_k z_k(\lambda))$, where

$$z_i(\lambda) = \frac{f_i(t_n)\sum_j f_j(t_n)}{\sum_j f_j^2(t_n)}\left(1 + \lambda n^{-\eta+\delta} + o_{\mathbb{P}}(n^{-\eta+\delta})\right).$$

Moreover, using Theorem 4.5.3, it follows that

$$n^{-\alpha}z_i(\lambda) \xrightarrow{\mathbb{P}} \frac{\theta_i}{\nu}, \qquad \frac{z_i(\lambda)}{\sum_j z_j(\lambda)} \xrightarrow{\mathbb{P}} \frac{\theta_i}{\mu\nu},$$

$$\nu_n(\boldsymbol{z}) = \frac{\sum_i z_i^2(\lambda)}{\sum_i z_i(\lambda)} = 1 + \lambda n^{-\eta+\delta} + o_{\mathbb{P}}(n^{-\eta+\delta}).$$

Now, we may consider the breadth-first exploration of the above graph and define the exploration process $S_n^{NR}(l) = \sum_i z_i(\lambda)\mathcal{I}_i^n(l) - l$, as in (4.71). The only thing to note here is that the component sizes are not necessarily encoded by the excursion lengths above the past minima of $\mathbf{S}_n^{NR}$. However, if $\tilde{S}_n^{NR}(l) = \sum_i \mathcal{I}_i^n(l) - l$, then it can be shown that (see [33, Lemma 3.1]) $\tilde{\mathbf{S}}_n^{NR}$ and $\mathbf{S}_n^{NR}$ have the same distributional limit. Thus, a conclusion identical to Proposition 4.5.12 follows for $\bar{\mathcal{G}}_n(t_c(\lambda))$. Due to the size-biased exploration of the components one can also obtain analogues of Lemmas 4.5.13 and 4.5.14 for $\bar{\mathcal{G}}_n(t_c(\lambda))$. This explains the following proposition:

**Proposition 4.5.15.** *For fixed* $K \geqslant 1$, $(n^{-\rho+\delta}\mathscr{B}(\bar{\mathscr{C}}_{(i)}(\lambda)), SP'(\bar{\mathscr{C}}_{(i)}(\lambda)))_{i \in [K]} \xrightarrow{d} (\xi_i, \mathscr{N}_i)_{i \in [K]},$ *as* $n \to \infty$.

### 4.5.5    Completing the proof of Theorem 4.1.1

In this section, we finally conclude the proof of Theorem 4.1.1. Recall Theorem 4.5.8 and the terminologies therein. Let $n^{-\eta}\mathscr{C}_{(i)}^{fr}(\lambda)$ denote the measured metric space with measure $\mu_{fr}^i$ and the distances multiplied by $n^{-\eta}$. At this moment, let us recall the relevant properties $\mathscr{C}_{(i)}(\lambda)$ and $\bar{\mathscr{C}}_{(i)}(\lambda)$:

(A) By Proposition 4.5.4, $\cup_{j\leqslant i}\mathscr{C}_{(j)}(\lambda) \subset \cup_{j\leqslant i}\bar{\mathscr{C}}_{(j)}(\lambda)$ almost surely for any $i \geqslant 1$. Therefore, applying Propositions 4.5.6 and 4.5.9, it follows that with high probability $\mathscr{C}_{(i)}(\lambda) \subset \bar{\mathscr{C}}_{(i)}(\lambda)$ for any fixed $i \geqslant 1$.

(B) By Proposition 4.5.10 (b), $\bar{\mathcal{F}}_i(\lambda) - \mathcal{F}_i(\lambda) \xrightarrow{\mathbb{P}} 0$ and consequently $\mu_{fr}^i(\bar{\mathscr{C}}_{(i)}(\lambda) \setminus \mathscr{C}_{(i)}(\lambda)) \xrightarrow{\mathbb{P}} 0.$

(C) By Propositions 4.5.10 (a) and 4.5.15, the number of surplus edges with one endpoint in $\mathscr{C}_{(i)}(\lambda) \setminus \bar{\mathscr{C}}_{(i)}(\lambda)$ converges in probability to zero. Moreover, with high probability there is no surplus edge within the blobs. This implies that, for any pair of vertices $u, v \in \mathscr{C}_{(i)}(\lambda)$, with high probability, the shortest path between them is exactly the same in $\mathscr{C}_{(i)}(\lambda)$ and in $\bar{\mathscr{C}}_{(i)}(\lambda)$.

Thus, from the definition of Gromov-weak convergence in Section 4.3.1, an application of Theorem 4.5.8 yields that $(n^{-\eta}\mathscr{C}_{(i)}^{\mathrm{fr}})_{i \geqslant 1} \xrightarrow{d} (M_i)_{i \geqslant 1}$, The only thing remaining to show is that we can replace the measure $\mu_{\mathrm{fr}}^i$ by $\mu_{\mathrm{ct},i}$. Now, using Propositions 4.5.9 and 4.5.10 (b), it is enough to show that

$$\sum_{b \in \mathscr{B}(\mathscr{C}_{(i)}(\lambda))} \left| f_b(t_n) - (\nu-1)|\mathscr{C}_{(b)}(t_n)| \right| = o_{\mathbb{P}}(n^\rho). \qquad (4.77)$$

Indeed, during the breadth-first exploration of the superstructure of $\mathcal{G}_n(t_c(\lambda))$, the blobs are explored in a size-biased manner with the sizes being $(f_i(t_n))_{i \geqslant 1}$. Therefore, one can again use Lemma 4.5.7. Recall that, by Lemma 4.5.14, for any $\varepsilon > 0$, one can choose $T > 0$ so large that the probability of exploring $\mathscr{C}_{(i)}(\lambda)$ within time $Tn^{\rho-\delta}$ is at least $1 - \varepsilon$. Thus, if $\mathscr{V}_l^b$ denotes the set of blobs explored before time $l$, then, for any $T > 0$,

$$\sum_{b \in \mathscr{V}_{Tn^{\rho-\delta}}^b} \left| f_b(t_n) - (\nu-1)|\mathscr{C}_{(b)}(t_n)| \right|$$

$$= (1 + o_{\mathbb{P}}(1)) Tn^{\rho-\delta} \sum_i \frac{f_i(t_n)}{\sum_i f_i(t_n)} \left| f_i(t_n) - (\nu-1)|\mathscr{C}_{(i)}(t_n)| \right|.$$

Using the Cauchy-Schwarz inequality and Theorem 4.5.3 it now follows that the above term is $o(n^\rho)$. Therefore (4.77) follows. Finally the proof of Theorem 4.1.1 is complete using Lemma 4.5.1. $\qquad \square$

**Remark 37.** The fact that the measure can be changed from $\mu_{\mathrm{fr}}^i$ to $\mu_{\mathrm{ct},i}$ in $n^{-\eta}\mathscr{C}_{(i)}^{\mathrm{fr}}$ follows only from (4.77), which again follows from the entrance boundary conditions. However, the entrance boundary conditions in Theorem 4.1.3 hold for weight sequences $\boldsymbol{w} = (w_i)_{i \in [n]}$ under general assumptions (see Assumption 4.2). Therefore, one could also replace the measure $\mu_{\mathrm{ct},i}$ by $\mu_{w,i}$, where $\mu_{w,i} = \sum_{i \in A} w_i / \sum_{k \in \mathscr{C}_{(i)}(\lambda)} w_i$ and $\boldsymbol{w}$ satisfies Assumption 4.2.

### 4.5.6    Proof of Theorem 4.1.2

The argument is related to Chapter 3.7. Using [85, Lemma 3.2], the random graph $CM_n(\mathbf{d}, p_n(\lambda))$, conditionally on its degree sequence $\mathbf{d}^p$, is distributed as $CM_n(\mathbf{d}^p)$. To complete the proof of Theorem 4.1.2, consider the exploration algorithm given by Algorithm 4.3, now on the graph $CM_n(\mathbf{d}, p_n(\lambda))$, conditionally on the degree sequence $\mathbf{d}^p$. The starting vertex is chosen in a size biased manner with sizes proportional to the degrees $\mathbf{d}^p$. For convenience, we denote $X = (\mathscr{C}_{(i)}^p(\lambda))_{i \leqslant K}$ in this section. Consider a bounded continuous function $f : (\mathscr{S}_*)^K \mapsto \mathbb{R}$, where we recall $\mathscr{S}_*$ from Section 4.3.1. Recall from [110, Theorem 1.1] that

$$\liminf_{n \to \infty} \mathbb{P}\left(CM_n(\mathbf{d}) \text{ is simple}\right) > 0.$$

Thus, it is enough to show that

$$\mathbb{E}\left[f(X)\mathbf{1}\{CM_n(\mathbf{d}) \text{ is simple}\}\right] - \mathbb{E}\left[f(X)\right]\mathbb{P}\left(CM_n(\mathbf{d}) \text{ is simple}\right) \to 0. \quad (4.78)$$

Now, for any $T > 0$, let $\mathcal{A}_{n,T}$ denote the event that $X$ is explored before time $Tn^\rho$ by the exploration algorithm. Using [69, Lemma 13], it follows that

$$\lim_{T \to \infty} \limsup_{n \to \infty} \mathbb{P}\left(\mathcal{A}_{n,T}^c\right) = 0.$$

Let $X_T$ denote the random vector consisting of $K$ largest ones among the components explored before time $Tn^\rho$. Thus,

$$\lim_{T \to \infty} \limsup_{n \to \infty} \mathbb{E}\left[f(X)\mathbf{1}\{CM_n(\mathbf{d}) \text{ is simple}\}\mathbf{1}_{\mathcal{A}_{n,T}^c}\right]$$
$$\leqslant \|f\|_\infty \lim_{T \to \infty} \limsup_{n \to \infty} \mathbb{P}\left(\mathcal{A}_{n,T}^c\right) = 0,$$

which implies that

$$\lim_{T \to \infty} \limsup_{n \to \infty} \left|\mathbb{E}\left[(f(X) - f(X_T))\mathbf{1}\{CM_n(\mathbf{d}) \text{ is simple}\}\right]\right| = 0. \quad (4.79)$$

Further, let $\mathcal{B}_{n,T}$ denote the event that a vertex $v$ is explored before time $Tn^\rho$ such that $v$ is involved in a self-loop or a multiple edge in $CM_n(\mathbf{d})$. For any fixed vertex $v$, the $i$-th half edge creates a self-loop in $CM_n(\mathbf{d})$ with probability at most $(d_v - i)/(\ell_n - 1)$ and creates a multiple edge with probability at most $(i-1)/(\ell_n - 1)$ so that the probability of $v$ creating a self-loop or a multiple edge is at most $d_v^2/(\ell_n - 1)$. Let $\mathcal{I}_i^n(l)$ denote the indicator that vertex $i$ is discovered upto time $l$ and note that Algorithm 4.3 will explore the vertices in a size-biased manner with sizes being $\mathbf{d}^p$. Let $\mathbb{P}_p$ (respectively $\mathbb{E}_p$) denote

the conditional probability (respectively expectation), conditionally on $\mathbf{d}^p$. Thus,

$$\mathbb{P}_p(\mathcal{B}_{n,T}) \leqslant \frac{1}{\ell_n - 1} \mathbb{E}_p \left[ \sum_{i \in [n]} d_i^2 \mathcal{I}_i^n(Tn^\rho) \right]$$

$$= \frac{1}{\ell_n - 1} \left( \mathbb{E}_p \left[ \sum_{i=1}^{K} d_i^2 \mathcal{I}_i^n(Tn^\rho) \right] + \mathbb{E}_p \left[ \sum_{i=K+1}^{n} d_i^2 \mathcal{I}_i^n(Tn^\rho) \right] \right).$$

Now, using Assumption 4.1, for every fixed $K \geqslant 1$,

$$\frac{1}{\ell_n - 1} \mathbb{E}_p \left[ \sum_{i=1}^{K} d_i^2 \mathcal{I}_i^n(Tn^\rho) \right] \leqslant \frac{2}{\ell_n} \sum_{i=1}^{K} d_i^2 \xrightarrow{\mathbb{P}} 0.$$

Further, $\mathbb{P}_p(\mathcal{I}_i^n(Tn^\rho) = 1) \leqslant Tn^\rho d_i^p / (\sum_{i \in [n]} d_i^p - 2Tn^\rho)$. Therefore,

$$\frac{2}{\ell_n} \mathbb{E}_p \left[ \sum_{i=K+1}^{n} d_i^2 \mathcal{I}_i^n(Tn^\rho) \right] \leqslant \frac{2Tn^\rho}{\ell_n \sum_{i \in [n]} d_i^p} \sum_{i=K+1}^{n} d_i^2 d_i^p$$

$$\leqslant O_\mathbb{P}(1) \left( n^{-3\alpha} \sum_{i=K+1}^{n} d_i^3 \right), \tag{4.80}$$

where the last step follows using $\sum_{i \in [n]} d_i^p \sim 2 \times \mathrm{Bin}(\ell_n/2, p_n(\lambda))$, standard concentration inequalities for the binomial distribution, and the fact that $d_i^p \leqslant d_i$ for all $i \in [n]$. Now, by Assumption 4.1, the final term in (4.80) tends to zero in probability if we first take $\limsup_{n \to \infty}$ and then take $\lim_{K \to \infty}$. Consequently, for any fixed $T > 0$,

$$\lim_{n \to \infty} \mathbb{P}(\mathcal{B}_{n,T}) = 0. \tag{4.81}$$

Let $\mathcal{E}_{n,T}$ denote the event that no self-loops or multiple edges are attached to the vertices in $CM_n(\mathbf{d})$ that are discovered after time $Tn^\rho$. Then (4.79) and (4.81) implies that

$$\lim_{n \to \infty} \mathbb{E}\left[ f(X) \mathbf{1} \{ CM_n(\mathbf{d}) \text{ is simple} \} \right] = \lim_{T \to \infty} \lim_{n \to \infty} \mathbb{E} \left[ f(X_T) \mathbf{1}_{\mathcal{E}_{n,T}} \right]$$

$$= \lim_{T \to \infty} \lim_{n \to \infty} \mathbb{E} \left[ f(X_T) \mathbb{P}(\mathcal{E}_{n,T} | \mathscr{F}_{Tn^\rho}) \right] \tag{4.82}$$

$$= \lim_{T \to \infty} \lim_{n \to \infty} \mathbb{E} \left[ f(X_T) \mathbb{P}(\mathcal{E}_{n,T} | \mathscr{F}_{Tn^\rho}, \mathcal{B}_{n,T}) \right].$$

Let $\mathcal{G}_{Tn^\rho}^*$ denote the graph obtained from $CM_n(\mathbf{d})$ after removing the vertices discovered upto time $Tn^\rho$. Then $\mathcal{G}_{Tn^\rho}^*$ is distributed as a configuration model conditional on its degree sequence. Thus conditional on $\mathscr{F}_{Tn^\rho} \cap \mathcal{B}_{n,T}$, $\mathcal{E}_{n,T}$

happens if and only if $\mathcal{G}^*_{\mathsf{T}n\rho}$ is simple. Now, an argument similar to (3.55) in Chapter 3 can be applied to conclude that

$$\mathbb{P}\left(\mathcal{G}^*_{\mathsf{T}n\rho} \text{ is simple}\,|\,\mathscr{F}_{\mathsf{t}n\rho}\right) - \mathbb{P}\left(\mathrm{CM}_n(\mathbf{d}) \text{ is simple}\right) \xrightarrow{\mathbb{P}} 0,$$

and using (4.82), (4.78) follows, and the proof of Theorem 4.1.2 is now complete. □

## 4.6    Conclusion

We have obtained the scaling limit for the metric structure of the ordered component sizes of the critical percolation clusters for $\mathrm{CM}_n(\mathbf{d})$ when the empirical degree distribution has diverging third moment. The key ingredient of the proof is a universality principle in Theorem 4.4.2, which basically says that after replacing each nodes by metric spaces having small diameter, the scaling limit for the rank-one inhomogeneous random graphs does not change even if typical distances change. This work provides a general framework to establish the scaling limit for networks which are in the same universality class as identified in [36]. The overall idea for percolation on $\mathrm{CM}_n(\mathbf{d})$ is not very specific to the underlying model, and could be applicable to other types of inhomogeneous random graphs. An analogous framework for the Erdős-Rényi universality class was established in [29]. The underlying topology for the convergence of metric spaces in [29] was taken to be Gromov-Hausdorff-Prokhorov topology, which turns out to be strictly stronger than the Gromov-weak topology considered here. In the next chapter, we strengthen Theorem 4.1.1 to Gromov-Hausdorff-Prokhorov topology under some additional mild assumptions on the degree sequence.

## 4.A    Proof of Proposition 4.3.4

Note that due to the difference in the choice of $p_n(\lambda)$ in Chapter 3 Assumption 3.1 and this chapter, $\lambda$ must be replaced by $\lambda\nu$. Let $\mathcal{E}(\cdot)$ denote the operator that maps a process to its ordered vector of excursion lengths, and $\mathcal{A}(\cdot)$ maps a process to the vector of areas under those excursions. Let us use $\mathrm{Exp}(b)$ as a generic notation to write an exponential random variable with rate $b$. All the different exponential random variables will be assumed to

independent. Now,

$$\frac{1}{\sqrt{\nu}}\mathcal{E}\left(\sum_{i\geqslant 1}\frac{\theta_i}{\sqrt{\nu}}\left(\mathbf{1}\{\text{Exp}(\theta_i/(\mu\sqrt{\nu}))\leqslant t\}-(\theta_i/(\mu\sqrt{\nu}))t)+\lambda\nu t\right)$$

$$\overset{d}{=}\frac{1}{\nu}\mathcal{E}\left(\sum_{i\geqslant 1}\frac{\theta_i}{\sqrt{\nu}}\left(\mathbf{1}\{\text{Exp}(\theta_i/(\mu\nu))\leqslant u\}-(\theta_i/(\mu\nu))u)+\lambda u\sqrt{\nu}\right)$$

$$\overset{d}{=}\frac{1}{\nu}\mathcal{E}\left(\sum_{i\geqslant 1}\frac{\theta_i}{\mu\nu}\left(\mathbf{1}\{\text{Exp}(\theta_i/(\mu\nu))\leqslant u\}-(\theta_i/(\mu\nu))u)+\frac{\lambda}{\mu}u\right),$$

where the last step follows by rescaling the space by $\mu\sqrt{\nu}$ and noting that the rescaling of space does not affect excursion lengths. Again,

$$\mathcal{A}\left(\sum_{i\geqslant 1}\frac{\theta_i}{\mu\sqrt{\nu}}\left(\mathbf{1}\{\text{Exp}(\theta_i/(\mu\sqrt{\nu}))\leqslant t\}-(\theta_i/(\mu\sqrt{\nu}))t)+\frac{\lambda\nu}{\mu}t\right)$$

$$\overset{d}{=}\mathcal{A}\left(\sum_{i\geqslant 1}\frac{\theta_i}{\mu\nu}\left(\mathbf{1}\{\text{Exp}(\theta_i/(\mu\nu))\leqslant u\}-(\theta_i/(\mu\nu))t)+\frac{\lambda}{\mu}u\right),$$

which is obtained by rescaling both the space and time by $\sqrt{\nu}$. Thus, the proof follows.

## 4.B   Computation for $\mathbb{E}[\mathscr{W}(V_n^*)^3]$

Recall that

$$\left(\mathscr{W}(V_n^*)\right)^r=\sum_{k_1,\cdots,k_r\in[n]}w_{k_1}\dots w_{k_r}\mathbf{1}\{V_n^*\rightsquigarrow k_1,\dots,V_n^*\rightsquigarrow k_r\}.$$

For the third moment, the leading contributions arise from the structures given in Figure 3. Thus,

$$\mathbb{E}\left[\left(\mathscr{W}(V_n^*)\right)^3\right]$$

$$\leqslant\mathbb{E}\left[(W_n^*)^3\right]+\frac{\mathbb{E}[D_n^*]\left(\mathbb{E}[D_n'W_n]\right)^3\sigma_3(n)^2}{(\mathbb{E}[D_n'])^5(1-\nu_n')^5}$$

$$+\frac{\mathbb{E}[D_n^*(D_n^*-1)(D_n^*-2)]\left(\mathbb{E}[D_n'W_n]\right)^3}{(\mathbb{E}[D_n'])^3(1-\nu_n')^3}$$

$$+\frac{\mathbb{E}[D_n^*]\left(\mathbb{E}[D_n'W_n]\right)^3\sigma_4(n)}{(\mathbb{E}[D_n'])^4(1-\nu_n')^4}+\frac{\mathbb{E}[D_n^*(D_n^*-1)]\left(\mathbb{E}[D_n'W_n]\right)^3\sigma_3(n)}{(\mathbb{E}[D_n'])^4(1-\nu_n')^4}$$

$$=O(n^{4\alpha-1}+n^{6\alpha-2+3\delta}+n^{4\alpha-1+3\delta}+n^{4\alpha-1+4\delta}+n^{6\alpha-2+4\delta})=o(n^{2\delta+1}),$$

and the proof follows.

## 4.C   Proofs of Lemmas 4.4.9 and 4.4.10

*Proof of Lemma 6.3.* Recall the representation of $\bar{S}_n^j(t)$ from (4.47). It is enough to show that

$$\sup_{t\in[0,T]} n^{-\alpha}\left|\sum_{i\in[n]} d_i'\left(\mathcal{I}_i^n(tn^{\alpha+\delta}) - \frac{d_i'}{\ell_n'}tn^{\alpha+\delta}\right)\right|$$

$$= \sup_{t\in[0,T]} n^{-\alpha}|M_n(tn^{\alpha+\delta})| \xrightarrow{\mathbb{P}} 0.$$

Fix any $T > 0$ and define $\ell_n'(T) = \ell_n' - 2Tn^{\alpha+\delta} - 1$, and $M_n'(l) = \sum_{i\in[n]} d_i'(\mathcal{I}_i^n(l) - (d_i/\ell_n'(T))l)$. Note that

$$\sup_{t\in[0,T]} n^{-\alpha}|M_n(tn^{\alpha+\delta}) - M_n'(tn^{\alpha+\delta})|$$

$$\leqslant Tn^\delta \frac{(2Tn^{\alpha+\delta} - 1)\sum_{i\in[n]} d_i'^2}{\ell_n'(T)^2} = o_{\mathbb{P}}(1),$$

and thus the proof reduces to showing that

$$\sup_{t\in[0,T]} n^{-\alpha}|M_n'(tn^{\alpha+\delta})| \xrightarrow{\mathbb{P}} 0. \tag{4.83}$$

Note that, uniformly over $l \leqslant Tn^{\alpha+\delta}$,

$$\mathbb{P}\left(\mathcal{I}_i^n(l+1) = 1 \mid \mathscr{F}_l\right) \leqslant \frac{d_i'}{\ell_n'(T)} \quad \text{on the set } \{\mathcal{I}_i^n(l) = 0\}.$$

Therefore,

$$\mathbb{E}\left[M_n'(l+1) - M_n'(l) \mid \mathscr{F}_l\right]$$

$$= \mathbb{E}\left[\sum_{i\in[n]} n^{-\alpha}d_i'\left(\mathcal{I}_i^n(l+1) - \mathcal{I}_i^n(l) - \frac{d_i'}{\ell_n'(T)}\right)\Big|\mathscr{F}_l\right]$$

$$= \sum_{i\in[n]} n^{-\alpha}d_i'\left(\mathbb{E}\left[\mathcal{I}_i^n(l+1)|\mathscr{F}_l\right]\mathbf{1}\{\mathcal{I}_i^n(l) = 0\} - \frac{d_i'}{\ell_n'(T)}\right) \leqslant 0.$$

Thus $(M_n'(l))_{l=1}^{Tn^{\alpha+\delta}}$ is a super-martingale. Further, uniformly for all $l \leqslant Tn^{\alpha+\delta}$,

$$\mathbb{P}\left(\mathcal{I}_i^n(l) = 0\right) \leqslant \left(1 - \frac{d_i'}{\ell_n'}\right)^l. \tag{4.84}$$

Thus, Assumption 2 gives

$$
n^{-\alpha}\big|\mathbb{E}[M'_n(l)]\big|
$$

$$
\leqslant n^{-\alpha}\sum_{i\in[n]}d'_i\left(1-\left(1-\frac{d'_i}{\ell'_n}\right)^l-\frac{d'_i}{\ell'_n}l\right)+n^{-\alpha}l\sum_{i\in[n]}d'^2_i\left(\frac{1}{\ell'_n(T)}-\frac{1}{\ell'_n}\right)
$$

$$
\leqslant \frac{l^2}{2\ell'^2_n n^{\alpha}}\sum_{i\in[n]}d'^3_i+o(1)=o(1),
$$

where we have used the fact that

$$
n^{-\alpha}l\sum_{i\in[n]}d'^2_i(1/\ell'_n(T)-1/\ell'_n)=O(n^{2\rho+1-\alpha-2})=O(n^{(\tau-4)/(\tau-1)}),
$$

uniformly for $l\leqslant Tn^{\alpha+\delta}$ and, in the last step, that fact that $\delta<\eta$. Therefore, uniformly over $l\leqslant Tn^{\alpha+\delta}$,

$$
\lim_{n\to\infty}\big|\mathbb{E}[M'_n(l)]\big|=0. \tag{4.85}
$$

Now, note that for any $(x_1,x_2,\dots)$, $0\leqslant a+b\leqslant x_i$ and $a,b>0$ one has $\prod_{i=1}^{R}(1-a/x_i)(1-b/x_i)\geqslant\prod_{i=1}^{R}(1-(a+b)/x_i)$. Thus, for all $l\geqslant 1$ and $i\neq j$,

$$
\mathbb{P}\left(\mathcal{I}^n_i(l)=0,\mathcal{I}^n_j(l)=0\right)\leqslant\mathbb{P}\left(\mathcal{I}^n_i(l)=0\right)\mathbb{P}\left(\mathcal{I}^n_j(l)=0\right) \tag{4.86}
$$

and therefore $\mathcal{I}^n_i(l)$ and $\mathcal{I}^n_j(l)$ are negatively correlated. Observe also that, uniformly over $l\leqslant Tb_n$,

$$
\mathrm{Var}\left(\mathcal{I}^n_i(l)\right)\leqslant\mathbb{P}\left(\mathcal{I}^n_i(l)=1\right)
$$

$$
\leqslant\sum_{l_1=1}^{l}\mathbb{P}\left(\text{vertex }i\text{ is first discovered at stage }l_1\right)\leqslant\frac{ld'_i}{\ell'_n(T)}. \tag{4.87}
$$

Therefore, using the negative correlation in (4.86), uniformly over $l\leqslant Tn^{\alpha+\delta}$,

$$
n^{-2\alpha}\mathrm{Var}\left(M'_n(l)\right)\leqslant\frac{l}{\ell'_n(T)n^{2\alpha}}\sum_{i\in[n]}d'^3_i=o(1). \tag{4.88}
$$

Now we can use the super-martingale inequality [147, Lemma 2.54.5] stating that for any super-martingale $(M(t))_{t\geqslant 0}$, with $M(0)=0$,

$$
\varepsilon\mathbb{P}\left(\sup_{s\leqslant t}|M(s)|>3\varepsilon\right)\leqslant 3\mathbb{E}\left[|M(t)|\right]\leqslant 3\left(|\mathbb{E}\left[M(t)\right]|+\sqrt{\mathrm{Var}\left(M(t)\right)}\right). \tag{4.89}
$$

Thus (4.83) follows using (4.85), (4.88), and (4.89).    $\square$

*Proof of Lemma 6.4.* Fix any $T > 0$ and recall that $\ell_n(T) = \ell'_n - 2Tn^{\alpha+\delta} - 1$. Denote $W(l) = \sum_{i\in[n]} w_i \mathcal{J}_i^n(l)$. Firstly, observe that

$$\mathbb{E}[W(l+1) - W(l) \mid \mathscr{F}_l]$$

$$= \sum_{i\in[n]} w_i \mathbb{E}[\mathcal{J}_i^n(l+1) \mid \mathscr{F}_l] \mathbf{1}\{\mathcal{J}_i^n(l) = 0\} \leqslant \frac{\sum_{i\in[n]} d'_i w_i}{\ell'_n(T)},$$

uniformly over $l \leqslant Tn^{\alpha+\delta}$. Therefore, $(\tilde{W}(l))_{l=1}^{Tn^{\alpha+\delta}}$ is a super-martingale, where $\tilde{W}(l) = W(l) - (\sum_{i\in[n]} d'_i w_i/\ell'_n)l$. Again, the goal is to use (4.89). Using (4.84), we can show that $|\mathbb{E}[\tilde{W}(l)]| = o(n^{\alpha+\delta})$, uniformly over $l \leqslant Tn^{\alpha+\delta}$. Also, using (4.86) and (4.87) and Assumption 4.2, $\mathrm{Var}(\tilde{W}(l)) \leqslant \sum_{i\in[n]} w_i^2 \mathrm{var}(\mathcal{J}_i^n(l)) = o(n^{2(\alpha+\delta)})$, uniformly over $l \leqslant Tn^{\alpha+\delta}$. Finally, using (4.89), we conclude the proof. □

## 4.D   Proof sketch for Proposition 4.5.12

The proof of Proposition 4.5.12 can be carried out using similar ideas as Chapter 3 Theorem 3.5.1. The key idea to prove Proposition 4.5.12 is that the scaling limit is governed by the vertices having large degrees only. More precisely, for any $\varepsilon > 0$ and $T > 0$,

$$\lim_{K\to\infty} \limsup_{n\to\infty} \mathbb{P}\left( \sup_{t\leqslant T} n^{-\alpha} \left| \sum_{i>K} a_i\left(\mathcal{J}_i^n(tn^{\rho-\delta}) - \frac{a_i}{\ell_n^a} tn^{\rho-\delta}\right) \right| > \varepsilon \right) = 0.$$

This can be proved using martingale estimates. Thus, if one considers the truncated sum

$$\sum_{i\leqslant K} a_i\left(\mathcal{J}_i^n(l) - \frac{a_i}{\ell_n^a}l\right) + (\nu_n(a) - 1)\,l,$$

with the first $K$ (fixed) terms it is enough to show that the iterated limit of the truncated process (first taking $\lim_{n\to\infty}$ and then $\lim_{K\to\infty}$) converges to $\mathbf{S}$ with respect to the Skorohod $J_1$ topology. Now, using the fact that $a_i / \sum_i a_i \xrightarrow{\mathbb{P}} \theta_i/(\mu\nu)$, and the fact that the vertices are explored in a size-biased manner with sizes being $(a_i)_{i\geqslant 1}$, it follows that (see Chapter 3 Lemma 3.5.2), for each fixed $K \geqslant 1$,

$$\left(\mathcal{J}_i^n(tn^{\rho-\delta})\right)_{i\in[K],t\geqslant 0} \xrightarrow{d} \left(\mathbf{1}\{\mathrm{Exp}(\theta_i/(\mu\nu)) \leqslant t\}\right)_{i\in[K],t\geqslant 0}.$$

This concludes the proof of Proposition 4.5.12.

# Chapter 5

# Global lower mass-bound for critical configuration models in the heavy-tailed regime

**Abstract.**   We establish the global lower mass-bound property for largest connected components in the critical window of phase transition for configuration model when the degree distribution has an infinite third moment. The scaling limit of the critical percolation clusters, viewed as measured metric spaces, was established in [31] with respect to the Gromov-weak topology. Our result extends those scaling limit results to hold under the stronger Gromov-Hausdorff-Prokhorov topology. This implies convergence of global functionals such as the diameters of the critical components. Further, our result establishes compactness of the random metric spaces, which arise as scaling limits of critical clusters in the heavy-tailed regime.

Any connected graph $\mathscr{C}$ can be viewed as a metric space with the distance between points given by $ad(\cdot,\cdot)$ for some constant $a > 0$, where $d(\cdot,\cdot)$ is used as a generic notation to denote the graph-distance (i.e., number of edges in the shortest path). Suppose that each vertex $i$ is assigned a *mass* $w_i$ so that there is a natural probability measure associated to the Borel sigma-algebra on $(\mathscr{C}, ad)$ with the measure given by $\mu(A) = \sum_{v \in A} w_v / \sum_{v \in \mathscr{C}} w_v$ for any $A \subset \mathscr{C}$. We denote the above metric space with a measure by $(\mathscr{C}, a, w)$. Fix any $\delta > 0$ and define the $\delta$-lower mass of $(\mathscr{C}, a, w)$ by

$$\mathfrak{m}(\delta) := \frac{\inf_{v \in \mathscr{C}} \sum_{u : ad(v,u) \leqslant \delta} w_u}{\sum_{u \in \mathscr{C}} w_u}. \tag{5.1}$$

For a sequence $(\mathscr{C}_n, a_n, w_n)_{n \geqslant 1}$ of graphs viewed as metric spaces endowed with a measure, the global lower mass-bound property is defined as follows:

**Definition 6** (Global lower mass-bound property [20]). For $\delta > 0$, let $\mathfrak{m}_n(\delta)$ denote the $\delta$-lower mass of $(\mathscr{C}_n, a_n, w_n)$. Then $(\mathscr{C}_n, a_n, w_n)_{n \geqslant 1}$ is said to satisfy the global lower mass-bound property if and only if $\sup_{n \geqslant 1} \mathfrak{m}_n(\delta)^{-1} < \infty$ for any $\delta > 0$. When the sequence $(\mathscr{C}_n)_{n \geqslant 1}$ is a collection of random graphs, $(\mathscr{C}_n, a_n, w_n)_{n \geqslant 1}$ is said to satisfy the global lower mass-bound property if and only if $(\mathfrak{m}_n(\delta)^{-1})_{n \geqslant 1}$ is a tight sequence of random variables for any $\delta > 0$.

The aim of this chapter is to prove the global lower mass-bound property for connected components of a configuration model at criticality, when the third moment of the empirical degree distribution tends to infinity. Informally speaking, the global lower mass-bound property ensures that all the small neighborhoods have mass bounded away from zero, so that the graph does not have any *light spots* and the total mass is well-distributed over the whole graph. This has several interesting consequences in the theory of critical random graphs, which we discuss in detail below after the formal statement of the result. We start by defining the configuration model and state the precise assumptions, followed by a formal statement of the main result. Subsequently, we discuss some implications of this result in the context of recent scaling limit results for critical percolation on a configuration model.

## 5.1    Main results

Fix $\tau \in (3,4)$. Throughout this chapter we will use the shorthand notation

$$\alpha = 1/(\tau - 1), \quad \rho = (\tau - 2)/(\tau - 1), \quad \eta = (\tau - 3)/(\tau - 1).$$

Further, we assume the following conditions on the degree sequences of $CM_n(\mathbf{d})$:

**Assumption 5.1** (Degree sequence)**.** For each $n \geqslant 1$, let $\mathbf{d} = \mathbf{d}_n = (d_1, \ldots, d_n)$ be a degree sequence satisfying $d_1 \geqslant d_2 \geqslant \ldots \geqslant d_n$. We assume the following about $(\mathbf{d}_n)_{n \geqslant 1}$ as $n \to \infty$:

  (i) (*High-degree vertices*) For each fixed $i \geqslant 1$,

$$n^{-\alpha} d_i \to \theta_i,$$

     where $\theta = (\theta_1, \theta_2, \ldots) \in \ell_\downarrow^3 \setminus \ell_\downarrow^2$.

 (ii) (*Moment assumptions*) Let $D_n$ denote the degree of a typical vertex, i.e. a vertex chosen uniformly at random, independently of $CM_n(\mathbf{d})$. Then, $D_n$ converges in distribution to some discrete random variable $D$ and

$$\frac{1}{n} \sum_{i \in [n]} d_i \to \mu := \mathbb{E}[D], \quad \frac{1}{n} \sum_{i \in [n]} d_i^2 \to \mu_2 := \mathbb{E}[D^2],$$

$$\lim_{K \to \infty} \limsup_{n \to \infty} n^{-3\alpha} \sum_{i=K+1}^{n} d_i^3 = 0.$$

(iii) For all sufficiently large $n$, the following holds uniformly over $i \in [n]$ and $x > 0$:

$$n^{-\alpha} \sum_{j : d_j > xn^\alpha} d_j \geqslant x^{-b}, \quad \left(\frac{d_i}{n^\alpha}\right)^{b-1} n^{-2\alpha} \sum_{j \leqslant i} d_j^2 > C \qquad (5.2)$$

    for some $b \in (1, 2)$. Further, $\limsup_{n \to \infty} \sum_{i \geqslant 1} e^{-n^{-2\alpha} \sum_{j=1}^{i} d_j^2} < \infty$.

(iv) Let $n_1$ be the number of degree-one vertices. Then $n_1 = \Theta(n)$, which is equivalent to assuming that $\mathbb{P}(D = 1) > 0$.

 (v) The weight sequence $w = (w_i)_{i \in [n]}$ satisfies

$$\lim_{n \to \infty} \frac{1}{\ell_n} \sum_{i \in [n]} d_i w_i = \mu_w, \quad \max \left\{ \sum_{i \in [n]} d_i w_i^2, \sum_{i \in [n]} d_i^2 w_i \right\} = O(n^{3\alpha}).$$

Assumption 5.1 (i)–(iii) are the general set of assumptions on the degree distribution under which the scaling limit for the component sizes, surplus edges and the metric structure of critical configuration model was proved in [31, 69]. These assumptions are applicable for a configuration model with

power-law degree distribution with exponent $\tau \in (3,4)$. More precisely, if $F$ is a distribution function on non-negative integers satisfying $(1 - F)(x) = Cx^{\tau-1}$, then Assumptions 5.1 (i)–(iv) is satisfied when (a) $d_i = (1 - F)^{-1}(i/n)$, (b) $d_i$'s are i.i.d. samples from $F$ [69, Section 2]. Thus, above assumptions are applicable for configuration model with power-law degree distribution with exponent $\tau \in (3,4)$. We note that Assumption 5.1 (iv) is required for technical purposes, which was not required in [31, 69]. Assumptions 5.1 (v) for the weight sequence is satisfied for $w_i = 1$ or $w_i = d_i$ for all $i \in [n]$. $w_i = 1$ is equivalent to the normalized counting measure on $\mathscr{C}$. Moreover, we assume that the configuration model lies within the critical window of the phase transition, i.e., for some $\lambda \in \mathbb{R}$,

$$\nu_n = \frac{\sum_{i\in[n]} d_i(d_i - 1)}{\sum_{i\in[n]} d_i} = 1 + \lambda n^{-\eta} + o(n^{-\eta}). \tag{5.3}$$

We denote the $i$-th largest connected component of $CM_n(\mathbf{d})$ by $\mathscr{C}_{(i)}$. For each $v \in [n]$ and $\delta > 0$, let $\mathcal{N}_v(\delta)$ denote the $\delta n^\eta$ neighborhood of $v$ in $CM_n(\mathbf{d})$. For each $i \geqslant 1$, define

$$\mathfrak{m}_i^n(\delta) = \inf_{v\in\mathscr{C}_{(i)}} n^{-\rho} \sum_{k\in\mathcal{N}_v(\delta)} w_k. \tag{5.4}$$

For $CM_n(\mathbf{d})$ satisfying Assumption 5.1 and (5.3), the total mass of components $n^{-\rho} \sum_{v\in\mathscr{C}_{(i)}} w_v$ is known to converge to some non-degenerate random variable with support $(0, \infty)$ [69, Theorem 21]. Therefore, it is enough to rescale by $n^\rho$ in (5.4) instead of the total weight of the components as given in (5.1). The following theorem is the main result of this chapter:

**Theorem 5.1.1** (Global lower mass-bound). *Suppose that Assumption 5.1 and (5.3) holds. Then, for each fixed $i \geqslant 1$, $(\mathscr{C}_{(i)}, n^{-\eta}, \mathbf{w})_{n\geqslant 1}$ satisfies global lower mass-bound, i.e., for any $\delta > 0$, the sequence $(\mathfrak{m}_i^n(\delta)^{-1})_{n\geqslant 1}$ is tight.*

By the results of [110], under Assumption 5.1, Therefore,

$$\liminf_{n\to\infty} \mathbb{P}(CM_n(\mathbf{d}) \text{ is simple}) > 0.$$

This immediately implies the following corollary:

**Corollary 5.1.2.** *Under Assumption 5.1 and (5.3), the largest components of $UM_n(\mathbf{d})$ also satisfies the global lower mass-bound property.*

Next we state another important corollary, which says that the global lower mass-bound property is also satisfied by critical percolation clusters of

$CM_n(\mathbf{d})$ and $UM_n(\mathbf{d})$. To this end, let us assume that

$$\lim_{n\to\infty} \frac{\sum_{i\in[n]} d_i(d_i - 1)}{\sum_{i\in[n]} d_i} = \nu > 1.$$

$CM_n(\mathbf{d})$ is super-critical in the sense that there exists a unique *giant* component whp for $\nu > 1$,, and when $\nu < 1$, all the components have size $o_{\mathbb{P}}(n)$ [114, 133]. Percolation refers to deleting each edge of a graph independently with probability $1 - p$. The critical window for percolation was studied in [31, 69], and is defined by the values of $p$ given by

$$p_c(\lambda) = \frac{1}{\nu_n} + \frac{\lambda}{n^\eta} + o(n^{-\eta}). \tag{5.5}$$

Let $\mathscr{C}_{(i)}(p_c(\lambda))$ denote the $i$-th largest component of the graph obtained by percolation with probability $p_c(\lambda)$ on the graph $CM_n(\mathbf{d})$. Then the following result holds:

**Corollary 5.1.3.** *Under Assumption 5.1 and (5.5), $(\mathscr{C}_{(i)}(p_c(\lambda)), n^{-\eta}, \mathbf{w})$ satisfies the global lower mass-bound property, for each fixed $i \geqslant 1$.*

### 5.1.1   Discussion

**Gap between Gromov-weak and GHP convergence.**   For formal definitions of the Gromov-weak topology, and Gromov-Hausdorff-Prokhorov (GHP) topology on the space of compact measured metric spcaes, we refer the reader to [20, 36, 91]. The Gromov-weak topology is an analogue of finite-dimensional convergence, since it takes into account distances between a finite number of sampled points from the underlying metric space. Thus, global functionals such as the diameter is not continuous with respect to this topology. Further, under the Gromov-weak convergence, the limit of compact measured metric spaces may not be compact. On the other hand, GHP convergence imposes a stronger topology which takes care of both the above points. The global lower mass (GLM) bound property acts as a bridge between these two notions of convergence. In fact, Gromov-weak convergence and GLM-bound together imply GHP-convergence when the support of the limiting measure is the entire limiting space [20, Theorem 6.1], in which case the limiting metric space is always compact. Thus, given Gromov-weak convergence, in order to derive convergence of global functionals like diameter, it is desirable to establish the GLM-bound.

**Scaling limit of critical percolation clusters.**   The scaling limit for largest critical percolation clusters $\mathscr{C}_{(i)}(p_c(\lambda))$, viewed as a measured metric space, was derived in Chapter 4 with respect to the Gromov-weak topology. Following the above discussion, Corollary 5.1.3 establishes that the convergence in Chapter 4 holds with respect to the GHP topology. This in particular establishes that the limiting metric spaces in [31, 36] are compact almost surely. Due to Assumption 5.1 (iv), some additional conditions are imposed on $\theta$. For example, the assumption is satisfied for $\theta_i \in [L_1(i)i^{-\alpha_1}, L_2(i)i^{-\alpha_2}]$, where $\alpha_1, \alpha_2 \in (1/3, 1/2)$, and $L_1, L_2$ are slowly varying functions. This is much less restrictive than assuming $\theta_i = i^{-\alpha}$ as in [36]. The compactness of the limiting metric spaces in [31, 36] was also established under some regularity conditions in a very recent preprint [56] using independent methods as in this chapter. In addition to the compactness of the limiting metric space, we also have the convergence of the diameters, i.e.,

$$\left(n^{-\eta}\mathrm{diam}(\mathscr{C}_{(i)}(p_c(\lambda)))\right)_{i\geqslant 1} \xrightarrow{d} (X_i)_{i\geqslant 1}$$

with respect to the product topology, where $(X_i)_{i\geqslant 1}$ is a non-degenerate random vector. In fact $X_i$ corresponds to the diameter of the limiting object of $\mathscr{C}_{(i)}(p_c(\lambda))$ from [31].

**Proof ideas and technical motivation for this work.**   The key idea of the proof of Theorem 5.1.1 consists of two main steps. The first step is to show that the neighborhoods of the high-degree vertices, called *hubs*, have mass $\Theta(n^\rho)$. Secondly, the probability of any small $\varepsilon n^\eta$ neighborhood not containing hubs is arbitrarily small. These two facts, summarized in Propositions 5.2.1 and 5.2.2 below, together ensure that the total mass of any neighborhood of $\mathscr{C}_{(i)}$ of radius $\varepsilon n^\eta$ is bounded away from zero. These two facts were proved in [36] in the context of inhomogeneous random graphs. However, the proof techniques are completely different here. The main advantage in [36] was that the breadth-first exploration of components could be dominated by a branching process with *mixed Poisson* progeny distribution that is *independent of* $n$. The above facts allow one to use existing literature and estimate the probabilities that a long path exists in the branching process in [36]. However, such a technique is specific to rank-one inhomogeneous random graphs and does not work in the cases where the above stochastic domination is not possible. This was partly a motivating reason for this work. Moreover, the final section contains many results about exponential bounds for the number of edges in the large components (Proposition 5.4.1),

a coupling of the neighborhood exploration with a branching process with stochastically larger progeny distribution (Section 5.4.2), which is interesting in its own right.

**Organization of this chapter.**   The rest of this chapter is organized as follows: In Section 5.2, we state two key propositions, one involving the total mass of small neighborhoods, and the second one involving a bound on the diameter. The proof of Theorem 5.1.1 is completed in Section 5.2. In Section 5.3 we derive the required bounds on the total mass of small neighborhoods. In Section 5.4 we obtain the required bounds on the diameter.

## 5.2    Proof of Theorem 5.1.1

In this section, we first state the two key propositions in Propositions 5.2.1, and 5.2.2, and then complete the proof of Theorem 5.1.1. The following shows that hub $i$ has sufficient mass close to it with high probability:

**Proposition 5.2.1.** *For each fixed $i \geqslant 1$ and $\varepsilon_2 > 0$, there exists $\delta_{i,\varepsilon_2} > 0$ and $n_{i,\varepsilon_2} \geqslant 1$ such that, for any $\delta \in (0, \delta_{i,\varepsilon_2}]$ and $n \geqslant n_{i,\varepsilon_2}$,*

$$\mathbb{P}\left( \sum_{k \in \mathcal{N}_i(\delta)} w_k \leqslant \theta_i \delta n^\rho \right) \leqslant \frac{\varepsilon_2}{2^{i+1}}. \tag{5.6}$$

Denote by $\mathcal{G}_n^{>K}$ the graph obtained by removing the vertices $1, \ldots, K$ having the largest degrees and the associated edges from $\mathrm{CM}_n(\mathbf{d})$. Note that $\mathcal{G}_n^{>K}$ is a configuration model conditional on its degree sequence. Let $\Delta^{>K}$ denote the maximum of the diameters of the connected components of $\mathcal{G}_n^{>K}$. For a component $\mathscr{C} \subset \mathrm{CM}_n(\mathbf{d})$, we write $\Delta(\mathscr{C})$ to denote its diameter. The following proposition shows that the diameter of all components of $\mathcal{G}_n^{>K}$ is small with high probability:

**Proposition 5.2.2.** *Assume that Assumption 5.1 holds. Then, for any $\varepsilon_1, \varepsilon_2 > 0$, there exists $K = K(\varepsilon_1, \varepsilon_2)$ and $n_0 = n_0(\varepsilon_1, \varepsilon_2)$ such that for all $n \geqslant n_0$,*

$$\mathbb{P}\left( \Delta^{>K} > \varepsilon_1 n^\eta \right) \leqslant \frac{\varepsilon_2}{4}. \tag{5.7}$$

*Proof of Theorem 5.1.1.* Fix any $i \geqslant 1$ and $\varepsilon_1, \varepsilon_2 > 0$. Let us choose $K$ and $n_0$ satisfying (5.7). In view of Proposition 5.2.1, let $\delta_0 = \min\{\varepsilon_1, \delta_{1,\varepsilon_2}, \ldots, \delta_{K,\varepsilon_2}\}/2$, and $n_0' = \max\{n_0, n_{1,\varepsilon_2}, \ldots, n_{K,\varepsilon_2}\}$. Thus, for all $n \geqslant n_0'$, (5.6) is satisfied for all $i \in [K]$. Define

$$F_1 := \{\Delta^{>K} < \varepsilon_1 n^\eta/2\}, \quad F_2 := \{\Delta(\mathscr{C}_{(i)}) > \varepsilon_1 n^\eta/2\}.$$

Notice that, on the event $F_1 \cap F_2$, it must be that one of the vertices $1, 2, \ldots, K$ belongs to $\mathscr{C}_{(i)}$, and the union of the neighborhoods of $[K]$ of radius $\lceil \varepsilon_1 n^\eta / 2 \rceil + 1 \approx \varepsilon_1 n^\eta / 2$ covers $\mathscr{C}_{(i)}$. Therefore, given any vertex $v \in \mathscr{C}_{(i)}$, $\mathcal{N}_v(\varepsilon_1)$ contains at least one of the neighborhoods $(\mathcal{N}_j(\varepsilon_1 / 2))_{j \in [K]}$. This observation yields that

$$\inf_{v \in \mathscr{C}_{(i)}} n^{-\rho} \sum_{k \in \mathcal{N}_v(\varepsilon_1)} w_k \geqslant \min_{j \in [K]} n^{-\rho} \sum_{k \in \mathcal{N}_j(\varepsilon_1/2)} w_k \geqslant \min_{j \in [K]} n^{-\rho} \sum_{k \in \mathcal{N}_j(\delta_0)} w_k.$$

Thus, for all $n \geqslant n_0'$,

$$\mathbb{P}\left( F_1 \cap F_2 \cap \left\{ \inf_{v \in \mathscr{C}_{(i)}} n^{-\rho} \sum_{k \in \mathcal{N}_v(\varepsilon_1)} w_k \leqslant \theta_K \delta_0 \right\} \right)$$
$$\leqslant \sum_{j \in [K]} \mathbb{P}\left( \sum_{k \in \mathcal{N}_j(\delta)} w_k \leqslant \theta_j \delta_0 n^\rho \right) \leqslant \frac{\varepsilon_2}{2}. \tag{5.8}$$

Further, on the event $F_2^c$, $\sum_{k \in \mathcal{N}_v(\varepsilon_1)} w_k = \sum_{k \in \mathscr{C}_{(i)}} w_k$ for all $v \in \mathscr{C}_{(i)}$. Moreover, using [69, Theorem 21], it follows that $n^{-\rho} \sum_{k \in \mathscr{C}_{(i)}} w_k$ converges in distribution to a random variable with strictly positive support. Using the Portmanteau theorem, the above implies that for any $\delta_0' > 0$, there exists $\tilde{n}_0 = \tilde{n}_0(\varepsilon_2, \delta_0')$ such that, for all $n \geqslant \tilde{n}_0$,

$$\mathbb{P}\left( n^{-\rho} \sum_{k \in \mathscr{C}_{(i)}} w_k \leqslant \delta_0' \right) \leqslant \frac{\varepsilon_2}{4}.$$

Therefore,

$$\mathbb{P}\left( F_2^c \cap \left\{ \inf_{v \in \mathscr{C}_{(i)}} n^{-\rho} \sum_{k \in \mathcal{N}_v(\varepsilon_1)} w_k \leqslant \delta_0' \right\} \right) \leqslant \frac{\varepsilon_2}{4}. \tag{5.9}$$

Now, using (5.8), (5.9) together with Proposition 5.2.2, it follows that, for any $n \geqslant \max\{n_0', \tilde{n}_0\}$,

$$\mathbb{P}\left( \inf_{v \in \mathscr{C}_{(i)}} n^{-\rho} \sum_{k \in \mathcal{N}_v(\varepsilon_1)} w_k \leqslant \min\{\delta_0', \theta_K \delta_0\} \right) \leqslant \varepsilon_2.$$

This completes the proof of Theorem 5.1.1. □

## 5.3    Lower bound on the total mass of neighborhoods of hubs

In this section, we prove Proposition 5.2.1.

*Proof of Proposition 5.2.1.* Let us denote the component of $CM_n(\mathbf{d})$ containing vertex $i$ by $\backslash(i)$. Consider the breadth-first exploration of $\backslash(i)$ starting from vertex $i$, given by the following algorithm:

**Algorithm 5.1** (Exploring the graph)**.** The algorithm carries along vertices that can be alive, active, exploring and killed and half-edges that can be alive, active or killed. We sequentially explore the graph as follows:

(S0)  At stage $l = 0$, all the vertices and the half-edges are *alive*, and only the half-edges associated to vertex $i$ are *active*. Also, there are no *exploring* vertices except $i$.

(S1)  At each stage $l$, if there is no active half-edge, choose a vertex $v$ proportional to its degree among the alive (not yet killed) vertices and declare all its half-edges to be *active* and declare $v$ to be *exploring*. If there is an active vertex but no exploring vertex, then declare the *smallest* vertex to be exploring.

(S2)  At each stage $l$, take an active half-edge $e$ of an exploring vertex $v$ and pair it uniformly to another alive half-edge $f$. Kill $e, f$. If $f$ is incident to a vertex $v'$ that has not been discovered before, then declare all the half-edges incident to $v'$ active, except $f$ (if any). If $degree(v') = 1$ (i.e. the only half-edge incident to $v'$ is $f$) then kill $v'$. Otherwise, declare $v'$ to be active and larger than all other vertices that are alive. After killing $e$, if $v$ does not have another active half-edge, then kill $v$ also.

(S3)  Repeat from (S1) at stage $l + 1$ if not all half-edges are already killed.

Call a vertex *discovered* if it is either active or killed. Let $\mathscr{V}_l$ denote the set of vertices discovered up to time $l$ and $\mathcal{I}_i^n(l) := \mathbf{1}\{i \in \mathscr{V}_l\}$. Define the exploration process by

$$S_n(l) = d_i + \sum_{j \neq i} d_j \mathcal{I}_j^n(l) - 2l = d_i + \sum_{j \neq i} d_j \left(\mathcal{I}_j^n(l) - \frac{d_j}{\ell_n}l\right) + \left(\frac{1}{\ell_n}\sum_{j \neq i} d_j^2 - 2\right)l.$$
(5.10)

Note that the exploration process keeps track of the number of active half-edges. Thus, $\backslash(i)$ is explored when $S_n$ hits zero. Moreover, since one edge is explored at each step, the hitting time to zero is the total number of edges in $\backslash(i)$. Define the re-scaled version $\bar{S}_n$ of $S_n$ by $\bar{S}_n(t) = n^{-\alpha}S_n(\lfloor tn^\rho \rfloor)$. Then,

by Assumption 5.1 and (5.3),

$$\bar{S}_n(t) = \theta_i - \frac{\theta_i^2 t}{\mu} + n^{-\alpha} \sum_{j \neq i} d_j \left( \mathfrak{I}_j^n(tn^\rho) - \frac{d_j}{\ell_n} tn^\rho \right) + \lambda t + o(1).$$

Using arguments similar to [69, Theorem 8], it can be shown that

$$\bar{S}_n \overset{d}{\to} S_\infty, \tag{5.11}$$

with respect to the Skorohod $J_1$ topology, where

$$S_\infty(t) = \theta_i - \frac{\theta_i^2 t}{\mu} + \sum_{j \neq i} \theta_j \left( \mathfrak{I}_j(t) - \frac{\theta_j t}{\mu} \right) + \lambda t,$$

with $\mathfrak{I}_j(s) := \mathbf{1}\{\xi_j \leqslant s\}$ and $\xi_j \sim \text{Exponential}(\theta_j/\mu)$ independently.

Let $h_n(u)$ (respectively $h_\infty(u)$) denote the first hitting time of $\bar{S}_n$ (respectively $S_\infty$) to $u$. More precisely,

$$h_n(u) := \inf\{t : \bar{S}_n(t) \leqslant u \text{ or } \bar{S}_n(t-) \leqslant u\},$$

and define $h_\infty(u)$ similarly by replacing $\bar{S}_n(t)$ by $\bar{S}_\infty(t)$ above. Note that for any $u > 0$, $h_\infty(u) < h_\infty(u-)$ implies that $\bar{S}_\infty(t)$ has a jump at $u$, which is a zero probability event. Thus, [107, Chapter VI.2, Proposition 2.11] is applicable and together with the convergence in (5.11), this yields

$$n^{-\rho} h_n(u) \overset{d}{\to} h_\infty(u)$$

for any $u > 0$. Further, the distribution of $h_\infty(u)$ do not contain any atoms. This follows using [33, Lemma 3.5]. Now an application of Portmanteau theorem yields that there exist $\beta_{\varepsilon_2,i} > 0$ and $n_{i,\varepsilon_2} \geqslant 1$ such that, for all $n \geqslant n_{i,\varepsilon_2}$,

$$\mathbb{P}(h_n(\theta_i/2) \leqslant n^\rho \beta_{\varepsilon_2,i}) \leqslant \frac{\varepsilon_2}{2^{i+1}}.$$

Firstly the goal is to show that there exists a $\delta_{i,\varepsilon}$ such that for any $\delta \in (0, \delta_{i,\varepsilon_2}]$,

$$\sum_{k \in \mathcal{N}_i(\delta)} d_k \leqslant \theta_i \delta n^\rho \quad \implies \quad h_n(\theta_i/2) \leqslant n^\rho \beta_{\varepsilon_2,i}.$$

Recall that $\mathcal{N}_v(\delta)$ denotes the $\delta n^\eta$ neighborhood of $v$ in $CM_n(\mathbf{d})$. To this end, let $\partial(j)$ denote the set of vertices at distance $j$ from $i$. Let $E_{j1}$ denote the total number of edges between vertices in $\partial(j)$ and $\partial(j-1)$, and let $E_{j2}$ denote the number of edges within $\partial(j-1)$. Define $E_j = E_{j1} + E_{j2}$. Fix any $\delta < 2\beta_{\varepsilon,i}/\theta_i$. Note that if $\sum_{k \in \mathcal{N}_i(\delta)} d_k \leqslant \theta_i \delta n^\rho$, then the total number

of edges in $\mathcal{N}_i(\delta)$ is at most $\theta_i \delta n^\rho / 2$. Thus there exists $j \leqslant \delta n^\eta$ such that $E_j \leqslant \theta_i \delta n^\rho / 2\delta n^\eta = \theta_i n^\alpha / 2$. This implies that $S_n$ must go below $\theta_i n^\alpha / 2$ before exploring all the vertices in $\mathcal{N}_i(\delta)$. This is because we are exploring the components in a breadth-first manner and $\bar{S}_n$ keeps track of the number of active half-edges which are the potential connections to vertices at the next level. Since one edge is explored in each time step, and we rescale time by $n^\rho$, this implies that

$$h_n(\theta_i/2) \leqslant \frac{1}{2} n^{-\rho} \sum_{k \in \mathcal{N}_i(\delta)} d_k \leqslant \delta\theta_i/2 \leqslant \beta_{\varepsilon,i}.$$

Therefore, for all $n \geqslant n_{i,\varepsilon}$,

$$\mathbb{P}\left( \sum_{k \in \mathcal{N}_i(\delta)} d_k \leqslant \theta_i \delta n^\rho \right) \leqslant \mathbb{P}(h_n(\theta_i/2) \leqslant \beta_{\varepsilon,i}) \leqslant \frac{\varepsilon}{2^{i+1}}. \tag{5.12}$$

Finally, to conclude Proposition 5.2.1 from (5.12), we use the following result from [69, Lemma 22]: For any $T > 0$,

$$\sup_{u \leqslant T} \left| \sum_{i \in [n]} w_i \mathcal{I}_i^n(un^\rho) - \frac{\sum_{i \in [n]} d_i w_i}{\ell_n} un^\rho \right| = o_{\mathbb{P}}(n^\rho).$$

Note that [69, Lemma 22] does not use $\sum_{i \in [n]} w_i = O(\ell_n)$ from [69, Assumption 3], and thus it is omitted in Assumption 5.1. The proof of Proposition 5.2.1 now follows.    $\square$

## 5.4   Diameter after removing hubs

Recall the definition of the graph $\mathcal{G}_n^{>K}$ from Proposition 5.2.2. If we keep on exploring $\mathcal{G}_n^{>K}$ in a breadth-first manner using Algorithm 5.1 and ignore the cycles created, we get a random tree. The idea is to couple neighborhoods of $i$ in $\mathcal{G}_n^{>K}$ with a suitable branching process such that the progeny distribution of the branching process dominates the number of children of each vertices in the breadth-first tree. Therefore, if there is a long path in $\mathcal{G}_n^{>K}$ which makes the diameter large, that long path must be present in the branching process as well under the above coupling. In this way, the question about the diameter of $\mathcal{G}_n^{>K}$ reduces to the question about the height of a branching process. To estimate the height suitably, we use a beautiful recent technique by Addario-Berry from [1] which allows one to relate the height of a branching process to the sum of inverses of the associated breadth-first random walk.

In Section 5.4.1, we establish large deviation bounds for the number of edges within components. This allows us to come up with the desired coupling in Section 5.4.2. In Section 5.4.3, we analyze the breadth-first random walk to show that the height of the branching process being larger than $\varepsilon n^\eta$ has small probability. These bounds are different from those derived in [1] since the branching process depends on $n$ and there is a joint scaling involved between the distances and the mean of the branching process.

### 5.4.1  Asymptotics for the number of edges

For a graph $G$, let $E(G)$ denote the number of edges in $G$.

**Proposition 5.4.1.** *There exists $\varepsilon_0 > 0$ such that the following holds: For all $\varepsilon \in (0, \varepsilon_0)$, there exists $\delta > 0$ such that for all sufficiently large $n$*

$$\mathbb{P}(E(\mathscr{C}(i)) > n^{\rho+\varepsilon}) \leqslant Ce^{-Cn^\delta},$$

*for some absolute constant $C > 0$ and for all $i \in [n]$.*

Consider exploring $CM_n(\mathbf{d})$ with Algorithm 5.1, and the associated exploration process defined in (5.10). Let us denote by $d_{(l)}$ the degree of the vertex found at step $l$. If no new vertex is found at step $l$, then $d_{(l)} = 0$. Also, let $\mathscr{F}_l$ denote the sigma algebra containing all the information revealed by the exploration process upto time $l$. Thus,

$$S_n(0) = d_i, \quad \text{and} \quad S_n(l) = S_n(l-1) + (d_{(l)} - 2).$$

Using the Doob-Meyer decomposition, one can write

$$S_n(l) = M_n(l) + A_n(l),$$

where $M_n$ is a martingale with respect to $(\mathscr{F}_l)_{l\geqslant 1}$. The drift $A_n$ and the quadratic variation $\langle M_n \rangle$ of $M_n$ are given by

$$A_n(l) = \sum_{j=1}^{l} \mathbb{E}\big[d_{(j)} - 2 | \mathscr{F}_{j-1}\big], \quad \langle M_n \rangle(l) = \sum_{j=1}^{l} \text{Var}\big(d_{(j)} | \mathscr{F}_{j-1}\big).$$

Fix $\varepsilon_0 = (4 - \tau)/(\tau - 1)$. We use $C$ as a generic notation for an absolute constant whose value can be different in different places. We will show that for any $\varepsilon \in (0, \varepsilon_0)$, there exists $\delta$ such that the following two lemmas hold with $t_n := n^{\rho+\varepsilon}$:

**Lemma 5.4.2.** *For all sufficiently large $n$, $\mathbb{P}(n^{-(\alpha+\varepsilon)} M_n(t_n) > 1) \leqslant Ce^{-Cn^\delta}$.*

**Lemma 5.4.3.** *For all* $K \geqslant 1$ *the following bound holds sufficiently large* $n$:

$$\mathbb{P}\left(n^{-(\alpha+\varepsilon)}A_n(t_n) \geqslant -C\sum_{i=1}^{K}\theta_i^2\right) \leqslant Ce^{-Cn^\delta}.$$

*Proof of Proposition 5.4.1 subject to Lemmas 5.4.2, 5.4.3.* Note that, we can choose $K \geqslant 1$ such that $\sum_{i=1}^{K}\theta_i^2$ is arbitrarily large as $\theta \notin \ell_\downarrow^2$. Thus, if $n^{-(\alpha+\varepsilon)}M_n(t_n) \leqslant 1$ and $n^{-(\alpha+\varepsilon)}A_n(t_n) \leqslant -\sum_{i=1}^{K}\theta_i^2$, then $n^{-(\alpha+\varepsilon)}S_n(t_n) < 0$, and therefore $\mathscr{C}(i)$ must be explored before time $t_n$. Thus, Lemmas 5.4.2 and 5.4.3 together complete the proof of Proposition 5.4.1. $\qquad\square$

*Proof of Lemma 5.4.2.* Firstly note that $\varepsilon_0 < \alpha$ and therefore $t_n = o(n)$. Thus, uniformly over $j \leqslant t_n$,

$$\mathrm{Var}\left(d_{(j)}|\mathscr{F}_{j-1}\right) \leqslant \mathbb{E}[d_{(j)}^2|\mathscr{F}_{j-1}] = \frac{\sum_{j\notin\mathscr{V}_{j-1}}d_j^3}{\ell_n - 2j + 2} \leqslant \frac{\sum_{j\in[n]}d_j^3}{\ell_n - 2t_n + 2} \leqslant Cn^{3\alpha-1},$$

so that almost surely,

$$\langle M_n\rangle(t_n) \leqslant t_n Cn^{3\alpha-1} = Cn^{2\alpha+\varepsilon}.$$

Also, $d_{(j)} \leqslant Cn^\alpha$ almost surely. We can now use Freedman's inequality [87, Proposition 2.1] to conclude that

$$\mathbb{P}(M_n(t_n) > n^{\alpha+\varepsilon}) \leqslant \exp\left(-\frac{n^{2\alpha+2\varepsilon}}{2(n^\alpha n^{\alpha+\varepsilon} + Cn^{2\alpha+\varepsilon})}\right) \leqslant Ce^{-Cn^\varepsilon},$$

and the proof follows. $\qquad\square$

*Proof of Lemma 5.4.3.* Note that

$$\mathbb{E}\left[d_{(i)} - 2|\mathscr{F}_{i-1}\right] = \frac{\sum_{j\notin\mathscr{V}_{i-1}}d_j^2}{\ell_n - 2i + 1} - 2$$

$$= \frac{1}{\ell_n}\sum_{j\in[n]}d_j(d_j - 2) - \frac{1}{\ell_n}\sum_{j\in\mathscr{V}_{i-1}}d_j^2 + \frac{(2i-1)\sum_{j\notin\mathscr{V}_{i-1}}d_j^2}{\ell_n(\ell_n - 2i + 1)}$$

$$\leqslant \lambda n^{-\eta} - \frac{1}{\ell_n}\sum_{j\in\mathscr{V}_{i-1}}d_j^2 + \frac{(2i-1)}{(\ell_n - 2i + 1)^2}\sum_{j\in[n]}d_j^2 + o(n^{-\eta})$$

uniformly over $i \leqslant t_n$. Therefore, for all sufficiently large $n$,

$$A_n(t_n) \leqslant \lambda n^{\alpha+\varepsilon} - \frac{1}{\ell_n}\sum_{i=1}^{t_n}\sum_{j\in\mathscr{V}_{i-1}}d_j^2 + \frac{Ct_n^2}{\ell_n} + o(n^{\alpha+\varepsilon})$$

$$= \lambda n^{\alpha+\varepsilon} - \frac{1}{\ell_n}\sum_{i=1}^{t_n}\sum_{j\in\mathscr{V}_{i-1}}d_j^2 + o(n^{\alpha+\varepsilon}),$$

(5.13)

where in the last step we have used the fact that $\varepsilon < (4-\tau)/(\tau-1)$. Let us denote the second term above by (A). To analyze (A), define the event

$$\mathcal{A}_n := \{\exists j : d_j > n^{\alpha-\varepsilon/2}, j \notin \mathscr{V}_{t_n/2}\}.$$

Thus, for all sufficiently large $n$,

$$\mathbb{P}(\mathcal{A}_n) \leqslant \sum_{j:d_j > n^{\alpha-\varepsilon/2}} \left(1 - \frac{d_i}{\ell_n - 2t_n}\right)^{t_n} \leqslant n e^{-n^{\varepsilon/2}}. \tag{5.14}$$

On the event $\mathcal{A}_n^c$,

$$\begin{aligned}
(A) &= \frac{1}{\ell_n} \sum_{i=1}^{t_n} \sum_{j \in [n]} d_j^2 \mathbf{1}\{j \in \mathscr{V}_{i-1}\} \\
&\geqslant \frac{1}{\ell_n} \sum_{i=\frac{t_n}{2}+1}^{t_n} \sum_{j=1}^{K} d_j^2 \mathbf{1}\{j \in \mathscr{V}_{i-1}\} \geqslant C n^{\alpha+\varepsilon} \sum_{j=1}^{K} \theta_j^2.
\end{aligned} \tag{5.15}$$

Combining (5.13), (5.14) and (5.15) now completes the proof. $\qquad\square$

## 5.4.2   Coupling with Branching processes

Define the event $\mathcal{K}_n := \{E(\mathscr{C}_{(i)}) > n^{\rho+\varepsilon}\}$. On the event $\mathcal{K}_n^c$, we can couple the breath-first exploration starting from vertex $i$ with a suitable branching process. Consider the branching process $\mathcal{X}_n(i)$ starting with $d_i$ individuals, and the progeny distribution $\bar{\xi}_n$ given by

$$\mathbb{P}\left(\bar{\xi}_n = k\right) = \bar{p}_k = \begin{cases} \frac{(k+1)n_{k+1}}{\underline{\ell}_n} & \text{for } k \geqslant 1, \\ \frac{n_1 - 2n^{\rho+\varepsilon}}{\underline{\ell}_n} & \text{for } k = 0, \end{cases} \tag{5.16}$$

where $\underline{\ell}_n = \ell_n - 2n^{\rho+\varepsilon}$. Note that, at each step of the exploration, we have at most $(k+1)n_{k+1}$ half-edges that are incident to vertices having $k$ further unpaired half-edges. Further, on the event $\mathcal{K}_n^c$, we have at least $\underline{\ell}_n$ choices for pairing. Therefore, the number of active half-edges discovered at each step in the breadth-first exploration of the neighborhoods of $i$ is stochastically dominated by $\bar{\xi}_n$. This proves the next proposition, which we state after setting up some notation. Recall that $\mathcal{G}_n^{>i-1}$ denotes the graph obtained by deleting vaertices $[i-1]$ and the associated edges from $CM_n(\mathbf{d})$. Let $\partial_i(r)$ denote the number of vertices at distance $r$ from $i$ in the graph $\mathcal{G}_n^{>i-1}$. Let $\bar{\xi}_n(i)$ denote the random variable with the distribution in (5.16) truncated in

such a way that $\{d_1, \ldots, d_{i-1}\}$ are excluded from the support. More precisely,

$$\mathbb{P}(\bar{\xi}_n(i) = k) = \begin{cases} \frac{(k+1)n_{k+1}}{L} & \text{for } 1 \leqslant k \leqslant d_i, \\ \frac{n_1 - 2n^{\rho + \varepsilon}}{L} & \text{for } k = 0, \end{cases}$$

where $L$ is the normalizing constant. Let $\mathcal{X}_{n,res}(i)$ denote the branching process starting with $d_i$ individuals and progeny distribution $\bar{\xi}_n(i)$ and let $\bar{\partial}_i(r)$ denotes the number of individuals at generation $r$ of $\mathcal{X}_n(i)$. Then the above stochastic domination argument immediately yields the next proposition:

**Proposition 5.4.4.** *For all $r \geqslant 1$ and $i \in [n]$ and $n \geqslant 1$:*

$$\mathbb{P}(\partial_i(r) \neq \varnothing) \leqslant \mathbb{P}(\bar{\partial}_i(r) \neq \varnothing) + \mathbb{P}(E(\mathscr{C}(i)) > n^{\rho + \varepsilon}).$$

Before going into the next section, we note that, by Assumption 5.1,

$$\bar{\nu}_n(i) = \mathbb{E}\left[\bar{\xi}^n(i)\right] = \frac{1}{\ell_n} \sum_{j \geqslant i} d_j(d_j - 1) = \frac{1}{\ell_n} \sum_{j \geqslant i} d_j(d_j - 1) + O(n^{-\alpha + \epsilon})$$

$$\leqslant 1 - \left(Cn^{-2\alpha} \sum_{j \leqslant i} d_j^2\right) n^{-\eta} + o(n^{-\eta}).$$

Thus for $i$ large and $n \geqslant n_0$,

$$\mathbb{E}\left[\bar{\xi}^n(i)\right] \leqslant 1 - \beta_i n^{-\eta}, \quad \text{where } \beta_i = C \sum_{j \leqslant i} \theta_j^2. \tag{5.17}$$

This fact will be crucially used in the next section.

### 5.4.3   Estimating heights of trees via random walks

Consider a branching process $\mathcal{X}_{n,res}(i)$ starting with $d_i$ individuals, and progeny distribution $\bar{\xi}_n(i)$ given by (5.16). Thus the progeny distribution satisfies

$$\mathbb{E}[\bar{\xi}_n(i)] \leqslant 1 - \beta_i n^{-\eta}, \quad \text{Var}\left(\bar{\xi}_n(i)\right) \leqslant Cn^{3\alpha - 1}, \tag{5.18}$$

where the choices of $\beta_i$'s are given by (5.17). We will prove the following theorem in this section:

**Theorem 5.4.5.** *Fix any $\varepsilon > 0$ and let $r_0 = \varepsilon n^{\eta}/2$. Then for all $i \in [n]$*

$$\mathbb{P}(\bar{\partial}_i(r_0) \neq \varnothing) \leqslant C2^{-\beta_i/C},$$

*for some large constant $C > 0$.*

The estimate in Theorem 5.4.5 is interesting in its own right and do not follow from previous asymptotic results in [1, 126]. This is due to the dependence of the branching process and the height on $n$. In the proof of Theorem 5.4.5, we leverage the high-level ideas from [1]. Define the breadth-first random walk by

$$s_n(u) = s_n(u-1) + \zeta_u - 1, \quad s_n(0) = d_i, \tag{5.19}$$

where $(\zeta_u)_{u \geqslant 0}$ are i.i.d. observations from the distribution of $\bar{\xi}_n(i)$. Define the function

$$H_n(t) = \sum_{u \in [0,t)} \frac{1}{s_n(u)},$$

and $\sigma = \inf\{u : s_n(u) = 0\}$. It was shown in [1, Proposition 1.7] that the height of a branching process is at most $3H_n(\sigma)$. Thus Theorem 5.4.5 can be concluded from the following estimate:

**Proposition 5.4.6.** *For any $\varepsilon > 0$ and $i \in [n]$,*

$$\mathbb{P}(H_n(\sigma) \geqslant \varepsilon n^\eta) \leqslant C2^{-\varepsilon \beta_i/C},$$

*for some large constant $C > 0$.*

Denote $I_l := [2^{l-1}d_i, 2^{l+2}d_i)$ for $l > 0$, and $I_l := [d_i 2^{l-2}, d_i 2^{l+1})$ for $l < 0$. Note that $I_l$'s are not disjoint intervals. We decompose the possible values of the random walk (5.19) into different scales. At each time $t$, the scale of $s_n(t)$, denoted by $\mathrm{scl}(s_n(t))$, is an integer. Suppose that $\mathrm{scl}(s_n(u)) = l$ for some $u > 0$. A change of scale occurs when $s_n$ leaves $I_l$. That is, at time $T := \inf\{t > u : s_n(t) \notin I_l\}$, a change of scale occurs, and the new scale is given by $\mathrm{scl}(s_n(T)) = l'$, where $l' \in \mathbb{Z}$ is such that $s_n(T) \in (2^{l'-1}d_i, 2^{l'}d_i]$. Now, the next change of scale occurs at time $T' := \inf\{t > T : s_n(t) \notin I_{l'}\}$, and the scale remains the same until $T'$, i.e., $\mathrm{scl}(s_n(t)) = l'$ for all $T \leqslant t < T'$. Define

$$H_{nl}(t) := \sum_{u \in [0,t),\ \mathrm{scl}(s_n(u))=l} \frac{1}{s_n(u)}, \quad \text{so that} \quad H_n(t) = \sum_{l \in \mathbb{Z}} H_{nl}(t).$$

Denote $T_{nl}(t) := \#\{u \in [0,t) : \mathrm{scl}(s_n(u)) = l\}$, and note that for $l > 0$

$$2^{l-1}d_i H_{nl}(t) \leqslant T_{nl}(t) \leqslant 2^{l+2}d_i H_{nl}(t),$$

and a similar inequality holds for $l < 0$. Therefore, for any $x > 0$ and $l > 0$,

$$\mathbb{P}\left(H_{nl}(\sigma) \geqslant \frac{x}{2^{l-1}d_i}\right) \leqslant \mathbb{P}(T_{nl}(\sigma) \geqslant x), \tag{5.20}$$

and a similar inequality holds for $l < 0$. Thus the proof of Proposition 5.4.6 follows from a careful estimate of the final term in (5.20), which is given by the next lemma. Let $s'_n$ be a random walk given by the same recursion relation as (5.19), except only that $s'_n(0) \in I_l$. Let $\sigma_{nl} := \min\{t \geqslant 1 : s'_n(t) \notin I_l\}$ and $r_{nl} := \min\{t \geqslant 1 : \sup_{x \in I_l} \mathbb{P}_x(\sigma_{nl} > t) \leqslant 1/2\}$.

**Lemma 5.4.7.** *For all $n \geqslant 1$, and $l \in \mathbb{Z}$:*

$$\mathbb{P}(T_{nl}(\sigma) \geqslant ar_{nl}) \leqslant C \min\{1, 2^{-l}\} 2^{-a/C},$$

*for some large constant $C > 0$.*

*Proof.* Firstly, note that $T_{nl}(\sigma) \neq 0$ if and only if $\mathrm{scl}(s_n(u)) = l$ for some $u < \sigma$. The number of upcrossings of an interval $[a, b]$ by $s_n$ is defined to be the supremum of the integers $k$ such that there exists times $(u_j, t_j)_{j=1}^k$ satisfying $u_1 < t_1 < u_2 < \cdots < t_k$, and $s_n(u_j) < a < b < s_n(t_j)$ for all $j \in [k]$. Now, for any $l \geqslant 2$, if $\mathrm{scl}(s_n(u)) = l$ occurs, then $s_n$ must have made an upcrossing of the intervals $((2^{j-1}d_i, 2^j d_i])_{1 \leqslant j \leqslant l}$. Using [1, Lemma 3.1], it follows that there exists a constant $C > 0$ such that for any $l \geqslant 2$,

$$\mathbb{P}(T_{nl}(\sigma) \neq 0) \leqslant C2^{-l}.$$

Moreover, we bound $\mathbb{P}(T_{nl}(\sigma) \neq 0)$ by 1 for $l \leqslant 1$. Next define $\mathrm{visit}(l, t)$ to be the number of visits to scale $l$, i.e., this is the supremum over $k \in \mathbb{N}$ such that one can find $(u_j, t_j)_{j=1}^k$ with $u_1 < t_1 < \cdots < u_k < t_k$ satisfying $\mathrm{scl}(s_n(u_j)) \neq l$ but $\mathrm{scl}(s_n(t_j)) = l$. Set $\mathrm{visit}(l, 0) = 1$ and $\mathrm{visit}(l, t) = 0$ if $\mathrm{scl}(s_n(t)) \neq l$. Further, define $M_{nl} = \mathrm{visit}(l, \sigma)$ (total number of visits to scale $l$) and $t_{jl} = \#\{t < \sigma : \mathrm{scl}(s_n(t)) = l, \mathrm{visit}(l, t) = j\}$ (the time spent at scale $l$ during the $j$-th visit). Thus $T_{nl}(\sigma) = \sum_{j=1}^{M_{nl}} t_{jl}$, and for $m \geqslant 2$,

$$\mathbb{P}\left(\sum_{j=1}^{M_{nl}} t_{jl} > ar_{nl}\right) \leqslant \mathbb{P}(M_{nl} > m) + \mathbb{P}\left(\sum_{j=1}^{m} t_{jl} > ar_{nl}\right).$$

Now $s_n$ can enter scale $l$ from below, which yields an upcrossing of the interval $[2^{l-1}d_i, 2^l d_i)$. Otherwise, $s_n$ can enter scale $l$ from above, whence it must be the case that while leaving the scale $l$ during the previous visit, the walk went from scale $l$ to $l+1$. The latter case yields an upcrossing of $[2^l d_i, 2^{l+1}d_i)$. Therefore, if $U_n(t, [a, b))$ denotes the number of upcrossings

of $[a, b)$ by $s_n$ before time $t$, then

$$\mathbb{P}(M_{nl} > m)$$
$$\leqslant \mathbb{P}\Big(U_n(\sigma, [2^{l-1}d_i, 2^l d_i)) \geqslant (m+1)/2\Big)$$
$$+ \mathbb{P}\Big(U_n(\sigma, [2^l d_i, 2^{l+1} d_i) \geqslant (m+1)/2\Big)$$
$$\leqslant \frac{1}{2^{(m-1)/2}}.$$

On the other hand, after each time $r_{nl}$, the probability of exiting from scale $l$ is at most $1/2$, by definition. Now, $\mathbb{P}(t_{jl} > kr_{nl}) \leqslant 2^{-k}$, which implies that $\lfloor t_{jl}/r_{nl} \rfloor$ can be stochastically dominated by Geometric$(1/2)$ random variable. Thus, if $(g_i)_{i \geqslant 1}$ denotes an i.i.d. collection of Geometric$(1/2)$ random variables,

$$\mathbb{P}\Big(\sum_{j=1}^m t_{jl} \geqslant (k+m)r_{nl}\Big) \leqslant \mathbb{P}\Big(\sum_{j=1}^m \Big\lfloor \frac{t_{jl}}{r_{nl}} \Big\rfloor \geqslant k\Big) \leqslant \mathbb{P}\Big(\sum_{i=1}^m g_i > k\Big)$$
$$= \mathbb{P}(\mathrm{Bin}(k, 1/2) < m) \leqslant e^{-(k-2m)^2/2k},$$

where the last step follows using standard concentration inequalities such as [115, Theorem 2.1]. Therefore, the proof follows by taking $k = m = a/2$.    □

For a sequence $a = (a_l)_{l \in \mathbb{Z}}$, define $V_n(a) = d_i^{-1} \sum_{l \in \mathbb{Z}} a_l r_{nl}/2^l$ and $\delta(a) = C \sum_{l \in \mathbb{Z}} \min\{1, 2^{-l}\}2^{-Ca_l}$. Using Lemma 5.4.7 and (5.20), we can now conclude that

$$\mathbb{P}(H_n(\sigma) > V_n(a)) \leqslant \delta(a). \tag{5.21}$$

To apply the above bound, we need a good estimate on $r_{nl}$. To apply the above bound, we need a good estimate on $r_{nl}$. Let $\mathbb{P}_x$ denote the law of the random walk $s_n'$, with $s_n'(0) = x$, but satisfying identical recurrence recurrence relation as (5.19). Suppose that $\sigma_{nl} := \min\{t \geqslant 1 : s_n(t) \notin I_l, s_n(0) \in I_l\}$, and $r_{nl} := \min\{t \geqslant 1 : \sup_{x \in I_l} \mathbb{P}_x(\sigma_{nl} > t) \leqslant 1/2\}$. The next two lemmas allow us to deduce such a result for $l > 0$ and $l < 0$ respectively:

**Lemma 5.4.8.** *Fix any $l \geqslant 0$ and let $\sigma_{nl} := \inf\{t : s_n(t) \notin I_l\}$. Then, for all $i \in [n]$,*

$$\limsup_{n \to \infty} \sup_{x \in I_l} \mathbb{P}_x\Big(\sigma_{nl} > Cn^\eta \frac{d_i 2^{l/2}}{\beta_i^{1/(b-1)}}\Big) \leqslant \frac{1}{2}.$$

*Proof.* Fix any $x \in I_l$. Note that for any $t > 0$,

$$\mathbb{P}_x(\sigma_{nl} > tn^\rho)$$
$$\leqslant \mathbb{P}_x(S_n(tn^\rho) \in I_l) \leqslant \mathbb{P}_x(S_n(tn^\rho) > 2^{l-1}d_i)$$
$$= \mathbb{P}_x(S_n(tn^\rho) + tn^\rho n^{-\eta}\beta_i > 2^{l-1}d_i + tn^\alpha\beta_i) \qquad (5.22)$$
$$\leqslant \frac{tn^\rho c_0 n^{3\alpha-1}}{(2^l d_i + tn^\alpha\beta_i)^2} \leqslant \frac{c_0 tn^{2\alpha}}{2^{2l}d_i^2},$$

where the last step follows from Chebyshev's inequality and the estimates in (5.18). For $l \geqslant 0$, by setting $t = \frac{d_i 2^{l/2}}{n^\alpha(\beta_i)^{1/(b-1)}}$ where b is given by (5.2), (5.22) reduces to

$$\mathbb{P}(\sigma_{nl} > tn^\rho) \leqslant \frac{C}{2^l(d_i/n^\alpha)\beta_i^{1/(b-1)}} \leqslant \frac{C}{2^l}, \qquad (5.23)$$

where we have used Assumption 5.1 in the last step. This completes the proof of Lemma 5.4.8 using (5.23). $\qquad \square$

**Lemma 5.4.9.** *For $l > 0$, and $i \in [n]$,*

$$\limsup_{n \to \infty} \sup_{x \in (d_i/2^{l+2}, d_i/2^{l-1})} \mathbb{P}_x\left(\sigma_{nl} > Cn^\eta \frac{d_i}{\beta_i 2^{bl}}\right) \leqslant \frac{1}{2}.$$

*Proof.* Recall from the definition of $\xi_n(i)$ from (5.18), and let $(\xi_{nj})_{j \geqslant 1}$ be an iid collection with the same distribution as $\xi_n(i)$. Note that, for $x \in I_l$,

$$\mathbb{P}_x(\sigma_{nl} > tn^\rho) \leqslant \mathbb{P}_x(S_n(tn^\rho) \in I_l) \leqslant \mathbb{P}\left(\xi_{nj} \leqslant \frac{d_i}{2^{l-1}} \, \forall j \leqslant tn^\rho\right)$$
$$\leqslant \left(1 - \frac{\sum_{j:d_j > d_i/2^{l-1}} d_j}{\ell_n(1 + o(1))}\right)^{tn^\rho} \leqslant \exp\left(-tn^{-\alpha} \sum_{j:d_j > d_i/2^{l-1}} d_j\right).$$

Putting $t = \frac{d_i}{n^\alpha \beta_i 2^{bl}}$ and using Assumption 5.1, it follows that the right-hand side above is at most

$$\exp\left(-\frac{C}{(d_i/n^\alpha)^{b-1}n^{-2\alpha}\sum_{j \leqslant i} d_j^2}\right),$$

and the proof follows. $\qquad \square$

*Proof of Proposition 5.4.6.* We will use the estimates of the terms appearing in (5.21). Firstly, note from Lemmas 5.4.8, 5.4.9 that, for all sufficiently large n,

$$r_{nl} \leqslant \begin{cases} \frac{d_i 2^{l/2}}{\beta_i^{1/(b-1)}} & \text{for } l > 0, \\ \frac{d_i}{2^{bl}\beta_i} & \text{for } l < 0. \end{cases}$$

Moreover, take $b_l = \varepsilon 2^{l/4}$, for all $l \geqslant 0$ and $b_l = \varepsilon 2^{-(b-1)l/2}$ for $l < 0$. In the above case, there exists an absolute constant $C > 0$ such that

$$V_n(b) \leqslant \frac{C\varepsilon n^\eta}{\beta_i},$$

where the constant only depends on the parameter $\tau$. Moreover, $\delta(b) \leqslant C2^{-\varepsilon}$. Therefore, (5.21) yields that

$$\mathbb{P}\left(H_n > \frac{C\varepsilon n^\eta}{\beta_i}\right) \leqslant C2^{-\varepsilon}.$$

Taking $\varepsilon' = C\varepsilon/\beta_i$, the proof of Proposition 5.4.6 follows. $\qquad\square$

### 5.4.4    Proof of Proposition 5.2.2

Let us now complete the proof of Proposition 5.2.2 using Propositions 5.4.4, and Theorem 5.4.5. Define $\mathscr{C}_{\mathrm{res}}(i)$ to be the connected component containing vertex $i$ in the graph $\mathcal{G}_n^{>i-1} = CM_n(\mathbf{d}) \setminus [i-1]$. Note that if $\Delta^{>K} > \varepsilon_1 n^\eta$, then there exists a path in $CM_n(\mathbf{d})$ avoiding all the vertices in $[K]$. Suppose that the minimum index among vertices on that path is $i_0$. Then $\Delta(\mathscr{C}_{\mathrm{res}}(i_0)) > \varepsilon n^\eta$. Therefore, $\Delta^{>K} > \varepsilon_1 n^\eta$ implies that there exists an $i > K$ satisfying $\Delta(\mathscr{C}_{\mathrm{res}}(i)) > \varepsilon n^\eta$. Let $\partial_i(r)$ denotes the number of vertices at distance $r$ starting from vertex $i$ in the graph $\mathcal{G}_n^{>i-1}$. Recall the definition of $\bar{\partial}$ in Proposition 5.4.4. Thus,

$$\mathbb{P}\left(\Delta^{>K} > \varepsilon_1 n^\eta\right) \leqslant \sum_{i>K} \mathbb{P}(\partial_i(\varepsilon_1 n^\eta/2) \neq \varnothing) \leqslant C\delta \sum_{i>K} e^{-\beta_i}$$

which tends to zero if we first take $n \to \infty$ and then take $K \to \infty$ using Assumption 5.1. Thus the proof follows. $\qquad\square$

## 5.5    Conclusion

We prove a global lower mass-bound property for the largest components of the critical configuration model when the third moment of the degree distribution diverges to infinity. Together with the results in Chapter 4, this proves that the scaling limits in Chapter 4 hold with respect to the Gromov-Hausdorff-Prokhorov topology, and the limiting metric space in Chapter 4 is almost surely compact under some regularity conditions. Also, this implies that the diameter of these components converge to some non-degenerate random variable after rescaling by $n^{(\tau-3)/(\tau-1)}$. The main proof technique involves an exponential bound on the probability that the height of a sequence of branching processes is large, which may be of independent interest.

# Chapter 6

# Critical percolation on scale-free random graphs: Effect of the single-edge constraint

**Abstract.** In this chapter, we study the percolation critical behavior for random graphs with degree distributions having a power-law with exponent $\tau \in (2,3)$. In this regime, the critical behavior is observed when the percolation probability tends to zero with the network size. We identify the critical window for the configuration model, the erased configuration model and the generalized random graph. The critical window turns out to be different for the multigraph version of the configuration model, a feature that is not observed for $\tau > 3$. We provide exact asymptotics of the rescaled component sizes, and describe many structural properties of these critical components. We also analyze the so-called barely sub/supercritical regimes, which establishes the relevance of the critical window identified in this chapter.

All the results and the relevant literature discussed in the previous chapters assume a finite second-moment condition on the degree distribution, and thus do not include the $\tau \in (2, 3)$ case, where the asymptotic degree distribution has infinite second moment but finite first moment. These networks are popularly known as *scale-free networks* [21] in the literature. One of the well-known features of scale-free networks is that these networks are *robust* under random edge-deletion, i.e., for any sequence $(p_n)_{n \geqslant 1}$ with $\liminf_{n \to \infty} p_n > 0$, the graph obtained by applying percolation with probability $p_n$ is always supercritical. This feature has been studied experimentally in [7], using heuristic arguments in [58, 63, 64, 74] (see also [51, 52, 95] in the context of optimal paths in the strong disorder regime), and mathematically in [45]. Thus, in order to observe the percolation critical behavior, one needs to take $p_c \to 0$ with the network size, even if the average degree of the network is finite. However, obtaining the right scaling exponents for the critical behavior was an open question in the mathematical literature.

In this chapter, we discuss the first mathematically rigorous results in the $\tau \in (2, 3)$ regime for the critical behavior of component sizes and their complexity. As canonical random graph models on which percolation acts, we take the multigraph generated by the configuration model, and the closely associated erased configuration model, obtained by deleting self-loops and multiple-edges in the configuration model. The latter model is often referred to in the literature as the configuration model with single-edge constraint. The most striking observation of this chapter is that the critical value changes depending on the *single-edge constraint*, a feature that has never surfaced in the finite second-moment setting. For the configuration model multigraph, the critical value turns out to be $p_c \sim n^{-(3-\tau)/(\tau-1)}$, whereas under the single-edge constraint $p_c \sim n^{-(3-\tau)/2}$, which is much larger $n^{-(3-\tau)/(\tau-1)}$. The largest component sizes in both the regimes are of the order $n^\alpha p_c$, and the scaling limits are in a completely different universality class than the $\tau \in (3, 4)$ and $\tau > 4$ case. We also study percolation on the generalized random graph, which gives uniformly chosen graph conditional on degrees. The contributions of this chapter can be summarized as follows:

1. For the configuration model multigraph, we obtain scaling limits for the largest component sizes and surplus edges under a strong topology. Further, the diameter of the largest components is shown to be tight random variables. To establish that the scaling limits correspond to the critical behavior, we further look at the near-critical behavior and derive the asymptotics for the component sizes in the so-called barely

sub/supercritical regimes.

2. Under the single-edge constraint, we identify the scaling limit of the largest component sizes in the part of the critical window, where the criticality parameter is sufficiently small.

3. This is the first work on critical percolation on random graphs in the $\tau \in (2,3)$ setting, thus the techniques are novel. The primary difficulty in this setting is that the exploration process approach does not work. For the configuration model, this difficulty is circumvented by sandwiching the percolated graphs by two configuration models, which yield the same scaling limits for the component sizes. The main novelty in the proof of the configuration model is the analysis of the limiting exploration process.

4. On the other hand, in the single-edge constraint scenario, the proofs require a more detailed understanding of the structure of the critical components. It turns out that the hubs (vertices of high-degree) do not connect to each other directly, but there are some special vertices that interconnect hubs. This interconnected structure forms the *core* of the critical components, and the 1-neighborhood of the core spans the critical components. We primarily use path counting techniques here since the exploration process approach does not work anymore. For path counting, we compare the connection probabilities between the hubs with the connection probabilities in a preferential attachment model, which is interesting in its own right.

## 6.1   Main results

### 6.1.1   The configuration model

**Notions of convergence and the limiting objects**

Recall the notations from Chapter 1.2.1. Consider a decreasing sequence $\theta = (\theta_1, \theta_2, \dots) \in \ell_\downarrow^2 \setminus \ell_\downarrow^1$. Denote by $\mathcal{I}_i(s) := \mathbf{1}\{\xi_i \leqslant s\}$ where $\xi_i \sim \mathrm{Exp}(\theta_i/\mu)$ independently, and $\mathrm{Exp}(r)$ denotes the exponential distribution with rate $r$. Consider the process

$$S_\infty^\lambda(t) = \lambda \sum_{i=1}^\infty \theta_i \mathcal{I}_i(t) - 2t, \tag{6.1}$$

for some $\lambda \in \mathbb{R}, \mu > 0$ and define the reflected version of $S_\infty^\lambda(t)$ by

$$\text{refl}(S_\infty^\lambda(t)) = S_\infty^\lambda(t) - \min_{0 \leqslant u \leqslant t} S_\infty^\lambda(u).$$

For any function $f \in \mathbb{D}[0, \infty)$, define $\underline{f}(x) = \inf_{y \leqslant x} f(y)$. $\mathbb{D}_+[0, \infty)$ is the subset of $\mathbb{D}[0, \infty)$ consisting of functions with positive jumps only. Note that $\underline{f}$ is continuous when $f \in \mathbb{D}_+[0, \infty)$. An *excursion* of a function $f \in \mathbb{D}_+[0, T]$ is an interval $(l, r)$ such that

$$\min\{f(l-), f(l)\} = \underline{f}(l) = \underline{f}(r) = \min\{f(r-), f(r)\}$$
$$\text{and} \quad f(x) > \underline{f}(r), \ \forall x \in (l, r) \subset [0, T].$$

Excursions of a function $f \in \mathbb{D}_+[0, \infty)$ are defined similarly. We will show that, for any $\lambda > 0$, the excursions of the process $\mathbf{S}_\infty^\lambda$ can be ordered almost surely as an element of $\ell_\downarrow^2$. We denote this ordered vector by $(\gamma_i(\lambda))_{i \geqslant 1}$.

Also, define the counting process $\mathbf{N}^\lambda$ to be the Poisson process that has intensity $\text{refl}(S_\infty^\lambda(t))$ at time $t$ conditional on $(S_\infty^\lambda(u))_{u \leqslant t}$. Formally, $\mathbf{N}^\lambda$ is characterized as the counting process for which

$$N(t) - \frac{\sum_{i=1}^\infty \theta_i^2}{\mu^2} \int_0^t \text{refl}(S_\infty^\lambda(u)) du \tag{6.2}$$

is a martingale. We use the notation $N(\gamma)$ to denote the number of marks in the interval $\gamma$. Let $\mathbf{Z}(\lambda)$ denote the vector $((\gamma_i(\lambda), N(\gamma_i(\lambda))))_{i \geqslant 1}$, ordered as an element of $\mathbb{U}_\downarrow^0$.

**Results for the critical window**

Fix any $\tau \in (2, 3)$. Throughout this chapter, we denote

$$\alpha = 1/(\tau - 1), \qquad \rho = (\tau - 2)/(\tau - 1), \qquad \eta = (3 - \tau)/(\tau - 1),$$

and assume the following about the degree sequences $(\mathbf{d}_n)_{n \geqslant 1}$:

**Assumption 6.1.**    (i) (*High-degree vertices*) For any $i \geqslant 1$, $n^{-\alpha} d_i \to \theta_i$, where the vector $\theta := (\theta_1, \theta_2, \dots) \in \ell_\downarrow^2 \setminus \ell_\downarrow^1$.

(ii) (*Moment assumptions*) $\lim_{n \to \infty} \frac{1}{n} \sum_{i \in [n]} d_i = \mu$, and

$$\lim_{K \to \infty} \limsup_{n \to \infty} n^{-2\alpha} \sum_{i=K+1}^n d_i^2 = 0.$$

In Section 6.1.3, we discuss that Assumption 6.1 is satisfied for power-law degrees with exponent $\tau \in (2,3)$. For $CM_n(\mathbf{d})$, the *criticality parameter* $\nu_n$ is defined as

$$\nu_n = \frac{\sum_{i \in [n]} d_i(d_i - 1)}{\sum_{i \in [n]} d_i}.$$

Molloy and Reed [133], and Janson and Luczak [114] showed that, under some regularity conditions, $CM_n(\mathbf{d})$ has a unique giant component (a component of size $\Theta(n)$) with high probability precisely when $\nu_n \to \nu > 1$. Under Assumption 6.1, $\nu_n \to \infty$, as $n \to \infty$ and $CM_n(\mathbf{d})$ always contains a giant component.

Let $CM_n(\mathbf{d}, p)$ denote the graph obtained from percolation with probability $p$ on the graphs $CM_n(\mathbf{d})$. Now, under Assumption 6.1 for any $p \in (0,1]$, $CM_n(\mathbf{d}, p)$ retains a giant component with high probability, i.e. $CM_n(\mathbf{d}, p)$ is always supercritical; see the remark below [99, Theorem 4.5]. Thus, in order to observe the critical behavior, one must take $p \to 0$, as $n \to \infty$. However, it is is not clear here how to define the critical window of phase-transition. One way to do it is to say that inside the critical window, the order of the sizes of largest connected components are same, and the rescaled vector of ordered component sizes converge to some *non-degenerate* random vector. This property has been observed universally for the critical window of phase transition; see [69, 70] and the references therein. In this chapter, we define the critical window to be those values of $p$ for which the re-scaled vector of component sizes converge to some non-degenerate random vector. It is worthwhile mentioning that there is a substantial literature on how to define the critical value, and the phase transition. See [48, 96, 99, 118, 134] for different definitions of critical probability and related discussions.

We will show that the critical window for percolation on $CM_n(\mathbf{d})$ is given by

$$p_c = p_c(\lambda) := \frac{\lambda}{\nu_n}(1 + o(1)), \quad \lambda \in (0, \infty). \tag{6.3}$$

Notice that, under Assumption 6.1, the parameter $\nu_n$ is of the order $n^{2\alpha - 1} = n^\eta$, where $\eta = (3 - \tau)/(\tau - 1) > 0$.

To avoid complicated notation, we will always write $\mathscr{C}_{(i)}(p)$ to denote the $i$-th largest component in the percolated graph. The random graph on which percolation acts will always be clear from the context. A vertex is called isolated if it has degree zero in the graph $CM_n(\mathbf{d}, p_c(\lambda))$. We define the component size corresponding to an isolated vertex to be zero (see Remark 38 below). For any component $\mathscr{C} \subset CM_n(\mathbf{d}, p_c(\lambda))$, let $SP(\mathscr{C})$ denote the number of surplus edges given by the number of edges in $\mathscr{C} - |\mathscr{C}| + 1$. Finally, let

$\mathbf{Z}_n(\lambda)$ denote the vector $(n^{-\rho}|\mathscr{C}_{(i)}(p_c(\lambda))|, \mathrm{SP}(\mathscr{C}_{(i)}(p_c(\lambda))))_{i \geqslant 1}$, ordered as an element of $\mathbb{U}_\downarrow^0$. The following theorem gives the asymptotics for the critical component sizes and the surplus edges of $\mathrm{CM}_n(\mathbf{d}, p_c(\lambda))$:

**Theorem 6.1.1** (Component sizes and surplus edges). *Under Assumption 6.1, as $n \to \infty$,*

$$\mathbf{Z}_n(\lambda) \xrightarrow{d} \mathbf{Z}(\lambda)$$

*with respect to the $\mathbb{U}_\downarrow^0$ topology, where $\mathbf{Z}(\lambda)$ is defined in Section 6.1.1.*

**Remark 38.** Note that if $\tau \in (2,3)$, then $2\rho < 1$. When percolation is performed with probability $p_c$, there are of the order $n$ isolated vertices and thus $n^{-2\rho}$ times the number of isolated vertices tends to infinity. This is the reason why we must ignore the contributions due to isolated vertices, when considering the convergence of the component sizes in the $\ell_\downarrow^2$ topology. Note that an isolated vertex with self-loops does not create an isolated component.

For a connected graph $G$, let $\mathrm{diam}(G)$ denote its diameter. Our next result shows that the diameter of the largest connected components is of constant order.

**Theorem 6.1.2** (Diameter of largest clusters). *Under Assumption 6.1, for any $i \geqslant 1$, $(\mathrm{diam}(\mathscr{C}_{(i)}(p_c(\lambda))))_{n \geqslant 1}$ is a tight sequence of random variables, .*

### Behavior in the near-critical regimes

We now present asymptotic results for the component sizes in the so-called barely subcritical ($p_n \ll p_c(\lambda)$) and barely supercritical regimes ($p_n \gg p_c(\lambda)$). The following two theorems summarize the near-critical behavior:

**Theorem 6.1.3** (Barely subcritical regime). *For $\mathrm{CM}_n(\mathbf{d}, p_n)$, let us assume that $\frac{\log(n)}{\ell_n} \ll p_n \ll p_c(\lambda)$ and that Assumption 6.1 holds. Then, for each fixed $i \geqslant 1$, as $n \to \infty$,*

$$\frac{|\mathscr{C}_{(i)}(p_n)|}{n^\alpha p_n} \xrightarrow{\mathbb{P}} \theta_i.$$

For the result about the barely supercritical regime, we need one further mild technical assumption, which is as follows: Let $D_n^*$ denote the degree of a vertex chosen in a size-biased manner with the sizes being $(d_i/\ell_n)_{i \in [n]}$. Then, there exists a constant $\kappa > 0$ such that

$$1 - \mathbb{E}[e^{-t p_n^{1/(3-\tau)} D_n^*}] = \kappa p_n^{(\tau-2)/(3-\tau)}(t^{\tau-2} + o(1)). \tag{6.4}$$

**Theorem 6.1.4** (Barely supercritical regime). *For* $CM_n(\mathbf{d}, p_n)$, *suppose that* $p_n \gg p_c(\lambda)$ *and Assumption 6.1, (6.4) holds. Then, as* $n \to \infty$,

$$\frac{|\mathscr{C}_{(1)}(p_n)|}{np_n^{1/(3-\tau)}} \xrightarrow{\mathbb{P}} \frac{\mu\kappa^{1/(3-\tau)}}{2^{(\tau-2)/(3-\tau)}}, \qquad \frac{E(\mathscr{C}_{(1)}(p_n))}{np_n^{1/(3-\tau)}} \xrightarrow{\mathbb{P}} \frac{\mu\kappa^{1/(3-\tau)}}{2^{(4-\tau)/(3-\tau)}},$$

*and for all* $i \geqslant 2$, $|\mathscr{C}_{(i)}(p_n)| = o_{\mathbb{P}}(np_n^{1/(3-\tau)})$, $E(\mathscr{C}_{(i)}(p_n)) = o_{\mathbb{P}}(np_n^{1/(3-\tau)})$, *where* $E(G)$ *denotes the number of edges in the graph* $G$.

**Remark 39.** The identity (6.4) is basically a version of the celebrated Abel-Tauberian theorem [84, Chapter XIII.5] (see also [38, Chapter 1.7]). However, since both $D_n^*$ and $p_n$ depend on $n$, the joint asymptotics needs to be stated as an assumption. In Section 6.1.3, we discuss how this assumption is satisfied for power-law degree distributions with $\tau \in (2, 3)$.

### 6.1.2    Effect of the single-edge constraint

In this section, we will consider two random graph models that do not allow for self-loops or multiple edges in the graph, namely the generalized random graph and the erased configuration model. We will see that for random graphs that generate simple graphs, the critical window for percolation is given by

$$p_c = p_c(\lambda) := \lambda n^{-\frac{3-\tau}{2}}(1 + o(1)), \quad \lambda \in (0, \infty). \tag{6.5}$$

We state the results about two different random graph models in two different subsections below, starting with the generalized random graph:

**Generalized random graphs**

Given a set of weights $(w_i)_{i \in [n]}$ on the vertex set $[n]$, the generalized random graph model [54], denoted by $GRG_n(\boldsymbol{w})$ is generated by creating an edge between vertex $i$ and $j$ independently with probability

$$p_{ij} = \frac{w_i w_j}{\ell_n + w_i w_j}, \tag{6.6}$$

where $\ell_n = \sum_{i \in [n]} w_i$. This model has the property that, conditionally on the degree sequence $\mathbf{d}$, the law of the obtained random graph is the same as a uniformly chosen graphs from the space of all simple graphs with degree distribution $\mathbf{d}$. The graph $GRG_n(\boldsymbol{w}, p)$ is obtained by keeping each edge of the graph independently with probability $p$. The later deletion process is also independent of the randomization of the graph.

**Assumption 6.2.** For some $\tau \in (2,3)$, consider the distribution function satisfying $(1-F)(x) = Cx^{-(\tau-1)}$ and let $w_i = (1-F)^{-1}(i/n)$.

In the above case, if $W_n$ denotes the weight of a typical vertex, then

$$\mathbb{E}[W_n] = \frac{1}{n} \sum_{i \in [n]} w_i \to \mu = \mathbb{E}[W].$$

Moreover,

$$n^{-\alpha} w_i = c_F i^{-\alpha},$$

for some constant $c_F > 0$. Throughout $c_F$ will denote the special constant appearing above. Assumption 6.2 is strictly stronger than Assumption 6.1 in the sense that Assumption 6.2 specifies not only the high-degree vertices but all the $w_i$'s. This is required in the proofs as one needs precise estimates of quantities like $\mathbb{E}[W_n \mathbf{1}\{W_n \geqslant K_n\}]$. However, Assumption 6.2 also yields that the weight sequence satisfies a power law with exponent $\tau \in (2,3)$.

Let $\mathscr{C}_{(i)}(p)$ denote the $i$-th largest component of $\mathrm{GRG}_n(w,p)$, and define $W_{(i)}(p) := \sum_{k \in \mathscr{C}_{(i)}(p)} w_k$. We will consider the scaling limits of $(W_{(i)}(p_c))_{i \geqslant 1}$ and $(\mathscr{C}_{(i)}(p_c))_{i \geqslant 1}$. To describe the limiting object, consider the graph $G_\infty(\lambda)$ on the vertex set $\mathbb{Z}_+$, where the vertices $i$ and $j$ are joined independently by $\mathrm{Poisson}(\lambda_{ij})$ many edges with $\lambda_{ij}$ given by

$$\lambda_{ij} := \lambda^2 \int_0^\infty \theta_i(x)\theta_j(x)dx, \quad \theta_i(x) := \frac{c_F^2 i^{-\alpha} x^{-\alpha}}{\mu + c_F^2 i^{-\alpha} x^{-\alpha}}. \tag{6.7}$$

Let $W_{(i)}^\infty(\lambda)$ denote the $i$-th largest element of the set

$$\left\{ \sum_{i \in \mathscr{C}} \theta_i : \mathscr{C} \text{ is a connected component of } G_\infty(\lambda) \right\}.$$

The following is the main result:

**Theorem 6.1.5** (Critical regime for $\mathbf{GRG_n(w)}$). *There exists an absolute constant $\lambda_0$ such that for any $\lambda \in (0, \lambda_0)$, under Assumption 6.2, as $n \to \infty$,*

$$n^{-\alpha}(W_{(i)}(p_c(\lambda)))_{i \geqslant 1} \xrightarrow{d} (W_{(i)}^\infty(\lambda))_{i \geqslant 1},$$

*and*

$$(n^\alpha p_c)^{-1}(|\mathscr{C}_{(i)}(p_c(\lambda))|)_{i \geqslant 1} \xrightarrow{d} (W_{(i)}^\infty(\lambda))_{i \geqslant 1},$$

*with respect to the $\ell^2_\downarrow$ topology.*

In the proofs we will also need to show that $W_{(i)}^\infty(\lambda) < \infty$ almost surely for all $i \geqslant 1$. In fact we will prove the following about the limiting object:

**Proposition 6.1.6.** *There exists an absolute constant $\lambda_0$ such that for any $\lambda \in (0, \lambda_0)$, $\mathbf{W}^\infty(\lambda) := (W^\infty_{(i)}(\lambda))_{i \geqslant 1}$ is in $\ell^2_\downarrow$ almost surely.*

**Remark 40.** In the proofs under the single-edge constraint, coming up with an analyzable exploration process for the clusters seems challenging. The only tool we have is an estimate of the connection probabilities of hubs via an intermediate vertex, which allows us to estimate expectations of several moments of component sizes and total weights of those component. These are often referred to as susceptibility functions. The susceptibility functions allow us to ignore negligible contributions on the total weights of cluster using the first-moment method. Unfortunately, the first-moment method does not work for high values of $\lambda$. Also, we do not know how to show the finiteness of the limiting object in Proposition 6.1.6 for large $\lambda$. This is the reason for assuming $\lambda \in (0, \lambda_0)$ in Theorem 6.1.5. The proof for general $\lambda$ is an open question.

**Remark 41.** Theorem 6.1.5 also holds under different choices of $p_{ij}$'s than given by (6.6). For example, for the Chung-Lu Model ($p_{ij} := \min\{w_i w_j / \ell_n, 1\}$) and the Norros-Reittu model (with $p_{ij} := 1 - e^{-w_i w_j / \ell_n}$), the statement of Theorem 6.1.5 holds, with the scaling limit obtained in an identical manner with $\theta_i(x)$ is (6.7) replaced respectively by $\min\{c_F^2 i^{-\alpha} x^{-\alpha} / \mu, 1\}$, and $1 - e^{-c_F^2 i^{-\alpha} x^{-\alpha} / \mu}$.

**Erased configuration model**

The erased configuration model is obtained by erasing self-loops and multiple edges of $\mathrm{CM}_n(\mathbf{d})$. We denote this random graph by $\mathrm{ECM}_n(\mathbf{d})$, and we denote the graph obtained after bond percolation on $\mathrm{ECM}_n(\mathbf{d})$ by $\mathrm{ECM}_n(\mathbf{d}, p)$. We will assume that $\mathbf{d}$ satisfies Assumption 6.2, where $F$ is a distribution function supported on the non-negative integers. Thus, we take $d_i = (1 - F)^{-1}(i/n)$, and we add an extra half-edge to vertex 1 if $\sum_{i \in [n]} d_i$ is odd. The limiting object for $\mathrm{ECM}_n(\mathbf{d})$ is identical to $\mathrm{GRG}_n(\mathbf{w})$ after replacing $\theta_i(x)$ in (6.7) by

$$\theta_i(x) := 1 - e^{-c_F^2 i^{-\alpha} x^{-\alpha} / \mu}. \tag{6.8}$$

**Theorem 6.1.7** (Critical regime for $\mathbf{ECM_n(d)}$). *There exists an absolute constant $\lambda_0$ such that for any $\lambda \in (0, \lambda_0)$, under Assumption 6.2, the scaling limit results in Theorem 6.1.5 hold for $\mathrm{ECM}_n(\mathbf{d}, p_c(\lambda))$ with limit objects described by (6.8) above.*

Notice that first performing percolation and then erasing self-loops and multiple-edges gives a different random graph than first erasing the self-
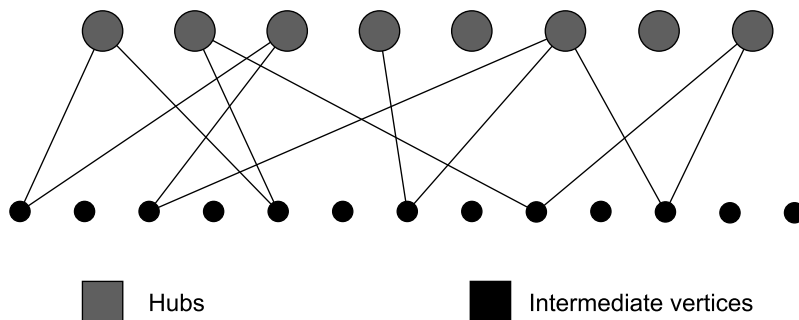
Figure 6.1: Visualization of the connectivity between the hubs (vertices of degree $\Theta(n^\alpha)$). The intermediate vertices have degree $\Theta(n^\rho)$.

loops and multiple-edges and then performing percolation. For the $\tau > 3$ case however, the order of these operations does not matter and leads to the same scaling limits as [69, 70]. However, the operations of deletion of self-loops and multiple-edges and performing percolation are not interchangeable in the $\tau \in (2,3)$ regime, as evidenced by Theorems 6.1.1 and 6.1.7. This can be understood intuitively. In $CM_n(\mathbf{d})$, vertices $i$ and $j$ share $d_i d_j/(\ell_n - 1)$ edges in expectation. Thus for hubs with $d_i = O(n^\alpha)$ and $d_j = O(n^\alpha)$, in expectation $O(1)$ many edges survive after percolation in the critical window (6.3). On the other hand, whenever $p \to 0$, hubs are never connected directly under the single-edge constraint. We will see in the proofs that the value $p_c$ in (6.5) is such that the hubs are connected to each other via intermediate vertices of degree $\Theta(n^\rho)$ (see Figure 6.1). This forms a core of the largest connected components, and the 1-neighborhood of this core spans the largest connected component asymptotically.

**Behavior in the near-critical regimes**

In this section, we state the results about the barely subcritical regime under the single-edge constraint. The result below holds for both percolation on $GRG_n(\mathbf{w})$ and $ECM_n(\mathbf{d})$:

**Theorem 6.1.8.** *Suppose that Assumption 6.2 holds and $p_n \ll p_c(\lambda)$. Then, for any fixed $i \geqslant 1$, as $n \to \infty$,*

$$\frac{|\mathscr{C}_{(i)}(p_n)|}{n^\alpha p_n} \xrightarrow{\mathbb{P}} c_F i^{-\alpha}, \quad and \quad \frac{W_{(i)}(p_n)}{n^\alpha} \xrightarrow{\mathbb{P}} c_F i^{-\alpha}.$$

Under the single-edge constraint, the exact asymptotics in the barely super-critical case is left to future work. In the proofs under the single-edge con-

straint, coming up with an analyzable exploration process for the clusters seems challenging. The only tool we have is an estimate of the connection probabilities of hubs via an intermediate vertex, which allows us to estimate expectations of several moments of component sizes and total weights of those component. These are often referred to as susceptibility functions. The susceptibility functions allow us to ignore negligible contributions on the total weights of cluster using first-moment method. Unfortunately, the first-moment method does not work in the barely-supercritical regime, or even high values of $\lambda$ in Theorems 6.1.5, 6.1.7. This is the reason for assuming $\lambda \in (0, \lambda_0)$ in those theorems.

### 6.1.3    Discussion

**Assumption on the degrees.**    Note that Assumption 6.1 is weaker than Assumption 6.2. Indeed, suppose that $d_i = (1-F)^{-1}(i/n)$, for some distribution function $F$ supported on non-negative integers, and $(1-F)(x) = Ck^{-(\tau-1)}$, for $k \leqslant x < k+1$. We ignore the effect due to adding a dummy half-edge to vertex 1 if necessary to make $\sum_{i \in [n]} d_i$ even, since this does not change any asymptotics. Now, Assumption 6.1 (i) is satisfied with $\theta_i = Ci^{-\alpha}$. One can also verify Assumption 6.1 (ii) using identical arguments as [69, Lemma 6]. For this specific choice of $d_i$, (6.4) holds as well. To see this, write $t_n = t p_n^{1/(3-\tau)}$, and note that

$$1 - \mathbb{E}[e^{-t_n D_n^*}] = \frac{1}{\ell_n} \sum_{k \in [n]} d_k (1 - e^{-t_n d_k}).$$

Let us split the last sum in two parts on the set $\{k : d_k < 1/(2t_n)\}$ and its complement, and denote them by (I) and (II) respectively. We write $a_n \sim b_n$ to denote that $\lim_{n \to \infty} a_n/b_n = 1$. Using the fact that $x - x^2/2 \leqslant 1 - e^{-x} \leqslant x$, it follows that for some constants $c_1, c_2 > 0$,

$$\frac{\text{(I)}}{t_n^{\tau-2}} \sim t_n^{3-\tau} \int_0^{1/(2t_n)} \frac{dx}{x^{\tau-2}} \sim c_0, \quad \text{and} \quad \frac{\text{(II)}}{t_n^{\tau-2}} \sim t_n^{-(\tau-2)} \int_{1/(2t_n)}^{\infty} \frac{dx}{x^{\tau-1}} \sim c_1,$$

which yields (6.4).

**Critical windows: emergence of hub connectivity.**    The critical window changes due to the single-edge constraint, as noted in (6.3) and (6.5). However, there are some common features. Firstly, the component sizes are of the order $n^\alpha p_c$ in both the regimes. This is due to the fact that the main

contribution to the component sizes comes from hubs and their finite neighborhood. Secondly, in both cases, the critical window is the regime in which hubs start getting connected. More precisely, if critical window is given by those values of $\pi$ such that for any fixed $i, j \geqslant 1$

$$\liminf_{n \to \infty} \mathbb{P}(i, j \text{ are in the same component in the } \pi\text{-percolated graph}) \in (0, 1).$$

For the configuration model, hubs are connected directly with strictly positive probability, while under the single-edge constraint, hubs are connected via intermediate vertices of degree $\Theta(n^\rho)$. Intuitively, in the barely subcritical regime, all the hubs are in different components. Hubs start forming the critical components as the $p$ varies over the critical window since most paths between hubs are of length 2 and go via intermediate vertices of degree $\Theta(n^\rho)$. Finally in the barely super-critical regime the giant component is formed which contains all the hubs. This feature is also observed in the $\tau \in (3, 4)$ case [33]. However, the distinction between $\tau \in (3, 4)$ and $\tau \in (2, 3)$ is that for $\tau \in (3, 4)$ the paths between the hubs have a length that grows as $n^{(\tau-3)/(\tau-1)}$.

## 6.2   Configuration model: Proofs

In this section, we prove our results related to critical percolation on $CM_n(\mathbf{d})$. We start by proving some properties of the process (6.1) in Section 6.2.1. In Section 6.2.2, we describe a way to approximate percolation on a configuration model by a suitable configuration model. In Section 6.2.3, we analyze the latter graph by setting up an exploration process and obtaining its scaling limit. Hence the proof of Theorem 6.1.1 is completed. In Section 6.2.4, we consider the near critical behavior and provide proofs of Theorems 6.1.3 and 6.1.4.

### 6.2.1   Properties of the excursions of the limiting process

In the following proposition, we summarize the properties of the limiting process $\mathbf{S}_\infty^\lambda$ that are required in our analysis.

**Proposition 6.2.1.** *(P1) As* $t \to \infty$, $S_\infty^\lambda(t) \xrightarrow{a.s.} -\infty$. *Thus,* $\mathbf{S}_\infty^\lambda$ *does not have an excursion of infinite length almost surely.*

*(P2) For any* $\delta > 0$, $\mathbf{S}_\infty^\lambda$ *has only finitely many excursions of length at least* $\delta$ *almost surely.*

*(P3) Let $\mathcal{R}$ denote the set of excursion end-points of $\mathbf{S}_\infty^\lambda$. Then $\mathcal{R}$ does not have an isolated point.*

*(P4) For any $t > 0$, $\mathbb{P}(S_\infty^\lambda(t) = \inf_{u \leqslant t} S_\infty^\lambda(u)) = 0$.*

The conditions in Proposition 6.2.1 form the bedrock of using [69, Lemma 14], which will be crucial in the next section. An inquisitive reader might note that the conditions are related to [11, Proposition 14]. The proof of Proposition 6.2.1 requires the analysis of the martingale decomposition for the process $\mathbf{S}_\infty^\lambda$. Consider the sigma-field $\mathscr{F}_t = \sigma(\{\xi_i \leqslant s\} : s \leqslant t, i \geqslant 1)$, where for a collection of sets $\mathcal{A}$, $\sigma(\mathcal{A})$ denotes the minimum sigma algebra containing all the sets in $\mathcal{A}$. Then $(\mathscr{F}_t)_{t \geqslant 0}$ is a filtration. All the martingales in this section will be with respect to $(\mathscr{F}_t)_{t \geqslant 0}$, unless stated otherwise. Without loss of generality we assume that $\mu = 1$ in this section to simplify notation. Below we summarize the martingale decomposition for $\mathbf{S}_\infty^\lambda$:

**Lemma 6.2.2.** *The process $\mathbf{S}_\infty^\lambda$ admits the Doob-Meyer decomposition $S_\infty^\lambda(t) = M(t) + A(t)$ with the drift term $A(t)$, and the quadratic variation for the martingale term $\langle M \rangle(t)$ given by*

$$A(t) = \lambda \sum_{i=1}^\infty \theta_i^2 \min\{\xi_i, t\} - 2t, \qquad \langle M \rangle(t) = \lambda^2 \sum_{i=1}^\infty \theta_i^3 \min\{\xi_i, t\}.$$

*Proof.* Define $M_i(t) = \mathbf{1}\{\xi_i \leqslant t\} - \theta_i \min\{\xi_i, t\}$. The proof follows if we show that $(M_i(t))_{t \geqslant 0}$ is a martingale with quadratic variation term given by

$$\langle M_i \rangle(t) = \theta_i \min\{\xi_i, t\}.$$

Denote $\mathcal{I}_i(t) = \mathbf{1}\{\xi_i \leqslant t\}$. Let $(\mathcal{N}(t))_{t \geqslant 0}$ denote a unit jump Poisson process. Then $\mathcal{I}_i(t)$ can be written in terms of the random time change of $(\mathcal{N}(t))_{t \geqslant 0}$ [80, 139] as follows

$$\mathcal{I}_i(t) = \mathcal{N}\left(\theta_i \int_0^t (1 - \mathcal{I}_i(s))ds\right).$$

Further, $\int_0^t (1 - \mathcal{I}_i(s))ds = \min\{\xi_i, t\}$ which completes the proof. $\qquad\square$

The rest of the section is devoted to proving the properties of $\mathbf{S}_\infty^\lambda$ stated in Proposition 6.2.1. We give the proofs of different conditions separately below:

*Proof of Proposition 6.2.1 (P1).* We use the martingale decomposition of $\mathbf{S}_\infty^\lambda$ from Lemma 6.2.2. Fix $K \geqslant 1$ such that $\lambda \sum_{i > K} \theta_i^2 < 1$. Such a choice of $K$ is

always possible as $\theta \in \ell_{\downarrow}^2$. Further define the stopping time $T := \inf\{t : \xi_i \leqslant t, \ \forall i \in [K]\}$, and observe that $T < \infty$ almost surely. Note that $\min\{\xi_i, t\} \leqslant t$ and thus,

$$\frac{1}{t}\lambda \sum_{i > K} \theta_i^2 \min\{\xi_i, t\} \leqslant 1, \quad \text{almost surely.}$$

Therefore, for any $t > T$,

$$A(t) = \lambda \sum_{i \in [K]} \theta_i^2 \xi_i + \lambda \sum_{i > K} \theta_i^2 \min\{\xi_i, t\} - 2t \leqslant \lambda \sum_{i \in [K]} \theta_i^2 \xi_i - t, \quad \text{almost surely.}$$

We conclude that, for any $r \in (0, 1)$, $t^{-r} A(t) \xrightarrow{a.s.} -\infty$. For the martingale part we will use the exponential concentration inequality [149, Inequality 1, Page 899], which is stated below:

**Lemma 6.2.3.** *If $M$ is any continuous time local martingale such that $M(0) = 0$, and $\sup_{t \in [0, \infty)} |M(t) - M(t-)| \leqslant c$, almost surely, then for any $t > 0$, $a > 0$ and $b > 0$,*

$$\mathbb{P}\left( \sup_{s \in [0, t]} M(s) > a, \text{ and } \langle M \rangle(t) \leqslant b \right) \leqslant \exp\left( -\frac{a^2}{2b}\psi\left(\frac{ac}{b}\right) \right),$$

*where $\psi(x) = ((1 + x)\log(1 + x) - x)/x^2$.*

In particular, $\psi(x) \geqslant 1/(2(1 + x/3))$ (see [115, Page 27]). Note that $\langle M \rangle(t) \leqslant \lambda^2 t \sum_{i=1}^{\infty} \theta_i^3$. We apply Lemma 6.2.3 with $a = \varepsilon t^r$, $b = \lambda^2 t \sum_{i=1}^{\infty} \theta_i^3$, and $c = \theta_1$. Now, $\psi(ac/b) \geqslant C/(1 + t^{r-1})$, and thus for any $\varepsilon > 0$, and $r \in (1/2, 1)$

$$\mathbb{P}\left( \sup_{s \in [0, t]} |M(s)| > \varepsilon t^r \right) \leqslant 2\exp(-Ct^{2r-1}),$$

for some constant $C > 0$, where the bound on the absolute value of $M$ follows from the fact that $-M$ is also a martingale, so Lemma 6.2.3 applies to $-M$ as well. Now an application of the Borel-Cantelli lemma proves that $t^{-r}|M(t)| \xrightarrow{a.s.} 0$, for any $r \in (1/2, 1)$. This fact, together with the asymptotics of the drift term, completes the proof. $\qquad\square$

*Proof of Proposition 6.2.1 (P2).* Let $t_k = (k-1)\delta/2$ and define the event

$$C_k^\delta := \left\{ \sup_{t \in (t_{k-1}, t_k]} S_\infty^\lambda(t_{k+1}) - S_\infty^\lambda(t) > 0 \right\}.$$

Suppose that there is an excursion $(l, r)$ with $r - l > \delta$ and $l \in (t_{k-1}, t_k]$ for some $k$. Since $r > t_{k+1}$, $S_\infty^\lambda(t_{k+1}) > S_\infty^\lambda(l) \geqslant \inf_{t \in (t_{k-1}, t_k]} S_\infty^\lambda(t)$, and therefore $C_k^\delta$ must occur. Therefore, if $S_\infty^\lambda$ has infinitely many excursions of length at least $\delta$, then $C_k^\delta$ must occur infinitely often. Using the Borel-Cantelli lemma, the proof follows if we can show that

$$\sum_{k=1}^\infty \mathbb{P}(C_k^\delta) < \infty.$$

As before, fix $K \geqslant 1$ such that $\lambda \sum_{i > K} \theta_i^2 < 1$, and let $T := \inf\{t : \xi_i \leqslant t, \forall i \in [K]\}$. Notice that for each $K \geqslant 1$,

$$\sum_{k=1}^\infty \mathbb{P}\left(T > t_{k-1}\right) = \sum_{k=1}^\infty \mathbb{P}\left(\exists i \in [K] : \xi_i > t_{k-1}\right) \leqslant \sum_{k=1}^\infty K e^{-\theta_K (k-1)\delta/2} < \infty,$$

and therefore it is enough to show that

$$\sum_{k=1}^\infty \mathbb{P}(C_k^\delta \cap \{T \leqslant t_{k-1}\}) < \infty.$$

Now,

$$\sup_{t \in [t_{k-1}, t_k]} S_\infty^\lambda(t_{k+1}) - S_\infty^\lambda(t)$$

$$\leqslant M(t_{k+1}) + \sup_{t \in [t_{k-1}, t_k]} -M(t) + \sup_{t \in [t_{k-1}, t_k]} A(t_{k+1}) - A(t)$$

$$\leqslant M(t_{k+1}) - M(t_{k-1}) + \sup_{t \in [t_{k-1}, t_k]} M(t_{k-1}) - M(t)$$

$$+ \sup_{t \in [t_{k-1}, t_k]} \left[ \lambda \sum_{i=1}^\infty \theta_i^2(\min\{\xi_i, t_{k+1}\} - \min\{\xi_i, t\}) - (t_{k+1} - t) \right] - \frac{\delta}{2}$$

$$\leqslant 2 \sup_{t \in [t_{k-1}, t_{k+1}]} |M(t) - M(t_{k-1})| - \frac{\delta}{2}$$

$$+ \sup_{t \in [t_{k-1}, t_k]} \left[ \lambda \sum_{i=1}^\infty \theta_i^2(\min\{\xi_i, t_{k+1}\} - \min\{\xi_i, t\}) - (t_{k+1} - t) \right].$$

The third term is negative on the event $\{\tau \leqslant t_{k-1}\}$. Thus we only need to estimate the probability

$$\mathbb{P}\left( \sup_{t \in [t_{k-1}, t_{k+1}]} |M(t) - M(t_{k-1})| > \frac{\delta}{4} \right).$$

Note that $M(t) - M(t_{k-1})$ is a martingale with respect to the filtration $(\mathscr{F}_t)_{t \geqslant t_{k-1}}$ with quadratic variation given by

$$\lambda^2 \sum_{i=1}^{\infty} \theta_i^3 \big( \min\{\xi_i, t\} - \min\{\xi_i, t_{k-1}\} \big).$$

Further, $\mathbb{E}[\min\{\xi_i, t\}] = \theta_i^{-1}(1 - e^{-\theta_i t})$. Therefore, Doob's martingale inequality [128, Theorem 1.9.1.3] implies

$$\sum_{k=1}^{\infty} \mathbb{P}\left( \sup_{t \in [t_{k-1}, t_{k+1}]} |M(t) - M(t_{k-1})| > \frac{\delta}{4} \right)$$
$$\leqslant \sum_{k=1}^{\infty} \frac{16}{\delta^2} \sum_{i=1}^{\infty} \theta_i^2 (e^{-\theta_i t_{k-1}} - e^{-\theta_i t_{k+1}}).$$

By interchanging the sums, the last term is finite and the proof now follows.

$\square$

*Proof of Proposition 6.2.1 (P3).* Define the process

$$L(t) = \lambda \sum_{i=1}^{\infty} \theta_i N_i(t) - 2t,$$

where $(N_i(t))_{t \geqslant 0}$ is a rate $\theta_i$ Poisson process, independently over $i$. We assume that $\mathbf{S}_{\infty}^{\lambda}$ and $\mathbf{L}$ are coupled by taking $\mathcal{I}_i(s) = \mathbf{1}\{N_i(s) \geqslant 1\}$, so that $S_{\infty}^{\lambda}(t) \leqslant L(t)$. Using [28, Chapter VII.1, Theorem 1] and the fact that $\sum_i \theta_i = \infty$,

$$\inf\{t > 0 : L(t) < 0\} = 0, \quad \text{almost surely.} \tag{6.9}$$

Moreover, for any stopping time $\mathcal{T} > 0$, $(S_{\infty}^{\lambda}(\mathcal{T} + t) - S_{\infty}^{\lambda}(\mathcal{T}))_{t \geqslant 0}$, conditioned on the sigma-field $\sigma(S_{\infty}^{\lambda}(s) : s \leqslant \mathcal{T})$, is distributed as a process defined in (6.1) for some random $\theta$. Now we can take $\mathcal{T}$ to be an excursion endpoint and the proof follows.

$\square$

*Proof of Proposition 6.2.1 (P4).* We leverage the proof techniques in [11, Proposition 14 (b)]. Define the process

$$Q_m(t) = \lambda \sum_{i=m+1}^{\infty} \theta_i \mathbf{1}_{A_i} M_i(t) - 2t,$$

where $(A_i)_{i \geqslant 1}$ is a sequence of independent events with $\mathbb{P}(A_i) = (1 - 2\theta_i t_0)^+$, and $(M_i)_{i \geqslant 1}$ are independent Poisson processes with rates $\theta_i e^{-\theta_i t_0}$. Now,

$\mathbf{Q}_m$ is a Lévy process, and thus we can apply (6.9) together with [28, Chapter VI.1, Proposition 3] (see also the remark below the statement of the cited proposition) to conclude that, for any $t_0 > 0$ and $m \geqslant 1$,

$$\mathbb{P}\Big(Q_m(t_0) = \inf_{t \in [0, t_0]} Q_m(t)\Big) = 0. \tag{6.10}$$

Now, if one can couple $(Q_{2m}(t))_{t \in [0, t_0]}$ and $(S_\infty(t))_{t \in [0, t_0]}$ in such a way that

$$\lim_{m \to \infty} \mathbb{P}\left(Q_{2m}(t_0) - Q_{2m}(t) \leqslant S_\infty(t_0) - S_\infty(t), \ \forall t \in [0, t_0]\right) = 0, \tag{6.11}$$

then (6.10) and (6.11) together complete the proof. To see (6.11), write $(\xi_{i,j}, j \in J_i)$ for the set of points of $\mathbf{M}_i$ in $[0, t_0]$ if $A_i$ occurs, but to be the empty set if $A_i$ does not occur. Call a coupling *successful* if $(\xi_{i,j} : i > 2m, j \in J_i)$ and $(\xi_i)_{i \geqslant 1}$ are coupled in such a way that $\xi_{i,j} = \xi_{h(i,j)}$ for some random variable $h(i, j) \leqslant i$, and the values $(h(i, j) : i > 2m, j \in J_i)$ are distinct. Aldous and Limic [11] showed the existence of a coupling such that the probability of the coupling is successful tends to 0 as $m \to \infty$. Although their proof is under a different setting with $\theta \in \ell_\downarrow^3 \setminus \ell_\downarrow^2$, the proof of the coupling holds under the assumption that $\theta \in \ell_\downarrow^2 \setminus \ell_\downarrow^1$. Moreover, (6.11) holds under a successful coupling and thus the proof is complete. $\qquad\square$

### 6.2.2   Sandwiching the percolated configuration model

A key step in all our proofs is to *approximate* $CM_n(\mathbf{d}, p_n)$ by a configuration model, which is given in Proposition 6.2.4 below. This idea has also appeared in the context of the finite third moment [70] and the infinite third moment case [69]. We emphasize that Proposition 6.2.4 holds for percolation on the configuration model without any specific assumption on the degree distribution, as long as $\ell_n p_n \gg \log(n)$. We start by describing the approximating configuration model below:

**Algorithm 6.1.** (S0) Keep each half-edge with probability $p_n$. If the total number of retained half-edges is odd, attach a *dummy* half-edge to vertex 1.

(S1) Perform a uniform perfect matching among the retained half-edges, i.e., within the retained half-edges, pair unpaired half-edges sequentially with a uniformly chosen unpaired half-edge until all half-edges are paired. The paired half-edges create edges in the graph, and we call the resulting graph $\mathcal{G}_n(p_n)$.

The following proposition formally states that $\mathcal{G}_n(p_n)$ approximates the percolated graph $CM_n(\mathbf{d}, p_n)$:

**Proposition 6.2.4.** *Let $p_n$ be such that $\ell_n p_n \gg \log(n)$. Then there exists a sequence $(\varepsilon_n)_{n \geqslant 1}$ with $\varepsilon_n \to 0$, and a coupling such that, with high probability,*

$$\mathcal{G}_n(p_n(1 - \varepsilon_n)) \subset CM_n(\mathbf{d}, p_n) \subset \mathcal{G}_n(p_n(1 + \varepsilon_n)).$$

*Proof.* The proof relies on an exact construction of $CM_n(\mathbf{d}, p_n)$ by Fountoulakis [85] which goes as follows:

**Algorithm 6.2.** (S0) Perform a binomial trial $X \sim \text{Bin}(\ell_n/2, p_n)$ and choose $2X$ half-edges uniformly at random from the set of all half-edges.

(S1) Perform a perfect matching of these $2X$ chosen half-edges. The resulting graph is distributed as $CM_n(\mathbf{d}, p_n)$.

Notice the similarity between Algorithm 6.1 (S1) and Algorithm 6.2 (S1). In Algorithm 6.1 (S0), given the number of retained half-edges, the choice of the half-edges can be performed sequentially uniformly at random without replacement. Thus, given the number of half-edges in the two algorithms, we can couple the choice of the half-edges, and their pairing (the restriction of a uniform matching to a subset remains). Let $\mathcal{H}_1$, $\mathcal{H}_2^-$ and $\mathcal{H}_2^+$ respectively denote the number of half-edges in $CM_n(\mathbf{d}, p_n)$, $\mathcal{G}_n(p_n(1 - \varepsilon_n))$ and $\mathcal{G}_n(p_n(1 + \varepsilon_n))$. From the above discussion, the proof is complete if we can show that as $n \to \infty$,

$$\mathbb{P}\big(\mathcal{H}_2^- \leqslant \mathcal{H}_1 \leqslant \mathcal{H}_2^+\big) \to 1.$$

Notice that $\mathcal{H}_1 = 2X$, where $X \sim \text{Bin}(\ell_n/2, p_n)$, and $\mathcal{H}_2^+ \sim \text{Bin}(\ell_n, p_n(1 + \varepsilon_n))$. Using standard concentration inequalities [115, Corollary 2.3], it follows that

$$\mathcal{H}_1 = \ell_n p_n + o_\mathbb{P}(\sqrt{\ell_n p_n \log(n)}),$$

and

$$\mathcal{H}_2^+ = \ell_n p_n + \ell_n p_n \varepsilon_n + o_\mathbb{P}(\sqrt{\ell_n p_n \log(n)}).$$

Now, if we choose $\varepsilon_n$ such that $\varepsilon_n \gg (\log(n)/(\ell_n p_n))^{1/2}$ and $\varepsilon_n \to 0$, then, with high probability, $\mathcal{H}_1 \leqslant \mathcal{H}_2^+$. Similarly we can conclude that $\mathcal{H}_2^- \leqslant \mathcal{H}_1$ with high probability. The proof is now complete. $\qquad \square$

We conclude this subsection by stating some properties of the degree sequence of the graph $\mathcal{G}_n(p_n)$ that will be crucial in the analysis below. Let

$\tilde{\mathbf{d}} = (\tilde{d}_1, \ldots, \tilde{d}_n)$ be the degree sequence induced by Algorithm 6.1 (S1). Then the following result holds for $\tilde{\mathbf{d}}$:

**Lemma 6.2.5.** *For each fixed* $i \geqslant 1$, $\tilde{d}_i = d_i p_n (1 + o_{\mathbb{P}}(1))$ *and* $\tilde{\ell}_n = \ell_n p_n (1 + o_{\mathbb{P}}(1))$. *Moreover, for* $p_n \ll p_c$, $\sum_{i \in [n]} \tilde{d}_i^2 = \tilde{\ell}_n (1 + o_{\mathbb{P}}(1))$, *whereas for* $p_n = p_c(\lambda)$ *the following holds: For any* $\varepsilon > 0$,

$$\lim_{K \to \infty} \limsup_{n \to \infty} \mathbb{P}\left( \sum_{i > K} \tilde{d}_i^2 > \varepsilon \tilde{\ell}_n \right) = 0. \tag{6.13}$$

*Proof.* Note that $\tilde{d}_i \sim \mathrm{Bin}(d_i, p_n)$, independently for $i \in [n]$. For each fixed $i \geqslant 1$, $d_i p_n = \Theta(n^\rho)$, which tends to infinity. Thus the first fact follows using [115, Theorem 2.1]. Since, $\tilde{\ell}_n \sim \mathrm{Bin}(\ell_n, p_n)$, the second fact also follows using the same bound. To prove (6.13), we first assume that $p_n = p_c(\lambda)$ given by (6.3). Then, for any $\varepsilon > 0$, the probability in (6.13) is at most

$$\mathbb{P}\left( \sum_{i > K} \tilde{d}_i^2 > \varepsilon \tilde{\ell}_n, \frac{\ell_n p_n}{2} \leqslant \tilde{\ell}_n \leqslant 2 \ell_n p_n \right) + o(1)$$

$$\leqslant \mathbb{P}\left( \sum_{i > K} \tilde{d}_i^2 > \frac{\varepsilon \ell_n p_n}{2} \right) \leqslant \frac{4 p_n \sum_{i > K} d_i^2}{\ell_n},$$

where the last step follows from Markov's inequality. The proof now follows using Assumption 6.1 and $p_n = O(n^{2\alpha - 1})$. The case for $p_n \ll p_c$ follows similarly. $\square$

### 6.2.3    Analysis in the critical window

**Convergence of the exploration process**

Let $\tilde{\mathbf{d}} = (\tilde{d}_1, \ldots, \tilde{d}_n)$ be the degree sequence induced by Algorithm 6.1 (S1) with $p_n = p_c(\lambda)$, and consider $\mathcal{G}_n(p_c(\lambda))$. Note that $\mathcal{G}_n(p_c(\lambda))$ has the same distribution as $\mathrm{CM}_n(\tilde{\mathbf{d}})$. We start by describing how the connected components in the graph can be explored while generating the random graph simultaneously:

**Algorithm 6.3** (Exploring the graph). The algorithm carries along vertices that can be alive, active, exploring and killed and half-edges that can be alive, active or killed. We sequentially explore the graph as follows:

(S0) At stage $i = 0$, all the vertices and the half-edges are *alive* but none of them are *active*. Also, there are no *exploring* vertices.

(S1) At each stage $i$, if there is no active half-edge at stage $i$, choose a vertex $v$ proportional to its degree among the alive (not yet killed) vertices and declare all its half-edges to be *active* and declare $v$ to be *exploring*. If there is an active vertex but no exploring vertex, then declare the *smallest* vertex to be exploring.

(S2) At each stage $i$, take an active half-edge $e$ of an exploring vertex $v$ and pair it uniformly to another alive half-edge $f$. Kill $e, f$. If $f$ is incident to a vertex $v'$ that has not been discovered before, then declare all the half-edges incident to $v'$ active, except $f$ (if any). If $\text{degree}(v') = 1$ (i.e. the only half-edge incident to $v'$ is $f$) then kill $v'$. Otherwise, declare $v'$ to be active and larger than all other vertices that are alive. After killing $e$, if $v$ does not have another active half-edge, then kill $v$ also.

(S3) Repeat from (S1) at stage $i+1$ if not all half-edges are already killed.

Algorithm 6.3 gives a *breadth-first* exploration of the connected components of $CM_n(\tilde{\mathbf{d}})$. Define the exploration process by

$$S_n(0) = 0, \quad S_n(l) = S_n(l-1) + \tilde{d}_{(l)} J_l - 2,$$

where $J_l$ is the indicator that a new vertex is discovered at time $l$ and $\tilde{d}_{(l)}$ is the degree of the new vertex chosen at time $l$ when $J_l = 1$. Suppose $\mathscr{C}_k$ is the $k^{\text{th}}$ connected component explored by the above exploration process and define $\tau_k = \inf\{i : S_n(i) = -2k\}$. Then $\mathscr{C}_k$ is discovered between the times $\tau_{k-1} + 1$ and $\tau_k$, and $\tau_k - \tau_{k-1} - 1$ gives the total number of edges in $\mathscr{C}_k$. Call a vertex *discovered* if it is either active or killed. Let $\mathscr{V}_l$ denote the set of vertices discovered up to time $l$ and $\mathscr{I}_i^n(l) := \mathbf{1}\{i \in \mathscr{V}_l\}$. Note that

$$S_n(l) = \sum_{i \in [n]} \tilde{d}_i \mathscr{I}_i^n(l) - 2l. \tag{6.14}$$

Define the re-scaled version $\bar{\mathbf{S}}_n$ of $\mathbf{S}_n$ by $\bar{S}_n(t) = n^{-\rho} S_n(\lfloor tn^\rho \rfloor)$. Then,

$$\bar{S}_n(t) = n^{-\rho} \sum_{i \in [n]} \tilde{d}_i \mathscr{I}_i^n(tn^\rho) - 2t + o(1). \tag{6.15}$$

Note the similarity between the expressions in (6.1) and (6.15). We will prove the following:

**Theorem 6.2.6.** *Consider the process $\bar{\mathbf{S}}_n := (\bar{S}_n(t))_{t \geqslant 0}$ defined in (6.15) and recall the definition of $\bar{\mathbf{S}}_\infty$ from (6.1). Then, under Assumption 6.1, as $n \to \infty$,*

$$\bar{\mathbf{S}}_n \xrightarrow{\text{d}} \bar{\mathbf{S}}_\infty$$

*with respect to the Skorohod $J_1$ topology.*

*Proof.* We denote $\tilde{\ell}_n(u) = \sum_{i \in [n]} \tilde{d}_i - 2un^\rho$. Since $\tilde{\ell}_n = \Theta_{\mathbb{P}}(n^{2\rho})$, $\tilde{\ell}_n(u) = \tilde{\ell}_n(1 + o_{\mathbb{P}}(1))$ uniformly over $u \leqslant t$. Let $\tilde{\mathbb{P}}(\cdot)$ (respectively $\tilde{\mathbb{E}}[\cdot]$) denote the conditional probability (respectively expectation) conditional on $(\tilde{d}_i)_{i \in [n]}$. Note that, for any $t \geqslant 0$, uniformly over $l \leqslant tn^\rho$

$$\tilde{\mathbb{P}}\left(\mathfrak{I}_i^n(l) = 0\right) \geqslant \left(1 - \frac{\tilde{d}_i}{\tilde{\ell}_n(t)}\right)^l, \quad \text{and} \quad \tilde{\mathbb{E}}\left[\mathfrak{I}_i^n(l)\right] \leqslant \frac{l\tilde{d}_i}{\tilde{\ell}_n(t)}.$$

Now, note that

$$n^{-\rho} \tilde{\mathbb{E}}\left[\sum_{i > K} \tilde{d}_i \mathfrak{I}_i^n(tn^\rho)\right] \leqslant t \frac{\sum_{i > K} \tilde{d}_i^2}{\tilde{\ell}_n(t)}.$$

Using (6.13), it is now enough to deduce the scaling limit for

$$\bar{S}_n^K(t) = n^{-\rho} \sum_{i=1}^{K} \tilde{d}_i \mathfrak{I}_i^n(tn^\rho) - 2t$$

and then taking $K \to \infty$. The next lemma gives the scaling limit for $\bar{S}_n^K$ and completes the proof of Theorem 6.2.6. $\qquad\square$

**Lemma 6.2.7.** *Fix any $K \geqslant 1$, and $\mathfrak{I}_i(s) := \mathbf{1}\{\xi_i \leqslant s\}$ where $\xi_i \sim \text{Exp}(\theta_i/\mu)$ independently for $i \in [K]$. Under Assumption 6.1, as $n \to \infty$,*

$$\left(\mathfrak{I}_i^n(tn^\rho)\right)_{i \in [K], t \geqslant 0} \xrightarrow{d} \left(\mathfrak{I}_i(t)\right)_{i \in [K], t \geqslant 0}.$$

*Proof.* By noting that $(\mathfrak{I}_i^n(tn^\rho))_{t \geqslant 0}$ are indicator processes, it is enough to show that

$$\tilde{\mathbb{P}}\left(\mathfrak{I}_i^n(t_i n^\rho) = 0, \forall i \in [K]\right) \to \tilde{\mathbb{P}}\left(\mathfrak{I}_i(t_i) = 0, \forall i \in [K]\right) = \exp\left(-\mu^{-1} \sum_{i=1}^{K} \theta_i t_i\right).$$

for any $t_1, \ldots, t_K \in \mathbb{R}$. Now,

$$\tilde{\mathbb{P}}\left(\mathfrak{I}_i^n(m_i) = 0, \forall i \in [K]\right) = \prod_{l=1}^{\infty} \left(1 - \sum_{i \leqslant K: l \leqslant m_i} \frac{\tilde{d}_i}{\tilde{\ell}_n - \Theta(l)}\right). \qquad (6.16)$$

Taking logarithms on both sides of (6.16) and using the fact that $l \leqslant \max m_i = \Theta(n^\rho)$ we get

$$\tilde{\mathbb{P}}\left(\mathfrak{I}_i^n(m_i) = 0 \, \forall i \in [K]\right) = \exp\left(-\sum_{l=1}^{\infty} \sum_{i \leqslant K: l \leqslant m_i} \frac{\tilde{d}_i}{\tilde{\ell}_n} + o(1)\right)$$
$$= \exp\left(-\sum_{i \in [K]} \frac{\tilde{d}_i m_i}{\tilde{\ell}_n} + o(1)\right). \qquad (6.17)$$

Putting $m_i = t_i n^\rho$, Assumption 6.1 (i), (ii) gives

$$\frac{m_i \tilde{d}_i}{\tilde{\ell}_n} = \frac{\theta_i t_i}{\mu}(1 + o_{\mathbb{P}}(1)). \tag{6.18}$$

Hence (6.18), and (6.17) complete the proof of Lemma 6.2.7.    □

**Large components are explored early**

Now, we prove two key results that allow us to deduce the convergence of
the component sizes. Firstly, we show that the rescaled vector of component
sizes is tight in $\ell^2_\downarrow$ (see Proposition 6.2.8). This result is then used to show that
the largest components of $\mathcal{G}_n(p_c(\lambda))$ are explored before time $\Theta(n^\rho)$. Let $\mathscr{C}_{(i)}$
denote the $i$-th largest component for $\mathcal{G}_n(p_c(\lambda))$.

**Proposition 6.2.8.** *Under  Assumption 6.1, for any $\varepsilon > 0$,*

$$\lim_{K \to \infty} \limsup_{n \to \infty} \mathbb{P}\left( \sum_{i > K} |\mathscr{C}_{(i)}|^2 > \varepsilon n^{2\rho} \right) = 0.$$

Denote $D_i = \sum_{k \in \mathscr{C}_{(i)}} \tilde{d}_k$. It is enough to show that

$$\lim_{K \to \infty} \limsup_{n \to \infty} \mathbb{P}\left( \sum_{i > K} D_i^2 > \varepsilon n^{2\rho} \right) = 0. \tag{6.19}$$

In the above, we have used our convention that the component size of an
isolated vertex is zero. For a vertex $v$, let $\mathscr{C}(v)$ denote the component con-
taining vertex $v$ in $\mathcal{G}_n(p_n)$ and $D(v) = \sum_{k \in \mathscr{C}(v)} \tilde{d}_k$. Let $\mathcal{G}^K$ be the random
graph obtained by removing all edges attached to vertices $1, \dots, K$ and let $\mathbf{d}'$
be the obtained degree sequence. Further, let $\mathscr{C}^K(v)$ and $\mathscr{C}^K_{(i)}$ denote the con-
nected component containing $v$ and the $i$-th largest component respectively,
and $D^K(v) = \sum_{k \in \mathscr{C}^K(v)} \tilde{d}_k$, $D^K_i = \sum_{k \in \mathscr{C}^K_{(i)}} \tilde{d}_k$. Suppose $V_n^*$ is a vertex of $\mathcal{G}^K$
chosen according to the size-biased distribution with sizes being $(\tilde{d}_k/\tilde{\ell}_n)_{k \geqslant 1}$,
independently of the graph. Denote the criticality parameter of $\mathcal{G}^K$ by $\nu_n^K$.

**Lemma 6.2.9.** *Suppose that  Assumption 6.1 holds. Then, for fixed $K \geqslant 1$ (suffi-
ciently large), with high probability*

$$\tilde{\mathbb{E}}\left[ D^K(V_n^*) \right] \leqslant \frac{\sum_{i > K} \tilde{d}_i^2}{\tilde{\ell}_n - 2 \sum_{i=1}^K \tilde{d}_i} \times \frac{1}{1 - \nu_n^K}. \tag{6.20}$$

*Proof.* We make use of path counting techniques [31, 111]. Note that the criti-
cality parameter of the graph $\mathcal{G}_n(p_c(\lambda))$ is $\tilde{\nu}_n = \lambda(1 + o_{\mathbb{P}}(1))$, by Lemma 6.2.5.

Now, conditional on the set of removed half-edges and $\tilde{\mathbf{d}}$, $\mathcal{G}^K$ is still a configuration model with some degree sequence $\mathbf{d}'$ with $d_i' \leqslant \tilde{d}_i$ for all $i \in [n] \setminus [K]$ and $d_i' = 0$ for $i \in [K]$. Further, the criticality parameter of $\mathcal{G}^K$ satisfies

$$
\begin{aligned}
\nu_n^K &= \frac{\sum_{i \in [n]} d_i'(d_i' - 1)}{\sum_{i \in [n]} d_i'} \leqslant \frac{\sum_{i > K} \tilde{d}_i(\tilde{d}_i - 1)}{\tilde{\ell}_n - 2\sum_{i=1}^K \tilde{d}_i} \\
&= \tilde{\nu}_n \frac{\sum_{i > K} \tilde{d}_i(\tilde{d}_i - 1)}{\sum_{i \in [n]} \tilde{d}_i(\tilde{d}_i - 1)}(1 + o_{\mathbb{P}}(1)) = \lambda \frac{\sum_{i > K} \tilde{d}_i(\tilde{d}_i - 1)}{\sum_{i \in [n]} \tilde{d}_i(\tilde{d}_i - 1)}(1 + o_{\mathbb{P}}(1)) \\
&= \lambda \frac{\sum_{i > K} d_i(d_i - 1)}{\sum_{i \in [n]} d_i(d_i - 1)}(1 + o_{\mathbb{P}}(1)).
\end{aligned}
$$

Now, by Assumption 6.1 and Lemma 6.2.5, it is possible to choose $K_0$ large such that,

$$\text{for all } K \geqslant K_0, \text{ with high probability } \nu_n^K < 1, \tag{6.21}$$

and then we can apply similar arguments as in [31, Section 7] for calculating weight-based susceptibility functions with weights being the degrees. The term $\sum_{i > K} \tilde{d}_i^2 / \sum_{i > K} \tilde{d}_i$ arises in (6.20) due to the fact that

$$\tilde{\mathbb{E}}[\text{degree of } V_n^*] = \frac{\sum_{i \in [n]} d_i'^2}{\sum_{i \in [n]} d_i'} \leqslant \frac{\sum_{i > K} \tilde{d}_i^2}{\tilde{\ell}_n - 2\sum_{i=1}^K \tilde{d}_i}.$$

Thus the proof of Lemma 6.2.9 follows.    $\square$

*Proof of Proposition 6.2.8.* Denote the sum of squares of the D-values excluding the components containing vertices $1, 2, \ldots, K$ by $\mathscr{S}_K$. Note that

$$\sum_{i > K} D_i^2 \leqslant \mathscr{S}_K \leqslant \sum_{i \geqslant 1} (D_i^K)^2.$$

Now, using Lemma 6.2.9 and (6.21), it follows that

$$
\begin{aligned}
\tilde{\mathbb{P}}\left( \sum_{i \geqslant 1} (D_i^K)^2 > \varepsilon n^{2\rho} \right) &\leqslant \frac{1}{\varepsilon n^{2\rho}} \tilde{\mathbb{E}}\left[ \sum_{i \geqslant 1} (D_i^K)^2 \right] \\
&= \frac{1}{\varepsilon n^{2\rho}} \tilde{\mathbb{E}}\left[ D^K(V_n^*) \right] \leqslant O_{\mathbb{P}}(1) n^{-2\rho} \sum_{i > K} \tilde{d}_i^2.
\end{aligned}
$$

Thus, (6.19) follows from Lemma 6.2.5.    $\square$

The next proposition shows that the large components are explored before time $\Theta(n^\rho)$ by Algorithm 6.3. The proof follows using similar arguments as [69, Lemma 13] and we skip it here. Let $\mathscr{C}_{\max}^{\geqslant T}$ denote the size of the largest component that is started exploring by Algorithm 6.3 after time $Tn^\rho$.

**Proposition 6.2.10.** *Under Assumption 6.1, for any $\varepsilon > 0$,*

$$\lim_{T \to \infty} \limsup_{n \to \infty} \mathbb{P}\big(|\mathscr{C}_{\max}^{\geqslant T}| > \varepsilon n^\rho\big) = 0.$$

**Convergence of the component sizes and the surplus edges**

We start by first showing the asymptotics of the component sizes:

**Lemma 6.2.11.** *Under Assumption 6.1, as $n \to \infty$,*

$$(n^{-\rho}|\mathscr{C}_{(i)}|)_{i \geqslant 1} \xrightarrow{d} (\gamma_i(\lambda))_{i \geqslant 1}, \tag{6.22}$$

*with respect to the $\ell^2_\downarrow$ topology.*

*Proof.* Recall [69, Lemma 14] and notice that the process $\mathbf{S}_\infty^\lambda$ satisfies all the *nice* properties stated therein by Proposition 6.2.1 (see [69, Lemma 15] for a similar application in a different context). This observation, together with Proposition 6.2.10 yields the finite-dimensional convergence in (6.22). Finally the proof is completed using Proposition 6.2.8. $\square$

**Lemma 6.2.12.** *Let $N_n^\lambda(k)$ be the number of surplus edges discovered up to time $k$ and $\bar{N}_n^\lambda(u) = N_n^\lambda(\lfloor un^\rho \rfloor)$. Then, as $n \to \infty$,*

$$(\bar{\mathbf{S}}_n, \bar{\mathbf{N}}_n^\lambda) \xrightarrow{d} (\mathbf{S}_\infty^\lambda, \mathbf{N}^\lambda),$$

*where $\mathbf{N}^\lambda$ is defined in (6.2).*

*Proof.* We write $N_n^\lambda(l) = \sum_{i=2}^{l} \xi_i$, where $\xi_i = \mathbf{1}\{\mathscr{V}_i = \mathscr{V}_{i-1}\}$. Let $A_i$ denote the number of active half-edges after stage $i$ while implementing Algorithm 6.3. Note that

$$\mathbb{P}\left(\xi_i = 1 | \mathscr{F}_{i-1}\right) = \frac{A_{i-1} - 1}{\tilde{\ell}_n - 2i - 1} = \frac{A_{i-1}}{\tilde{\ell}_n}(1 + O(i/n)) + O(n^{-1}),$$

uniformly for $i \leqslant Tn^\rho$ for any $T > 0$. Therefore, the instantaneous rate of change of the re-scaled process $\bar{N}^\lambda$ at time $t$, conditional on the past, is

$$n^\rho \frac{A_{\lfloor tn^\rho \rfloor}}{n^{2\rho} \frac{\mu^2}{\sum_{i \geqslant 1} \theta_i^2}}(1 + o(1)) + o(1) = \frac{\sum_{i \geqslant 1} \theta_i^2}{\mu^2} \mathrm{refl}(\bar{S}_n(t))(1 + o(1)) + o(1).$$

Theorem 6.2.6 yields that $\mathrm{refl}(\bar{\mathbf{S}}_n) \xrightarrow{d} \mathrm{refl}(\mathbf{S}_\infty^\lambda)$. Then, by the Skorohod representation theorem, we can assume that $\mathrm{refl}(\bar{\mathbf{S}}_n) \to \mathrm{refl}(\mathbf{S}_\infty^\lambda)$ almost surely on some probability space. Observe that $(\int_0^t \mathrm{refl}(S_\infty^\lambda(u))du)_{t \geqslant 0}$ has continuous sample paths. Therefore, the conditions of [128, Corollary 1, Page 388] are satisfied and the proof is complete. $\square$

**Lemma 6.2.13.** *For any $\varepsilon > 0$,*

$$\lim_{\delta \to 0} \limsup_{n \to \infty} \mathbb{P}\left( \sum_{i : |\mathscr{C}_{(i)}| \leqslant \delta n^\rho} |\mathscr{C}_{(i)}| \times \mathrm{SP}(\mathscr{C}_{(i)}) > \varepsilon b_n \right) = 0.$$

Again, the proof of Lemma 6.2.13 can be carried out in an identical manner as [69, Proposition 19] and therefore is skipped here. The only crucial thing to observe here is that $V_n$ has to be replaced by $V_n^*$ for an analogue of [69, Lemma 20] and one has to consider the cases $\lambda < 1$, and $\lambda > 1$ separately instead of $\lambda < 0$ and $\lambda > 0$ in [69].

*Proof of Theorem 6.1.1.* Let $\mathbf{Z}_n'(\lambda)$ denote the vector $(n^{-\rho}|\mathscr{C}_{(i)}|, \mathrm{SP}(\mathscr{C}_{(i)}))_{i \geqslant 1}$, ordered as an element in $\mathbb{U}_\downarrow^0$. Then, Lemmas 6.2.11, 6.2.12, and 6.2.13 together imply that

$$\mathbf{Z}_n'(\lambda) \xrightarrow{\mathrm{d}} \mathbf{Z}(\lambda).$$

Finally the proof is complete using Proposition 6.2.4. □

**Analysis of the diameter**

In this section, we deduce the asymptotics of the diameter of the components of $\mathscr{G}_n(p_c(\lambda))$, and hence complete the proof of Theorem 6.1.2.

*Proof of Theorem 6.1.2.* Firstly, we can leverage Janson's path counting technique again (see [111, Lemma 5.1]) to show that, given a vertex $v \in [n]$, the expected number of paths of length $l$ from $v$ (conditional on $\tilde{\mathbf{d}}$) is at most $\tilde{\nu}_n^l$. Recall the notations $D(v) = \sum_{k \in \mathscr{C}(v)} \tilde{d}_k$, $D_{(i)} = \sum_{k \in \mathscr{C}_{(i)}} \tilde{d}_k$, and $V_n^* = $ a vertex chosen according to the size-biased distribution $(\tilde{d}_i / \tilde{\ell}_n)_{i \geqslant 1}$. For any fixed $\delta > 0$, define

$$X_n(K) = \sum_{v \in [n]} \mathbf{1}\{D(v) > \delta n^\rho, |\mathscr{C}(v)| > \delta n^\rho, \mathrm{diam}(\mathscr{C}(v)) > K\}.$$

Observe that, for $\lambda < 1$,

$$\tilde{\mathbb{P}}\left(\exists i \geqslant 1 : D_{(i)} > \delta n^\rho, |\mathscr{C}_{(i)}| > \delta n^\rho, \operatorname{diam}(\mathscr{C}_{(i)}) > K\right)$$

$$\leqslant \sum_{i \geqslant 1} \tilde{\mathbb{P}}\left(D_{(i)} > \delta n^\rho, |\mathscr{C}_{(i)}| > \delta n^\rho, \operatorname{diam}(\mathscr{C}_{(i)}) > K\right)$$

$$= \tilde{\ell}_n \sum_{v \in [n]} \frac{\tilde{d}_v}{\tilde{\ell}_n} \tilde{\mathbb{E}}\left[\frac{1}{D(v)}\mathbf{1}\{D(v) > \delta n^\rho, |\mathscr{C}(v)| > \delta n^\rho, \operatorname{diam}(\mathscr{C}(v)) > K\}\right]$$

$$\leqslant \frac{\tilde{\ell}_n}{\delta n^\rho} \tilde{\mathbb{P}}\left(D(V_n^*) > \delta n^\rho, |\mathscr{C}(V_n^*)| > \delta n^\rho, \operatorname{diam}(\mathscr{C}(V_n^*)) > K\right)$$

$$= \frac{\tilde{\ell}_n}{\delta n^\rho} \tilde{\mathbb{P}}\left(X_n(K) \geqslant \delta n^\rho\right) \leqslant \frac{\tilde{\ell}_n}{\delta^2 n^{2\rho}} \sum_{l > K} \tilde{\nu}_n^l = \frac{C\tilde{\ell}_n}{\delta^2 n^{2\rho}} \frac{\lambda^K}{1-\lambda} \xrightarrow{\mathbb{P}} 0,$$

if we first take $\lim_{n \to \infty}$ and then $\lim_{K \to \infty}$. The proof can be generalized naturally to the case $\lambda > 1$. In that case, we delete $R$ high-degree vertices to obtain a new graph $\mathcal{G}^{>R}$, for which the above result holds (see the proof of Lemma 6.2.9). However, after putting back the $R$ deleted vertices, the diameter of $\mathcal{G}^{>R}$ can change by a factor of at most $R$. This implies the tightness of the diameter for the largest connected components of $\mathcal{G}_n(p_c(\lambda))$ for $\lambda > 1$. Finally the proof of Theorem 6.1.2 follows by invoking Proposition 6.2.4 again. $\qquad\square$

### 6.2.4   Near-critical behavior

Finally we consider the near-critical behavior for $\mathrm{CM}_n(\mathbf{d}, p)$ in this section. The analysis for the barely subcritical and supercritical regimes are given separately below.

**Barely-subcritical regime**

In this section, we analyze the barely-subcritical regime ($p_n \ll p_c$) for percolation and complete the proof of Theorem 6.1.3. Recall the exploration process from Algorithm 6.3 on the graph $\mathcal{G}_n(p_n)$, starting with vertex $j$. Let $\mathscr{C}(j, p_n)$ denote the connected component in $\mathcal{G}_n(p_n)$ containing vertex $j$. We will use the same notation for the quantities defined in Section 6.2.3, but the reader should keep in mind that we now deal with different $p_n$ values. We avoid augmenting $p_n$ in the notation for the sake of simplicity. Let $\mathcal{N}_1$ denote the 1-neighborhood of $j$ in $\mathcal{G}_n(p_n)$ and define $\mathcal{D}(j)$ to be the number of half-edges at vertices incident to $\mathcal{N}_1$ which are not paired with $j$. Define the exploration process $\mathbf{S}_n^j$ similar to (6.14) but starting with $\mathcal{D}(j)$ half-edges as

given below:

$$S_n^j(0) = \mathcal{D}(j), \quad S_n^j(l) = \mathcal{D}(j) + \sum_{i \notin \mathcal{N}_1} \tilde{d}_i \mathcal{I}_i^n(l) - 2l.$$

Thus the exploration process starts from $\tilde{d}_j$ now. Consider the re-scaled process $\bar{S}_n^j$ defined as $\bar{S}_n^j(t) = (n^\alpha p_n)^{-1} S_n^j(\lfloor t n^\alpha p_n \rfloor)$. Then,

$$\bar{S}_n^j(t) = (n^\alpha p_n)^{-1} \mathcal{D}(j) + (n^\alpha p_n)^{-1} \sum_{i \notin \mathcal{N}_1} \tilde{d}_i \mathcal{I}_i^n(t n^\alpha p_n) - 2t + o_\mathbb{P}(1).$$

Recall that $\tilde{\mathbb{E}}$ is the conditional expectation conditionally on $(\tilde{d}_i)_{i \in [n]}$. Now, since the vertices are explored in a size-biased manner with the sizes being $(\tilde{d}_i / \tilde{\ell}_n)_{i \in [n]}$, for any $t \geqslant 0$,

$$\tilde{\mathbb{E}}\left[ \frac{1}{n^\alpha p_n} \sum_{i \notin \mathcal{N}_1} \tilde{d}_i \mathcal{I}_i^n\left(\lfloor t n^\alpha p_n \rfloor\right) \right] \leqslant \frac{t n^\alpha p_n}{n^\alpha p_n \tilde{\ell}_n} \sum_{i \in [n]} \tilde{d}_i^2 = o_\mathbb{P}(1),$$

where the last step follows from Lemma 6.2.5. Moreover, $\mathcal{N}_1 = \tilde{d}_j - 2L_j$, where $L_j$ denotes the number of self-loops associated to vertex $j$ in $\mathcal{G}_n(p_n)$. Since $\tilde{\mathbb{E}}[L_j] \leqslant \tilde{d}_j^2 / \tilde{\ell}_n = o_\mathbb{P}(1)$, it follows that $\tilde{\mathbb{E}}[\mathcal{N}_1] = \tilde{d}_j(1 + o_\mathbb{P}(1))$. Further, $\tilde{\mathbb{E}}[\mathcal{D}(j)] \leqslant \tilde{v}_n \tilde{\mathbb{E}}[|\mathcal{N}_1|] = o_\mathbb{P}(n^\alpha p_n)$. Consequently, $\bar{S}_n^j \xrightarrow{\mathbb{P}} 0$, and therefore, the number of edges in $\mathscr{C}(j, p_n) \setminus \mathcal{N}_1 = o_\mathbb{P}(1)$. Since the number of vertices in $\mathscr{C}(j, p_n) \setminus \mathcal{N}_1$ is at most the number of edges $+1$, this yields that $|\mathscr{C}(j, p_n) \setminus \mathcal{N}_1| = o_\mathbb{P}(n^\alpha p_n)$. Also, one can use Lemma 6.2.5 to show that $\mathrm{Var}(\mathcal{N}_1 \mid \tilde{d}) = o_\mathbb{P}(n^{2\alpha} p_n^2)$, which yields $\mathcal{N}_1 = \tilde{d}_j(1 + o_\mathbb{P}(1))$. Thus,

$$(n^\alpha p_n)^{-1} |\mathscr{C}(j, p_n)| \xrightarrow{\mathbb{P}} \theta_j.$$

To conclude Theorem 6.1.3, it remains to prove that for each fixed $j \geqslant 1$,

$$|\mathscr{C}(j, p_n)| = |\mathscr{C}_{(j)}(p_n)|, \text{ with high probability.} \tag{6.23}$$

For that, we show that the rescaled vector of ordered component sizes converges in $\ell_\downarrow^2$. It is enough to show that for any $\varepsilon > 0$

$$\lim_{K \to \infty} \limsup_{n \to \infty} \mathbb{P}\left( \sum_{i > K} |\mathscr{C}_{(i)}(p_n)|^2 > \varepsilon n^{2\alpha} p_n^2 \right) = 0.$$

This can be concluded using identical arguments as Proposition 6.2.8 above. Now, (6.23) follows using [31, Lemma 7.6]. The proof of Theorem 6.1.3 is now complete.

**Barely-supercritical regime**

In this section, we provide the proof of Theorem 6.1.4 by leveraging techniques from [101, 114]. Using Proposition 6.2.4, it is enough to prove Theorem 6.1.4 for the graph $\mathcal{G}_n(p_n)$ generated by Algorithm 6.1. Let $\tilde{d}$ denote the degree sequence obtained after performing Algorithm 6.1 (S1). Thus, $\mathcal{G}_n(p_n)$ is distributed as $\mathrm{CM}_n(\tilde{d})$. We will verify Assumptions (B1)–(B8) from [101] on the graph $\mathcal{G}_n(p_n)$, which allows us to conclude Theorem 6.1.4 from [101, Theorem 5.3]. Consider the following exploration process on $\mathcal{G}_n(p_n)$ from [101, Section 5.1]:

**Algorithm 6.4.** (S0) Associate an independent Exponential(1) clock $\xi_e$ to each half-edge $e$. Any half-edge can be in one of the *states* among sleeping, active, and dead. Initially at time 0, all the half-edges are sleeping. Whenever the set of active half-edges is empty, select a sleeping half-edge $e$ uniformly at random among all sleeping half-edges and declare it to be active. If $e$ is incident to $v$, declare all the other half-edges of $v$ to be active as well. The process stops when there is no sleeping half-edge left; the remaining sleeping vertices are all isolated and we have explored all other components.

(S1) Pick an active half-edge (which one does not matter) and kill it, i.e., change its status to dead.

(S2) Wait until the next half-edge dies (spontaneously). This half-edge is paired to the one killed in the previous step (S1) to form an edge of the graph. If the vertex it belongs to is sleeping, then we declare this vertex awake and all of its other half-edges active. Repeat from (S1) if there is any active half-edge; otherwise from (S0).

Denote the number of living half-edges upto time t by $L_n(t)$. Let $V_{n,k}$ and $\tilde{V}_{n,k}(t)$ respectively denote the number of sleeping vertices of degree $k$ such that all the $k$ associated exponential clocks ring after time $t$. Define

$$\tilde{V}_n(t) = \sum_{k=0}^{\infty} \tilde{V}_{n,k}(t), \quad \tilde{S}_n(t) = \sum_{k=0}^{\infty} k\tilde{V}_{n,k}(t), \quad \tilde{A}_n(t) = L_n(t) - \tilde{S}_n(t).$$

We show that Assumptions (B1)–(B8) from [101] holds with

$$\zeta = (\kappa/2)^{\frac{1}{3-\tau}}, \quad \gamma_n = \alpha_n = p_n^{\frac{\tau-2}{3-\tau}}, \quad \psi(t) = 2t - \kappa t^{\tau-2},$$
$$\hat{g}(t) = \mu\kappa t^{\tau-2}, \quad \hat{h}(t) = \mu t. \tag{6.24}$$

The $\zeta$ in our notation corresponds to $\tau$ in [101, Theorem 5.3], but we have avoided that since $\tau$ denotes the power-law exponent in our case.

**Remark 42.** Notice that the rate at which an edge is created in the system is $\ell_n p_n$, instead of rate $n$, as in the setting of [101]. Thus, in time $\alpha_n$, the total number of explored half-edges in $\ell_n p_n \alpha_n$ which is of the order $n p_n^{1/(3-\tau)}$. For this reason, the largest component size is of the order $n p_n^{1/(3-\tau)}$ in Theorem 6.1.4.

Conditions (B1)–(B4) and (B8) in [101] are straightforward, and are left to the reader. To verify Conditions (B5)–(B7), we first obtain below the asymptotics of the mean-curve and then show that the processes $\tilde{S}_n$, $\tilde{V}_n$, $\tilde{A}_n$ remain uniformly close to their expected curves. These are summarized in the following two propositions:

**Proposition 6.2.14.** *For any fixed* $t \geqslant 0$, *as* $n \to \infty$,

$$\sup_{u \leqslant t} \left| \frac{1}{n \alpha_n p_n} \left( \mathbb{E}[\tilde{S}_n(0) - \mathbb{E}[\tilde{S}_n(\alpha_n t)] \right) - \hat{g}(t) \right| \to 0, \tag{6.25}$$

$$\sup_{u \leqslant t} \left| \frac{1}{n \alpha_n p_n} \left( \mathbb{E}[\tilde{V}_n(0)] - \mathbb{E}[\tilde{V}_n(\alpha_n t)] \right) - \hat{h}(t) \right| \to 0, \tag{6.26}$$

$$\sup_{u \leqslant t} \left| \frac{1}{n \alpha_n p_n} \mathbb{E}[\tilde{A}_n(\alpha_n t)] - \psi(t) \right| \to 0.$$

**Proposition 6.2.15.** *For any fixed* $t \geqslant 0$, *as* $n \to \infty$, *all the terms* $\sup_{u \leqslant t} |\tilde{S}_n(\alpha_n t) - \mathbb{E}[\tilde{S}_n(\alpha_n t)]|$, $\sup_{u \leqslant t} |\tilde{V}_n(\alpha_n t) - \mathbb{E}[\tilde{V}_n(\alpha_n t)]|$, *and* $\sup_{u \leqslant t} |\tilde{A}_n(\alpha_n t) - \mathbb{E}[\tilde{A}_n(\alpha_n t)]|$ *are* $o(n \alpha_n p_n)$.

To prove Propositions 6.2.14 and 6.2.15, we make crucial use of the following lemma:

**Lemma 6.2.16.** *For any* $t > 0$, *as* $n \to \infty$,

$$\mathbb{E}\left[ \sum_{i \in [n]} \tilde{d}_i e^{-t \alpha_n \tilde{d}_i} \right] = (1 + o(1)) p_n \sum_{i \in [n]} d_i e^{-t \alpha_n p_n d_i},$$

$$\mathbb{E}\left[ \sum_{i \in [n]} e^{-t \alpha_n \tilde{d}_i} \right] = (1 + o(1)) \sum_{i \in [n]} e^{-t \alpha_n p_n d_i}.$$

*Proof.* Note that if $X \sim \text{Bin}(m, p)$, then

$$\mathbb{E}\left[ X e^{-sX} \right] = mp e^{-s} (1 - p + p e^{-s})^{m-1}.$$

267

Putting $m = d_i$, $p = p_n$, and $s = t\alpha_n$, it follows that

$$\mathbb{E}\big[\tilde{d}_i e^{-t\alpha_n \tilde{d}_i}\big] = d_i p_n e^{-t\alpha_n}\left(1 - p_n\left(1 - e^{-t\alpha_n}\right)\right)^{d_i-1}$$
$$= (1 + o(1))d_i p_n (1 - p_n t\alpha_n)^{d_i}$$
$$= (1 + o(1))d_i p_n e^{-t\alpha_n p_n d_i}.$$

Thus the proof follows. □

*Proof of Proposition 6.2.14.* Note that by Lemma 6.2.16

$$\mathbb{E}\big[\tilde{S}_n(t)\big] = \mathbb{E}\Bigg[\sum_{i\in[n]} \tilde{d}_i e^{-t\alpha_n \tilde{d}_i}\Bigg] = (1 + o(1))\ell_n p_n \mathbb{E}\big[e^{-t\alpha_n p_n D_n^*}\big],$$

$$\mathbb{E}\big[\tilde{V}_n(t)\big] = \mathbb{E}\Bigg[\sum_{i\in[n]} e^{-t\alpha_n \tilde{d}_i}\Bigg] = (1 + o(1))\ell_n \mathbb{E}\big[e^{-t\alpha_n p_n D_n}\big],$$

where $D_n^*$ has a size-biased distribution with the sizes being $(d_i/\ell_n)_{i\in[n]}$, and $D_n$ is the degree of a vertex chosen uniformly at random from $[n]$. By the convergence of $\mathbb{E}[D_n]$ in Assumption 6.1,

$$\mathbb{E}\big[1 - e^{-t\alpha_n p_n D_n}\big] = (1 + o(1))t\alpha_n p_n \mathbb{E}[D_n],$$

by noting that $1 - e^{-x} = x(1 + o(1))$ as $x \to 0$. Further, by using (6.4),

$$\mathbb{E}\big[1 - e^{-t\alpha_n p_n D_n^*}\big] = \kappa\alpha_n(t^{\tau-2} + o(1)). \tag{6.27}$$

Thus, (6.25) and (6.26) follow. Moreover, $L_n(t)$ is a pure death process, where $L(0) = \sum_{i\in[n]} \tilde{d}_i$, and the jumps occur at rate $L_n(t)$, and at each jump $L_n(t)$ decreases by 2. Therefore, $\mathbb{E}[L_n(t)] = \mathbb{E}[L(0)]e^{-2t}$ and consequently,

$$\mathbb{E}[\tilde{A}(\alpha_n t)] = \ell_n p_n\big(e^{-2\alpha_n t} - \mathbb{E}\big[e^{-tp_n\alpha_n D_n^*}\big]\big) + o(\ell_n\alpha_n p_n)$$
$$= \ell_n p_n \alpha_n(2t - \kappa t^{\tau-2}) + +o(\ell_n\alpha_n p_n).$$

Thus the proof follows. □

*Proof of Proposition 6.2.15.* Let us consider $\tilde{S}_n$ only and the other inequalities follow using identical arguments. We will use precise bounds in [101, Lemma 5.13]. In fact, using the fact that $1 - e^{-x} \geqslant (1 \wedge x)/3$, it follows that

$$\mathbb{E}\Big[\sup_{u\leqslant t\alpha_n} |\tilde{S}_n(u) - \mathbb{E}[\tilde{S}_n(u)]|\Big] \leqslant C\mathbb{E}\Bigg[\sum_{i\in[n]} \tilde{d}_i^2(1 - e^{-t\alpha_n \tilde{d}_i})\Bigg]. \tag{6.28}$$

Now, using standard concentration inequalities for tails of binomial distributions [115, Theorem 2.1], for any $i \in [n]$,

$$\mathbb{P}(\tilde{d}_i > 2d_1 p_n) \leqslant Ce^{-Cd_1 p_n} = Ce^{-Cn^\rho \lambda_n},$$

and therefore $\max_{i \in [n]} \tilde{d}_i \leqslant 2d_1 p_n$, a.s. $\mathbb{P}_p$. Now, using (6.24), the bounds (6.27) and (6.28) yield

$$\frac{1}{(\ell_n p_n \alpha_n)^2} \mathbb{E}\Big[ \sup_{u \leqslant t\alpha_n} |\tilde{S}_n(u) - \mathbb{E}[\tilde{S}_n(u)]| \Big]$$
$$\leqslant \frac{C2d_1 p_n \ell_n p_n \alpha_n}{(\ell_n p_n \alpha_n)^2} = \frac{C}{\lambda_n^{(\tau-2)/(3-\tau)}} \to 0,$$

since $\lambda_n \to 0$, as $n \to \infty$. Thus the proof follows.    □

*Proof of Theorem 6.1.4.* The proof follows by applying [101, Theorem 5.3]. Proposition 6.2.14 verifies conditions (B5)–(B7) in [101], and the rest of the conditions are straightforward to verify.    □

## 6.3    Generalized random graphs: Proofs

In this section, we prove our results related to critical percolation on $\mathrm{GRG}_n(w)$. In Section 6.3.1, we set up the fundamental technical ingredients required for the proof. In Section 6.3.2, we use the first-moment method to identify the primary contributions on the total weight of the components. The connectivity structure between the hubs is described in detail in Section 6.3.3, which allows us to deduce the component sizes of components containing hubs in Section 6.3.4. Finally, Theorem 6.1.5 is proved in Section 6.3.5.

### 6.3.1    Key Ingredients

In this section, we provide the two key ingredients that will play a pivotal role in the proof of Theorem 6.1.5. The first one, stated in Lemma 6.3.1, provides the estimates for different moments of the $w$. Next, in Lemma 6.3.2, we estimate the probability of two vertices being connected via another vertex. The later result forms the conceptual bedrock of our analysis (see Section 6.3.3). For example, one can set up path-counting techniques in Corollary 6.3.3 using Lemma 6.3.2. We use a generic notation C to denote a positive constant. Also we write $a_n \sim b_n$ to mean that $a_n/b_n \to 1$, as $n \to \infty$.

**Lemma 6.3.1** (Moment estimates). *Under  Assumption 6.2, there exists a constant $C_F > 0$ (depending only on F) such that, for all $i \in [n]$,*

$$w_i = \left(\frac{c_F n}{i}\right)^\alpha. \tag{6.29}$$

*For any $a > 0, b \leqslant \alpha$,*

$$
\begin{aligned}
\#\{r : w_r > a\ell_n/w_j\} &\sim a^{-(\tau-1)}n\left(\frac{w_j}{\ell_n}\right)^{\tau-1}, \\
\sum_{k:w_k \leqslant a\ell_n/w_j} w_k^2 &\sim Ca^{3-\tau}n\left(\frac{\ell_n}{w_j}\right)^{3-\tau}, \\
\sum_{w_k > a\ell_n/w_j} w_k &\sim Ca^{-(\tau-2)}n^{3-\tau}w_j^{\tau-2}, \\
\sum_{k:w_k \leqslant an^b} w_k^{\tau-2} &\sim Cn^\rho(an^b)^\alpha,
\end{aligned}
\tag{6.30}
$$

*where $C > 0$ is considered as a generic notation for a constant.*

*Proof.* (6.29) follows directly from Assumption 6.2. Next, note that

$$\left(\frac{n}{r}\right)^\alpha \leqslant a\frac{\ell_n}{w_j} \iff r \geqslant a^{-(\tau-1)}n\left(\frac{w_j}{\ell_n}\right)^{\tau-1}.$$

Thus, (6.30) follows by noting that

$$\sum_{r:w_r \leqslant a\ell_n/w_j} w_r^2 = n^{2\alpha} \sum_{r \geqslant a^{-(\tau-1)}n(w_j/\ell_n)^{\tau-1}} r^{-2\alpha} \sim Ca^{3-\tau}n\left(\frac{\ell_n}{w_j}\right)^{3-\tau},$$

and

$$
\begin{aligned}
\sum_{w_k > a\ell_n/w_j} w_k &\sim Cn^\alpha \sum_{k \leqslant Ca^{-(\tau-1)}(w_j/\ell_n)^{\tau-1}} k^{-\alpha} \\
&\sim Cn^\alpha\left(a^{-(\tau-1)}n\left(\frac{w_j}{\ell_n}\right)^{\tau-1}\right)^\rho \\
&\sim Ca^{-(\tau-2)}n^{3-\tau}w_j^{\tau-2}.
\end{aligned}
$$

The last expression is also similar. $\qquad\square$

**Lemma 6.3.2** (Two-hop connection probabilities). *There exists an absolute constant $C > 0$ such that for all $n \geqslant 1$,*

$$p_{ij}(2) := p_c^2 \sum_{v \in [n]} \frac{w_i w_v^2 w_j}{(\ell_n + w_i w_v)(\ell_n + w_j w_v)} \leqslant \frac{C\lambda^2}{(i \wedge j)^{1-\alpha}(i \vee j)^\alpha}, \quad \forall i, j \in [n].$$

*Proof.* Without loss of generality, we assume that $w_i \geqslant w_j$, i.e., $i \leqslant j$. Let us split the sum in three parts with $\{v : w_i w_v \leqslant \ell_n\}$, $\{v : w_j w_v \leqslant \ell_n < w_i w_v\}$, and $\{v : w_j w_v > \ell_n\}$, and denote them by (I), (II) and (III) respectively. Note that using Lemma 6.3.1,

$$(\text{I}) \leqslant \frac{p_c^2 w_i w_j}{\ell_n} \sum_{v : w_v \leqslant \ell_n / w_i} w_v^2 = C\lambda^2 n^{-(3-\tau)} \frac{w_i w_j}{\ell_n^2} n \frac{\ell_n^{3-\tau}}{w_i^{3-\tau}}$$

$$\leqslant C\lambda^2 \frac{w_i^{\tau-2} w_j}{\ell_n} \leqslant \frac{C\lambda^2}{i^{1-\alpha} j^\alpha},$$

$$(\text{II}) \leqslant \frac{p_c^2 w_j}{\ell_n} \sum_{v : \frac{\ell_n}{w_i} < w_v \leqslant \frac{\ell_n}{w_j}} w_v \leqslant C\lambda^2 n^{-(3-\tau)} \frac{w_j}{\ell_n} \ell_n^{3-\tau} w_i^{\tau-2}$$

$$\leqslant C\lambda^2 \frac{w_i^{\tau-2} w_j}{\ell_n} \leqslant \frac{C\lambda^2}{i^{1-\alpha} j^\alpha},$$

$$(\text{III}) \leqslant p_c^2 \#\{w_v > \ell_n / w_j\} = p_c^2 n \frac{w_j^{\tau-1}}{\ell_n^{\tau-1}} \leqslant C\lambda^2 \frac{w_i^{\tau-2} w_j}{\ell_n} \leqslant \frac{C\lambda^2}{i^{1-\alpha} j^\alpha}.$$

Thus the proof follows. □

**Corollary 6.3.3** (Path counting estimate). *Let $f_k(i,j)$ denote the probability that there exists a path of length $k$ in $\mathrm{GRG}_n(\boldsymbol{w})$ from $i$ to $j$. For all $1 - \alpha < b < \alpha$, there exists a constant $c_0 > 0$ such that for all $n \geqslant 1$,*

$$f_{2k}(i,j) \leqslant \frac{(c_0 \lambda^2)^k}{(i \wedge j)^b (i \vee j)^{1-b}}. \tag{6.31}$$

*Proof.* Note that

$$f_{2k}(i,j) \leqslant \sum_{\substack{(v_r)_{r=0}^{2k} : v_r's \text{ distinct} \\ v_0 = i, v_{2k} = j}} p_c^{2k} \prod_{r=0}^{2k} \frac{w_{v_r} w_{v_{r+1}}}{\ell_n + w_{v_r} w_{v_{r+1}}},$$

and the proof follows directly from [73, Lemma 2.4] using Lemma 6.3.2.   □

### 6.3.2   Negligible contributions on the total weight

Suppose that $\mathscr{C}(i)$ denotes the component in $\mathrm{GRG}_n(\boldsymbol{w}, p_c(\lambda))$ containing vertex $i$ and $W_k(i) = \sum_{v \in \mathscr{C}(i), d(v,i)=k} w_v$, where $d(\cdot, \cdot)$ is used as a notation for graph distance (the number of edges on the shortest path) throughout. In this section, we identify the terms that have negligible contributions to $W(i)$. The next proposition states that the total contribution on the total weight coming from vertices in the odd neighborhood is small. Moreover, the total weight outside a large but finite neighborhood of $i$ is also negligible.

**Proposition 6.3.4.** *Suppose that* $\lambda < c_0^{-1/2}$, *where* $c_0$ *is defined in Corollary 6.3.3. For any fixed* $i \geqslant 1$ *and* $\varepsilon > 0$,

$$\lim_{K \to \infty} \limsup_{n \to \infty} \mathbb{P}\left( \sum_{k > K} W_{2k}(i) > \varepsilon n^\alpha \right) = 0,$$

$$\lim_{n \to \infty} \mathbb{P}\left( \sum_{k=0}^\infty W_{2k+1}(i) > \varepsilon n^\alpha \right) = 0. \tag{6.32}$$

*Proof.* Recall the definition of $f_k(i,j)$ from Corollary 6.3.3 and note that $c_0 \lambda^2 < 1$. Therefore, using Corollary 6.3.3,

$$\mathbb{E}[W_{2k}(i)] \leqslant \sum_{j \in [n]} w_j f_{2k}(i,j)$$

$$\leqslant n^\alpha (c_0 \lambda^2)^k \left[ \sum_{j < i} \frac{1}{j^\alpha} \frac{1}{j^b i^{1-b}} + \sum_{j > i} \frac{1}{j^\alpha} \frac{1}{i^b j^{1-b}} \right] \tag{6.33}$$

$$\leqslant C n^\alpha (c_0 \lambda^2)^k \left[ \frac{1}{i^{1-b}} + \frac{1}{i^\alpha} \right] \leqslant C(c_0 \lambda^2)^k \frac{n^\alpha}{i^{1-b}}$$

$$\leqslant C(c_0 \lambda^2)^k w_i i^{b-(1-\alpha)}.$$

Now, an application of Markov's inequality proves the first part of (6.32). We stress that (6.33) holds uniformly over $i \in [n]$, which we will use in the next part of the proof.

The proof of the second part in (6.32) is complete if we can show that, for any $i \in [n]$,

$$n^{-\alpha} \mathbb{E}\left[ \sum_{k=0}^\infty W_{2k+1}(i) \right] \leqslant \frac{C n^{-\epsilon}}{i^\rho}, \tag{6.34}$$

for absolute constants $C > 0$, $\epsilon > 0$. Firstly, note that for any vertex $i \in [n]$,

$$p_c \sum_{v \in [n]} \frac{w_i w_v^2}{\ell_n + w_i w_v} \leqslant C \lambda w_i^{\tau-2} n^{3-\tau}. \tag{6.35}$$

To see this, let us split the above sum in two parts with $\{v : w_v \leqslant \ell_n / w_i\}$ and $\{v : w_v > \ell_n / w_i\}$, and denote them by (I) and (II) respectively. Then, by Lemma 6.3.1,

$$(I) \leqslant \frac{p_c}{\ell_n} w_i n \frac{\ell_n^{3-\tau}}{w_i^{3-\tau}} \leqslant C \lambda w_i^{\tau-2} n^{(3-\tau)/2},$$

$$(II) \leqslant p_c \sum_{v : w_v > \ell_n / w_i} w_v \leqslant C \lambda w_i^{\tau-2} n^{(3-\tau)/2},$$

and (6.35) follows. Now, we will use the precise bound in (6.33). Choose $\epsilon = \frac{1}{2}(\frac{3-\tau}{\tau-1} - \frac{3-\tau}{2})$ and $b$ such that $b - (1-\alpha) = \epsilon$. Thus,

$$n^{-\alpha}\mathbb{E}[W_{2k+1}(i)] \leqslant n^{-\alpha} \sum_{v\in[n]} \mathbb{P}(i \text{ and } v \text{ create an edge})\mathbb{E}[W_{2k}(v)]$$

$$\leqslant n^{-\alpha}n^\epsilon \sum_{v\in[n]} p_c \frac{w_i w_v}{\ell_n + w_i w_v} C(c_0\lambda^2)^k w_v$$

$$= \frac{C(c_0\lambda^2)^k n^\epsilon}{n^\alpha}\left(p_c \sum_{v\in[n]} \frac{w_i w_v^2}{\ell_n + w_i w_v}\right)$$

$$\leqslant \frac{C(c_0\lambda^2)^k}{i^\rho}n^\epsilon n^{\frac{3-\tau}{2} - \frac{3-\tau}{\tau-1}} = \frac{C(c_0\lambda^2)^k n^{-\epsilon}}{i^\rho},$$

where the last-but-one step follows from (6.35), and in the final step we have used the choice of $\epsilon > 0$. The proof of (6.34) now follows using the fact that $c_0\lambda^2 < 1$, which also concludes the proof of Proposition 6.3.4. □

We will be interested in obtaining the limit of $\sum_{k=1}^\infty W_{2k}(i)$. Using Proposition 6.3.4, it is enough to find the limit, as $n \to \infty$, of the quantity $\sum_{k=1}^K W_{2k}(i)$ for each fixed $K \geqslant 1$. The next proposition states that for each fixed $k \geqslant 1$, the primary contribution to $W_{2k}(i)$ arises only due to the hubs. For $\delta > 0$, define $V_L(\delta) := \{v : w_v > \delta n^\alpha\}$, and $W_k(i, \delta) := \sum_{v\notin V_L(\delta), d(v,i)=k} w_v$.

**Proposition 6.3.5.** *For any fixed $i \geqslant 1$, $K \geqslant 1$, and $\varepsilon > 0$,*

$$\lim_{\delta\to 0} \limsup_{n\to\infty} \mathbb{P}\left(\sum_{k=1}^K W_{2k}(i, \delta) > \varepsilon n^\alpha\right) = 0.$$

*Proof.* Suppose that we choose $\delta > 0$ to be so small that $i \in V_L(\delta)$. Therefore, $v \notin V_L$ implies that $v > i$. Using Corollary 6.3.3, it follows that

$$\mathbb{E}[W_{2k}(i, \delta)] \leqslant \sum_{v\notin V_L(\delta)} \frac{n^\alpha}{v^\alpha} \frac{(c_0\lambda^2)^k}{i^b v^{1-b}}$$

$$\leqslant \frac{Cn^\alpha}{i^b} \sum_{v>\delta^{-(\tau-1)}} \frac{1}{v^{1+\alpha-b}} \leqslant \frac{Cn^\alpha \delta^{(\tau-1)(\alpha-b)}}{i^b}.$$

Therefore,

$$n^{-\alpha}\mathbb{E}\left[\sum_{k=1}^K W_{2k}(i, \delta)\right] \leqslant CKi^{-b}\delta^{(\tau-1)(\alpha-b)}. \tag{6.36}$$

Now, an application of Markov's inequality completes the proof. □

### 6.3.3    Total weight of components containing hubs

To simplify writing we will always assume that $c_F = 1$ without loss of generality. Recall the definition of the graph $G_\infty(\lambda)$ from Section 6.1.2. In the rest of this section, we write $\theta_i = i^{-\alpha}$. Consider the following objects defined on the graph $G_\infty$: $W_k^\infty(i) = \sum_{j \in C(i),\, d(i,j)=k} \theta_j$, where $C(i)$ is defined as the component in $G_\infty(\lambda)$ containing vertex $i$, and $W_{\leqslant K}^\infty(i) = \sum_{k=1}^K W_k^\infty(i)$, and $W^\infty(i) = \sum_{k=1}^\infty W_k^\infty(i)$. The main result of this subsection is the following:

**Theorem 6.3.6.** *Suppose that $\lambda < c_0^{-1/2}$, where $c_0$ is defined in Corollary 6.3.3. For each fixed $i \geqslant 1$, as $n \to \infty$, $n^{-\alpha} W(i) \xrightarrow{d} W^\infty(i)$.*

The key ingredient in the proof is the proposition below. We immediately give the proof of Theorem 6.3.6 after stating the proposition, and devote the rest of this section to the proof of Proposition 6.3.7:

**Proposition 6.3.7.** *For each fixed $i \geqslant 1$ and $K \geqslant 1$, as $n \to \infty$,*

$$n^{-\alpha} \sum_{k=1}^K W_{2k}(i) \xrightarrow{d} W_{\leqslant K}^\infty(i).$$

*Proof of Theorem 6.3.6.* Propositions 6.3.4, 6.3.5, and 6.3.7 together directly concludes the proof. □

Throughout this subsection, we will use the notation $V_L := \{i : w_i > \delta n^\alpha\}$, $V_{SS} = \{i : w_i < \delta n^\rho\}$, $V_{SI} = \{i : \delta n^\rho \leqslant w_i \leqslant \delta^{-1} n^\rho\}$ and $V_{SL} = \{i : w_i > \delta^{-1} n^\rho\}$. We have tacitly avoided augmenting $\delta > 0$ in the notation to simplify notation. Note that for $v \in V_{SI}$ and $i \in V_L$, $w_i w_v = \Theta(\ell_n)$. Consider the following multigraph $G_{n,\delta}$ on the vertex set $V_L$, where the number of edges $X_{ij}$ between $i$ and $j$ is number of distinct $v \in [n]$ such that both $(i,v)$ and $(v,j)$ are edges of $GRG_n(w, p_c)$. Note that, for any $i \neq j$,

$$X_{ij} = \sum_{v \neq i,j} \mathrm{Ber}\left(\frac{w_i w_j w_v^2 p_c^2}{(\ell_n + w_i w_v)(\ell_n + w_j w_v)}\right), \tag{6.37}$$

with the different Bernoulli random variables in the sum (6.37) being independent. We can split the above sum in three parts with $v \in V_{SS}$, $v \in V_{SI}$, and $v \in V_{SL}$ and denote them by (I), (II), and (III) respectively. Now, using Lemma 6.3.1,

$$\mathbb{E}[(I)] \leqslant \frac{w_i w_j p_c^2}{\ell_n^2} \sum_{v:w_v < \delta n^\rho} w_v^2 \tag{6.38}$$

$$\leqslant C_0 \theta_i^2 \delta^{3-\tau} n^{2\alpha-2+1+(3-\tau)\rho} p_c^2 \leqslant C_1 \delta^{3-\tau},$$

$$\mathbb{E}[(III)] \leqslant p_c^2 \times \#\{v : w_v > \delta^{-1} n^\rho\} = C\delta^{\tau-1}. \tag{6.39}$$

Using the above and Markov's inequality, it follows that

$$X_{ij} = \sum_{v:w_v \in [\delta n^\rho, \delta^{-1} n^\rho]} \mathrm{Ber}\left(\frac{w_i w_j w_v^2 p_c^2}{(\ell_n + w_i w_v)(\ell_n + w_j w_v)}\right) + E(\delta, n)$$

$$= X_{ij}(\delta) + E(\delta, n),$$

where for any $\varepsilon > 0$

$$\lim_{\delta \to 0} \limsup_{n \to \infty} \mathbb{P}(E(\delta, n) > \varepsilon) = 0. \tag{6.40}$$

Define

$$\lambda_{ij}(\delta) = \sum_{v:w_v \in [\delta n^\rho, \delta^{-1} n^\rho]} \frac{w_i w_j w_v^2 p_c^2}{(\ell_n + w_i w_v)(\ell_n + w_j w_v)}.$$

The proof of Proposition 6.3.7 is decomposed into three key lemmas below. After stating these lemmas, we first prove Proposition 6.3.7, and subsequently prove the lemmas.

**Lemma 6.3.8.** *For any $\delta \in (0, 1)$, and $i, j \geqslant 1$*

$$\lim_{n \to \infty} d_{\mathrm{TV}}\big(X_{ij}(\delta), \mathrm{Poi}(\lambda_{ij}(\delta))\big) = 0.$$

**Lemma 6.3.9.** *For any fixed $i, j \geqslant 1$*

$$\lim_{\delta \to 0} \lim_{n \to \infty} \lambda_{ij}(\delta) = \lambda^2 \int_0^\infty \frac{\theta_i \theta_j x^{-2\alpha}}{(\mu + \theta_i x^{-\alpha})(\mu + \theta_j x^{-\alpha})} dx.$$

**Lemma 6.3.10.** *For any $\delta \in (0, 1)$, the collection of random variables $(X_{ij}(\delta))_{i,j \in V_L}$ is asymptotically independent.*

*Proof of Proposition 6.3.7.* The proof follows directly from Lemmas 6.3.8, 6.3.9, and 6.3.10. $\square$

*Proof of Lemma 6.3.8.* Using standard inequalities from Stein's method [98, Theorem 2.10], it follows that, as $n \to \infty$,

$$d_{\mathrm{TV}}\big(X_{ij}(\delta), \mathrm{Poi}(\lambda_{ij}(\delta))\big) \leqslant \sum_{v:w_v \in [\delta n^\rho, \delta^{-1} n^\rho]} \left(\frac{w_i w_j w_v^2 p_c^2}{(\ell_n + w_i w_v)(\ell_n + w_j w_v)}\right)^2$$

$$\leqslant C n^{4\alpha-4} p_c^4 \sum_{v:w_v \in [\delta n^\rho, \delta^{-1} n^\rho]} w_v^4 \leqslant \frac{C}{\delta^2} n^{2\alpha-2} p_c^4 \sum_{v:w_v \in [\delta n^\rho, \delta^{-1} n^\rho]} w_v^2$$

$$= \frac{C}{\delta^2} n^{2\alpha-2} n^{-2(3-\tau)} n^{1+(3-\tau)\rho} = \frac{C n^{-(3-\tau)}}{\delta^2} \to 0,$$

and the proof follows. $\square$

*Proof of Lemma 6.3.9.* Observe that

$$\lambda_{ij}(\delta) = \sum_{v: w_v \in [\delta n^\rho, \delta^{-1} n^\rho]} \frac{w_i w_j w_v^2 p_c^2}{(\ell_n + w_i w_v)(\ell_n + w_j w_v)}$$

$$= p_c^2 n^{4\alpha - 2} \sum_{k=\delta^{\tau-1} n^{3-\tau}}^{\delta^{-(\tau-1)} n^{3-\tau}} \frac{\theta_i \theta_j k^{-2\alpha}}{(\mu + \theta_i n^{-\frac{3-\tau}{\tau-1}} k^{-\alpha})(\mu + \theta_j n^{-\frac{3-\tau}{\tau-1}} k^{-\alpha})}$$

$$= p_c^2 n^{4\alpha - 2} \int_{\delta^{\tau-1} n^{3-\tau}}^{\delta^{-(\tau-1)} n^{3-\tau}} \frac{\theta_i \theta_j k^{-2\alpha}}{(\mu + \theta_i n^{-\frac{3-\tau}{\tau-1}} k^{-\alpha})(\mu + \theta_j n^{-\frac{3-\tau}{\tau-1}} k^{-\alpha})} dk$$

$$= (1 + o(1))\lambda^2 \int_{\delta^{\tau-1}}^{\delta^{-(\tau-1)}} \frac{\theta_i \theta_j x^{-2\alpha}}{(\mu + \theta_i x^{-\alpha})(\mu + \theta_j x^{-\alpha})} dx,$$

and the proof follows.  □

*Proof of Lemma 6.3.10.* Note that for pairs $(i, j)$, and $(k, l)$ with $\{i, j\} \cap \{k, l\} = \varnothing$, $X_{ij}$ and $X_{kl}$ are independent due to the independence of the occupancy of edges in $GRG_n(w, p_c)$. The only dependence between $X_{ij}$ and $X_{ik}$ arises due to connections $(i, v)$, $(v, j)$ and $(v, k)$. Thus, the lemma is proved if we can show that the above does not arise with high probability. Note that

$$\mathbb{P}(i, j, k \in V_L : \exists v \in [n] \text{ such that } (i, v), (v, j), (v, k) \text{ are edges in } GRG_n(w, p_c))$$

$$\leqslant \sum_{i,j,k \in V_L} \sum_{v \in [n]} p_c^3 \frac{w_i w_v}{\ell_n + w_i w_v} \frac{w_j w_v}{\ell_n + w_j w_v} \frac{w_k w_v}{\ell_n + w_k w_v}.$$

Again, let us split the above sum in two pars with $\{v : w_v \leqslant n^\rho\}$ and $\{v : w_v > n^\rho\}$, and denote them by (I) and (II) respectively. Thus, using Lemma 6.3.1,

$$\text{(I)} \leqslant \frac{p_c^3 w_1^3 |V_L|^3}{\ell_n^3} \sum_{v: w_v \leqslant n^\rho} w_v^3 \leqslant \frac{p_c^3 w_1^3 |V_L|^3 n^\rho}{\ell_n^3} \sum_{v: w_v \leqslant n^\rho} w_v^2$$

$$\leqslant C p_c^3 n^{3\alpha - 3 + \rho + 1 + (3-\tau)\rho} = O(p_c),$$

$$\text{(II)} \leqslant |V_L|^3 p_c^3 \#\{v : w_v > n^\rho\} = O(p_c).$$

This completes the proof of Lemma 6.3.10.  □

### 6.3.4   Sizes of components containing hubs

In this section, we consider the asymptotic size of $\mathscr{C}(i)$, the component containing vertex $i$. We will prove the following theorem:

**Theorem 6.3.11.** *For each fixed* $i \geqslant 1$, *as* $n \to \infty$, $(n^\alpha p_c)^{-1}|\mathscr{C}(i)| \overset{d}{\to} W^\infty(i)$.

Lemmas 6.3.12, and 6.3.13 identify the primary contribution to the component sizes. Since the proof of the lemmas are short, they are given immediately. We conclude the section with the proof of Theorem 6.3.11 using these two lemmas. Define $\mathscr{C}_k(i) := \{v \in \mathscr{C}(i) : d(v, i) = k\}$. Thus $\mathscr{C}_k(i)$ denotes the set of vertices at exactly distance $k$ from vertex $i$.

**Lemma 6.3.12.** *Suppose that* $\lambda < c_0^{-1/2}$, *where* $c_0$ *is defined in Corollary 6.3.3. For any fixed* $i \geqslant 1$, *and* $\varepsilon > 0$,

$$\lim_{n \to \infty} \mathbb{P}(\sum_{k=0}^{\infty} |\mathscr{C}_{2k}(i)| > \varepsilon n^\alpha p_c) = 0,$$

$$\lim_{K \to \infty} \limsup_{n \to \infty} \mathbb{P}\left(\sum_{k > K} |\mathscr{C}_{2k+1}(i)| > \varepsilon n^\alpha p_c\right) = 0.$$

*Proof.* Note that

$$\mathbb{E}\left[|\mathscr{C}_{k+1}(i)| \Big| \bigcup_{r=1}^{k} \mathscr{C}_r(i)\right] \leqslant p_c \sum_{v_1 \in \mathscr{C}_k(i)} \sum_{v_2 \in [n]} \frac{w_{v_1} w_{v_2}}{\ell_n + w_{v_1} w_{v_2}} \quad (6.41)$$

$$\leqslant p_c W_k(i),$$

and therefore $\mathbb{E}[|\mathscr{C}_{k+1}(i)|] \leqslant p_c \mathbb{E}[W_k(i)]$. Now the estimates (6.33), (6.34) conclude the proof. $\quad\square$

Let $\mathscr{C}_k'(i) \subset \mathscr{C}_k(i)$ denote the vertices of $\mathscr{C}_k(i)$ that are neighbors of some vertex in $\mathscr{C}_{k-1}(i) \cap V_L$, where $V_L := \{v : w_v > \delta n^\alpha\}$ for some $\delta > 0$. Then the following lemma estimates the contribution to the component size due to the non-hubs at distance $2k + 1$:

**Lemma 6.3.13.** *For each fixed* $i \geqslant 1$, $k \geqslant 1$, *and* $\varepsilon > 0$,

$$\lim_{\delta \to 0} \limsup_{n \to \infty} \mathbb{P}(|\mathscr{C}_{2k+1}(i) \setminus \mathscr{C}_{2k+1}'(i)| > \varepsilon n^\alpha p_c) = 0. \quad (6.42)$$

*Proof.* Using an identical argument as (6.41) yields $\mathbb{E}[|\mathscr{C}_{2k+1}(i) \setminus \mathscr{C}_{2k+1}'(i)|] \leqslant p_c \mathbb{E}[W_{2k}(i, \delta)]$, where $W_k(i, \delta) := \sum_{v \in \mathscr{C}_k(i) \cap V_L^c} w_v$. Now, (6.42) follows from (6.36). $\quad\square$

**Fact 3.** *Given a matrix* $(p_{ij})_{i \in [m], j \in [n]}$, *let* $I_{ij} \sim \text{Ber}(p_{ij})$, *independently. For all* $i \in [m]$, *construct the random set* $V_i := \{j : I_{ij} = 1\}$, *and let* $V = \cup_{i \in [m]} V_i$. *Then* $\text{Var}(|V|) \leqslant \sum_{i \in [m]} \text{Var}(|V_i|)$.

*Proof.* Note that $\mathbb{P}(v_j \in V) = 1 - \prod_{i \in [m]}(1 - p_{ij})$, and the events $\{u \in V\}$ and $\{w \in V\}$ are independent for all $u \neq w$. Further, $\text{Var}(|V_i|) = \sum_{j \in [n]} p_{ij}(1 - p_{ij})$. Therefore,

$$\text{Var}(|V|) = \sum_{j \in [n]} \left(1 - \prod_{i \in [m]}(1 - p_{ij})\right) \prod_{i \in [m]}(1 - p_{ij})$$

$$\leqslant \sum_{j \in [n]} \sum_{i \in [m]} p_{ij} \prod_{i \in [m]}(1 - p_{ij})$$

$$\leqslant \sum_{j \in [n]} \sum_{i \in [m]} p_{ij}(1 - p_{ij}) = \sum_{i \in [m]} \text{Var}(|V_i|),$$

where the third step follows using the fact the $1 - \prod(1 - x_r) \leqslant \sum x_r$, whenever $0 \leqslant x_r \leqslant 1$ for every $r \geqslant 1$. $\qquad\square$

*Proof of Theorem 6.3.11.* Let $F_k$ denote the minimal sigma-algebra with respect to which $\cup_{r=1}^{k} \mathscr{C}_r(i)$ is measurable. Define $W'_k(i) := \sum_{v \in \mathscr{C}_k(i) \cap V_L} w_v$. Using Lemmas 6.3.12, 6.3.13, it is now enough to show that $|\mathscr{C}'_{2k+1}(i)| = p_c W'_{2k}(i)(1 + E(\delta, n))$, where the random variable $E(\delta, n)$ satisfies (6.40). This follows from Chebyshev's inequality if we can show that

$$\mathbb{E}[|\mathscr{C}'_{2k+1}(i)| \| F_{2k}] = p_c W'_{2k}(i)(1 + o(1)), \quad \text{and} \quad \text{Var}\left(|\mathscr{C}'_{2k+1}(i)| \| F_{2k}\right) \leqslant E_n,$$
(6.43)

where $\mathbb{E}[E_n] = o(n^{2\alpha} p_c^2)$. For $v \in [n] \setminus \mathscr{C}_{2k}(i)$, let $I_v$ denote the indicator that there exists $u \in \mathscr{C}_{2k}(i) \cap V_L$ such that $(v, u)$ creates an edge. Thus, for any $v \in [n] \setminus \mathscr{C}_{2k}(i)$

$$\mathbb{P}(I_v = 1 | F_{2k}) = 1 - \prod_{u \in \mathscr{C}_{2k}(i) \cap V_L}\left(1 - \frac{p_c w_u w_v}{\ell_n + w_u w_v}\right).$$

Using inclusion-exclusion with respect to the union of $u \in \mathscr{C}_{2k}(i) \cap V_L$, it now follows that

$$\mathbb{E}[|\mathscr{C}'_{2k+1}(i)| \| F_{2k}]$$

$$\geqslant \sum_{v \notin \mathscr{C}_{2k}(i)} \sum_{u \in \mathscr{C}_{2k}(i) \cap V_L} \frac{p_c w_u w_v}{\ell_n + w_u w_v}$$

$$- \sum_{v \notin \mathscr{C}_{2k}(i)} \sum_{\substack{u_1, u_2 \in \mathscr{C}_{2k}(i) \cap V_L, \\ u_1 < u_2}} \frac{p_c^2 w_{u_1} w_v^2 w_{u_2}}{(\ell_n + w_{u_1} w_v)(\ell_n + w_{u_2} w_v)}.$$

Let us denote the first and second terms above by (I) and (II) respectively. Note that

$$(\text{II}) \leqslant \sum_{u_1, u_2 \in V_L} \sum_{v \in [n]} \frac{p_c^2 w_{u_1} w_v^2 w_{u_2}}{(\ell_n + w_{u_1} w_v)(\ell_n + w_{u_2} w_v)} = O(1) = o(n^{\alpha} p_c),$$

almost surely, where the third equality above follows using (6.38), (6.39) and Lemma 6.3.9. Further, by observing

$$\frac{1}{n^\alpha p_c} \sum_{u \in \mathscr{C}_{2k}(i) \cap V_L} \sum_{v \in \mathscr{C}_{2k}(i)} \frac{p_c w_u w_v}{\ell_n + w_u w_v}$$

$$\leqslant \frac{1}{n^\alpha \ell_n} (W_{2k}(i))^2 = O_\mathbb{P}(n^{\alpha-1}) = o_\mathbb{P}(1),$$

it follows that

$$\mathbb{E}[|\mathscr{C}_{2k+1}(i)||F_{2k}] = \sum_{u \in \mathscr{C}_{2k}(i) \cap V_L} \sum_{v \in [n]} \frac{p_c w_u w_v}{\ell_n + w_u w_v} + o_\mathbb{P}(n^\alpha p_c).$$

Now for $\varepsilon > 0$ (sufficiently small), let us split the above term in two parts with $\{v : w_v \leqslant n^{\rho-\varepsilon}\}$, $\{v : w_v > n^{\rho-\varepsilon}\}$, and call them (I) and (II) respectively. Now, using Lemma 6.3.1,

$$\frac{\text{(II)}}{n^\alpha p_c} \leqslant C|V_L|^2 \frac{n^{1-(\tau-1)\rho+\varepsilon(\tau-1)}}{n^\alpha} \leqslant C|V_L|^2 n^{\rho-\rho(\tau-1)+\varepsilon(\tau-1)} = o(1),$$

almost surely, and

$$\text{(I)} = p_c \sum_{u \in \mathscr{C}_{2k}(i) \cap V_L} \sum_{v : w_v \leqslant n^{\rho-\varepsilon}} \frac{w_u w_v}{\ell_n(1 + o(1))} = p_c W'_{2k}(i)(1 + o(1)).$$

The estimate for the expectation term in (6.43) now follows. For $u \in \mathscr{C}_{2k}(i)$, let $N_u$ denote the number of neighbors of $u$ in $\mathscr{C}_{2k+1}(i)$. For the variance term, Fact 3 implies that

$$\text{Var}\left(|\mathscr{C}'_{2k+1}(i)||F_{2k}\right) \leqslant \sum_{u \in \mathscr{C}_{2k}(i) \cap V_L} \text{Var}(N_u)$$

$$\leqslant \sum_{u \in \mathscr{C}_{2k}(i) \cap V_L} \sum_{v \in [n]} \frac{p_c w_u w_v}{\ell_n + w_u w_v} \leqslant p_c W'_{2k}(i),$$

and the required estimate in (6.43) follows using Theorem 6.3.6. □

### 6.3.5  Proof of Theorem 6.1.5

To conclude Theorem 6.1.5 using Theorems 6.3.6, 6.3.11, it is enough to show that $(n^{-\alpha}(W_{(i)})_{i \geqslant 1})_{n \geqslant 1}$ and $((n^\alpha p_c)^{-1}(|\mathscr{C}_{(i)}(p_c)|)_{i \geqslant 1})_{n \geqslant 1}$ are tight in $\ell^2_\downarrow$, and the limiting object in Theorem 6.1.5 is finite almost surely. We state the tightness below and defer the finiteness of the limiting object to Proposition 6.3.15 in the next section.

**Proposition 6.3.14.** *Suppose that* $\lambda < c_0^{-1/2}$*, where* $c_0$ *is defined in Corollary 6.3.3. Then,* $(n^{-\alpha}(W_{(i)})_{i \geqslant 1})_{n \geqslant 1}$ *and* $((n^{\alpha}p_c)^{-1}(|\mathscr{C}_{(i)}(p_c)|)_{i \geqslant 1})_{n \geqslant 1}$ *are tight in* $\ell_{\downarrow}^2$*.*

*Proof.* To show the $\ell_{\downarrow}^2$-tightness of $(n^{-\alpha}(W_{(i)})_{i \geqslant 1})_{n \geqslant 1}$, it is enough to show that for any $\varepsilon > 0$

$$\lim_{K \to \infty} \limsup_{n \to \infty} \mathbb{P}\left( \sum_{i > K} W_{(i)}^2 > \varepsilon n^{2\alpha} \right) = 0.$$

Consider the graph $\mathrm{GRG}_n(\boldsymbol{w}, p_c) \setminus [K]$, and define $W_{(i)K}$, $W(v; K)$ on this graph analogously as $W_{(i)}$ and $W(v)$. It is enough to show that

$$\lim_{K \to \infty} \limsup_{n \to \infty} \mathbb{P}\left( \sum_{i \geqslant 1} W_{(i),K}^2 > \varepsilon n^{2\alpha} \right) = 0. \tag{6.44}$$

Let $V_n^*(K)$ denote a vertex chosen in a size-biased manner from $[n] \setminus [K]$ with the sizes being proportional to $(w_i)_{i > K}$, chosen independently from $\mathrm{GRG}_n(\boldsymbol{w}, p_c)$. Let $\ell_n(K) := \sum_{i > K} w_i$. Then, $\ell_n(K) = \ell_n(1 + o(1))$ for each fixed $K \geqslant 1$. Note that

$$\mathbb{E}\left[ \sum_{i \geqslant 1} W_{(i),K}^2 \right] = \ell_n(K)\mathbb{E}[W(V_n^*(K); K)]$$

$$\leqslant \ell_n(K)\left[ \sum_{k \geqslant 0} \sum_{v \in [n]} w_v f_{2k}(V_n^*(K), v) + \sum_{k \geqslant 0} \sum_{v \in [n]} w_v f_{2k+1}(V_n^*(K), v) \right]. \tag{6.45}$$

Let us denote the two sums above by (I) and (II) respectively. We can now use the estimates from (6.33) and (6.34). Note that (6.33) implies that, for $\lambda < c_0^{-1/2}$,

$$n^{-2\alpha}(\mathrm{I}) \leqslant Cn^{-2\alpha}\ell_n(K)\mathbb{E}[w_{V_n^*(K)}] = C\sum_{i > K} i^{-2\alpha},$$

which tends to zero as $K \to \infty$. Moreover, (6.34) implies that for $\lambda < c_0^{-1/2}$,

$$n^{-2\alpha}(\mathrm{II}) \leqslant Cn^{-\epsilon} \sum_{i > K} \frac{1}{i} = \frac{C \log(n)}{n^{\epsilon}},$$

which goes to zero as $n \to \infty$. Thus, (6.45) follows, and (6.44) follows from Markov's inequality.

For the $\ell_{\downarrow}^2$-tightness of $((n^{\alpha}p_c)^{-1}(|\mathscr{C}_{(i)}(p_c)|)_{i \geqslant 1})_{n \geqslant 1}$, note that, for any vertex $i \in [n]$, $\mathbb{E}[|\mathscr{C}_{k+1}(i)|] \leqslant p_c\mathbb{E}[W_k(i)]$ for all $k \geqslant 1$, and therefore $\mathbb{E}[|\mathscr{C}(i)|] \leqslant$

$p_c \mathbb{E}[W(i)]$. Thus, if $V_n^*$ a vertex chosen in a size-biased manner with the sizes being $(w_i/\ell_n)_{i \in [n]}$, chosen independently of $\mathrm{GRG}_n(w, p_c)$,

$$\mathbb{E}\left[\sum_{i \geqslant 1} |\mathscr{C}_{(i)}|^2\right] = \mathbb{E}\left[\sum_{i \in [n]} |\mathscr{C}(i)|\right] \leqslant p_c \mathbb{E}\left[\sum_{i \in [n]} W(i)\right] = p_c \mathbb{E}\left[\sum_{i \geqslant 1} |\mathscr{C}_{(i)}| W_{(i)}\right]$$

$$= \ell_n p_c \mathbb{E}[|\mathscr{C}(V_n^*)|] \leqslant \ell_n p_c^2 \mathbb{E}[W(V_n^*)] = p_c^2 \mathbb{E}\left[\sum_{i \geqslant 1} W_{(i)}^2\right].$$

We can apply this quantity to $\mathrm{GRG}_n(w, p_c) \setminus [K]$ as above and the proof of Proposition 6.3.14 is now complete. $\qquad\square$

### 6.3.6    Finiteness of the limiting object

We write $\theta_i = i^{-\alpha}$. Recall that the graph $G_\infty(\lambda)$ with vertex set $\mathbb{Z}_+$ is created by creating $\mathrm{Poisson}(\lambda_{ij})$ many edges vertices $i$ and $j$, where

$$\lambda_{ij} = \lambda^2 \int_0^\infty \frac{\theta_i \theta_j x^{-2\alpha}}{(\mu + \theta_i x^{-\alpha})(\mu + \theta_j x^{-\alpha})} dx. \tag{6.46}$$

Let $C(i)$ denote the connected component containing vertex $i$ and define $W^\infty(i) = \sum_{j \in C(i)} \theta_j$. We will show the following and the fact that $(W_{(i)}^\infty)_{i \geqslant 1} \in \ell_\downarrow^2$ then follows from Proposition 6.3.14 using Fatou's lemma.

**Proposition 6.3.15.** *Consider $\theta_i = i^{-\alpha}$. There exists and absolute constant $\lambda_0$ such that the following holds for any $\lambda \in (0, \lambda_0)$: For each $i \in \mathbb{Z}_+$, $W^\infty(i) < \infty$ almost surely.*

**Lemma 6.3.16.** $\lambda_{ij} \leqslant \frac{C\lambda^2}{(i \wedge j)^{1-\alpha}(i \vee j)^\alpha}$ *for some absolute constant $C > 0$.*

*Proof.* Without loss of generality, let us assume that $\theta_i > \theta_j$ (i.e., $i < j$), and $\mu = 1$. Let us split the integral (6.46) in three parts with $\{x : \theta_i x^{-\alpha} < 1\}$, $\{x : 0 < \theta_j x^{-\alpha} < 1 < \theta_i x^{-\alpha}\}$ and $\{x : \theta_j x^{-\alpha} > 1\}$ and denote them by (I), (II) and (III) respectively. Then,

$$\text{(I)} \leqslant \lambda^2 \theta_i \theta_j \int_{\theta_i^{1/\alpha}}^\infty x^{-2\alpha} dx \leqslant C\lambda^2 \theta_i \theta_j (\theta_i^{1/\alpha})^{1-2\alpha} = \frac{C\lambda^2}{i^{1-\alpha} j^\alpha},$$

$$\text{(II)} \leqslant \lambda^2 \theta_j \int_{\theta_j^{1/\alpha}}^{\theta_i^{1/\alpha}} x^{-\alpha} dx = \frac{C\lambda^2}{j^\alpha}\left[\frac{1}{i^{1-\alpha}} - \frac{1}{j^{1-\alpha}}\right] \leqslant \frac{C\lambda^2}{i^{1-\alpha} j^\alpha},$$

$$\text{(III)} \leqslant \lambda^2 \int_0^{\theta_j^{1/\alpha}} dx \leqslant \frac{C\lambda^2}{j} \leqslant \frac{C\lambda^2}{i^{1-\alpha} j^\alpha}.$$

Thus the proof follows. $\qquad\square$

*Proof of Proposition 6.3.15.* We will use the path counting estimates from [73]. We estimate the probability that there exists a non self-intersecting path of length $k$ from $i$ to $j$ in $G_\infty$. Using Lemma 6.3.16, note that

$$\mathbb{P}(j \in C(i), d(i,j) = k)$$

$$\leqslant (c_2\lambda^2)^k \sum_{\substack{(\nu_r)_{r=0}^k : \nu_r\text{'s distinct} \\ \nu_0=i, \nu_k=j}} \prod_{r=0}^{k-1} \frac{1}{(\nu_r \wedge \nu_{r+1})^{1-\alpha}(\nu_r \vee \nu_{r+1})^\alpha}.$$

Using [73, Lemma 2.4], for any $b < \alpha$,

$$\mathbb{P}(j \in C(i), d(i,j) = k) \leqslant \frac{(c_3\lambda^2)^k}{i^b j^{1-b}},$$

for some absolute constant $c_3 > 0$. Thus, for $\lambda < 1/\sqrt{c_3}$,

$$\mathbb{E}[W^\infty(i)] \leqslant C \sum_{j \geqslant 1} j^{-\alpha} \frac{(c_3\lambda^2)^k}{i^b j^{1-b}} \leqslant \frac{C}{i^b} < \infty,$$

which implies $W^\infty(i) < \infty$ almost surely. □

### 6.3.7   Near-critical behavior

*Proof of Theorem 6.1.8.* The proof can be completed by modifying the arguments in Section 6.3. In fact, if $p_n = \lambda_n n^{-(3-\tau)/2}$, for some $\lambda_n \to 0$, then Lemma 6.3.2 holds with $\lambda$, replaced by $\lambda_n$. One can use identical arguments as Proposition 6.3.4 to show that $W(i) = w_i(1 + o_\mathbb{P}(1))$. Finally, one can use identical arguments as Proposition 6.3.14 to deduce the $\ell_\downarrow^2$ tightness of the vector of component sizes and weights. Thus, the proof of Theorem 6.1.8 follows. □

## 6.4   Erased configuration model: Proofs

In this section, we provide the necessary adaptations required to the arguments in Section 6.3 to complete the proof of Theorem 6.1.7. Let $e_{ij}$ denote the number of edges between vertices $i$ and $j$ in $CM_n(\mathbf{d})$. Note that an edge $\{i, j\}$ appears in $ECM_n(\mathbf{d})$ if and only if $e_{ij} \geqslant 1$. We start by describing two elementary properties of the occurrence of edges in Lemmas 6.4.1, 6.4.2, which will be the key to the required adaptations:

**Lemma 6.4.1.** *For all distinct* $i, v, j$, $\mathbb{P}(e_{iv} \geqslant 1, e_{vj} \geqslant 1) \leqslant \mathbb{P}(e_{iv} \geqslant 1)\mathbb{P}(e_{vj} \geqslant 1)$

.

*Proof.* For any $m_1 \geqslant 1$,

$$\mathbb{P}(e_{iv} = m_1) = \frac{\binom{d_i}{m_1}\binom{d_v}{m_1}m_1!}{(\ell_n - 1)\dots(\ell_n - 2m_1 + 1)}. \tag{6.47}$$

Further, fix an $m_2 \geqslant 1$. Now conditionally on the first $m_1$ edge, the probability of $\{e_{vj} = m_2\}$ is also given by (6.47), with $d_v$ changed to $d_v - m_2$ and the product in the denominator being $(\ell_n - 2m_1 - 1)\cdots(\ell_n - 2(m_1 + m_2) + 1)$. Therefore, for any $m_1, m_2 \geqslant 1$,

$$\mathbb{P}(e_{iv} = m_1, e_{vj} = m_2) = \frac{\binom{d_i}{m_1}\binom{d_v}{m_1}m_1!\binom{d_v - m_1}{m_2}\binom{d_j}{m_2}m_2!}{(\ell_n - 1)\cdots(\ell_n - 2(m_1 + m_2) + 1)}.$$

Moreover, for any $a > 0$, $x/y \leqslant (x + a)/(y + a)$, iff $x \leqslant y$, which yields that $(d_v - m_1 - k)/(\ell_n - 2m_1 - k - 1) \leqslant (d_v - k)/(\ell_n - k - 1)$. Therefore,

$$\frac{\mathbb{P}(e_{iv} = m_1, e_{vj} = m_2)}{\mathbb{P}(e_{iv} = m_1)\mathbb{P}(e_{vj} = m_2)} = \frac{\binom{d_v - m_1}{m_2}}{\binom{d_v}{m_2}} \prod_{i=1}^{m} \frac{\ell_n - 2m_1 + 1 - 2i}{\ell_n - 2i + 1} \leqslant 1.$$

The proof of Lemma 6.4.1 thus follows.   $\square$

**Lemma 6.4.2.** *For any* $i \neq j$, $\mathbb{P}(e_{ij} \geqslant 1) \leqslant \min\{\frac{d_i d_j}{\ell_n - 1}, 1\}$.

*Proof.* Obviously, $\mathbb{P}(e_{ij} \geqslant 1) \leqslant 1$, and by Markov's inequality,

$$\mathbb{P}(e_{ij} \geqslant 1) \leqslant \mathbb{E}[e_{ij}] = \frac{d_i d_j}{\ell_n - 1},$$

and the proof follows.   $\square$

Define $p_{ij}(2)$ to be the probability that there exists some intermediate vertex $v$ such that $\{i, v\}$ and $\{v, j\}$ form edges in $\mathrm{ECM}_n(\mathbf{d}, p_c)$. We have seen that the bound in Lemma 6.3.2 forms the bedrock for all the error estimates for $\mathrm{GRG}_n(\mathbf{w}, p_c(\lambda))$, while using the first moment method. The next lemma provides an analogue of Lemma 6.3.2 for $\mathrm{ECM}_n(\mathbf{d}, p_c(\lambda))$:

**Lemma 6.4.3.** $p_{ij}(2) \leqslant \frac{C\lambda^2}{(i \wedge j)^{1-\alpha}(i \vee j)^\alpha}$, $\forall i, j \in [n]$ *for some constant* $C > 0$.

*Proof.* Without loss of generality, we assume that $d_i > d_j$, i.e., $i < j$. Note that

$$p_{ij}(2) \leqslant p_c^2 \sum_{v \in [n]} \mathbb{P}(e_{iv} \geqslant 1, e_{iv} \geqslant 1) \leqslant p_c^2 \sum_{v \in [n]} \mathbb{P}(e_{iv} \geqslant 1)\mathbb{P}(e_{iv} \geqslant 1),$$

where we have used Lemma 6.4.1. Let us split the sum in three parts with $\{v : d_i d_v \leqslant \ell_n\}$, $\{v : d_j d_v \leqslant \ell_n < d_i d_v\}$, and $\{v : d_j d_v > \ell_n\}$, and bound $\mathbb{P}(e_{iv \geqslant 1})\mathbb{P}(e_{vj} \geqslant 1)$ from above by $4d_i d_v^2 d_j/\ell_n^2$, $2d_v d_j/\ell_n$ and $1$ respectively on those sets. The rest of the argument is identical to Lemma 6.3.2.   $\square$

For the proof of the results for $\mathrm{ECM}_n(\mathbf{d}, p_c(\lambda))$, we replace the $w_i$'s by the $d_i$'s in all the notations in Section 6.3. The inequality in Lemma 6.4.3 again establishes that the two-hop connection probabilities are upper-bounded by the connection probabilities of the preferential attachment model. Thus, we can use Lemma 6.4.3 together with Lemma 6.4.1 to get an identical estimate as (6.31), and therefore Proposition 6.3.4 holds. Let us now identify the connectivity structure between the hubs to establish an analog of Proposition 6.3.7.

We use the notation $V_L := \{i : d_i > \delta n^\alpha\}$, $V_{SS} = \{i : d_i < \delta n^\rho\}$, $V_{SI} = \{i : \delta n^\rho \leqslant d_i \leqslant \delta^{-1} n^\rho\}$ and $V_{SL} = \{i : d_i > \delta^{-1} n^\rho\}$. Consider the following multigraph $G_{n,\delta}$ on the vertex set $V_L$, where the number of edges $X_{ij}$ between $i$ and $j$ is number of distinct $v \in [n]$ such that both $(i, v)$ and $(v, j)$ are edges of $\mathrm{ECM}_n(\mathbf{d}, p_c)$. Note that, for any $i \neq j$,

$$X_{ij} = \sum_{v \neq i,j} \mathrm{Ber}\big(p_c^2 \mathbb{P}(e_{iv} \geqslant 1, \ e_{vj} \geqslant 1)\big).$$

We can split the above sum in three parts with $v \in V_{SS}$, $v \in V_{SI}$, and $v \in V_{SL}$ and denote them by (I), (II), and (III) respectively. Now, using Lemma 6.3.1 and Lemma 6.4.1,

$$\mathbb{E}[(\mathrm{I})] \leqslant \frac{d_i d_j p_c^2}{\ell_n^2} \sum_{v:d_v<\delta n^\rho} d_v^2 \leqslant C_0 \theta_1^2 \delta^{3-\tau} n^{2\alpha-2+1+(3-\tau)\rho} p_c^2 \leqslant C_1 \delta^{3-\tau},$$

$$(\mathrm{III}) \leqslant p_c^2 \times \#\{v : d_v > \delta^{-1} n^\rho\} = \delta^{\tau-1}.$$

Using the above calculation of (I) and (III) and Markov's inequality, it follows that

$$X_{ij} = \sum_{v:d_v \in [\delta n^\rho, \delta^{-1} n^\rho]} \mathrm{Ber}\big(p_c^2 \mathbb{P}(e_{iv} \geqslant 1, \ e_{vj} \geqslant 1)\big) + \mathrm{Err}_{\mathbb{P}}(\delta, n)$$

$$= X_{ij}(\delta) + \mathrm{Err}_{\mathbb{P}}(\delta, n). \tag{6.48}$$

Define the quantity

$$\lambda_{ij}(\delta) = p_c^2 \sum_{v:d_v \in [\delta n^\rho, \delta^{-1} n^\rho]} \left(1 - e^{-\frac{d_i d_v}{\ell_n}}\right)\left(1 - e^{-\frac{d_j d_v}{\ell_n}}\right). \tag{6.49}$$

**Lemma 6.4.4.** *For any $\delta \in (0,1)$, and $i, j \geqslant 1$*

$$\lim_{n\to\infty} d_{\mathrm{TV}}\big(X_{ij}(\delta), \mathrm{Poi}(\lambda_{ij}(\delta))\big) = 0.$$

*Proof.* Let us first estimate the probabilities of the Bernoulli random variables in (6.48). Firstly, note that using [105, (36)], for any $v \neq i$

$$\left| \mathbb{P}(e_{iv} = 0) - e^{-\frac{d_i d_v}{\ell_n}} \right| \leqslant \frac{d_i d_v^2}{(\ell_n - d_1)^2},$$

$$\left| \mathbb{P}(e_{iv} = 0, e_{vj} = 0) - e^{-\frac{(d_i + d_j)d_v}{\ell_n}} \right| \leqslant \frac{(d_i + d_j)d_v^2}{(\ell_n - d_1)^2}. \tag{6.50}$$

The second inequality also follows from [105, (36)], because if we merge two vertices $i$ and $j$ into one single vertex, then $\mathbb{P}(e_{iv} = 0, e_{vj} = 0)$ becomes the probability that the merged vertex does not have an edge with $v$, and one can use the first bound to deduce the second. Now, using Lemma 6.3.1,

$$\frac{d_i}{(\ell_n - d_1)^2} \sum_{v \leqslant \delta^{-1} n^\rho} d_v^2 \leqslant C n^{-\rho(\tau - 2)} \to 0,$$

and therefore one can conclude that

$$E_1 := p_c^2 \sum_{v: d_v \in [\delta n^\rho, \delta^{-1} n^\rho]} \left| \mathbb{P}(e_{iv} \geqslant 1, \ e_{vj} \geqslant 1) - \left(1 - e^{-\frac{d_i d_v}{\ell_n}}\right)\left(1 - e^{-\frac{d_j d_v}{\ell_n}}\right) \right|$$

which goes to zero as $n \to \infty$. We will now use Stein's method for convergence for sums of negatively correlated Bernoulli random variables [115, Theorem 6.24], which states that for $I_i \sim \text{Bernoulli}(p_i)$ with $\text{cov}(I_i, I_j) < 0$, for all $i \neq j$, then

$$d_{TV}\left( \sum_i I_i, \text{Poi}\left( \sum_i p_i \right) \right) \leqslant \max_i p_i.$$

Note that $\mathbf{1}\{e_{iv} \geqslant 1, e_{vj} \geqslant 1\}$ and $\mathbf{1}\{e_{iv'} \geqslant 1, e_{v'j} \geqslant 1\}$ are negatively correlated for $v \neq v'$ which can be established using similar arguments as Lemma 6.4.1. Therefore

$$d_{TV}\left( X_{ij}(\delta), \text{Poi}(\lambda_{ij}(\delta)) \right)$$

$$\leqslant p_c^2 E_1 + p_c^2 \max_{v: w_v \in [\delta n^\rho, \delta^{-1} n^\rho]} \left(1 - e^{-\frac{d_i d_v}{\ell_n}}\right)\left(1 - e^{-\frac{d_j d_v}{\ell_n}}\right) = O(p_c^2),$$

and the proof follows. $\qquad \square$

**Lemma 6.4.5.** *For any fixed* $i, j \geqslant 1$

$$\lim_{\delta \to 0} \lim_{n \to \infty} \lambda_{ij}(\delta) = \lambda^2 \int_0^\infty \left(1 - e^{-\frac{\theta_i x^{-\alpha}}{\mu}}\right)\left(1 - e^{-\frac{\theta_j x^{-\alpha}}{\mu}}\right) dx$$

*Proof.* Using (6.49), the proof is identical to Lemma 6.3.9. $\qquad \square$

For the asymptotic independence of the two-hop connections between the hubs, we need an estimate of the joint connection probabilities between distinct vertices in the configuration model, as given in the following lemma:

**Lemma 6.4.6.** *Consider four distinct vertices $i$, $u$, $k$ and $v$ such that $d_u, d_v \in [an^\rho, bn^\rho]$ and $d_i, d_j \in [an^\alpha, bn^\alpha]$, for some $a, b > 0$. Then, as $n \to \infty$,*

$$\left| \mathbb{P}(e_{iu} = 0, e_{kv} = 0) - e^{-\frac{d_i d_u}{\ell_n}} e^{-\frac{d_k d_v}{\ell_n}} \right| = O\left( \frac{d_i d_u^2 + d_k d_v^2}{(\ell_n - 2d_1)^2} \right).$$

*Proof.* We first sequentially pair the half-edges of $u$, and then in the next step pair the half-edges of $v$. Let $\mathcal{A}_n$ denote the event that after the first stage of pairing $u$ does not create more than $n^\varepsilon$ edges with $k$ or $v$, where $0 < \varepsilon < \rho$. Firstly, note that $\mathbb{P}(\mathcal{A}_n^c)$ is exponentially small in $n$. Indeed, using (6.47), the probability that $u$ and $k$ share at least $n^\varepsilon$ edges is at most

$$\sum_{m > n^\varepsilon} \left( \frac{d_k d_u}{\ell_n - 2d_1} \right)^{m_1} \frac{1}{m_1!} \leqslant Ce^{-Cn^\varepsilon},$$

where we have used the fact that $d_k d_u = O(\ell_n)$, and the last bound follows from the tail probabilities of a Poisson distribution. A similar bound holds for the connections between $u$ and $v$ as well, and therefore,

$$\mathbb{P}(\mathcal{A}_n^c) \leqslant Ce^{-cn^\varepsilon}. \tag{6.51}$$

Now, using (6.50) and (6.51)

$$\mathbb{P}(\mathcal{A}_n \cap \{e_{iu} = 0\}) = e^{-\frac{d_i d_u}{\ell_n}} + O\left( \frac{d_i d_u^2}{(\ell_n - 2d_1)^2} \right),$$

and considering the second step of pairing of the remaining half-edges of $v$

$$\mathbb{P}(e_{kv} = 0 | \mathcal{A}_n \cap \{e_{iu} = 0\}) = e^{-\frac{d_k d_v}{\ell_n}} + O\left( \frac{d_k d_v^2}{(\ell_n - 2d_1)^2} \right).$$

Thus the proof follows. $\qquad \square$

**Lemma 6.4.7.** *For any $\delta \in (0,1)$, the collection of random variables $(X_{ij}(\delta))_{i,j \in V_L}$ is asymptotically independent.*

*Proof.* Consider vertices $i, j, k, l$ from $V_L$. Let $\mathcal{E}_{iuj}$ denote the event that $\{e_{iu} \geqslant 1, e_{uj} \geqslant 1\}$. Suppose that we can show as $n \to \infty$

$$E := p_c^2 \sum_{i,j,k.l \in V_L} \sum_{d_u, d_v \in [\delta n^\rho, \delta^{-1} n^\rho]} \left| \mathbb{P}(\mathcal{E}_{iuj} \cap \mathcal{E}_{kvl}) - \mathbb{P}(\mathcal{E}_{iuj})\mathbb{P}(\mathcal{E}_{kvl}) \right| = o(1).$$

$$\tag{6.52}$$

Then, with high probability, we can couple all the Bernoulli random variable in $(X_{ij}(\delta))_{i,j\in V_L}$ with an independent collection of Bernoulli random variables, and then the proof will be complete. Suppose that $i, j, k, l$ are distinct. To estimate $\mathbb{P}(\mathcal{E}_{iuj} \cap \mathcal{E}_{kvl})$, we use Lemma 6.4.6. Indeed, if we merge vertices $i$ with $j$ and $k$ with $l$, and denote them by $v_1$ and $v_2$ respectively, then the required probability is the same as the probability that $(v_1, u)$ and $(v_2, v)$ share an edge. Therefore, Lemma 6.4.6 implies that $E$ defined by (6.52) is of the order

$$p_c^2 \sum_{u,v:\ d_u,d_v\in[\delta n^\rho,\delta^{-1}n^\rho]} \frac{(d_i + d_j)d_u^2 + (d_k + d_l)d_v^2}{(\ell_n - 2d_1)^2}$$

$$\leqslant p_c^2 \frac{Cn^\alpha}{\ell_n^2} n^{1-(\tau-1)\rho} n^{1+\rho(3-\tau)} = Cn^{-\rho(\tau-2)} \to 0.$$

Since $V_L$ is a finite collection, this proves (6.52) on the partial sum with $i, j, k, l$ being distinct. The case where $i = k$ can be dealt with similarly using (6.50), and we do not repeat the argument again. This completes the proof of Lemma 6.4.7.    $\square$

## 6.5    Conclusion

In this chapter, we have provided to the best of our knowledge the first mathematically rigorous analysis for the critical window for random graphs that satisfy a power law distribution with exponent $\tau \in (2,3)$. The main surprise is that the critical window changes depending on whether the network is constrained to have single-edges between the vertices. We identify the critical window and the scaling limits for the component sizes within the critical window. The main technical obstacle under the single-edge constraint is that the exploration process approach does not work and we had to resort to path counting techniques to identify the primary contributions on the component sizes. For this reason, the proof does not work for sufficiently large $\lambda$ values, which we leave as an open question.

**Chapter 7**

# Open problems

In the final chapter of this thesis, we discuss several related research questions, which are open to the best of our knowledge.

## 7.1   Component sizes and complexity.

### 7.1.1   Finite and infinite third moment cases

**Tail bound on the component sizes.**   While the scaling limit results yield an exact asymptotic distribution of the largest component sizes, these results do not give any explicit probability bounds for large, but finite $n$. It is often interesting to obtain explicit bounds for the probabilities of the events involving "tails" such as the largest component is quite large ($\{|\mathscr{C}_{(1)}(p_c(\lambda))| > An^\rho\}$ for large $A > 0$), or it is quite small ($\{|\mathscr{C}_{(1)}(p_c(\lambda))| \leqslant \delta n^\rho\}$ for small $\delta > 0$), where $\rho$ is the scaling exponent for the component sizes. For critical Erdős-Rényi random graphs with $p = 1/n(1 + \lambda n^{-1/3})$, Nachmias and Peres [136] showed that

$$\mathbb{P}(|\mathscr{C}_{(1)}(p_c(\lambda))| > An^{2/3}) \leqslant \frac{c_1}{A}e^{-c_2 A^3}, \tag{7.1}$$

$$\mathbb{P}(|\mathscr{C}_{(1)}(p_c(\lambda))| \leqslant \delta n^{2/3}) \leqslant c_3 \delta^{3/5}, \tag{7.2}$$

for any $A > A_0$, $\delta \in (0, \delta_0)$, and $n \geqslant n_0$, where $c_1, c_2, c_3, A_0, \delta_0 > 0$ and $n_0 \geqslant 1$ can be calculated explicitly. See also [143] and [148, Corollary 19]. In the context of quantum random graphs [66], the bound in (7.1) was proven to be $A^{-3/2}$, with the exponential term missing, while the bound for (7.2) involves $\delta^{3/5}$. Even for Erdős-Rényi random graphs, whether the bounds in (7.1) and (7.2) are optimal is not known to the best of our knowledge. Using results from [143] (see also [102, (1.2)]), [8, Corollary 2] and Portmanteau theorem, we can obtain a lower bound on the probability (7.1) of $c_1 A^{-3/2}e^{-c_2 A^3}$, which differs from (7.1) in the polynomial term.

There are several challenges in order to derive such estimates for random graphs with general moment assumptions on the degrees such as Chapter 2. Specifically, the techniques involving Chernoff bound in [136, Section 4] does not work for general degree distributions. Bounds like (7.1) and (7.2) in the $\tau \in (3,4)$ case has never been studied. The properties of the scaling limit was studied in [5, 102] for $\tau \in (3,4)$. Using the results from [102, Theorem 1.6], Theorem 3.2.2 and Portmanteau theorem, the probability in (7.1) is bounded from below by $c_1 A^{-(\tau-1)/2}e^{-c_2 A^{\tau-1}}$. But an upper-bound with the same exponential term could also possibly be proved leading to the following conjecture:

**Conjecture 7.1.** *Under Assumption 3.1, there exists* $A_0 > 0$, $n_0 \geqslant 1$ *and constant* $c_1, c_2$ *such that for all* $A > A_0$ *and* $n \geqslant n_0$

$$\mathbb{P}(|\mathscr{C}_{(1)}(p_c(\lambda))| > Ab_n) \leqslant \frac{c_1}{p(A)} e^{-c_2 A^{\tau-1}},$$

*for some polynomial* $p$.

We do not have a good guess about the bound of the form (7.2) for $\tau \in (3,4)$. It is further interesting to learn about these tail probabilities when $A = A_n \to \infty$ or $\delta = \delta_n \to 0$. For $A_n \ll n^{1/12}$, suitable bounds were derived for Erdős-Rényi random graphs in [146, Proposition 3.1]. In Chapter 5, we consider the $A_n \to \infty$ case (see Theorem 5.4.1), but we did not focus on obtaining the optimal bound, which is an interesting question.

**Comparison to Joseph's scaling limit.**    As observed in Section 3.3.2, Assumption 3.1 is satisfied almost surely when the degrees are an iid sample from a power-law distribution with exponent $\tau \in (3,4)$. Thus, conditionally on the observed degree sequence, the exploration process converges to the process

$$S_{\infty,1}(t) := \sum_{i=1}^{\infty} \frac{C_F}{\Gamma_i^\alpha} \left( \mathbf{1}_{\left\{ X_i \leqslant \frac{s \mu \Gamma_i^\alpha}{C_F} \right\}} - \frac{C_F}{\mu \Gamma_i^\alpha} \int_{\Gamma_{i-1}}^{\Gamma_i} u^{-2\alpha} du \right), \qquad (7.3)$$

where $(X_j)_{j \geqslant 1}$ and $\Gamma_i = \sum_{j \leqslant i} E_j$ where $(E_j)_{j \geqslant 1}$ is an independent collections of i.i.d unit rate exponential random variables; see Theorem 3.5.1. On the other hand, the iid degree setting has been studied in [65, 121], where the scaling limit of the exploration process turns out to be different. More precisely, the scaling limit is given by

$$S_{\infty,2}(t) = Y(t) + A(t), \qquad (7.4)$$

where

$$A(t) = -\frac{C_F \Gamma(4-\tau)}{(\tau-3)(\tau-2)\mu^{\tau-2}} t^{\tau-2},$$

and $Y(t)$ is the unique process with independent increments such that for every $t \geqslant 0$ and $u \in \mathbb{R}$,

$$\mathbb{E}[\exp(iuY(t))] = \exp\left( \int_0^t ds \int_0^\infty dx (e^{iux} - 1 - iux) \frac{C_F}{\mu x^{\tau-1}} e^{-xs/\mu} \right).$$

Since the second process is the limit of the same exploration process averaged out over the degrees, this indirectly implies that the law of $\mathbf{S}_{\infty,2}$ is the

same as the law of $\mathbf{S}_{\infty,1}$, averaged out over the $\Gamma$-values (although we use a different exploration process from [121], the fact that the component sizes are huge compared to the number of cycles in a component, one can prove Theorem 3.5.1 for the exploration process in [121] also). This is remarkable given the vastly different descriptions of the scaling limits (7.3) and (7.4). For example, the martingale part of (7.3) does not have independent increments due to thinning of the Poisson processes. However, after averaging out over $\Gamma$-values, the dependence goes away. It may be worthwhile investigating whether there is a direct approach to show that $\mathbf{S}_{\infty,1}$, after averaged out over the $\Gamma$-values, yields the same law as $\mathbf{S}_{\infty,2}$.

**Joint convergence over the critical window.**    While studying the joint convergence over the critical window in Theorems 2.2.4 and 3.2.5, we considered finite dimensional convergence. It will be interesting to show that the process $(\mathbf{Z}_n(\lambda))_{\lambda \in \mathbb{R}}$ converges in $\mathbb{D}(\mathbb{R}, \mathbb{U}_\downarrow^0)$, where we recall that $\mathbf{Z}_n(\lambda)$ is the vector of rescaled component sizes and the surplus edges and $\mathbb{D}(\mathbb{R}, \mathbb{U}_\downarrow^0)$ denotes the set of càdlàg functions equipped with the Skorohod $J_1$-topology. The proof will follow if one can verify a suitable tightness criterion $\mathbb{D}(\mathbb{R}, \mathbb{U}_\downarrow^0)$-valued stochastic processes, but we were unable to find a suitable tightness criterion.

**Dynamically evolving critical random graphs.**    In [146], Roberts and Şengül considered a dynamically evolving version of critical Erdős-Rényi random graphs. The dynamic graph process $(G_t)_{t \geqslant 0}$ starts with $G_0$ which is distributed as $\mathrm{ERRG}_n(1/n)$, and each pair $\{u, v\}$ is equipped with an independent rate-one Poisson process $\mathcal{N}_{uv}$. At each event time of $\mathcal{N}_{uv}$, an edge is resampled according to an independent Bernoulli$(1/n)$ random variables. Then, for each fixed $t \geqslant 0$, $G_t$ is distributed as $\mathrm{ERRG}_n(1/n)$. Let $\mathscr{C}_{(1)}(t)$ denote the largest component of $G_t$. It was shown in [146] that $M_n := \sup_{t \in [0,1]} |\mathscr{C}_{(1)}(t)| \leqslant \beta n^{2/3} (\log(n))^{1/3}$ with high probability for $\beta < 2/3^{2/3}$. There are several further interesting questions that arise for this dynamic graph process.

1. Does $M_n / n^{2/3} (\log(n))^{1/3}$ converge in probability to some $\beta_0 > 0$?

2. What is the behavior of $I_n := \inf_{t \in [0,1]} |\mathscr{C}_{(1)}(t)|$?

3. What happens in the heavy-tailed universality class of the multiplicative coalescent regime? How does the exponent of $\log(n)$ change depending on the power-law exponent $\tau$ of the degree distribution.

Questions 1 and 2 are discussed in [146]. While $\beta_0$ was conjectured to be $2/3^{1/3}$, not much was known for question 2. For question 3, the tail asymptotics in (7.1) is intimately related to the asymptotics of $M_n$, as noted in [146]. For each $t \in [0, 1]$, $\mathscr{C}_{(1)}(t)$ satisfies (7.1), and therefore a simple union bound yields an upperbound of $M_n = O(n^{2/3}(\log(n))^{1/3})$ with high probability (proving the lower bound is considerably difficult which was accomplished in [146]). Following the prediction of tail bounds in Conjecture 7.1, this leads us to the the conjecture below:

**Conjecture 7.2.** *Under Assumption 3.1, $M_n = \Theta_{\mathbb{P}}(n^\rho (\log(n))^{1/(\tau-1)})$.*

**Simulation guarantee for sample paths.**    The scaling limits for the component sizes are described by largest excursions of certain stochastic processes with negative drift. It is difficult to generate a sample from this distribution due to the lack of availability of the precise distribution function. It will be interesting in the stochastic simulation literature to develop techniques for generating a sample from this distribution, and obtain exact error bounds if the simulation method is approximate. For the heavy-tailed scaling limits, simulating sample paths of the thinned Lévy process is not standard. A natural strategy could be to truncate the sum in (3.1) upto first K terms for a large K. However, there is a more accurate approach using the techniques in [19]. Here, one can approximate the small jumps by a Brownian motion using the following theorem: Recall that $\mathcal{I}_i(t) = \mathbf{1}\{\xi_i \leqslant t\}$, $\xi_i \sim \text{Exp}(\theta_i)$ (let $\mu = 1$) and define

$$E_K(t) = \sum_{i=K+1}^{\infty} \theta_i (\mathcal{I}_i(t) - \theta_i t), \quad \sigma_K^2 = \sum_{i=K+1}^{\infty} \theta_i^3.$$

**Theorem 7.1.1** ([67]). *Suppose that $\frac{\max_{i>K} \theta_i}{\sum_{i=K+1}^{\infty} \theta_i^3} \to 0$. Then, as $K \to \infty$,*

$$(\sigma_K^{-1} E_K(t))_{t \geqslant 0} \xrightarrow{d} W,$$

*where $W$ is a standard Brownian motion.*

See Appendix 7.A for a proof. Note that if we take $\theta_i = C_F i^{-\alpha}$ for some $\alpha \in (1/3, 1/2)$, we have $\sum_{i=K+1}^{\infty} \theta_i^3 = \Theta(K^{1-3\alpha})$. Therefore,

$$\frac{\max_{i>K} \theta_i}{\sum_{i=K+1}^{\infty} \theta_i^3} = \Theta(K^{-\alpha-1+3\alpha}) = \Theta(K^{2\alpha-1}) \to 0.$$

Thus, the assumption of Theorem 7.1.1 is satisfied. However, we did not pursue the question of simulation guarantees further.

**Independent proof for the scaling limit of diameter.** In Chapter 5, we have seen that the largest components converge as measured metric spaces under the Gromov-Hausdorff-Prokhorov topology yields the convergence of the diameters of these components. This is yields that $\mathrm{diam}(\mathscr{C}_{(1)}(p_c(\lambda)))$ converges in distribution to some random variable. The above approach of proving the convergence of diameters is indirect and considerably difficult. Till date there is no direct approach available to show the convergence of $\mathrm{diam}(\mathscr{C}_{(1)}(p_c(\lambda)))$ even for the Erdős-Rényi random graphs. Only some bounds were derived in [134] establishing the tightness of $\mathrm{diam}(\mathscr{C}_{(1)}(p_c(\lambda)))$ and $(\mathrm{diam}(\mathscr{C}_{(1)}(p_c(\lambda))))^{-1}$. A direct proof is expected to yield a simpler expression for the diameter of the limiting metric spaces in [3, 36], and it is expected to require novel techniques as well.

**Critical behavior on random geometric graphs.** A random geometric graph is obtained by throwing $n$ points uniformly at random in the d-dimensional box $[0,1]^d$, and creating an edge between two points if their euclidean distance is at most $\lambda n^{-d}$. Random geometric graphs are known to exhibit phase transition as $\lambda$ increases [141]. The phase transition has also been studied for random geometric graphs on hyperbolic spaces [39]. However, analyzing the critical behavior on random geometric graphs is an open question. The inherent structure of these graphs are fundamentally different than the random graph models that do not depend on an underlying geometry. For example, the probability that a random vertex is involved in a clique of size $k$ is bounded away from zero for each fixed $k \geqslant 1$, showing that random geometric graphs cannot be approximated by a branching process locally. Thus, the critical components are not expected to have $O(1)$ many surplus edges anymore. Technically, it is challenging to deal with the exploration process since the drift and the quadratic variation terms depend on the area covered by the spheres centered at the active vertices, which is difficult to track.

**Concentration of total size and total number of large components.** Let $\Xi_n$ denote the point process $\{n^{-\rho}|\mathscr{C}_{(i)}(p_c(\lambda))|\}_{i \geqslant 1}$, and let $\Xi$ be the point process $\{|\gamma_i|\}_{i \geqslant 1}$, where $|\gamma_i|$ denotes the scaling limit of $|\mathscr{C}_{(i)}(p_c(\lambda))|$ in Theorems 2.2.1, or 3.2.1. Fix $\varepsilon > 0$. Then the convergence of the exploration processes imply that as $n \to \infty$

$$Z_{n,\varepsilon} := \int_\varepsilon^\infty x \, \Xi_n(dx) \xrightarrow{d} Z_\varepsilon := \int_\varepsilon^\infty x \, \Xi(dx)$$

$$W_{n,\varepsilon} := \Xi_n([\varepsilon,\infty)) \xrightarrow{d} W_\varepsilon := \Xi([\varepsilon,\infty));$$

see [117, Proposition 1.4]. As $\varepsilon \to 0$, $Z_\varepsilon$ gives the *total mass* of the largest components, and $W_\varepsilon$ gives the number of largest components. Thus it is desirable to understand the asymptotics of $Z_\varepsilon$ and $W_\varepsilon$ as $\varepsilon \to 0$. Note that $Z_0 = \infty$ and $W_0 = \infty$, and thus by monotone convergence theorem, $Z_\varepsilon \xrightarrow{\mathbb{P}} \infty$ and $W_\varepsilon \xrightarrow{\mathbb{P}} \infty$, as $\varepsilon \to 0$. For the Erdős-Rényi universality class, Janson and Spencer [117] showed that as $\varepsilon \to 0$

$$\varepsilon^{1/2} Z_\varepsilon \xrightarrow{\mathbb{P}} \sqrt{\frac{2}{\pi}} \quad \text{and} \quad \varepsilon^{3/2} W_\varepsilon \xrightarrow{\mathbb{P}} \sqrt{\frac{2}{9\pi}}. \tag{7.5}$$

Thus, even if both $Z_\varepsilon$ and $W_\varepsilon$ are non-degenerate random variables, these concentrate as $\varepsilon \to 0$. The asymptotic normality is still an open question. Also, it will be interesting to derive the asymptotics (7.5) for the general description of the multiplicative coalescent given in [11].

## 7.2   Infinite second moment case

In this section, we state the open problems related to the critical behavior in the infinite third moment case.

**Barely super-critical regime and the large $\lambda$ case.**   While the critical behavior was studied in detail for the configuration model in Chapter 6, the lack of an exploration process approach under the single-edge constraint limited our analysis to the barely subcritical regime and small values of $\lambda$ (i.e., $\lambda \in (0, \lambda_0)$ for some constant $\lambda_0$ which is independent of the model) within the critical window. In a future work, we wish to address the $\lambda > \lambda_0$ and the barely supercritical case. This will complete the analysis for the critical behavior of the component sizes within under the single-edge constraint.

**Novel evolution dynamics in the $\tau \in (2,3)$ case.**   In Chapters 2 and 3, we have seen that the evolution of the component sizes over the critical window is always guided by the multiplicative coalescent process, but apparently one would get a completely different coalescent process in the infinite second moment case, especially under the single-edge constraint. In this case, components merge when the hubs get connected via some intermediate vertex. Now the evolution of the total weights of is not Markovian. One has to keep track of all the hubs within components (which gives rise to an infinite dimensional vector) rather than some statistic of the components (like the total mass) to describe the process. This gives rise to novel evolution dynamics in the context of critical random graphs.

**Bounds on the diameter in the infinite second moment case.**    Although the diameter of critical components is $O(1)$ for $CM_n(\mathbf{d})$, we did not derive such a result under the single-edge constraint. In fact it is possible that the diameters in the latter case is diverging to infinity. We do not have a concrete intuition for this problem so far and it requires further investigation.

**Uniformly chosen graphs with given degrees.**    It will be interesting to study the critical behavior in the infinite second moment case for uniformly chosen graphs with given degrees (denoted by $UM_n(\mathbf{d})$). We expect the same scaling critical exponents as under the single edge-constraint, and the scaling limit is expected to be the same as critical percolation on generalized random graphs. The reason behind this is as follows: Suppose that $\mathbf{D} := (D_i)_{i \in [n]}$ denotes the degree sequence of $GRG_n(\mathbf{w})$. Conditionally on $\mathbf{D} = \mathbf{d}$, the distribution of $GRG_n(\mathbf{w})$ is same as that of $UM_n(\mathbf{d})$. Now the "core" of the components consists of vertices with weight $\Theta(n^\rho)$ and $\Theta(n^\alpha)$. Call these special vertices. One can probably use concentration arguments to show that $D_i \approx d_i$ for all special vertices. Now if one can show that perturbing the degrees of the special vertices does not change the connection probabilities in $UM_n(\mathbf{d})$ significantly, then it will be possible to show that, conditionally on $\mathbf{D} = \mathbf{d}$, the core for the connected components of $GRG_n(\mathbf{w})$ is the same graph as the core of the connected components of $UM_n(\mathbf{d})$. However, formalizing this is not straightforward.

## 7.3    Global structure.

**Joint convergence of metric spaces in $\mathfrak{l}^4$ topology.**    Let $\mathcal{M}$ denote the space of measured compact metric spaces endowed with the Gromov-Hausdorff-Prokhorov (GHP) topology. In Chapters 4 and 5, we have considered the product topology on the $\mathcal{M}^{\mathbb{N}}$ for the joint convergence of the components. A stronger result was shown for Erdős-Rényi random graphs in [3] under the metric defined below. Define the metric $d_{\mathfrak{l}^4}$ on $\mathcal{M}^{\mathbb{N}}$ by

$$d_{\mathfrak{l}^4}(\mathbf{X}, \mathbf{Y}) = \left( \sum_{i \geqslant 1} \left( d_{\mathrm{GHP}}(X_i, Y_i) \right)^4 \right)^{1/4},$$

where $d_{\mathrm{GHP}}$ denotes the GHP distance. It is desirable to extend the results in Chapters 4 and 5 under this stronger topology which requires suitable bounds on the diameter of small components.

**Evolution as metric space-valued stochastic process.**    The augmented multiplicative coalescent process only tracks the evolution of the component sizes and the surplus edges as $\lambda$ increases over the critical window. The metric structures of the largest connected components also evolve as $\lambda$ increases. It will be interesting to describe the evolution of the infinite dimensional measured metric spaces.

## 7.4    More challenges.

**Critical behavior for general graphs.**    Studying the critical behavior for more general sequence of graphs is an open direction. Of course the phrase "general graphs" is too vague, and one must impose regularity conditions to see the critical behavior and scaling limits. For example, one may consider percolation on sequences of dense graphs (with $\Theta(n^2)$ many edges) that converge in the so-called cut metric [49, 50] and impose some restrictions on the limiting graphon. The critical value for the phase transition was identified in [43] under mild conditions, while the critical behavior is a completely open question.

**The minimum spanning tree problem.**    The study of critical percolation has experienced a renewed interest after a recent seminal work by Addario-Berry et al. [4]. They studied the limit as a metric space of the minimum spanning tree (MST) on a complete graph with iid edge weights under the GHP-topology. Exploiting the relation between Kruskal's algorithm for generating MST, they showed that the MST can be *approximated* by $\mathscr{C}_{(1)}(p_c(\lambda))$ after removing the cycles in a specified manner. Using the results about the limit of $\mathscr{C}_{(1)}(p_c(\lambda))$ from [3], they could describe the metric structure of the MST on complete graph. The results in [36] and in Chapter 4 forms a basis of studying MST in the heavy-tailed regime where the scaling limit is expected to be different than MST on complete graph. However, the study of minimal spanning trees is an open question, even for the simple models like random regular graphs.

## 7.A    Proof of Theorem 7.1.1

Note that $(E_K(t))_{t \geqslant 0}$ is a super-martingale with the Doob-Meyer decomposition ([11, Section 2.1])

$$E_K(t) = M_K(t) + A_K(t),$$

where $A_K(t) = -\sum_{i=K+1}^{\infty} \theta_i^2(t - \xi_i)^+$ and the quadratic variation of $M_K$ is given by

$$\langle M_K \rangle(t) = \sum_{i=K+1}^{\infty} \theta_i^3 \min\{t, \xi_i\}.$$

For any fixed $T > 0$, we show that $\sigma_K^{-1} \sup_{t \leqslant T} |A_K(t)| \xrightarrow{P} 0$ and use the martingale functional central limit theorem [151, Theorem 2.1] to conclude the theorem. To see the first part, note that

$$\mathbb{E}\left[(T - \xi_i)^+\right] \leqslant T\mathbb{P}\left(\xi_i \leqslant T\right) = T(1 - e^{-\theta_i T}) \leqslant \theta_i T^2.$$

Noting that $A_K(\cdot)$ is negative and monotonically decreasing we have, for any $\varepsilon > 0$,

$$\mathbb{P}\left(\sup_{t \leqslant T} |A_K(t)| > \varepsilon \sigma_K\right) = \mathbb{P}\left(|A_K(T)| > \varepsilon \sigma_K\right) \leqslant \frac{1}{\varepsilon \sigma_K} \mathbb{E}\left[-A_K(T)\right]$$

$$\leqslant \frac{T^2 \sigma_K^2}{\varepsilon \sigma_K} = \frac{T^2 \sigma_K}{\varepsilon} \to 0.$$

Let $J(M_K, T)$ denote the value of the maximum jump of $M_K$ before time $T$. Now to show that the martingale parts converge to a Brownian motion, it is enough to show that, as $K \to \infty$,

(1) $\sigma_K^{-2} \langle M_K \rangle(t) \xrightarrow{P} t$, for each fixed $t > 0$,

(2) $\sigma_K^{-1} \mathbb{E}\left[|J(M_K, T)|\right] \to 0$.

Let us first verify (2). Note that $J(M_K, T) = \theta_i^2$ if $\xi_i \leqslant T$ and $\xi_j > T$ for all $K < j < i$ (since $\theta_i$'s are non-increasing). Therefore,

$$\sigma_K^{-1} \mathbb{E}\left[|J(M_K, T)|\right] \leqslant \sigma_K^{-1} \sum_{i=K+1}^{\infty} \theta_i^2 \mathbb{P}\left(\xi_i \leqslant T\right) \leqslant T\sigma_K \to 0.$$

To see (1) we will use Chebyshev's inequality. Note that

$$\mathbb{E}\left[\min\{t, \xi_i\}\right] = \int_0^t \theta_i x e^{-\theta_i x} dx + t\mathbb{P}\left(\xi_i > t\right) = \frac{1 - e^{-\theta_i t}}{\theta_i}.$$

Using the fact that $x - x^2/2 \leqslant 1 - e^{-x} \leqslant x$, we compute

$$t - \frac{t^2}{2} \frac{\sum_{i=K+1}^{\infty} \theta_i^4}{\sum_{i=K+1}^{\infty} \theta_i^3} \leqslant \sigma_K^{-2} \mathbb{E}\left[\langle M_K \rangle(t)\right] = \sigma_K^{-2} \sum_{i=K+1}^{\infty} \theta_i^2(1 - e^{-\theta_i t}) \leqslant t,$$

and therefore,

$$\left| \sigma_K^{-2} \mathbb{E}\left[\langle M_K \rangle(t)\right] - t \right| \leqslant \frac{t^2}{2} \frac{\sum_{i=K+1}^{\infty} \theta_i^4}{\sum_{i=K+1}^{\infty} \theta_i^3} \leqslant \frac{t^2}{2} \max_{i>K} \theta_i \to 0.$$

Moreover, since $\mathrm{Var}\left(\min\{t, \xi_i\}\right) \leqslant Ct^2 \theta_i^{-2}$ for some uniform constant $C > 0$,

$$\mathrm{Var}\left(\sigma_K^{-2} \langle M_K \rangle(t)\right) = \sigma_K^{-4} \sum_{i=K+1}^{\infty} \theta_i^6 \mathrm{Var}\left(\min\{t, \xi_i\}\right) \leqslant Ct^2 \frac{\max_{i>K} \theta_i}{\sum_{i=K+1}^{\infty} \theta_i^3} \to 0,$$

where the last part follows by our assumption. Thus, (1) follows and the proof is complete. $\qquad\square$

# Bibliography

[1] Addario-Berry, L. (2018). Most trees are short and fat. *Probab. Theory Relat. Fields*, pages 1–26.

[2] Addario-Berry, L., Bhamidi, S., and Sen, S. (2017). A probabilistic approach to the leader problem in random graphs. *arXiv:1703.09908*.

[3] Addario-Berry, L., Broutin, N., and Goldschmidt, C. (2012). The continuum limit of critical random graphs. *Probab. Theory Relat. Fields*, 152(3):367–406.

[4] Addario-Berry, L., Broutin, N., Goldschmidt, C., and Miermont, G. (2013). The scaling limit of the minimum spanning tree of the complete graph. *Ann. Probab.*, 45(5):3075–3144.

[5] Aïdékon, E., van der Hofstad, R., Kliem, S., and van Leeuwaarden, J. S. H. (2016). Large deviations for power-law thinned Lévy processes. *Stoch. Proc. Appl.*, 126(5):1353–1384.

[6] Albert, R. and Barabási, A.-L. (2002). Statistical mechanics of complex networks. *Rev Mod. Phys.*, 74(1):47–97.

[7] Albert, R., Jeong, H., and Barabási, A.-L. (2000). Error and attack tolerance of complex networks. *Nature*, 406:378.

[8] Aldous, D. (1997). Brownian excursions, critical random graphs and the multiplicative coalescent. *Ann. Probab.*, 25(2):812–854.

[9] Aldous, D. (1999). Deterministic and stochastic models for coalescence (aggregation and coagulation): A review of the mean-field theory for probabilists. *Bernoulli*, 5(1):3–48.

[10] Aldous, D. (2016). The incipient giant component in bond percolation on general finite weighted graphs. *Electron. Commun. Probab.*, 21:1–9.

[11] Aldous, D. and Limic, V. (1998). The entrance boundary of the multiplicative coalescent. *Electron. J. Probab.*, 3(3):1–59.

[12] Aldous, D., Miermont, G., and Pitman, J. (2004). The exploration process of inhomogeneous continuum random trees, and an extension of Jeulin's

local time identity. *Probab. Theory Relat. Fields*, 129(2):182–218.

[13] Aldous, D. and Pitman, J. (1999). A family of random trees with random edge lengths. *Random Struct. Algor.*, 15(2):176–195.

[14] Aldous, D. and Pitman, J. (2000). Inhomogeneous continuum random trees and the entrance boundary of the additive coalescent. *Probab. Theory Relat. Fields*, 118(4):455–482.

[15] Aldous, D. and Pittel, B. (2000). On a random graph with immigrating vertices: emergence of the giant component. *Random Struct. Algor.*, 17(2):79–102.

[16] Aldous, D. and Steele, J. M. (2004). The objective method: probabilistic combinatorial optimization and local weak convergence. In Kesten, H., editor, *Probability on Discrete Structures*, pages 1–72. Springer Berlin Heidelberg, Berlin, Heidelberg.

[17] Alon, N., Benjamini, I., and Stacey, A. (2004). Percolation on finite graphs and isoperimetric inequalities. *Ann. Probab.*, 32(3):1727–1745.

[18] Angel, O., van der Hofstad, R., and Holmgren, C. (2016). Limit laws for self-loops and multiple edges in the configuration model. *arXiv:1603.07172*.

[19] Asmussen, S. and Rosinski, J. (2001). Approximations of Small Jumps of Levy Processes with a View Towards Simulation. *J. Appl. Probab.*, 38(2):482–493.

[20] Athreya, S., Löhr, W., and Winter, A. (2016). The gap between Gromov-vague and Gromov–Hausdorff-vague topology. *Stoch. Proc. Appl.*, 126(9):2527–2553.

[21] Barabási, A.-L. (2016). *Network Science*. Cambridge University Press.

[22] Bender, E. A. and Canfield, E. R. (1978). The asymptotic number of labeled graphs with given degree sequences. *J. Combin. Theory Ser. A*, 24(3):296–307.

[23] Benjamini, I., Boucheron, S., Lugosi, G., and Rossignol, R. (2012). Sharp threshold for percolation on expanders. *Ann. Probab.*, 40(1):130–145.

[24] Benjamini, I., Nachmias, A., and Peres, Y. (2011). Is the critical percolation probability local? *Probab. Theory Relat. Fields*, 149(1):261–269.

[25] Benjamini, I. and Schramm, O. (2001). Recurrence of distributional limits of finite planar graphs. *Electron. J. Probab.*, 6:1–13.

[26] Berger, N., Borgs, C., Chayes, J. T., and Saberi, A. (2014). Asymptotic behavior and distributional limits of preferential attachment graphs. *Ann. Probab.*, 42(1):1–40.

[27] Beringer, D., Pete, G., and Timár, Á. (2017). On percolation critical probabilities and unimodular random graphs. *Electron. J. Probab.*, 22:1–26.

[28] Bertoin, J. (1996). *Lévy Processes*. Cambridge University Press.

[29] Bhamidi, S., Broutin, N., Sen, S., and Wang, X. (2014a). Scaling limits of random graph models at criticality: Universality and the basin of attraction of the Erdős-Rényi random graph. *arXiv:1411.3417*.

[30] Bhamidi, S., Budhiraja, A., and Wang, X. (2014b). The augmented multiplicative coalescent and critical dynamic random graph models. *Probab. Theory Relat. Fields*, 160(3):733–796.

[31] Bhamidi, S., Dhara, S., van der Hofstad, R., and Sen, S. (2017a). Universality for critical heavy-tailed random graphs: Metric structure of maximal components. *arXiv:1703.07145*.

[32] Bhamidi, S., van der Hofstad, R., and van Leeuwaarden, J. S. H. (2010). Scaling limits for critical inhomogeneous random graphs with finite third moments. *Electron. J. Probab.*, 15(6):1682–1702.

[33] Bhamidi, S., van der Hofstad, R., and van Leeuwaarden, J. S. H. (2012). Novel scaling limits for critical inhomogeneous random graphs. *Ann. Probab.*, 40(6):2299–2361.

[34] Bhamidi, S. and Sen, S. (2016). Geometry of the vacant set left by random walk on random graphs, Wright's constants, and critical random graphs with prescribed degrees. *arXiv:1608.07153*.

[35] Bhamidi, S., Sen, S., and Wang, X. (2017b). Continuum limit of critical inhomogeneous random graphs. *Probab. Theory Relat. Fields*, 169(1):565–641.

[36] Bhamidi, S., van der Hofstad, R., and Sen, S. (2018). The multiplicative coalescent, inhomogeneous continuum random trees, and new universality classes for critical random graphs. *Probab. Theory Relat. Fields*, 170(1):387–474.

[37] Billingsley, P. (1999). *Convergence of Probability Measures*. John Wiley & Sons, Inc.

[38] Bingham, N. H., Goldie, C. M., and Teugels, J. L. (1989). *Regular Variation*. Cambridge University Press.

[39] Bode, M., Fountoulakis, N., and Müller, T. (2015). On the largest component of a hyperbolic model of complex networks. *Electr. J. Comb.*, 22(3):P3.24.

[40] Bollobás, B. (1980). A probabilistic proof of an asymptotic formula for the number of labelled regular graphs. *European J. Combin.*, 1(4):311–316.

[41] Bollobás, B. (1984). The evolution of random graphs. *Trans. Amer. Math. Soc.*, 286(1):257.

[42] Bollobás, B. (2001). *Random Graphs*. Cambridge University Press.

[43] Bollobás, B., Borgs, C., Chayes, J., and Riordan, O. (2010). Percolation on dense graph sequences. *Ann. Probab.*, 38(1):150–183.

[44] Bollobás, B., Janson, S., and Riordan, O. (2007). The phase transition in inhomogeneous random graphs. *Random Struct. Algor.*, 31(1):3–122.

[45] Bollobás, B. and Riordan, O. (2003). Robustness and vulnerability of scale-free random graphs. *Internet Math.*, 1(1):1–35.

[46] Bollobás*, B. and Riordan, O. (2004). The diameter of a scale-free random-graph. *Combinatorica*, 24(1):5–34.

[47] Bollobás, B. and Riordan, O. (2006). *Percolation*. Cambridge University Press.

[48] Borgs, C., Chayes, J. T., van der Hofstad, R., Slade, G., and Spencer, J. (2005). Random subgraphs of finite graphs: I. The scaling window under the triangle condition. *Random Struct. Algor.*, 27(2):137–184.

[49] Borgs, C., Chayes, J. T., Lovász, L., Sós, V. T., and Vesztergombi, K. (2008). Convergent sequences of dense graphs I: Subgraph frequencies, metric properties and testing. *Adv. Math.*, 219(6):1801–1851.

[50] Borgs, C., Chayes, J. T., Lovász, L., Sós, V. T., and Vesztergombi, K. (2012). Convergent sequences of dense graphs II. Multiway cuts and statistical physics. *Ann. Math.*, 176(1):151–219.

[51] Braunstein, L. A., Buldyrev, S. V., Cohen, R., Havlin, S., and Stanley, H. E. (2003). Optimal paths in disordered complex networks. *Phys. Rev. Lett.*, 91(16):168701.

[52] Braunstein, L. A., Wu, Z., Chen, Y., Buldyrev, S. V., Kalisky, T., Sreenivasan, S., Cohen, R., López, E., Havlin, S., and Stanley, H. E. (2007). Optimal path and minimal spanning trees in random weighted networks. *Int. J. Bifurc. Chaos*, 17(07):2215–2255.

[53] Breiman, L. (1968). *Probability*. SIAM: Society for Industrial and Applied Mathematics.

[54] Britton, T., Deijfen, M., and Martin-Löf, A. (2006). Generating simple random graphs with prescribed degree distribution. *J. Stat. Phys.*, 124(6):1377–1397.

[55] Broido, A. D. and Clauset, A. (2018). Scale-free networks are rare. *arXiv:1801.03400*.

[56] Broutin, N., Duquesne, T., and Wang, M. (2018). Limits of multiplicative

inhomogeneous random graphs and Lévy trees. *arXiv:1804.05871*.

[57] Broutin, N. and Marckert, J.-F. (2016). A new encoding of coalescent processes. Applications to the additive and multiplicative cases. *Probab. Theory Relat. Fields*, 166(1):515–552.

[58] Callaway, D. S., Newman, M. E. J., Strogatz, S. H., and Watts, D. J. (2000). Network robustness and fragility: Percolation on random graphs. *Phys. Rev. Lett.*, 85:5468–5471.

[59] Camarri, M. and Pitman, J. (2000). Limit distributions and random trees derived from the birthday problem with unequal probabilities. *Electron. J. Probab.*, 5(2):1–18.

[60] Chung, F., Horn, P., and Lu, L. (2009). Percolation in general graphs. *Internet Math.*, 6(3):331–347.

[61] Chung, F. and Lu, L. (2002a). Connected components in random graphs with given expected degree sequences. *Ann. Comb.*, 6(2):125–145.

[62] Chung, F. and Lu, L. (2002b). The average distances in random graphs with given expected degrees. *Proc. Natl. Acad. Sci. USA*, 99(25):15879–82.

[63] Cohen, R., ben Avraham, D., and Havlin, S. (2002). Percolation critical exponents in scale-free networks. *Phys. Rev. E*, 66(3):36113.

[64] Cohen, R., Erez, K., ben Avraham, D., and Havlin, S. (2000). Resilience of the internet to random breakdowns. *Phys. Rev. Lett.*, 85:4626–4628.

[65] Conchon-Kerjan, G. and Goldschmidt, C. (2017). Stable graphs: the metric space scaling limits of critical random graphs with i.i.d. power-law degrees. *In Preparation*.

[66] Dembo, A., Levit, A., and Vadlamani, S. (2014). Component sizes for large quantum Erdős Rényi graph near criticality. *arXiv:1404.5705*.

[67] Dhara, S., van der Hofstad, R., and Mandjes, M. (2015). Personal communication.

[68] Dhara, S., van der Hofstad, R., and van Leeuwaarden, J. S. H. (2018). Critical percolation on scale-free random graphs: effect of the single-edge constraint. *Ongoing work*.

[69] Dhara, S., van der Hofstad, R., van Leeuwaarden, J. S. H., and Sen, S. (2016). Heavy-tailed configuration models at criticality. *arXiv:1612.00650*.

[70] Dhara, S., van der Hofstad, R., van Leeuwaarden, J. S. H., and Sen, S. (2017a). Critical window for the configuration model: finite third moment degrees. *Electron. J. Probab.*, 22(16):1–33.

[71] Dhara, S., Mukherjee, D., and Sen, S. (2017b). Phase transitions of extremal cuts for the configuration model. *Electron. J. Probab.*, 22:1–29.

[72] Ding, J., Kim, J. H., Lubetzky, E., and Peres, Y. (2010). Anatomy of a young giant component in the random graph. *Random Struct. Algor.*, 39(2):139–178.

[73] Dommers, S., van der Hofstad, R., and Hooghiemstra, G. (2010). Diameters in Preferential Attachment Models. *J. Stat. Phys.*, 139(1):72–107.

[74] Dorogovtsev, S. N., Goltsev, A. V., and Mendes, J. F. F. (2008). Critical phenomena in complex networks. *Rev. Mod. Phys.*, 80:1275–1335.

[75] Duminil-Copin, H. and Tassion, V. (2017). A note on Schramm's locality conjecture for random-cluster models. *arXiv:1707.07626*.

[76] Duquesne, T. and Le Gall, J.-F. (2002). *Random trees, Lévy processes and spatial branching processes*, volume 281. Société mathématique de France.

[77] Durrett, R. (2010). *Random graph dynamics*. Cambridge University Press.

[78] Erdős, P. and Rényi, A. (1959). On random graphs: I. *Publicationes Mathematicae*, 6:290–297.

[79] Erdős, P. and Rényi, A. (1960). On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci.*, pages 17–61.

[80] Ethier, S. N. and Kurtz, T. G. (1986). *Markov Processes: Characterization and Convergence*. John Wiley & Sons, Inc., New Jersey.

[81] Evans, S. N. (2006). Probability and real trees. *Ecole d'Eté de Probabilités de Saint-Flour XXXV-2005*.

[82] Federico, L., van der Hofstad, R., den Hollander, F., and Hulshof, T. (2017). Expansion of percolation critical points for Hamming graphs. *arXiv:1701.02099*.

[83] Federico, L., van der Hofstad, R., den Hollander, F., and Hulshof, T. (2018). The scaling limit for critical percolation on the Hamming graph. *In preparation*.

[84] Feller, W. (1991). *An Introduction to Probability Theory and Its Applications: Volume 2*. John Wiley & Sons, Inc.

[85] Fountoulakis, N. (2007). Percolation on sparse random graphs with given degree sequence. *Internet Math.*, 4(1):329–356.

[86] Fountoulakis, N., Joos, F., and Perarnau, G. (2016). Percolation on random graphs with a fixed degree sequence. *arXiv:1611.08496*.

[87] Freedman, D. A. (1975). On tail probabilities for martingales. *Ann. Probab.*, 3(1):100–118.

[88] G.H. Hardy (1949). *Divergent Series*. Clarendon (Oxford University) Press, Oxford.

[89] Gilbert, E. N. (1959). Random graphs. *Ann. Math. Statist.*, 30(4):1141–

1144.

[90] Goldschmidt, C., Haas, B., and Sénizergues, D. (2017). Stable graphs: distributions and line-breaking constructions. *In Preparation*.

[91] Greven, A., Pfaffelhuber, P., and Winter, A. (2009). Convergence in distribution of random metric measure spaces (λ-coalescent measure trees). *Probab. Theory Relat. Fields*, 145(1):285–322.

[92] Grimmett, G. R. (1999). *Percolation*. Springer-Verlag Berlin Heidelberg.

[93] Gromov, M.L. (2007). *Metric Structures for Riemannian and Non-Riemannian Spaces*. Translated by Bates, S.M. Birkhǎ user Basel.

[94] Hatami, H. and Molloy, M. (2012). The scaling window for a random graph with a given degree sequence. *Random Struct. Algor.*, 41(1):99–123.

[95] Havlin, S., Braunstein, L. A., Buldyrev, S. V., Cohen, R., Kalisky, T., Sreenivasan, S., and Eugene Stanley, H. (2005). Optimal path in random networks with disorder: A mini review. *Physica A: Statistical Mechanics and its Applications*, 346(1-2):82–92.

[96] Heydenreich, M. and van der Hofstad, R. (2017). *Progress in High-Dimensional Percolation and Random Graphs*. Springer International Publishing Switzerland.

[97] van der Hofstad, R. (2013). Critical behavior in inhomogeneous random graphs. *Random Struct. Algor.*, 42:480–508.

[98] van der Hofstad, R. (2016). *Random Graphs and Complex Networks*, volume 1. Cambridge university press.

[99] van der Hofstad, R. (2017). *Stochastic Processes on Random Graphs*. Lecture notes for the 47th Summer School in Probability Saint-Flour 2017.

[100] van der Hofstad, R. (2018). *Random Graphs and Complex Networks*, volume 2.

[101] van der Hofstad, R., Janson, S., and Luczak, M. (2016a). Component structure of the configuration model: barely supercritical case. *arXiv:1611.05728*.

[102] van der Hofstad, R., Kliem, S., and van Leeuwaarden, J. S. H. (2018). Cluster tails for critical power-law inhomogeneous random graphs. *J. Stat. Phys.*, 171(1):38–95.

[103] van der Hofstad, R. and Nachmias, A. (2017). Hypercube percolation. *J. Eur. Math. Soc.*, 19(3):725–814.

[104] van der Hofstad, R. and Slade, G. (2005). Asymptotic expansions in $n^{-1}$ for percolation critical values on the $n$-Cube and $\mathbb{Z}^n$. *Random Struct. Algor.*, 27(3):331–357.

[105] van der Hofstad, R., van der Hoorn, P., Litvak, N., and Stegehuis, C. (2017). Limit theorems for assortativity and clustering in the configuration model with scale-free degrees. *arxiv:1712.08097*.

[106] van der Hofstad, R., van Leeuwaarden, J. S. H., and Stegehuis, C. (2016b). Mesoscopic scales in hierarchical configuration models. *To appear Stoch. Proc. Appl.*

[107] Jacod, J. and Shiryaev, A. N. (2003). *Limit Theorems for Stochastic Processes*. Grundlehren der mathematischen Wissenschaften. Springer-Verlag Berlin Heidelberg.

[108] Janson, S. (2008). The largest component in a subcritical random graph with a power law degree distribution. *Ann. Appl. Probab.*, 18(4):1651–1668.

[109] Janson, S. (2009a). On percolation in random graphs with given vertex degrees. *Electron. J. Probab.*, 14:87–118.

[110] Janson, S. (2009b). The probability that a random multigraph is simple. *Comb. Probab. Comp.*, 18(1-2):205–225.

[111] Janson, S. (2010). Susceptibility of random graphs with given vertex degrees. *J. Combin.*, 1(3-4):357–387.

[112] Janson, S., Knuth, D. E., Łuczak, T., and Pittel, B. (1993). The birth of the giant component. *Random Struct. Algor.*, 4(3):233–358.

[113] Janson, S. and Luczak, M. J. (2008). Susceptibility in subcritical random graphs. *J. Math. Phys.*, 49(12):125207.

[114] Janson, S. and Luczak, M. J. (2009). A new approach to the giant component problem. *Random Struct. Algor.*, 34(2):197–216.

[115] Janson, S., Łuczak, T., and Rucinski, A. (2000). *Random Graphs.* Wiley, New York.

[116] Janson, S. and Riordan, O. (2012). Susceptibility in inhomogeneous random graphs. *Electron. J. Combin*, 19(1):1–59.

[117] Janson, S. and Spencer, J. (2007). A point process describing the component sizes in the critical window of the random graph evolution. *Comb. Probab. Comput.*, 16(4):631–658.

[118] Janson, S. and Warnke, L. (2018). On the critical probability in percolation. *Electron. J. Probab.*, 23:1–25.

[119] Joos, F. and Perarnau, G. (2018). Critical percolation on random regular graphs. *Proc. Amer. Math. Soc.*, 146:3321–3332.

[120] Joos, F., Perarnau, G., Rautenbach, D., and Reed, B. (2018). How to determine if a random graph with a fixed degree sequence has a giant component. *Probab. Theory Relat. Fields*, 170(1):263–310.

[121] Joseph, A. (2014). The component sizes of a critical random graph with given degree sequence. *Ann. Appl. Probab.*, 24(6):2560–2594.

[122] Kallenberg, O. (1997). *Foundations of Modern Probability*. Springer, New York.

[123] Kang, M. and Seierstad, T. G. (2008). The critical phase for random graphs with a given degree sequence. *Comb. Probab. Comp.*, 17(01):67–86.

[124] Karatzas, I. and Shreve, S. E. (1991). *Brownian Motion and Stochastic Calculus*, volume 113 of *Graduate Texts in Mathematics*. Springer-Verlag, New York.

[125] Kesten, H. (1982). Percolation Theory for Mathematicians. In *Progress in Probability and Statistics*. Birkhäuser, Boston.

[126] Kortchemski, I. (2017). Sub-exponential tail bounds for conditioned stable Bienaymé–Galton–Watson trees. *Probab. Theory Relat. Fields*, 168(1):1–40.

[127] Le Gall, J. F. (2005). Random trees and applications. *Probab. Surveys*, 2:245–311.

[128] Lipster, R. S. and Shiryayev, A. N. (1989). *Theory of Martingales*. Springer, Dordrecht.

[129] Łuczak, T. (1990). Component behavior near the critical point of the random graph process. *Random Struct. Algor.*, 1(3):287–310.

[130] Łuczak, T., Pittel, B., and Wierman, J. C. (1994). The structure of a random graph at the point of the phase transition. *Trans. Amer. Math. Soc.*, 341:721–748.

[131] Martin, R. H. (1986). *Nonlinear operators and differential equations in Banach spaces*. Krieger Publishing Co., Inc., Melbourne, FL.

[132] Massart, P. (1990). The tight constant in the dvoretzky-kiefer-wolfowitz inequality. *Ann. Probab.*, 18(3):1269–1283.

[133] Molloy, M. and Reed, B. (1995). A critical-point for random graphs with a given degree sequence. *Random Struct. Algor.*, 6(2-3):161–179.

[134] Nachmias, A. and Peres, Y. (2008). Critical random graphs: Diameter and mixing time. *Ann. Probab.*, 36(4):1267–1286.

[135] Nachmias, A. and Peres, Y. (2010a). Critical percolation on random regular graphs. *Random Struct. Algor.*, 36(2):111–148.

[136] Nachmias, A. and Peres, Y. (2010b). The critical random graph, with martingales. *Israel J. Math.*, 176(1):29–41.

[137] Newman, M. E. J. (2010). *Networks: An Introduction*. Oxford University Press.

[138] Norros, I. and Reittu, H. (2006). On a conditionally Poissonian graph process. *Adv. Appl. Probab.*, 38(1):59–75.

[139] Pang, G., Talreja, R., and Whitt, W. (2007). Martingale proofs of many-server heavy-traffic limits for Markovian queues. *Probab. Surveys*, 4:193–267.

[140] Pastor-Satorras, R. and Vespignani, A. (2007). *Evolution and structure of the Internet: A statistical physics approach*. Cambridge University Press.

[141] Penrose, M. (2003). *Random Geometric Graphs*. Oxford University Press.

[142] Pitman, J. (2001). Random mappings, forests, and subsets associated with abel-cayley-hurwitz multinomial expansions. *Séminaire Lotharingien de Combinatoire*, 46.

[143] Pittel, B. (2001). On the largest component of the random graph at a nearcritical stage. *J. Combin. Theory Ser. B*, 82(2):237–269.

[144] Riordan, O. (2012). The phase transition in the configuration model. *Comb. Probab. Comp.*, 21:265–299.

[145] Riordan, O. and Warnke, L. (2012). Achlioptas process phase transitions are continuous. *Ann. Appl. Probab.*, 22(4):1450–1464.

[146] Roberts, M. I. and Sengul, B. (2016). Exceptional times of the critical dynamical Erdős-Rényi graph. *To appear in Ann. Appl. Probab.*

[147] Rogers, L. C. G. and Williams, D. (1994). *Diffusions, Markov Processes, and Martingales.*, volume 1. John Wiley & Sons, Ltd., Chichester, 2nd edition.

[148] Scott, A. D. and Sorkin, G. B. (2006). Solving sparse random instances of Max Cut and Max 2-CSP in linear expected time. *Comb. Probab. Comput.*, 15(1-2):281–315.

[149] Shorack, G. R. and Wellner, J. A. (1986). *Empirical Processes with Applications to Statistics*. John Wiley & Sons.

[150] Whitt, W. (2002). *Stochastic-Process Limits: An Introduction to Stochastic-Process Limits and Their Application to Queues*. Springer-Verlag, New York.

[151] Whitt, W. (2007). Proofs of the martingale FCLT. *Probab. Surveys*, 4:269–302.

[152] Wormald, N. C. (1995). Differential equations for random processes and random graphs. *Ann. Appl. Probab.*, pages 1217–1235.

# Summary

Random graphs have played an instrumental role in modelling real-world networks arising from the internet topology, social networks, or even protein-interaction networks within cells. Percolation, on the other hand, has been the fundamental model for understanding robustness and spread of epidemics on these networks. From a mathematical perspective, percolation is the simplest model that exhibits phase transition, and fascinating features are observed around the critical point. In this thesis, we prove limit theorems about structural properties of the connected components obtained from percolation on random graphs at criticality. The results are obtained for random graphs with general degree sequence, and we identify different universality classes for the critical behavior based on moment assumptions on the degree distribution.

In Chapter 1, we start with an introduction to this attractive branch of probability which has spurred interest among mathematicians for several decades, with many of the interesting questions being still open. We briefly review the history of the percolation phase transition on finite graphs, and describe the emerging literature for the critical behavior. Subsequently, we describe our results from a high-level, and discuss the general proof ideas. Three types of fundamentally different critical behaviors are observed depending on whether degree distribution satisfies (a) a finite third moment condition, (b) an infinite third moment condition, (c) and an infinite second moment condition. In all these regimes, we ask questions about the component sizes and structures of the critical components. The goal of this chapter is to convey the main challenges in the upcoming chapters for each of the above regimes without going into the technical framework.

In Chapter 2, we state and prove results about the component sizes and surplus edges when the degree distribution satisfies a finite third moment condition. The evolution of component sizes and surplus edges over the critical window is also shown to converge to the augmented multiplicative coa-

lescent. The results show that only a finite third moment condition ensures that the critical behavior lies in the same universality class as classical homogeneous random graph models like Erdős-Rényi random graph or random regular graph.

In Chapter 3, we investigate the infinite third moment case. In this setting, the critical behavior for component sizes and surplus edges turns out to be in a completely different universality class. The key difference lies in the fact that the asymptotics of high degree vertices play a pivotal role in describing the scaling limits. For example, if the highest degree vertex is deleted, the scaling limit changes in this regime which is in sharp contrast to the finite third moment case. The results in Chapters 2 and 3 observe both the possible scaling limits for multiplicative coalescent processes that was predicted by Aldous and Limic.

In Chapter 4, we describe the global structure of components in the infinite third moment case. More precisely, one can view any connected graph as a metric space equipped with a measure, where the elements of the metric space given by the vertices, the metric given by the graph distance, and the measure proportional to the counting measure. With all the above ingredients, the critical components can be viewed as a random element from the space of all complete metric spaces equipped with a measure. In this chapter we show that after rescaling the distances suitably, the largest critical components converge with respect to Gromov-weak topology. These results yield joint convergence of several functionals related to distances within these connected components.

In Chapter 5, we establish the global lower mass bound property. This property ensures that the convergence results in Chapter 4 could be strengthened to hold under the Gromov-Hausdorff-Prokhorov topology. The primary outcome of the later stronger form of convergence is that it yields convergence of global distance related functionals like the diameter.

In Chapter 6, we investigate the case where the degree distribution has infinite second moment. Even defining the critical window for percolation is challenging in this case and all the questions related to the critical behavior in this regime were completely open question till date. We initiate this study by identifying critical values and scaling limits of the component sizes. The striking observation that we make in this regime is that the critical exponents and the scaling limits depend crucially on the so-called single-edge constraint, i.e., the critical behavior for the configuration model and the erased configuration model are fundamentally different. We also establish

the uniqueness of the critical exponents by analyzing the barely sub/super-critical regimes.

In the final Chapter, we conclude with many open problems and future directions.

The results in this thesis are strongest in terms of the topology of convergence and the results are proved under minimal assumptions. The results are expected to have potential impact on understanding spread of epidemics, minimum spanning trees on random networks with arbitrary degree sequence. The proof ideas are also robust and we hope that many of the core ideas would work for more many other random graph models.

# About the author

Souvik Dhara was born on May 9, 1991 in Kolkata, India. Souvik grew up in the southern part of Kolkata, and obtained his high school degree from Harinavi D.V.A.S. High School. During 2009 - 2012, he obtained his bachelor's degree with Statistics major. In 2012, he joined Indian Statistical Institute in the master's program, and obtained his M.Stat degree with Mathematical Statistics and Probability specialization in 2014. In August 2014, he joined a PhD program in the Eindhoven University of Technology under the supervision of Remco van der Hofstad and Johan van Leeuwaarden. His PhD project was part of the NETWORKS program funded by the Netherlands Organisation for Scientific Research (NWO).

Souvik's research interests lie in the intersection of probability theory and combinatorics, applied probability and operations research. During PhD, his primary aim has been understanding the interplay between the structural properties of networks and stochastic processes on them. Souvik has explored and applied several recent concepts in both probability theory and combinatorics related to local and global structures of random networks. His main contribution in the field is to derive limit laws for critical percolation on graphs with arbitrary degree distribution. His findings is the topic of this PhD thesis. At the same time, he has made contributions to combinatorics, and applied probability by analyzing graph limits and cut properties of random graphs, and providing asymptotic analysis of stochastic process arising from modern cloud computing systems and wireless networks.

In July 2018, Souvik joined as a Schramm fellow, which is offered jointly by Microsoft Research New England and MIT Mathematics. He will be at Microsoft Research during 2018 - 2019, and at MIT during 2019 - 2021.

# Publications and preprints

[1] Critical behavior of percolation on random graphs with given degree: A survey (2018+); Souvik Dhara, Remco van der Hofstad, Johan S.H. van Leeuwaarden. (Preprint)

[2] Limits of sparse configuration models and beyond: graphexes and multi-graphexes (2018+); Christian Borgs, Jennifer T. Chayes, Souvik Dhara, Subhabrata Sen. (Preprint)

[3] Critical percolation on scale-free random graphs: Effect of the single-edge constraint (2018+); Souvik Dhara, Remco van der Hofstad, Johan S.H. van Leeuwaarden. (Preprint)

[4] Global lower mass-bound for critical configuration models in the heavy-tailed regime (2018+); Shankar Bhamidi, Souvik Dhara, Remco van der Hofstad, Sanchayan Sen. (Preprint)

[5] Universality for critical heavy-tailed network models: Metric structure of maximal components (2017); Shankar Bhamidi, Souvik Dhara, Remco van der Hofstad, Sanchayan Sen. arXiv:1703.07145

[6] Heavy-tailed configuration models at criticality (2016); Souvik Dhara, Remco van der Hofstad, Johan S.H. van Leeuwaarden, Sanchayan Sen; arXiv:1612.00650

[7] Corrected mean-field model for random sequential adsorption on random geometric graphs (2016); Souvik Dhara, Johan S.H. van Leeuwaarden, Debankur Mukherjee; To appear with Journal of Statistical Physics.

[8] Phase transitions of extremal cuts for the configuration model (2017); Souvik Dhara, Debankur Mukherjee, Subhabrata Sen; Electronic Journal of Probability 22, no. 86, 1–29.

[9] Optimal Service Elasticity in Large-Scale Distributed Systems (2017); Debankur Mukherjee, Souvik Dhara, Sem Borst, Johan S.H. van Leeuwaarden; SIGMETRICS'17, Urbana-Champaign, Illinois, USA. Proceedings of the ACM on Measurement and Analysis of computing systems.

[10] Critical window for the configuration model: finite third moment degrees (2016); Souvik Dhara, Remco van der Hofstad, Johan S.H. van Leeuwaarden, Sanchayan Sen; Electronic Journal of Probability 22, no. 16, 1–33.

[11] Generalized random sequential adsorption on Erdos-Renyi random graphs (2016); Souvik Dhara, Johan S.H. van Leeuwaarden, Debankur Mukherjee; Journal of Statistical Physics 164, 1217-1232.