Annales de l'Institut Henri Poincaré - Probabilités et Statistiques
2020, Vol. 56, No. 3, 1515–1558
https://doi.org/10.1214/19-AIHP980

ANNALES DE L'INSTITUT HENRI POINCARÉ
PROBABILITÉS ET STATISTIQUES
www.imstat.org/aihp

Heavy-tailed configuration models at criticality

Souvik Dhara[a,b], Remco van der Hofstad[c], Johan S. H. van Leeuwaarden[c,d] and Sanchayan Sen[e]

[a]*Department of Mathematics, Massachusetts Institute of Technology, Cambridge, USA. E-mail: sdhara@mit.edu*
[b]*Microsoft Research – New England, Cambridge, USA*
[c]*Department of Mathematics and Computer Science, Eindhoven University of Technology, Eindhoven, The Netherlands. E-mail: r.w.v.d.hofstad@tue.nl*
[d]*Department of Econometrics and Operations Research, Tilburg University, Tilburg, The Netherlands. E-mail: j.s.h.vanleeuwaarden@uvt.nl*
[e]*Department of Mathematics, Indian Institute of Science, Bengaluru, India. E-mail: sanchayan.sen1@gmail.com*

Received 6 September 2017; accepted 13 March 2019

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.* **24** (2014) 2560–2594) about the scaling limits of uniform simple graphs with i.i.d. degrees in the critical window, and sheds light on the relation between the scaling limits obtained by Joseph and in this 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* **160** (2014) 733–796). 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.

Résumé. Nous étudions le comportement critique des tailles des composantes du modèle de configuration lorsque la queue de la loi du degré d'un sommet choisi uniformément au hasard est une fonction à variation régulière d'exposant $\tau-1$, où $\tau \in (3,4)$. Nous montrons que les tailles des composantes sont d'ordre $n^{(\tau-2)/(\tau-1)}L(n)^{-1}$ où $L(\cdot)$ est une fonction à variation lente. Nous montrons également que les tailles des composantes, une fois ordonnées et remises à l'échelle, convergent en loi vers les longueurs ordonnées des excursions d'un processus de Lévy raréfié. Ceci montre que les limites d'échelle des tailles des composantes pour ces modèles de configuration à queue lourde sont dans une classe d'universalité différente des graphes aléatoires d'Erdős–Rényi. De plus nous montrons que le vecteur de ces tailles de composantes remises à l'échelle et de leurs excès respectifs convergent en loi pour une topologie forte. Notre approche résout une conjecture de Joseph (*Ann. Appl. Probab.* **24** (2014) 2560–2594) sur les limites d'échelle de graphes simples uniformes à degrés i.i.d. dans la fenêtre critique, et met en lumière les relations entre les limites d'échelle obtenues par Joseph et celles considérées dans cet article, qui se révèlent très différentes. Par ailleurs, nous utilisons un modèle de percolation pour étudier l'évolution des tailles des composantes et des arêtes en excès à l'intérieur de la fenêtre critique, dont nous montrons qu'elle converge au sens des marginales de dimension finie vers le coalescent multiplicatif augmenté introduit par Bhamidi et al. (*Probab. Theory Related Fields* **160** (2014) 733–796). Les résultats principaux de cet article sont montrés sous des hypothèses assez générales sur les degrés des sommets, et nous discutons des situations déjà considérées où ces hypothèses sont vérifiées.

MSC: 60C05; 05C80

Keywords: Critical configuration model; Heavy-tailed degree; Thinned Lévy process; Augmented multiplicative coalescent; Universality; Critical percolation

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 [18], Łuczak [37], Janson et al. [31] and Aldous [4] for Erdős–Rényi random graphs. A large body of subsequent work in [9,10,14,22,33,40,41,47] 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 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 [22]. However, empirical studies of real-world networks from various fields including physics, biology, computer science [3,7,24,35,43,44] show that the degree distribution is often heavy-tailed and of power-law type. A first work towards understanding the critical behavior in the *heavy-tailed* network models is [15], 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 [17] 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 [39] 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 $\mathbb{P}(D_n \geq k) \sim L_0(k)/k^{\tau-1}$ 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 in 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 [15]. 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 [33, 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 $[n] := \{1, 2, \dots, n\}$ 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. Thus, an edge is a pair of half-edges. Initially we have $\ell_n = \sum_{i \in [n]} d_i$ unpaired 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 $\boldsymbol{d}$ [45, Proposition 7.13]. It was further shown in [29] 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. The topology needed for the convergence in distribution will always be specified unless it is clear from the context. We often use the Bachmann–Landau notation $O(\cdot)$, $o(\cdot)$ for large n asymptotics of real numbers. 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 $\mathbb{P}_n(\mathcal{E}_n) \to 1$. 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 $f_n/g_n \xrightarrow{\mathbb{P}} 0$ w.h.p.; $f_n = \Theta_{\mathbb{P}}(g_n)$ 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

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

with the p-norm metric $d(\mathbf{x}, \mathbf{y}) = (\sum_{i=1}^{\infty} |x_i - y_i|^p)^{1/p}$. 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

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

with the metric

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

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

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

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 an interval I to the metric space $E = (E, \mathrm{d})$ equipped with the Skorohod J_1-topology. Consider a decreasing sequence $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots) \in \ell^3_{\downarrow} \setminus \ell^2_{\downarrow}$. Denote $\mathcal{I}_i(s) := \mathbb{1}_{\{\xi_i \leq s\}}$ where $\xi_i \sim \mathrm{Exp}(\theta_i/\mu)$ independently, and $\mathrm{Exp}(r)$ denotes the exponential distribution with rate r. Consider the process

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

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

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

The process of the form (1.5) was termed *thinned* Lévy processes in [15] (see also [2,48]), 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

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

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 $\mathrm{refl}(\bar{S}_\infty^\lambda(t))$ at time t conditionally on $(\bar{S}_\infty^\lambda(u))_{u\leq t}$. Formally, $\mathbf{N}$ is characterized as the counting process for which

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

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 $\mathbb{D}(\mathbb{R},\mathbb{U}_\downarrow^0)$, called the augmented multiplicative coalescent (AMC) process. Think of a collection of particles in a system with $\mathbf{X}(s)$ describing their masses and $\mathbf{Y}(s)$ describing an additional attribute at time s. Let $K_1, K_2 > 0$ be constants. The evolution of the system takes place according to the following rule at time s:

- ▷ For $i \neq j$, at rate $K_1 X_i(s) X_j(s)$, the ith and jth components merge and create a new component of mass $X_i(s)+X_j(s)$ and attribute $Y_i(s)+Y_j(s)$.
- ▷ For any $i\geq 1$, at rate $K_2 X_i^2(s)$, $Y_i(s)$ increases to $Y_i(s)+1$.

Of course, at each event time, the indices are re-organized to give a proper element of $\mathbb{U}_\downarrow^0$. 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

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

where $\xi_i \sim \mathrm{Exp}(\theta_i)$ independently over i and $\boldsymbol{\theta}\in \ell_\downarrow^3\setminus\ell_\downarrow^2$. Therefore, by [5, Lemma 1], the process $\mathrm{refl}(\bar{\mathbf{S}}_\infty^\lambda)$ has no infinite excursions almost surely and only finitely many excursions of 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

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

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

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

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

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

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

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

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

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,*

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

with respect to the $\ell^2_{\downarrow}$*-topology where* $\gamma_i(\lambda)$ *is the length of the* i*th largest excursion of the process* $\bar{\mathbf{S}}^{\lambda}_{\infty}$, *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* $\mathbf{Z}_n := \mathrm{ord}(b_n^{-1}|\mathscr{C}_{(i)}|, \mathrm{SP}(\mathscr{C}_{(i)}))_{i\geq 1}$, $\mathbf{Z} := \mathrm{ord}(\gamma_i(\lambda), N(\gamma_i(\lambda)))_{i\geq 1}$. *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 [33] where the degrees were assumed to be an i.i.d. sample from an exact power-law distribution and the results were obtained for the component sizes of $\mathrm{CM}_n(\boldsymbol{d})$ (as in Theorem 1). We will see that Assumption 1 is satisfied for i.i.d. degrees in Section 2.2. Thus, a quenched version (conditional on the degrees) of [33, Theorem 8.3] follows from our results. Further, if the degrees are chosen approximately as the weights chosen in [15], then our results continue to hold. This sheds light on the relation between the scaling limits in [15] and [33] (see Remark 11). Moreover, Theorem 3 resolves [33, 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 $\mathbf{Z} = \mathrm{ord}(\kappa\gamma_i(\lambda), N(\gamma_i(\lambda)))_{i\geq 1}$, 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 ith 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 super-critical, i.e.,

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

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

$$p_n = p_n(\lambda) := \frac{1}{\nu_n}\big(1 + \lambda c_n^{-1} + o\big(c_n^{-1}\big)\big) \quad \text{for some } \lambda \in \mathbb{R}. \tag{1.17}$$

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* $\theta_i/\sqrt{\nu}$, *let* $\mathscr{C}_{(i)}^p$ *denote the ith 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[0, 1] random variable $U_{(ij)}$. 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. This is known as the *Harris coupling*. 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* $n \to \infty$,

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

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 [22, Theorem 3.7], 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 [10,20], since the scaling limits in those settings were described in terms of the excursions of a Brownian motion with negative 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 minor, 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 with $\mathbb{P}(D_n \geq k) \sim L_0(k)k^{-(\tau-1)}$ for some $\tau \in (3, 4)$. Here, 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 [15] 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

$$G(x) = 1 - F(x) = \frac{C_F L_0(x)}{x^{\tau-1}}, \tag{2.1}$$

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 $f^{-1}(x) := \inf\{y : f(y) \le x\}$. Therefore, using [16, Theorem 1.5.12],

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

where $L(\cdot)$ is another slowly-varying function. Note that [16, 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

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

Define the degree sequence $\boldsymbol{d}_\lambda$ by taking the degree of the ith 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 $K \ge 1$. 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

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

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

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

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

Therefore,

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

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

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

From our choice of δ, $-3\alpha + 3\delta < -1$ and therefore $\sum_{i\ge 1} i^{-3\alpha+3\delta} < \infty$. By [16, Lemma 1.3.2], $\sup_{1\le x\le C} L(x)^3 < \infty$. 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*

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

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 [15, 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

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

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

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

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,

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

Again, using (2.4),

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

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,

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

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 [26, page 333] it can be seen that λ_0 is finite which completes the proof. □

Remark 9. Note that if we add approximately cnc_n^{-1} ($c > 0$ is a constant) ones in the degree sequence given in (2.4), then we end up with another configuration model for which $\lim_{n\to\infty} c_n(\nu_n - \nu) = \zeta'$ with $\zeta > \zeta'$. Similarly, deleting cnc_n^{-1} 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 [33]. Let $\Delta_1, \dots, \Delta_n$ 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 $d_i = \Delta_{(i)}$, $\Delta_{(i)}$ being the ith order statistic of $(\Delta_1, \dots, \Delta_n)$. We show that $\boldsymbol{d}$ satisfies Assumption 1 almost surely under a suitable coupling. We use a coupling from [19, 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

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

It can be checked that $(d_1, \dots, d_n) \stackrel{d}{=} (\bar{d}_1, \dots, \bar{d}_n)$ and therefore we will ignore the bar in the subsequent notation. Note that, by the strong law of large numbers, $\Gamma_{n+1}/n \xrightarrow{\text{a.s.}} 1$. 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 $\theta_i = (C_F/\Gamma_i)^\alpha$. To see that $\boldsymbol{\theta} \in \ell^3_{\downarrow} \setminus \ell^2_{\downarrow}$ almost surely, we need to show that

$$\mathbb{P}\left(\sum_{i=1}^{\infty} \Gamma_i^{-3\alpha} < \infty\right) = 1, \qquad \mathbb{P}\left(\sum_{i=1}^{\infty} \Gamma_i^{-2\alpha} = \infty\right) = 1. \tag{2.20}$$

(2.20) follows easily from the observations $2\alpha < 1 < 3\alpha$ and $\Gamma_i/i \xrightarrow{\text{a.s.}} 1$ as $i \to \infty$.

Next, 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, using (2.20), and $\Gamma_{n+1}/n \xrightarrow{\text{a.s.}} 1$, 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

$$\lim_{n\to\infty} c_n(\nu_n - \nu) \xrightarrow{\text{a.s.}} \Lambda_0, \tag{2.21}$$

where

$$\Lambda_0 := -\frac{C_F^{2\alpha}}{\mathbb{E}[D]} \sum_{i=1}^{\infty} \left(\int_{\Gamma_{i-1}}^{\Gamma_i} u^{-2\alpha}\,\mathrm{d}u - \Gamma_i^{-2\alpha}\right). \tag{2.22}$$

Therefore, the results in Section 1.3 hold conditionally on the degree sequence if we assume the degrees to be an i.i.d. sample 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 [34, Corollary 3.22].

Remark 11. Let us recall the limiting object obtained in [33, Theorem 8.1] and compare this with the limiting object $\bar{\mathbf{S}}_\infty^{\Lambda_0}$, defined in (1.5) with $\theta_i = (C_F/\Gamma_i)^\alpha$ and Λ_0 given by (2.22). We will prove an analogue of [33, Theorem 8.1] in Theorem 8. Although we use a different exploration process from [33], 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 [33] 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 [22], 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 [4,14,15,22,41]. 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 $\lambda_n = o(c_n)$. Janson [27] showed that the size of the kth largest cluster for a subcritical configuration model (i.e., the case $\nu_n \to \nu$ and $\nu < 1$) is $d_k/(1-\nu)$ (see [27, 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 kth 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 $\mathcal{X}_n$ [46]. 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 of the percolation clusters is necessary to understand the minimal spanning tree of the giant component with i.i.d. continuous weights on each edge [1]. Also, since the percolated configuration model is again a configuration model [25,28], 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 [15, Theorem 1.1] for rank-one inhomogeneous random graphs. Thus, $\mathrm{CM}_n(\boldsymbol{d})$ with regularly-varying tails lies in the domain of attraction of the new universality class studied in [15]. This is again confirming the predictions made by statistical physicists that the nature of the phase transition does not depend on the precise details of the model. Our scaling limit fits into the general class of limits predicted in [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 [4,6,14,22,33], the limiting component sizes are described by the excursions of a Brownian motion with a parabolic drift. All these models have 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) has been established in [4, (10)], [6, (4)]. In the case of [15] and this paper, the barely subcritical component sizes do not become negligible due to the existence of the high-degree vertices (see [15, 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 an 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 [15] 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, [13] studied this problem for rank-1 inhomogeneous random graphs for heavy-tailed weights. In [11,12], 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 [13]. This is the first step in understanding the minimal spanning tree problem (see [1]) for $\mathrm{CM}_n(\boldsymbol{d})$.

(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 [33, 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, that we call 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 configuration model 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 (if any), except f. If degree$(v')=1$ (i.e. the only half-edge incident to v' is f) then kill v'. Otherwise, declare v' to be active and larger than all other vertices that are alive. After killing e, if v does not have another active half-edge, then kill v also.

(S3) Repeat from (S1) at stage $i+1$ 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

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

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 kth 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

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

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),

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

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* $\bar{\mathbf{S}}_\infty^\lambda := (\bar{S}_\infty^\lambda(t))_{t\geq 0}$ *from* (1.5). *Then,*

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

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 $\{\mathcal{I}_i^n(l) = 0\}$,

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

Let $\ell_n(T) = \ell_n - 2Tb_n + 1$. Then, uniformly over $l \leq Tb_n$,

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

Also, $\mathcal{I}_i^n(l+1) - \mathcal{I}_i^n(l) = 0$ on the set $\{\mathcal{I}_i^n(l) = 1\}$. Denote $M_n^K(l) = a_n^{-1} \sum_{i\in[n]} d_i(\mathcal{I}_i^n(l) - \frac{d_i}{\ell_n(T)} l)$. Then,

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

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

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

Thus, Assumption 1(ii) gives

$$\begin{aligned} \big|\mathbb{E}\big[M_n^K(l)\big]\big| &= a_n^{-1} \sum_{i=K+1}^{n} d_i \left(\frac{d_i}{\ell_n(T)} l - \mathbb{P}\big(\mathcal{I}_i^n(l) = 1\big) \right) \\ &\leq a_n^{-1} \sum_{i=K+1}^{n} d_i \left| 1 - \left(1 - \frac{d_i}{\ell_n}\right)^l - \frac{d_i}{\ell_n} l \right| + a_n^{-1} l \sum_{i\in[n]} d_i^2 \left(\frac{1}{\ell_n(T)} - \frac{1}{\ell_n} \right) \\ &\leq \frac{l^2}{2\ell_n^2 a_n} \sum_{i=K+1}^{n} d_i^3 + o(1) \\ &\leq \frac{T^2 n^{2\rho} n^{3\alpha} L(n)^3}{2\ell_n^2 L(n)^2 n^\alpha L(n)} \left(a_n^{-3} \sum_{i=K+1}^{n} d_i^3 \right) + o(1) = C \left(a_n^{-3} \sum_{i=K+1}^{n} d_i^3 \right) + o(1), \end{aligned} \tag{4.9}$$

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

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

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

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

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 $i\neq j$,

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

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

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

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

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

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

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

uniformly for $l\leq Tb_n$. Now we can use the super-martingale inequality [42, Lemma 2.54.5] stating that for any super-martingale $(M(t))_{t\geq 0}$, with $M(0)=0$,

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

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 $\varepsilon>0$,

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

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 $n\to\infty$,*

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

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

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

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

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

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

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

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

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

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 $K \to \infty$. □

For future purposes, we also describe the scaling limit of the reflected process:

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

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

Proof. This follows from Theorem 8 and the fact that the reflection is Lipschitz continuous with respect to the Skorohod J_1-topology (see [49, 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* $\varepsilon > 0$,

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

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 [22]. 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 aproach with size-biased point processes originally proposed 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,*

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

Consequently, for any $\varepsilon>0$,

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

Proof. We make use of a result due to Janson [30] regarding bounds on the susceptibility functions for the configuration model. In fact, [30, Lemma 5.2] shows that, for any configuration model $\mathrm{CM}_n(\boldsymbol{d})$ with $\nu_n<1$,

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

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 $\nu_n^{[K]}$ of $\mathcal{G}^{[K]}$ satisfies

$$\begin{aligned}\nu_n^{[K]}&=\frac{\sum_{i\in[n]}d_i'(d_i'-1)}{\sum_{i\in[n]}d_i'}\leq\frac{\sum_{i\in[n]}d_i(d_i-1)-\sum_{i=1}^K d_i(d_i-1)}{\ell_n-2\sum_{i=1}^K d_i}\\&=\nu_n-C_1n^{2\alpha-1}L(n)^2\sum_{i\leq K}\theta_i^2=\nu_n-C_1c_n^{-1}\sum_{i\leq K}\theta_i^2\end{aligned} \tag{5.5}$$

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

$$\mathbb{E}_K\big[\big|\mathscr{C}^{[K]}(V_n)\big|\big]\leq\frac{C_2}{1-\nu_n^{[K]}}\leq\frac{C_2}{1-\nu_n+C_1c_n^{-1}\sum_{i\leq K}\theta_i^2}\leq\frac{C_2c_n}{-\lambda+C_1\sum_{i\leq K}\theta_i^2}, \tag{5.6}$$

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

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

as well as $b_n^2=nc_n$ by (1.10b). □

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

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

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 $\varepsilon > 0$,*

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

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

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

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

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

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

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

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

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

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

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

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 $\bigcup_{(l,r)\in\mathcal{E}}(l,r)$ 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 ith largest excursion of f and define $\Phi_m : \mathbb{D}_+[0,T] \to \mathbb{R}^m$ by

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

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 [40, 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,

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

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,

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

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 $r_1, \dots, r_k \in \mathcal{Y}$ such that $r_1 \leq \Phi_1(f) + \varepsilon$, $T - r_k < \Phi_1(f) + \varepsilon$, and $r_i - r_{i-1} \leq \Phi_1(f) + \varepsilon$, $\forall i = 2, \dots, k$. 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,

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

Now, $\Lambda_n(x_i) \in (r_i, r_i + 2\varepsilon)$ for all sufficiently large n, since $x_i \in (r_i, r_i + \varepsilon)$. Thus, for all large enough n, there exists a point $z_i^n \in (r_i, r_i + 2\varepsilon)$ such that

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

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 in $(r_i, r_i + 2\varepsilon)$ for all large enough n. Also, using the fact that the complement of $\bigcup_{(l,r)\in\mathcal{E}}(l, r)$ 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 $(r_i^0, r_i^0 + 2\varepsilon)$ and thus in $(l_i - \varepsilon, l_i + 2\varepsilon)$, 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 $(\mathcal{A}_1(f_n),\dots,\mathcal{A}_m(f_n))$ converges 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* $\bar{\mathbf{S}}_\infty^\lambda$ *defined in* (1.5) *is good.*

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

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

which verifies Definition 1(a). Next, in order to verify Definition 1(b), let $r \in \mathcal{R}$ be any excursion end-point of $\bar{\mathbf{S}}_\infty^\lambda$. It is enough to show that

$$\inf\bigl\{t > 0 : \bar{S}_\infty^\lambda(r+t) - \bar{S}_\infty^\lambda(r) < 0\bigr\} = 0, \quad \text{almost surely.} \tag{5.26}$$

Let $V_r = \{i : \mathcal{I}_i(r) = 1\}$. Thus conditional on the sigma-field $\sigma(\bar{S}_\infty^\lambda(s) : s \leq r)$, the process $(\bar{S}_\infty^\lambda(r+t) - \bar{S}_\infty^\lambda(r))_{t\geq 0}$ is distributed as $\hat{\mathbf{S}}_\infty^\lambda$ given by

$$\hat{S}_\infty^\lambda(t) = \sum_{i\notin V_r}\theta_i\bigl(\mathcal{I}_i(t) - (\theta_i/\mu)t\bigr) + \lambda t. \tag{5.27}$$

Now, let $\mathbf{L}$ be the Lévy process defined as

$$L(t) = \sum_{i\notin V_r}\theta_i\bigl(\mathcal{N}_i(t) - (\theta_i/\mu)t\bigr) + \lambda t, \tag{5.28}$$

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 $\hat{S}_\infty^\lambda(t) \leq L(t)$ for all $t > 0$. Using [8, Theorem VII.1],

$$\inf\bigl\{t > 0 : L(t) < 0\bigr\} = 0, \quad \text{almost surely,} \tag{5.29}$$

and thus, (5.26) follows, and the proof is complete. □

5.4. *Finite-dimensional convergence*

As described in Section 4, the excursion lengths of the exploration process $\bar{\mathbf{S}}_n$ give 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* $n \to \infty$,

$$\bigl(\bar{\mathbf{S}}_n, \bar{\mathbf{N}}_n^\lambda\bigr) \xrightarrow{d} \bigl(\bar{\mathbf{S}}_\infty^\lambda, \mathbf{N}\bigr), \tag{5.30}$$

where $\mathbf{N}$ *is the counting process 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

$$\mathbb{P}(\xi_i=1|\mathscr{F}_{i-1})=\frac{A_{i-1}-1}{\ell_n-2i-1}=\frac{A_{i-1}}{\ell_n}\big(1+O(i/n)\big)+O\big(n^{-1}\big), \tag{5.31}$$

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

$$b_n\frac{A_{\lfloor tb_n\rfloor}}{n\mu}\big(1+o(1)\big)+o(1)=\frac{1}{\mu}\operatorname{refl}\big(\bar{S}_n(t)\big)\big(1+o(1)\big)+o(1). \tag{5.32}$$

We first argue that, for any fixed $u>0$,

$$\big(\bar{N}_n^\lambda(u)\big)_{n\geq 1}\text{ is tight in }\mathbb{R}_+. \tag{5.33}$$

Fix $\varepsilon>0$. Using Theorem 10, and the fact that the supremum of a process is continuous with respect to the Skorohod J_1-topology [49, Theorem 13.4.1], we can choose $K\geq 1$ large enough so that

$$\limsup_{n\to\infty}\mathbb{P}\Big(\sup_{i\leq\lfloor ub_n\rfloor}A_i>Ka_n\Big)<\varepsilon. \tag{5.34}$$

Fix times $0<l_1<\cdots<l_m\leq\lfloor ub_n\rfloor$, and let $\mathcal{A}(l_1,\dots,l_m)$ denote the event that the surplus edges appear at times $l_1,\dots,l_m$ and $A_{l_j-1}\leq Ka_n$ for all $j\in[m]$. Then,

$$\begin{aligned}
&\mathbb{P}\bigg(\sum_{i=2}^{\lfloor ub_n\rfloor}\xi_i\geq m,\text{ and }\sup_{i\leq\lfloor ub_n\rfloor}A_i\leq Ka_n\bigg)\\
&\quad\leq\sum_{0<l_1<\cdots<l_m\leq\lfloor ub_n\rfloor}\mathbb{P}\big(\mathcal{A}(l_1,\dots,l_m)\big)\\
&\quad\leq\sum_{0<l_1<\cdots<l_m\leq\lfloor ub_n\rfloor}\mathbb{E}\big[\mathbb{P}(\text{surplus created at }l_m|\mathscr{F}_{l_m-1})\mathbb{1}_{\{A_{l_m-1}\leq Ka_n\}}\mathbb{1}_{\mathcal{A}(l_1,\dots,l_{m-1})}\big]\\
&\quad\leq\frac{Ka_n}{\ell_n-2\lfloor ub_n\rfloor+1}\sum_{0<l_1<\cdots<l_m\leq\lfloor ub_n\rfloor}\mathbb{P}\big(\mathcal{A}(l_1,\dots,l_{m-1})\big).
\end{aligned} \tag{5.35}$$

Continuing the iteration in the last step, it follows that

$$\mathbb{P}\bigg(\sum_{i=2}^{\lfloor ub_n\rfloor}\xi_i\geq m,\text{ and }\sup_{i\leq\lfloor ub_n\rfloor}A_i\leq Ka_n\bigg)\leq\big(1+o(1)\big)\bigg(\frac{Ka_n}{\ell_n}\bigg)^m\frac{\lfloor ub_n\rfloor\cdots(\lfloor ub_n\rfloor-m+1)}{m!}, \tag{5.36}$$

which tends to zero in the iterated limit $\lim_{m\to\infty}\limsup_{n\to\infty}$. An application of (5.34) now yields (5.33).

Next, let $\mathbf{S}_n'$ be the process obtained by discarding the points where a surplus edge was added. More precisely, if $\zeta_l=S_n(l)-S_n(l-1)$, then we can define $S_n'(l)=S_n'(l-1)+\zeta_l'$, where

$$\zeta_l'=\zeta_{k_l},\quad\text{with }k_l=\inf\{j>k_{l-1}:\zeta_j\neq -2\}, k_0=0. \tag{5.37}$$

Let $\bar{S}_n'(t)=a_n^{-1}S_n'(\lfloor tb_n\rfloor)$. Also, let $d_{J_1,T}$ denote the metric for the Skorohod J_1-topology on $\mathbb{D}([0,T],\mathbb{R})$. We claim that, for any $T>0$ and $\varepsilon>0$,

$$\lim_{n\to\infty}\mathbb{P}\big(d_{J_1,T}\big(\bar{\mathbf{S}}_n',\bar{\mathbf{S}}_n\big)>\varepsilon\big)=0. \tag{5.38}$$

First, let $1\leq l_1<\cdots<l_K\leq\lfloor Tb_n\rfloor$ denote the times where the surplus edges have occurred. Also, let $\mathcal{A}$ be the good event that $l_j+1<l_{j+1}$ for all $j\leq K$, i.e., none of the surplus edges occur in consecutive steps. Note that

$$\mathbb{P}\Big(\mathcal{A}^c\cap\Big\{\sup_{i\leq\lfloor Tb_n\rfloor}A_i\leq Ka_n\Big\}\Big)\leq Tb_n\bigg(\frac{Ka_n}{\ell_n}\bigg)^2=O\big(b_n^{-1}\big), \tag{5.39}$$

and thus using (5.34), $\mathbb{P}(\mathcal{A}^c) \to 0$. We now restrict ourselves to $\mathcal{A}$. Putting $l_0 = 0$ and $l_{K+1} = \lfloor Tb_n \rfloor + 1$, let

$$\Lambda_n(l) = \begin{cases} l + j - 1 & \text{for } l_{j-1} < l < l_j, \\ l_j + j - 1 & \text{for } l = l_j - 0.5, \\ l_j + j & \text{for } l = l_j. \end{cases} \tag{5.40}$$

$\Lambda_n(t)$ is obtained by linearly interpolating between the values specified by (5.40). Also, note that the definition of Λ_n works well on $\mathcal{A}$, and on $\mathcal{A}^c$, we define $\Lambda_n(t) = t$. Using (5.33) and (5.39) it immediately follows that

$$\sup_{l \le Tb_n} \left|\Lambda_n(l) - l\right| = o_{\mathbb{P}}(b_n). \tag{5.41}$$

Moreover, the occurrence of each surplus edge causes $|S_n'(l) - S_n(\Lambda_n(l))|$ to increase by at most 2, so that

$$\sup_{l \le Tb_n} \left|S_n'(l) - S_n\big(\Lambda_n(l)\big)\right| = o_{\mathbb{P}}(a_n). \tag{5.42}$$

Then, (5.38) follows by combining (5.41) and (5.42). We now proceed to complete the proof of Lemma 16. Let us start by setting up some notation for the rest of the proof. Fix $T > 0$, $k \ge 0$ and let $\mathrm{Surp}_T = \{l_1, \dots, l_k\}$ be the surplus generation times, where $1 \le l_1 < l_2 < \cdots < l_k \le \lfloor Tb_n \rfloor + k$. Let $(z_l)_{l \le \lfloor Tb_n \rfloor + k}$ be a sequence of integers such that $z_{l_i} = -2$ and $z_l \ge -1$ for $l \notin \{l_1, \dots, l_k\}$. Thus $(z_l)_{l \le \lfloor Tb_n \rfloor + k}$ represents the increments of a sample path of S_n which has explored k surplus edges, and Surp_T is the set of times when surplus edges are found. Next, $(z_l')_{l \le \lfloor Tb_n \rfloor}$ denote the sequence obtained from $(z_l)_{l \le \lfloor Tb_n \rfloor + k}$ by deleting the -2's. Thus, $(z_l')_{l \le \lfloor Tb_n \rfloor}$ corresponds to the increments of a sample path of S_n'. Recall that $\zeta_l = S_n(l) - S_n(l-1)$. Thus,

$$\begin{aligned}
&\mathbb{P}\left(N_n^\lambda\big(\lfloor Tb_n \rfloor + k\big) = k \,\middle|\, \big(S_n'(l)\big)_{l \le \lfloor Tb_n \rfloor} = \left(\sum_{j \le l} z_j'\right)_{l \le \lfloor Tb_n \rfloor}, N_n^\lambda\big(\lfloor Tb_n \rfloor + k\big) \le \omega_n\right) \\
&= \sum_{1 \le l_1 < \cdots < l_k \le Tb_n + k} \mathbb{P}\left(\text{surplus occurs only at times } l_1, \dots, l_k \,\middle|\, \begin{array}{l} \big(S_n'(l)\big)_{l \le Tb_n} = \left(\sum_{j \le l} z_j'\right)_{l \le \lfloor Tb_n \rfloor}, \\ N_n^\lambda\big(\lfloor Tb_n \rfloor + k\big) \le \omega_n \end{array}\right) \\
&= \sum_{1 \le l_1 < \cdots < l_k \le Tb_n + k} \frac{\mathbb{P}(\zeta_l = z_l, \text{ for all } 1 \le l \le \lfloor Tb_n \rfloor + k)}{\mathbb{P}((S_n'(l))_{l \le Tb_n} = (\sum_{j \le l} z_j')_{l \le \lfloor Tb_n \rfloor}, N_n^\lambda(\lfloor Tb_n \rfloor + k) \le \omega_n)}.
\end{aligned} \tag{5.43}$$

Define $m_1 = \#\{i \in [n] : d_i = z_1 + 2\}$, and for $l \notin \mathrm{Surp}_T$, we denote $m_l = \#\{i \in [n] : d_i = z_l + 2\} - \#\{j < l : z_j = z_l\}$. Thus, m_l gives the number of degree $z_l + 2$ vertices in the system at time l, which are potential candidates to cause a jump of z_l at time l. Next, let a_l denote the number of active half-edges at time l when the exploration process takes the path $(z_l)_{l \le \lfloor Tb_n \rfloor + k}$, and $a_l' = S_n'(l) - \min_{j<l} S_n'(j)$. Now,

$$\begin{aligned}
\mathbb{P}\big(\zeta_l = z_l, \forall l \le \lfloor Tb_n \rfloor + k\big) &= \frac{\prod_{l \notin \mathrm{Surp}_T} m_l \times \prod_{j=1}^k (a_{l_j - 1} - 1)}{(\ell_n - 1)(\ell_n - 3) \cdots (\ell_n - 2\lfloor Tb_n \rfloor - 2k + 1)} \\
&= \frac{\prod_{l \notin \mathrm{Surp}_T} m_l \times \prod_{j=1}^k (a_{l_j - 1} - 1)}{(\ell_n - 1) \cdots (\ell_n - 2\lfloor Tb_n \rfloor + 1)} \times \big(1 + o(1)\big) \prod_{j=1}^k \frac{a_{l_j - 1}'}{\ell_n},
\end{aligned} \tag{5.44}$$

where the $o(1)$ term above is uniform in $k \le \omega_n = \log n$. Thus,

$$(5.43) = \big(1 + o(1)\big) \frac{\sum_{1 \le l_1 < \cdots < l_k \le \lfloor Tb_n \rfloor + k} \prod_{j=1}^k \frac{a_{l_j - 1}'}{\ell_n^k}}{\sum_{r=0}^{\omega_n} \sum_{1 \le l_1 < \cdots < l_r \le \lfloor Tb_n \rfloor + r} \prod_{j=1}^r \frac{a_{l_j - 1}'}{\ell_n^r}} =: \big(1 + o(1)\big) \frac{\beta_{n,k}}{\sum_{r=0}^\infty \beta_{n,r}}, \tag{5.45}$$

where $\beta_{n,r} = 0$ for $r > \omega_n$. Now, Theorem 8 together with (5.38) also implies that $\bar{\mathbf{S}}_n' \xrightarrow{d} \mathbf{S}_\infty^\lambda$ with respect to the Skorohod J_1-topology. Since the reflection of a process is continuous in Skorohod J_1-topology (see [49, Lemma 13.5.1]) it also follows that $\mathrm{refl}(\bar{\mathbf{S}}_n') \xrightarrow{d} \mathrm{refl}(\mathbf{S}_\infty^\lambda)$. Thus,

$$\big((\beta_{n,r})_{r\geq 0}, \big(\bar{S}_n'(u)\big)_{u\leq T}\big) \xrightarrow{d} \left(\left(\frac{1}{r!}\left(\frac{1}{\mu}\int_0^T \mathrm{refl}\big(\bar{S}_\infty^\lambda(u)\big)\,\mathrm{d}u\right)^r\right)_{r\geq 0}, \big(\bar{S}_\infty^\lambda(u)\big)_{u\leq T}\right), \tag{5.46}$$

where the convergence of $(\beta_{n,r})_{r\geq 0}$ holds with respect to the product topology on $\mathbb{R}^\infty$. Next, let us ensure that $\sum_{r=0}^\infty \beta_{n,r}$ in (5.43) converges to the desired quantity. To this end, consider a probability space where the convergence of (5.46) holds almost surely. On this space, $\sup_{l\leq Tb_n+k}\mathrm{refl}(S_n'(l)) \leq 2(\sup_{l\leq Tb_n+k} S_n'(l)+\omega_n) =: X_n(T)$, and thus

$$\beta_{n,r} \leq \frac{(Tb_n+\omega_n)^r}{r!}\frac{X_n(T)^r}{\ell_n^r}. \tag{5.47}$$

Since $a_n^{-1}\sup_{l\leq Tb_n+k} S_n'(l)$ converges, an application of Dominated Convergence Theorem yields that

$$\sum_{r\geq 0}\beta_{n,r} \xrightarrow{\text{a.s.}} \sum_{r\geq 0}\frac{1}{r!}\left(\frac{1}{\mu}\int_0^T \mathrm{refl}\big(\bar{S}_\infty^\lambda(u)\big)\,\mathrm{d}u\right)^r = \exp\left(\frac{1}{\mu}\int_0^T \mathrm{refl}\big(\bar{S}_\infty^\lambda(u)\big)\,\mathrm{d}u\right). \tag{5.48}$$

Next, for bounded continuous functions $\phi_1 : \mathbb{D}([0,T],\mathbb{R})\to\mathbb{R}$ and $\phi_2:\mathbb{N}\to\mathbb{R}$,

$$\begin{aligned}
&\mathbb{E}\big[\phi_1\big(\big(\bar{S}_n'(u)\big)_{u\leq T}\big)\phi_2\big(\bar{N}_n^\lambda(T)\big)\big]\\
&\quad= o(1)+\mathbb{E}\big[\phi_1\big(\big(\bar{S}_n'(u)\big)_{u\leq T}\big)\phi_2\big(\bar{N}_n^\lambda(T)\big)\mathbb{1}_{\{N_n^\lambda(\lfloor Tb_n\rfloor+k)\leq\omega_n\}}\big]\\
&\quad= o(1)+\mathbb{E}\left[\phi_1\big(\big(\bar{S}_n'(u)\big)_{u\leq T}\big)\mathbb{1}_{\{N_n^\lambda(\lfloor Tb_n\rfloor+k)\leq\omega_n\}}\times\big(1+o(1)\big)\frac{\sum_{k\geq 0}\phi_2(k)\beta_{n,k}}{\sum_{r\geq 0}\beta_{n,r}}\right]\\
&\quad= o(1)+\mathbb{E}\left[\phi_1\big(\big(\bar{S}_n'(u)\big)_{u\leq T}\big)\times\frac{\sum_{k\geq 0}\phi_2(k)\beta_{n,k}}{\sum_{r\geq 0}\beta_{n,r}}\right]\to\mathbb{E}\big[\phi_1\big(\big(\bar{S}_\infty^\lambda(u)\big)_{u\leq T}\big)\phi_2\big(N(T)\big)\big],
\end{aligned}\tag{5.49}$$

where $N(T)$, conditionally on $(\bar{S}_\infty^\lambda(u))_{u\leq T}$, is distributed as $\mathrm{Poisson}(\frac{1}{\mu}\int_0^T\mathrm{refl}(\bar{S}_\infty^\lambda(u))\,\mathrm{d}u)$. We have used (5.33) in the third step, and the final step follows by combining (5.46) and (5.48). Hence, we have shown that, for any $T>0$,

$$\big(\big(\bar{S}_n'(u)\big)_{u\leq T}, \bar{N}_n^\lambda(T)\big) \xrightarrow{d} \big(\big(\bar{S}_\infty^\lambda(u)\big)_{u\leq T}, N(T)\big). \tag{5.50}$$

Next, let $U_1^n < U_2^n < \cdots$ denote the location of surplus edges in the process S_n. Then, using (5.44) yields

$$\begin{aligned}
&\mathbb{P}\big(U_j^n = l_j, \text{for all } j\in[k]\mid\big(\bar{S}_n'(u)\big)_{u\leq T}, \bar{N}_n^\lambda(T)=k\big)\\
&\quad=\big(1+o(1)\big)\frac{\frac{1}{\ell_n^k}\prod_{j=1}^k(A_{l_j}-1)}{\sum_{1\leq l_1'<\cdots<l_k'\leq\lfloor Tb_n\rfloor+k}\frac{1}{\ell_n^k}\prod_{j=1}^k(A_{l_j'}-1)}.
\end{aligned}\tag{5.51}$$

From this, it can be seen that the law of $b_n^{-1}(U_j^n)_{j\in[k]}$, conditionally on $(\bar{S}_n'(u))_{u\leq T}$, and $\bar{N}_n^\lambda(T)=k$, converges to the order-statistics of k i.i.d. random variables with density $\frac{\mathbb{1}_{\{u\in[0,T]\}}\mathrm{refl}(\bar{S}_\infty^\lambda(u))}{\int_0^T\mathrm{refl}(\bar{S}_\infty^\lambda(u))\,\mathrm{d}u}$. Now, combining (5.50), (5.51) together with (5.38) completes the proof of Lemma 16. □

Remark 14. In [22] and an earlier version of the paper, we concluded the proof of Lemma 16 from the convergence of compensators only. We thank Lorenzo Federico and Tim Hulshof for pointing out a gap in this argument. Indeed, for Erdős–Rényi random graphs [4] or rank-one inhomogeneous random graphs [14,15], showing the convergence of compensators suffices using [21, Theorem 1] since the surplus edges can be added independently after we have observed the whole exploration process. However, this is not true for the configuration model because the surplus edges occur precisely at places with jumps -2. For this reason, we have modified the proof and the identical argument also completes the proof in [22, Lemma 5.16].

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

$$b_n^{-1}\big(|\mathscr{C}_{(1)}|, |\mathscr{C}_{(2)}|,\ldots,|\mathscr{C}_{(m)}|\big) \xrightarrow{d} \big(\gamma_1(\lambda),\gamma_2(\lambda),\ldots,\gamma_m(\lambda)\big) \tag{5.52}$$

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

Proof. Fix any $m \geq 1$. Let $\mathscr{C}_{(i)}^T$ be the ith largest component explored by Algorithm 1 up to time Tb_n. Denote by $\mathscr{D}_{(i)}^{\mathrm{ord},T}$ the ith 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

$$\lim_{n\to\infty} \mathbb{E}\big[g\big((2b_n)^{-1}\big(\mathscr{D}_{(1)}^{\mathrm{ord},T}, \mathscr{D}_{(2)}^{\mathrm{ord},T}, \dots, \mathscr{D}_{(m)}^{\mathrm{ord},T}\big)\big)\big] = \mathbb{E}\big[g\big(\gamma_1^T(\lambda), \gamma_2^T(\lambda), \dots, \gamma_m^T(\lambda)\big)\big], \tag{5.53}$$

where $\gamma_i^T(\lambda)$ is the ith largest excursion of $\bar{\mathbf{S}}_\infty$ restricted to the time interval $[0,T]$. Now the support of the joint distribution of $(\gamma_i^T(\lambda))_{i\geq 1}$ is concentrated on $\{(x_1,x_2,\dots): x_1 > x_2 > \cdots\}$. Thus, using Lemma 16, it follows that

$$\lim_{n\to\infty} \mathbb{E}\big[g\big(b_n^{-1}\big(|\mathscr{C}_{(1)}^T|, |\mathscr{C}_{(2)}^T|, \dots, |\mathscr{C}_{(m)}^T|\big)\big)\big] = \mathbb{E}\big[g\big(\gamma_1^T(\lambda), \gamma_2^T(\lambda), \dots, \gamma_m^T(\lambda)\big)\big]. \tag{5.54}$$

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

$$\lim_{T\to\infty} \mathbb{E}\big[g\big(\gamma_1^T(\lambda), \gamma_2^T(\lambda), \dots, \gamma_m^T(\lambda)\big)\big] = \mathbb{E}\big[g\big(\gamma_1(\lambda), \gamma_2(\lambda), \dots, \gamma_m(\lambda)\big)\big]. \tag{5.55}$$

Finally, using Lemma 13, the proof of Theorem 17 is completed by (5.54) and (5.55). □

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

6. Convergence in the $\mathbb{U}_\downarrow^0$ 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* $\ell_\downarrow^2 \times \mathbb{N}^\infty$ *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:

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

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

Proof of Theorem 2. First $(X_{ni}, Y_{ni})_{i\geq 1} \xrightarrow{d} (X_i, Y_i)_{i\geq 1}$ in $\mathbb{U}_\downarrow^0$-topology if the following three conditions hold:

(i) $(X_{ni}, Y_{ni})_{i=1}^k \xrightarrow{d} (X_i, Y_i)_{i=1}^k$ for all $k \geq 1$.
(ii) $\lim_{\delta\to 0}\limsup_{n\to\infty} \mathbb{P}(\sum_{i:X_{ni}\leq\delta} X_{ni}^2 > \varepsilon) = 0$, for any $\varepsilon > 0$.
(iii) $\lim_{\delta\to 0}\limsup_{n\to\infty} \mathbb{P}(\sum_{i:X_{ni}\leq\delta} X_{ni}Y_{ni} > \varepsilon) = 0$, for any $\varepsilon > 0$.

To see this, note that $\|(X_{ni}, Y_{ni})_{i\geq 1}\|_{\mathbb{U}^0} = \|(X_{ni})_{i\geq 1}\|_2 + \sum_i X_{ni}Y_{ni}$, where $\|\cdot\|_{\mathbb{U}^0}$ denotes the norm induced by metric in (1.3). Conditions (i) and (ii) together implies that $(\|(X_{ni})_{i\geq 1}\|_2)_{n\geq 1}$ is tight, and (i) and (iii) together implies that $(\sum_i X_{ni}Y_{ni})_{n\geq 1}$ is tight. Thus under (i)–(iii), $((X_{ni}, Y_{ni})_{i\geq 1})_{n\geq 1}$ is tight in $\mathbb{U}_\downarrow^0$-topology, and the finite dimensional convergence in (i) yields that $(X_{ni}, Y_{ni})_{i\geq 1} \xrightarrow{d} (X_i, Y_i)_{i\geq 1}$ in $\mathbb{U}_\downarrow^0$-topology.

To complete the proof of Theorem 2, note that Lemma 18 yields conditions (i) and (ii), and Proposition 19 yields condition (iii). Thus the proof follows. □

The remainder of this section is devoted to the proof of Proposition 19. 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 used in [22].

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,*

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

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 $\varepsilon, \delta > 0$. Note that

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

The third step in (6.3) follows using

$$\begin{aligned}
&\mathbb{E}\big[\mathrm{SP}\big(\mathscr{C}(V_n)\big)\mathbb{1}_{\{|\mathscr{C}(V_n)|\leq \delta b_n\}}\big] \\
&\quad = \sum_{k=1}^{\infty}\sum_{j\geq k}\mathbb{P}\big(\mathrm{SP}\big(\mathscr{C}(V_n)\big) = j, \big|\mathscr{C}(V_n)\big| \leq \delta b_n\big) \\
&\quad = \sum_{k=1}^{\infty}\sum_{j\geq k}\sum_{i\geq \log_2(1/(k^{0.12}\delta))} \mathbb{P}\big(\mathrm{SP}\big(\mathscr{C}(V_n)\big) = j, \big|\mathscr{C}(V_n)\big| \in \big(2^{-(i+1)}k^{-0.12}b_n, 2^{-i}k^{-0.12}b_n\big]\big),
\end{aligned} \tag{6.4}$$

and the second-last step in (6.3) follows from Lemma 20. The proof of Proposition 19 now follows for $\lambda < 0$.

Next, 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 $\delta > 0$, it is possible to choose $T > 0$ such that for all sufficiently large n,

$$\mathbb{P}(\text{all the vertices } 1, \dots, R \text{ are explored within time } Tb_n) > 1 - \delta. \tag{6.5}$$

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

$$\sum_{i\notin \mathscr{V}_{T_e}} d_i \geq \ell_n - 2Tb_n \quad \Longrightarrow \quad \sum_{i\notin \mathscr{V}_{T_e}} d_i = \ell_n + o_{\mathbb{P}}(n), \tag{6.6}$$

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

$$\nu_n^* \leq \frac{\sum_{i\in[n]} d_i(d_i - 1) - \sum_{i=1}^{R} d_i(d_i - 1)}{\sum_{i\notin \mathscr{V}_{T_e}} d_i} - 1 = 1 + c_n^{-1}\bigg(\lambda - \sum_{i=1}^{R}\theta_i^2\bigg) + o\big(c_n^{-1}\big). \tag{6.7}$$

Therefore, combining (6.5), (6.7), we can use Lemma 20 on $\mathcal{G}_T^*$ since $c_n(\nu_n^*-1)<0$. Thus, if $\mathscr{C}_{(i)}^*$ denotes the ith largest component of $\mathcal{G}_T^*$, then

$$\lim_{T\to\infty}\lim_{\delta\to 0}\limsup_{n\to\infty}\mathbb{P}\bigg(\sum_{i:|\mathscr{C}_{(i)}^*|\leq\delta b_n}\big|\mathscr{C}_{(i)}^*\big|\times\mathrm{SP}\big(\mathscr{C}_{(i)}^*\big)>\varepsilon b_n\bigg)=0. \tag{6.8}$$

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

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

Note that

$$\begin{aligned}\sum_{i\in\mathcal{K}_n^T}|\mathscr{C}_{(i)}|\times\mathrm{SP}(\mathscr{C}_{(i)})&\leq\bigg(\sum_{i\in\mathcal{K}_n^T}|\mathscr{C}_{(i)}|^2\bigg)^{1/2}\times\bigg(\sum_{i\in\mathcal{K}_n^T}\mathrm{SP}(\mathscr{C}_{(i)})^2\bigg)^{1/2}\\&\leq\bigg(\sum_{i:|\mathscr{C}_{(i)}|\leq\delta b_n}|\mathscr{C}_{(i)}|^2\bigg)^{1/2}\times\mathrm{SP}(T_e),\end{aligned} \tag{6.9}$$

where $\mathrm{SP}(t)$ is the number of surplus edges explored up to time tb_n and we have used the fact that $\sum_{i\in\mathcal{K}_n^T}\mathrm{SP}(\mathscr{C}_{(i)})^2\leq(\sum_{i\in\mathcal{K}_n^T}\mathrm{SP}(\mathscr{C}_{(i)}))^2\leq\mathrm{SP}(T_e)^2$. From Lemma 16 and the $\ell^2_{\downarrow}$ tightness in Theorem 1, we can conclude that for any $T>0$,

$$\lim_{\delta\to 0}\limsup_{n\to\infty}\mathbb{P}\bigg(\sum_{i\in\mathcal{K}_n^T}|\mathscr{C}_{(i)}|\times\mathrm{SP}(\mathscr{C}_{(i)})>\varepsilon b_n\bigg)=0. \tag{6.10}$$

The proof is now complete for the case $\lambda>0$ by combining (6.8) and (6.10). □

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 [29, 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 $f:\mathbb{D}([0,T],\mathbb{R})\mapsto\mathbb{R}$,

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

Let $\ell_n':=\ell_n-2Tb_n-d_1+1$. 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, and let d_{v_l} denote the degree of v_l. Note that, uniformly over $l\leq Tb_n$, any half-edge of v_l creates a self-loop with probability at most d_{v_l}/ℓ_n'. Thus, the expected number of self-loops attached to v_l, conditionally on $\mathscr{F}_{l-1}$, is at most $d_{v_l}^2/\ell_n'$. Moreover, the expected number of multiple edge attached to v_l, conditionally on $\mathscr{F}_{l-1}$, is at most

$$\frac{d_{v_l}(d_{v_l}-1)\sum_{i\in[n]}d_i(d_i-1)}{\ell_n'(\ell_n'-1)}=\big(1+o(1)\big)\frac{d_{v_l}^2}{\ell_n'}, \tag{7.2}$$

where we have used that $\nu_n=1+o(1)$. Therefore,

$$\mathbb{E}\big[\#\{\text{self-loops or multiple edges discovered while }v_l\text{ is exploring}\}\mid\mathscr{F}_{l-1}\big]\leq\frac{3d_{v_l}^2}{\ell_n'}. \tag{7.3}$$

Thus, for any $T > 0$,

$$\begin{aligned}
&\mathbb{E}\big[\#\{\text{self-loops or multiple edges discovered up to time } Tb_n\}\big] \\
&\quad\le \frac{3}{\ell_n'}\mathbb{E}\bigg[\sum_{i\in[n]} d_i^2\mathcal{I}_i^n(Tb_n)\bigg] = \frac{3}{\ell_n'}\mathbb{E}\bigg[\sum_{i=1}^{K} d_i^2\mathcal{I}_i^n(Tb_n)\bigg] + \frac{3}{\ell_n'}\mathbb{E}\bigg[\sum_{i=K+1}^{n} d_i^2\mathcal{I}_i^n(Tb_n)\bigg],
\end{aligned}\tag{7.4}$$

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

$$\frac{3}{\ell_n'}\mathbb{E}\bigg[\sum_{i=1}^{K} d_i^2\mathcal{I}_i^n(Tb_n)\bigg] \le \frac{3}{\ell_n'}\sum_{i=1}^{K} d_i^2 \to 0,\tag{7.5}$$

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

$$\frac{3}{\ell_n'}\mathbb{E}\bigg[\sum_{i=K+1}^{n} d_i^2\mathcal{I}_i^n(Tb_n)\bigg] \le \frac{3Tb_n}{\ell_n'^2}\sum_{i=K+1}^{n} d_i^3 \le C\bigg(a_n^{-3}\sum_{i=K+1}^{n} d_i^3\bigg),\tag{7.6}$$

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 $n \to \infty$,

$$\mathbb{P}(\text{at least one self-loop or multiple edge is discovered before time } Tb_n) \to 0.\tag{7.7}$$

Now,

$$\begin{aligned}
&\mathbb{E}\big[f(\bar{\mathbf{S}}_n^T)\mathbb{1}_{\{\mathrm{CM}_n(\boldsymbol{d})\text{ is simple}\}}\big] \\
&\quad= \mathbb{E}\big[f(\bar{\mathbf{S}}_n^T)\mathbb{1}_{\{\text{no self-loops or multiple edges found after time } Tb_n\}}\big] + o(1) \\
&\quad= \mathbb{E}\big[f(\bar{\mathbf{S}}_n^T)\mathbb{P}(\text{no self-loops or multiple edges found after time } Tb_n | \mathscr{F}_{Tb_n})\big] + o(1),
\end{aligned}\tag{7.8}$$

Define $T_e = \inf\{l \ge Tb_n : \text{a component is finished exploring at time } l\}$. Using the fact that $(b_n^{-1}T_e)_{n\ge1}$ is a tight sequence, the limit of the expected number of self-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 of $\mathcal{G}^*$. 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 $t > 0$,

$$\limsup_{n\to\infty}\mathbb{E}\bigg[a_n^{-2}\sum_{i\in[n]} d_i^2\mathcal{I}_i^n(tb_n)\bigg] \le \limsup_{n\to\infty} a_n^{-2}tb_n\frac{\sum_{i\in[n]} d_i^3}{\ell_n - 2tb_n} < \infty,\tag{7.9}$$

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 [45, Theorem 7.11] we get

$$\mathbb{P}\big(\mathcal{G}^* \text{ is simple} \mid \mathscr{F}_{T_e}\big) \xrightarrow{\mathbb{P}} \mathrm{e}^{-3/4}.\tag{7.10}$$

Now using (7.8), (7.10) and the dominated convergence theorem, we conclude that

$$\mathbb{E}\big[f(\bar{\mathbf{S}}_n^T)\mathbb{1}_{\{\mathrm{CM}_n(\boldsymbol{d})\text{ is simple}\}}\big] = \mathbb{E}\big[f(\bar{\mathbf{S}}_n^T)\big]\mathbb{P}(\mathrm{CM}_n(\boldsymbol{d})\text{ is simple}) + o(1).\tag{7.11}$$

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 $(\mathscr{W}_i)_{i\ge1}$ when the weight sequence satisfies some regularity conditions:

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 into two main steps: the first one is to obtain the finite-dimensional limits of $\mathbf{Z}_n^w$ and secondly to 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 by Algorithm 1 upto time l.

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

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

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

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

Proof. Fix any $T>0$. Let $\ell_n(T) = \ell_n - 2Tb_n + 1$. Define

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

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

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

uniformly over $l\leq Tb_n$ and therefore, $(W_n(l))_{l=1}^{Tb_n}$ is a supermartingale. Using (4.8), we compute

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

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

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

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 $\varepsilon > 0$

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

and

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

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 of 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

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

Now, [30, 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 35 given in Appendix A. The proof of (8.8b) can also be completed using an identical argument as in the proof of Proposition 19 by observing that

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

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 in 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* $\mathbf{Z}_n^k = \mathrm{ord}(b_n^{-1} v_k(\mathscr{C}_{(i)}), \mathrm{SP}(\mathscr{C}_{(i)}))_{i\geq 1}$, $\mathbf{Z}^k := \mathrm{ord}(\frac{kr_k}{\mu}\gamma_i(\lambda), N(\gamma_i))_{i\geq 1}$, *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 with $\mu_w = kr_k/\mu$ 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 of 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 [25]. 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 [28] 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 [28] 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 15. 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 remains 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 jth half-edge corresponding to vertex i being kept after the explosion. Then $I_{ij} \sim \mathrm{Ber}(\sqrt{p_n})$ 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 [32, Corollary 2.27] yields

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

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

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

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 ith 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

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

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 $n_+ \sim \mathrm{Bin}(\ell_n, 1 - \sqrt{p_n})$, and $a_n \gg n^{1/3}$ for $\tau \in (3,4)$. To see that $\tilde{\boldsymbol{d}}$ satisfies Assumption 1(iii) note that by (9.4),

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

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 $\tilde{d}_i \sim \mathrm{Bin}(d_i, \sqrt{p_n})$. A standard concentration inequality for the binomial distribution [32, (2.9)] yields that, for any $0 < \varepsilon \leq 3/2$,

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

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 ith 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 $\mathbb{U}_{\downarrow}^0$ 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 $\mathrm{CM}_{\tilde{n}}(\tilde{\boldsymbol{d}})$. Recall that $\tilde{n}_1$ is the number of degree-one vertices left after Algorithm 2(S2). Since the vertices are to be chosen uniformly from all degree-one vertices as described in (S3′),

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

where the penultimate 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 surplus edge counts. 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. For all the results in this section, we assume that the assumptions of Theorem 5 hold.

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 *sandwiches* the graph process $(\mathrm{CM}_n(\boldsymbol{d}, p_n(\lambda)))_{\lambda\in\mathbb{R}}$. This result was proved in [22, Proposition 8.4]. 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*

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

and

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

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

We next prove an important consequence of Proposition 24 which says that the rescaled component sizes and surplus edges for $\mathrm{CM}_n(\boldsymbol{d}, p_n(\lambda))$ and $\mathcal{G}_n(t_n(\lambda))$ are close in $\mathbb{U}^0_\downarrow$. This allows us to focus our attention to $\mathcal{G}_n(t_n(\lambda))$ in the subsequent analysis. From here onward, we augment λ to a predefined notation to emphasize the dependence on λ. Let $\mathscr{C}_{(i)}(\lambda)$ and $\mathscr{C}^p_{(i)}(\lambda)$ denote respectively the ith largest component of $\mathrm{CM}_n(\boldsymbol{d}, p_n(\lambda))$ and $\mathcal{G}_n(t_n(\lambda))$. Also, let $\mathbf{Z}_n(\lambda)$ and $\mathbf{Z}^p_n(\lambda)$ denote the vectors $(b_n^{-1}|\mathscr{C}_{(i)}(\lambda)|, \mathrm{SP}(\mathscr{C}_{(i)}(\lambda)))_{i\geq 1}$ and $(b_n^{-1}|\mathscr{C}^p_{(i)}(\lambda)|, \mathrm{SP}(\mathscr{C}^p_{(i)}(\lambda)))_{i\geq 1}$ respectively, ordered as an element of $\mathbb{U}^0_\downarrow$. Recall the definition of the metric $\mathrm{d}_{\mathbb{U}}$ from (1.3).

Proposition 25. *Under the coupling in Proposition 24, as $n \to \infty$, $\mathrm{d}_{\mathbb{U}}(\mathbf{Z}_n(\lambda), \mathbf{Z}^p_n(\lambda)) \xrightarrow{\mathbb{P}} 0$ for any $\lambda \in \mathbb{R}$.*

We first prove the following elementary fact that will be crucial in our analysis:

Fact 1. $\mathbb{P}(\forall i \geq 1 : \gamma_{i+1} < \gamma_i) = 1$, where γ_i denotes the length of the ith largest excursion in (1.5).

Proof. It is enough to show that, with probability one, no two excursions of a process in (1.5) have the same length. For any rational q, let $e(q)$ be the excursion containing q. Thus it is enough to show that for two rationals q_1, q_2,

$$\mathbb{P}\big(e(q_1) \neq e(q_2), \text{but } \big|e(q_1)\big| = \big|e(q_2)\big|\big) = 0. \tag{10.3}$$

Without loss of generality, suppose that $e(q_1)$ appears earlier than $e(q_2)$. Let $V_{q_2} = \{i : \mathcal{I}_i(q_2) = 1\}$. Thus conditional on the sigma-field $\sigma(\bar{S}_\infty^\lambda(s) : s \leq q_2)$, the process $(\bar{S}_\infty^\lambda(q_2 + t) - \bar{S}_\infty^\lambda(q_2))_{t\geq 0}$ is distributed as $\hat{\mathbf{S}}_\infty^\lambda$ given by

$$\hat{S}_\infty^\lambda(t) = \sum_{i\notin V_{q_2}} \theta_i\big(\mathcal{I}_i(t) - (\theta_i/\mu)t\big) + \lambda t. \tag{10.4}$$

Therefore, the process in (10.4) again has the form (1.5). Now, for any $x > 0$, the probability that $|e(q_2)| = x$, conditionally on $\sigma(\bar{S}_\infty^\lambda(s) : s \leq q_2)$ and $|e(q_1)| = x$, is zero using Lemma 36 from Appendix C. This concludes the proof of Fact 1. ☐

Proof of Proposition 25. Let $\mathscr{C}_{(i)}^{p,+}(\lambda)$ (respectively, $\mathscr{C}_{(i)}^{p,-}(\lambda)$) denote the ith largest component of $\mathrm{CM}_n(\boldsymbol{d}, p_n(\lambda) + \varepsilon_n)$ (respectively, $\mathrm{CM}_n(\boldsymbol{d}, p_n(\lambda) - \varepsilon_n)$) where we have chosen ε_n according to Proposition 24. Let $\mathbf{Z}_n^{p,+}(\lambda)$ (respectively, $\mathbf{Z}_n^{p,-}(\lambda)$) denote the vector $(b_n^{-1}|\mathscr{C}_{(i)}^{p,+}(\lambda)|, \mathrm{SP}(\mathscr{C}_{(i)}^{p,+}(\lambda)))_{i\geq 1}$ (respectively, $(b_n^{-1}|\mathscr{C}_{(i)}^{p,-}(\lambda)|, \mathrm{SP}(\mathscr{C}_{(i)}^{p,-}(\lambda)))_{i\geq 1}$), ordered as an element of $\mathbb{U}_\downarrow^0$. Using Theorem 4, both $\mathbf{Z}_n^{p,+}(\lambda)$ and $\mathbf{Z}_n^{p,-}(\lambda)$ have the same scaling limit.

Next, for any $K \geq 1$,

$$\lim_{n\to\infty} \mathbb{P}\big(\mathscr{C}_{(i)}^{p,-}(\lambda) \subset \mathscr{C}_{(i)}^{p,+}(\lambda), \forall i \leq K\big) = 1. \tag{10.5}$$

If $\mathscr{C}_{(1)}^{p,-}(\lambda)$ is not contained in $\mathscr{C}_{(1)}^{p,+}(\lambda)$, then $|\mathscr{C}_{(1)}^{p,-}(\lambda)| \leq |\mathscr{C}_{(j)}^{p,+}(\lambda)|$ for some $j \geq 2$, which implies that $|\mathscr{C}_{(1)}^{p,-}(\lambda)| \leq |\mathscr{C}_{(2)}^{p,+}(\lambda)|$. Suppose that there is a subsequence $(n_{0k})_{k\geq 1}$ along which

$$\lim_{n_{0k}\to\infty} \mathbb{P}\big(\big|\mathscr{C}_{(1)}^{p,-}(\lambda)\big| \leq \big|\mathscr{C}_{(2)}^{p,+}(\lambda)\big|\big) > 0. \tag{10.6}$$

If (10.6) yields a contradiction, then (10.5) is proved for $K = 1$. To this end, first note that $(b_n^{-1}(|\mathscr{C}_{(i)}^{p,-}(\lambda)|, |\mathscr{C}_{(i)}^{p,+}(\lambda)|)_{i\geq 1})_{n\geq 1}$ is tight in $(\ell_\downarrow^2)^2$. Thus taking a subsequence $(n_k)_{k\geq 1} \subset (n_{0k})_{k\geq 1}$ along which the random vector converges, it follows that

$$b_{n_k}^{-1}\big(\big|\mathscr{C}_{(i)}^{p,-}(\lambda)\big|, \big|\mathscr{C}_{(i)}^{p,+}(\lambda)\big|\big)_{i\geq 1} \xrightarrow{d} (\gamma_i, \bar{\gamma}_i)_{i\geq 1} \quad \text{in } \big(\ell_\downarrow^2\big)^2, \tag{10.7}$$

where $(\gamma_i)_{i\geq 1} \stackrel{d}{=} (\bar{\gamma}_i)_{i\geq 1}$. Thus, along the subsequence $(n_k)_{k\geq 1}$,

$$\lim_{n_k\to\infty} \mathbb{P}\big(\big|\mathscr{C}_{(1)}^{p,-}(\lambda)\big| \leq \big|\mathscr{C}_{(2)}^{p,+}(\lambda)\big|\big) = \mathbb{P}(\gamma_1 \leq \bar{\gamma}_2). \tag{10.8}$$

Fact 2. For all $i \geq 1$, $\gamma_i = \bar{\gamma}_i$ almost surely.

Proof. Under the coupling in Proposition 24, $\sum_{j\leq i} |\mathscr{C}_{(j)}^{p,-}(\lambda)| \leq \sum_{j\leq i} |\mathscr{C}_{(j)}^{p,+}(\lambda)|$ and therefore $\mathbb{P}(\sum_{j\leq i} \gamma_j \leq \sum_{j\leq i} \bar{\gamma}_j) = 1$, for each fixed $i \geq 1$. In particular, $\gamma_1 \leq \bar{\gamma}_1$. But, since γ_1, $\bar{\gamma}_1$ have the same distribution, it must be the case that $\gamma_1 = \bar{\gamma}_1$ almost surely. Inductively, we can prove that $\gamma_i = \bar{\gamma}_i$ almost surely. ☐

Thus, using Fact 2, (10.8) reduces to

$$\lim_{n_k\to\infty} \mathbb{P}\big(\big|\mathscr{C}_{(1)}^{p,-}(\lambda)\big| \leq \big|\mathscr{C}_{(2)}^{p,+}(\lambda)\big|\big) = \mathbb{P}(\gamma_1 \leq \gamma_2) = \mathbb{P}(\gamma_1 = \gamma_2) = 0, \tag{10.9}$$

where the last equality follows from Fact 1. Now, (10.9) contradicts (10.6), and thus (10.5) follows for $K = 1$.

For $K \geq 2$, we can use a similar argument to show that, with high probability, $\bigcup_{i\leq K} \mathscr{C}_{(i)}^{p,-}(\lambda) \subset \bigcup_{i\leq K} \mathscr{C}_{(i)}^{p,+}(\lambda)$. Now, if both $\mathscr{C}_{(1)}^{p,-}(\lambda)$ and $\mathscr{C}_{(2)}^{p,-}(\lambda)$ are contained in $\mathscr{C}_{(1)}^{p,+}(\lambda)$, then $|\mathscr{C}_{(1)}^{p,+}(\lambda)| \geq |\mathscr{C}_{(1)}^{p,-}(\lambda)| + |\mathscr{C}_{(2)}^{p,-}(\lambda)|$, which occurs with probability tending to zero. This follows using Fact 2 and $\mathbb{P}(\bar{\gamma}_1 \geq \gamma_1 + \gamma_2) = \mathbb{P}(\gamma_1 \geq \gamma_1 + \gamma_2) = 0$. Thus, $\mathscr{C}_{(2)}^{p,-}(\lambda) \subset \mathscr{C}_{(2)}^{p,+}(\lambda)$ with high probability and we can use similar arguments to conclude (10.5) for $i \leq K$.

Next, we show that, for any $K \geq 1$,

$$\lim_{n\to\infty} \mathbb{P}\big(\mathscr{C}_{(i)}^{p,-}(\lambda) \subset \mathscr{C}_{(i)}(\lambda) \subset \mathscr{C}_{(i)}^{p,+}(\lambda), \forall i \leq K\big) = 1. \tag{10.10}$$

If $\mathscr{C}_{(1)}(\lambda)$ is not contained in $\mathscr{C}_{(1)}^{p,+}(\lambda)$, then $|\mathscr{C}_{(1)}(\lambda)| \leq |\mathscr{C}_{(2)}^{p,+}(\lambda)|$. However, since $|\mathscr{C}_{(1)}^{p,-}(\lambda)| \leq |\mathscr{C}_{(1)}(\lambda)|$, it follows that $|\mathscr{C}_{(1)}^{p,-}(\lambda)| \leq |\mathscr{C}_{(2)}^{p,+}(\lambda)|$. Now, one can repeat identical argument as in (10.5) to prove that $\mathscr{C}_{(i)}(\lambda) \subset \mathscr{C}_{(i)}^{p,+}(\lambda)$ for all

$i \le K$ with high probability. Moreover, since $\mathrm{CM}_n(\boldsymbol{d}, p_n(\lambda) - \varepsilon_n) \subset \mathcal{G}_n(t_n(\lambda))$ and $\mathscr{C}_{(i)}^{p,-}(\lambda) \subset \mathscr{C}_{(i)}^{p,+}(\lambda)$ for all $i \le K$ with high probability, it must also be the case that $\mathscr{C}_{(i)}^{p,-}(\lambda) \subset \mathscr{C}_{(i)}(\lambda) \subset \mathscr{C}_{(i)}^{p,+}(\lambda)$ for all $i \le K$ with high probability. Thus we conclude (10.10).

Similarly, one can also show that

$$\lim_{n\to\infty} \mathbb{P}\big(\mathscr{C}_{(i)}^{p,-}(\lambda) \subset \mathscr{C}_{(i)}^{p}(\lambda) \subset \mathscr{C}_{(i)}^{p,+}(\lambda), \forall i \le K\big) = 1. \tag{10.11}$$

Finally, since $\mathbf{Z}_n^{p,-}(\lambda)$ and $\mathbf{Z}_n^{p,+}(\lambda)$ have the same distributional limit, it follows using (10.5) that, for all $i \le K$,

$$\big|\mathscr{C}_{(i)}^{p,+}(\lambda)\big| - \big|\mathscr{C}_{(i)}^{p,-}(\lambda)\big| = o_{\mathbb{P}}(b_n) \quad \text{and} \quad \mathrm{SP}\big(\mathscr{C}_{(i)}^{p,+}(\lambda)\big) - \mathrm{SP}\big(\mathscr{C}_{(i)}^{p,-}(\lambda)\big) \xrightarrow{\mathbb{P}} 0. \tag{10.12}$$

Thus, (10.10) and (10.11) yields

$$\big|\big|\mathscr{C}_{(i)}^{p}(\lambda)\big| - \big|\mathscr{C}_{(i)}(\lambda)\big|\big| = o_{\mathbb{P}}(b_n) \quad \text{and} \quad \big|\mathrm{SP}\big(\mathscr{C}_{(i)}^{p}(\lambda)\big) - \mathrm{SP}\big(\mathscr{C}_{(i)}(\lambda)\big)\big| \xrightarrow{\mathbb{P}} 0. \tag{10.13}$$

Moreover, $(\mathbf{Z}_n^p(\lambda))_{n\ge1}$ is tight in $\mathbb{U}_\downarrow^0$, and since both $(\mathbf{Z}_n^{p,-}(\lambda))_{n\ge1}$ and $(\mathbf{Z}_n^{p,+}(\lambda))_{n\ge1}$ are tight in $\mathbb{U}_\downarrow^0$, it also follows that $(\mathbf{Z}_n(\lambda))_{n\ge1}$ is tight in $\mathbb{U}_\downarrow^0$.

Let $\pi_k, T_k : \mathbb{U}_\downarrow^0 \mapsto \mathbb{U}_\downarrow^0$ be the functions such that for $\mathbf{z} = ((x_i, y_i))_{i\ge1}$, $\pi_k(\mathbf{z})$ consists of only (x_i, y_i) for $i \le k$ and zeroes in other coordinates, and $T_k(\mathbf{z})$ consists only of (x_i, y_i) for $i > k$. Thus,

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

For each fixed $K \ge 1$ the first term in the right hand side of (10.14) converges in probability to zero by (10.13). Also, using the tightness of both $(\mathbf{Z}_n(\lambda))_{n\ge1}$ and $(\mathbf{Z}_n^p(\lambda))_{n\ge1}$ with respect to the $\mathbb{U}_\downarrow^0$ topology, it follows that for any $\varepsilon > 0$,

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

Thus, the proof of Proposition 25 now follows. □

Let us now describe the evolution of the components under the dynamic construction Algorithm 3. We write $\mathscr{C}_{(i)}(\lambda)$ for the ith largest component of $\mathcal{G}_n(t_n(\lambda))$ and define

$$\mathcal{O}_i(\lambda) = \#\text{ open half-edges in } \mathscr{C}_{(i)}(\lambda). \tag{10.16}$$

Think of $\mathcal{O}_i(\lambda)$ as the *mass* of the component $\mathscr{C}_{(i)}(\lambda)$. Let $\mathbf{Z}_n^o(\lambda)$ denote the vector of the number of open half-edges (re-scaled by b_n) and surplus edges of $\mathcal{G}_n(t_n(\lambda))$, ordered as an element of $\mathbb{U}_\downarrow^0$. For a process $\mathbf{X}$, we will write $\mathbf{X}[\lambda_\star, \lambda^\star]$ to denote the restricted process $(X(\lambda))_{\lambda\in[\lambda_\star,\lambda^\star]}$. Let $\ell_n^o(\lambda) = \sum_{i\ge1} \mathcal{O}_i(\lambda)$. Note that

$$\ell_n^o(\lambda) = \frac{n\mu(\nu-1)}{\nu}\big(1 + o_{\mathbb{P}}(1)\big). \tag{10.17}$$

Indeed, (10.17) follows from Lemma 37 in Appendix D, and $\sum_{i\in[n]} d_i(d_i-1)/\ell_n \to \nu$. Observe that during the evolution of the graph process generated by Algorithm 3 in the time interval $[t_n(\lambda), t_n(\lambda + \mathrm{d}\lambda)]$, the ith and jth $(i > j)$ largest components merge at rate

$$2\mathcal{O}_i(\lambda)\mathcal{O}_j(\lambda) \times \frac{1}{\ell_n^o(\lambda) - 1} \times \frac{1}{2(\nu_n - 1)c_n} \approx \frac{\nu}{\mu(\nu-1)^2}\big(b_n^{-1}\mathcal{O}_i(\lambda)\big)\big(b_n^{-1}\mathcal{O}_j(\lambda)\big), \tag{10.18}$$

and create 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

$$\mathcal{O}_i(\lambda)\big(\mathcal{O}_i(\lambda) - 1\big) \times \frac{1}{\ell_n^o(\lambda) - 1} \times \frac{1}{2(\nu_n - 1)c_n} \approx \frac{\nu}{2\mu(\nu-1)^2}\big(b_n^{-1}\mathcal{O}_i(\lambda)\big)^2, \tag{10.19}$$

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 that the masses (the number of open half-edges) of the components and the system 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 [22, 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. Denote the resulting graph by $\bar{\mathcal{G}}_n(t_n(\lambda))$.

Remark 16. 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 let $\bar{\mathbf{Z}}_n^{o,\mathrm{scl}}(\lambda)$ 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 (10.18) and (10.19), 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. *Multiplicative coalescent with mass and weight*

To deduce the scaling limits involving the components sizes, let us consider a dynamic process that additionally tracks the evolution of some weights of components. Initially, the system consists of particles (possibly infinitely many) where particle i has mass x_i and weight z_i. Think of the x_i's as the number of open half-edges, and z_i's as the component sizes. Let $(X_i(t), Z_i(t))_{i\geq 1}$ denote masses, and weights at time t. We always order $(Z_i(t))_{i\geq 1}$ in a decreasing manner. 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 $Z_i(t) + Z_j(t)$.

For $\boldsymbol{x}, \boldsymbol{z} \in \ell^2_\downarrow$, we denote the vector $(Z_i(t))_{i\geq 1}$ by $\mathrm{MC}_2(\boldsymbol{x}, \boldsymbol{z}, t)$. We will need the following theorem:

Theorem 26. *Suppose that $(\boldsymbol{x}_n, \boldsymbol{z}_n) \to (\boldsymbol{x}, \boldsymbol{x})$ in $\ell^2 \times \ell^2_\downarrow$. Then, for any $t \geq 0$,*

$$\mathrm{MC}_2(\boldsymbol{x}_n, \boldsymbol{z}_n, t) \xrightarrow{d} \mathrm{MC}_2(\boldsymbol{x}, \boldsymbol{x}, t). \tag{10.20}$$

Proof. For $\boldsymbol{x}_n = (x_i^n)_{i\geq 1}$ and $\boldsymbol{z}_n = (z_i^n)_{i\geq 1}$, let $\boldsymbol{w}_n^+ = \mathrm{ord}(x_i^n \vee z_i^n)$, $\boldsymbol{w}_n^- = \mathrm{ord}(x_i^n \wedge z_i^n)$, where ord denotes the decreasing ordering of the elements. Notice that $\boldsymbol{w}_n^+ \to \boldsymbol{x}$, and $\boldsymbol{w}_n^- \to \boldsymbol{x}$ in $\ell^2_\downarrow$. Using the Feller property of the multiplicative coalescent [4, Proposition 5], it follows that

$$\mathrm{MC}_2\big(\boldsymbol{w}_n^+, \boldsymbol{w}_n^+, t\big) \xrightarrow{d} \mathrm{MC}_2(\boldsymbol{x}, \boldsymbol{x}, t) \quad \text{and} \quad \mathrm{MC}_2\big(\boldsymbol{w}_n^-, \boldsymbol{w}_n^-, t\big) \xrightarrow{d} \mathrm{MC}_2(\boldsymbol{x}, \boldsymbol{x}, t), \tag{10.21}$$

with respect to the $\ell^2_\downarrow$ topology. Suppose that $\mathrm{MC}_2(\boldsymbol{w}_n^+, \boldsymbol{w}_n^+, t)$ and $\mathrm{MC}_2(\boldsymbol{w}_n^-, \boldsymbol{w}_n^-, t)$ are coupled through the subgraph coupling (see [4, page 838]). Under the subgraph coupling, (10.21) yields

$$\big\|\mathrm{MC}_2\big(\boldsymbol{w}_n^+, \boldsymbol{w}_n^+, t\big)\big\|_2^2 - \big\|\mathrm{MC}_2\big(\boldsymbol{w}_n^-, \boldsymbol{w}_n^-, t\big)\big\|_2^2 \xrightarrow{\mathbb{P}} 0. \tag{10.22}$$

Moreover,

$$\big\|\mathrm{MC}_2\big(\boldsymbol{w}_n^-, \boldsymbol{w}_n^-, t\big)\big\|_2^2 \leq \big\|\mathrm{MC}_2(\boldsymbol{x}_n, \boldsymbol{z}_n, t)\big\|_2^2 \leq \big\|\mathrm{MC}_2\big(\boldsymbol{w}_n^+, \boldsymbol{w}_n^+, t\big)\big\|_2^2. \tag{10.23}$$

Hence, using [4, Lemma 17], under the subgraph coupling,

$$\big\|\mathrm{MC}_2\big(\boldsymbol{w}_n^+, \boldsymbol{w}_n^+, t\big) - \mathrm{MC}_2(\boldsymbol{x}_n, \boldsymbol{z}_n, t)\big\|_2^2 \leq \big\|\mathrm{MC}_2\big(\boldsymbol{w}_n^+, \boldsymbol{w}_n^+, t\big)\big\|_2^2 - \big\|\mathrm{MC}_2(\boldsymbol{x}_n, \boldsymbol{z}_n, t)\big\|_2^2 \xrightarrow{\mathbb{P}} 0, \tag{10.24}$$

and thus the proof of (10.20) follows. □

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 26 for deducing the scaling limits of the required quantities for the graph $\bar{\mathcal{G}}_n(t_n(\lambda))$. Let $\mathcal{O}_{(i)}(\lambda)$ to denote the ith largest number in the sequence $(\mathcal{O}_j(\lambda))_{j\geq 1}$. Also recall that $\mathbf{Z}_n^o(\lambda)$ denotes the vector $(\mathcal{O}_i(\lambda), \mathrm{SP}(\mathscr{C}_{(i)}(\lambda)))_{i\geq 1}$, ordered as an element of $\mathbb{U}_{\downarrow}^0$.

Lemma 27. *There exists a constant $\kappa > 0$ such that, for any $i \geq 1$,*

$$\mathcal{O}_i(\lambda) = \kappa\left|\mathscr{C}_{(i)}(\lambda)\right| + o_{\mathbb{P}}(b_n). \tag{10.25}$$

Also, for any fixed $K \geq 1$,

$$\lim_{n\to\infty} \mathbb{P}\big(\mathcal{O}_i(\lambda) = \mathcal{O}_{(i)}(\lambda), \forall i \leq K\big). \tag{10.26}$$

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

Proof. Let $\boldsymbol{d}^\lambda = (d_k^\lambda)_{k\in[n]}$ be the degree sequence of $\mathcal{G}_n(t_n(\lambda))$. Now, conditionally on $\boldsymbol{d}^\lambda$, $\mathcal{G}_n(t_n(\lambda))$ is distributed as a configuration model. We apply Theorem 21 with w_i being the number of open half-edges incident to vertex i. To this end, it is enough to show that Assumption 3 holds with high probability. Since $w_i \leq d_i$ and $d_i^\lambda \leq d_i$, Assumption 3 (ii) is obvious. Thus it is enough to show that $\sum_{i\in[n]} d_i^\lambda w_i / \sum_{i\in[n]} d_i^\lambda$ converges in probability to some constant. This can be verified using the differential equation method. See Appendix D for details. Thus, (10.25) follows from (8.3). The $\mathbb{U}_{\downarrow}^0$-tightness also follows from Theorem 21.

Next, we prove (10.26). Using (8.8a), for any $\varepsilon, \eta > 0$, there exists $M = M(\varepsilon, \eta) \geq 1$ such that

$$\mathbb{P}\bigg(\sum_{i>M} \mathcal{O}_i(\lambda)^2 > \varepsilon b_n^2\bigg) < \eta. \tag{10.27}$$

Now, if $\mathcal{O}_1(\lambda) \neq \mathcal{O}_{(1)}(\lambda)$, then there exists a $j \geq 2$ such that $\mathcal{O}_1(\lambda) \leq \mathcal{O}_j(\lambda)$. Thus,

$$\mathbb{P}\big(\mathcal{O}_1(\lambda) \neq \mathcal{O}_{(1)}(\lambda)\big) \leq \mathbb{P}\big(\exists 2 \leq j \leq M : \mathcal{O}_1(\lambda) \leq \mathcal{O}_j(\lambda)\big) + \mathbb{P}\big(\exists j > M : \mathcal{O}_1(\lambda) \leq \mathcal{O}_j(\lambda)\big). \tag{10.28}$$

Using (10.25) and Theorem 4 together with Proposition 25, it follows that $(b_n^{-1}\mathcal{O}_i(\lambda))_{i\leq M}$ converge in distribution to $(\gamma_i)_{i\in[M]}$, where γ_i denotes the ith largest excursion length of a process of the form (1.5). An application of Fact 1 yields

$$\lim_{n\to\infty} \mathbb{P}\big(\exists 2 \leq j \leq M : \mathcal{O}_1(\lambda) \leq \mathcal{O}_j(\lambda)\big) \leq \mathbb{P}(\gamma_1 \leq \gamma_2) = \mathbb{P}(\gamma_1 = \gamma_2) = 0. \tag{10.29}$$

Moreover,

$$\begin{aligned}\mathbb{P}\big(\exists j > M : \mathcal{O}_1(\lambda) \leq \mathcal{O}_j(\lambda)\big) &\leq \mathbb{P}\bigg(b_n^{-2}\sum_{i>M} \mathcal{O}_i(\lambda)^2 > b_n^{-2}\mathcal{O}_1(\lambda)^2\bigg) \\ &\leq \mathbb{P}\big(b_n^{-1}\mathcal{O}_1(\lambda) \leq \varepsilon\big) + \mathbb{P}\bigg(\sum_{i>M} \mathcal{O}_i(\lambda)^2 > \varepsilon b_n^2\bigg).\end{aligned} \tag{10.30}$$

The limsup as $n \to \infty$ of the first term is at most $\mathbb{P}(\gamma_1 \leq \varepsilon)$. Since the distribution of γ_1 does not have an atom at zero, $\mathbb{P}(\gamma_1 \leq \varepsilon)$ can be taken to be at most η by choosing $\varepsilon > 0$ sufficiently small. Combining (10.29) and (10.30) yields that $\mathbb{P}(\mathcal{O}_1(\lambda) \neq \mathcal{O}_{(1)}(\lambda)) \leq 2\eta$ for all sufficiently large n. Similarly,

$$\mathbb{P}\big(\mathcal{O}_2(\lambda) \neq \mathcal{O}_{(2)}(\lambda) \text{ and } \mathcal{O}_1(\lambda) = \mathcal{O}_{(1)}(\lambda)\big) \leq \mathbb{P}\big(\exists j \geq 3 : \mathcal{O}_2(\lambda) \leq \mathcal{O}_j(\lambda)\big) + \mathbb{P}\big(\mathcal{O}_2(\lambda) = \mathcal{O}_1(\lambda)\big). \tag{10.31}$$

Both of these terms can be bounded using similar arguments as above. The proof for bounding $\mathbb{P}(\mathcal{O}_i(\lambda) \neq \mathcal{O}_{(i)}(\lambda))$ for $i \geq 2$ is similar. $\square$

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 27 states that, for each fixed λ, $\mathbf{Z}_n^o(\lambda)$ is close to $\kappa\mathbf{Z}_n(\lambda)$. The following result states that formally:

Corollary 28. *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. Using Lemma 27, the proof follows using identical arguments as shown after (10.13) in the proof of Proposition 25. □

10.3. *Comparison between the dynamic construction and the modified process*

The goal of this section is to prove that for large components, the modified construction does not change the number of open half-edges, component sizes and surplus edges considerably. To this end, recall Algorithm 4 and the associated notation. Thus $\bar{\mathscr{C}}_{(i)}(\lambda)$ is the ith largest component of $\bar{\mathcal{G}}_n(t_n(\lambda))$, and $\bar{\mathcal{O}}_i(\lambda)$ is the number of open half-edges in $\bar{\mathscr{C}}_{(i)}(\lambda)$. Note that we stop discarding the open half-edges in Algorithm 4, so that $\bar{\mathcal{O}}_i(\lambda)$ counts the number of open half-edges associated to vertices at time $t_n(\lambda_\star)$. Also, $\bar{\mathcal{O}}_{(i)}(\lambda)$ and $\mathcal{O}_{(i)}(\lambda)$ denote the ith largest numbers in the sequences $(\bar{\mathcal{O}}_j(\lambda))_{j\geq 1}$ and $(\mathcal{O}_j(\lambda))_{j\geq 1}$ respectively.

Proposition 29. *For each fixed* $i \geq 1$, *as* $n \to \infty$,

(a) $\bar{\mathcal{O}}_{(i)}(\lambda) - \mathcal{O}_i(\lambda) = o_{\mathbb{P}}(b_n)$,

(b) $\mathrm{SP}(\bar{\mathscr{C}}_{(i)}(\lambda)) - \mathrm{SP}(\mathscr{C}_{(i)}(\lambda)) \xrightarrow{\mathbb{P}} 0$.

Comparison of component sizes

We start by showing that the component sizes and open half-edges in $\mathcal{G}_n(t_n(\lambda))$ and $\bar{\mathcal{G}}_n(t_n(\lambda))$ have identical distributional limit:

Lemma 30. *For any* $\lambda > \lambda_\star$, *the sequences* $(b_n^{-1}\bar{\mathcal{O}}_{(i)}(\lambda))_{i\geq 1}$ *and* $(b_n^{-1}|\bar{\mathscr{C}}_{(i)}(\lambda)|)_{i\geq 1}$ *converges in distribution with respect to the* $\ell^2_\downarrow$*-topology, and the distributional limits are identical to those of* $(b_n^{-1}\mathcal{O}_i(\lambda))_{i\geq 1}$ *and* $(b_n^{-1}|\mathscr{C}_{(i)}(\lambda)|)_{i\geq 1}$ *respectively.*

Proof. Let $\boldsymbol{\xi}(\boldsymbol{\theta}, \lambda)$ denote the ordered vector of excursion lengths of the process (1.5) by putting $\mu = 1$. Recall the discussion after Algorithm 4 that, if $\beta_n = (\bar{s}_1(\nu_n - 1)c_n)^{1/2} = cb_n(1 + o_{\mathbb{P}}(1))$, then $((\beta_n^{-1}\bar{\mathcal{O}}_{(i)}(\lambda))_{i\geq 1})_{\lambda > \lambda_\star}$ evolves exactly as a standard multiplicative coalescent.

Now, using Lemma 27, together with Proposition 24 and the scaling limit from Theorem 4, we know that

$$\left(\beta_n^{-1}\left|\mathcal{O}_i(\lambda)\right|\right)_{i\geq 1} \xrightarrow{d} \boldsymbol{\xi}(c_1\boldsymbol{\theta}, c_2\lambda), \tag{10.32}$$

with respect to the ℓ^2-topology, for some constants $c_1, c_2 > 0$. Here we have used the fact from Lemma 27 that $\mathcal{O}_i(\lambda) = \mathcal{O}_{(i)}(\lambda)$ with high probability for each fixed $i \geq 1$. Also, to adjust a multiplicative $\sqrt{\nu}$ factor in the scaling limit in Theorem 4, we are using the fact that, for $\eta_1, \eta_2 > 0$, $\boldsymbol{\theta} \in \ell^3_\downarrow \setminus \ell^2_\downarrow$ and $\lambda \in \mathbb{R}$, $\boldsymbol{\xi}(\eta_1\boldsymbol{\theta}, \eta_2\lambda) \stackrel{d}{=} \frac{1}{\eta_1}\boldsymbol{\xi}(\boldsymbol{\theta}, \frac{\eta_2}{\eta_1^2}\lambda)$. Thus, in particular, $(\beta_n^{-1}|\mathcal{O}_i(\lambda_\star)|)_{i\geq 1} \xrightarrow{d} \boldsymbol{\xi}(c_1\boldsymbol{\theta}, c_2\lambda_\star)$. Using [5, Theorem 2], there exists a version of the standard multiplicative coalescent $(\mathrm{MC}_1(\lambda))_{-\infty<\lambda<\infty}$ process such that $\mathrm{MC}_1(\lambda)$ has the same distribution as $\boldsymbol{\xi}(c_1\boldsymbol{\theta}, c_2\lambda)$. Using the Feller property of the multiplicative coalescent [4, Proposition 5], it follows that

$$\left(\beta_n^{-1}\left|\bar{\mathcal{O}}_{(i)}(\lambda)\right|\right)_{i\geq 1} \xrightarrow{d} \boldsymbol{\xi}(c_1\boldsymbol{\theta}, c_2\lambda), \tag{10.33}$$

with respect to the $\ell^2_\downarrow$-topology. For the component sizes, one can use (10.20) and (10.25) to conclude the proof. □

Next we show that with high probability the largest components of $\mathcal{G}_n(t_n(\lambda))$ contain those of $\bar{\mathcal{G}}_n(t_n(\lambda))$:

Lemma 31. *For any* $K \geq 1$,

$$\lim_{n\to\infty} \mathbb{P}\left(\mathscr{C}_{(i)}(\lambda) \subset \bar{\mathscr{C}}_{(i)}(\lambda), \forall i \leq K\right) = 1. \tag{10.34}$$

Proof. Under the coupling in described in Remark 16, and using Lemma 30, the proof is identical to the proof of (10.5). We omit further details. □

Proof of Proposition 29(a). Let $\mathcal{A}_K$ denote the event in (10.34). On $\mathcal{A}_K$, $\mathcal{O}_1(\lambda) \leq \bar{\mathcal{O}}_1(\lambda) \leq \bar{\mathcal{O}}_{(1)}(\lambda)$. Using the fact from Lemma 30 that $b_n^{-1}\bar{\mathcal{O}}_{(1)}(\lambda)$ and $b_n^{-1}\mathcal{O}_1(\lambda)$ have the same scaling limit, it follows that $\bar{\mathcal{O}}_{(1)}(\lambda) - \mathcal{O}_1(\lambda) = o_{\mathbb{P}}(b_n)$. Thus, the proof for $i = 1$ follows by applying Lemma 31. For $i \geq 2$, note that on $\mathcal{A}_K$, $\sum_{j\leq i}\mathcal{O}_j(\lambda) \leq \sum_{j\leq i}\bar{\mathcal{O}}_j(\lambda) \leq \sum_{j\leq i}\bar{\mathcal{O}}_{(j)}(\lambda)$. Using this relation the proof can be completed inductively as before by applying Lemmas 30 and 31. □

Comparison of surplus edges

Next, we compare the surplus edges. Using Lemma 31, it follows that $\mathrm{SP}(\mathscr{C}_{(i)}(\lambda)) \le \mathrm{SP}(\bar{\mathscr{C}}_{(i)}(\lambda))$ with high probability for each fixed $i \ge 1$. Let $B_{\mathrm{SP}}(\lambda)$ denote the total number of surplus edges created during the formation of some bad edges. Then, with high probability,

$$\mathrm{SP}\big(\bar{\mathscr{C}}_{(i)}(\lambda)\big) \le \mathrm{SP}\big(\mathscr{C}_{(i)}(\lambda)\big) + B_{\mathrm{SP}}(\lambda) + \sum_{j \ne i: \mathscr{C}_{(j)}(\lambda) \subset \bar{\mathscr{C}}_{(i)}(\lambda)} \mathrm{SP}\big(\mathscr{C}_{(j)}(\lambda)\big). \tag{10.35}$$

Thus, it is enough to show that the last two terms converge in probability to zero. We first bound $B_{\mathrm{SP}}(\lambda)$:

Lemma 32. *For any $\lambda \ge \lambda_\star$, $B_{\mathrm{SP}}(\lambda) \xrightarrow{\mathbb{P}} 0$.*

Proof. Before going into the proof, recall Algorithms 3 and 4 and all the definitions therein. We write $C > 0$ for a generic constant whose value can be different in different lines. Firstly, let $\mathcal{B}_i(\lambda)$ denote the number of open half-edges in $\bar{\mathscr{C}}_{(i)}(\lambda)$ that caused the creation of at least one edge in $[\lambda_\star, \lambda)$. We refer to these as bad half-edges since pairing one of them again would give rise to a bad edge. Let $\boldsymbol{\mathcal{B}}(\lambda)$ be the corresponding vector obtained by ordering them as an element of $\ell^2_\downarrow$. We will show that

$$\sup_{\lambda' \in [\lambda_\star, \lambda]} b_n^{-1} \big\| \boldsymbol{\mathcal{B}}\big(\lambda'\big) \big\|_2 = b_n^{-1} \big\| \boldsymbol{\mathcal{B}}(\lambda) \big\|_2 \xrightarrow{\mathbb{P}} 0. \tag{10.36}$$

For any fixed $K \ge 1$, define the event $\mathcal{A}_K$ that $\mathscr{C}_{(i)}(\lambda) \subset \bar{\mathscr{C}}_{(i)}(\lambda)$ for all $i \le K$. Then $\mathbb{P}(\mathcal{A}_K) \to 1$, by Lemma 31. On the event $\mathcal{A}_K$, one can bound $\mathcal{B}_i(\lambda) \le \bar{\mathcal{O}}_i(\lambda) - \mathcal{O}_i(\lambda)$ for all $i \le K$ and $\mathcal{B}_i(\lambda) \le \bar{\mathcal{O}}_i(\lambda)$ for all $i > K$. Further, on $\mathcal{A}_K$, $\sum_{j \le i} \mathcal{O}_j(\lambda) \le \sum_{j \le i} \bar{\mathcal{O}}_j(\lambda) \le \sum_{j \le i} \bar{\mathcal{O}}_{(j)}(\lambda)$ for all $i \le K$. Now, by Lemma 30, $(b_n^{-1} \bar{\mathcal{O}}_{(i)}(\lambda))_{i \ge 1}$ and $(b_n^{-1} \mathcal{O}_i(\lambda))_{i \ge 1}$ have the same distributional limit. Therefore, $b_n^{-1} \mathcal{O}_j(\lambda)$ and $b_n^{-1} \bar{\mathcal{O}}_j(\lambda)$ have the same distributional limits for all $j \le K$. Moreover, since $(b_n^{-1} \bar{\mathcal{O}}_{(i)}(\lambda))_{i \ge 1}$ converges in $\ell^2_\downarrow$, we can choose K sufficiently large so that $\sum_{i > K} \bar{\mathcal{O}}_i(\lambda)^2 < \varepsilon b_n^2$ with probability at most ε, for any $\varepsilon > 0$. Thus, (10.36) follows.

For any semi-martingale $(Y_t)_{t \ge 0}$, we write $\mathrm{D}(Y_t)$ for its compensator. Let us now turn to the asymptotics of $B_{\mathrm{SP}}(\lambda)$. Note that $B_{\mathrm{SP}}(\lambda)$ has jump size at most 1. Moreover, a surplus is bad if at least one of the end-points is bad. Therefore,

$$\mathrm{D}\big(B_{\mathrm{SP}}(\lambda)\big) \le C \int_{\lambda_\star}^{\lambda} \frac{\sum_{i \ge 1} \bar{\mathcal{O}}_i(\lambda') \mathcal{B}_i(\lambda')}{\bar{s}_1 c_n} \,\mathrm{d}\lambda'. \tag{10.37}$$

The Cauchy–Schwarz inequality together with Theorem 26 and (10.36) shows that $\mathrm{D}(B_{\mathrm{SP}}(\lambda)) = o_{\mathbb{P}}(1)$. We next use Lenglart inequality [36] which says that, for any non-negative process $(X_t)_{t \ge 0}$ with $X_0 = 0$, and $\varepsilon, \delta > 0$,

$$\mathbb{P}\Big(\sup_{s \in [0,t]} X_s > \varepsilon\Big) \le \frac{\mathbb{E}[\min\{\mathrm{D}(X_t), \delta\}]}{\varepsilon} + \mathbb{P}\big(\mathrm{D}(X_t) > \varepsilon\big). \tag{10.38}$$

Finally, using (10.38) together with $\mathrm{D}(B_{\mathrm{SP}}(\lambda)) = o_{\mathbb{P}}(1)$, implies the required statement for bad surplus edges. □

Next, we bound the final term in (10.35). 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 get connected to one of the blue components during the evolution. Let $\mathcal{C}_M(\lambda)$, $\bar{\mathcal{C}}_M(\lambda)$ denote the union of all such blue components in $\mathcal{G}_n(t_n(\lambda))$ and $\bar{\mathcal{G}}_n(t_n(\lambda))$. We track the surplus edges in $\bar{\mathcal{C}}_M(\lambda)$ that were created when a bad edge caused some component with a surplus edge to merge with a component in $\bar{\mathcal{C}}_M(\lambda)$. Let $F_M(\lambda)$ denote the number of times when such surplus edges were created upto time λ (note that $F_M(\lambda)$ does not count the total number of these unwanted surplus edges). We show that $F_M(\lambda)$ is zero with high probability.

Lemma 33. *For any $\lambda \ge \lambda_\star$ and $M \ge 1$, $F_M(\lambda) \xrightarrow{\mathbb{P}} 0$.*

Proof. The argument is similar to Lemma 32. Let $\mathcal{I}_M := \{i : \bar{\mathscr{C}}_{(i)}(\lambda) \subset \bar{\mathcal{C}}_M(\lambda)\}$. Then, the rate at which $F_M(\lambda)$ increases by 1 is given by

$$\mathrm{D}\big(F_M(\lambda)\big) \leq C \int_{\lambda_\star}^{\lambda} \bigg(\frac{\sum_{i\in\mathcal{I}_M} \mathcal{B}_i(\lambda') \sum_{j\notin\mathcal{I}_M:\mathrm{SP}(\bar{\mathscr{C}}_{(j)}(\lambda'))\geq 1} \bar{\mathcal{O}}_j(\lambda')}{\bar{s}_1 c_n} + \frac{\sum_{i\in\mathcal{I}_M} \bar{\mathcal{O}}_i(\lambda') \sum_{j\notin\mathcal{I}_M:\mathrm{SP}(\bar{\mathscr{C}}_{(j)}(\lambda'))\geq 1} \mathcal{B}_j(\lambda')}{\bar{s}_1 c_n} \bigg) \mathrm{d}\lambda'. \tag{10.39}$$

Let us write the two terms above by (I) and (II) respectively. Then, using the fact that $|\mathcal{I}_M| \leq M$,

$$(I) \leq \frac{C b_n^2}{\bar{s}_1 c_n} \int_{\lambda_\star}^{\lambda} b_n^{-2} \sum_{i\leq M} \mathcal{B}_i(\lambda') \times \sum_{j\geq 1} \bar{\mathcal{O}}_j(\lambda') \,\mathrm{SP}\big(\bar{\mathscr{C}}_{(j)}(\lambda')\big)\,\mathrm{d}\lambda' \xrightarrow{\mathbb{P}} 0, \tag{10.40}$$

where the last step follows using (10.36), $\bar{s}_1 c_n = O_{\mathbb{P}}(b_n^2)$, and the $\mathbb{U}_{\downarrow}^0$ tightness of $(\bar{\mathbf{Z}}_n^o(\lambda'))_{n\geq 1}$. The last statement is a consequence of the near-Feller property of the augmented multiplicative coalescent [10, Theorem 3.1]. Similarly,

$$(II) \leq \frac{C b_n^2}{\bar{s}_1 c_n} \int_{\lambda_\star}^{\lambda} b_n^{-2} \sum_{i\leq M} \bar{\mathcal{O}}_i(\lambda') \times \sum_{j\geq 1} \mathcal{B}_i(\lambda') \,\mathrm{SP}\big(\bar{\mathscr{C}}_{(j)}(\lambda')\big)\,\mathrm{d}\lambda' \xrightarrow{\mathbb{P}} 0, \tag{10.41}$$

where we have used that $b_n^{-1} \sum_{j\geq 1} \mathcal{B}_i(\lambda')\,\mathrm{SP}(\bar{\mathscr{C}}_{(j)}(\lambda')) \xrightarrow{\mathbb{P}} 0$, which can be proved using identical arguments as in the proof of (10.36) and the $\mathbb{U}_{\downarrow}^0$ tightness of $(\bar{\mathbf{Z}}_n^o(\lambda'))_{n\geq 1}$. Now we conclude the proof using Lenglart's inequality similarly as in the proof of Lemma 32. □

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

Lemma 34. *Fix any $\lambda > \lambda_\star$. For any $\varepsilon > 0$, and $K \geq 1$, there exists $M = M(\varepsilon, K, \lambda)$ such that*

$$\limsup_{n\to\infty} \mathbb{P}\big(\bar{\mathscr{C}}_{(1)}(\lambda), \dots, \bar{\mathscr{C}}_{(K)}(\lambda) \textit{ are not contained in } \bar{\mathcal{C}}_M(\lambda)\big) \leq \varepsilon. \tag{10.42}$$

Proof. Recall that $\mathcal{I}_M = \{i : \bar{\mathscr{C}}_{(i)}(\lambda) \subset \bar{\mathcal{C}}_M(\lambda)\}$. If $\bar{\mathscr{C}}_{(i_0)}(\lambda)$ is not contained in $\bar{\mathcal{C}}_M(\lambda)$ for some $i_0 \leq K$, then that would imply that $\sum_{i\notin\mathcal{I}_M} |\bar{\mathscr{C}}_{(i)}(\lambda)|^2 \geq |\bar{\mathscr{C}}_{(i_0)}(\lambda)|^2$ is not negligible. Thus, it is enough to show that, for any $\varepsilon > 0$, there exists M such that

$$\limsup_{n\to\infty} \mathbb{P}\bigg(\sum_{i\notin\mathcal{I}_M} \big|\bar{\mathscr{C}}_{(i)}(\lambda)\big|^2 > \varepsilon b_n^2\bigg) \leq \varepsilon. \tag{10.43}$$

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 the 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 such 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 $\bigcup_{i=1}^M \bar{\mathscr{C}}_{(i)}(\lambda_\star)$. Under this coupling,

$$\sum_{i\notin\mathcal{I}_M} \big|\bar{\mathscr{C}}_{(i)}(\lambda)\big|^2 \leq \sum_{i\geq 1} \big|\bar{\mathscr{C}}_{(i)}^{>M}(\lambda)\big|^2. \tag{10.44}$$

Now, $(\bar{\mathbf{Z}}_n(\lambda_\star))_{n\geq 1}$ is tight in $\mathbb{U}_{\downarrow}^0$. Thus, the $\ell_{\downarrow}^2$-norm of $(b_n^{-1}|\bar{\mathscr{C}}_{(i)}(\lambda_\star)|)_{i>M}$ can be made aritrarily small. Therefore, using the Feller property of the multiplicative coalescent process,

$$\lim_{M\to\infty} \limsup_{n\to\infty} \mathbb{P}\bigg(\sum_{i\geq 1} \big|\bar{\mathscr{C}}_{(i)}^{>M}(\lambda)\big|^2 > \varepsilon b_n^2\bigg) = 0, \tag{10.45}$$

the proof of (10.43) now follows using (10.44). □

Proof of Proposition 29(b). Recall the expression in the right hand side of (10.35). The second term goes to zero in probability using Lemma 32 and the third term goes to zero in probability using Lemmas 33 and 34. Thus the proof follows. □

10.4. *Proof of Theorem 5*

We now have all the ingredients to complete the proof of Theorem 5. Take $\lambda_\star = \lambda_1$.

Recall that, for an element $\mathbf{z} = (x_i, y_i)_{i\geq 1} \in \mathbb{U}^0_\downarrow$ and a constant $c > 0$, we denote $c\mathbf{z} = (cx_i, y_i)_{i\geq 1}$. Recall the definition of $\mathbf{Z}_n^o(\lambda)$ after (10.16). Denote $\beta_n = (\bar{s}_1(\nu_n - 1)c_n)^{1/2}$ and let $\mathbf{Z}_n^{o,\mathrm{scl}}(\lambda) = \beta_n^{-1} b_n \mathbf{Z}_n^o(\lambda)$. We define $\bar{\mathbf{Z}}_n^o(\lambda)$ and $\bar{\mathbf{Z}}_n^{o,\mathrm{scl}}(\lambda)$ analogously for the modified graph $\bar{\mathcal{G}}_n(t_n(\lambda))$. Thus, for any $\lambda \in \mathbb{R}$,

$$\bar{\mathbf{Z}}_n^{o,\mathrm{scl}}(\lambda) \xrightarrow{d} c\kappa \mathbf{Z}(\lambda), \tag{10.46}$$

where $b_n\beta_n^{-1} \xrightarrow{\mathbb{P}} c$ (see Lemma 37), κ is as in Corollary 28, and $\mathbf{Z}(\lambda)$ is given by Theorem 4. (10.46) follows by applying Theorem 4, Corollary 28, Proposition 29, together with Proposition 24 and (10.26).

Next, we had argued after Remark 16 that $(\bar{\mathbf{Z}}_n^{o,\mathrm{scl}}(\lambda))_{\lambda\in[\lambda_\star,\lambda^\star]}$ evolves as a standard augmented multiplicative coalescent. Let $(\mathcal{T}_\lambda)_{\lambda\in\mathbb{R}}$ denote the semigroup associated to the standard augmented multiplicative coalescent process. Using the near-Feller property for the augmented multiplicative coalescent [10, Theorem 3.1], it follows from (10.46) that

$$\big(\bar{\mathbf{Z}}_n^{o,\mathrm{scl}}(\lambda_1), \bar{\mathbf{Z}}_n^{o,\mathrm{scl}}(\lambda_2)\big) \xrightarrow{d} \big(c\kappa\mathbf{Z}(\lambda_1), \mathcal{T}_{\lambda_2-\lambda_1}\big(c\kappa\mathbf{Z}(\lambda_1)\big)\big). \tag{10.47}$$

Define $\mathcal{T}'_\lambda = \mathcal{T}_{\lambda/(c\kappa)^2}$. Applying Proposition 29, it follows that

$$\big(\mathbf{Z}_n^o(\lambda_1), \mathbf{Z}_n^o(\lambda_2)\big) \xrightarrow{d} \big(\kappa\mathbf{Z}(\lambda_1), \kappa\mathcal{T}'_{\lambda_2-\lambda_1}\big(\mathbf{Z}(\lambda_1)\big)\big). \tag{10.48}$$

Corollary 28 yields that

$$\big(\mathbf{Z}_n(\lambda_1), \mathbf{Z}_n(\lambda_2)\big) \xrightarrow{d} \big(\mathbf{Z}(\lambda_1), \mathcal{T}'_{\lambda_2-\lambda_1}\big(\mathbf{Z}(\lambda_1)\big)\big). \tag{10.49}$$

Repeating the above argument inductively, there exists a version of the augmented multiplicative coalescent $\mathbf{AMC} = (\mathrm{AMC}(\lambda))_{\lambda\in\mathbb{R}}$, with the semigroup given by $(\mathcal{T}'_\lambda)_{\lambda\in\mathbb{R}}$, such that for any $k \geq 1$

$$\big(\mathbf{Z}_n(\lambda_1), \ldots, \mathbf{Z}_n(\lambda_k)\big) \xrightarrow{d} \big(\mathrm{AMC}(\lambda_1), \ldots, \mathrm{AMC}(\lambda_k)\big). \tag{10.50}$$

Moreover, the scaling limit in Theorem 2 shows that $\mathcal{T}'_{\lambda-\lambda_1}(\mathbf{Z}(\lambda_1))$ and $\mathbf{Z}(\lambda)$ have the same distribution. This ensures that there exists a version of the augmented multiplicative coalescent whose distribution at each fixed time λ is identical to $\mathbf{Z}(\lambda)$. Finally, the proof of Theorem 5 is completed by using Proposition 25.

Appendix A: Path counting

In this section, we derive a generalization of [30, Lemma 5.1] by extending the argument therein. Let V_n' denote the vertex chosen according to the distribution G_n on $[n]$, independently of the graph. We will later take G_n to be the uniform distribution on $[n]$, and the size-biased distribution with the sizes being proportional to the degrees. 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 35. *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,*

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

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,

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

Now, using the exactly same arguments as [30, Lemma 5.1], it follows that

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

and this completes the proof. □

Appendix B: Proof of Lemma 20

Proof. The proof is an adaptation of the proof of [22, Lemma 20]. Let V_n' denote the vertex chosen according to the distribution G_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

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

Note that

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

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

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

Thus,

$$\mathbb{P}\big(S_n(\gamma) \geq H\big) \leq \frac{\mathbb{E}[D_n']}{H}. \tag{B.4}$$

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

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

Now,

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

where

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

Therefore, using induction, (B.5) yields

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

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 $(K-1)!$, as well as $2\mathrm{e}\sqrt{\delta} < 1$ in the last step. Since $\lambda < 0$, we can use Lemma 35 to conclude that for all sufficiently large n

$$\mathbb{E}\big[\big|\mathscr{C}(V_n')\big|\big] \leq Cc_n, \tag{B.9}$$

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

Appendix C: Non-atomic distribution for the excursion process

Below we prove the scaling limit of the exploration process (4.1) has a density at each fixed time. This was proved in [15, Lemma 3.5] for $\theta_i = i^{-\alpha}$. However, the argument does not generalize directly for general $\boldsymbol{\theta} \in \ell^3_\downarrow \setminus \ell^2_\downarrow$. Below we give a proof for the general case $\boldsymbol{\theta}$:

Lemma 36. *Let $S(t) = \sum_{i=1}^\infty \theta_i(\mathcal{I}_i(t) - \theta_i t)$, where $\xi_i \sim \mathrm{Exp}(\theta_i)$ independently, and $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots) \in \ell^3_\downarrow \setminus \ell^2_\downarrow$. Then the distribution of S_t has no atoms for all $t > 0$.*

Proof. Let $\phi_t(v) = \mathbb{E}[\mathrm{e}^{\mathrm{i}vS(t)}]$ for $v \in \mathbb{R}$. Then,

$$\phi_t(v) = \prod_{j=1}^\infty \mathrm{e}^{-\mathrm{i}v\theta_j^2 t}\big(1 + \big(\mathrm{e}^{\mathrm{i}v\theta_j} - 1\big)\big(1 - \mathrm{e}^{-\theta_j t}\big)\big). \tag{C.1}$$

Therefore,

$$\begin{aligned} \big|\phi_t(v)\big|^2 &\leq \prod_{j=1}^\infty \big|1 + \big(\cos(v\theta_j) - 1\big)\big(1 - \mathrm{e}^{-t\theta_j}\big) + \mathrm{i}\big(1 - \mathrm{e}^{-t\theta_j}\big)\sin(v\theta_j)\big|^2 \\ &= \prod_{j=1}^\infty \big(\big(\cos(v\theta_j) + \mathrm{e}^{-t\theta_j}\big(1 - \cos(v\theta_j)\big)\big)^2 + \big(1 - \mathrm{e}^{-t\theta_j}\big)^2\sin^2(v\theta_j)\big) \\ &= \prod_{j=1}^\infty \big(\mathrm{e}^{-2t\theta_j}\big(1 - 2\cos(v\theta_j)\big) + 2\mathrm{e}^{-t\theta_j}\cos(v\theta_j) + \big(1 - \mathrm{e}^{-t\theta_j}\big)^2\big) \\ &= \prod_{j=1}^\infty \big(1 - 2\mathrm{e}^{-t\theta_j}\big(1 - \mathrm{e}^{-t\theta_j}\big)\big(1 - \cos(v\theta_j)\big)\big) \\ &\leq \mathrm{e}^{-\sum_{j=1}^\infty 2\mathrm{e}^{-t\theta_j}(1-\mathrm{e}^{-t\theta_j})(1-\cos(v\theta_j))}, \end{aligned} \tag{C.2}$$

where in the last step we have used the fact that $1 - x \leq \mathrm{e}^{-x}$ for all $x > 0$. Let $j_0(v,t) \geq 1$ be such that $\max\{|v|\theta_j, t\theta_j\} \leq 1$ for all $j \geq j_0(v,t)$. Now, for $j \geq j_0(v,t)$, we have that $\mathrm{e}^{-t\theta_j} \geq \mathrm{e}^{-1}$, $(1 - \mathrm{e}^{-t\theta_j}) \geq t\theta_j/2$ and $1 - \cos(v\theta_j) \geq \frac{2}{\pi}v^2\theta_j^2$. Thus,

$$\big|\phi_t(v)\big| \leq \mathrm{e}^{-\frac{2t}{\mathrm{e}\pi}v^2\sum_{j\geq j_0(v,t)}\theta_j^3}. \tag{C.3}$$

Let $j_v = j_0(v,t)$ and $g(|v|) := \sum_{j \ge j_v} \theta_j^3$. Since t is fixed, we have ignored t in the notation here. We need the following facts about g:

Fact 3. The function $g : [0,\infty) \mapsto (0,\infty)$ is non-increasing and left continuous. Further, $\int_0^\infty g(v)\,\mathrm{d}v = \infty$.

Proof. The non-increasingness and left continuity follow easily from the definition. Let $C_0 = \sum_{j=1}^\infty \theta_j^3$ and $f(x) = C_0^{-1/3}\theta_j$ for $j-1 \le x < j$ and $f(x) = 0$ for $x < 0$. Let X be a random variable with density function f^3. Then

$$\int_0^\infty \mathbb{P}\left(\frac{1}{f(X)C_0^{1/3}} \ge v\right) \mathrm{d}v = \mathbb{E}\left[\frac{1}{f(X)C_0^{1/3}}\right] = C_0^{-1/3} \int_0^\infty f^2(x)\,\mathrm{d}x = C_0^{-1} \sum_{j=1}^\infty \theta_j^2 = \infty. \tag{C.4}$$

Moreover,

$$\int_0^\infty \mathbb{P}\left(\frac{1}{f(X)C_0^{1/3}} \ge v\right) \mathrm{d}v = \sum_{j=1}^\infty C_0^{-1}\theta_j^3 \mathbb{1}_{\{\frac{1}{\theta_j} \ge v\}} = C_0^{-1} g(v), \tag{C.5}$$

for $v \ge t$. Thus, $\int_0^\infty g(v)\,\mathrm{d}v = \infty$. □

Fact 4. For every $M > 0$, there exists $T_n = T_n(M) \nearrow \infty$ such that

$$\lim_{n\to\infty} \frac{1}{T_n} \Lambda\left(\left\{0 \le u \le T_n : u^2 g(u) \le M\right\}\right) = 0, \tag{C.6}$$

where Λ denotes the Lebesgue measure.

Proof. Assume the contrary, i.e., there exists $\varepsilon > 0$ and $T_0 > 0$ such that

$$\frac{1}{T} \Lambda\left(\left\{0 \le u \le T : u^2 g(u) \le M\right\}\right) \ge \varepsilon, \quad \forall T \ge T_0. \tag{C.7}$$

Thus, for any $T \ge T_0$,

$$u_T := \sup\left\{0 \le u \le T : u^2 g(u) \le M\right\} \ge \varepsilon T. \tag{C.8}$$

Since g is left continuous, $u_T^2 g(u_T) \le M$ and thus

$$T^2 g(T) \le T^2 g(u_T) \le \frac{T^2 M}{u_T^2} \le \frac{M}{\varepsilon^2}, \tag{C.9}$$

where in the first step, we have used that g is non-decreasing from Fact 3. Thus $g(T) \le M/(\varepsilon T)^2$ for all $T \ge T_0$ and therefore $\int_0^\infty g(v)\,\mathrm{d}v < \infty$, which contradicts Fact 3. □

Continuing to the proof of Lemma 36, note that for any $a \in \mathbb{R}$ and $M > 0$, and T_n chosen according to Fact 4,

$$\begin{aligned}
\mathbb{P}\big(S(t) = a\big) &= \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} \mathrm{e}^{-\mathrm{i}va} \phi_t(v)\,\mathrm{d}v = \lim_{n\to\infty} \frac{1}{2T_n} \int_{-T_n}^{T_n} \mathrm{e}^{-\mathrm{i}va} \phi_t(v)\,\mathrm{d}v \\
&\le \limsup_{n\to\infty} \frac{1}{2T_n} \int_{-T_n}^{T_n} \mathrm{e}^{-\frac{2t}{\mathrm{e}\pi} v^2 g(|v|)}\,\mathrm{d}v = \limsup_{n\to\infty} \frac{1}{T_n} \int_0^{T_n} \mathrm{e}^{-\frac{2t}{\mathrm{e}\pi} v^2 g(v)}\,\mathrm{d}v \\
&\le \limsup_{n\to\infty} \frac{1}{T_n} \Lambda\left(\left\{0 \le u \le T_n : u^2 g(u) \le M\right\}\right) + \mathrm{e}^{-\frac{2t}{\mathrm{e}\pi} M} = \mathrm{e}^{-\frac{2t}{\mathrm{e}\pi} M},
\end{aligned} \tag{C.10}$$

where in the third step we have used (C.3), and the last-but-one step follows from Fact 4. Since $M > 0$ is arbitrary, $\mathbb{P}(S(t) = a) = 0$ and the proof follows. □

Appendix D: Evolution of moments in the dynamic construction

Let $w_i(t)$ denote the number of unpaired/open half-edges incident to vertex i at time t in Algorithm 3. Let $s_1(t)$ denote the total number of unpaired half-edges at time t. Denote also $s_2(t) = \sum_{i\in[n]} w_i(t)^2$, $s_{d,w}(t) = \sum_{i\in[n]} d_i w_i(t)$. Further, we write $\mu_n = \ell_n/n$.

Lemma 37. *Under Assumption 1, the quantities* $\sup_{t\le T} |\frac{1}{n}s_1(t) - \mu_n \mathrm{e}^{-2t}|$, $\sup_{t\le T} |\frac{1}{n}s_2(t) - \mu_n \mathrm{e}^{-4t}(\nu_n + \mathrm{e}^{2t})|$, $\sup_{t\le T} |\frac{1}{n}s_{d,w}(t) - \mu_n(1+\nu_n)\mathrm{e}^{-2t}|$ *are all* $O_{\mathbb{P}}(n^{-1/2})$, *for any* $T > 0$.

Proof. The proof uses the differential equation method [50]. Notice that, after each ring of an exponential clock in Algorithm 3, $s_1(t)$ decreases by two. Let Y denote a unit rate Poisson process. Using the random time change representation [23],

$$s_1(t) = \ell_n - 2Y\left(\int_0^t s_1(u)\,\mathrm{d}u\right) = \ell_n + M_n(t) - 2\int_0^t s_1(u)\,\mathrm{d}u, \tag{D.1}$$

where $\boldsymbol{M}_n$ is a martingale. Now, the quadratic variation of $\boldsymbol{M}_n$ satisfies $\langle M_n\rangle(t) \le 4t\ell_n = O(n)$, which implies that $\sup_{t\le T}|M_n(t)| = O_{\mathbb{P}}(\sqrt{n})$. Moreover, notice that the function $f(t) = \mu_n \mathrm{e}^{-2t}$ satisfies $f(t) = \mu_n - 2\int_0^t f(u)\,\mathrm{d}u$. Therefore,

$$\sup_{t\le T}\left|\frac{1}{n}s_1(t) - \mu_n \mathrm{e}^{-2t}\right| \le \sup_{t\le T}\frac{|M_n(t)|}{n} + 2\int_0^T \sup_{t\le u}\left|\frac{1}{n}s_1(t) - \mu_n \mathrm{e}^{-2t}\right|\mathrm{d}u. \tag{D.2}$$

Using Grőnwall's inequality [38, Proposition 1.4], it follows that

$$\sup_{t\le T}\left|\frac{1}{n}s_1(t) - \mu_n \mathrm{e}^{-2t}\right| \le \mathrm{e}^{2T}\sup_{t\le T}\frac{|M_n(t)|}{n} = O_{\mathbb{P}}\big(n^{-1/2}\big), \tag{D.3}$$

as required. For $s_2(t)$, note that if half-edges corresponding to vertices i and j are paired, s_2 changes by $-2w_i - 2w_j + 2$ and if two half-edges corresponding to i are paired, s_2 changes by $-4w_i + 4$. Thus,

$$\begin{aligned}\sum_{i\in[n]} w_i(t)^2 &= \sum_{i\in[n]} d_i^2 + M_n'(t) + \int_0^t \sum_{i\ne j}\frac{w_i(u)w_j(u)(-2w_i(u) - 2w_j(u) + 2)}{s_1(u) - 1}\,\mathrm{d}u \\ &\quad + \int_0^t \sum_{i\in[n]}\frac{w_i(u)(w_i(u)-1)(-4w_i(u)+4)}{s_1(u)-1}\,\mathrm{d}u \\ &= n\mu_n(1+\nu_n) + M_n'(t) + \int_0^t \big(-4s_2(u) + 2s_1(u)\big)\,\mathrm{d}u + O(1),\end{aligned} \tag{D.4}$$

where $\boldsymbol{M}_n'$ is a martingale with quadratic variation given by $\langle M_n'\rangle(t) = O(n)$. Again, an estimate equivalent to (D.3) follows using Grőnwall's inequality. Notice also that when a clock corresponding to vertex i rings and it is paired to vertex j, then $s_{d,w}$ decreases by $d_i + d_j$. Thus,

$$\begin{aligned} s_{d,w}(t) &= \sum_{i\in[n]} d_i^2 + M_n''(t) - \int_0^t \sum_{i\ne j}\frac{w_i(u)w_j(u)(d_i+d_j)}{s_1(u)-1}\,\mathrm{d}u - \int_0^t\sum_{i\in[n]}\frac{w_i(u)(w_i(u)-1)2d_i}{s_1(u)-1}\,\mathrm{d}u \\ &= n\mu_n(1+\nu_n) + M_n''(t) - 2\int_0^t s_{d,w}(u)\,\mathrm{d}u,\end{aligned} \tag{D.5}$$

where $\boldsymbol{M}_n''$ is a martingale with quadratic variation given by $\langle M_n''\rangle(t) \le 2t\sum_{i\in[n]} d_i^2 = O(n)$. We can now apply Grőnwall's inequality as before. The proof of Lemma 37 is now complete. □

Acknowledgements

We sincerely thank Shankar Bhamidi for 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, a CRM-ISM fellowship, and a UGC CAS-II grant, Grant number F.510/25/CAS-II/2018(SAP-I). We thank Lorenzo Federico and Tim Hulshof for pointing out an error in the proof of Lemma 16 in the previous arXiv version of this paper.

References

[1] L. Addario-Berry, N. Broutin, C. Goldschmidt and G. Miermont. The scaling limit of the minimum spanning tree of the complete graph. *Ann. Probab.* **45** (5) (2017) 3075–3144. MR3706739 https://doi.org/10.1214/16-AOP1132

[2] E. Aïdékon, R. van der Hofstad, S. Kliem and J. S. H. van Leeuwaarden. Large deviations for power-law thinned Lévy processes. *Stochastic Process. Appl.* **126** (5) (2016) 1353–1384. MR3473098 https://doi.org/10.1016/j.spa.2015.11.006

[3] R. Albert and A. L. Barabási. Statistical mechanics of complex networks. *Rev. Modern Phys.* **74** (1) (2002) 47–97. MR1895096 https://doi.org/10.1103/RevModPhys.74.47

[4] D. Aldous. Brownian excursions, critical random graphs and the multiplicative coalescent. *Ann. Probab.* **25** (2) (1997) 812–854. MR1434128 https://doi.org/10.1214/aop/1024404421

[5] D. Aldous and V. Limic. The entrance boundary of the multiplicative coalescent. *Electron. J. Probab.* **3** (3) (1998) 1–59. MR1491528 https://doi.org/10.1214/EJP.v3-25

[6] D. Aldous and B. Pittel. On a random graph with immigrating vertices: Emergence of the giant component. *Random Structures Algorithms* **17** (2) (2000) 79–102. MR1774745 https://doi.org/10.1002/1098-2418(200009)17:2<79::AID-RSA1>3.3.CO;2-N

[7] A. L. Barabási and R. Albert. Emergence of scaling in random networks. *Science* **286** (1999) 509–512. MR2091634 https://doi.org/10.1126/science.286.5439.509

[8] J. Bertoin. *Lévy Processes*. Cambridge University Press, Cambridge, 1996. MR1406564

[9] S. Bhamidi, N. Broutin, S. Sen and X. Wang. Scaling limits of random graph models at criticality: Universality and the basin of attraction of the Erdős–Rényi random graph, 2014. Available at arXiv:1411.3417.

[10] S. Bhamidi, A. Budhiraja and X. Wang. The augmented multiplicative coalescent, bounded size rules and critical dynamics of random graphs. *Probab. Theory Related Fields* **160** (3) (2014) 733–796. MR3278920 https://doi.org/10.1007/s00440-013-0540-x

[11] S. Bhamidi, S. Dhara, R. van der Hofstad and S. Sen. Universality for critical heavy-tailed random graphs: Metric structure of maximal components, 2017. Available at arXiv:1703.07145.

[12] S. Bhamidi, S. Dhara, R. van der Hofstad and S. Sen. Global lower mass-bound for critical configuration models in the heavy-tailed regime. In preparation.

[13] S. Bhamidi, R. van der Hofstad and S. Sen. The multiplicative coalescent, inhomogeneous continuum random trees, and new universality classes for critical random graphs. *Probab. Theory Related Fields* **170** (1) (2018) 387–474. MR3748328 https://doi.org/10.1007/s00440-017-0760-6

[14] S. Bhamidi, R. van der Hofstad and J. S. H. van Leeuwaarden. Scaling limits for critical inhomogeneous random graphs with finite third moments. *Electron. J. Probab.* **15** (6) (2010) 1682–1702. MR2735378 https://doi.org/10.1214/EJP.v15-817

[15] S. Bhamidi, R. van der Hofstad and J. S. H. van Leeuwaarden. Novel scaling limits for critical inhomogeneous random graphs. *Ann. Probab.* **40** (6) (2012) 2299–2361. MR3050505 https://doi.org/10.1214/11-AOP680

[16] N. H. Bingham, C. M. Goldie and J. L. Teugels. *Regular Variation*. Cambridge University Press, Cambridge, 1989. MR1015093

[17] B. Bollobás. A probabilistic proof of an asymptotic formula for the number of labelled regular graphs. *European J. Combin.* **1** (4) (1980) 311–316. MR0595929 https://doi.org/10.1016/S0195-6698(80)80030-8

[18] B. Bollobás. The evolution of random graphs. *Trans. Amer. Math. Soc.* **286** (1) (1984) 257–274. MR0756039 https://doi.org/10.2307/1999405

[19] L. Breiman. *Probability*. Society for Industrial and Applied Mathematics, Philadelphia, 1968. MR1163370 https://doi.org/10.1137/1.9781611971286

[20] N. Broutin and J.-F. Marckert. A new encoding of coalescent processes: Applications to the additive and multiplicative cases. *Probab. Theory Related Fields* **166** (1) (2016) 515–552. MR3547745 https://doi.org/10.1007/s00440-015-0665-1

[21] T. C. Brown. Compensators and Cox convergence. *Math. Proc. Cambridge Philos. Soc.* **90** (02) (1981) 305. MR0620740 https://doi.org/10.1017/S0305004100058760X

[22] S. Dhara, R. van der Hofstad, J. S. H. van Leeuwaarden and S. Sen. Critical window for the configuration model: Finite third moment degrees. *Electron. J. Probab.* **22** (16) (2017) 1–33. MR3622886 https://doi.org/10.1214/17-EJP29

[23] S. N. Ethier and T. G. Kurtz. *Markov Processes: Characterization and Convergence*. John Wiley & Sons, Inc., New York, 1986. MR0838085 https://doi.org/10.1002/9780470316658

[24] M. Faloutsos, P. Faloutsos and C. Faloutsos. On power-law relationships of the Internet topology. *Comput. Commun. Rev.* **29** (4) (1999) 251–262.

[25] N. Fountoulakis. Percolation on sparse random graphs with given degree sequence. *Internet Math.* **4** (1) (2007) 329–356. MR2522948

[26] G. H. Hardy. *Divergent Series*. Clarendon (Oxford University) Press, Oxford, 1949. MR0030620

[27] S. Janson. The largest component in a subcritical random graph with a power law degree distribution. *Ann. Appl. Probab.* **18** (4) (2008) 1651–1668. MR2434185 https://doi.org/10.1214/07-AAP490

[28] S. Janson. On percolation in random graphs with given vertex degrees. *Electron. J. Probab.* **14** (2009) 87–118. MR2471661 https://doi.org/10.1214/EJP.v14-603

[29] S. Janson. The probability that a random multigraph is simple. *Combin. Probab. Comput.* **18** (1–2) (2009) 205–225. MR2497380 https://doi.org/10.1017/S0963548308009644

[30] S. Janson. Susceptibility of random graphs with given vertex degrees. *J. Comb.* **1** (3–4) (2010) 357–387. MR2799217 https://doi.org/10.4310/JOC.2010.v1.n4.a2

[31] S. Janson, D. E. Knuth, T. Łuczak and B. Pittel. The birth of the giant component. *Random Structures Algorithms* **4** (3) (1993) 233–358. MR1220220 https://doi.org/10.1002/rsa.3240040303

[32] S. Janson, T. Łuczak and A. Rucinski. *Random Graphs*. Wiley, New York, 2000. MR1782847 https://doi.org/10.1002/9781118032718

[33] A. Joseph. The component sizes of a critical random graph with given degree sequence. *Ann. Appl. Probab.* **24** (6) (2014) 2560–2594. MR3262511 https://doi.org/10.1214/13-AAP985

[34] O. Kallenberg. *Foundations of Modern Probability*. Springer, New York, 1997. MR1464694

[35] J. M. Kleinberg, R. Kumar, P. Raghavan, S. Rajagopalan and A. S. Tomkins. The Web as a graph: Measurements, models and methods. In *COCOON '99 Proceedings* 1–17. *Lecture Notes in Computer Science* **1627**. Springer, Berlin, 1999. MR1730317 https://doi.org/10.1007/3-540-48686-0_1

[36] E. Lenglart. Relation de domination entre deux processus. *Ann. Inst. Henri Poincaré Probab. Stat.* **13** (2) (1977) 171–179. MR0471069

[37] T. Łuczak. Component behavior near the critical point of the random graph process. *Random Structures & Algorithms* **1** (3) (1990) 287–310. MR1099794 https://doi.org/10.1002/rsa.3240010305

[38] R. H. Martin. *Nonlinear Operators and Differential Equations in Banach Spaces*. Krieger Publishing Co., Inc., Melbourne, FL, 1986. MR0887947

[39] M. Molloy and B. Reed. A critical-point for random graphs with a given degree sequence. *Random Structures & Algorithms* **6** (2–3) (1995) 161–179. MR1370952 https://doi.org/10.1002/rsa.3240060204

[40] A. Nachmias and Y. Peres. Critical percolation on random regular graphs. *Random Structures & Algorithms* **36** (2) (2010) 111–148. MR2583058 https://doi.org/10.1002/rsa.20277

[41] O. Riordan. The phase transition in the configuration model. *Combinatorics, Probability and Computing* **21** (2012) 265–299. MR2900063 https://doi.org/10.1017/S0963548311000666

[42] L. C. G. Rogers and D. Williams. *Diffusions, Markov Processes, and Martingales*, Vol. 1. 2nd edition. John Wiley & Sons Ltd., Chichester, 1994. MR1331599

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

[44] C. Stegehuis, R. van der Hofstad and J. S. H. van Leeuwaarden. Epidemic spreading on complex networks with community structures. *Scientific Reports* **6** (2016) 29748.

[45] R. van der Hofstad. *Random Graphs and Complex Networks*, Vol. I. Cambridge University Press, Cambridge, 2016. MR3617364 https://doi.org/10.1017/9781316779422

[46] R. van der Hofstad, S. Janson and M. Luczak. Component structure of the configuration model: Barely supercritical case. *Random Structures Algorithms* **55** (1) (2019) 3–55. MR3974190 https://doi.org/10.1002/rsa.20837

[47] R. van der Hofstad, A. J. E. M. Janssen and J. S. H. van Leeuwaarden. Critical epidemics, random graphs, and Brownian motion with a parabolic drift. *Adv. in Appl. Probab.* **42** (4) (2010) 1187–1206. MR2796682 https://doi.org/10.1239/aap/1293113157

[48] R. van der Hofstad, S. Kliem and J. S. H. van Leeuwaarden. Cluster tails for critical power-law inhomogeneous random graphs. *J. Stat. Phys.* **171** (1) (2018) 38–95. MR3773851 https://doi.org/10.1007/s10955-018-1978-0

[49] W. Whitt. *Stochastic-Process Limits: An Introduction to Stochastic-Process Limits and Their Application to Queues*. Springer-Verlag, New York, 2002. MR1876437

[50] N. C. Wormald. Differential equations for random processes and random graphs. *Ann. Appl. Probab.* **5** (4) (1995) 1217–1235. MR1384372