Heavy-tailed configuration models at criticality

Heavy-tailed configuration models at criticality

Souvik Dhara*, Remco van der Hofstad*, Johan S.H. van Leeuwaarden*, Sanchayan Sen[†]

*Department of Mathematics and Computer Science, Eindhoven University of Technology
[†] Department of Mathematics and Statistics, McGill University

December 14, 2016

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 $\tau - 1$, where $\tau \in (3, 4)$. 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 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 this paper, 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, which is shown to converge in finite dimension to the augmented multiplicative coalescent process introduced by Bhamidi et. al., *Probab. Theory Related Fields* (2014). The main results of this paper 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.

1 Introduction

Most random graph models posses a phase-transition property: there is a model-dependent parameter θ and a critical value θ_c such that whenever $\theta > \theta_c$, the largest component of the graph contains a positive proportion of vertices with high probability (w.h.p) and when $\theta \leq \theta_c$, the largest component is of smaller order than the size of the graph w.h.p The random graph is called *critical* when $\theta = \theta_c$. The study of critical random graphs started in the 1990s with the works of Bollobás [17], Łuczak [34], Janson et al. [28] and Aldous [4] for Erdős-Rényi random graphs. A large body of subsequent work in [10, 13, 20, 30, 36, 37, 43] showed that the behavior of a wide array of random graphs at criticality is universal in the sense that certain graph properties do not depend on the precise description of the model. One of these universal features is that the scaling

Correspondence to: S. Dhara.
Email addresses: s.dhara@tue.nl, r.w.v.d.hofstad@tue.nl, j.s.h.v.leeuwaarden@tue.nl, sanchayan.sen1@gmail.com
2010 *Mathematics Subject Classification.* Primary: 60C05, 05C80.
Keywords and phrase. Critical configuration model, heavy-tailed degree, thinned Lévy process, augmented multiplicative coalescent, universality, critical percolation

limit of the large component sizes, for many graph models, is identical to that of the Erdős-Rényi random graphs. All these universality results are obtained under the common assumption that the degree distribution is light-tailed, i.e., the asymptotic degree distribution has sufficiently large moments. For critical configuration models, a finite third-moment condition proves to be crucial [20]. However, empirical studies of real-world networks from various fields including physics, biology, computer science [3, 7, 21, 32, 39, 40] show that the degree distribution is heavy-tailed and of power-law type. A first work towards understanding the critical behavior in the *heavy-tailed* network models is [14], which showed that, for rank-one inhomogeneous random graphs, when the weight distribution follows a power-law with exponent $\tau \in (3,4)$, the component sizes and the scaling limits turn out to be quite different from that of the Erdős-Rényi random graph. This revealed an entirely new universality class for the phase transition of heavy-tailed random graphs. In this paper, we study the configuration model with heavy-tailed power-law degrees. The configuration model is the canonical model for generating a random *multi*-graph with a prescribed degree sequence. This model was introduced by Bollobás [16] to generate a uniform simple d-regular graph on n vertices, when dn is even. The idea was later generalized to general degree sequences by Molloy and Reed [35] and others. We will denote the multi-graph generated by the configuration model on the vertex set $[n] = \{1, \dots, n\}$ with the degree sequence $\boldsymbol{d}$ by $\mathrm{CM}_n(\boldsymbol{d})$. The configuration model, conditioned on simplicity, yields a uniform simple graph with the same degree sequence, which explains its popularity.

Our main contribution. Let D_n be the degree of a uniformly chosen vertex, independently of the random graph $\mathrm{CM}_n(\boldsymbol{d})$. The main goal of this paper is to obtain various asymptotic results for the component sizes of $\mathrm{CM}_n(\boldsymbol{d})$ when $\mathbb{P}\left(D_n \geq k\right) \sim L_0(k)/k^{\tau-1}$ for some $\tau \in (3,4)$ and $L_0(\cdot)$ a slowly-varying function (here $\sim$ denotes an unspecified approximation that will be defined in more detail below). In fact, under a general set of assumptions (see Assumptions 1 and 2), we prove the following:

(1) The largest connected components are 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 $L(\cdot)$.

(2) The joint distribution of the re-scaled component sizes and the surplus edges converges in distribution to a suitable limiting random vector under a strong topology. It turns out that 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, which is also the case in [14]. 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.

(3) The results hold conditioned on the graph being simple, thus solving [30, Conjecture 8.5] in this case.

(4) The scaling limits also hold for the graphs obtained by performing critical percolation on a supercritical graph. The percolation clusters can be coupled in a natural way using the Harris coupling. This enables us to study the *evolution* of the component sizes and the surplus edges as a dynamic process in the critical window. The evolution of the component sizes and surplus edges is shown to converge to a version of the *augmented multiplicative coalescent* process that was first introduced in [10]. In fact, 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.

Thus, this paper provides a detailed understanding about the critical component sizes and surplus edges for the heavy-tailed graphs in the critical window. Before stating our main results precisely, we introduce some notation and concepts.

1.1 The model

Consider n vertices labeled by $[n] := \{1, 2, ..., n\}$ and a non-increasing sequence of degrees $\boldsymbol{d} = (d_i)_{i\in[n]}$ such that $\ell_n = \sum_{i\in[n]} d_i$ is even. For notational convenience, we suppress the dependence of the degree sequence on n. The configuration model on n vertices having degree sequence $\boldsymbol{d}$ is constructed as follows:

> 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.

Note that the graph constructed by the above procedure may contain self-loops or multiple edges. It can be shown that conditionally on $\mathrm{CM}_n(\boldsymbol{d})$ being simple, the law of such graphs is uniform over all possible simple graphs with degree sequence $\boldsymbol{d}$ [41, Proposition 7.13]. It was further shown in [26] that, under very general assumptions, the asymptotic probability of the graph being simple is positive.

1.2 Definition and notation

We use the standard notation of $\xrightarrow{\mathbb{P}}$, and $\xrightarrow{d}$ to denote convergence in probability and in distribution, respectively. We often use the Bachmann–Landau notation $O(\cdot)$, $o(\cdot)$ for large n asymptotics of real numbers. The topology needed for the convergence in distribution will always be specified unless it is clear from the context. The notation $A_n \sim B_n$ will be used to say that $A_n/B_n \to 1$. We say that a sequence of events $(\mathcal{E}_n)_{n\geq 1}$ occurs with high probability (w.h.p) with respect to the probability measures $(\mathbb{P}_n)_{n\geq 1}$ when $\mathbb{P}_n\big(\mathcal{E}_n\big) \to 1$. Define $f_n = O_{\mathbb{P}}(g_n)$ when $(|f_n|/|g_n|)_{n\geq 1}$ is tight; $f_n = o_{\mathbb{P}}(g_n)$ when $f_n/g_n \xrightarrow{\mathbb{P}} 0$ whp; $f_n = \Theta_{\mathbb{P}}(g_n)$ if both $f_n = O_{\mathbb{P}}(g_n)$ and $g_n = O_{\mathbb{P}}(f_n)$. For a random variable X and a distribution F, we write $X \sim F$ to denote that X has distribution F. Denote

$$\ell^p_{\downarrow} := \Big\{\mathbf{x} = (x_1, x_2, x_3, ...) : x_1 \geq x_2 \geq x_3 \geq ... \text{ and } \sum_{i=1}^{\infty} x_i^p < \infty\Big\} \tag{1.1}$$

with the p-norm metric $d(\mathbf{x}, \mathbf{y}) = \big(\sum_{i=1}^{\infty} |x_i - y_i|^p\big)^{1/p}$. Let $\ell^2_{\downarrow} \times \mathbb{N}^{\infty}$ denote the product topology of $\ell^2_{\downarrow}$ and $\mathbb{N}^{\infty}$ with $\mathbb{N}^{\infty}$ denoting the sequences on $\mathbb{N}$ endowed with the product topology. Define also

$$\mathbb{U}_{\downarrow} := \Big\{((x_i, y_i))_{i=1}^{\infty} \in \ell^2_{\downarrow} \times \mathbb{N}^{\infty} : \sum_{i=1}^{\infty} x_i y_i < \infty \text{ and } y_i = 0 \text{ whenever } x_i = 0\ \forall i\Big\} \tag{1.2}$$

with the metric

$$\mathrm{d}_{\mathbb{U}}((\mathbf{x}_1, \mathbf{y}_1), (\mathbf{x}_2, \mathbf{y}_2)) := \bigg(\sum_{i=1}^{\infty} (x_{1i} - x_{2i})^2\bigg)^{1/2} + \sum_{i=1}^{\infty} \big|x_{1i}y_{1i} - x_{2i}y_{2i}\big|. \tag{1.3}$$

Further, let $\mathbb{U}^0_{\downarrow} \subset \mathbb{U}_{\downarrow}$ be given by

$$\mathbb{U}^0_{\downarrow} := \big\{((x_i, y_i))_{i=1}^{\infty} \in \mathbb{U}_{\downarrow} : \text{ if } x_k = x_m, k \leq m, \text{ then } y_k \geq y_m\big\}. \tag{1.4}$$

Let $(\mathbb{U}^0_{\downarrow})^k$ denote the k-fold product space of $\mathbb{U}^0_{\downarrow}$. For any $\mathbf{z} \in \mathbb{U}_{\downarrow}$, $\mathrm{ord}(\mathbf{z})$ will denote the element of $\mathbb{U}^0_{\downarrow}$ obtained by suitably ordering the coordinates of $\mathbf{z}$.

We often use the boldface notation $\mathbf{X}$ for the process $(X(s))_{s\geq 0}$, unless stated otherwise. $\mathbb{D}[I,E]$ will denote the space of càdlàg functions from a locally compact second countable Hausdorff space I to the metric space $E=(E,\mathrm{d})$ equipped with Skorohod J_1-topology. $\mathbb{D}[0,t]$ (resp. $\mathbb{D}[0,\infty)$) simply denotes the case $I=[0,t]$ (resp. $[0,\infty)$) with $E=\mathbb{R}$. Consider a decreasing sequence $\boldsymbol{\theta}=(\theta_1,\theta_2,\dots)\in\ell^3_\downarrow\setminus\ell^2_\downarrow$. Denote by $\mathcal{I}_i(s):=\mathbb{1}_{\{\xi_i\leq s\}}$ where $\xi_i\sim\mathrm{Exp}(\theta_i/\mu)$ independently, and $\mathrm{Exp}(r)$ denotes the exponential distribution with rate r. Consider the process

$$\bar{S}^\lambda_\infty(t)=\sum_{i=1}^\infty\theta_i\left(\mathcal{I}_i(t)-(\theta_i/\mu)t\right)+\lambda t, \tag{1.5}$$

for some $\lambda\in\mathbb{R},\mu>0$ and define the reflected version of $\bar{S}^\lambda_\infty(t)$ by

$$\mathrm{refl}(\bar{S}^\lambda_\infty(t))=\bar{S}^\lambda_\infty(t)-\min_{0\leq u\leq t}\bar{S}^\lambda_\infty(u). \tag{1.6}$$

The process of the form (1.5) was termed *thinned* Lévy processes in [14] (see also [2, 44]), since the summands are thinned versions of Poisson processes. For any function $f\in\mathbb{D}[0,\infty)$, define $\underline{f}(x)=\inf_{y\leq x}f(y)$. $\mathbb{D}_+[0,\infty)$ is the subset of $\mathbb{D}[0,\infty)$ consisting of functions with positive jumps only. Note that $\underline{f}$ is continuous when $f\in\mathbb{D}_+[0,\infty)$. An *excursion* of a function $f\in\mathbb{D}_+[0,T]$ is an interval (l,r) such that

$$\min\{f(l-),f(l)\}=\underline{f}(l)=\underline{f}(r)=\min\{f(r-),f(r)\}\quad\text{and}\quad f(x)>\underline{f}(r),\ \forall x\in(l,r)\subset[0,T]. \tag{1.7}$$

Excursions of a function $f\in\mathbb{D}_+[0,\infty)$ are defined similarly. We will use γ to denote an excursion, as well as the length of the excursion γ to simplify notation.

Also, define the counting process $\mathbf{N}$ to be the Poisson process that has intensity $\mathrm{refl}(\bar{S}^\lambda_\infty(t))$ at time t conditional on $(\bar{S}^\lambda_\infty(u))_{u\leq t}$. Formally, $\mathbf{N}$ is characterized as the counting process for which

$$N(t)-\int_0^t\mathrm{refl}(\bar{S}^\lambda_\infty(u))\mathrm{d}u \tag{1.8}$$

is a martingale. We use the notation $N(\gamma)$ to denote the number of marks in the interval γ.

Finally, we define a Markov process $(\mathbf{Z}(s))_{s\in\mathbb{R}}$ on $\mathbb{D}(\mathbb{R},\mathbb{U}^0_\downarrow)$, 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 $K_1X_i(s)X_j(s)$, 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 $i\geq 1$, at rate $K_2X_i^2(s)$, $Y_i(s)$ increases to $Y_i(s)+1$.

Of course, at each event time, the indices are re-organized to give a proper element of $\mathbb{U}^0_\downarrow$. This process was first introduced in [10] to study the joint behavior of the component sizes and the surplus edges over the critical window. In [10], 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 [10, Theorem 3.1] that this is a (nearly) Feller process, a property that will play a crucial rule in the final part of this paper.

Remark 1. Notice that the summation term in (1.5), after replacing θ_i by $\mu\theta_i$, is of the form

$$V^{\boldsymbol{\theta}}(s)=\mu^\alpha\sum_{i=1}^\infty\left(\theta_i\mathbb{1}_{\{\xi_i\leq s\}}-\theta_i^2s\right), \tag{1.9}$$

where $\xi_i\sim\mathrm{Exp}(\theta_i)$ independently over i and $\boldsymbol{\theta}\in\ell^3_\downarrow\setminus\ell^2_\downarrow$. Therefore, by [5, Lemma 1], the process $\mathrm{refl}(\bar{\mathbf{S}}^\lambda_\infty)$ has no infinite excursions almost surely and only finitely many excursions with length at least δ, for any $\delta>0$.

1.3 Main results for critical configuration models

Throughout this paper we will use the shorthand notation

$$\alpha = 1/(\tau-1), \qquad \rho = (\tau-2)/(\tau-1), \qquad \eta = (\tau-3)/(\tau-1), \tag{1.10a}$$

$$a_n = n^{\alpha} L(n), \qquad b_n = n^{\rho}(L(n))^{-1}, \qquad c_n = n^{\eta}(L(n))^{-2}, \tag{1.10b}$$

where $\tau \in (3,4)$ and $L(\cdot)$ is a slowly-varying function. We state our results under the following assumptions:

Assumption 1. Fix $\tau \in (3,4)$. Let $\boldsymbol{d} = (d_1,\dots,d_n)$ be a degree sequence such that the following conditions hold:

(i) (*High-degree vertices*) For any fixed $i \geq 1$,

$$\frac{d_i}{a_n} \to \theta_i, \tag{1.11}$$

where $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots) \in \ell^3_{\downarrow} \setminus \ell^2_{\downarrow}$.

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

$$\frac{1}{n}\sum_{i\in[n]} d_i \to \mu := \mathbb{E}\left[D\right], \qquad \frac{1}{n}\sum_{i\in[n]} d_i^2 \to \mathbb{E}[D^2], \qquad \lim_{K\to\infty}\limsup_{n\to\infty} a_n^{-3}\sum_{i=K+1}^{n} d_i^3 = 0. \tag{1.12}$$

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

$$\nu_n(\lambda) := \frac{\sum_{i\in[n]} d_i(d_i-1)}{\sum_{i\in[n]} d_i} = 1 + \lambda c_n^{-1} + o(c_n^{-1}). \tag{1.13}$$

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

Note that Assumption 1 (i)-(ii) implies $\liminf_{n\to\infty} \mathbb{E}[D_n^3] = \infty$. The following three results hold for any $\mathrm{CM}_n(\boldsymbol{d})$ satisfying Assumption 1:

Theorem 1. *Consider $\mathrm{CM}_n(\boldsymbol{d})$ with the degrees satisfying* Assumption 1. *Denote the i^{th}-largest cluster of $\mathrm{CM}_n(\boldsymbol{d})$ by $\mathscr{C}_{(i)}$. Then,*

$$\left(b_n^{-1}|\mathscr{C}_{(i)}|\right)_{i\geq 1} \xrightarrow{d} \left(\gamma_i(\lambda)\right)_{i\geq 1}, \tag{1.14}$$

with respect to the $\ell^2_{\downarrow}$-topology where $\gamma_i(\lambda)$ is the length of the i^{th} largest excursion of the process $\bar{\mathbf{S}}_{\infty}^{\lambda}$, while b_n and the constants λ, μ are defined in (1.10b) and Assumption 1.

Theorem 2. *Consider $\mathrm{CM}_n(\boldsymbol{d})$ with the degrees satisfying* Assumption 1. *Let $\mathrm{SP}(\mathscr{C}_{(i)})$ denote the number of surplus edges in $\mathscr{C}_{(i)}$ and let $\mathbf{Z}_n := \mathrm{ord}(b_n^{-1}|\mathscr{C}_{(i)}|, \mathrm{SP}(\mathscr{C}_{(i)}))_{i\geq 1}$ and $\mathbf{Z} := \mathrm{ord}(\gamma_i(\lambda), N(\gamma_i))_{i\geq 1}$. Then, as $n\to\infty$,*

$$\mathbf{Z}_n \xrightarrow{d} \mathbf{Z} \tag{1.15}$$

with respect to the $\mathbb{U}^0_{\downarrow}$ topology, where $\mathbf{N}$ is defined in (1.8).

Theorem 3. *The results in* Theorem 1 *and* Theorem 2 *also hold for $\mathrm{CM}_n(\boldsymbol{d})$ conditioned on simplicity.*

Remark 2. The only previous work to understand the critical behavior of the configuration model with heavy-tailed degrees was by Joseph [30] 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 $\mathrm{CM}_n(\boldsymbol{d})$ (Theorem 1). We will see that Assumption 1 is satisfied for i.i.d degrees in Section 2.2. Thus, a quenched version of [30, Theorem 8.3] follows from our results. Further, if the degrees are chosen approximately as the weights chosen in [14], then our results continue to hold. This sheds light on the relation between the scaling limits in [14] and [30] (see Remark 11). Moreover, Theorem 3 resolves [30, Conjecture 8.5].

Remark 3. The conclusions of Theorems 1, 2, and 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 $\boldsymbol{w} = (w_i)_{i\in[n]}$, in Section 8 we will show that the scaling limit for $\mathbf{Z}_n^w := \mathrm{ord}(b_n^{-1}\mathscr{W}_i, \mathrm{SP}(\mathscr{C}_{(i)}))_{i\geq 1}$ is given by $\mathbf{Z} = \mathrm{ord}(\kappa\gamma_i(\lambda), N(\gamma_i))_{i\geq 1}$, where the constant κ is given by $\kappa = \lim_{n\to\infty}\sum_{i\in[n]} d_i w_i/\sum_{i\in[n]} d_i$. Observe that, for $w_i = \mathbb{1}_{\{d_i=k\}}$, $\mathscr{W}_i$ gives the asymptotic number of vertices of degree k in the i^{th} largest component.

Remark 4. It might not be immediate why we should work with Assumption 1. We will see in Section 2.1 that Assumption 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 1 in more detail in Section 3.

1.4 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 $\mathrm{CM}_n(\boldsymbol{d})$ under the following assumptions:

Assumption 2. (i) Assumption 1 (i), and (ii) hold for the degree sequence and $\mathrm{CM}_n(\boldsymbol{d})$ is supercritical, i.e.,

$$\nu_n = \frac{\sum_{i\in[n]} d_i(d_i-1)}{\sum_{i\in[n]} d_i} \to \nu = \frac{\mathbb{E}\left[D(D-1)\right]}{\mathbb{E}\left[D\right]} > 1. \tag{1.16}$$

(ii) (*Critical window for percolation*) The percolation parameter p_n satisfies

$$p_n = p_n(\lambda) := \frac{1}{\nu_n}\big(1+\lambda c_n^{-1} + o(c_n^{-1})\big) \tag{1.17}$$

for some $\lambda \in \mathbb{R}$.

Let $\mathrm{CM}_n(\boldsymbol{d}, p_n(\lambda))$ denote the graph obtained through percolation on $\mathrm{CM}_n(\boldsymbol{d})$ with bond retention probability $p_n(\lambda)$. The following result gives the asymptotics for the ordered component sizes and the surplus edges for $\mathrm{CM}_n(\boldsymbol{d}, p_n(\lambda))$:

Theorem 4. *Consider* $\mathrm{CM}_n(\boldsymbol{d}, p_n(\lambda))$ *satisfying* Assumption 2. *Let* $\tilde{\mathbf{S}}_\infty^\lambda$ *denote the process in* (1.5) *with* θ_i *replaced by* $\theta_i/\sqrt{\nu}$, *and* $\mathscr{C}_{(i)}^p$ *denote the* i^{th} *largest component of* $\mathrm{CM}_n(\boldsymbol{d}, p_n)$ *and let* $\mathbf{Z}_n^p(\lambda) := \mathrm{ord}(b_n^{-1}|\mathscr{C}_{(i)}^p|, \mathrm{SP}(\mathscr{C}_{(i)}^p))_{i\geq 1}$, $\mathbf{Z}^p(\lambda) := \mathrm{ord}((\nu^{1/2}\tilde{\gamma}_i(\lambda), N(\tilde{\gamma}_i(\lambda)))_{i\geq 1}$, *where* $\tilde{\gamma}_i(\lambda)$ *is the largest excursion of* $\tilde{\mathbf{S}}_\infty^\lambda$. *Then, for any* $\lambda \in \mathbb{R}$, *as* $n \to \infty$,

$$\mathbf{Z}_n^p(\lambda) \xrightarrow{d} \mathbf{Z}^p(\lambda) \tag{1.18}$$

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

Now, consider a graph $\mathrm{CM}_n(\boldsymbol{d})$ satisfying Assumption 2 (i). To any edge (ij) between vertices i and j (if any), associate an independent uniform random $U_{(ij)}$. Note that the graph obtained by keeping only those edges satisfying $U_{(ij)} \leq p_n(\lambda)$ is distributed as $\mathrm{CM}_n(\boldsymbol{d}, p_n(\lambda))$. This construction naturally couples the graphs $(\mathrm{CM}_n(\boldsymbol{d}, p_n(\lambda)))_{\lambda\in\mathbb{R}}$ using the same set of uniform random variables. Our next result shows that the evolution of the component sizes and the surplus edges of $\mathrm{CM}_n(\boldsymbol{d}, p_n(\lambda))$, as λ varies, can be described by a version of the augmented multiplicative coalescent process described in Section 1.2:

Theorem 5. *Suppose that* Assumption 2 *holds, and $\ell_n/n = \mu + o(n^{-\zeta})$ for some $\eta < \zeta < 1/2$. Fix any $k \geq 1$, $-\infty < \lambda_1 < \cdots < \lambda_k < \infty$. Then, there exists a version $\mathbf{AMC} = (\mathrm{AMC}(\lambda))_{\lambda\in\mathbb{R}}$ of the augmented multiplicative coalescent such that, as $n \to \infty$,*

$$(\mathbf{Z}_n^p(\lambda_1), \ldots \mathbf{Z}_n^p(\lambda_k)) \xrightarrow{d} (\mathbf{AMC}(\lambda_1), \ldots, \mathbf{AMC}(\lambda_k)) \tag{1.19}$$

with respect to the $(\mathbb{U}_\downarrow^0)^k$ topology, where at each λ, $\mathrm{AMC}(\lambda)$ is distributed as the limiting object in (1.18).

Remark 5. Theorem 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 [20, Theorem 4], which was proved only for the cluster sizes.

Remark 6. Theorem 5, in fact, shows that there exists a version of the AMC process whose distribution at each fixed λ can be described by the excursions of a thinned Lévy process and an associated Poisson process. This did not appear in [10, 19], since the scaling limits in their settings were described in terms of the excursions of a Brownian motion with parabolic drift.

Remark 7. The additional assumption in Theorem 5 about the asymtotics ℓ_n/n is required only in one place for Proposition 24 and the rest of the proof works under Assumption 2 only. That is why we have separated this assumption from the set of conditions in Assumption 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 2.1.

Remark 8. As we will see in Section 10, the proof of Theorem 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.

2 Important examples

2.1 Power-law degrees with small perturbation

As discussed in the introduction, our main goal is to obtain results for the critical configuration model satisfying $\mathbb{P}(D_n \geq k) \sim L_0(k)k^{-(\tau-1)}$ for some $\tau \in (3,4)$. In this section, we consider such an example and show that the conditions of Assumption 1 are satisfied. Thus, the results in Section 1.3 hold for $\mathrm{CM}_n(\boldsymbol{d})$ in the following set-up that is closely related to the model studied in [14] 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

$$G(x) = 1 - F(x) = \frac{C_F L_0(x)}{x^{\tau-1}}(1 + o(1)) \quad \text{as } x \to \infty, \tag{2.1}$$

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 $f^{-1}(x):=\inf\{y: f(y)\leq x\}$. Therefore, using [15, Theorem 1.5.12],

$$G^{-1}(x)=\frac{C_F^{1/(\tau-1)}L(1/x)}{x^{1/(\tau-1)}}(1+o(1)) \quad \text{as } x\to 0, \tag{2.2}$$

where $L(\cdot)$ is another slowly-varying function. Note that [15, Theorem 1.5.12] is stated for positive exponents only. Since our exponent is negative, the asymptotics in (2.2) holds for $x\to 0$. Suppose that the random variable D is such that

$$\nu=\frac{\mathbb{E}\left[D(D-1)\right]}{\mathbb{E}\left[D\right]}=1. \tag{2.3}$$

Define the degree sequence $\boldsymbol{d}_\lambda$ by taking the degree of the i^{th} vertex to be

$$d_i=d_i(\lambda):=G^{-1}(i/n)+\delta_{i,n}(\lambda), \tag{2.4}$$

where the $\delta_{i,n}(\lambda)$'s are non-negative integers satisfying the asymptotic equivalence

$$\delta_{i,n}(\lambda)\sim\lambda G^{-1}(i/n)c_n^{-1}, \quad \text{as } n\to\infty. \tag{2.5}$$

The $\delta_{i,n}(\lambda)$'s are chosen in such a way that Assumption 1 (iv) is satisfied. Fix any $K\geq 1$. Notice that (2.2) and (2.5) imply that, for all large enough n (independently of K), the first K largest degrees $(d_i)_{i\in[K]}$ satisfy

$$d_i=\left(\frac{n^\alpha C_F^\alpha L(n/i)}{i^\alpha}\right)\left(1+\lambda c_n^{-1}+o(c_n^{-1})\right). \tag{2.6}$$

Therefore, $\boldsymbol{d}_\lambda$ satisfies Assumption 1 (i) with $\theta_i=(C_F/i)^\alpha$. We next address Assumption 1 (ii), (iii) in the next two lemmas:

Lemma 6. *The degree sequence $\boldsymbol{d}_\lambda$ defined in* (2.4) *satisfies* Assumption 1 (ii).

Proof. Note that, by (2.6), $d_1^2=o(n)$. Also, since G^{-1} is non-increasing

$$\int_0^1 G^{-1}(x)\mathrm{d}x-\frac{d_1}{n}\leq\frac{1}{n}\sum_{i\in[n]}G^{-1}(i/n)\leq\int_0^1 G^{-1}(x)\mathrm{d}x. \tag{2.7}$$

Therefore,

$$\frac{1}{n}\sum_{i\in[n]}d_i=\frac{1}{n}\sum_{i\in[n]}G^{-1}(i/n)(1+O(c_n^{-1}))=\int_0^1 G^{-1}(x)\mathrm{d}x+O(d_1/n)+O(c_n^{-1})=\mathbb{E}\left[D\right]+O(b_n^{-1}). \tag{2.8}$$

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 [15, 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 $i\leq nC^{-1}$, $L(n/i)/L(n)<Ai^\delta$. Therefore, (2.2) implies

$$a_n^{-3}\sum_{i>K}d_i^3\leq A\sum_{i>K}i^{-3\alpha+3\delta}+\frac{\sup_{1\leq x\leq C}L(x)^3}{L(n)^3}\sum_{i>nC^{-1}}i^{-3\alpha}. \tag{2.9}$$

From our choice of δ, $-3\alpha+3\delta<-1$ and therefore $\sum_{i\geq 1}i^{-3\alpha+3\delta}<\infty$. By [15, Lemma 1.3.2], $\sup_{1\leq x\leq C}L(x)^3<\infty$. 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 $K\to\infty$. ∎

Lemma 7. *The degree sequence $\boldsymbol{d}_\lambda$ defined in* (2.4) *satisfies* Assumption 1 (iii), *i.e., there exists $\lambda_0 \in \mathbb{R}$ such that*

$$\nu_n(\lambda) = 1 + (\lambda + \lambda_0)c_n^{-1} + o(c_n^{-1}). \tag{2.10}$$

Proof. Firstly, Lemma 6 guarantees the convergence of the second moment of the degree sequence. However, (2.10) is more about obtaining sharper asymptotics for $\nu_n(\lambda)$. We use similar arguments as in [14, 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

$$\nu_n = 1 + \lambda_0 c_n^{-1} + o(c_n^{-1}). \tag{2.11}$$

Consider $d_i(0)$ as given in (2.4) with $\lambda = 0$. Lemma 6 implies

$$\nu_n = \frac{\sum_{i\in[n]} d_i(0)^2}{n\mathbb{E}[D]} - 1 + o(c_n^{-1}). \tag{2.12}$$

Fix any $K \geq 1$. We have

$$\int_{K/n}^{1} G^{-1}(u)^2 \mathrm{d}u - \frac{d_K^2}{n} \leq \frac{1}{n}\sum_{i=K+1}^{n} d_i^2 \leq \int_{K/n}^{1} G^{-1}(u)^2 \mathrm{d}u. \tag{2.13}$$

Now by (2.4), $d_K^2/n = \Theta(K^{-2\alpha}L(n/K)^2 n^{-\eta})$. Therefore,

$$\nu - \nu_n = \frac{1}{\mathbb{E}[D]}\left(\sum_{i=1}^{K}\int_{(i-1)/n}^{i/n} G^{-1}(u)^2 \mathrm{d}u - \frac{1}{n}\sum_{i=1}^{K} d_i^2\right) + O(K^{-2\alpha}L(n/K)^2 n^{-\eta}). \tag{2.14}$$

Again, using (2.4),

$$\frac{1}{n}\sum_{i=1}^{K} d_i^2 = n^{-\eta}\sum_{i=1}^{K}\left(\frac{C_F}{i}\right)^{2\alpha} L(n/i)^2 + o(c_n^{-1}) = c_n^{-1}\sum_{i=1}^{K}\left(\frac{C_F}{i}\right)^{2\alpha} + \varepsilon(c_n, K), \tag{2.15}$$

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 (2.15) satisfies $\lim_{n\to\infty} c_n\varepsilon(c_n, K) = 0$ for each fixed $K \geq 1$. Again,

$$\begin{aligned}\sum_{i=1}^{K}\int_{(i-1)/n}^{i/n} G^{-1}(u)^2 \mathrm{d}u &= n^{-\eta}\sum_{i=1}^{K}\int_{(i-1)}^{i}\left(\frac{C_F}{u}\right)^{2\alpha} L(n/u)^2 \mathrm{d}u + o(c_n^{-1}) \\ &= c_n^{-1}\sum_{i=1}^{K}\int_{(i-1)}^{i}\left(\frac{C_F}{u}\right)^{2\alpha} \mathrm{d}u + \varepsilon'(c_n, K),\end{aligned} \tag{2.16}$$

where $\lim_{n\to\infty} c_n\varepsilon'(c_n, K) = 0$ for each fixed $K \geq 1$. Thus combining (2.14), (2.15), and (2.16) and first letting $n \to \infty$ and then $K \to \infty$, we get

$$\lim_{n\to\infty} c_n(\nu_n - \nu) = \lambda_0, \tag{2.17}$$

where

$$\lambda_0 = -\frac{C_F^{2\alpha}}{\mathbb{E}[D]}\sum_{i=1}^{\infty}\left(\int_{i-1}^{i} u^{-2\alpha}\mathrm{d}u - i^{-2\alpha}\right). \tag{2.18}$$

Using Euler-Maclaurin summation [23, Page 333] it can be seen that λ_0 is finite which completes the proof. ☐

Remark 9. Note that if we add approximately $cn^{1-\eta}$ ($c > 0$ is a constant) ones in the degree sequence given in (2.4), 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 ζ 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 10. If ν in (2.3) is larger than one, then the degree sequence satisfies Assumption 2. Therefore, the results for critical percolation also hold in this setting. (2.8) implies that the additional assumption in Theorem 5 is also satisfied.

2.2 Random degrees sampled from a power-law distribution

We now consider the set-up discussed in [30]. Let $D_1, \dots, D_n$ be i.i.d samples from a distribution F, where F is defined in (2.1). Therefore, the asymptotic relation in (2.2) holds. Consider the random degree sequence $\boldsymbol{d}$ where $d_i = D_{(i)}$, $D_{(i)}$ being the i^{th} order statistic of $(D_1, \dots, D_n)$. We show that $\boldsymbol{d}$ satisfies Assumption 1 almost surely under a suitable coupling. We use a coupling from [18, Section 13.6]. Let $(E_1, E_2, \dots)$ be an i.i.d sequence of unit rate exponential random variables and let $\Gamma_i := \sum_{j=1}^{i} E_j$. Let

$$\bar{d}_i = \bar{D}_{(i)} = G^{-1}(\Gamma_i/\Gamma_{n+1}). \tag{2.19}$$

It can be checked that $(d_1, \dots, d_n) \overset{d}{=} (\bar{d}_1, \dots, \bar{d}_n)$ and therefore, we will ignore the bar in the subsequent notation. Note that, by the stong law of large numbers, $\Gamma_{n+1}/n \xrightarrow{\text{a.s.}} 1$. Thus, for each fixed $i \geq 1$, $\Gamma_{n+1}/(n\Gamma_i) \xrightarrow{\text{a.s.}} 1/\Gamma_i$. Using (2.2), we see that $\boldsymbol{d}$ satisfies Assumption 1 (i) almost surely under this coupling with $\theta_i = (C_F/\Gamma_i)^{\alpha}$. The first two conditions of Assumption 1 (ii) are trivially satisfied by $\boldsymbol{d}$ almost surely using the strong law of large numbers. To see the third condition, we first claim that

$$\mathbb{P}\bigg(\sum_{i=1}^{\infty} \Gamma_i^{-3\alpha} < \infty\bigg) = 1. \tag{2.20}$$

To see (2.20), note that Γ_i has a Gamma distribution with shape parameter i and scale parameter 1. Thus, for $i > 3\alpha$,

$$\mathbb{E}[\Gamma_i^{-3\alpha}] = \frac{\Gamma(i-3\alpha)}{\Gamma(i)} = i^{-3\alpha}(1 + O(1/i)), \tag{2.21}$$

where $\Gamma(x)$ is the Gamma function and the last equality follows from Stirling's approximation. Therefore,

$$\mathbb{E}\bigg[\sum_{i=1}^{\infty} \Gamma_i^{-3\alpha}\bigg] = \sum_{i=1}^{\infty} \mathbb{E}\big[\Gamma_i^{-3\alpha}\big] < \infty \tag{2.22}$$

and (2.20) follows. Now, using the fact that $\Gamma_{n+1}/n \xrightarrow{\text{a.s.}} 1$, we can use arguments identical to (2.9) 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 1 (ii) holds almost surely.

To see Assumption 1 (iii), an argument similar to Lemma 7 can be carried out to prove that

$$\lim_{n\to\infty} c_n(\nu_n - \nu) \xrightarrow{\text{a.s.}} \Lambda_0, \tag{2.23}$$

where

$$\Lambda_0 := -\frac{C_F^{2\alpha}}{\mathbb{E}\left[D\right]} \sum_{i=1}^{\infty} \bigg(\int_{\Gamma_{i-1}}^{\Gamma_i} u^{-2\alpha} \mathrm{d}u - \Gamma_i^{-2\alpha}\bigg). \tag{2.24}$$

Therefore, the results in Section 1.3 hold conditionally on the degree sequence if we assume the degrees to be i.i.d samples from a distribution of the form (2.1). For the percolation results, notice that the additional condition in Theorem 5 is a direct consequence of the convergence rates of sums of i.i.d sequence of random variables [31, Corollary 3.22].

Remark 11. Let us recall the limiting object obtained in [30, Theorem 8.1] and compare this with the limiting object $\bar{\mathbf{S}}_{\infty}^{\Lambda_0}$, defined in (1.5) with Λ_0 given by (2.24). We will prove an analogue of [30, Theorem 8.1] in Theorem 8. Although we use a different exploration process from [30], the fact that the component sizes are *huge* compared to the number of cycles in a component, one can prove Theorem 8 for the exploration process in [30] also. This will indirectly imply that Joseph's limiting object obeys the law of $\bar{\mathbf{S}}_{\infty}^{\Lambda_0}$, averaged out over the Γ-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 Discussion

Assumptions on the degree distribution. Let us now briefly explain the significance of Assumption 1. Unlike the finite third-moment case [20], the high-degree vertices dictate the scaling limit in Theorem 1 and therefore it is essential to fix their asymptotics through Assumption 1 (i). Assumption 1 (iii) defines the critical window of the phase transition and Assumption 1 (iv) is reminiscent of the fact that a configuration model with negligibly small amount of degree-one vertices is always supercritical. Assumption 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 [4, 13, 14, 20, 37]. 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 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 [24] showed that the size of the k^{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 [24, Remark 1.4]). In [11], we show that this is indeed the case for the entire barely subcritical regime, i.e., the size of the k^{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 ρ_n is the survival probability of $\mathcal{X}_n$ [42]. 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 paper is to understand the evolution of the component sizes and the surplus edges over the critical window in Theorem 5. It turns out that a precise characterization of the evolution the percolation clusters is necessary for understanding the minimal spanning tree of the giant component with i.i.d weights on each edge [1]. Also, since the percolated configuration model is again a configuration model [22, 25], 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 1 is identical to that in [14, Theorem 1.1] for rank-1 inhomogeneous random graphs. Thus, $\mathrm{CM}_n(\boldsymbol{d})$ with regularly-varying tails falls onto the domain of attraction of the new universality class studied in [14]. This is again conforming to 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 [5]. In the notation of [5, (6)], the scaling limits $\mathrm{CM}_n(\boldsymbol{d})$, under Assumption 1, give rise to the case $\kappa = 0$. To understand this, let us discuss some existing works. In [4, 6, 13, 20, 30], 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 [4, (10)], [6, (4)]. In the case of [14] and this paper, the barely subcritical component sizes do not become negligible due to the existence of the high-degree vertices (see [14, Theorem 1.3]). As discussed in [5, 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 [5] 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 to explain 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 [14] if the size of the high-degree vertices increases by a factor of $L(n)$ (see (1.10b)). 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.

Open problems.
(i) A natural question is to study what the component sizes, viewed as metric spaces, look like. Recently, [12] studied this problem for rank-1 inhomogeneous random graphs for heavy-tailed weights. In a work in progress Bhamidi et al. [11], we show that the metric space structure of $\mathrm{CM}_n(\boldsymbol{d})$ is in the same universality class of the rank-one inhomogeneous model, as shown in [12]. This is the first step in understanding the minimal spanning tree problem (see [1]).
(ii) As discussed in Section 2.2 (see Remark 11), it will be interesting to get a direct proof of the fact that the limiting object in [30, Theorem 8.1] is obtained by averaging the distribution of $\mathbf{S}_\infty^{\Lambda_0}$ over the collections $(\Gamma_i)_{i\geq 1}$.
(iii) We have only shown the finite-dimensional convergence in Theorem 5. It is an open question to obtain a suitable tightness criterion that would imply the process level convergence of the vector of component sizes and surplus edges over the whole critical window.

Overview of the proofs. The proofs of Theorems 1 and 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 4, 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 1 follows from the asymptotics of the susceptibility function in Section 5. The joint convergence of the component sizes and surplus edges is proved by verifying a uniform tightness condition on the surplus edges in Section 6. Then, in Section 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. The proof of Theorem 4 is completed by showing that the percolated degree sequence is again a configuration model satisfying Assumption 1. Section 10 is devoted to the proof of Theorem 5 which exploits different properties of the augmented multiplicative coalescent process.

4 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 1 (Exploring the graph)**.** Consider the configuration model $\mathrm{CM}_n(\boldsymbol{d})$. The algorithm carries along vertices that can be alive, active, exploring and killed and half-edges that can be alive, active or killed. We sequentially explore the graph as follows:

(S0) At stage $i = 0$, all the vertices and the half-edges are *alive* but none of them are *active*. Also, there are no *exploring* vertices.

(S1) At each stage i, if there is no active half-edge at stage i, 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 $\mathrm{degree}(v') = 1$ (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 1 gives a *breadth-first* exploration of the connected components of $\mathrm{CM}_n(\boldsymbol{d})$. Define the exploration process by

$$S_n(0) = 0, \quad S_n(l) = S_n(l-1) + d_{(l)} J_l - 2, \tag{4.1}$$

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^{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 τ_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 $\mathcal{I}_i^n(l) := \mathbb{1}_{\{i \in \mathscr{V}_l\}}$. Note that

$$S_n(l) = \sum_{i\in[n]} d_i \mathcal{I}_i^n(l) - 2l = \sum_{i\in[n]} d_i \left(\mathcal{I}_i^n(l) - \frac{d_i}{\ell_n} l\right) + (\nu_n(\lambda) - 1)\, l. \tag{4.2}$$

Recall the notation in (1.10b). 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 (iii),

$$\bar{S}_n(t) = a_n^{-1} \sum_{i\in[n]} d_i \left(\mathcal{I}_i^n(t b_n) - \frac{d_i}{\ell_n} t b_n\right) + \lambda t + o(1). \tag{4.3}$$

Note the similarity between the expressions in (1.5) and (4.3). We will prove the following:

Theorem 8. *Consider the process* $\bar{\mathbf{S}}_n := (\bar{S}_n(t))_{t\geq 0}$ *defined in* (4.3) *and recall the definition of* $\bar{\mathbf{S}}_\infty := (\bar{S}_\infty(t))_{t\geq 0}$ *from* (1.5). *Then,*

$$\bar{\mathbf{S}}_n \xrightarrow{d} \bar{\mathbf{S}}_\infty \tag{4.4}$$

with respect to the Skorohod J_1 topology.

The proof of Theorem 8 is completed by showing that the summation term in (4.3) is predominantly carried by the first *few* terms and the limit of the first few terms gives rise to the limiting process given in (1.5). Fix $K \geq 1$ to be large. Denote by $\mathscr{F}_l$ the sigma-field containing the information generated up to time l by Algorithm 1. Also, let Υ_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 $\{\mathcal{I}_i^n(l)=0\}$,

$$\mathbb{P}\left(\mathcal{I}_i^n(l+1)=1\big|\mathscr{F}_l\right) = \begin{cases} \frac{d_i}{\ell_n-2(l-\upsilon_l)-1} & \text{if } l\notin\Upsilon_l,\\ \frac{d_i}{\ell_n-2(l-\upsilon_l)} & \text{otherwise}\end{cases} \tag{4.5}$$

and, uniformly over $l \leq Tb_n$,

$$\mathbb{P}\left(\mathcal{I}_i^n(l+1)=1\big|\mathscr{F}_l\right) \geq \frac{d_i}{\ell_n} \quad \text{on the set } \{\mathcal{I}_i^n(l)=0\}. \tag{4.6}$$

Denote $M_n^K(l) = a_n^{-1}\sum_{i\in[n]} d_i\big(\mathcal{I}_i^n(l)-\frac{d_i}{\ell_n}l\big)$. Then,

$$\begin{aligned}\mathbb{E}\big[M_n^K(l+1)-M_n^K(l)\big|\mathscr{F}_l\big] &= \mathbb{E}\Big[\sum_{i=K+1}^n a_n^{-1}d_i\left(\mathcal{I}_i^n(l+1)-\mathcal{I}_i^n(l)-\frac{d_i}{\ell_n}\right)\Big|\mathscr{F}_l\Big]\\ &= \sum_{i=K+1}^n a_n^{-1}d_i\left(\mathbb{E}\big[\mathcal{I}_i^n(l+1)\big|\mathscr{F}_l\big]\mathbb{1}_{\{\mathcal{I}_i^n(l)=0\}}-\frac{d_i}{\ell_n}\right)\geq 0.\end{aligned} \tag{4.7}$$

Thus $(M_n^K(l))_{l=1}^{Tb_n}$ is a sub-martingale. Further, (4.5) implies that, uniformly for all $l \leq Tb_n$,

$$\mathbb{P}\left(\mathcal{I}_i^n(l)=0\right) \geq \left(1-\frac{d_i}{\ell_n'}\right)^l, \tag{4.8}$$

where $\ell_n' = \ell_n - 2Tb_n - 1$. Thus, Assumption 1 (ii) gives

$$\begin{aligned}\big|\mathbb{E}[M_n^K(l)]\big| &= a_n^{-1}\sum_{i=K+1}^n d_i\left(\mathbb{P}\left(\mathcal{I}_i^n(l)=1\right)-\frac{d_i}{\ell_n}l\right)\\ &\leq a_n^{-1}\sum_{i=K+1}^n d_i\left(1-\left(1-\frac{d_i}{\ell_n'}\right)^l-\frac{d_i}{\ell_n'}l\right)+a_n^{-1}l\sum_{i\in[n]}d_i^2\left(\frac{1}{\ell_n'}-\frac{1}{\ell_n}\right)\\ &\leq \frac{l^2}{2\ell_n'^2 a_n}\sum_{i=K+1}^n d_i^3+o(1)\\ &\leq \frac{T^2n^{2\rho}n^{3\alpha}L(n)^3}{2\ell_n'^2L(n)^2n^\alpha L(n)}\left(a_n^{-3}\sum_{i=K+1}^n d_i^3\right)+o(1) = C\left(a_n^{-3}\sum_{i=K+1}^n d_i^3\right)+o(1),\end{aligned} \tag{4.9}$$

for some constant $C>0$, where we have used the fact that

$$a_n^{-1}l\sum_{i\in[n]}d_i^2\Big(\frac{1}{\ell_n'}-\frac{1}{\ell_n}\Big) = O(n^{2\rho+1-\alpha-2}/L(n)^3) = O(n^{(\tau-4)/(\tau-1)}/L(n)^3) = o(1), \tag{4.10}$$

uniformly for $l \leq Tb_n$. Therefore, uniformly over $l \leq Tb_n$,

$$\lim_{K\to\infty} \limsup_{n\to\infty} \left|\mathbb{E}[M_n^K(l)]\right| = 0. \tag{4.11}$$

Now, note that for any $(x_1, x_2, \dots)$, $0 \leq a+b \leq x_i$ and $a, b > 0$ one has $\prod_{i=1}^R (1-a/x_i)(1-b/x_i) \geq \prod_{i=1}^R (1-(a+b)/x_i)$. Thus, by (4.5), for all $l \geq 1$ and $i \neq j$,

$$\mathbb{P}\left(\mathcal{I}_i^n(l) = 0, \mathcal{I}_j^n(l) = 0\right) \leq \mathbb{P}\left(\mathcal{I}_i^n(l) = 0\right) \mathbb{P}\left(\mathcal{I}_j^n(l) = 0\right) \tag{4.12}$$

and therefore $\mathcal{I}_i^n(l)$ and $\mathcal{I}_j^n(l)$ are negatively correlated. Observe also that, uniformly over $l \leq Tb_n$,

$$\mathrm{Var}\left(\mathcal{I}_i^n(l)\right) \leq \mathbb{P}\left(\mathcal{I}_i^n(l) = 1\right) \leq \sum_{l_1=1}^{l} \mathbb{P}\left(\text{vertex } i \text{ is first discovered at stage } l_1\right) \leq \frac{ld_i}{\ell_n'}. \tag{4.13}$$

Therefore, using the negative correlation in (4.12), uniformly over $l \leq Tb_n$,

$$\mathrm{Var}\left(M_n^K(l)\right) \leq a_n^{-2} \sum_{i=K+1}^{n} d_i^2 \mathrm{Var}\left(\mathcal{I}_i^n(l)\right) \leq \frac{l}{\ell_n' a_n^2} \sum_{i=K+1}^{n} d_i^3 \leq C a_n^{-3} \sum_{i=K+1}^{n} d_i^3, \tag{4.14}$$

for some constant $C > 0$ and by using Assumption 1 (ii) again,

$$\lim_{K\to\infty} \limsup_{n\to\infty} \mathrm{Var}\left(M_n^K(l)\right) = 0, \tag{4.15}$$

uniformly for $l \leq Tb_n$. Now we can use the super-martingale inequality [38, Lemma 2.54.5] stating that for any super-martingale $(M(t))_{t\geq 0}$, with $M(0) = 0$,

$$\varepsilon \mathbb{P}\left(\sup_{s\leq t} |M(s)| > 3\varepsilon\right) \leq 3\mathbb{E}\left[|M(t)|\right] \leq 3\left(\left|\mathbb{E}\left[M(t)\right]\right| + \sqrt{\mathrm{Var}\left(M(t)\right)}\right). \tag{4.16}$$

Using (4.11), (4.14), and (4.16), together with the fact that $(-M_n^K(l))_{l=1}^{Tb_n}$ is a super-martingale, we get, for any $\varepsilon > 0$,

$$\lim_{K\to\infty} \limsup_{n\to\infty} \mathbb{P}\left(\sup_{l\leq Tb_n} |M_n^K(l)| > \varepsilon\right) = 0. \tag{4.17}$$

Define the truncated exploration process

$$\bar{S}_n^K(t) = a_n^{-1} \sum_{i=1}^{K} d_i \left(\mathcal{I}_i^n(tb_n) - \frac{d_i}{\ell_n} tb_n\right) + \lambda t. \tag{4.18}$$

Define $\mathcal{I}_i^n(tb_n) = \mathcal{I}_i^n(\lfloor tb_n \rfloor)$ and recall that $\mathcal{I}_i(s) := \mathbb{1}_{\{\xi_i \leq s\}}$ where $\xi_i \sim \mathrm{Exp}(\theta_i/\mu)$.

Lemma 9. *Fix any $K \geq 1$. As $n \to \infty$,*

$$\left(\mathcal{I}_i^n(tb_n)\right)_{i\in[K], t\geq 0} \xrightarrow{d} \left(\mathcal{I}_i(t)\right)_{i\in[K], t\geq 0}. \tag{4.19}$$

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

$$\mathbb{P}\left(\mathcal{I}_i^n(t_i b_n) = 0,\ \forall i \in [K]\right) \to \mathbb{P}\left(\mathcal{I}_i(t_i) = 0,\ \forall i \in [K]\right) = \exp\Big(-\mu^{-1} \sum_{i=1}^{K} \theta_i t_i\Big). \tag{4.20}$$

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

$$\mathbb{P}\left(\mathcal{I}_i^n(m_i) = 0,\ \forall i \in [K]\right) = \prod_{l=1}^{\infty} \Big(1 - \sum_{i\leq K: l\leq m_i} \frac{d_i}{\ell_n - \Theta(l)}\Big), \tag{4.21}$$

where the $\Theta(l)$ term arises from the expression in (4.5) and noting that $v_l \leq l$. Taking logarithms on both sides of (4.21) and using the fact that $l \leq \max m_i = \Theta(b_n)$ we get

$$\mathbb{P}\left(\mathcal{I}_i^n(m_i) = 0\,\forall i \in [K]\right) = \exp\Big(-\sum_{l=1}^{\infty}\sum_{i\leq K: l\leq m_i}\frac{d_i}{\ell_n} + o(1)\Big) = \exp\Big(-\sum_{i\in[K]}\frac{d_i m_i}{\ell_n} + o(1)\Big). \tag{4.22}$$

Putting $m_i = t_i b_n$, Assumption 1 (i), (ii) gives

$$\frac{m_i d_i}{\ell_n} = \frac{\theta_i t_i}{\mu}(1 + o(1)). \tag{4.23}$$

Hence (4.23), and (4.22) complete the proof of Lemma 9. □

Proof of Theorem 8. The proof of Theorem 8 now follows from (4.3), (4.17) and Lemma 9 by first taking the limit as $n \to \infty$ and then taking the limit as $K \to \infty$. □

Theorem 10. *Recall the definition of* $\mathrm{refl}(\bar{\mathbf{S}}_\infty)$ *from* (1.6)*. As* $n \to \infty$*,*

$$\mathrm{refl}(\bar{\mathbf{S}}_n) \xrightarrow{d} \mathrm{refl}(\bar{\mathbf{S}}_\infty). \tag{4.24}$$

Proof. This follows from Theorem 8 and the fact that the reflection is Lipschitz continuous with respect to the Skorohod J_1 topology (see [45, Theorem 13.5.1]). □

5 Convergence of component sizes

In this section, we complete the proof of Theorem 1. First, we prove a tail summability condition that ensures that the vector of ordered component sizes is tight in $\ell^2_\downarrow$. This also implies that Algorithm 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 8 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 1 follows.

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 size with respect to the $\ell^2_\downarrow$ topology:

Proposition 11. *For any* $\varepsilon > 0$*,*

$$\lim_{K\to\infty}\limsup_{n\to\infty}\mathbb{P}\Big(\sum_{i>K}|\mathscr{C}_{(i)}|^2 > \varepsilon b_n^2\Big) = 0. \tag{5.1}$$

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 12 below to complete the proof of Proposition 11. This fact is not true for the finite third-moment setting [20]. However, since the large-degree vertices guide the scaling behavior in the infinite third-moment case, the observation in Lemma 12 saves some computational complexity, and gives a different proof of the $\ell^2_\downarrow$ tightness than the arguments with size-biased point processes originally described in [4].

Lemma 12. *Consider* $\mathrm{CM}_n(\boldsymbol{d})$ *satisfying* Assumption 1. *Let* $\mathcal{G}^{[K]}$ *be the random graph obtained by removing all edges attached to vertices* $1,\dots,K$ *and let* $\boldsymbol{d}'$ *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* $\{\mathscr{C}^{[K]}_{(i)} : i \geq 1\}$ *be the components of* $\mathcal{G}^{[K]}$*, ordered according to their sizes. Then,*

$$\lim_{K\to\infty} \limsup_{n\to\infty} c_n^{-1}\mathbb{E}\left[|\mathscr{C}^{[K]}(V_n)|\right] = 0. \tag{5.2}$$

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

$$\lim_{K\to\infty} \limsup_{n\to\infty} \mathbb{P}\left(\sum_{i\geq 1} \left|\mathscr{C}^{[K]}_{(i)}\right|^2 > \varepsilon b_n^2\right) = 0. \tag{5.3}$$

Proof. We make use of a result due to Janson [27] regarding bounds on the susceptibility functions for the configuration model. In fact, [27, Lemma 5.2] shows that, for any configuration model $\mathrm{CM}_n(\boldsymbol{d})$ with $\nu_n < 1$,

$$\mathbb{E}\left[|\mathscr{C}(V_n)|\right] \leq 1 + \frac{\mathbb{E}\left[D_n\right]}{1-\nu_n}. \tag{5.4}$$

Now, conditional on the set of removed half-edges, $\mathcal{G}^{[K]}$ is still a configuration model with some degree sequence $\boldsymbol{d}'$ with $d_i' \leq d_i$ for all $i \in [n] \setminus [K]$ and $d_i' = 0$ for $i \in [K]$. Further, the criticality parameter of $\mathcal{G}^{[K]}$ satisfies

$$\begin{aligned} \nu_n^{[K]} = \frac{\sum_{i\in[n]} d_i'(d_i'-1)}{\sum_{i\in[n]} d_i'} &\leq \frac{\sum_{i\in[n]} d_i(d_i-1) - \sum_{i=1}^K d_i(d_i-1)}{\ell_n - 2\sum_{i=1}^K d_i} \\ &= \nu_n - C_1 n^{2\alpha-1}L(n)^2 \sum_{i\leq K}\theta_i^2 = \nu_n - C_1 c_n^{-1}\sum_{i\leq K}\theta_i^2 \end{aligned} \tag{5.5}$$

for some constant $C_1 > 0$. Since $\boldsymbol{\theta} \notin \ell^2_{\downarrow}$, 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 (5.4) on $\mathcal{G}^{[K]}$, we get

$$\mathbb{E}_K\left[|\mathscr{C}^{[K]}(V_n)|\right] \leq \frac{C_2}{1-\nu_n^{[K]}} \leq \frac{C_2}{1-\nu_n + C_1 c_n^{-1}\sum_{i\leq K}\theta_i^2} \leq \frac{C_2 c_n}{-\lambda + C_1\sum_{i\leq K}\theta_i^2}, \tag{5.6}$$

for some constant $C_2 > 0$. Using the fact that $\boldsymbol{\theta} \notin \ell^2_{\downarrow}$, this concludes the proof of (5.2). The proof of (5.3) follows from (5.2) by using the Markov inequality and the observation that

$$\mathbb{E}\left[\sum_{i\geq 1} |\mathscr{C}^{[K]}_{(i)}|^2\right] = n\mathbb{E}\left[|\mathscr{C}^{[K]}(V_n)|\right]. \tag{5.7}$$

□

Proof of Proposition 11. Denote the sum of squares of the component sizes excluding the components containing vertices $1,2,\dots,K$ by $\mathscr{S}_K$. Note that

$$\sum_{i>K} |\mathscr{C}_{(i)}|^2 \leq \mathscr{S}_K \leq \sum_{i\geq 1} |\mathscr{C}^{[K]}_{(i)}|^2. \tag{5.8}$$

Thus, Proposition 11 follows from Lemma 12. □

5.2 Large components are explored early

As remarked at the beginning of Section 5, an important consequence of Proposition 11 is that after time $\Theta(b_n)$, Algorithm 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 8 only gives information about the components explored on the time scale of the order b_n.

Lemma 13. *Let $\mathscr{C}_{\max}^{\geq T}$ be the largest among those components which are started exploring after time Tb_n by* Algorithm 1*. Then, for any $\varepsilon > 0$,*

$$\lim_{T\to\infty} \limsup_{n\to\infty} \mathbb{P}\left(|\mathscr{C}_{\max}^{\geq T}| > \varepsilon b_n\right) = 0. \tag{5.9}$$

Proof. Define the event $\mathscr{A}_{K,T}^n := \{$all the vertices of $[K]$ are explored before time $Tb_n\}$. Recall the definition of $\mathscr{C}_{(i)}^{[K]}$ from Lemma 12. Firstly, note that

$$\mathbb{P}\left(|\mathscr{C}_{\max}^{\geq T}| > \varepsilon b_n,\ \mathscr{A}_{K,T}^n\right) \leq \mathbb{P}\bigg(\sum_{i\geq 1} \big|\mathscr{C}_{(i)}^{[K]}\big|^2 > \varepsilon^2 b_n^2\bigg). \tag{5.10}$$

Moreover, using (4.5) and the fact that $d_j b_n = \Theta(n)$, we get

$$\begin{aligned} \mathbb{P}\left((\mathscr{A}_{K,T}^n)^c\right) &= \mathbb{P}\left(\exists j \in [K] : j \text{ is not explored before } Tb_n\right) \\ &\leq \sum_{j=1}^{K} \mathbb{P}\left(j \text{ is not explored before } Tb_n\right) \\ &\leq \sum_{j=1}^{K} \left(1 - \frac{d_j}{\ell_n - \Theta(Tb_n)}\right)^{Tb_n} \leq \sum_{j=1}^{K} \mathrm{e}^{-CT}, \end{aligned} \tag{5.11}$$

where $C > 0$ is a constant that may depend on K. Now, by (5.10),

$$\mathbb{P}\left(|\mathscr{C}_{\max}^{\geq T}| > \varepsilon b_n\right) \leq \mathbb{P}\bigg(\sum_{i\geq 1} \big|\mathscr{C}_{(i)}^{[K]}\big|^2 > \varepsilon^2 b_n^2\bigg) + \mathbb{P}\left((\mathscr{A}_{K,T}^n)^c\right). \tag{5.12}$$

The proof follows by taking $\limsup_{n\to\infty}$, $\lim_{T\to\infty}$, $\lim_{K\to\infty}$ respectively and using (5.3), (5.11). ☐

5.3 Sample path properties

Recall the definition of an excursion from (1.7). Define the set of excursions of a function f by

$$\mathcal{E} := \{(l, r) : (l, r) \text{ is an excursion of } f\}. \tag{5.13}$$

We also denote the set of excursion end-points by $\mathcal{Y}$, i.e.,

$$\mathcal{Y} := \{r > 0 : (l, r) \in \mathcal{E}\}. \tag{5.14}$$

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 12. 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 δ 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^{th} largest excursion of f and define $\Phi_m : \mathbb{D}_+[0,T] \to \mathbb{R}^m$ by

$$\Phi_m(f) = (\mathcal{L}_1(f), \mathcal{L}_2(f), \ldots, \mathcal{L}_m(f)). \tag{5.15}$$

Note that $\Phi_m(\cdot)$ is well-defined for any good function defined in Definition 2.

Lemma 14. *Suppose that $f \in \mathbb{D}_+[0,T]$ is good. Then, Φ_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 [36, 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

$$f(x) > \min\{f(r-), f(r)\} + \delta \quad \forall x \in (l+\varepsilon, r-\varepsilon). \tag{5.16}$$

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 $\{\Lambda_n\}_{n\geq 1} \subset \mathfrak{L}$ such that for all large enough n,

$$||f_n \circ \Lambda_n - f|| < \frac{\delta}{6} \text{ and } ||\Lambda_n - I|| < \varepsilon, \tag{5.17}$$

where I is the identity function. Now, by Remark 12, f is continuous at r. This implies that $f(r-) = f(r)$, and using (5.16) and (5.17), for all large enough n,

$$f_n(y) > f_n \circ \Lambda_n(r) + \frac{2\delta}{3} \quad \forall y \in (l+2\varepsilon, r-2\varepsilon). \tag{5.18}$$

Further, using the continuity of f at r, $f_n(r) \to f(r)$ and thus, for all sufficiently large n,

$$|f_n \circ \Lambda_n(r) - f_n(r)| \leq |f_n \circ \Lambda_n(r) - f(r)| + |f_n(r) - f(r)| < \frac{\delta}{3}. \tag{5.19}$$

Hence, (5.18) implies that, for all sufficiently large n,

$$f_n(y) > f_n(r) + \frac{\delta}{3} \quad \forall y \in (l+2\varepsilon, r-2\varepsilon). \tag{5.20}$$

Thus, for any $\varepsilon > 0$, we have

$$\liminf_{n\to\infty} \Phi_1(f_n) \geq r - l - 4\varepsilon = \Phi_1(f) - 4\varepsilon. \tag{5.21}$$

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 $r_1 \leq \Phi_1(f) + \varepsilon, T - r_k < \Phi_1(f) + \varepsilon$, and $r_i - r_{i-1} \leq \Phi_1(f) + \varepsilon, \forall i = 2, \ldots, k$. 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 (5.17), for all large enough n,

$$f_n(r_i) - f_n(\Lambda_n(x_i)) > \frac{\delta}{2}. \tag{5.22}$$

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

$$f_n(r_i) - f_n(z_i^n) > \frac{\delta}{2}. \tag{5.23}$$

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 $\underline{f}(x) = \inf_{y\leq x} f(y)$. Therefore, f_n must have an excursion ending 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 $\varepsilon > 0$,

$$\limsup_{n\to\infty} \Phi_1(f_n) \leq \Phi_1(f) + 3\varepsilon. \tag{5.24}$$

Hence the proof follows from (5.21) and (5.24). □

Remark 13. For $f \in \mathbb{D}_+[0,T]$, let $\mathcal{A}_i(f)$ denote the area under the excursion $\mathcal{L}_i(f)$. Let $(f_n)_{n\geq 1}$ 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, (5.17), (5.21) and (5.24) also implies that $(\mathcal{A}_1(f_n), \dots, \mathcal{A}_m(f_n)) \to (\mathcal{A}_1(f), \dots, \mathcal{A}_m(f))$, for any $m \geq 1$.

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 ε almost surely.

Lemma 15. *The thinned Lévy process $\mathbf{S}_\infty^\lambda$ defined in (1.5) is good.*

Proof. Let us make use of the properties of the process $\mathbf{S}_\infty^\lambda$ that were established in [5]. $\mathbf{S}_\infty^\lambda$ satisfies Definition 2 (b),(c) by [5, (8)]. The fact that the excursion endpoints of $\mathbf{S}_\infty^\lambda$ do not have any isolated points almost surely follows directly from [5, Proposition 14 (d)]. Further, [5, Proposition 14 (b)] implies that, for any $u > 0$, $\mathbb{P}\left(S_\infty^\lambda(u) = \inf_{u'\leq u} S_\infty^\lambda(u')\right) = 0$. Taking the integral with respect to the Lebesgue measure and interchanging the limit by using Fubini's theorem, we conclude that almost surely

$$\int_0^T \mathbb{1}_{\left\{S_\infty^\lambda(u) = \inf_{u'\leq u} S_\infty^\lambda(u')\right\}} \mathrm{d}u = 0, \tag{5.25}$$

which verifies Definition 1 (a). Now, let $\mathbf{L}$ be the Lévy process defined as

$$L(t) = \sum_{i=1}^{\infty} \theta_i \left(\mathcal{N}_i(t) - (\theta_i/\mu)t\right) + \lambda t, \tag{5.26}$$

where $(\mathcal{N}_i(t))_{t\geq 0}$ is a Poisson process with rate θ_i which are independent for different i. Via the natural coupling that states $\mathcal{I}_i(t) \leq \mathcal{N}_i(t)$, we can assume that $S_\infty^\lambda(t) \leq L(t)$ for all $t > 0$. Using [8, Theorem VII.1],

$$\inf\{t > 0 : L(t) < 0\} = 0, \quad \text{almost surely.} \tag{5.27}$$

Moreover, for any stopping time $\mathcal{T} > 0$, $(S_\infty^\lambda(\mathcal{T}+t) - S_\infty^\lambda(\mathcal{T}))_{t\geq 0}$, conditioned on the sigma-field $\sigma(S_\infty^\lambda(s) : s \leq \mathcal{T})$, is distributed as a process defined in (1.5) for some random $\boldsymbol{\theta}$ and Λ. Now we can take $\mathcal{T}$ to be an excursion endpoint and an application of (5.27) verifies Definition 1 (b). □

5.4 Finite-dimensional convergence

As described in Section 4, the excursion lengths of the exploration process $\bar{\mathbf{S}}_n$ gives the total number of edges in the explored components. Lemma 16 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 1.

Lemma 16. *Let $N_n^\lambda(k)$ be the number of surplus edges discovered up to time k and $\bar{N}_n^\lambda(u) = N_n^\lambda(\lfloor ub_n \rfloor)$. Then, as $n \to \infty$,*

$$(\bar{\mathbf{S}}_n, \bar{\mathbf{N}}_n^\lambda) \xrightarrow{d} (\mathbf{S}_\infty^\lambda, \mathbf{N}^\lambda), \tag{5.28}$$

where $\mathbf{N}^\lambda$ is defined in (1.8).

Proof. We write $N_n^\lambda(l) = \sum_{i=2}^{l} \xi_i$, where $\xi_i = \mathbb{1}_{\{\mathscr{V}_i = \mathscr{V}_{i-1}\}}$. Let A_i denote the number of active half-edges after stage i while implementing Algorithm 1. Note that

$$\mathbb{P}\left(\xi_i = 1 | \mathscr{F}_{i-1}\right) = \frac{A_{i-1} - 1}{\ell_n - 2i - 1} = \frac{A_{i-1}}{\ell_n}(1 + O(i/n)) + O(n^{-1}), \tag{5.29}$$

uniformly for $i \leq Tb_n$ for any $T > 0$. Therefore, the instantaneous rate of change of the re-scaled process $\bar{\mathbf{N}}^\lambda$ at time t, conditional on the past, is

$$b_n \frac{A_{\lfloor tb_n \rfloor}}{n\mu}\left(1 + o(1)\right) + o(1) = \frac{1}{\mu}\mathrm{refl}(\bar{S}_n(t))\left(1 + o(1)\right) + o(1). \tag{5.30}$$

Recall from Theorem 10 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 $(\int_0^t \mathrm{refl}(\bar{S}_\infty(u))du)_{t \geq 0}$ has continuous sample paths. Therefore, the conditions of [33, Corollary 1, Page 388] are satisfied and the proof is complete. ☐

Theorem 17. *For any $m \geq 1$, as $n \to \infty$*

$$b_n^{-1}\left(|\mathscr{C}_{(1)}|, |\mathscr{C}_{(2)}|, \dots, |\mathscr{C}_{(m)}|\right) \xrightarrow{d} (\gamma_1(\lambda), \gamma_2(\lambda), \dots, \gamma_m(\lambda)) \tag{5.31}$$

with respect to the product topology, where $\gamma_i(\lambda)$ is the i^{th} largest excursion of $\bar{\mathbf{S}}_\infty$ defined in (1.5).

Proof. Fix any $m \geq 1$. Let $\mathscr{C}_{(i)}^T$ be the i^{th} largest component explored by Algorithm 1 up to time Tb_n. Denote by $\mathscr{D}_{(i)}^{\mathrm{ord},T}$ the i^{th} largest value of $(\sum_{k \in \mathscr{C}_{(i)}^T} d_k)_{i \geq 1}$. Let $g : \mathbb{R}^m \mapsto \mathbb{R}$ be a bounded continuous function. By Lemma 15 the sample paths of $\bar{\mathbf{S}}_\infty$ are almost surely good. Thus, using Theorem 8, Lemma 14 gives

$$\lim_{n\to\infty} \mathbb{E}\left[g\Big((2b_n)^{-1}\big(\mathscr{D}_{(1)}^{\mathrm{ord},T}, \mathscr{D}_{(2)}^{\mathrm{ord},T}, \dots, \mathscr{D}_{(m)}^{\mathrm{ord},T}\big)\Big)\right] = \mathbb{E}\left[g\big(\gamma_1^T(\lambda), \gamma_2^T(\lambda), \dots, \gamma_m^T(\lambda)\big)\right], \tag{5.32}$$

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 $(\gamma_i^T(\lambda))_{i \geq 1}$ is concentrated on$\{(x_1, x_2, \dots) : x_1 > x_2 > \dots\}$. Thus, using Lemma 16, it follows that

$$\lim_{n\to\infty} \mathbb{E}\left[g\Big(b_n^{-1}\big(|\mathscr{C}_{(1)}^T|, |\mathscr{C}_{(2)}^T|, \dots, |\mathscr{C}_{(m)}^T|\big)\Big)\right] = \mathbb{E}\left[g\big(\gamma_1^T(\lambda), \gamma_2^T(\lambda), \dots, \gamma_m^T(\lambda)\big)\right]. \tag{5.33}$$

Since $\mathbf{S}_\infty^\lambda$ satisfies Definition 2 (b), (c), it follows that

$$\lim_{T\to\infty} \mathbb{E}\left[g\big(\gamma_1^T(\lambda), \gamma_2^T(\lambda), \dots, \gamma_m^T(\lambda)\big)\right] = \mathbb{E}\left[g\big(\gamma_1(\lambda), \gamma_2(\lambda), \dots, \gamma_m(\lambda)\big)\right] \tag{5.34}$$

Finally, using Lemma 13, the proof of Theorem 17 is completed by (5.33) and (5.34).

☐

Proof of Theorem 1. The proof of Theorem 1 now follows directly from Theorem 17 and Proposition 11. ☐

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

The goal of this section is to prove the joint convergence of the component sizes and the surplus edges as described in Theorem 2. We start with a preparatory lemma:

Lemma 18. *The convergence in* (1.15) *holds with respect to the* $\ell^2_\downarrow \times \mathbb{N}^\infty$ *topology.*

Proof. Note that Lemma 13 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 14, 16, Remark 13 and Theorem 1. ☐

Recall the definition of the metric $\mathrm{d}_{\mathbb{U}}$ from (1.3). Using Lemma 18, it now remains to obtain a uniform summability condition on the tail of the sum of products of the scaled component sizes and the surplus edges. This is formally stated in Proposition 19 below. The proof is completed in the similar spirit as the finite third-moment case [20].

Proposition 19. *For any* $\varepsilon > 0$,

$$\lim_{\delta\to 0}\limsup_{n\to\infty}\mathbb{P}\bigg(\sum_{i:|\mathscr{C}_{(i)}|\leq \delta b_n}|\mathscr{C}_{(i)}|\times \mathrm{SP}(\mathscr{C}_{(i)}) > \varepsilon b_n\bigg) = 0. \tag{6.1}$$

The following estimate will be the crucial ingredient to complete the proof of Proposition 19. The proof of Lemma 20 is postponed to Appendix B since this uses similar ideas as [20].

Lemma 20. *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* $\mathrm{CM}_n(\boldsymbol{d})$ *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,*

$$\mathbb{P}\left(\mathrm{SP}(\mathscr{C}(V_n)) \geq K, |\mathscr{C}(V_n)| \in (\delta_K b_n, 2\delta_K b_n)\right) \leq \frac{C\sqrt{\delta}}{a_n K^{1.1}} \tag{6.2}$$

where C *is a fixed constant independent of* n, δ, K.

Proof of Proposition 19 using Lemma 20. First consider the case $\lambda < 0$. Fix any $\varepsilon, \eta > 0$. Note that

$$\begin{aligned}
&\mathbb{P}\bigg(\sum_{|\mathscr{C}_{(i)}|\leq\varepsilon b_n}|\mathscr{C}_{(i)}|\mathrm{SP}(\mathscr{C}_{(i)}) > \eta b_n\bigg) \leq \frac{1}{\eta b_n}\mathbb{E}\bigg[\sum_{i=1}^{\infty}|\mathscr{C}_{(i)}|\mathrm{SP}(\mathscr{C}_{(i)})\mathbb{1}_{\{|\mathscr{C}_{(i)}|\leq\varepsilon b_n\}}\bigg] \\
&= \frac{a_n}{\eta}\mathbb{E}\left[\mathrm{SP}(\mathscr{C}(V_n))\mathbb{1}_{\{|\mathscr{C}(V_n)|\leq\varepsilon b_n\}}\right] \\
&= \frac{a_n}{\eta}\sum_{k=1}^{\infty}\sum_{i\geq\log_2(1/(k^{0.12}\varepsilon))}\mathbb{P}\big(\mathrm{SP}(\mathscr{C}(V_n))\geq k, |\mathscr{C}(V_n)|\in(2^{-(i+1)}k^{-0.12}b_n, 2^{-i}k^{-0.12}b_n]\big) \\
&\leq \frac{C}{\eta}\sum_{k=1}^{\infty}\frac{1}{k^{1.1}}\sum_{i\geq\log_2(1/(k^{0.12}\varepsilon))}2^{-i/2} \leq \frac{C}{\eta}\sum_{k=1}^{\infty}\frac{\sqrt{\varepsilon}}{k^{1.04}} = O(\sqrt{\varepsilon}),
\end{aligned} \tag{6.3}$$

where the last-but-two step follows from Lemma 20. The proof of Proposition 19 now follows for $\lambda < 0$.

Now consider the case $\lambda > 0$. Fix a large integer $R \geq 1$ such that $\lambda - \sum_{i=1}^{R}\theta_i^2 < 0$. This can be done because $\boldsymbol{\theta} \notin \ell^2_\downarrow$. Using (5.10), for any $\eta > 0$, it is possible to choose $T > 0$ such that for all sufficiently large n,

$$\mathbb{P}\left(\text{all the vertices } 1,\dots,R \text{ are explored within time } Tb_n\right) > 1-\eta. \tag{6.4}$$

Let T_e denote the first time after Tb_n when we finish exploring a component. By Theorem 8, $(b_n^{-1}T_e)_{n\geq 1}$ is a tight sequence. Let $\mathcal{G}_T^*$ denote the graph obtained by removing the components explored up to time T_e. Then, $\mathcal{G}_T^*$ is again a configuration model conditioned on its degrees. Let ν_n^* denote the value of the criticality parameter for $\mathcal{G}^*$. Note that

$$\sum_{i\notin \mathscr{V}_{T_e}} d_i \geq \ell_n - 2Tb_n \implies \sum_{i\notin \mathscr{V}_{T_e}} d_i = \ell_n + o_{\mathbb{P}}(n), \tag{6.5}$$

and thus conditionally on $\mathscr{F}_{T_e}$ and the fact that $(1,\dots,R)$ are explored within time Tb_n,

$$\nu_n^* \leq \frac{\sum_{i\in[n]} d_i^2 - \sum_{i=1}^R d_i^2}{\sum_{i\notin \mathscr{V}_{T_e}} d_i} - 1 = 1 + c_n^{-1}\big(\lambda - \sum_{i=1}^R \theta_i^2\big) + o(c_n^{-1}). \tag{6.6}$$

Therefore, combining (6.4), (6.6), we can use Lemma 20 on $\mathcal{G}_T^*$ since $c_n(\nu_n^*-1)<0$. Thus, if $\mathscr{C}_{(i)}^*$ denotes the i^{th} largest component of $\mathcal{G}_T^*$, then

$$\lim_{T\to\infty}\lim_{\delta\to 0}\limsup_{n\to\infty}\mathbb{P}\bigg(\sum_{i:|\mathscr{C}_{(i)}^*|\leq \delta b_n} |\mathscr{C}_{(i)}^*|\times \mathrm{SP}(\mathscr{C}_{(i)}^*) > \varepsilon b_n\bigg) = 0. \tag{6.7}$$

To conclude the proof for the whole graph $\mathrm{CM}_n(\boldsymbol{d})$ (with $\lambda>0$), let

$$\mathcal{K}_n^T := \{i : |\mathscr{C}_{(i)}| \leq \delta b_n, |\mathscr{C}_{(i)}| \text{ is explored before the time } T_e\}.$$

Note that

$$\begin{aligned}\sum_{i\in\mathcal{K}_n^T} |\mathscr{C}_{(i)}|\cdot \mathrm{SP}(\mathscr{C}_{(i)}) &\leq \Big(\sum_{i\in\mathcal{K}_n}|\mathscr{C}_{(i)}|^2\Big)^{1/2}\times\Big(\sum_{i\in\mathcal{K}_n}\mathrm{SP}(\mathscr{C}_{(i)})^2\Big)^{1/2}\\ &\leq \Big(\sum_{|\mathscr{C}_{(i)}|\leq\delta b_n}|\mathscr{C}_{(i)}|^2\Big)^{1/2}\times \mathrm{SP}(T_e),\end{aligned} \tag{6.8}$$

where $\mathrm{SP}(t)$ is the number of surplus edges explored up to time tb_n and we have used the fact that $\sum_{i\in\mathcal{K}_n}\mathrm{SP}(\mathscr{C}_{(i)})^2 \leq (\sum_{i\in\mathcal{K}_n}\mathrm{SP}(\mathscr{C}_{(i)}))^2 \leq \mathrm{SP}(T_e)^2$. From Lemma 16 and Proposition 11 we can conclude that for any $T>0$,

$$\lim_{\delta\to 0}\limsup_{n\to\infty}\mathbb{P}\bigg(\sum_{i\in\mathcal{K}_n^T}|\mathscr{C}_{(i)}|\cdot\mathrm{SP}(\mathscr{C}_{(i)}) > \varepsilon b_n\bigg) = 0. \tag{6.9}$$

The proof is now complete for the case $\lambda>0$ by combining (6.7) and (6.9). ☐

7 Proof for simple graphs

In this section, we give a proof of Theorem 3. Let $\mathbb{P}_s(\cdot)$ (respectively $\mathbb{E}_s[\cdot]$) denote the probability measure (respectively the expectation) conditionally on the graph $\mathrm{CM}_n(\boldsymbol{d})$ being simple. For any process $\mathbf{X}$ on $\mathbb{D}([0,\infty),\mathbb{R})$, we define $\mathbf{X}^T := (X(t))_{t\leq T}$. Thus the truncated process $\mathbf{X}^T$ is $\mathbb{D}([0,T],\mathbb{R})$-valued. Now, by [26, Theorem 1.1], $\liminf_{n\to\infty}\mathbb{P}(\mathrm{CM}_n(\boldsymbol{d}) \text{ is simple}) > 0$. This fact ensures that, under the conditional measure $\mathbb{P}_s$, $(b_n^{-1}|\mathscr{C}_{(i)}|)_{i\geq 1}$ is tight with respect to the $\ell^2_{\downarrow}$ topology. Therefore, to conclude Theorem 3, it suffices to show that the exploration process $\bar{\mathbf{S}}_n$, defined in (4.3), has the same limit (in distribution) under $\mathbb{P}_s$ as obtained in Theorem 8 so that the finite-dimensional limit of $(b_n^{-1}|\mathscr{C}_{(i)}|)_{i\geq 1}$ remains unchanged under $\mathbb{P}_s$. Thus, it is enough to show that for any bounded continuous function $f:\mathbb{D}([0,T],\mathbb{R})\mapsto\mathbb{R}$,

$$\big|\mathbb{E}[f(\bar{\mathbf{S}}_n^T)] - \mathbb{E}_s[f(\bar{\mathbf{S}}_n^T)]\big| \to 0. \tag{7.1}$$

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 1, d_{v_l} the degree of v_l and $(e_1, \dots, e_r)$ the ordered set of active half-edges of v_l when v_l is declared to be exploring. Note that, for $l \leq Tb_n$, 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,

$$\mathbb{E}\left[\#\{\text{self-loops or multiple edges discovered while } v_l \text{ is exploring}\} | \mathscr{F}_{l-1}\right] \leq \frac{2d_{v_l}^2}{\ell_n'}. \tag{7.2}$$

Thus, for any $T > 0$,

$$\begin{aligned}
&\mathbb{E}\left[\#\{\text{self-loops or multiple edges discovered up to time } Tb_n\right] \\
&\leq \frac{2}{\ell_n'}\mathbb{E}\bigg[\sum_{i\in[n]} d_i^2 \mathcal{I}_i^n(Tb_n)\bigg] = \frac{2}{\ell_n'}\mathbb{E}\bigg[\sum_{i=1}^{K} d_i^2 \mathcal{I}_i^n(Tb_n)\bigg] + \frac{2}{\ell_n'}\mathbb{E}\bigg[\sum_{i=K+1}^{n} d_i^2 \mathcal{I}_i^n(Tb_n)\bigg],
\end{aligned} \tag{7.3}$$

where $\mathcal{I}_i^n(l) = \mathbb{1}_{\{i \in \mathscr{V}_l\}}$. Now, using Assumption 1 (i), for every fixed $K \geq 1$,

$$\frac{2}{\ell_n'}\mathbb{E}\bigg[\sum_{i=1}^{K} d_i^2 \mathcal{I}_i^n(Tb_n)\bigg] \leq \frac{2}{\ell_n'}\sum_{i=1}^{K} d_i^2 \to 0, \tag{7.4}$$

since $2\alpha - 1 < 0$. Moreover, recall from (4.6) that $\mathbb{P}\left(\mathcal{I}_i^n(Tb_n) = 1\right) \leq Tb_n d_i/\ell_n'$. Therefore, for some constant $C > 0$,

$$\frac{2}{\ell_n'}\mathbb{E}\bigg[\sum_{i=K+1}^{n} d_i^2 \mathcal{I}_i^n(Tb_n)\bigg] \leq \frac{Tb_n}{\ell_n'^2}\sum_{i=K+1}^{n} d_i^3 \leq C\bigg(a_n^{-3}\sum_{i=K+1}^{n} d_i^3\bigg), \tag{7.5}$$

which, by Assumption 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 $n \to \infty$,

$$\mathbb{P}\left(\text{at least one self-loop or multiple edge is discovered before time } Tb_n\right) \to 0. \tag{7.6}$$

Now,

$$\begin{aligned}
&\mathbb{E}\left[f(\bar{\mathbf{S}}_n^T)\mathbb{1}_{\{\mathrm{CM}_n(\boldsymbol{d}) \text{ is simple}\}}\right] \\
&= \mathbb{E}\left[f(\bar{\mathbf{S}}_n^T)\mathbb{1}_{\{\text{no self-loops or multiple edges found after time } Tb_n\}}\right] + o(1) \\
&= \mathbb{E}\left[f(\bar{\mathbf{S}}_n^T)\mathbb{P}\left(\text{no self-loops or multiple edges found after time } Tb_n | \mathscr{F}_{Tb_n}\right)\right] + o(1),
\end{aligned} \tag{7.7}$$

Define, $T_e = \inf\{l \geq Tb_n : \text{a component is finished exploring at time } l\}$. Using the fact that $(b_n^{-1}T_e)_{n\geq 1}$ 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 19, 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 ν_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 1 (ii) (4.5), for any $t > 0$,

$$\limsup_{n\to\infty} \mathbb{E}\bigg[a_n^{-2}\sum_{i\in[n]} d_i^2 \mathcal{I}_i(tb_n)\bigg] \leq \limsup_{n\to\infty} a_n^{-2} tb_n \frac{\sum_{i\in[n]} d_i^3}{\ell_n - 2tb_n} < \infty, \tag{7.8}$$

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 [41, Theorem 7.11] we get

$$\mathbb{P}\left(\mathcal{G}^* \text{ is simple}\middle| \mathscr{F}_{T_e}\right) \xrightarrow{\mathbb{P}} \mathrm{e}^{-3/4}. \tag{7.9}$$

Now using (7.7), (7.9) and the dominated convergence theorem, we conclude that

$$\mathbb{E}\left[f(\bar{\mathbf{S}}_n^T)\mathbb{1}_{\{\mathrm{CM}_n(\boldsymbol{d}) \text{ is simple}\}}\right] = \mathbb{E}\left[f(\bar{\mathbf{S}}_n^T)\right]\mathbb{P}\left(\mathrm{CM}_n(\boldsymbol{d}) \text{ is simple}\right) + o(1). \tag{7.10}$$

Therefore, (7.1) follows and the proof of Theorem 3 is complete. ∎

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 $(\mathscr{W}_i)_{i\geq 1}$ when the weight sequence satisfies some regularity conditions given below:

Assumption 3. The weight sequences $\boldsymbol{w} = (w_i)_{i\in[n]}$ 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 21. *Consider* $\mathrm{CM}_n(\boldsymbol{d})$ *satisfying* Assumption 1 *and a weight sequence* $\boldsymbol{w}$ *satisfying* Assumption 3. *Denote* $\mathbf{Z}_n^w = \mathrm{ord}(b_n^{-1}\mathscr{W}_i, \mathrm{SP}(\mathscr{C}_{(i)}))_{i\geq 1}$ *and* $\mathbf{Z}^w := \mathrm{ord}(\mu_w\gamma_i(\lambda), N(\gamma_i))_{i\geq 1}$, *where* $\gamma_i(\lambda)$, *and* $N(\gamma_i)$ *are defined in* Theorem 2. *As* $n\to\infty$,

$$\mathbf{Z}_n^w \xrightarrow{d} \mathbf{Z}^w, \tag{8.1}$$

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

The proof Theorem 21 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 11 and 19 and therefore we only provide a sketch with pointers to all the necessary ingredients. Recall that $\mathcal{I}_i^n(l) = \mathbb{1}_{\{i\in\mathscr{V}_l\}}$, where $\mathscr{V}_l$ is the set of discovered vertices upto time l by Algorithm 1.

Lemma 22. *Under* Assumptions 1, 3, *for any* $T>0$,

$$\sup_{u\leq T}\bigg|\sum_{i\in[n]} w_i\mathcal{I}_i^n(ub_n) - \frac{\sum_{i\in[n]} d_i w_i}{\ell_n} ub_n\bigg| = O_{\mathbb{P}}(a_n). \tag{8.2}$$

Consequently, for each fixed $i\geq 1$,

$$\mathscr{W}_i = \mu_w\big|\mathscr{C}_{(i)}\big| + o_{\mathbb{P}}(b_n). \tag{8.3}$$

Proof. Fix any $T>0$. Define ,

$$W_n(l) = \sum_{i\in[n]} w_i\mathcal{I}_i^n(l) - \frac{\sum_{i\in[n]} d_i w_i}{\ell_n} l. \tag{8.4}$$

The goal is to use the supermartingale inequality (4.16) in the same spirit as in the proof of (4.17). Firstly, observe from (4.6) that

$$\begin{aligned}\mathbb{E}[W_n(l+1) - W_n(l)|\mathscr{F}_l] &= \mathbb{E}\bigg[\sum_{i\in[n]} w_i\left(\mathcal{I}_i^n(l+1) - \mathcal{I}_i^n(l)\right)\Big|\mathscr{F}_l\bigg] - \frac{\sum_{i\in[n]} d_i w_i}{\ell_n}\\ &= \sum_{i\in[n]} w_i\mathbb{E}\left[\mathcal{I}_i^n(l+1)\middle|\mathscr{F}_l\right]\mathbb{1}_{\{\mathcal{I}_i^n(l)=0\}} - \frac{\sum_{i\in[n]} d_i w_i}{\ell_n} \geq 0,\end{aligned} \tag{8.5}$$

uniformly over $l \leq Tb_n$ and therefore, $(W_n(l))_{l=1}^{Tb_n}$ is a sub-martingale. Let $\ell_n' = \ell_n - 2Tb_n - 1$. Using (4.8), we compute

$$\begin{aligned} \big|\mathbb{E}[W_n(l)]\big| &= \sum_{i\in[n]} w_i \left(\mathbb{P}\left(\mathcal{I}_i^n(l) = 1\right) - \frac{d_i}{\ell_n}\right) \\ &\leq \sum_{i\in[n]} w_i \left(1 - \left(1 - \frac{d_i}{\ell_n'}\right)^l - \frac{d_i}{\ell_n'} l\right) + l \sum_{i\in[n]} w_i \left(\frac{d_i}{\ell_n'} - \frac{d_i}{\ell_n}\right) \\ &\leq 2(2Tb_n)^2 \frac{\sum_{i\in[n]} d_i^2 w_i}{\ell_n'^2} = O(b_n^2 a_n^3/n^2) = O(a_n), \end{aligned} \tag{8.6}$$

uniformly over $l \leq Tb_n$. Also, using (4.12), (4.13), and Assumption 3 (ii),

$$\mathrm{Var}(W_n(l)) \leq \sum_{i\in[n]} w_i^2 \mathrm{var}(\mathcal{I}_i^n(l)) \leq Tb_n \frac{\sum_{i\in[n]} d_i w_i^2}{\ell_n'} = O(a_n^2), \tag{8.7}$$

uniformly over $l \leq Tb_n$. Using (4.16), (8.6) and (8.7), we conclude the proof of (8.2). The proof of (8.3) follows using Lemma 13 and simply observing that $a_n = o(b_n)$. ☐

Proof of Theorem 21. Lemma 22 ensures the finite-dimensional convergence in (8.1). Thus, the proof is complete if we can show that, for any $\varepsilon > 0$

$$\lim_{K\to\infty} \limsup_{n\to\infty} \mathbb{P}\left(\sum_{i>K} \mathscr{W}_i^2 > \varepsilon b_n^2\right) = 0, \tag{8.8a}$$

and

$$\lim_{\delta\to 0} \limsup_{n\to\infty} \mathbb{P}\left(\sum_{|\mathscr{C}_{(i)}|\leq \delta b_n} \mathscr{W}_i \times \mathrm{SP}(\mathscr{C}_{(i)}) > \varepsilon b_n\right) = 0. \tag{8.8b}$$

The arguments for proving (8.8a), and (8.8b) are similar to Propositions 11 and 19 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 11 is Lemma 12, and the proof Lemma 12 uses the fact that the expected sum of squares of the cluster sizes can be written in terms of susceptibility functions in (5.7) and then we made use of the estimate for the susceptibility function in (5.4). 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 $\mathrm{CM}_n(\boldsymbol{d})$,

$$\mathbb{E}\Big[\sum_{i\geq 1} \mathscr{W}_i^2\Big] = \ell_n^w \mathbb{E}\big[\mathscr{W}(V_n')\big]. \tag{8.9}$$

Now, [27, Lemma 5.2] can be extended using an identical argument to compute the weight-based susceptibility function in the right hand side of (8.9). See Lemma 31 given in Appendix A. The proof of (8.8b) can also be completed using an identical argument as Proposition 19 by observing that

$$\mathbb{P}\left(\sum_{|\mathscr{C}_{(i)}|\leq \delta b_n} \mathscr{W}_i \times \mathrm{SP}(\mathscr{C}_{(i)}) > \varepsilon b_n\right) \leq \frac{\ell_n^w}{\varepsilon b_n} \mathbb{E}\left[\mathrm{SP}(\mathscr{C}(V_n')) \mathbb{1}_{\{|\mathscr{C}(V_n')|\leq \delta b_n\}}\right]. \tag{8.10}$$

Moreover, an analog of Lemma 20 also holds for V_n' (see Appendix B), and the proof of (8.8b) can now be completed in an identical manner as the proof of Proposition 19. ☐

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 21, yields the following result about the degree composition of the largest clusters:

Corollary 23. *Consider* $\mathrm{CM}_n(\boldsymbol{d})$ *satisfying* Assumption 1. *Let* $v_k(G)$ *denote the number of vertices of degree* k *in the graph* G*. Then, for any fixed* $i \geq 1$,

$$v_k(\mathscr{C}_{(i)}) = \frac{kr_k}{\mu}|\mathscr{C}_{(i)}| + o_{\mathbb{P}}(b_n), \tag{8.11}$$

where $r_k = \mathbb{P}(D = k)$*. Denote* $\mathbf{Z}_n^k = \mathrm{ord}(b_n^{-1}v_k(\mathscr{C}_{(i)}), \mathrm{SP}(\mathscr{C}_{(i)}))_{i\geq 1}$ *and* $\mathbf{Z}^k := \mathrm{ord}(\frac{kr_k}{\mu}\gamma_i(\lambda), N(\gamma_i))_{i\geq 1}$*, where* $\gamma_i(\lambda)$*, and* $N(\gamma_i)$ *are defined in* Theorem 2*. As* $n \to \infty$,

$$\mathbf{Z}_n^k \xrightarrow{d} \mathbf{Z}^k, \tag{8.12}$$

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

Proof. The proof follows directly from Theorem 21 by putting $w_i = \mathbb{1}_{\{d_i=k\}}$. The fact that this weight sequence satisfies Assumption 3 is a consequence of Assumption 1. ☐

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 4. As discussed earlier, $\mathrm{CM}_n(\boldsymbol{d}, p)$ is obtained by first constructing $\mathrm{CM}_n(\boldsymbol{d})$ and then deleting each edge with probability $1-p$, independently of each other, and the graph $\mathrm{CM}_n(\boldsymbol{d})$. An interesting property of the configuration model is that $\mathrm{CM}_n(\boldsymbol{d}, p)$ is also distributed as a configuration model conditional on the degrees [22]. The rough idea here is to show that the degree distribution of $\mathrm{CM}_n(\boldsymbol{d}, p_n(\lambda))$ satisfies Assumption 1, where $p_n(\lambda)$ is given by Assumption 2. This allows us to invoke Theorem 2 and complete the proof of Theorem 4. Recall from Assumption 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 [25] that is easier to work with.

Algorithm 2 (Construction of $\mathrm{CM}_n(\boldsymbol{d}, p_n)$)**.** 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 $\tilde{\boldsymbol{d}} = (\tilde{d}_i)_{i\in[\tilde{n}]}$, i.e., $\tilde{d}_i \sim \mathrm{Bin}(d_i, \sqrt{p_n})$ for $i \in [n]$ and $\tilde{d}_i = 1$ for $i \in [\tilde{n}] \setminus [n]$;

(S2) Construct $\mathrm{CM}_{\tilde{n}}(\tilde{\boldsymbol{d}})$ independently of (S1);

(S3) Delete all the red vertices and the edges attached to them.

It was also shown in [25] that the obtained multigraph has the same distribution as $\mathrm{CM}_n(\boldsymbol{d}, p)$ 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 14. Notice that Algorithm 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 $\tilde{\boldsymbol{d}}$ also satisfies Assumption 1 (ii). Note that the total number of half-edges remain unchanged during the explosion in Algorithm 2 (S1) and therefore, $\sum_{i\in[\tilde{n}]}\tilde{d}_i=\sum_{i\in[n]}d_i$ and by Assumption 2 (i),

$$\frac{1}{n}\sum_{i\in[\tilde{n}]}\tilde{d}_i\to\mu\quad \mathbb{P}_p \text{ a.s.} \tag{9.1}$$

This verifies the first moment condition in Assumption 1 (ii) for the percolated degree sequence $\mathbb{P}_p$ a.s. Let I_{ij}:= the indicator of the j^{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

$$\mathbf{I}:=(I_{ij})_{j\in[d_i],i\in[n]}\quad \text{and}\quad f_1(\mathbf{I}):=\sum_{i\in[n]}\tilde{d}_i(\tilde{d}_i-1). \tag{9.2}$$

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 [29, Corollary 2.27] yields

$$\mathbb{P}_p\bigg(\Big|\sum_{i\in[n]}\tilde{d}_i(\tilde{d}_i-1)-p_n\sum_{i\in[n]}d_i(d_i-1)\Big|>t\bigg)\leq 2\exp\bigg(-\frac{t^2}{2\sum_{i\in[n]}d_i(d_i+1)^2}\bigg). \tag{9.3}$$

Now by Assumption 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 (9.3) and the Borel-Cantelli lemma we conclude that

$$\sum_{i\in[n]}\tilde{d}_i(\tilde{d}_i-1)=p_n\sum_{i\in[n]}d_i(d_i-1)+o(nc_n^{-1})\quad \mathbb{P}_p \text{ a.s.} \tag{9.4}$$

Thus, using Assumption 2, the second moment condition in Assumption 1 (ii) is verified for the percolated degree sequence $\mathbb{P}_p$ a.s. Let $\tilde{d}_{(i)}$ denote the i^{th} largest value of $(\tilde{d}_i)_{i\in[\tilde{n}]}$. The third-moment condition in Assumption 1 (ii) is obtained by noting that $\tilde{d}_i\leq d_i$ for all $i\in[n]$ and

$$\begin{aligned}
&\lim_{K\to\infty}\limsup_{n\to\infty}a_n^{-3}\sum_{i=K+1}^{\tilde{n}}\tilde{d}_{(i)}^3\leq \lim_{K\to\infty}\limsup_{n\to\infty}a_n^{-3}\sum_{i=K+1}^{\tilde{n}}\tilde{d}_i^3\\
&\leq \lim_{K\to\infty}\limsup_{n\to\infty}a_n^{-3}\Big(\sum_{i=K+1}^{n}\tilde{d}_i^3+n_+\Big)\leq \lim_{K\to\infty}\limsup_{n\to\infty}a_n^{-3}\Big(\sum_{i=K+1}^{n}d_i^3+n_+\Big)\to 0\quad \mathbb{P}_p \text{ a.s.,}
\end{aligned} \tag{9.5}$$

where we have used Assumption 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 $\tilde{\boldsymbol{d}}$ satisfies Assumption 1 (iii) note that by (9.4),

$$\frac{\sum_{i\in[\tilde{n}]}\tilde{d}_i(\tilde{d}_i-1)}{\sum_{i\in[\tilde{n}]}\tilde{d}_i}=p_n\frac{\sum_{i\in[n]}d_i(d_i-1)}{\sum_{i\in[n]}d_i}+o(c_n^{-1})=1+\lambda c_n^{-1}+o(c_n^{-1})\quad \mathbb{P}_p \text{ a.s.,} \tag{9.6}$$

where the last step follows from Assumption 2 (ii). Assumption 1 (iv) is trivially satisfied by $\tilde{\boldsymbol{d}}$. Finally, in order to verify Assumption 1 (i), it suffices to show that

$$\frac{\tilde{d}_{(i)}}{a_n}\to\theta_i\sqrt{p},\quad \mathbb{P}_p \text{ a.s.,} \tag{9.7}$$

where $p=1/\nu$. Recall that $\tilde{d}_i\sim\text{Bin}(d_i,\sqrt{p_n})$. A standard concentration inequality for the binomial distribution [29, (2.9)] yields that, for any $0<\varepsilon\leq 3/2$,

$$\mathbb{P}(|\tilde{d}_i-d_i\sqrt{p_n}|>\varepsilon d_i\sqrt{p_n})\leq 2\exp(-\varepsilon^2 d_i\sqrt{p_n}/3), \tag{9.8}$$

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 (9.5) yields that

$$\lim_{K\to\infty} \limsup_{n\to\infty} a_n^{-3} \max_{i>K} \tilde{d}_i^3 = 0. \tag{9.9}$$

Now, since $\boldsymbol{\theta}$ is an ordered vector, the proof of (9.7) follows.

To summarize, the above discussion in (9.1), (9.4), (9.5), and (9.7) yields that the degree sequence $\tilde{d}$ satisfies all the conditions in Assumption 1. Therefore, Theorem 2 can be applied to $\mathrm{CM}_{\tilde{n}}(\tilde{\boldsymbol{d}})$. Denote by $\tilde{\mathscr{C}}_{(i)}$ the i^{th} largest component of $\mathrm{CM}_{\tilde{n}}(\tilde{\boldsymbol{d}})$. Let $\tilde{\mathbf{Z}}_n = \mathrm{ord}(b_n^{-1}|\tilde{\mathscr{C}}_{(i)}|, \mathrm{SP}(\tilde{\mathscr{C}}_{(i)})_{i\geq 1}$ and $\tilde{\mathbf{Z}} := \mathrm{ord}(\tilde{\gamma}_i(\lambda), N(\tilde{\gamma}_i))_{i\geq 1}$, where $\gamma_i(\lambda)$, and $N(\gamma_i)$ are defined in Theorem 4. Now, Theorem 2 implies

$$\tilde{\mathbf{Z}}_n \xrightarrow{d} \tilde{\mathbf{Z}}, \tag{9.10}$$

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

Since the percolated degree sequence satisfies Assumption 1 $\mathbb{P}_p$ a.s., (8.11) holds for $\tilde{\mathscr{C}}_{(i)}$ also. Let $v_1^d(\tilde{\mathscr{C}}_{(i)})$ be the number of degree-one vertices of $\tilde{\mathscr{C}}_{(i)}$ which are deleted while creating the graph $\mathrm{CM}_n(\boldsymbol{d}, p_n)$ from $\mathrm{CM}_{\tilde{n}}(\tilde{\boldsymbol{d}})$. Since the vertices are to be chosen uniformly from all degree-one vertices as described in (S3′),

$$\begin{aligned} v_1^d(\tilde{\mathscr{C}}_{(i)}) &= \frac{n_+}{\tilde{n}_1} v_1(\tilde{\mathscr{C}}_{(i)}) + o_{\mathbb{P}}(b_n) = \frac{n_+}{\tilde{n}_1}\frac{\tilde{n}_1}{\ell_n}|\tilde{\mathscr{C}}_{(i)}| + o_{\mathbb{P}}(b_n) = \frac{n_+}{\ell_n}|\tilde{\mathscr{C}}_{(i)}| + o_{\mathbb{P}}(b_n) \\ &= \frac{\mu(1-\sqrt{p}_n) + o(1)}{\mu + o(1)}|\tilde{\mathscr{C}}_{(i)}| + o_{\mathbb{P}}(b_n) = \big(1-\sqrt{p}_n\big)|\tilde{\mathscr{C}}_{(i)}| + o_{\mathbb{P}}(b_n), \end{aligned} \tag{9.11}$$

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 $\mathrm{CM}_{\tilde{n}}(\tilde{\boldsymbol{d}})$. Moreover, the removal of degree-one vertices does not effect the count of surplus edges. Therefore, the proof of Theorem 4 is complete by using Corollary 23.

10 Convergence to AMC

Let us give an overview of the organization of this section: In Section 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 [20, Section 8], and we only give the overall idea and the necessary details specific to this paper.

10.1 The dynamic construction and the coupling

Let us consider graphs generated dynamically as follows:

Algorithm 3. Let $s_1(t)$ be the total number of unpaired or *open* half-edges at time t, and Ξ_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 Ξ_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 $\mathrm{CM}_n(\boldsymbol{d})$ 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 *sandwich* the graph process $(\mathrm{CM}_n(\boldsymbol{d}, p_n(\lambda)))_{\lambda\in\mathbb{R}}$. This result was proved in [20, Proposition 28]. The proof is identical under Assumption 2 and therefore is omitted here. Define,

$$t_n(\lambda) = \frac{1}{2}\log\left(\frac{\nu_n}{\nu_n - 1}\right) + \frac{1}{2(\nu_n - 1)}\frac{\lambda}{c_n}. \tag{10.1}$$

Proposition 24. *Fix $-\infty < \lambda_\star < \lambda^\star < \infty$. There exists a coupling such that with high probability*

$$\mathcal{G}_n(t_n(\lambda) - \varepsilon_n) \subset \mathrm{CM}_n(\boldsymbol{d}, p_n(\lambda)) \subset \mathcal{G}_n(t_n(\lambda) + \varepsilon_n), \quad \forall \lambda \in [\lambda_\star, \lambda^\star] \tag{10.2a}$$

and

$$\mathrm{CM}_n(\boldsymbol{d}, p_n(\lambda) - \varepsilon_n) \subset \mathcal{G}_n(t_n(\lambda)) \subset \mathrm{CM}_n(\boldsymbol{d}, p_n(\lambda) + \varepsilon_n), \quad \forall \lambda \in [\lambda_\star, \lambda^\star] \tag{10.2b}$$

where $\varepsilon_n = cn^{-\gamma_0}$, for some $\eta < \gamma_0 < 1/2$ and the constant c does not depend on λ.

From here onward, we augment λ to a predefined notation to emphasize the dependence on λ. We write $\mathscr{C}_{(i)}(\lambda)$ for the i^{th} largest component of $\mathcal{G}_n(t_n(\lambda))$ and define

$$\mathcal{O}_i(\lambda) = \#\text{ open half-edges in } \mathscr{C}_{(i)}(\lambda). \tag{10.3}$$

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 $\mathbb{U}^0_\downarrow$. 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 $\ell_n^o(\lambda) = \sum_{i\geq 1}\mathcal{O}_i(\lambda)$. Note that

$$\ell_n^o(\lambda) = \frac{n\mu(\nu-1)}{\nu}(1 + o_{\mathbb{P}}(1)). \tag{10.4}$$

(10.4) is a consequence of [9, 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, between time $[t_n(\lambda), t_n(\lambda + \mathrm{d}\lambda)]$, the i^{th} and j^{th} $(i > j)$ largest components, merge at rate

$$2\mathcal{O}_i(\lambda)\mathcal{O}_j(\lambda) \times \frac{1}{\ell_n^o(\lambda) - 1} \times \frac{1}{2(\nu_n - 1)c_n} \approx \frac{\nu}{\mu(\nu-1)^2}\left(b_n^{-1}\mathcal{O}_i(\lambda)\right)\left(b_n^{-1}\mathcal{O}_j(\lambda)\right), \tag{10.5}$$

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

$$\mathcal{O}_i(\lambda)(\mathcal{O}_i(\lambda) - 1) \times \frac{1}{\ell_n^o(\lambda) - 1} \times \frac{1}{2(\nu_n - 1)c_n} \approx \frac{\nu}{2\mu(\nu-1)^2}\left(b_n^{-1}\mathcal{O}_i(\lambda)\right)^2, \tag{10.6}$$

and $\mathscr{C}_{(i)}(\lambda)$ becomes a component with surplus edges $\mathrm{SP}(\mathscr{C}_{(i)}(\lambda))+1$ 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 and the system to 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 [20, Algorithm 7] that in fact evolves as augmented multiplicative coalescent:

Algorithm 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 $\mathscr{O}$ and create an edge between the corresponding vertices. However, the selected half-edges are kept alive, so that they can be selected again.

Remark 15. The only difference between Algorithms 3 and 4, is that the *paired* half-edges are not discarded and thus more edges are created by Algorithm 4. Thus, there is a natural coupling between the graphs generated by Algorithms 3 and 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 4 will be called *bad* edges.

In the subsequent part of this paper, 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 $\mathrm{ord}(\beta_n^{-1}\bar{\mathcal{O}}_i(\lambda), \mathrm{SP}(\bar{\mathscr{C}}_{(i)}(\lambda)))_{i\geq 1}$. Using an argument identical to (10.5), and (10.6), 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 $\bar{\mathbf{Z}}_n^{o,\mathrm{scl}}[\lambda_\star, \lambda^\star]$.

10.1.1 Augmented multiplicative coalescent with mass and weight

The near Feller property of the augmented multiplicative coalescent [10, 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,

- ▷ 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 $Y_i(t) + Y_j(t)$.
- ▷ for each i, attribute $Y_i(t)$ increases by 1 at rate $Y_i^2(t)/2$.

For $(\boldsymbol{x}, \boldsymbol{y}), (\boldsymbol{z}, \boldsymbol{y}) \in \mathbb{U}^0_\downarrow$, we write $(\boldsymbol{x}, \boldsymbol{z}, \boldsymbol{y})$ for $((\boldsymbol{x}, \boldsymbol{y}), (\boldsymbol{z}, \boldsymbol{y})) \in (\mathbb{U}^0_\downarrow)^2$. Denote by $\mathrm{MC}_2(\boldsymbol{x}, \boldsymbol{z}, t)$ and $\mathrm{AMC}_2(\boldsymbol{x}, \boldsymbol{z}, \boldsymbol{y}, t)$ 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 $\boldsymbol{x}$, weight $\boldsymbol{z}$ and attribute value $\boldsymbol{y}$. We will need the following theorem:

Theorem 25. *Suppose that $(\boldsymbol{x}_n, \boldsymbol{z}_n, \boldsymbol{y}_n) \to (\boldsymbol{x}, \boldsymbol{x}, \boldsymbol{y})$ in $(\mathbb{U}^0_\downarrow)^2$ and $\sum_i x_i = \infty$. Then, for any $t \geq 0$*

$$\mathrm{AMC}_2(\boldsymbol{x}_n, \boldsymbol{z}_n, \boldsymbol{y}_n) \xrightarrow{d} \mathrm{AMC}_2(\boldsymbol{x}, \boldsymbol{x}, \boldsymbol{y}). \tag{10.7}$$

Proof. By [20, Theorem 29],

$$\mathrm{MC}_2(\boldsymbol{x}_n, \boldsymbol{z}_n, t) \xrightarrow{d} \mathrm{MC}_2(\boldsymbol{x}, \boldsymbol{x}, t). \tag{10.8}$$

For $\boldsymbol{x}_n = (x_i^n)_{i\geq 1}$, and $\boldsymbol{z}_n = (z_i^n)_{i\geq 1}$ 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 $\ell^2_\downarrow$. Let us denote by $\mathrm{AMC}_1(\boldsymbol{x}, \boldsymbol{y}, t)$ 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 [10, Theorem 3.1] to conclude that $\mathrm{AMC}_1(\boldsymbol{x}_n, \boldsymbol{y}_n, t) \xrightarrow{d} \mathrm{AMC}_1(\boldsymbol{x}, \boldsymbol{y}, t)$. Moreover, $\mathrm{AMC}_2(\boldsymbol{w}_n^+, \boldsymbol{w}_n^+, y_n, t)$ and $\mathrm{AMC}_2(\boldsymbol{w}_n^-, \boldsymbol{w}_n^-, y_n, t)$ converges to the same limit. For $(\boldsymbol{x}, \boldsymbol{z}, \boldsymbol{y}) \in (\mathbb{U}^0_\downarrow)^2$, if $S_{\mathrm{pr}}(\boldsymbol{x}, \boldsymbol{z}, \boldsymbol{y}) = \sum_i z_i y_i$, then under the subgraph coupling

$$S_{\mathrm{pr}}(\mathrm{AMC}_2(\boldsymbol{w}_n^+, \boldsymbol{w}_n^+, y_n, t)) - S_{\mathrm{pr}}(\mathrm{AMC}_2(\boldsymbol{w}_n^-, \boldsymbol{w}_n^-, y_n, t)) \xrightarrow{\mathbb{P}} 0, \tag{10.9}$$

which implies that

$$(\mathrm{AMC}_1(\boldsymbol{x}_n, \boldsymbol{y}_n, t), S_{\mathrm{pr}}(\mathrm{AMC}_2(\boldsymbol{x}_n, \boldsymbol{z}_n, y_n, t))) \xrightarrow{d} (\mathrm{AMC}_1(\boldsymbol{x}, \boldsymbol{y}, t), S_{\mathrm{pr}}(\mathrm{AMC}_2(\boldsymbol{x}, \boldsymbol{x}, y, t))). \tag{10.10}$$

Now, using (10.8), (10.10), an application of [10, Lemma 4.11] concludes the proof. ☐

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 25 for deducing the scaling limits of the required quantities for the graph $\bar{\mathcal{G}}_n(t_n(\lambda))$.

Lemma 26. *There exists a constant $\kappa > 0$ such that, for any $i \geq 1$,*

$$\mathcal{O}_i(\lambda) = \kappa|\mathscr{C}_{(i)}(\lambda)| + o_{\mathbb{P}}(b_n). \tag{10.11}$$

Further, $(\mathbf{Z}_n^o(\lambda))_{n\geq 1}$ is tight in $\mathbb{U}_\downarrow^0$.

Proof. Let $(d_k^\lambda)_{k\in[n]}$ denote the degree sequence of $\mathrm{CM}_n(\boldsymbol{d}, p_n(\lambda))$ and define

$$\mathcal{O}_i^p(\lambda) = \sum_{k\in\mathscr{C}_{(i)}^p(\lambda)} (d_k - d_k^\lambda) = \sum_{k\in\mathscr{C}_{(i)}^p(\lambda)} d_k - 2(|\mathscr{C}_{(i)}^p(\lambda)| - 1 + \mathrm{SP}(\mathscr{C}_{(i)}^p(\lambda))). \tag{10.12}$$

Using (10.2b) 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 $\tilde{\mathscr{C}}$ of $\mathrm{CM}_{\tilde{n}}(\tilde{\boldsymbol{d}})$, the corresponding component in the percolated graph is obtained by cleaning up $R(\tilde{\mathscr{C}})$ red degree-one vertices. Thus, the degree deficiency of that percolated cluster is given by

$$\sum_{k\in\tilde{\mathscr{C}}\cap[n]} d_k - \sum_{k\in\tilde{\mathscr{C}}\cap[n]} \tilde{d}_k + 2R(\tilde{\mathscr{C}}). \tag{10.13}$$

Now, all the three terms appearing in the right hand side of (10.13) can be estimated using Theorem 21, where we recall from Section 9 that $\tilde{\boldsymbol{d}}$ satisfies Assumption 1. The proof is now complete. ☐

For an element $\mathbf{z} = (x_i, y_i)_{i\geq 1} \in \mathbb{U}_\downarrow^0$ and a constant $c > 0$, denote $c\mathbf{z} = (cx_i, y_i)_{i\geq 1}$. Thus, Lemma 26 states that, for each fixed λ, $\mathbf{Z}_n^o(\lambda)$ is close to $\kappa\mathbf{Z}_n(\lambda)$. The following lemma states that formally:

Corollary 27. *For each fixed λ, as $n \to \infty$, $\mathrm{d}_{\mathbb{U}}(\mathbf{Z}_n^o(\lambda), \kappa\mathbf{Z}_n(\lambda)) \xrightarrow{\mathbb{P}} 0$.*

Proof. Let $\pi_k, T_k : \mathbb{U}_\downarrow^0 \mapsto \mathbb{U}_\downarrow^0$ be the functions such that for $\mathbf{z} = ((x_i, y_i))_{i\geq 1}$, $\pi_k(\mathbf{z})$ consists of only (x_i, y_i) for $i \leq k$ and zeroes in other coordinates, and $T_k(\mathbf{z})$ consists only of (x_i, y_i) for $i > k$. Thus,

$$\mathrm{d}_{\mathbb{U}}(\mathbf{Z}_n^o(\lambda), \kappa\mathbf{Z}_n(\lambda)) \leq \mathrm{d}_{\mathbb{U}}(\pi_K(\mathbf{Z}_n^o(\lambda)), \pi_K(\kappa\mathbf{Z}_n(\lambda))) + \|T_K(\mathbf{Z}_n^o(\lambda))\|_{\mathbb{U}} + \|T_K(\kappa\mathbf{Z}_n(\lambda))\|_{\mathbb{U}}. \tag{10.14}$$

Now, for each fixed $K \geq 1$ the first term in the right hand side of (10.14) converges in probability to zero, by (10.11). Also, using the tightness of both $(\mathbf{Z}_n(\lambda))_{n\geq 1}$ and $(\mathbf{Z}_n^o(\lambda))_{n\geq 1}$ with respect to the $\mathbb{U}_\downarrow^0$ topology, it follows that for any $\varepsilon > 0$,

$$\lim_{K\to\infty}\lim_{n\to\infty} \mathbb{P}\left(\|T_K(\mathbf{Z}_n(\lambda))\|_{\mathbb{U}} > \varepsilon\right) = \lim_{K\to\infty}\lim_{n\to\infty} \mathbb{P}\left(\|T_K(\mathbf{Z}_n^o(\lambda))\|_{\mathbb{U}} > \varepsilon\right) = 0, \tag{10.15}$$

and the proof is now complete. ☐

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 and 4 evolve. Additionally, we color all the components blue that gets 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, and 4. Thus, in the modified setup, 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 $B_{\mathrm{SP}}(\lambda)$ the number of bad-edges that are created as surplus edge of some component.

Lemma 28. *For any $\lambda \geq \lambda_\star$, $B_{\mathrm{SP}}(\lambda) \xrightarrow{\mathbb{P}} 0$ and for all $M \geq 1$, $b_n^{-1}\mathcal{B}_M(\lambda) \xrightarrow{\mathbb{P}} 0$.*

Proof. Before going into the proof, recall Algorithm 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 $[t_n(\lambda), t_n(\lambda + \mathrm{d}\lambda)]$ at rate $(2(\nu_n - 1)\bar{s}_1 c_n)^2/(2\bar{s}_1 - 3)$. Denote $\mathcal{I}_M = \mathcal{I}_M(\lambda) = \{i : \bar{\mathscr{C}}_{(i)}(\lambda) \subset \bar{\mathcal{C}}_M(\lambda)\}$. 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 $(Y_t)_{t\geq 0}$, we write $\mathrm{D}(Y)(t)$ and $\mathrm{QV}(Y)(t)$, respectively to denote the compensator and the quadratic variation, i.e.,

$$Y_t - \mathrm{D}(Y)(t), \quad \text{and} \quad (Y_t - \mathrm{D}(Y)(t))^2 - \mathrm{QV}(Y)(t) \tag{10.16}$$

are both martingales. Now, $\mathrm{D}(B_{\mathrm{SP}}(\lambda)) \geq 0$, $\mathrm{D}(b_n^{-1}\mathcal{B}_1(\lambda)) \geq 0$, and for some constants $C_1, C_2 > 0$

$$\begin{aligned} \mathrm{D}(B_{\mathrm{SP}})(\lambda) &= \int_{\lambda_\star}^{\lambda} \frac{2\bar{s}_1 - 3}{4(\nu_n - 1)^2 \bar{s}_1^2 c_n^2} \sum_{i\geq 1} \binom{\bar{\mathcal{O}}_i(\lambda')}{2} \mathrm{d}\lambda' \leq \frac{C_1 n}{b_n^2} \int_{\lambda_\star}^{\lambda} \|\bar{\mathbf{O}}_n(\lambda')\|_2^2 \mathrm{d}\lambda' + o_{\mathbb{P}}(1) \\ &\leq \frac{C_1 n}{b_n^2}(\lambda^\star - \lambda_\star)\|\bar{\mathbf{O}}_n(\lambda^\star)\|_2^2 + o_{\mathbb{P}}(1), \end{aligned} \tag{10.17a}$$

$$\begin{aligned} \mathrm{D}(b_n^{-1}\mathcal{B}_1)(\lambda) &\leq b_n^{-1} \int_{\lambda_\star}^{\lambda} \frac{2\bar{s}_1 - 3}{4(\nu_n - 1)^2 \bar{s}_1^2 c_n^2} \sum_{i\in\mathcal{I}_M(\lambda)} \bar{\mathcal{O}}_i(\lambda') \sum_{i\geq 1} \bar{\mathcal{O}}_i(\lambda') |\bar{\mathscr{C}}_{(i)}(\lambda')| \mathrm{d}\lambda' \\ &\leq \frac{C_2 n}{b_n^2} \int_{\lambda_\star}^{\lambda} \Big(b_n^{-1} \sum_{i=1}^{M} \bar{\mathcal{O}}_{(i)}(\lambda')\Big) \|\bar{\mathbf{O}}_n(\lambda')\|_2 \|\bar{\mathbf{C}}_n(\lambda')\|_2 \mathrm{d}\lambda' + o_{\mathbb{P}}(1) \\ &\leq \frac{C_2 n}{b_n^2}(\lambda^\star - \lambda_\star)\Big(b_n^{-1} \sum_{i=1}^{M} \bar{\mathcal{O}}_{(i)}(\lambda^\star)\Big) \|\bar{\mathbf{O}}_n(\lambda^\star)\|_2 \|\bar{\mathbf{C}}_n(\lambda^\star)\|_2 + o_{\mathbb{P}}(1), \end{aligned} \tag{10.17b}$$

where $\bar{\mathcal{O}}_{(i)}$ denotes the i^{th} largest value of $(\bar{\mathcal{O}}_i)_{i\geq 1}$. Further,

$$\mathrm{QV}(B_{\mathrm{SP}})(\lambda) \leq \frac{C_1 n}{b_n^2}(\lambda^\star - \lambda_\star)\|\bar{\mathbf{O}}_n(\lambda^\star)\|_2^2 + o_{\mathbb{P}}(1), \tag{10.18a}$$

and

$$\begin{aligned} \mathrm{QV}(b_n^{-1}\mathcal{B}_1)(\lambda) &\leq b_n^{-2} \int_{\lambda_\star}^{\lambda} \frac{2\bar{s}_1 - 3}{4(\nu_n - 1)^2 \bar{s}_1^2 c_n^2} \sum_{i\in\mathcal{I}_M(\lambda)} \bar{\mathcal{O}}_i(\lambda') \sum_{i\geq 1} \bar{\mathcal{O}}_i(\lambda') |\bar{\mathscr{C}}_{(i)}(\lambda')|^2 \mathrm{d}\lambda' \\ &\leq \frac{C_2 n}{b_n^2} \int_{\lambda_\star}^{\lambda} \Big(b_n^{-1} \sum_{i=1}^{M} \bar{\mathcal{O}}_{(i)}(\lambda')\Big) \big(b_n^{-1}|\bar{\mathscr{C}}_{(1)}(\lambda')|\big) \|\bar{\mathbf{O}}_n(\lambda')\|_2 \|\bar{\mathbf{C}}_n(\lambda')\|_2 \mathrm{d}\lambda' + o_{\mathbb{P}}(1) \\ &\leq \frac{C_2 n}{b_n^2}(\lambda^\star - \lambda_\star)\Big(b_n^{-1} \sum_{i=1}^{M} \bar{\mathcal{O}}_{(i)}(\lambda^\star)\Big) \big(b_n^{-1}|\bar{\mathscr{C}}_{(1)}(\lambda^\star)|\big) \|\bar{\mathbf{O}}_n(\lambda^\star)\|_2 \|\bar{\mathbf{C}}_n(\lambda^\star)\|_2 + o_{\mathbb{P}}(1). \end{aligned} \tag{10.18b}$$

Recall that using Lemma 26, an application of Theorem 25 yields that $(\bar{\mathbf{Z}}_n(\lambda))_{n\geq 1}$ is tight in $\mathbb{U}^0_\downarrow$. The proof now follows using the fact that $n/b_n^2 \to 0$. □

Suppose that a bad edge is being created at time λ'. Now, this bad edge may be created by choosing the open half-edges from $\mathscr{C}_{(i)}(\lambda')$ and $\mathscr{C}_{(j)}(\lambda')$ for $1 \leq i,j \leq M$. For fixed M, let $F_M(\lambda)$ denote the number of such bad-edges created upto time λ. Using an argument identical to Lemma 28 one can show the following:

Lemma 29. *For any $\lambda \geq \lambda_\star$ and $M \geq 1$, $F_M(\lambda) \xrightarrow{\mathbb{P}} 0$.*

The following is the last ingredient that will be needed in the proof:

Lemma 30. *Fix any $\lambda \in [\lambda_\star, \lambda^\star]$. For any $\varepsilon > 0$, and $K \geq 1$, there exists $M = M(\varepsilon, K)$ such that*

$$\limsup_{n\to\infty} \mathbb{P}\left(\bar{\mathscr{C}}_{(1)}(\lambda), \dots, \bar{\mathscr{C}}_{(K)}(\lambda) \text{ are not contained in } \bar{\mathcal{C}}_M(\lambda)\right) \leq \varepsilon. \tag{10.19}$$

Proof. Let $\mathcal{I}_M := \{i : \bar{\mathscr{C}}_{(i)}(\lambda) \subset \bar{\mathcal{C}}_M(\lambda)\}$. It is enough to show that, for any $\varepsilon > 0$, there exists M such that

$$\limsup_{n\to\infty} \mathbb{P}\bigg(\sum_{i\notin\mathcal{I}_M} |\bar{\mathscr{C}}_{(i)}(\lambda)|^2 > \varepsilon b_n^2\bigg) \leq \varepsilon. \tag{10.20}$$

For any $M \geq 1$, consider the merging dynamics of Algorithm 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,

$$\sum_{i\notin\mathcal{I}_M} |\bar{\mathscr{C}}_{(i)}(\lambda)|^2 \leq \sum_{i\geq 1} |\bar{\mathscr{C}}^{>M}_{(i)}(\lambda)|^2. \tag{10.21}$$

Now, using Lemma 26, an application of Theorem 25 yields that $(\bar{\mathbf{Z}}_n(\lambda))_{n\geq 1}$ is tight in $\mathbb{U}^0_\downarrow$. Thus the proof now follows. □

10.4 Proof of Theorem 5

We now have all the ingredients to complete the proof of Theorem 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 26, Theorem 25 implies

$$(\bar{\mathbf{Z}}_n(\lambda_1), \bar{\mathbf{Z}}_n(\lambda_2)) \xrightarrow{d} (\bar{\mathbf{Z}}(\lambda_1), \bar{\mathbf{Z}}(\lambda_1, \lambda_2)), \tag{10.22}$$

for some random elements $\mathbf{Z}(\lambda_1), \mathbf{Z}(\lambda_1, \lambda_2)$ of $\mathbb{U}^0_\downarrow$. Now, $\bar{\mathbf{Z}}_n(\lambda_1) = \mathbf{Z}_n(\lambda_1)$. Moreover, using Lemmas 28, 29, and 30 and the facts that both $(\bar{\mathbf{Z}}_n(\lambda_2))_{n\geq 1}$ and $(\mathbf{Z}_n(\lambda_2))_{n\geq 1}$ converge, it follows that (see the argument in Corollary 27)

$$\mathrm{d}_{\mathbb{U}}(\bar{\mathbf{Z}}_n(\lambda_2), \mathbf{Z}_n(\lambda_2)) \xrightarrow{\mathbb{P}} 0. \tag{10.23}$$

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 (10.22) does not depend on λ_1 by Theorem 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 $k \geq 1$

$$(\mathbf{Z}_n(\lambda_1), \dots, \mathbf{Z}_n(\lambda_k)) \xrightarrow{d} (\mathrm{AMC}(\lambda_1), \dots, \mathrm{AMC}(\lambda_k)). \tag{10.24}$$

Finally, the proof of Theorem 2 is completed by using Proposition 24. □

A Path counting

In this section, we derive a generalization of [27, 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 31. *Let $\boldsymbol{w} = (w_i)_{i\in[n]}$ be a weight sequence and consider $\mathrm{CM}_n(\boldsymbol{d})$ such that $\nu_n < 1$. Then,*

$$\mathbb{E}\Big[\sum_{i\in\mathscr{C}(V_n')} w_i\Big] \leq \mathbb{E}\big[w_{V_n'}\big] + \frac{\mathbb{E}\left[D_n'\right]\mathbb{E}\big[D_n w_{V_n}\big]}{\mathbb{E}\left[D_n\right](1-\nu_n)}. \tag{A.1}$$

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,

$$\mathbb{E}_v\Big[\sum_{i\in\mathscr{C}(V_n')} w_i\Big] \leq w_v + d_v \sum_{l\geq 1} \sum_{\substack{x_1,\dots,x_l\\ x_i\neq x_j,\forall i\neq j}} \frac{\prod_{i=1}^{l-1} d_{x_i}(d_{x_i}-1) d_{x_l} w_{x_l}}{(\ell_n-1)\dots(\ell_n-2l+1)}. \tag{A.2}$$

Now, using the exactly same arguments as [27, Lemma 5.1], it follows that

$$\mathbb{E}\Big[\sum_{i\in\mathscr{C}(V_n')} w_i\Big] \leq \mathbb{E}\big[w_{V_n'}\big] + \frac{\mathbb{E}\left[D_n'\right]\mathbb{E}\left[D_n w_{V_n}\right]}{\mathbb{E}\left[D_n\right]} \sum_{l\geq 1} \nu_n^{l-1}, \tag{A.3}$$

and this completes the proof. ∎

B Appendix: Proof of Lemma 20

The proof is an adaptation of the proof of [20, 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, δ, K. Consider the graph exploration described in Algorithm 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 (4.1) with $S_n(0) = D_n'$. Note that $\mathscr{C}(V_n')$ is explored when $\mathbf{S}_n$ hits zero. For $H > 0$, let

$$\gamma := \inf\{l \geq 1 : S_n(l) \geq H \text{ or } S_n(l) = 0\} \wedge 2\delta_K b_n. \tag{B.1}$$

Note that

$$\begin{aligned}
\mathbb{E}\left[S_n(l+1) - S_n(l) \middle| (\mathcal{I}_i^n(l))_{i=1}^n\right] &= \sum_{i\in[n]} d_i \mathbb{P}\left(i\notin\mathscr{V}_l, i\in\mathscr{V}_{l+1} \middle| (\mathcal{I}_i^n(l))_{i=1}^n\right) - 2\\
&= \frac{\sum_{i\notin\mathscr{V}_l} d_i^2}{\ell_n-2l-1} - 2 \leq \frac{\sum_{i\in[n]} d_i^2}{\ell_n-2l-1} - 2\\
&:= \lambda c_n^{-1} + o(c_n^{-1}) + \frac{2l+1}{\ell_n-2l-1}\times\frac{\sum_{i\in[n]} d_i^2}{\ell_n} \leq 0
\end{aligned} \tag{B.2}$$

uniformly over $l \leq 2\delta_K b_n$ 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

$$\mathbb{E}\left[D_n'\right] \geq \mathbb{E}\left[S_n(\gamma)\right] \geq H\mathbb{P}\left(S_n(\gamma)\geq H\right). \tag{B.3}$$

Thus,

$$\mathbb{P}\left(S_n(\gamma) \geq H\right) \leq \frac{\mathbb{E}\left[D'_n\right]}{H}. \tag{B.4}$$

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

$$\begin{aligned}
&\mathbb{P}\left(\mathrm{SP}(\mathscr{C}(V'_n)) \geq K, |\mathscr{C}(V'_n)| \in (\delta_K b_n, 2\delta_K b_n)\right) \\
&\leq \mathbb{P}\left(S_n(\gamma) \geq H\right) + \mathbb{P}\left(\mathrm{SP}(\mathscr{C}(V'_n)) \geq K, S_n[0, 2\delta_K b_n] < H, S_n[0, \delta_K b_n] > 0\right).
\end{aligned} \tag{B.5}$$

Now,

$$\begin{aligned}
&\mathbb{P}\left(\mathrm{SP}(\mathscr{C}(V'_n)) \geq K, S_n[0, 2\delta_K b_n] < H, S_n[0, \delta_K b_n] > 0\right) \\
&\leq \sum_{1 \leq l_1 < \cdots < l_K \leq 2\delta_K b_n} \mathbb{P}\left(\text{surpluses occur at times } l_1, \dots, l_K, S_n[0, 2\delta_K b_n] < H, S_n[0, \delta_K b_n] > 0\right) \\
&= \sum_{1 \leq l_1 < \cdots < l_K \leq 2\delta_K b_n} \mathbb{E}\left[\mathbb{1}_{\{0 < S_n[0, l_K - 1] < H, \mathrm{SP}(l_K - 1) = K-1\}} Y\right],
\end{aligned} \tag{B.6}$$

where

$$\begin{aligned}
Y &= \mathbb{P}\left(K^{th} \text{ surplus occurs at time } l_K, S_n[l_K, 2\delta_K b_n] < H, S_n[l_K, \gamma] > 0 \mid \mathscr{F}_{l_K - 1}\right) \\
&\leq \frac{CK^{1.1} a_n}{\ell_n \sqrt{\delta}} \leq \frac{CK^{1.1}}{b_n \sqrt{\delta}}.
\end{aligned} \tag{B.7}$$

Therefore, using induction, (B.5) yields

$$\begin{aligned}
&\mathbb{P}\left(\mathrm{SP}(\mathscr{C}(V'_n)) \geq K, S_n[0, 2\delta_K b_n] < H, S_n[0, \delta_K b_n] > 0\right) \\
&\leq C\left(\frac{K^{1.1}}{\sqrt{\delta} b_n}\right)^K \frac{(2\delta b_n)^{K-1}}{K^{0.12(K-1)}(K-1)!} \sum_{l_1=1}^{2\delta_K b_n} \mathbb{P}\left(|\mathscr{C}(V'_n)| \geq l_1\right) \leq C \frac{\delta^{K/2}}{K^{1.1} b_n} \mathbb{E}\left[|\mathscr{C}(V'_n)|\right],
\end{aligned} \tag{B.8}$$

where we have used the fact that $\#\{1 \leq l_2, \dots, l_K \leq 2\delta b_n\} = (2\delta b_n)^{K-1}/(K-1)!$ and Stirling's approximation for $(K-1)!$ in the last step. Since $\lambda < 0$, we can use Lemma 31 to conclude that for all sufficiently large n

$$\mathbb{E}\left[|\mathscr{C}(V_n)|\right] \leq C c_n, \tag{B.9}$$

for some constant $C > 0$ and we get the desired bound for (B.5). The proof of Lemma 20 is now complete. □

Acknowledgement

We sincerely thank Shankar Bhamidi for several helpful discussions. This research have been supported by the Netherlands Organisation for Scientific Research (NWO) through Gravitation Networks grant 024.002.003. In addition, RvdH has been supported by VICI grant 639.033.806, JvL has been supported by the European Research Council (ERC), and SS has been supported by EPSRC grant EP/J019496/1 and a CRM-ISM fellowship.

References

[1] Addario-Berry, L., Broutin, N., Goldschmidt, C., and Miermont, G. (2013). The scaling limit of the minimum spanning tree of the complete graph. *arXiv:1301.1664*.

[2] Aidekon, E., van der Hofstad, R., Kliem, S., and van Leeuwaarden, J. S. H. (2014). Large deviations for power-law thinned Levy processes. *arXiv:1404.1692*.

[3] Albert, R. and Barabási, A. L. (2002). Statistical mechanics of complex networks. *Reviews of Modern Physics*, 74(1):47–97.

[4] Aldous, D. (1997). Brownian excursions, critical random graphs and the multiplicative coalescent. *Annals of Probability*, 25(2):812–854.

[5] Aldous, D. and Limic, V. (1998). The entrance boundary of the multiplicative coalescent. *Electronic Journal of Probability*, 3(3):1–59.

[6] Aldous, D. and Pittel, B. (2000). On a random graph with immigrating vertices: emergence of the giant component. *Random Structures & Algorithms*, 17(2):79–102.

[7] Barabási, A. L. and Albert, R. (1999). Emergence of scaling in random networks. *Science*, 286:509–512.

[8] Bertoin, J. (1996). *Lévy Processes*. Cambridge University Press.

[9] 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*.

[10] Bhamidi, S., Budhiraja, A., and Wang, X. (2014b). The augmented multiplicative coalescent, bounded size rules and critical dynamics of random graphs. *Probability Theory and Related Fields*, 160(3):733–796.

[11] Bhamidi, S., Dhara, S., van der Hofstad, R., and Sen, S. (2016). Metric space limits for the heavy-tailed critical configuration model. *In preparation*.

[12] Bhamidi, S., van der Hofstad, R., and Sen, S. (2015). The multiplicative coalescent, inhomogeneous continuum random trees, and new universality classes for critical random graphs. *arXiv:1508.04645*.

[13] Bhamidi, S., van der Hofstad, R., and van Leeuwaarden, J. S. H. (2010). Scaling limits for critical inhomogeneous random graphs with finite third moments. *Electronic Journal of Probability*, 15(6):1682–1702.

[14] Bhamidi, S., van der Hofstad, R., and van Leeuwaarden, J. S. H. (2012). Novel scaling limits for critical inhomogeneous random graphs. *Annals of Probability*, 40(6):2299–2361.

[15] Bingham, N. H., Goldie, C. M., and Teugels, J. L. (1989). *Regular Variation*. Cambridge University Press.

[16] Bollobás, B. (1980). A probabilistic proof of an asymptotic formula for the number of labelled regular graphs. *European Journal of Combinatorics*, 1(4):311–316.

[17] Bollobás, B. (1984). The evolution of random graphs. *Transactions of the American Mathematical Society*, 286(1):257–274.

[18] Breiman, L. (1968). *Probability*. Society for Industrial and Applied Mathematics.

[19] Broutin, N. and Marckert, J.-F. (2016). A new encoding of coalescent processes: applications to the additive and multiplicative cases. *Probability Theory and Related Fields*, 166(1):515–552.

[20] Dhara, S., van der Hofstad, R., van Leeuwaarden, J. S. H., and Sen, S. (2016). Critical window for the configuration model: finite third moment degrees. *arXiv:1605.02868*.

[21] Faloutsos, M., Faloutsos, P., and Faloutsos, C. (1999). On power-law relationships of the Internet topology. *Computer Communication Review*, 29(4):251–262.

[22] Fountoulakis, N. (2007). Percolation on sparse random graphs with given degree sequence. *Internet Mathematics*, 4(1):329–356.

[23] G.H. Hardy (1949). *Divergent Series*. Clarendon (Oxford University) Press, Oxford.

[24] Janson, S. (2008). The largest component in a subcritical random graph with a power law degree distribution. *Annals of Applied Probability*, 18(4):1651–1668.

[25] Janson, S. (2009a). On percolation in random graphs with given vertex degrees. *Electronic Journal of Probability*, 14:87–118.

[26] Janson, S. (2009b). The probability that a random multigraph is simple. *Combinatorics, Probability and Computing*, 18(1-2):205–225.

[27] Janson, S. (2010). Susceptibility of random graphs with given vertex degrees. *Journal of Combinatorics*, 1(3-4):357–387.

[28] Janson, S., Knuth, D. E., Łuczak, T., and Pittel, B. (1993). The birth of the giant component. *Random Structures & Algorithms*, 4(3):233–358.

[29] Janson, S., Łuczak, T., and Rucinski, A. (2000). *Random Graphs.* Wiley, New York.

[30] Joseph, A. (2014). The component sizes of a critical random graph with given degree sequence. *Annals of Applied Probability*, 24(6):2560–2594.

[31] Kallenberg, O. (1997). *Foundations of Modern Probability*. Springer, New York.

[32] Kleinberg, J. M., Kumar, R., Raghavan, P., Rajagopalan, S., and Tomkins, A. S. (1999). The Web as a graph : measurements, models, and methods. *COCOON'99 Proceedings*, 1627(2):1–17.

[33] Lipster, R. S. and Shiryayev, A. N. (1989). *Theory of Martingales*. Springer, Dordrecht.

[34] Łuczak, T. (1990). Component behavior near the critical point of the random graph process. *Random Structures & Algorithms*, 1(3):287–310.

[35] Molloy, M. and Reed, B. (1995). A critical-point for random graphs with a given degree sequence. *Random Structures & Algorithms*, 6(2-3):161–179.

[36] Nachmias, A. and Peres, Y. (2010). Critical percolation on random regular graphs. *Random Structures & Algorithms*, 36(2):111–148.

[37] Riordan, O. (2012). The phase transition in the configuration model. *Combinatorics, Probability and Computing*, 21:265–299.

[38] Rogers, L. C. G. and Williams, D. (1994). *Diffusions, Markov Processes, and Martingales.*, volume 1. John Wiley & Sons, Ltd., Chichester, 2nd edition.

[39] Siganos, G., Faloutsos, M., Faloutsos, P., and Faloutsos, C. (2003). Power laws and the AS-level internet topology. *IEEE/ACM Transactions on Networking*, 11(4):514–524.

[40] Stegehuis, C., van der Hofstad, R., and van Leeuwaarden, J. S. H. (2016). Epidemic spreading on complex networks with community structures. *Scientific Reports*, 6.

[41] van der Hofstad, R. (2016). *Random Graphs and Complex Networks*, volume I. *To appear with* Cambridge University Press.

[42] van der Hofstad, R., Janson, S., and Luczak, M. (2016). Component structure of the configuration model: barely supercritical case. *arXiv:1611.05728*.

[43] van der Hofstad, R., Janssen, A. J. E. M., and van Leeuwaarden, J. S. H. (2010). Critical epidemics, random graphs, and Brownian motion with a parabolic drift. *Advances in Applied Probability*, 42(4):1187–1206.

[44] van der Hofstad, R., Kliem, S., and van Leeuwaarden, J. S. H. (2014). Cluster tails for critical power-law inhomogeneous random graphs. *arXiv:1404.1727*.

[45] Whitt, W. (2002). *Stochastic-Process Limits: An Introduction to Stochastic-Process Limits and Their Application to Queues*. Springer-Verlag, New York.