



Critical Percolation on Scale-Free Random Graphs: New Universality Class for the Configuration Model

Souvik Dhara¹ , Remco van der Hofstad², Johan S. H. van Leeuwen³

¹ Department of Mathematics, Massachusetts Institute of Technology, Microsoft Research, Cambridge, USA.
E-mail: sdhara@mit.edu

² Department of Mathematics and Computer Science, Eindhoven University of Technology, Eindhoven, The Netherlands. E-mail: r.w.v.d.hofstad@tue.nl

³ Stochastic Operations Research, Tilburg University, Tilburg, The Netherlands.
E-mail: j.s.h.vanleeuwen@tilburguniversity.edu

Received: 12 November 2019 / Accepted: 12 January 2021

Published online: 17 February 2021 – © The Author(s), under exclusive licence to Springer-Verlag GmbH, DE part of Springer Nature 2021

Abstract: In this paper, we study the critical behavior of percolation on a configuration model with degree distribution satisfying an infinite second-moment condition, which includes power-law degrees with exponent $\tau \in (2, 3)$. It is well known that, in this regime, many canonical random graph models, such as the configuration model, are *robust* in the sense that the giant component is not destroyed when the percolation probability stays bounded away from zero. Thus, the critical behavior is observed when the percolation probability tends to zero with the network size, despite of the fact that the average degree remains bounded. In this paper, we initiate the study of critical random graphs in the infinite second-moment regime by identifying the critical window for the configuration model. We prove scaling limits for component sizes and surplus edges, and show that the maximum diameter the critical components is of order $\log n$, which contrasts with the previous universality classes arising in the literature. This introduces a third and novel universality class for the critical behavior of percolation on random networks, that is not covered by the multiplicative coalescent framework due to Aldous and Limic (Electron J Probab 3(3):1–59, 1998). We also prove concentration of the component sizes outside the critical window, and that a unique, complex *giant component* emerges after the critical window. This completes the picture for the percolation phase transition on the configuration model.

1. Introduction

Bond percolation, or simply percolation, refers to the random graph obtained by independently keeping each edge of a graph with some fixed probability p (and deleting with probability $1 - p$). Percolation is a classical and important model in statistical physics and network science, as it serves as a canonical model for assessing robustness of a network when the edges of the underlying network are randomly damaged, and also as

This project was supported by the Netherlands Organisation for Scientific Research (NWO) through Gravitation Networks Grant 024.002.003. In addition, RvdH was supported by VICI Grant 639.033.806.

a basic model of vaccination for the prevention of an epidemic on networks. A detailed account of many of these applications can be found in [7,50]. From a theoretical perspective, percolation is one of the most elementary models that exhibits a phase transition, i.e., there exist values $p_c = p_c(n)$ such that for $p > p_c(1 + \varepsilon)$ and $\varepsilon > 0$, the proportion of vertices in the largest connected component is bounded away from zero with high probability, whereas for $p < p_c(1 - \varepsilon)$ this proportion becomes negligible. The critical behavior is observed when $p \approx p_c$, and fascinating behavior starts to emerge for the percolation process around this critical value.

It turns out that there is a window of values of p where the component functionals show intermediate and unique behavior. For example, rescaled component functionals converge to non-degenerate scaling limits, in contrast to the fact that they always concentrate for other values of p . Also, the large components in this window are structurally intermediate in the sense that neither there is a giant component with a growing number of cycles, nor do the components look like trees. This regime is called the *critical window* of the percolation phase-transition. Starting with the pioneering work of Aldous [4], deriving scaling limits for critical component functionals has been the ground for an enormous literature with several interesting scaling-limit results over the past decades [5,6,11,12,27,28,44,48,49,53]. We refer the reader to [26, Chapter 1] and references therein for an elaborate discussion of the nature of this transition, and a literature overview.

In the literature, two fundamentally different types of behavior have been proved for the scaling limits and the critical exponents associated to the critical window and component sizes depending on whether the asymptotic degree distribution satisfies a finite third-moment condition [11,28] or an infinite third - but a finite second-moment condition [12,27]. However, the study of critical behavior in the infinite second-moment setting was an open question.

When the degree distribution is asymptotically a power-law with exponent $\tau \in (2, 3)$, then the finite second-moment condition fails. These networks are popularly known as *scale-free networks* [7] in the literature. Many real-world networks are observed to be scale-free [2,30,36,50]. One of the well-known features of scale-free networks is that they are *robust* under random edge-deletion, i.e., for any sequence $(p_n)_{n \geq 1}$ satisfying $\liminf_{n \rightarrow \infty} p_n > 0$, the graph obtained by performing percolation with probability p_n is supercritical. This feature has been studied experimentally in [3], using heuristic arguments in [23–25,29] (see also [19,20,34] in the context of optimal paths in the strong disorder regime), and mathematically in [17]. Thus, in order to observe the percolation critical behavior, one needs to have $p_n \rightarrow 0$ with the network size, despite of the fact that the average degree of the network remains bounded.

In this paper, we initiate the study of critical behavior in the scale-free regime. As a canonical random graph model on which percolation acts, we take the multigraph generated by the configuration model. When the degree distribution satisfies a power-law with exponent $\tau \in (2, 3)$, it was heuristically argued in [24,29] that the critical value is $p_c \sim n^{-(3-\tau)/(\tau-1)}$, so that the critical window is given by the collection of values $p_c = p_c(\lambda) = \lambda n^{-(3-\tau)/(\tau-1)}$, where $\lambda > 0$ indicates the location inside the critical window. We establish that the scaling exponents from [24,29] are indeed true, and discuss asymptotics of component functionals inside the critical window. We also show that $p_c = p_c(\lambda) = \lambda n^{-(3-\tau)/(\tau-1)}$ with $\lambda > 0$ gives the right critical window, by showing that a giant component emerges at the end of the critical window ($\lambda \rightarrow \infty$), while components have a trivial star-like structure before the critical window ($\lambda \rightarrow 0$). The main contributions of this paper can be summarized as follows:

Critical window. At criticality, we obtain scaling limits for the largest component sizes and surplus edges in a strong topology. The result displays a completely new universality class of scaling limits of critical components. The scaling limits here are different from the general multiplicative coalescent framework in [5]. In particular, the limiting exploration process has bounded variation, so that the general tools from [5] cannot be applied. We also study the diameter of these components and show that the maximum diameter is of order $\log n$.

Near-critical behavior. For $p_n = \lambda_n n^{-(3-\tau)/(\tau-1)}$ with $\lambda_n \rightarrow 0$, the graph is subcritical and we show that the largest components sizes, rescaled by $n^\alpha p_n$, concentrate. On the other hand, when $\lambda_n \rightarrow \infty$, the largest component size, rescaled by $np_n^{1/(3-\tau)}$, concentrates, and this is the *unique giant component* in the sense that the size of the second largest component is much smaller than $np_n^{1/(3-\tau)}$. The nature of the emergence of this giant component for $p_n \gg p_c$ is markedly different compared to the universality classes in the $\tau \in (3, 4)$ and the $\tau > 4$ regimes, where the giant emerges when the percolation probability satisfies $(p_n - p_c(\lambda_1)) \gg (p_c(\lambda_2) - p_c(\lambda_1))$, for some strictly positive p_c and $-\infty < \lambda_1 < \lambda_2 < \infty$ [38].

Methods. Technically, analyzing percolation on random graphs like the configuration model is challenging, because in order to make Aldous's exploration process approach [4] work, one is required to keep track of many functionals of the unexplored part of the graph [48], resulting in a high-dimensional exploration process. This difficulty was circumvented in [27, 28] by using Janson's algorithm [39]. Unfortunately, Janson's algorithm does not work here due to the fact that the algorithm creates $n - o(n)$ degree-one vertices. Instead, we sandwich the percolated graph in between two configuration models, which yield the same scaling limits for the component sizes. Also, in order to deduce scaling limits of the component sizes from that of the exploration process, we prove several properties of the limiting exploration process, which are interesting from an independent perspective.

Remark 1 (Single-edge constraint). In a parallel work [10], Bhamidi and the first two authors consider critical percolation on simple random graphs, i.e., random graphs having no multiple-edges, namely generalized random graphs. It turns out that the critical window there is $p_c \sim n^{-(3-\tau)/2} \gg n^{-(3-\tau)/(\tau-1)}$. This is a distinctive feature in the infinite second-moment case that never surfaced in the other two universality classes of critical random graphs.

Organization of the paper. In Sect. 2, we state our results precisely. In Sect. 2.1, we give the precise definitions of the model and the scaling limits. Section 2.2 is devoted to comments about the heuristics, and some important special cases. In Sect. 3, we study excursions of the limiting exploration process. Section 4 contains the proofs of the results at criticality, and in Sect. 5, we analyze the near-critical regimes.

2. Main Results

2.1. The configuration model.

2.1.1. Model description The configuration model generates random multigraphs with any given degree sequence. Consider n vertices labeled by $[n] := \{1, 2, \dots, n\}$ and a non-increasing sequence of degrees $\mathbf{d} = \mathbf{d}_n = (d_i)_{i \in [n]}$ such that $\ell_n = \sum_{i \in [n]} d_i$ is even. The configuration model on n vertices having degree sequence \mathbf{d} is constructed as follows [8, 16]:

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

Let $\text{CM}_n(\mathbf{d})$ denote the graph constructed by the above procedure. Note that $\text{CM}_n(\mathbf{d})$ may contain self-loops and multiple edges. In fact, the probability that $\text{CM}_n(\mathbf{d})$ is a simple graph tends to zero in our setting with an infinite second-moment condition on the degree distribution [36, Proposition 7.12]. Before stating the main results about the configuration model, we set up some necessary notation.

2.1.2. Notions of convergence and the limiting objects To describe the main results of this paper, we need some definitions and notations. We use the Bachmann-Landau asymptotic notation $O(\cdot)$, $o(\cdot)$, $\Theta(\cdot)$ for large- n asymptotics of real numbers. For $(a_n)_{n \geq 1}, (b_n)_{n \geq 1} \subset (0, \infty)$, we write $a_n \ll b_n$, $a_n \sim b_n$ and $a_n \gg b_n$ as a shorthand for $\lim_{n \rightarrow \infty} a_n/b_n = 0, 1, \infty$, respectively. We often use C as a generic notation for a positive constant whose value can be different in different lines. We also 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 say that a sequence of events $(\mathcal{E}_n)_{n \geq 1}$ occurs with high probability (whp) with respect to the probability measures $(\mathbb{P}_n)_{n \geq 1}$ when $\mathbb{P}_n(\mathcal{E}_n) \rightarrow 1$. 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 $f_n/g_n \xrightarrow{\mathbb{P}} 0$; $f_n = \Theta_{\mathbb{P}}(g_n)$ if both $f_n = O_{\mathbb{P}}(g_n)$ and $g_n = O_{\mathbb{P}}(f_n)$. Denote

$$\ell_{\downarrow}^p := \left\{ \mathbf{x} = (x_i)_{i=1}^{\infty} \subset [0, \infty) : x_{i+1} \leq x_i \ \forall i, \text{ and } \sum_{i=1}^{\infty} x_i^p < \infty \right\}$$

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

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

endowed with the metric

$$d_{\mathbb{U}}((\mathbf{x}_1, \mathbf{y}_1), (\mathbf{x}_2, \mathbf{y}_2)) := \left(\sum_{i=1}^{\infty} (x_{1i} - x_{2i})^2 \right)^{1/2} + \sum_{i=1}^{\infty} |x_{1i} y_{1i} - x_{2i} y_{2i}|. \quad (2.1)$$

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

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

Let $(\mathbb{U}_{\downarrow}^0)^k$ denote the k -fold product space of $\mathbb{U}_{\downarrow}^0$.

Throughout, we write $\mathbb{D}[0, \infty)$ to denote the space of càdlàg functions $[0, \infty) \mapsto \mathbb{R}$ equipped with the Skorohod J_1 -topology. Also, let $\mathbb{D}_+[0, \infty) \subset \mathbb{D}[0, \infty)$ be the collection of functions with positive jumps only, and $\mathbb{C}[0, \infty) \subset \mathbb{D}[0, \infty)$ be the collection of continuous functions. For any fixed $T > 0$, $\mathbb{D}[0, T], \mathbb{D}_+[0, T], \mathbb{C}[0, T]$ are

defined similarly for functions $[0, T] \mapsto \mathbb{R}$. For any function $f \in \mathbb{D}[0, \infty)$, define $\underline{f}(t) = \inf_{s \leq t} f(s)$. Note that \underline{f} is non-increasing. Moreover,

$$\underline{f} \in \mathbb{C}[0, \infty), \quad \text{whenever } f \in \mathbb{D}_+[0, \infty). \tag{2.2}$$

Indeed, if \underline{f} is discontinuous at some point t , then $\underline{f}(t-) > \underline{f}(t)$, but that would mean that f has a negative jump of size $\underline{f}(t-) - \underline{f}(t)$ at t . Thus (2.2) holds. Next, for any $f \in \mathbb{D}_+[0, \infty)$, define the zero set of f by $\mathcal{Z}_f = \{t \geq 0 : f(t) - \underline{f}(t) = 0\}$, and let $\text{cl}(\mathcal{Z}_f)$ denote the closure of \mathcal{Z}_f . An interval (l, r) is called an excursion above the past minimum of f , or simply excursion of f (see [9, Sect. IV.2]) if

$$f(t) - \underline{f}(t) > 0, \quad \forall t \in (l, r), \quad \text{where } l \in \text{cl}(\mathcal{Z}_f) \text{ and } r \in \text{cl}(\mathcal{Z}_f) \cup \{\infty\}. \tag{2.3}$$

For $f \in \mathbb{D}_+[0, T]$, we consider $(l, r) \subset [0, T]$, and define an excursion similarly as in (2.3).

We often use boldface notation \mathbf{X} for the stochastic process $(X(s))_{s \geq 0}$, unless stated otherwise. Consider a decreasing sequence $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots) \in \ell_{\downarrow}^2 \setminus \ell_{\downarrow}^1$. Denote by $\mathcal{S}_i(s) := \mathbb{1}_{\{\xi_i \leq s\}}$ where $\xi_i \sim \text{Exp}(\theta_i/\mu)$ independently, and $\text{Exp}(r)$ denotes the exponential distribution with rate r . Consider the process

$$S_{\infty}^{\lambda}(t) = \frac{\lambda\mu}{\|\boldsymbol{\theta}\|_2^2} \sum_{i=1}^{\infty} \theta_i \mathcal{S}_i(t) - t, \tag{2.4}$$

for some $\lambda, \mu > 0$. Note that, for all $t > 0$, $\mathbb{E}[S_{\infty}^{\lambda}(t)] < \infty$ since $\sum_i \theta_i^2 < \infty$, and consequently $S_{\infty}^{\lambda}(t) < \infty$, almost surely. Also, for any $u < t$,

$$\mathbb{E}[|S_{\infty}^{\lambda}(t) - S_{\infty}^{\lambda}(u)|] \leq \frac{\lambda\mu}{\|\boldsymbol{\theta}\|_2^2} \sum_{i=1}^{\infty} \theta_i e^{-\theta_i u} (1 - e^{-\theta_i(t-u)}) + |t - u| \leq (\lambda\mu + 1)|t - u|,$$

so that $\mathbf{S}_{\infty}^{\lambda}$ has bounded variation almost surely. However, since $\sum_i \theta_i = \infty$, the process experiences infinitely many jumps in any bounded interval of time. Define the reflected version of $S_{\infty}^{\lambda}(t)$ by

$$\text{refl}(S_{\infty}^{\lambda}(t)) = S_{\infty}^{\lambda}(t) - \min_{0 \leq u \leq t} S_{\infty}^{\lambda}(u).$$

We will show that, for any $\lambda > 0$, the excursion lengths of the process $\mathbf{S}_{\infty}^{\lambda} = (S_{\infty}^{\lambda}(t))_{t \geq 0}$ can be ordered almost surely as an element of ℓ_{\downarrow}^2 . We denote this ordered vector of excursion lengths by $(\gamma_i(\lambda))_{i \geq 1}$. For $v, t > 0$, define $M_t(v) := \sum_{j: v\theta_j \leq 1, t\theta_j \leq 1} \theta_j^3$. We will assume that for any $t > 0$,

$$\int_0^{\infty} e^{-tv^2 M_t(v)} dv < \infty. \tag{2.5}$$

The technical condition in (2.5) on top of $\boldsymbol{\theta} \in \ell_{\downarrow}^2 \setminus \ell_{\downarrow}^1$ will be used to ensure that the distribution of $S_{\infty}^{\lambda}(t)$ is non-atomic for all $t > 0$ (see Lemma 4 below), which in turn implies that we have strict ordering between excursion lengths, i.e., $\gamma_{i+1}(\lambda) < \gamma_i(\lambda)$ for all $i \geq 1$ almost surely. The condition (2.5) is relatively weak, and is, for example, satisfied for $\theta_j = j^{-\alpha}$ for $\alpha \in (1/2, 1)$. To see this, note that $v^2 M_t(v)$ is of the same

order as $v^{-1+1/\alpha}$. However, this also shows that (2.5) is not satisfied for the extreme case $\alpha = 1$, i.e., $\theta_j = j^{-1}$.

Also, define the counting process $\mathbf{N}^\lambda = (N^\lambda(t))_{t \geq 0}$ to be the Poisson process that has intensity $(\lambda\mu^2)^{-1} \|\boldsymbol{\theta}\|_2^2 \times \text{refl}(S_\infty^\lambda(t))$ at time t , conditionally on $(S_\infty^\lambda(u))_{u \leq t}$. Formally, \mathbf{N}^λ is characterized as the counting process for which

$$N^\lambda(t) - \frac{\|\boldsymbol{\theta}\|_2^2}{\lambda\mu^2} \int_0^t \text{refl}(S_\infty^\lambda(u)) du \tag{2.6}$$

is a martingale. We use the notation $N_i(\lambda)$ to denote the number of marks of \mathbf{N}^λ in the i -th largest excursion of $\mathbf{S}_\infty^\lambda$. Define

$$\mathbf{Z}(\lambda) := ((\gamma_i(\lambda), N_i(\lambda)))_{i \geq 1}, \text{ ordered as an element of } \mathbb{U}_\downarrow^0. \tag{2.7}$$

2.1.3. Results for the critical window Fix $\tau \in (2, 3)$. Throughout this paper, we denote

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

Also, let D_n be the degree of a vertex chosen uniformly at random from $[n]$. We start by stating our assumptions on the degree sequences:

Assumption 1. For each $n \geq 1$, let $\mathbf{d} = \mathbf{d}_n = (d_1, \dots, d_n)$ be a degree sequence satisfying $d_1 \geq d_2 \geq \dots \geq d_n$. We assume the following about $(\mathbf{d}_n)_{n \geq 1}$ as $n \rightarrow \infty$:

- (i) (*High-degree vertices*) For any $i \geq 1$, $n^{-\alpha} d_i \rightarrow \theta_i$, where $\boldsymbol{\theta} := (\theta_i)_{i \geq 1} \in \ell_\downarrow^2 \setminus \ell_\downarrow^1$ is such that (2.5) holds.
- (ii) (*Moment assumptions*) $(D_n)_{n \geq 1}$ is uniformly integrable, $\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i \in [n]} d_i = \mu$ for some $\mu > 0$, and

$$\lim_{K \rightarrow \infty} \limsup_{n \rightarrow \infty} n^{-2\alpha} \sum_{i=K+1}^n d_i^2 = 0. \tag{2.8}$$

In Sect. 2.2, we discuss the generality of Assumption 1 and show that power-law degrees satisfy these assumptions. For $\text{CM}_n(\mathbf{d})$, the *criticality parameter* ν_n is defined as

$$\nu_n = \frac{\sum_{i \in [n]} d_i(d_i - 1)}{\sum_{i \in [n]} d_i}.$$

Molloy and Reed [46], and Janson and Luczak [41] showed that, under some regularity conditions, $\text{CM}_n(\mathbf{d})$ has a unique giant component (a component of size $\Theta(n)$) with high probability precisely when $\nu_n \rightarrow \nu > 1$. Under Assumption 1, $\nu_n \rightarrow \infty$, as $n \rightarrow \infty$ since $\sum_{i \in [n]} d_i^2 \geq d_1^2 = \Theta(n^{2\alpha}) \gg n$, and $\text{CM}_n(\mathbf{d})$ always contains a giant component (see the remark below [37, Theorem 4.5] and consider $\pi = 1$).

We study *percolation*, which refers to deleting each edge of a graph independently with probability $1 - p$. In case of percolation on random graphs, the deletion of edges is also independent from the underlying graph. The percolation probability is allowed to depend on the network size, i.e., $p = p_n$. Let $\text{CM}_n(\mathbf{d}, p_n)$ denote the graph obtained from percolation with probability p_n on the graphs $\text{CM}_n(\mathbf{d})$. Fountoulakis [32] showed that $\text{CM}_n(\mathbf{d}, p_n)$ is distributed as $\text{CM}_n(\mathbf{d}^p)$, where \mathbf{d}^p is the degree sequence of the percolated graph. Note that the degrees in \mathbf{d}^p could be correlated, so later Janson [39]

gave an explicit construction which is simpler to analyze. This construction was used to identify the percolation phase transition in [39] and to study the critical window in [27,28]. An interested reader is also referred to [28, Algorithm 4] where a construction of the whole percolation process $(\text{CM}_n(\mathbf{d}, p))_{p \in [0,1]}$ is provided.

Now, under Assumption 1, if $\liminf_{n \rightarrow \infty} p_n > 0$, then $\text{CM}_n(\mathbf{d}, p_n)$ retains a giant component with high probability, i.e., $\text{CM}_n(\mathbf{d}, p_n)$ is always supercritical; see the remark below [37, Theorem 4.5]. Thus, in order to see the critical behavior, one must take $p_n \rightarrow 0$, as $n \rightarrow \infty$. For $p_n \rightarrow 0$, the graph always contains $n - o_{\mathbb{P}}(n)$ degree-zero or isolated vertices, which makes Janson’s construction inconvenient to work with.

For a sequence of finite graphs, the critical behavior is where we see intermediate behavior in the sense that it inherits some features from the subcritical (such as the absence of the giant component) and the supercritical regimes (the largest component is not a tree). The collection of such values of p is called the critical window. However, due to our lack of knowledge about the subcritical phase and the structural properties therein, it is not a priori evident here how to define the critical window. One way to define the subcritical regime and the critical window would be to say that inside the critical window, the rescaled vector of ordered component sizes converge to some *non-degenerate* random vector, whereas the component sizes concentrate in the subcritical regime. This property has been observed quite universally for the percolation critical window. In this paper, we take this as our definition of the critical window. It is worthwhile to mention that there is a substantial literature on how to define the critical value. See [18,35,37,43,47] for different definitions of the critical probability and related discussions.

We will show that the critical window for percolation on $\text{CM}_n(\mathbf{d})$ is given by

$$p_c = p_c(\lambda) := \frac{\lambda}{v_n}(1 + o(1)), \quad \lambda \in (0, \infty). \tag{2.9}$$

Notice that, under Assumption 1, $p_c \sim n^{-2\alpha+1} \sim n^{-\eta}$, where $\eta = (3 - \tau)/(\tau - 1) > 0$. The case where $p \ll p_c$ will be called the barely subcritical regime and the case $p_c \ll p \ll 1$ will be called the barely supercritical regime. We will show that a unique giant component emerges in the barely supercritical regime. We first state the results about the component sizes and the complexity in the critical window, and then discuss the barely sub-/supercritical regimes.

We will always write $\mathcal{C}_{(i)}(p)$ to denote the i -th largest component in the percolated graph. The random graph on which percolation acts will always be clear from the context. A vertex is called isolated if it has degree zero in the graph $\text{CM}_n(\mathbf{d}, p_c(\lambda))$. We define the component size corresponding to an isolated vertex to be zero (see Remark 2 below). For any component $\mathcal{C} \subset \text{CM}_n(\mathbf{d}, p_c(\lambda))$, let $\text{SP}(\mathcal{C})$ denote the number of surplus edges given by $\#\{\text{edges in } \mathcal{C}\} - |\mathcal{C}| + 1$. Finally, let

$$\mathbf{Z}_n(\lambda) := \left(n^{-\rho} |\mathcal{C}_{(i)}(p_c(\lambda))|, \text{SP}(\mathcal{C}_{(i)}(p_c(\lambda))) \right)_{i \geq 1}, \text{ ordered as an element of } \mathbb{U}_{\downarrow}^0.$$

The following theorem gives the asymptotics for the critical component sizes and the surplus edges of $\text{CM}_n(\mathbf{d}, p_c(\lambda))$:

Theorem 1 (Critical component sizes and surplus edges). *Under Assumption 1, as $n \rightarrow \infty$,*

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

with respect to the $\mathbb{U}_{\downarrow}^0$ topology, where $\mathbf{Z}(\lambda)$ is defined in (2.7).

Remark 2 (Ignoring isolated components). Note that $2\rho < 1$ for $\tau \in (2, 3)$. When percolation is performed with probability p_c , there are of the order n isolated vertices and thus $n^{-2\rho}$ times the number of isolated vertices tends to infinity. This is the reason why we must ignore the contributions due to isolated vertices, when considering the convergence of the component sizes in the ℓ_{\downarrow}^2 -topology. Note that an isolated vertex with self-loops does not create an isolated component.

For a connected graph G , $\text{diam}(G)$ denotes the diameter of the graph, i.e., the maximum graph distance between any pair of vertices. For an arbitrary graph G , $\text{diam}(G) := \max \text{diam}(\mathcal{C})$, where the maximum is taken over all connected components. Our next result shows that the diameter of the largest connected components is of order $\log n$:

Theorem 2 (Diameter of largest critical clusters). *Under Assumption 1,*

$$\text{diam}(\text{CM}_n(\mathbf{d}, p_c(\lambda))) = O_{\mathbb{P}}(\log n).$$

Thus, the maximum diameter scales logarithmically in the $\tau \in (2, 3)$, in contrast to the other universality classes in the $\tau \in (3, 4)$ and $\tau > 4$ regimes, where graph distances scale as a positive power of n [1, 13].

2.1.4. Behavior in the near-critical regimes We now discuss asymptotic results for the component sizes in the barely subcritical ($p_n \ll p_c(\lambda)$) and barely supercritical ($p_n \gg p_c(\lambda)$) regimes. The next two theorems summarize the behavior outside the critical window:

Theorem 3 (Barely subcritical regime). *For $\text{CM}_n(\mathbf{d}, p_n)$, suppose that $n^{-\alpha} \ll p_n \ll p_c(\lambda)$ and that Assumption 1 holds. Then, as $n \rightarrow \infty$,*

$$\left((n^{\alpha} p_n)^{-1} |\mathcal{C}_{(i)}(p_n)| \right)_{i \geq 1} \xrightarrow{\mathbb{P}} (\theta_i)_{i \geq 1},$$

in ℓ_{\downarrow}^2 topology, and $\mathbb{P}(\text{SP}(\mathcal{C}_{(i)}(p_n)) = 0) \rightarrow 1$, for all $i \geq 1$.

Remark 3 (Components and hubs). In the barely subcritical regime, we show that the i -th largest component is essentially the component containing the i -th largest degree vertex, or the i -th hub. Since the hubs have degree $\Theta(n^{\alpha})$, we need the assumption that $p_n \gg n^{-\alpha}$ in Theorem 3, as otherwise the hubs become isolated, in which case components are likely to be extremely small.

For the result in the barely supercritical regime, let $p_c(\lambda) \ll p_n \ll 1$. The exact asymptotics of the high-degree vertices and the tail behavior in (2.8) will not be required. Below, we state the sufficient conditions for the concentration of the size of the giant component. In Sect. 2.2, we will see that these conditions are satisfied when the degrees are sampled from a power-law distribution:

Assumption 2. For each $n \geq 1$, let $\mathbf{d} = \mathbf{d}_n = (d_1, \dots, d_n)$ be a degree sequence satisfying $d_1 \geq d_2 \geq \dots \geq d_n$. We assume the following about $(\mathbf{d}_n)_{n \geq 1}$:

- (i) $d_1 = O(n^{\alpha})$.
- (ii) $(D_n)_{n \geq 1}$ is uniformly integrable, and $\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i \in [n]} d_i = \mu$ for some $\mu > 0$.
- (iii) Let D_n^* denote the degree of a vertex chosen in a size-biased manner with the sizes being $(d_i / \ell_n)_{i \in [n]}$. Then, there exists a constant $\kappa > 0$ such that

$$1 - \mathbb{E}[e^{-t p_n^{1/(3-\tau)} D_n^*}] = \kappa p_n^{(\tau-2)/(3-\tau)} (t^{\tau-2} + o(1)). \quad (2.10)$$

Let $E(G)$ denote the number of edges in the graph G .

Theorem 4 (Barely supercritical regime). *For $\text{CM}_n(\mathbf{d}, p_n)$, suppose that $p_c(\lambda) \ll p_n \ll 1$ and that Assumption 2 hold. Then, as $n \rightarrow \infty$,*

$$\frac{|\mathcal{C}_{(i)}(p_n)|}{np_n^{1/(3-\tau)}} \xrightarrow{\mathbb{P}} \mu\kappa^{1/(3-\tau)}, \quad \frac{E(\mathcal{C}_{(i)}(p_n))}{np_n^{1/(3-\tau)}} \xrightarrow{\mathbb{P}} \mu\kappa^{1/(3-\tau)},$$

and for all $i \geq 2$, $|\mathcal{C}_{(i)}(p_n)| = o_{\mathbb{P}}(np_n^{1/(3-\tau)})$, $E(\mathcal{C}_{(i)}(p_n)) = o_{\mathbb{P}}(np_n^{1/(3-\tau)})$.

Remark 4 (Relation to Abel-Tauberian theorem). The infinite second-moment assumption is captured by (2.10). The identity (2.10) is basically a version of the celebrated Abel-Tauberian theorem [31, Chapter XIII.5] (see also [15, Chapter 1.7]). However, since both D_n^* and p_n depend on n , the joint asymptotics needs to be stated as an assumption. In Sect. 2.2, we discuss how this assumption is satisfied when (i) $d_i = (1 - F)^{-1}(i/n)$ (ii) d_i is the i -th order statistic of an i.i.d sample, where F is a power-law distribution with $\tau \in (2, 3)$.

2.2. Discussion.

Critical window: emergence of hub connectivity. The critical window is the regime in which hubs start getting connected. Hubs are the high-degree vertices, whose asymptotic degree is determined by Assumption 1(i). To understand the above remark more precisely, let us denote the probability that i and j are in the same component in the p -percolated graph by $\pi(i, j, p)$. Then, for any fixed $i, j \geq 1$,

$$\limsup_{n \rightarrow \infty} \pi(i, j, p_n) = 0 \quad \text{for } p_n \ll p_c, \tag{2.11}$$

$$0 < \liminf_{n \rightarrow \infty} \pi(i, j, p_n) \leq \limsup_{n \rightarrow \infty} \pi(i, j, p_n) < 1 \quad \text{for } p_n = \Theta(p_c), \tag{2.12}$$

$$\limsup_{n \rightarrow \infty} \pi(i, j, p_n) = 1 \quad \text{for } p_n \gg p_c, \tag{2.13}$$

Indeed, any two vertices i and j share $p_n d_i d_j / (\ell_n - 1)$ edges in expectation. This expectation is $o(1)$, $\Theta(1)$, or $\omega(1)$ depending on whether $p_n \ll p_c$, $p_n \sim p_c$, or $p_n \gg p_c$. In the subcritical regime, this observation and a simple union bound yields (2.11). For the critical case, a method of moment computation shows that the number of edges between hubs i and j converges in distribution to $\text{Poisson}(\lambda\theta_i\theta_j/\mu)$. We don't prove this here, but instead refer the reader to [36, Proposition 7.13] where similar Poisson approximation computations have been done for the configuration model. This shows (2.12). In the super-critical regime,

$$\mathbb{P}((i, j) \text{ don't share any edge}) \leq \prod_{l=1}^{d_j} \left(1 - \frac{p_n d_i}{\ell_n - 2l + 1} \right) \leq e^{-p_n d_i d_j / 2\ell_n} \rightarrow 0,$$

so that $1 - \pi(i, j, p_n) \rightarrow 0$ which yields (2.13). Intuitively, in the barely subcritical regime, all the hubs are in different components. Hubs start getting connected to each other directly, forming the critical components as the p varies over the critical window. Finally in the barely super-critical regime the giant component, which contains all the hubs, is formed. The features (2.11), (2.12) and (2.13) are also observed in the $\tau \in (3, 4)$

case [12]. However, the key distinction between $\tau \in (3, 4)$ and $\tau \in (2, 3)$ is that for $\tau \in (3, 4)$ the paths between the hubs have lengths that grow as $n^{(\tau-3)/(\tau-1)}$, whereas they are directly connected in the $\tau \in (2, 3)$ regime.

Intuitive explanation for the exploration process. Suppose that we explore the critically percolated configuration model sequentially in a breadth-first manner. The reflected version of the stochastic process in (2.4) turns out to be the limit of the process that counts the number of unpaired half-edges incident to the discovered vertices. This limiting process can be intuitively understood as follows. When we explore hubs, the exploration process increases drastically, causing the jumps in the first term in (2.4). The negative linear drift is an accumulation of two effects. (1) Because we explore two vertices at each time, we get a negative drift $-2t$. (2) The exploration of the low-degree vertices cumulatively causes a linear positive drift $+t$. The main contribution in the latter case comes due to the degree-one vertices in the system. Thus in total, we get a drift of $-t$ in the exploration process (2.4).

Assumption on the degrees. Assumptions 1, 2 hold for two interesting special cases of power-law degrees that have received special attention in the literature: Case (I) $d_i = (1 - F)^{-1}(i/n)$, Case (II) d_i 's are the order statistics of an i.i.d sample from F . Here F is some distribution function supported on non-negative integers and $(1 - F)(x) = c_F k^{-(\tau-1)}$, for $k \leq x < k+1$, and we recall that the inverse of a bounded non-increasing function $f : \mathbb{R} \mapsto \mathbb{R}$ is defined as

$$f^{-1}(x) := \inf\{y : f(y) \leq x\}.$$

We add a dummy half-edge to vertex 1 if necessary to make $\sum_{i \in [n]} d_i$ even. However, we ignore this contribution since this does not change any asymptotic calculation below. Recall that we use C as a generic notation for a constant whose value can be different between expressions, and $a_n \sim b_n$ denotes $\lim_{n \rightarrow \infty} a_n/b_n = 1$.

For Case (I), $d_i \sim (c_F n/i)^\alpha$ for all $i = o(n)$ and $d_i \leq C(n/i)^\alpha$ for all $i \in [n]$. Consequently, Assumption 1(i) is satisfied with $\theta_i = c_F^\alpha i^{-\alpha}$. To see Assumption 1(ii), note that

$$\frac{1}{n} \sum_{i \in [n]} d_i \sim \int_0^1 (1 - F)^{-1}(x) dx = \mathbb{E}[D],$$

where D has distribution function F , and

$$n^{-2\alpha} \sum_{i > K} d_i^2 \leq C \sum_{i > K} i^{-2\alpha} \sim CK^{1-2\alpha} \rightarrow 0 \quad \text{as } K \rightarrow \infty. \tag{2.14}$$

Also, $D_n \xrightarrow{d} D$, and $\mathbb{E}[D_n] \rightarrow \mathbb{E}[D]$ implies that $(D_n)_{n \geq 1}$ is uniformly integrable. To see Assumption 2, with the above computations, we have already verified all the conditions in Assumption 2(i), (ii). To verify Assumption 2(iii), we now show that, for $t_n = t p_n^{1/(3-\tau)}$ with fixed $t > 0$,

$$1 - \mathbb{E}[e^{-t_n D_n^*}] = \frac{1}{\ell_n} \sum_{k \in [n]} d_k (1 - e^{-t_n d_k}) \sim t_n^{\tau-2} \int_0^\infty c_F z^{-\alpha} (1 - e^{-c_F z^{-\alpha}}) dz,$$

and thus (2.10) holds as well. Let us split the last sum in three parts by restricting to the set $\{k : d_k < \varepsilon(t_n)^{-1}\}$, $\{k : d_k \in [\varepsilon(t_n)^{-1}, (\varepsilon t_n)^{-1}]\}$, and $\{k : d_k > (\varepsilon t_n)^{-1}\}$ and denote them by (I), (II) and (III) respectively. Using the fact that $1 - e^{-x} \leq x$, it follows that

$$\begin{aligned} \frac{(I)}{t_n^{\tau-2}} &\leq \frac{t_n^{3-\tau}}{\ell_n} \sum_{k:d_k < \varepsilon(t_n)^{-1}} d_k^2 \sim Cn^{2\alpha-1} t_n^{3-\tau} \sum_{k \geq Cn(t_n/\varepsilon)^{1/\alpha}} k^{-2\alpha} \\ &\sim Cn^{2\alpha-1} t_n^{3-\tau} \int_{Cn(t_n/\varepsilon)^{\tau-1}}^{\infty} x^{-2\alpha} dx \sim Cn^{2\alpha-1} t_n^{3-\tau} (Cn(t_n/\varepsilon)^{\tau-1})^{1-2\alpha} \sim C\varepsilon^{3-\tau}, \end{aligned} \quad (2.15)$$

and

$$\frac{(III)}{t_n^{\tau-2}} \leq \frac{C}{t_n^{(\tau-2)} \ell_n} \sum_{k:d_k > (\varepsilon t_n)^{-1}} d_k \leq \frac{Cn^{\alpha-1}}{t_n^{\tau-2}} \int_1^{Cn(t_n/\varepsilon)^{\tau-1}} \frac{dx}{x^\alpha} \sim C\varepsilon^{\tau-2}. \quad (2.16)$$

Also, we compute (II) by

$$\begin{aligned} \frac{(II)}{t_n^{\tau-2}} &= \frac{1}{t_n^{\tau-2} \ell_n} \sum_{k:d_k \in [\varepsilon t_n^{-1}, (\varepsilon t_n)^{-1}]} d_k (1 - e^{-t_n d_k}) \\ &\sim \frac{n^{\alpha-1}}{\mu t_n^{\tau-2}} \sum_{k \in [c_0 n(t_n/\varepsilon)^{\tau-1}, c_1(t_n/\varepsilon)^{\tau-1}]} c_F k^{-\alpha} (1 - e^{-t_n(c_F n/k)^\alpha}) \\ &= \frac{1}{n t_n^{\tau-1}} \sum_{z \in [c_0 \varepsilon^{\tau-1}, c_1/\varepsilon^{\tau-1}]} c_F z^{-\alpha} (1 - e^{-c_F z^{-\alpha}}), \end{aligned}$$

where we have put $k = n t_n^{\tau-1} z$, so that the z values increase by $1/(n t_n^{\tau-1})$ in the final sum. Thus, in the iterated limit $\lim_{\varepsilon \rightarrow 0} \limsup_{n \rightarrow \infty}$,

$$\frac{(II)}{t_n^{\tau-2}} \rightarrow \int_0^\infty c_F z^{-\alpha} (1 - e^{-c_F z^{-\alpha}}) dz = \kappa,$$

which yields (2.10) by combining it with (2.15) and (2.16).

Let us now consider Case (II), i.e., the i.i.d degree setup. We have assumed that the degree sequence is ordered in a non-decreasing manner, i.e., d_i is the i -th order statistic of the i.i.d samples. We use the following construction from [21, Sect. 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

$$\bar{d}_i = (1 - F)^{-1}(\Gamma_i/\Gamma_{n+1}). \quad (2.17)$$

Then $(d_1, \dots, d_n) \stackrel{d}{=} (\bar{d}_1, \dots, \bar{d}_n)$. Now, Γ_i 's follow a Gamma distribution with shape parameter n and scale parameter 1. 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 (2.17), we see that \mathbf{d} satisfies Assumption 1(i) almost surely with $\theta_i = (C_F/\Gamma_i)^\alpha$. To see that $(\theta_i)_{i \geq 1} \in \ell_{\downarrow}^2 \setminus \ell_{\downarrow}^1$, observe that $\Gamma_i/i \xrightarrow{\text{a.s.}} 1$, and $\alpha \in (1/2, 1)$. Next, the first condition in Assumption 1(ii) follows from the strong law of large numbers. To see the second condition, we note that $\sum_i \Gamma_i^{-2\alpha} < \infty$ almost surely. Now using the fact that $\Gamma_{n+1}/n \xrightarrow{\text{a.s.}} 1$, we can

use arguments identical to (2.14) to show that $\lim_{K \rightarrow \infty} \limsup_{n \rightarrow \infty} n^{-2\alpha} \sum_{i>K} d_i^2 = 0$ on the event $\{\sum_{i=1}^{\infty} \Gamma_i^{-2\alpha} < \infty\} \cap \{\Gamma_{n+1}/n \rightarrow 1\}$. Thus, we have shown that the third condition of Assumption 1(ii) holds almost surely. The verification of Assumption 2 is also identical to Case-(I) if we do the computations conditionally on the Gamma random variables and use the above asymptotics.

Extension to the Norros–Reittu model. A related model where one would expect the same behavior as the configuration model is the multigraph version of the Norros-Reittu model or the Poisson graph process [51]. Given a weight sequence $(w_i)_{i \in [n]}$, the Norros-Reittu multigraph is the multigraph generated by putting Poisson($w_i w_j / L_n$) many edges between vertices i and j , where $L_n = \sum_{i \in [n]} w_i$. If Assumptions 1, 2 holds with $(d_i)_{i \in [n]}$ replaced by $(w_i)_{i \in [n]}$, then we expect the same results for percolation on the Norros-Reittu multigraph about the critical and near critical regimes as described above. We do not pursue the Norros-Reittu multigraph here.

Open Problems. We next state some open problems:

Open Problem 1. Theorem 1 studies convergence of $\mathbf{Z}_n(\lambda)$ for each fixed λ . It will be interesting to study the distribution of $(\mathbf{Z}_n(\lambda))_{\lambda>0}$ as a stochastic process, when the percolated graphs are coupled through the Harris coupling. In the $\tau > 4$ and $\tau \in (3, 4)$ regimes, such evolution of critical components is described by the so-called augmented multiplicative coalescent process. However, we do not expect the limit to be the augmented multiplicative coalescent here. This is clear from the fact that the scaling limit in (2.4) is not related to the general characterization of exploration processes that arise in relation to multiplicative coalescent in [5]. Heuristically, one would expect that if $\sum_{i \in \mathcal{C}} d_i \mathbb{1}\{i \text{ is hub}\}$ denotes the mass of a component, then the components would merge at rate proportional to their masses, but additionally, there are immigrating vertices of degree-one that keep on increasing the component sizes as well. The description of the process, and proving its Feller properties and entrance boundary conditions, are interesting open challenges.

Open Problem 2. Is it possible to prove that the metric structure of components converges in a suitable topology? This question is motivated by a strong notion of structural convergence of critical components that was first established in [1] ($\tau > 4$) and [13] ($\tau \in (3, 4)$). Since the components have small distances, it may be natural to consider the local-weak convergence framework. However, the hubs within components have unbounded degrees, which is not covered directly in the local-weak convergence framework.

3. Properties of the Excursions of the Limiting Process

In this section, we prove some good properties of the process (2.4) that allows us to conclude the convergence of largest excursion lengths from the stochastic process convergence. In Sect. 3.1, we identify these good properties for functions in $\mathbb{D}_+[0, \infty)$ that ensure continuity of the largest excursion map. Then, we prove in Sect. 3.2 that $\mathbf{S}_{\infty}^{\lambda}$ satisfies these good properties almost surely.

3.1. Continuity of the largest excursion map. Recall the definitions of excursions from (2.3). Also, recall from Sect. 2.1.2 that $\underline{f}(t) = \inf_{u \leq t} f(u)$ and $\mathcal{E}_f = \{t : f(t) = \underline{f}(t)\}$. Define the set of excursions of f as

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

We denote the set of excursion begin-points (or left-points) and end-points (or right-points) by \mathcal{L}_f and \mathcal{R}_f respectively, i.e.,

$$\mathcal{L}_f := \{l \geq 0 : (l, r) \in \mathcal{E}_f \text{ for some } r\} \quad \text{and} \quad \mathcal{R}_f := \{r \geq 0 : (l, r) \in \mathcal{E}_f \text{ for some } l\}.$$

We will use the following elementary fact:

Fact 1. *Let $f \in \mathbb{D}_+[0, \infty)$. Then, for all $r \in \mathcal{R}_f \setminus \{\infty\}$, f is continuous at r . Consequently, $r \in \mathcal{L}_f$.*

Proof. Using the right-continuity of f , it suffices to show that $f(r) = f(r-)$. Suppose that is not the case. Since f has positive jumps only, we must have that $f(r-) < f(r)$. Since r is an excursion ending point, there exists $\varepsilon > 0$ such that $f(t) - \underline{f}(t) > 0$ for all $t \in (r - \varepsilon, r)$. On the other hand, using the right-continuity of f and the fact that $f(r) > f(r-)$, we obtain that $f(t) - \underline{f}(t) > 0$ for all $t \in [r, r + \varepsilon)$ for some $\varepsilon > 0$. Thus, there exists a sufficiently small $\varepsilon > 0$ such that $f(t) - \underline{f}(t) > 0$ for all $t \in (r - \varepsilon, r + \varepsilon)$. This contradicts the fact that $r \in \text{cl}(\mathcal{L}_f) \setminus \{\infty\}$. \square

For $f \in \mathbb{D}_+[0, \infty)$, let $\phi_i(f)$ be the length of the i -th largest excursion of f . Also, let $\mathcal{A}_i(f)$ denote the area under i -th largest excursion of f . We will show that if $f_n \rightarrow f$ in $\mathbb{D}[0, \infty)$ then ϕ_i and \mathcal{A}_i converge when the limiting function has some good properties. Let us start by describing these *good* properties:

Definition 1 (*Good functions*). A function $f \in \mathbb{D}_+[0, \infty)$ is said to be *good* if the following holds:

- (a) For all $r \in \mathcal{R}_f \setminus \{\infty\}$, r is not a local minimum of f .
- (b) There does not exist any interval (q_1, q_2) with $q_1, q_2 \in \mathbb{Q}_+$ such that $(q_1, q_2) \subset \mathcal{L}_f$.
- (c) For all $(l, r) \in \mathcal{E}_f$ with $r < \infty$, there exists $\varepsilon_0 = \varepsilon_0(l, r) > 0$ such that the following holds for all $\varepsilon \in (0, \varepsilon_0)$: There exists $\delta = \delta(\varepsilon, l, r) > 0$ such that

$$f(t) > f(r) + \delta \quad \forall t \in (l + \varepsilon, r - \varepsilon). \quad (3.1)$$

- (d) f does not have any infinite excursion, i.e., $\phi_1(f) < \infty$.
- (e) For any $\delta > 0$, f has only finitely many excursions of length at least δ .
- (f) For all $i \geq 1$, $\phi_{i+1}(f) < \phi_i(f)$.

Lemma 1. *Suppose that $f \in \mathbb{D}_+[0, \infty)$ is good. Further, let $(f_n)_{n \geq 1} \subset \mathbb{D}[0, \infty)$ be such that $f_n \rightarrow f$ in $\mathbb{D}[0, \infty)$. Moreover, let $\limsup_{n \rightarrow \infty} \phi_1(f_n) < \infty$, and if $z_n(T)$ denotes the length of the largest excursion of f_n starting after T , then $\lim_{T \rightarrow \infty} \limsup_{n \rightarrow \infty} z_n(T) = 0$. Then, for all $m \geq 1$, as $n \rightarrow \infty$,*

$$(\phi_i(f_n))_{i \in [m]} \rightarrow (\phi_i(f))_{i \in [m]}, \quad \text{and} \quad (\mathcal{A}_i(f_n))_{i \in [m]} \rightarrow (\mathcal{A}_i(f))_{i \in [m]}.$$

Proof. The proof here is for $m = 1$, and for $m > 1$, we can proceed inductively. Using Definitions 1(d), (e), as well as the assumptions of the lemma, we can take $T > 0$ and $n_0 \geq 1$ large so that the largest excursions of f_n and f end before T for all $n \geq n_0$. Let \mathcal{L} denote the set of continuous functions $\Lambda : [0, \infty) \rightarrow [0, \infty)$ that are strictly increasing and satisfy $\Lambda(0) = 0$, $\Lambda(T) = T$. Suppose (l, r) is the longest excursion of f on $[0, T]$, and thus $\phi_1(f) = r - l$. We will first show that $\lim_{n \rightarrow \infty} \phi_1(f_n) = \phi_1(f)$.

Fix $\varepsilon, \delta > 0$ such that (3.1) holds. Let $\|\cdot\|_T$ denote the sup-norm on $[0, T]$. Recall the definition of the metric for Skorohod J_1 -topology from [14, (12.13)]. Since $f_n \rightarrow f$ in $\mathbb{D}[0, T]$, there exists $(\Lambda_n)_{n \geq 1} \subset \mathcal{L}$, and $n_1 \geq n_0$ such that for all $n \geq n_1$,

$$\|f_n \circ \Lambda_n - f\|_T < \frac{\delta}{2} \quad \text{and} \quad \|\Lambda_n - I\|_T < \varepsilon, \quad (3.2)$$

where I is the identity function. Using (3.1) and (3.2), for all $t \in (l + \varepsilon, r - \varepsilon)$ and $n \geq n_1$,

$$f_n \circ \Lambda_n(t) > \underline{f}(t) - \frac{\delta}{2} > f(r) + \frac{\delta}{2} = \underline{f}(r) + \frac{\delta}{2},$$

where the last equality is due to $r \in \mathcal{Z}_f$ from Fact 1. Thus, using $\|\Lambda_n - I\|_T < \varepsilon$ from (3.2),

$$f_n(t) > \underline{f}(r) + \frac{\delta}{2} \quad \forall t \in (l + 2\varepsilon, r - 2\varepsilon). \tag{3.3}$$

Next, note that the infimum operation is continuous in the Skorohod J_1 -topology [56, Theorem 13.4.1], and thus $\underline{f}_n \rightarrow \underline{f}$ in $\mathbb{D}[0, T]$. Moreover, using (2.2), $\underline{f} \in \mathbb{C}[0, T]$, and therefore, there exists $n_2 \geq n_0$, such that for all $n \geq n_2$

$$\|\underline{f}_n - \underline{f}\|_T < \frac{\delta}{4}. \tag{3.4}$$

Using $\underline{f}(t) = \underline{f}(r)$ for all $t \in [l, r]$, this implies that, for all $n \geq n_2$,

$$\underline{f}(r) = \underline{f}(t) > \underline{f}_n(t) - \frac{\delta}{4} \quad \forall t \in (l + 2\varepsilon, r - 2\varepsilon),$$

and consequently (3.3) yields that for all $n \geq \max\{n_1, n_2\}$

$$f_n(t) - \underline{f}_n(t) > \frac{\delta}{4} \quad \forall t \in (l + 2\varepsilon, r - 2\varepsilon). \tag{3.5}$$

Thus,

$$\liminf_{n \rightarrow \infty} \phi_1(f_n) \geq r - l - 4\varepsilon = \phi_1(f) - 4\varepsilon, \tag{3.6}$$

which provides the required lower bound. We now turn to a suitable upper bound on the quantity $\limsup_{n \rightarrow \infty} \phi_1(f_n)$. We claim that, using Definition 1(b), one can find $r_1, \dots, r_k \in \mathcal{R}_f$ such that $r_1 \leq \phi_1(f) + \varepsilon$, $T - r_k < \phi_1(f) + \varepsilon$, and $r_i - r_{i-1} \leq \phi_1(f) + \varepsilon$, $\forall i = 2, \dots, k$. Indeed, since $\phi_1(f)$ is the largest excursion length of f , if there is no excursion end-point in between 0 and $\phi_1(f) + \varepsilon$, then there is no excursion begin-point in $[0, \varepsilon)$. The latter shows that the interval $[0, \varepsilon)$ is contained in $\mathcal{Z}_f = \{t : f(t) = \underline{f}(t)\}$, which contradicts Definition 1(b). The existence of the points r_2, \dots, r_k can be shown inductively using similar argument as above. Let l_1, \dots, l_k be the excursion begin-points corresponding to the endpoints r_1, \dots, r_k . We will show that, for all i , f_n will have an excursion begin-point in $(l_i - 4\varepsilon, l_i + 2\varepsilon)$ and end-point in $(r_i - 3\varepsilon, r_i + 2\varepsilon)$, so that the largest excursion of f_n is contained inside one of the intervals $(l_i - 4\varepsilon, r_i + 2\varepsilon)$ for $i \in [k]$.

Using Definition 1(a), r_i is not a local minimum, and thus for any $\varepsilon > 0$ (sufficiently small), there exists $\delta > 0$ and $t_i \in (r_i, r_i + \varepsilon)$ such that $f(r_i) - f(t_i) > \delta$. We also let $\delta > 0$ be sufficiently small such that (3.2) holds. Thus, using (3.2), for all $n \geq n_1$,

$$f(r_i) - f_n(\Lambda_n(t_i)) \geq f(r_i) - f(t_i) - \frac{\delta}{2} > \frac{\delta}{2}.$$

Since $t_i \in (r_i, r_i + \varepsilon)$, we have that $t_i^n = \Lambda_n(t_i) \in (r_i - \varepsilon, r_i + 2\varepsilon)$. Thus, for all $n \geq n_1$, there exists a point $t_i^n \in (r_i - \varepsilon, r_i + 2\varepsilon)$ such that

$$f(r_i) - f_n(t_i^n) > \frac{\delta}{2}. \quad (3.7)$$

Next, using (3.4),

$$\underline{f}_n(r_i - 3\varepsilon) \rightarrow \underline{f}(r_i - 3\varepsilon) \geq \underline{f}(r_i) = f(r_i), \quad (3.8)$$

since $r_i \in \mathcal{L}_f$. Combining (3.7) and (3.8), we see that $f_n(r_i - 3\varepsilon) > f_n(t_i^n)$, and by (3.5), we also have that $f_n(r_i - 3\varepsilon) - \underline{f}_n(r_i - 3\varepsilon) > 0$, when $\varepsilon > 0$ is so small that $r_i - 3\varepsilon > l_i + 2\varepsilon$. Thus f_n must have an excursion end-point in $(r_i - 3\varepsilon, r_i + 2\varepsilon)$. Also, using Definition 1(b), f has an excursion end-point $r_i^0 \in (l_i - \varepsilon, l_i)$. The previous argument shows that f_n has to have an excursion end-point $r_i^{0n} \in (r_i^0 - 3\varepsilon, r_i^0 + 2\varepsilon)$, and thus by (3.5), f_n must have an excursion begin-point in $(r_i^{0n}, l_i + 2\varepsilon) \subset (l_i - 4\varepsilon, l_i + 2\varepsilon)$. Therefore,

$$\limsup_{n \rightarrow \infty} \phi_1(f_n) \leq \max_{i \in [k]} (r_i - l_i) + 6\varepsilon \leq \max_{i \in [k]} (r_i - r_{i-1}) + 6\varepsilon \leq \phi_1(f) + 7\varepsilon. \quad (3.9)$$

Hence, the convergence of the largest excursion length follows from (3.6) and (3.9).

Next, we show that $\lim_{n \rightarrow \infty} \mathcal{A}_1(f_n) = \mathcal{A}_1(f)$. Let $e = (l, r)$ be the largest excursion of f . Using (3.5), the interval $(l - 2\varepsilon, r + 2\varepsilon)$ is part of some excursion of f_n . Let us denote this excursion by $e_n = (L_n(e), R_n(e))$. We will show that e_n is the largest excursion of f_n when n is large. Indeed, the arguments above already show that

$$l - 4\varepsilon \leq L_n(e) \leq l + 2\varepsilon, \quad \text{and} \quad r - 3\varepsilon \leq R_n(e) \leq r + 2\varepsilon,$$

and thus $R_n(e) - L_n(e) \geq r - l - 5\varepsilon$. Now, using Definition 1(f), we can take $\varepsilon > 0$ sufficiently small such that $\phi_2(f_n) < r - l - 5\varepsilon$ for all sufficiently large n . Thus, $e_n = (L_n(e), R_n(e))$ must be the largest excursion of f_n . The convergence of $\mathcal{A}_1(f_n)$ follows by using $f_n \rightarrow f$ in $\mathbb{D}[0, \infty)$ together with $L_n(e) \rightarrow l$ and $R_n(e) \rightarrow r$ as $n \rightarrow \infty$. \square

Remark 5. We emphasize that the strict ordering between excursion lengths in Definition 1(f) is only used in the convergence of $\mathcal{A}_i(f_n)$. This ensures that the location of largest excursions of f_n and f approximately coincide, which is strictly stronger than requiring the convergence of excursion lengths.

Next, we define what it means for a stochastic process $\mathbf{X} \in \mathbb{D}_+[0, \infty)$ to be good:

Definition 2 (*Good stochastic process*). A stochastic process \mathbf{X} with sample paths in $\mathbb{D}_+[0, \infty)$ is said to be *good* if the sample path satisfies all the conditions of Definition 1 almost surely.

The following is a direct consequence of Lemma 1:

Proposition 1. *Consider a sequence of stochastic processes $(\mathbf{X}_n)_{n \geq 1}$ and a good stochastic process \mathbf{X} such that $\mathbf{X}_n \xrightarrow{d} \mathbf{X}$. Also, let $(\phi_1(\mathbf{X}_n))_{n \geq 1}$ be tight, and if $Z_n(T)$ denotes the length of the largest excursion of \mathbf{X}_n starting after time T , then assume that, for any $\varepsilon > 0$, $\lim_{T \rightarrow \infty} \limsup_{n \rightarrow \infty} \mathbb{P}(Z_n(T) > \varepsilon) = 0$. Then, for all $m \geq 1$,*

$$(\phi_i(\mathbf{X}_n), \mathcal{A}_i(\mathbf{X}_n))_{i \in [m]} \xrightarrow{d} (\phi_i(\mathbf{X}), \mathcal{A}_i(\mathbf{X}))_{i \in [m]}.$$

3.2. *The limiting process is good almost surely.* In this section, we will show that the sample paths of $\mathbf{S}_\infty^\lambda$ are good almost surely. Throughout this section, we assume without loss of generality that $\mu = 1$ and $\sum_i \theta_i^2 = 1$ to simplify writing. An identical proof works for the general μ and θ by replacing λ with $\lambda' = \lambda\mu / \sum_i \theta_i^2$. Consider the sigma-field $\mathcal{F}_t = \sigma(\{\xi_i \leq s\} : s \leq t, i \geq 1)$, where $(\xi_i)_{i \geq 1}$ are the exploration random variables used in the definition of $\mathbf{S}_\infty^\lambda$ in (2.4), and, for a collection of sets \mathcal{A} , $\sigma(\mathcal{A})$ denotes the minimum sigma-algebra containing all the sets in \mathcal{A} . Then $(\mathcal{F}_t)_{t \geq 0}$ is a filtration and $\mathbf{S}_\infty^\lambda$ is adapted to $(\mathcal{F}_t)_{t \geq 0}$. Our goal is stated formally in the following proposition:

Proposition 2. *The sample paths of $\mathbf{S}_\infty^\lambda$ satisfy the conditions of Definition 1 almost surely.*

Proof of Proposition 2. The verification of each of the conditions in Definition 1 are given separately below.

Verification of Definition 1(a). Let $q \in \mathbb{Q}_+$ and define the random time $T_q = \inf\{t \geq q : S_\infty^\lambda(t) = \inf_{u \leq q} S_\infty^\lambda(u)\}$. We will show that, almost surely,

$$\inf\{t > 0 : S_\infty^\lambda(T_q + t) - S_\infty^\lambda(T_q) < 0\} = 0, \quad \text{on } \{T_q < \infty\}, \text{ for all } q \in \mathbb{Q}_+ \tag{3.10}$$

Note that if q lies in the interior of some finite-length excursion then $T_q \in (q, \infty)$, and also T_q is the end-point of that excursion. Therefore, $\{T_q : q \in \mathbb{Q}_+ \text{ and } T_q < \infty\}$ contains the set of excursion end-points of $\mathbf{S}_\infty^\lambda$. Now, (3.10) ensures that T_q is not a local minimum because we can find u arbitrarily close to T_q such that $S_\infty^\lambda(u) < S_\infty^\lambda(T_q)$. Hence, Definition 1(a) holds for $\mathbf{S}_\infty^\lambda$ almost surely.

Thus it suffices to prove (3.10). Since \mathbb{Q}_+ is countable, it is enough to prove (3.10) for each fixed $q \in \mathbb{Q}_+$. Let $V_q = \{i : \mathcal{I}_i(T_q) = 1\}$. Note that T_q is a stopping time. Moreover, conditionally on the sigma-field \mathcal{F}_{T_q} , the process $(S_\infty^\lambda(T_q + t) - S_\infty^\lambda(T_q))_{t \geq 0}$ is distributed as $\hat{\mathbf{S}}_\infty^\lambda$ given by

$$\hat{\mathbf{S}}_\infty^\lambda(t) = \lambda \sum_{i \notin V_q} \theta_i \mathcal{I}_i(t) - t. \tag{3.11}$$

Define $L(t) = \lambda \sum_{i=1}^\infty \theta_i \mathcal{N}_i(t) - t$, where $(\mathcal{N}_i(t))_{t \geq 0}$ is a rate- θ_i Poisson process, independently for different i . We assume that $\hat{\mathbf{S}}_\infty^\lambda$ and L are coupled by taking $\mathcal{I}_i(s) = \mathbb{1}\{\mathcal{N}_i(s) \geq 1\}$, so that $\hat{\mathbf{S}}_\infty^\lambda(t) \leq L(t)$ for all $t \geq 0$ almost surely. Thus, if $R_0 = \inf\{t > 0 : L(t) < 0\}$, then it suffices to show that

$$\mathbb{P}(R_0 = 0) = 1, \tag{3.12}$$

and (3.10) follows. Fix $\varepsilon > 0$ and $K \geq 1$. Then,

$$\begin{aligned} \mathbb{P}(R_0 \leq \varepsilon) &\geq \mathbb{P}(L(\varepsilon) < 0) \\ &\geq \mathbb{P}\left(\lambda \sum_{i=K+1}^\infty \theta_i \mathcal{N}_i(\varepsilon) < \varepsilon, \text{ and } \mathcal{N}_i(\varepsilon) = 0, \forall i \in [K]\right) \\ &= \prod_{i=1}^K \mathbb{P}(\mathcal{N}_i(\varepsilon) = 0) \times \mathbb{P}\left(\lambda \sum_{i=K+1}^\infty \theta_i \mathcal{N}_i(\varepsilon) < \varepsilon\right) \\ &= e^{-\varepsilon \sum_{i=1}^K \theta_i} \left(1 - \mathbb{P}\left(\lambda \sum_{i=K+1}^\infty \theta_i \mathcal{N}_i(\varepsilon) \geq \varepsilon\right)\right) \\ &\geq e^{-\varepsilon \sum_{i=1}^K \theta_i} \left(1 - \frac{\lambda}{\varepsilon} \mathbb{E}\left[\sum_{i=K+1}^\infty \theta_i \mathcal{N}_i(\varepsilon)\right]\right) = e^{-\varepsilon \sum_{i=1}^K \theta_i} \left(1 - \lambda \sum_{i=K+1}^\infty \theta_i^2\right), \end{aligned} \tag{3.13}$$

where the one-but-last step follows from Markov's inequality. Thus, using the fact that $\{R_0 \leq \varepsilon\} \searrow \{R_0 = 0\}$, as $\varepsilon \searrow 0$,

$$\mathbb{P}(R_0 = 0) = \lim_{\varepsilon \searrow 0} \mathbb{P}(R_0 \leq \varepsilon) \geq 1 - \lambda \sum_{i=K+1}^{\infty} \theta_i^2,$$

and since the above holds for any $K \geq 1$, and $\sum_i \theta_i^2 < \infty$, we have proved (3.12).

Verification of Definition 1(b). Next, we verify that Definition 1(b) holds almost surely for $\mathbf{S}_{\infty}^{\lambda}$. Since \mathbb{Q}_+ is countable, we may again work with fixed $q_1, q_2 \in \mathbb{Q}_+$, i.e., it suffices to prove that $(q_1, q_2) \not\subset \{t : S_{\infty}^{\lambda}(t) = \inf_{u \leq t} S_{\infty}^{\lambda}(u)\}$ almost surely. By the description of our thinned Lévy process, it has positive jumps only, and if there is a jump of size θ_i at time t , then $S_{\infty}^{\lambda}(t + \theta_i/2) > \inf_{u \leq t} S_{\infty}^{\lambda}(u) = \inf_{u \leq t + \theta_i/2} S_{\infty}^{\lambda}(u)$. Therefore, if $(q_1, q_2) \subset \{t : S_{\infty}^{\lambda}(t) = \inf_{u \leq t} S_{\infty}^{\lambda}(u)\}$, then there is no ξ_i such that $\xi_i \in (q_1, q_2)$. We compute

$$\begin{aligned} \mathbb{P}(\forall i \geq 1 : \xi_i \notin (q_1, q_2)) &= \prod_{i=1}^{\infty} \mathbb{P}(\xi_i \notin (q_1, q_2)) = \prod_{i=1}^{\infty} (1 - e^{-\theta_i q_1} + e^{-\theta_i q_2}) \\ &= \exp\left(\sum_{i=1}^{\infty} \log\left(1 - e^{-\theta_i q_1} (1 - e^{-\theta_i (q_2 - q_1)})\right)\right) \\ &\leq \exp\left(-e^{-\theta_1 q_1} \sum_{i=1}^{\infty} (1 - e^{-\theta_i (q_2 - q_1)})\right) = 0, \end{aligned}$$

where the one-but-last step follows using $\log(1-x) \leq -x$ for all $x \in (0, 1)$ and $e^{-\theta_i q_1} \geq e^{-\theta_1 q_1}$ for all $i \geq 1$, and the last step uses the fact that $\sum_{i=1}^{\infty} (1 - e^{-\theta_i (q_2 - q_1)}) = \infty$, which follows by applying the limit comparison test together with $(1 - e^{-\theta_i (q_2 - q_1)})/\theta_i \rightarrow q_2 - q_1$ as $i \rightarrow \infty$, and $\sum_{i=1}^{\infty} \theta_i = \infty$. Thus we have verified that Definition 1(b) holds almost surely for $\mathbf{S}_{\infty}^{\lambda}$.

Verification of Definition 1(c). Similarly as above, for any $q \in \mathbb{Q}_+$, define the stopping time $T_q(\varepsilon) = \inf\{t \geq q : S_{\infty}^{\lambda}(t) \leq \inf_{u \leq q} S_{\infty}^{\lambda}(u) + \varepsilon\}$. Thus, $T_q(\varepsilon) > q$ if $S_{\infty}^{\lambda}(q) > \inf_{u \leq q} S_{\infty}^{\lambda}(u) + \varepsilon$ and $T_q(\varepsilon) = q$ otherwise. Observe that $T_q(\varepsilon) < \infty$ almost surely since there are no infinite excursions. We claim that it is sufficient to prove

$$\lim_{\varepsilon \searrow 0} \mathbb{P}(\mathbf{S}_{\infty}^{\lambda} \text{ has an excursion end-point in } (T_q(\varepsilon), T_q(\varepsilon) + 2\varepsilon), \text{ and } T_q(\varepsilon) < \infty) = 1. \quad (3.14)$$

Let $T_q^- := \inf\{t > q : S_{\infty}^{\lambda}(t-) = \inf_{u \leq q} S_{\infty}^{\lambda}(u)\}$. Indeed, if q lies inside some excursion, i.e., $S_{\infty}^{\lambda}(q) > \inf_{u \leq q} S_{\infty}^{\lambda}(u)$, then $T_q(\varepsilon) \nearrow T_q^-$ as $\varepsilon \searrow 0$, and (3.14) shows that T_q^- must be an excursion end-point with probability 1. Now, if $\mathbf{S}_{\infty}^{\lambda}$ contains an excursion (l, r) having a point $t \in (l, r)$ such that $S_{\infty}^{\lambda}(t-) = \inf_{u \leq t} S_{\infty}^{\lambda}(u) = \inf_{u \leq t} S_{\infty}^{\lambda}(u)$, then there exists some $q \in \mathbb{Q}_+ \cap (l, r)$ such that T_q^- is not an excursion endpoint. The later has zero probability as we showed above. This completes the verification of Definition 1(c).

It remains to prove (3.14). As before, let $L(t) = \lambda \sum_{i=1}^{\infty} \theta_i \mathcal{N}_i(t) - t$, and let us also work under the coupling under which $S_{\infty}^{\lambda}(T_q(\varepsilon) + t) - S_{\infty}^{\lambda}(T_q(\varepsilon)) \leq L(t)$ for all $t \geq 0$ almost surely. We have $S_{\infty}^{\lambda}(T_q(\varepsilon)) \leq \inf_{u \leq q} S_{\infty}^{\lambda}(u) + \varepsilon$ on $\{T_q(\varepsilon) < \infty\}$, since the process has only positive jumps. Also, if $L(2\varepsilon) < -\varepsilon$, then $S_{\infty}^{\lambda}(T_q(\varepsilon) + 2\varepsilon) < S_{\infty}^{\lambda}(T_q(\varepsilon)) - \varepsilon \leq \inf_{u \leq q} S_{\infty}^{\lambda}(u)$, and consequently $\inf_{u \leq T_q(\varepsilon) + 2\varepsilon} S_{\infty}^{\lambda}(u) <$

$\inf_{u \leq T_q(\varepsilon)} S_\infty^\lambda(u) = \inf_{u \leq q} S_\infty^\lambda(u)$. Therefore, the event in (3.14) holds. Thus, using identical computations as (3.13), it follows that

$$\begin{aligned} & \mathbb{P}(S_\infty^\lambda \text{ has an excursion end-point in } (T_q(\varepsilon), T_q(\varepsilon) + 2\varepsilon), \text{ and } T_q(\varepsilon) < \infty) \\ & \geq \mathbb{P}(L(2\varepsilon) < -\varepsilon) = \mathbb{P}\left(\lambda \sum_{i=1}^\infty \theta_i \mathcal{N}_i(2\varepsilon) < \varepsilon\right) \geq e^{-2\varepsilon \sum_{i=1}^K \theta_i} \left(1 - 2\lambda \sum_{i=K+1}^\infty \theta_i^2\right), \end{aligned}$$

and (3.14) follows by taking the iterated limit $\lim_{K \rightarrow \infty} \lim_{\varepsilon \rightarrow 0}$, and using $\sum_i \theta_i^2 < \infty$. *Verification of Definition 1(d)*. We start by providing the martingale decomposition for S_∞^λ :

Lemma 2. *The process S_∞^λ admits the Doob-Meyer decomposition $S_\infty^\lambda(t) = M(t) + A(t)$ with the drift term $A(t)$ and the quadratic variation for the martingale term $\langle M \rangle(t)$ given by*

$$A(t) = \lambda \sum_{i=1}^\infty \theta_i^2 \min\{\xi_i, t\} - t, \quad \langle M \rangle(t) = \lambda^2 \sum_{i=1}^\infty \theta_i^3 \min\{\xi_i, t\}.$$

Proof. Define $M_i(t) = \mathbb{1}_{\{\xi_i \leq t\}} - \theta_i \min\{\xi_i, t\}$. Then

$$(M_i(t))_{t \geq 0} \text{ is a martingale.} \tag{3.15}$$

Indeed, note that $M_i(t+s) - M_i(t) = 0$ if $\xi_i \leq t$. Thus,

$$\begin{aligned} \mathbb{E}[M_i(t+s) - M_i(t) \mid \mathcal{F}_t] &= \mathbb{E}[\mathbb{1}_{\{t < \xi_i \leq t+s\}} - \theta_i(\min\{\xi_i, t+s\} - \min\{\xi_i, t\}) \mid \xi_i > t] \\ &= \mathbb{E}[\mathbb{1}_{\{t < \xi_i \leq t+s\}} - \theta_i \min\{\xi_i - t, s\} \mid \xi_i > t] \\ &= \mathbb{P}(0 < \xi_i \leq s) - \theta_i \mathbb{E}[\min\{\xi_i, s\}], \end{aligned} \tag{3.16}$$

where the last step follows from the memoryless property of the exponential distributions. Now, using the fact that $\int x e^{-ax} dx = -e^{-ax}(ax + 1)/a^2$, one can verify that $\theta_i \mathbb{E}[\min\{\xi_i, s\}] = 1 - e^{-\theta_i s}$. Applying this to (3.16), we can conclude that $\mathbb{E}[M_i(t+s) - M_i(t) \mid \mathcal{F}_t] = 0$, thus verifying (3.15). Moreover, the quadratic variation of $(M_i(t))_{t \geq 0}$ is given by

$$\langle M_i \rangle(t) = \theta_i \min\{\xi_i, t\}. \tag{3.17}$$

This follows from the characterization of unit-jump processes given in [52, Lemma 3.1], together with the fact that $\theta_i \min\{\xi_i, t\}$, the compensator of $\mathbb{1}_{\{\xi_i \leq t\}}$, is continuous in t . Then (3.15) and (3.17) completes the proof of Lemma 2. \square

We are now ready to verify Definition 1(d). In order to prove that S_∞^λ does not have an excursion of infinite length almost surely, it suffices to show that

$$\lim_{t \rightarrow \infty} S_\infty^\lambda(t) = -\infty \text{ almost surely.} \tag{3.18}$$

Fix $K \geq 1$ such that $\lambda \sum_{i > K} \theta_i^2 < 1/2$. Such a choice of K is always possible as $\theta \in \ell^2_\downarrow$. Further define the stopping time $T := \inf\{t : \xi_i \leq t, \forall i \in [K]\} = \max_{i \leq K} \xi_i$. Thus, $T < \infty$ almost surely. Note that $\min\{\xi_i, t\} \leq t$ and thus,

$$\frac{1}{t} \lambda \sum_{i > K} \theta_i^2 \min\{\xi_i, t\} < \frac{1}{2}, \text{ almost surely.}$$

Therefore, for any $t > T$,

$$A(t) = \lambda \sum_{i \in [K]} \theta_i^2 \xi_i + \lambda \sum_{i > K} \theta_i^2 \min\{\xi_i, t\} - t < \lambda \sum_{i \in [K]} \theta_i^2 \xi_i - \frac{t}{2}, \quad \text{almost surely.}$$

We conclude that, for any $r \in (0, 1)$, $t^{-r} A(t) \xrightarrow{\text{a.s.}} -\infty$. For the martingale part we will use the exponential concentration inequality [55, Inequality 1, p. 899], which is stated below:

Lemma 3. *If M is any continuous time local martingale such that $M(0) = 0$, and $\sup_{t \in [0, \infty)} |M(t) - M(t-)| \leq c$, almost surely, then for any $t > 0$, $a > 0$ and $b > 0$,*

$$\mathbb{P}\left(\sup_{s \in [0, t]} M(s) > a, \text{ and } \langle M \rangle(t) \leq b\right) \leq \exp\left(-\frac{a^2}{2b} \psi\left(\frac{ac}{b}\right)\right),$$

where $\psi(x) = ((1+x) \log(1+x) - x)/x^2$.

In particular, $\psi(x) \geq 1/(2(1+x/3))$ (see [42, p. 27]). Note that $\langle M \rangle(t) \leq \lambda^2 t \sum_{i=1}^{\infty} \theta_i^3$. We apply Lemma 3 with $a = \varepsilon t^r$, $b = \lambda^2 t \sum_{i=1}^{\infty} \theta_i^3$, and $c = \theta_1$. Using Lemma 2, $\langle M \rangle(t) \leq b$ almost surely. Now, $\psi(ac/b) \geq C/(1+t^{r-1})$, and thus for any $\varepsilon > 0$, and $r \in (1/2, 1)$

$$\mathbb{P}\left(\sup_{s \in [0, t]} |M(s)| > \varepsilon t^r\right) \leq 2 \exp(-Ct^{2r-1}),$$

for some constant $C > 0$, where the bound on the absolute value of M follows from the fact that $-M$ is also a martingale, so Lemma 3 applies to $-M$ as well. Now an application of the Borel–Cantelli lemma proves that $t^{-r} |M(t)| \xrightarrow{\text{a.s.}} 0$, for any $r \in (1/2, 1)$. This fact, together with the asymptotics of the drift term, completes the proof of (3.18). \square
Verification of Definition 1(e). Fix $\delta > 0$. Let $t_k = (k-1)\delta/2$ and define the event

$$C_k^\delta := \left\{ \sup_{t \in (t_{k-1}, t_k]} S_\infty^\lambda(t_{k+1}) - S_\infty^\lambda(t) > 0 \right\}.$$

Suppose that there is an excursion (l, r) with $r - l > \delta$ and $l \in (t_{k-1}, t_k]$ for some k . Since $r > t_{k+1}$ and $l \in (t_{k-1}, t_k]$, we have that $\inf_{u \leq t_{k+1}} S_\infty^\lambda(u) = \inf_{u \in (t_{k-1}, t_k]} S_\infty^\lambda(u)$. Consequently, $S_\infty^\lambda(t_{k+1}) > \inf_{t \in (t_{k-1}, t_k]} S_\infty^\lambda(t)$, and therefore C_k^δ must occur. Therefore, if S_∞^λ has infinitely many excursions of length at least δ , then C_k^δ must occur infinitely often. Using the Borel–Cantelli lemma, the proof follows if we can show that

$$\sum_{k=1}^{\infty} \mathbb{P}(C_k^\delta) < \infty.$$

As before, fix $K \geq 1$ such that $\lambda \sum_{i > K} \theta_i^2 < 1/2$, and let $T := \inf\{t : \xi_i \leq t, \forall i \in [K]\} = \max_{i \leq K} \xi_i$. Notice that for each $K \geq 1$,

$$\sum_{k=1}^{\infty} \mathbb{P}(T > t_{k-1}) = \sum_{k=1}^{\infty} \mathbb{P}(\exists i \in [K] : \xi_i > t_{k-1}) \leq \sum_{k=1}^{\infty} K e^{-\theta_K (k-1)\delta/2} < \infty,$$

and therefore it is enough to show that

$$\sum_{k=1}^{\infty} \mathbb{P}(C_k^\delta \cap \{T \leq t_{k-1}\}) < \infty. \quad (3.19)$$

Now,

$$\begin{aligned} \sup_{t \in [t_{k-1}, t_k]} [S_\infty^\lambda(t_{k+1}) - S_\infty^\lambda(t)] &\leq M(t_{k+1}) + \sup_{t \in [t_{k-1}, t_k]} -M(t) + \sup_{t \in [t_{k-1}, t_k]} [A(t_{k+1}) - A(t)] \\ &\leq M(t_{k+1}) - M(t_{k-1}) + \sup_{t \in [t_{k-1}, t_k]} [M(t_{k-1}) - M(t)] \\ &\quad + \sup_{t \in [t_{k-1}, t_k]} \left[\lambda \sum_{i=1}^{\infty} \theta_i^2 (\min\{\xi_i, t_{k+1}\} - \min\{\xi_i, t\}) - (t_{k+1} - t) \right] \\ &\leq 2 \sup_{t \in [t_{k-1}, t_{k+1}]} |M(t) - M(t_{k-1})| \\ &\quad + \sup_{t \in [t_{k-1}, t_k]} \left[\lambda \sum_{i=1}^{\infty} \theta_i^2 (\min\{\xi_i, t_{k+1}\} - \min\{\xi_i, t\}) - (t_{k+1} - t) \right]. \end{aligned}$$

On the event $\{T \leq t_{k-1}\}$, the second term inside the supremum above reduces to

$$\lambda \sum_{i>K} \theta_i^2 (\min\{\xi_i, t_{k+1}\} - \min\{\xi_i, t\}) - (t_{k+1} - t) \leq (t_{k+1} - t) \lambda \sum_{i>K} \theta_i^2 - (t_{k+1} - t) < -\frac{\delta}{2}, \quad (3.20)$$

using $\lambda \sum_{i>K} \theta_i^2 < 1/2$. Thus we only need to estimate

$$\mathbb{P}\left(\sup_{t \in [t_{k-1}, t_{k+1}]} |M(t) - M(t_{k-1})| > \frac{\delta}{4} \right).$$

Note that $(M(t) - M(t_{k-1}))_{t \geq t_{k-1}}$ is a martingale with respect to the filtration $(\mathcal{F}_t)_{t \geq t_{k-1}}$ starting from zero. Moreover, using an identical argument as Lemma 2 yields that the quadratic variation of $(M(t) - M(t_{k-1}))_{t \geq t_{k-1}}$ is given by

$$\lambda^2 \sum_{i=1}^{\infty} \theta_i^3 (\min\{\xi_i, t\} - \min\{\xi_i, t_{k-1}\}).$$

Further, $\mathbb{E}[\min\{\xi_i, t\}] = \theta_i^{-1}(1 - e^{-\theta_i t})$. Therefore, Doob's martingale inequality [45, Theorem 1.9.1.3] implies

$$\begin{aligned} &\sum_{k=1}^{\infty} \mathbb{P}\left(\sup_{t \in [t_{k-1}, t_{k+1}]} |M(t) - M(t_{k-1})| > \frac{\delta}{4} \right) \\ &\leq \sum_{k=1}^{\infty} \frac{16\lambda^2}{\delta^2} \sum_{i=1}^{\infty} \theta_i^2 (e^{-\theta_i t_{k-1}} - e^{-\theta_i t_{k+1}}) = \frac{16\lambda^2}{\delta^2} \sum_{i=1}^{\infty} \theta_i^2 (1 - e^{-\theta_i \delta}) \sum_{k=1}^{\infty} e^{-\theta_i t_{k-1}} < \infty, \end{aligned}$$

and the proof of (3.19) now follows using (3.20).

Verification of Definition 1(f). We first prove the following:

Lemma 4. *The distribution of $S_\infty^\lambda(t)$ has no atoms for all $t > 0$.*

Proof. Let $\phi_t(v) = \mathbb{E}[e^{ivS(t)}]$ for $v \in \mathbb{R}$. Using the sufficient condition for random variables to have non-atomic distribution stated in [33, p. 189], it suffices to prove that

$$\int_{-\infty}^{\infty} |\phi_t(v)| dv < \infty.$$

Note that

$$\begin{aligned} \phi_t(v) &= e^{-ivt} \prod_{j=1}^{\infty} \mathbb{E}[e^{iv\lambda\theta_j} \mathbb{1}\{\xi_j \leq t\}] = e^{-ivt} \prod_{j=1}^{\infty} (e^{iv\lambda\theta_j} (1 - e^{-\theta_j t}) + e^{-\theta_j t}) \\ &= e^{-ivt} \prod_{j=1}^{\infty} ((1 - e^{-t\theta_j}) \cos(v\lambda\theta_j) + e^{-t\theta_j} + i(1 - e^{-t\theta_j}) \sin(v\lambda\theta_j)). \end{aligned}$$

Therefore,

$$\begin{aligned} |\phi_t(v)|^2 &= \prod_{j=1}^{\infty} \left(((1 - e^{-t\theta_j}) \cos(v\lambda\theta_j) + e^{-t\theta_j})^2 + (1 - e^{-t\theta_j})^2 \sin^2(v\lambda\theta_j) \right) \\ &= \prod_{j=1}^{\infty} \left(e^{-2t\theta_j} + 2 \cos(v\lambda\theta_j) e^{-t\theta_j} (1 - e^{-t\theta_j}) + (1 - e^{-t\theta_j})^2 \right) \\ &= \prod_{j=1}^{\infty} \left(1 - 2e^{-t\theta_j} (1 - e^{-t\theta_j}) (1 - \cos(v\lambda\theta_j)) \right) \\ &\leq e^{-\sum_{j=1}^{\infty} 2e^{-t\theta_j} (1 - e^{-t\theta_j}) (1 - \cos(v\lambda\theta_j))}, \end{aligned}$$

where in the last step we have used the fact that $1 - x \leq e^{-x}$ for all $x > 0$. Recall (2.5). Let $j_0 \geq 1$ be such that $\max\{(2|v|\lambda/\pi)\theta_j, t\theta_j\} \leq 1$ for all $j \geq j_0$. Now, for $j \geq j_0$, we have that $e^{-t\theta_j} \geq e^{-1}$, $(1 - e^{-t\theta_j}) \geq t\theta_j/2$ and $1 - \cos(v\lambda\theta_j) \geq \frac{\lambda^2}{\pi} v^2 \theta_j^2$. Thus, using (2.5),

$$\int_{-\infty}^{\infty} |\phi_t(v)| dv \leq \int_{-\infty}^{\infty} e^{-\frac{\lambda^2 t}{e\pi} v^2 M_t(2|v|\lambda/\pi)} dv < \infty,$$

and the proof now follows. \square

In order to prove the strict ordering between excursion lengths, it is enough to show that no two excursions of S_∞^λ have the same length almost surely. For any $q \in \mathbb{Q}_+$, if q is in some excursion of S_∞^λ , i.e., if $S_\infty^\lambda(q) > \inf_{t \leq q} S_\infty^\lambda(t)$, then we let $e(q)$ be the excursion containing q , and otherwise we let $e(q) = \emptyset$. Thus it is enough to show that for any $q_1, q_2 \in \mathbb{Q}_+$,

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

Without loss of generality, let $q_1 < q_2$. Thus, if $e(q_1) \neq e(q_2)$, then $e(q_1)$ appears earlier than $e(q_2)$. Let $V_{q_2} = \{i : \mathcal{S}_i(q_2) = 1\}$. As before, conditionally on \mathcal{F}_{q_2} , the process $(S_\infty^\lambda(q_2 + t) - S_\infty^\lambda(q_2))_{t \geq 0}$ is distributed as \hat{S}_∞^λ given by

$$\hat{S}_\infty^\lambda(t) = \sum_{i \notin V_{q_2}} \theta_i (\mathcal{S}_i(t) - (\theta_i/\mu)t) + \lambda t. \tag{3.21}$$

Therefore, the process in (3.21) again has the form (2.4) (see (3.11)). Now, for any $x > 0$, the probability that $|e(q_2)| = x$, conditionally on \mathcal{F}_{q_2} and $|e(q_1)| = x$, is zero using Lemma 4 together with the fact that $|V_{q_2}^c| = \infty$. This concludes the verification of Definition 1(f).

4. The Critical Window

In this section, we prove our results related to critical percolation on $\text{CM}_n(d)$. In Sect. 4.1, we start by describing a way to approximate percolation on a configuration model by a suitable alternative configuration model. In Sect. 4.2, we analyze the latter graph. The first step is to set up an exploration process that approximately encodes the component sizes in terms of excursion lengths above past minima. This exploration process is shown to converge to $\mathbf{S}_\infty^\lambda$ (Sect. 4.2.1). We must also ensure that the exploration process does not have large excursions appearing beyond the time scale of the exploration process, which allows us to prove that the largest component sizes converge to largest excursion lengths of $\mathbf{S}_\infty^\lambda$ (Sect. 4.2.2). Next we analyze the surplus edges (Sect. 4.2.3) and the proof of Theorem 1 is completed in Sect. 4.2.4. Finally, we analyze the diameter of the critical components in Sect. 4.3 and complete the proof of Theorem 2.

4.1. Sandwiching the percolated configuration model. Following the pioneering work of Aldous [4], the main tool to prove scaling limits of the component sizes is to set up an appropriate exploration process. The idea is to explore the graph sequentially, and the exploration process keeps track of some functional of vertices that have been discovered but their neighborhoods have not been explored. For percolation on the configuration model, this could be the number of unpaired half-edges of those vertices. Now, for random graphs with independent connection probabilities, the exploration process is usually Markovian, but not for the configuration model. Indeed, one has to keep track of the degree-profile outside the explored graph in order to know the distribution of the degree of a newly discovered vertex. For d -regular graphs, Nachmias and Peres [48] used the above approach, but this becomes difficult in the unbounded degree case. In earlier papers with Sen [27,28], we have used a construction by Janson [39] which says that the percolated configuration model can be viewed as a configuration model satisfying some criticality condition, so that it is enough to analyze the behavior of these critical configuration models. However, in the $\tau \in (2, 3)$ regime, this construction does not work because it gives rise to $n - o(n)$ many degree-one vertices. As a remedy to this problem, we use a result of Fountoulakis [32] to show that the critical configuration model can be sandwiched between two approximately equal configuration models, as stated in Proposition 3 below. We emphasize that Proposition 3 holds for percolation on the configuration model without any specific assumption on the degree distribution, as long as $\ell_n p_n \gg \log(n)$, and this will be used in the proofs for the near-critical results as well. We start by describing the approximating configuration model below:

Algorithm 1. (S0) Keep each half-edge with probability p_n , independently, and delete the half-edges otherwise. If the total number of retained half-edges is odd, then attach a *dummy* half-edge to vertex 1.

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

The following proposition formally states that $\mathcal{G}_n(p_n)$ approximates $\text{CM}_n(\mathbf{d}, p_n)$:

Proposition 3. *Let p_n be such that $\ell_n p_n \gg \log(n)$. There exists $(\varepsilon_n)_{n \geq 1} \subset (0, \infty)$ with $\varepsilon_n \rightarrow 0$, and a coupling such that, with high probability,*

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

Proof. The proof relies on an exact construction of $\text{CM}_n(\mathbf{d}, p_n)$ by Fountoulakis [32] which goes as follows:

Algorithm 2. (S0) Perform a binomial trial $X \sim \text{Bin}(\ell_n/2, p_n)$ and choose $2X$ half-edges uniformly at random from the set of all half-edges.

(S1) Perform a perfect matching of these $2X$ chosen half-edges. The resulting graph is distributed as $\text{CM}_n(\mathbf{d}, p_n)$.

Notice the similarity between Algorithm 1 (S1) and Algorithm 2 (S1). In both algorithms, given the number of retained half-edges, the choice of the half-edges can be performed sequentially uniformly at random without replacement. Thus, given the number of half-edges in the two algorithms, we can couple the choice of the half-edges, and their pairing (the restriction of a uniform matching to a subset of half-edge remains uniform matching on that subset). Let $\mathcal{H}_1, \mathcal{H}_2^-$ and \mathcal{H}_2^+ , respectively, denote the number of half-edges in $\text{CM}_n(\mathbf{d}, p_n), \mathcal{G}_n(p_n(1 - \varepsilon_n))$ and $\mathcal{G}_n(p_n(1 + \varepsilon_n))$. From the above discussion, the proof is complete if we can show that, as $n \rightarrow \infty$,

$$\mathbb{P}(\mathcal{H}_2^- \leq \mathcal{H}_1 \leq \mathcal{H}_2^+) \rightarrow 1.$$

We ignore the contribution due to the possible addition of *only one* dummy edge in Algorithm 3 (S0), as it does not affect asymptotic computations. Notice that $\mathcal{H}_1 = 2X$, where $X \sim \text{Bin}(\ell_n/2, p_n)$, and $\mathcal{H}_2^{\pm} \sim \text{Bin}(\ell_n, p_n(1 \pm \varepsilon_n))$. Using standard concentration inequalities [42, Corollary 2.3], it follows that

$$\mathcal{H}_1 = \ell_n p_n + o_{\mathbb{P}}(\sqrt{\ell_n p_n \log(n)}),$$

and

$$\mathcal{H}_2^{\pm} = \ell_n p_n + \ell_n p_n \varepsilon_n + o_{\mathbb{P}}(\sqrt{\ell_n p_n \log(n)}).$$

If we choose ε_n such that $\varepsilon_n \gg (\log(n)/(\ell_n p_n))^{1/2}$ and $\varepsilon_n \rightarrow 0$, then, with high probability, $\mathcal{H}_1 \leq \mathcal{H}_2^+$. Similarly we can conclude that $\mathcal{H}_2^- \leq \mathcal{H}_1$ with high probability, and the proof of Proposition 3 follows. \square

We conclude this section by stating some properties of the degree sequence of the graph $\mathcal{G}_n(p_n)$ that will be crucial in the analysis below. Let $\tilde{\mathbf{d}} = (\tilde{d}_1, \dots, \tilde{d}_n)$ be the degree sequence induced by Algorithm 1 (S1), and let $\tilde{\ell}_n = \sum_i \tilde{d}_i$ be the number of retained half-edges. Then the following result holds for $\tilde{\mathbf{d}}$:

Lemma 5. (Degrees of $\mathcal{G}_n(p_n)$) *Suppose that $p_n \gg n^{-\alpha}$, and Assumption 1 holds. For each fixed $i \geq 1$, $\tilde{d}_i = d_i p_n(1 + o_{\mathbb{P}}(1))$, $\tilde{\ell}_n = \ell_n p_n(1 + o_{\mathbb{P}}(1))$, and $\sum_{i \in [n]} \tilde{d}_i(\tilde{d}_i - 1) = p_n^2 \sum_{i \in [n]} d_i(d_i - 1)(1 + o_{\mathbb{P}}(1))$. Consequently, for $p_n \ll p_c(\lambda)$, $\sum_{i \in [n]} \tilde{d}_i^2 = \tilde{\ell}_n(1 + o_{\mathbb{P}}(1))$, whereas for $p_n = p_c(\lambda)$,*

$$\tilde{v}_n = \frac{\sum_{i \in [n]} \tilde{d}_i(\tilde{d}_i - 1)}{\sum_{i \in [n]} \tilde{d}_i} = \lambda(1 + o_{\mathbb{P}}(1)), \quad \text{and} \quad \lim_{K \rightarrow \infty} \limsup_{n \rightarrow \infty} \mathbb{P}\left(\sum_{i > K} \tilde{d}_i(\tilde{d}_i - 1) > \varepsilon \tilde{\ell}_n\right) = 0, \tag{4.2}$$

for any $\varepsilon > 0$.

Proof. Note that $\tilde{d}_i \sim \text{Bin}(d_i, p_n)$, independently for $i \in [n]$. For each fixed $i \geq 1$, $d_i p_n \rightarrow \infty$, as $p_n \gg n^{-\alpha}$. Thus the first fact follows using [42, Theorem 2.1]. Since, $\tilde{\ell}_n \sim \text{Bin}(\ell_n, p_n)$, the second fact also follows using the same bound. To see the asymptotics for $\tilde{m}_2 := \sum_{i \in [n]} \tilde{d}_i(\tilde{d}_i - 1)$, note that $\mathbb{E}[\tilde{m}_2] = p_n^2 m_2$, where $m_2 = \sum_{i \in [n]} d_i(d_i - 1)$. Also, $\text{Var}(\tilde{d}_i(\tilde{d}_i - 1)) = 2d_i(d_i - 1)p_n^2(1 - p_n)(1 + (2d_i - 3)p_n)$. Thus,

$$\frac{\text{Var}\left(\sum_{i \in [n]} \tilde{d}_i(\tilde{d}_i - 1)\right)}{\left(\mathbb{E}\left[\sum_{i \in [n]} \tilde{d}_i(\tilde{d}_i - 1)\right]\right)^2} \leq \frac{4d_1 p_n^3 m_2}{p_n^4 m_2^2} = O\left(\frac{1}{p_n n^\alpha}\right) = o(1),$$

where the penultimate step uses the fact that $m_2 = \Theta(n^{2\alpha})$, $d_1 = \Theta(n^\alpha)$, and in the last step we have again used the fact that $p_n \gg n^{-\alpha}$. Using Chebyshev’s inequality, it now follows that $\tilde{m}_2 = p_n^2 m_2(1 + o_{\mathbb{P}}(1))$. Thus,

$$\tilde{v}_n = (1 + o_{\mathbb{P}}(1)) p_n \frac{\sum_{i \in [n]} d_i(d_i - 1)}{\sum_{i \in [n]} d_i} = (1 + o_{\mathbb{P}}(1)) p_n v_n.$$

For $p_n \ll p_c(\lambda)$, $p_n v_n = o(1)$. Thus, $\sum_{i \in [n]} \tilde{d}_i^2 = \tilde{\ell}_n(1 + o_{\mathbb{P}}(1))$. For $p_n = p_c(\lambda)$, the first equality in (4.2) follows using (2.9).

We now prove the second inequality in (4.2). For any $\varepsilon > 0$, the required probability is at most

$$\begin{aligned} & \mathbb{P}\left(\sum_{i > K} \tilde{d}_i(\tilde{d}_i - 1) > \varepsilon \tilde{\ell}_n, \frac{\ell_n p_n}{2} \leq \tilde{\ell}_n \leq 2\ell_n p_n\right) + o(1) \\ & \leq \mathbb{P}\left(\sum_{i > K} \tilde{d}_i(\tilde{d}_i - 1) > \frac{\varepsilon \ell_n p_n}{2}\right) + o(1) \\ & \leq \frac{4p_n^2 \sum_{i > K} d_i(d_i - 1)}{\varepsilon \ell_n p_n} + o(1) = \frac{4p_n \sum_{i > K} d_i^2}{\varepsilon \ell_n} + o(1), \end{aligned}$$

where the penultimate step follows from Markov’s inequality. The proof now follows using (2.8) and $p_n = \Theta(n^{1-2\alpha})$ for $p_n = p_c(\lambda)$. \square

4.2. Scaling limits of critical components.

4.2.1. Convergence of the exploration process Let $\tilde{\mathbf{d}} = (\tilde{d}_1, \dots, \tilde{d}_n)$ be the degree sequence induced by Algorithm 1 (S1) with $p_n = p_c(\lambda)$, and consider $\mathcal{G}_n(p_c(\lambda))$. Note that $\mathcal{G}_n(p_c(\lambda))$ has the same distribution as $\text{CM}_n(\tilde{\mathbf{d}})$. We start by describing how the connected components in the graph can be explored while generating the random graph simultaneously:

Algorithm 3 (*Exploring the graph*). The algorithm carries along vertices that can be alive, active, exploring and killed, and half-edges that can be alive, active or killed. Alive and killed half-edges correspond to unpaired and paired half-edges respectively, whereas active half-edges correspond to half-edges that have been found during the exploration, but have not been paired yet. Thus a half-edge can be alive and active simultaneously. Similarly, a vertex is killed when all its half-edges have been explored,

otherwise the vertex is alive. An active vertex is an alive vertex that has been found already during the exploration, whereas an exploring vertex is currently being explored. 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*. Proceed to step $i + 1$.
- (S2) At each stage i , if the set of active half-edges is non-empty, then take an active half-edge e of an exploring vertex v and pair it with a half-edge f chosen uniformly among the alive half-edges. 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 $\text{degree}(v') = 1$ (i.e. the only half-edge incident to v' is f) then kill v' . Otherwise, declare v' to be active and *larger* than all other vertices that are active. After killing e , if v does not have another active half-edge, then kill v also, and declare the *smallest* vertex to be exploring.
- (S3) Repeat from (S1) at stage $i + 1$ if not all half-edges are already killed.

Algorithm 3 gives a *breadth-first* exploration of the connected components of $\text{CM}_n(\tilde{\mathbf{d}})$. Define the exploration process by

$$S_n(0) = 0, \quad S_n(l) = S_n(l - 1) + \tilde{d}_{(l)} J_l - 2, \tag{4.3}$$

where J_l is the indicator that a new vertex is discovered at time l and $\tilde{d}_{(l)}$ is the degree of the new vertex chosen at time l when $J_l = 1$. The -2 in (4.3) takes into account the fact that two half-edges are killed whenever two half-edges are paired at some step. However, at the beginning of exploring a component when Algorithm 3 (S1) is carried out, we do not pair half-edges but the exploration process subtracts -2 nonetheless. For this reason, there is an additional -2 in (4.3) at the beginning of exploring each component, and thus the first component is explored when the exploration process hits -2 , the second component is explored when the process hits -4 and so on. More formally, suppose that \mathcal{C}_k is the k -th connected component explored by the above exploration process and define $\tau_k = \inf \{i : S_n(i) = -2k\}$. Then \mathcal{C}_k is discovered between the times $\tau_{k-1} + 1$ and τ_k , and $\tau_k - \tau_{k-1} - 1$ gives the total number of edges in \mathcal{C}_k . Call a vertex *discovered* if it is either active or killed. Let \mathcal{V}_l denote the set of vertices discovered up to time l and $\mathcal{S}_i^n(l) := \mathbb{1}_{\{i \in \mathcal{V}_l\}}$. Note that

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

In the rest of this section, we often use the asymptotics in Lemma 5 even if it is not stated explicitly. Recall that we write $\mathcal{F}_l^n = \sigma(\mathcal{S}_i^n(l) : i \in [n])$. All the martingales and related computations will be done with respect to the filtration $(\mathcal{F}_l^n)_{l \geq 0}$.

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

$$\bar{S}_n(t) = n^{-\rho} \sum_{i \in [n]} (\tilde{d}_i - 1) \mathcal{S}_i^n(tn^\rho) + n^{-\rho} \sum_{i \in [n]} \mathcal{S}_i^n(tn^\rho) - 2t + o(1), \tag{4.4}$$

where we have used the convention that $\mathcal{S}_i^n(tn^\rho) = \mathcal{S}_i^n(\lfloor tn^\rho \rfloor)$ when tn^ρ is not an integer. The following theorem describes the scaling limit of this rescaled process:

Theorem 5. Consider the process $\bar{\mathbf{S}}_n := (\bar{S}_n(t))_{t \geq 0}$ defined in (4.4) and recall the definition of $\mathbf{S}_\infty^\lambda$ in (2.4). Then, under Assumption 1, as $n \rightarrow \infty$,

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

with respect to the Skorohod J_1 -topology.

To prove Theorem 5, we need to obtain asymptotics of the first two terms in (4.4). The first term accounts for the contribution due to the non-degree-one vertices during the exploration. The first term is dominated by the contributions due to hubs, which allows us to use a truncation argument. The convergence of the truncated sum is given by the following lemma:

Lemma 6. Fix any $K \geq 1$, and $\mathcal{I}_i(s) := \mathbb{1}_{\{\xi_i \leq s\}}$ where $\xi_i \sim \text{Exp}(\theta_i/\mu)$ independently for $i \in [K]$. Under Assumption 1, as $n \rightarrow \infty$,

$$(\mathcal{I}_i^n(tn^\rho))_{i \in [K], t \geq 0} \xrightarrow{d} (\mathcal{I}_i(t))_{i \in [K], t \geq 0}$$

with respect to the Skorohod J_1 -topology.

The second term in (4.4) describes the proportion of time when a new vertex is found. Since we see a new vertex of degree one in most steps of the exploration process, this term is shown to converge to the constant function t , which is proved using martingale arguments. This is summarized in the next lemma:

Lemma 7. For any $u > 0$, as $n \rightarrow \infty$, $\sup_{u \leq t} n^{-\rho} \left| \sum_{i \in [n]} \mathcal{I}_i^n(un^\rho) - un^\rho \right| \xrightarrow{\mathbb{P}} 0$.

We first prove Theorem 5 using Lemmas 6 and 7. The lemmas will be proved subsequently. Let $\tilde{\ell}_n(u)$ denote the number of unpaired half-edges at time $\lfloor un^\rho \rfloor$. Thus, $\tilde{\ell}_n(u) = \tilde{\ell}_n - 2(\lfloor un^\rho \rfloor - c_{\lfloor un^\rho \rfloor}) + 1$, where c_l is the number of components explored up to time l . Note that $\tilde{\ell}_n - 2un^\rho + 1 \leq \tilde{\ell}_n(u) \leq \tilde{\ell}_n$. Since $\tilde{\ell}_n = \Theta_{\mathbb{P}}(n^{2\rho})$, we have $\tilde{\ell}_n(u) = \tilde{\ell}_n(1 + o_{\mathbb{P}}(1))$ uniformly over $u \leq t$. Let $\tilde{\mathbb{P}}(\cdot)$ (respectively $\tilde{\mathbb{E}}[\cdot]$) denote the conditional probability (respectively expectation) conditionally on $(\tilde{d}_i)_{i \in [n]}$.

Proof of Theorem 5. Note that, $\mathcal{I}_i^n(l) = 0$ for all $l \geq 1$ if $\tilde{d}_i = 0$. Now, if $\tilde{d}_i \geq 1$, then for any $t \geq 0$, uniformly over $l \leq tn^\rho$,

$$\tilde{\mathbb{E}}[\mathcal{I}_i^n(l)] = \tilde{\mathbb{P}}(\mathcal{I}_i^n(l) = 1) \leq \frac{l\tilde{d}_i}{\tilde{\ell}_n - 2un^\rho + 1}. \tag{4.5}$$

Let $X_{n,K}(t) := n^{-\rho} \sup_{u \leq t} \sum_{i > K} (\tilde{d}_i - 1) \mathcal{I}_i^n(un^\rho)$. Note that $\mathcal{I}_i^n(un^\rho) \leq \mathcal{I}_i^n(tn^\rho)$. Also, using $\mathcal{I}_i^n(un^\rho) = 0$ whenever $\tilde{d}_i = 0$, it follows that $(\tilde{d}_i - 1) \mathcal{I}_i^n(un^\rho) \geq 0$ for all $i \in [n]$ and $u > 0$. Thus,

$$\tilde{\mathbb{E}}[X_{n,K}(t)] \leq n^{-\rho} \tilde{\mathbb{E}} \left[\sum_{i > K} (\tilde{d}_i - 1) \mathcal{I}_i^n(tn^\rho) \right] \leq t \frac{\sum_{i > K} \tilde{d}_i (\tilde{d}_i - 1)}{\tilde{\ell}_n(t)} := \varepsilon_{n,K}(t),$$

where $\lim_{K \rightarrow \infty} \limsup_{n \rightarrow \infty} \mathbb{P}(\varepsilon_{n,K}(t) > \delta) = 0$ for any $\delta > 0$, due to Lemma 5. Therefore, for any $\varepsilon, \delta > 0$, using Markov's inequality,

$$\lim_{K \rightarrow \infty} \limsup_{n \rightarrow \infty} \mathbb{P}(\tilde{\mathbb{P}}(X_{n,K}(t) > \varepsilon) > \delta) \leq \lim_{K \rightarrow \infty} \limsup_{n \rightarrow \infty} \mathbb{P}(\tilde{\mathbb{E}}[X_{n,K}(t)] > \delta\varepsilon) = 0.$$

Let $\mathcal{B}_{n,K} := \{\tilde{\mathbb{P}}(X_{n,K}(t) > \varepsilon) > \delta\}$. It follows that

$$\mathbb{P}(X_{n,K}(t) > \varepsilon) = \mathbb{E}[\tilde{\mathbb{P}}(X_{n,K}(t) > \varepsilon)] \leq \mathbb{P}(\mathcal{B}_{n,K}) + \delta.$$

Taking the iterated limit $\lim_{\delta \rightarrow 0} \limsup_{K \rightarrow \infty} \limsup_{n \rightarrow \infty}$ yields, for any $\varepsilon > 0$,

$$\lim_{K \rightarrow \infty} \limsup_{n \rightarrow \infty} \mathbb{P}(X_{n,K}(t) > \varepsilon) = 0. \quad (4.6)$$

Using (4.6) and Lemma 7, it is now enough to deduce the scaling limit, as $n \rightarrow \infty$, for

$$\tilde{S}_n^K(t) = n^{-\rho} \sum_{i=1}^K \tilde{d}_i \mathcal{I}_i^n(tn^\rho) - t$$

and then taking $K \rightarrow \infty$. But for any fixed $K \geq 1$, Lemma 6 yields the limit of S_n^K , and the proof of Theorem 5 follows. \square

Proof of Lemma 6. By noting that $(\mathcal{I}_i^n(tn^\rho))_{t \geq 0}$ are indicator processes, for any $m_1 \leq m_2 \leq m_3$, it follows that $\min\{\mathcal{I}_i^n(m_2) - \mathcal{I}_i^n(m_1), \mathcal{I}_i^n(m_3) - \mathcal{I}_i^n(m_2)\} = 0$, and thus [14, Theorem 13.5] implies tightness of $(\mathcal{I}_i^n(tn^\rho))_{t \geq 0, n \geq 1}$ for each fixed $i \geq 1$. Thus, it is enough to show that

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

for any $t_1, \dots, t_K \in [0, \infty)$. Now,

$$\tilde{\mathbb{P}}(\mathcal{I}_i^n(m_i) = 0, \forall i \in [K]) = \prod_{l=1}^{\infty} \left(1 - \sum_{i \leq K: l \leq m_i} \frac{\tilde{d}_i}{\tilde{\ell}_n - \Theta(l)}\right). \quad (4.7)$$

Taking logarithms on both sides of (4.7) and using the fact that $l \leq \max m_i = \Theta(n^\rho)$ we get

$$\begin{aligned} \tilde{\mathbb{P}}(\mathcal{I}_i^n(m_i) = 0 \forall i \in [K]) &= \exp\left(-\sum_{l=1}^{\infty} \sum_{i \leq K: l \leq m_i} \frac{\tilde{d}_i}{\tilde{\ell}_n} + o(1)\right) \\ &= \exp\left(-\sum_{i \in [K]} \frac{\tilde{d}_i m_i}{\tilde{\ell}_n} + o(1)\right). \end{aligned} \quad (4.8)$$

Putting $m_i = t_i n^\rho$, Assumption 1 (i), (ii) give

$$\frac{m_i \tilde{d}_i}{\tilde{\ell}_n} = \frac{\theta_i t_i}{\mu} (1 + o_{\mathbb{P}}(1)). \quad (4.9)$$

Hence (4.8) and (4.9) complete the proof of Lemma 6. \square

Proof of Lemma 7. Define $W_n(l) = \sum_{i \in [n]} \mathcal{S}_i^n(l) - l$. Recall that \mathcal{V}_l denotes the set of vertices discovered up to time l , τ_k is the time when the k -th component has been explored, and c_l is the number of components explored up to time l . Observe that

$$\begin{aligned} \tilde{\mathbb{E}}[W_n(l+1) - W_n(l) \mid \mathcal{F}_l] &= \sum_{i \in [n]} \tilde{\mathbb{E}}[\mathcal{S}_i^n(l+1) \mid \mathcal{F}_l] \mathbb{1}_{\{i \notin \mathcal{V}_l\}} - 1 \\ &= \sum_{i \notin \mathcal{V}_l} \frac{\tilde{d}_i}{\tilde{\ell}_n - 2l + 2c_l + 1} - 1 = \frac{2l - 1 - \sum_{i \in \mathcal{V}_l} \tilde{d}_i - 2c_l}{\tilde{\ell}_n - 2l + 2c_l + 1}. \end{aligned} \quad (4.10)$$

To see that the final term in (4.10) is negative, note that if $l = \tau_k$ for some k , then $\sum_{i \in \mathcal{V}_k} \tilde{d}_i - 2\tau_k = 2k$, and $c_{\tau_k} = k$ so that

$$2\tau_k - 1 - \sum_{i \in \mathcal{V}_k} \tilde{d}_i - 2c_{\tau_k} = -1 < 0. \quad (4.11)$$

If $\tau_k < l < \tau_{k+1}$, then $\sum_{i \in \mathcal{V}_l \setminus \mathcal{V}_{\tau_k}} \tilde{d}_i - 2(l - \tau_k) \geq -1$, and also $c_l = c_{\tau_k} + 1$. Therefore, using (4.11), we conclude that the final term in (4.10) is negative for all $l \geq 1$, and consequently, $(W_n(l))_{l \geq 1}$ is a super-martingale. We will use the martingale-inequality [54, Lemma 2.54.5] stating that for any sub/super-martingale $(M(t))_{t \geq 0}$, with $M(0) = 0$,

$$\varepsilon \mathbb{P} \left(\sup_{s \leq t} |M(s)| > 3\varepsilon \right) \leq 3\mathbb{E}[|M(t)|] \leq 3 \left(|\mathbb{E}[M(t)]| + \sqrt{\text{Var}(M(t))} \right). \quad (4.12)$$

Using Taylor expansion,

$$\tilde{\mathbb{P}}(\mathcal{S}_i^n(l) = 1) \geq 1 - \left(1 - \frac{\tilde{d}_i}{\tilde{\ell}_n}\right)^l \geq \left(\frac{l\tilde{d}_i}{\tilde{\ell}_n} - \frac{l^2\tilde{d}_i^2}{\tilde{\ell}_n^2}\right) \mathbb{1}_{\{\tilde{d}_i < \tilde{\ell}_n\}},$$

and thus, using Lemma 5, and $l = tn^\rho$,

$$\begin{aligned} n^{-\rho} |\tilde{\mathbb{E}}[W_n(tn^\rho)]| &= t - n^{-\rho} \sum_{i \in [n]} \tilde{\mathbb{P}}(\mathcal{S}_i^n(tn^\rho) = 1) \\ &\leq t \sum_{i \in [n]} \frac{\tilde{d}_i \mathbb{1}_{\{\tilde{d}_i > \tilde{\ell}_n / tn^\rho\}}}{\tilde{\ell}_n} + \frac{t^2 n^\rho \sum_{i \in [n]} \tilde{d}_i^2}{\tilde{\ell}_n^2}. \end{aligned} \quad (4.13)$$

Let \mathcal{E}_n denote the good event that $\ell_n p_c(\lambda)/2 \leq \tilde{\ell}_n \leq 2\ell_n p_c(\lambda)$ and $p_c(\lambda)d_i/2 \leq \tilde{d}_i \leq 2p_c(\lambda)d_i$ for all i such that $d_i > C_0 n^\rho$ for some C_0 (sufficiently small). Using standard concentration inequalities for the binomial distribution [42, Theorem 2.1], $\mathbb{P}(\mathcal{E}_n^c) < 2e^{-n^\varepsilon}$ for some $\varepsilon > 0$. On the event \mathcal{E}_n , $\tilde{d}_i > Cn^\rho$, and thus $d_i > Cn^\rho$. We can bound

$$\sum_{i \in [n]} \frac{\tilde{d}_i \mathbb{1}_{\{\tilde{d}_i > \tilde{\ell}_n / tn^\rho\}}}{\tilde{\ell}_n} \leq \frac{C_1}{\ell_n} \sum_{i \in [n]} d_i \mathbb{1}_{\{d_i > Cn^\rho\}} = o(1),$$

where the final step follows using the uniform integrability from Assumption 1. The second term in (4.13) is $o_{\mathbb{P}}(1)$ using Lemma 5. Thus,

$$n^{-\rho} |\tilde{\mathbb{E}}[W_n(tn^\rho)]| = o_{\mathbb{P}}(1). \quad (4.14)$$

Next, 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, for all $l \geq 1$ and $i \neq j$,

$$\tilde{\mathbb{P}}(\mathcal{S}_i^n(l) = 0, \mathcal{S}_j^n(l) = 0) \leq \tilde{\mathbb{P}}(\mathcal{S}_i^n(l) = 0)\tilde{\mathbb{P}}(\mathcal{S}_j^n(l) = 0),$$

and thus

$$\begin{aligned} & \tilde{\mathbb{P}}(\mathcal{S}_i^n(l) = 1, \mathcal{S}_j^n(l) = 1) \\ &= 1 - \tilde{\mathbb{P}}(\mathcal{S}_i^n(l) = 0) - \tilde{\mathbb{P}}(\mathcal{S}_j^n(l) = 0) + \tilde{\mathbb{P}}(\mathcal{S}_i^n(l) = 0, \mathcal{S}_j^n(l) = 0) \\ &\leq 1 - \tilde{\mathbb{P}}(\mathcal{S}_i^n(l) = 0) - \tilde{\mathbb{P}}(\mathcal{S}_j^n(l) = 0) + \tilde{\mathbb{P}}(\mathcal{S}_i^n(l) = 0)\tilde{\mathbb{P}}(\mathcal{S}_j^n(l) = 0) \\ &= \tilde{\mathbb{P}}(\mathcal{S}_i^n(l) = 1)\tilde{\mathbb{P}}(\mathcal{S}_j^n(l) = 1). \end{aligned} \quad (4.15)$$

Therefore $\mathcal{S}_i^n(l)$ and $\mathcal{S}_j^n(l)$ are negatively correlated. Using (4.5), it follows that

$$\text{Var}(\mathcal{S}_i^n(l) | (\tilde{d}_i)_{i \in [n]}) \leq \tilde{\mathbb{P}}(\mathcal{S}_i^n(l) = 1) \leq \frac{l\tilde{d}_i}{\tilde{\ell}_n(t)},$$

uniformly over $l \leq tn^\rho$. Therefore, using the negative correlation in (4.15),

$$\begin{aligned} n^{-2\rho} \text{Var}(W_n(tn^\rho) | (\tilde{d}_i)_{i \in [n]}) &\leq n^{-2\rho} \sum_{i \in [n]} \text{Var}(\mathcal{S}_i^n(tn^\rho) | (\tilde{d}_i)_{i \in [n]}) \\ &= n^{-2\rho} tn^\rho \frac{\sum_{i \in [n]} \tilde{d}_i}{\tilde{\ell}_n(t)} = \Theta_{\mathbb{P}}(n^{-\rho}) = o_{\mathbb{P}}(1). \end{aligned} \quad (4.16)$$

Using (4.14) and (4.16), the proof now follows by an application of (4.12). \square

4.2.2. Large components are explored early In this section, we prove two key results that allow us to deduce the convergence of the component sizes. Firstly, we show that the rescaled vector of component sizes is tight in ℓ^2_{\downarrow} (see Proposition 4). This result is then used to show that the largest components of $\mathcal{G}_n(p_c(\lambda))$ are explored before time $\Theta(n^\rho)$ (Proposition 5). The latter allows us to apply Proposition 1. Let $\mathcal{C}_{(i)}$ denote the i -th largest component for $\mathcal{G}_n(p_c(\lambda))$. Recall that our convention is to take $|\mathcal{C}| = 0$, if the component consists of one vertex and no edges.

Proposition 4. *Under Assumption 1, for any $\varepsilon > 0$,*

$$\lim_{K \rightarrow \infty} \limsup_{n \rightarrow \infty} \mathbb{P}\left(\sum_{i > K} |\mathcal{C}_{(i)}|^2 > \varepsilon n^{2\rho}\right) = 0.$$

Let \mathcal{G}_n^K be the random graph obtained by removing all edges attached to vertices $1, \dots, K$ and let \mathbf{d}' be the obtained degree sequence. Further, let $\mathcal{C}^K(v)$ and $\mathcal{C}_{(i)}^K$ denote the connected component containing v and the i -th largest component respectively in \mathcal{G}_n^K . Let $D^K(v) = \sum_{k \in \mathcal{C}^K(v)} \tilde{d}_k$ and $D_i^K = \sum_{k \in \mathcal{C}_{(i)}^K} \tilde{d}_k$. Let $V_n^{*,K}$ be chosen according to the following size-biased distribution:

$$\mathbb{P}(V_n^{*,K} = i) = \frac{\tilde{d}_i}{\tilde{\ell}_n - \sum_{i=1}^K \tilde{d}_i}, \quad \text{for } i \in [n] \setminus [K].$$

Also, denote the criticality parameter of \mathcal{G}_n^K by ν_n^K .

Lemma 8. *Suppose that Assumption 1 holds. Then, for any $\varepsilon > 0$,*

$$\lim_{K \rightarrow \infty} \limsup_{n \rightarrow \infty} \mathbb{P} \left(\tilde{\mathbb{E}} \left[\sum_{k \in \mathcal{C}^K(V_n^{*,K})} (\tilde{d}_k - 1) \right] > \varepsilon \right) = 0.$$

Proof. Note that the criticality parameter of $\mathcal{G}_n(p_c(\lambda))$ is $\tilde{v}_n = \lambda(1 + o_{\mathbb{P}}(1))$, by Lemma 5. Now, conditionally on the set of removed half-edges, \mathcal{G}_n^K is still a configuration model with some degree sequence \mathbf{d}' with $d'_i \leq \tilde{d}_i$ for all $i \in [n] \setminus [K]$ and $d'_i = 0$ for $i \in [K]$. Further, the criticality parameter of \mathcal{G}_n^K satisfies

$$v_n^K = \frac{\sum_{i \in [n]} d'_i (d'_i - 1)}{\sum_{i \in [n]} d'_i} \leq \frac{\sum_{i > K} \tilde{d}_i (\tilde{d}_i - 1)}{\tilde{\ell}_n - 2 \sum_{i=1}^K \tilde{d}_i} = \lambda \frac{\sum_{i > K} \tilde{d}_i (\tilde{d}_i - 1)}{\sum_{i \in [n]} \tilde{d}_i (\tilde{d}_i - 1)} (1 + o_{\mathbb{P}}(1)), \quad (4.17)$$

where we have used $\tilde{v}_n = \lambda(1 + o_{\mathbb{P}}(1))$ in the last step. Now, by Assumption 1 and Lemma 5, it is possible to choose K_0 large such that for all $K \geq K_0$

$$v_n^K < 1 \quad \text{with high probability.}$$

This yields

$$\tilde{\mathbb{E}} \left[\sum_{k \in \mathcal{C}^K(V_n^{*,K})} (\tilde{d}_k - 1) \right] \leq \tilde{\mathbb{E}}[\tilde{d}_{V_n^{*,K}} - 1] \left(1 + \frac{\tilde{\mathbb{E}}[\tilde{d}_{V_n^{*,K}}]}{(1 - v_n^K)} + o_{\mathbb{P}}(1) \right), \quad (4.18)$$

where $\tilde{d}_{V_n^{*,K}}$ is the degree of the vertex $V_n^{*,K}$ in $\mathcal{G}_n(p_c(\lambda))$. The proof of (4.18) uses path-counting techniques for the configuration model [40]. Since the arguments are adaptations of [27], we move the proof to Appendix A.1. We now use Lemma 5 to compute the asymptotics of the different terms in (4.18). Note that $\tilde{\mathbb{E}}[\tilde{d}_{V_n^{*,K}}] \leq (1 + o_{\mathbb{P}}(1)) \sum_{i > K} \tilde{d}_i^2 / \tilde{\ell}_n = O_{\mathbb{P}}(1)$, and

$$\tilde{\mathbb{E}}[\tilde{d}_{V_n^{*,K}} - 1] = \frac{\sum_{i > K} \tilde{d}_i (\tilde{d}_i - 1)}{\tilde{\ell}_n - \sum_{i=1}^K \tilde{d}_i} = (1 + o_{\mathbb{P}}(1)) \frac{p_n \sum_{i > K} d_i (d_i - 1)}{\sum_{i \in [n]} d_i} \xrightarrow{\mathbb{P}} 0,$$

in the iterated limit $\lim_{K \rightarrow \infty} \lim_{n \rightarrow \infty}$. Thus the proof of Lemma 8 follows. \square

Proof of Proposition 4. Recall that $\mathcal{C}_{(i)}^K$ denotes the i -th largest component in \mathcal{G}_n^K and $D_i^K = \sum_{k \in \mathcal{C}_{(i)}^K} \tilde{d}_k$. Denote by \mathcal{S}_K , the squared sum of the component sizes after removing components containing $1, \dots, K$. Note that

$$\sum_{i > K} |\mathcal{C}_{(i)}|^2 = \sum_{i \geq 1} |\mathcal{C}_{(i)}|^2 - \sum_{i=1}^K |\mathcal{C}_{(i)}|^2 \leq \mathcal{S}_K \leq \sum_{i \geq 1} |\mathcal{C}_{(i)}^K|^2 \leq 4 \sum_{i \geq 1} D_i^K \sum_{k \in \mathcal{C}_{(i)}^K} (\tilde{d}_k - 1), \quad (4.19)$$

where the last step uses $d'_i \leq \tilde{d}_i$ and the fact that for any connected component \mathcal{C} with total degree D , we must have $D - |\mathcal{C}| \geq |\mathcal{C}|/4$. The last fact can be seen for $|\mathcal{C}| \geq 2$ by $D - |\mathcal{C}| \geq 2(|\mathcal{C}| - 1) - |\mathcal{C}| = |\mathcal{C}| - 2 \geq |\mathcal{C}|/4$, and for $|\mathcal{C}| = 1$ and $D \geq 2$,

this follows trivially. Note here that we do not consider components with $|\mathcal{C}| = 1$ and $D = 0$; see Remark 2. Thus it is enough to bound the final term in (4.19). Now,

$$\begin{aligned} \tilde{\mathbb{P}}\left(\sum_{i \geq 1} D_i^K \sum_{k \in \mathcal{C}_{(i)}^K} (\tilde{d}_k - 1) > \varepsilon n^{2\rho}\right) &\leq \frac{1}{\varepsilon n^{2\rho}} \tilde{\mathbb{E}}\left[\sum_{i \geq 1} D_i^K \sum_{k \in \mathcal{C}_{(i)}^K} (\tilde{d}_k - 1)\right] \\ &= \frac{\tilde{\ell}_n - \sum_{i \in [K]} \tilde{d}_i}{\varepsilon n^{2\rho}} \tilde{\mathbb{E}}\left[\sum_{k \in \mathcal{C}^K(V_n^{*,K})} (\tilde{d}_k - 1)\right]. \end{aligned} \quad (4.20)$$

Thus, the proof follows using Lemma 8, and the fact that $\tilde{\ell}_n - \sum_{i \in [K]} \tilde{d}_i \leq \tilde{\ell}_n = O_{\mathbb{P}}(n^{2\rho})$. \square

The next proposition shows that, in Algorithm 3, the large components are explored before time $\Theta(n^\rho)$. Let $\mathcal{C}_{\max}^{\geq T}$ denote the size of the largest component whose exploration is started by Algorithm 3 after time Tn^ρ , and let $D_{\max}^{\geq T} = \sum_{k \in \mathcal{C}_{\max}^{\geq T}} \tilde{d}_k$.

Proposition 5. *Under Assumption 1, for any $\varepsilon > 0$,*

$$\lim_{T \rightarrow \infty} \limsup_{n \rightarrow \infty} \mathbb{P}(|\mathcal{C}_{\max}^{\geq T}| > \varepsilon n^\rho) = 0 \quad \text{and} \quad \lim_{T \rightarrow \infty} \limsup_{n \rightarrow \infty} \mathbb{P}(D_{\max}^{\geq T} > \varepsilon n^\rho) = 0.$$

Proof. Define $\mathcal{A}_{K,T}^n := \{\text{all the vertices of } [K] \text{ are explored before time } Tn^\rho\}$. Let $\mathcal{C}_{(i)}^K$ denote the i -th largest component of \mathcal{G}_n^K so that

$$\begin{aligned} \tilde{\mathbb{P}}(|\mathcal{C}_{\max}^{\geq T}| > \varepsilon n^\rho, \mathcal{A}_{K,T}^n) &\leq \tilde{\mathbb{P}}\left(\sum_{i \geq 1} |\mathcal{C}_{(i)}^K|^2 > \varepsilon^2 n^{2\rho}\right) \\ &\leq \tilde{\mathbb{P}}\left(\sum_{i \geq 1} D_i^K \sum_{k \in \mathcal{C}_{(i)}^K} (\tilde{d}_k - 1) > \frac{\varepsilon^2 n^{2\rho}}{4}\right). \end{aligned} \quad (4.21)$$

The final term tends to zero in probability in the iterated limit $\lim_{K \rightarrow \infty} \limsup_{n \rightarrow \infty}$, as shown in (4.20). Next, using the fact that $\tilde{d}_j n^\rho = \Theta(\tilde{\ell}_n)$, we get

$$\begin{aligned} \tilde{\mathbb{P}}((\mathcal{A}_{K,T}^n)^c) &= \tilde{\mathbb{P}}(\exists j \in [K] : j \text{ is not explored before } Tn^\rho) \\ &\leq \sum_{j=1}^K \tilde{\mathbb{P}}(j \text{ is not explored before } Tn^\rho) \leq \sum_{j=1}^K \left(1 - \frac{\tilde{d}_j}{\tilde{\ell}_n - \Theta(Tn^\rho)}\right)^{Tn^\rho} \leq \sum_{j=1}^K e^{-CT}, \end{aligned} \quad (4.22)$$

where $C > 0$ is a constant that may depend on K , and the final step holds with high probability. Now, by (4.21),

$$\tilde{\mathbb{P}}(|\mathcal{C}_{\max}^{\geq T}| > \varepsilon n^\rho) \leq \tilde{\mathbb{P}}\left(\sum_{i \geq 1} |\mathcal{C}_{(i)}^K|^2 > \varepsilon^2 n^{2\rho}\right) + \tilde{\mathbb{P}}((\mathcal{A}_{K,T}^n)^c).$$

The proof for $\mathbb{P}(|\mathcal{C}_{\max}^{\geq T}| > \varepsilon n^\rho)$ follows by taking the iterated limit $\lim_{K \rightarrow \infty} \lim_{T \rightarrow \infty} \limsup_{n \rightarrow \infty}$.

For the upper bound on $\tilde{\mathbb{P}}(D_{\max}^{\geq T} > \varepsilon n^\rho, \mathcal{A}_{k,T}^n)$, note that

$$\begin{aligned} \tilde{\mathbb{P}}(D_{\max}^{\geq T} > \varepsilon n^\rho, |\mathcal{C}_{\max}^{\geq T}| \leq \varepsilon n^\rho/2, \mathcal{A}_{k,T}^n) &\leq \tilde{\mathbb{P}}\left(D_{\max}^{\geq T} (D_{\max}^{\geq T} - |\mathcal{C}_{\max}^{\geq T}|) > \varepsilon^2 n^{2\rho}/2, \mathcal{A}_{k,T}^n\right) \\ &\leq \tilde{\mathbb{P}}\left(\sum_{i \geq 1} D_i^K \sum_{k \in \mathcal{C}_{(i)}^K} (\tilde{d}_k - 1) > \frac{\varepsilon^2 n^{2\rho}}{2}\right). \end{aligned}$$

Hence, the proof for $\mathbb{P}(D_{\max}^{\geq T} > \varepsilon n^\rho)$ also follows. \square

4.2.3. Counting process that counts surplus Let $N_n^\lambda(k)$ be the number of surplus edges discovered up to time k and $\tilde{N}_n^\lambda(u) = N_n^\lambda(\lfloor un^\rho \rfloor)$. Below, we prove the asymptotics for the process \tilde{N}_n^λ :

Lemma 9. *Under Assumption 1, as $n \rightarrow \infty$,*

$$(\tilde{S}_n, \tilde{N}_n^\lambda) \xrightarrow{d} (S_\infty^\lambda, \mathbf{N}^\lambda),$$

where \mathbf{N}^λ is defined in (2.6).

Proof. We write $N_n^\lambda(l) = \sum_{i=2}^l \xi_i$, where $\xi_i = \mathbb{1}_{\{\mathcal{Y}_i = \mathcal{Y}_{i-1}\}}$. Let A_i denote the number of active half-edges after stage i while implementing Algorithm 3. Note that

$$\tilde{\mathbb{P}}(\xi_i = 1 \mid \mathcal{F}_{i-1}) = \frac{A_{i-1} - 1}{\tilde{\ell}_n - 2i - 1} = \frac{A_{i-1}}{\tilde{\ell}_n} (1 + O(i/\tilde{\ell}_n)) + O(\tilde{\ell}_n^{-1}),$$

uniformly for $i \leq Tn^\rho$ for any $T > 0$. By Lemma 5, $\tilde{\ell}_n = \ell_n p_c(\lambda)(1 + o_{\mathbb{P}}(1)) = n^{2\rho} \lambda \mu^2 / \sum_i \theta_i^2 (1 + o_{\mathbb{P}}(1))$. Therefore, the instantaneous rate of change of the re-scaled process \tilde{N}^λ at time t , conditional on the past, is

$$n^\rho \frac{A_{\lfloor tn^\rho \rfloor}}{n^{2\rho} \frac{\lambda \mu^2}{\sum_{i \geq 1} \theta_i^2}} (1 + o_{\mathbb{P}}(1)) + o_{\mathbb{P}}(1) = \frac{\sum_{i \geq 1} \theta_i^2}{\lambda \mu^2} \text{refl}(\tilde{S}_n(t)) (1 + o_{\mathbb{P}}(1)) + o_{\mathbb{P}}(1). \tag{4.23}$$

Since the reflection of a process is continuous in Skorohod J_1 -topology (see [56, Lemma 13.5.1]), we can use Theorem 5 to conclude that $\text{refl}(\tilde{S}_n) \xrightarrow{d} \text{refl}(S_\infty^\lambda)$, so that the compensator of \tilde{N}_n^λ converges. The convergence of the compensators is usually enough for convergence of Poisson processes. Indeed, for Erdős-Rényi random graphs [4] or rank-one inhomogeneous random graphs [11, 12], showing the convergence of compensators suffices using [22, Theorem 1]. This is because 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 . This difficulty was circumvented in [27] for the $\tau \in (3, 4)$ regime. In Appendix A.2, we adapt the arguments from [27] in the $\tau \in (2, 3)$ setting, which completes the proof of Lemma 9. \square

4.2.4. *Convergence of the component sizes and the surplus edges.* We first show the asymptotics of the component sizes and surplus edges of $\mathcal{G}_n(p_c(\lambda))$ generated by Algorithm 1. Recall that $\text{SP}(\mathcal{C})$ denotes the number of surplus of \mathcal{C} . The following lemma states the tightness of the vector of component sizes and surplus edges of $\mathcal{G}_n(p_c(\lambda))$ in the \mathbb{U}_\downarrow^0 -topology:

Lemma 10. *For any $\varepsilon > 0$,*

$$\lim_{\delta \rightarrow 0} \limsup_{n \rightarrow \infty} \mathbb{P} \left(\sum_{i: |\mathcal{C}_{(i)}| \leq \delta n^\rho} |\mathcal{C}_{(i)}| \times \text{SP}(\mathcal{C}_{(i)}) > \varepsilon n^\rho \right) = 0.$$

The proof of Lemma 10 is an adaptation of [27, Proposition 19] in this setting. We provide a proof of Lemma 10 in Appendix A.3. Next, let $\mathbf{Z}'_n(\lambda)$ denote the vector $(n^{-\rho} |\mathcal{C}_{(i)}|, \text{SP}(\mathcal{C}_{(i)}))_{i \geq 1}$, ordered as an element in \mathbb{U}_\downarrow^0 . Below, we prove the scaling limit of $\mathbf{Z}'_n(\lambda)$:

Proposition 6. *Under Assumption 1, as $n \rightarrow \infty$,*

$$\mathbf{Z}'_n(\lambda) \xrightarrow{d} \mathbf{Z}(\lambda) \tag{4.24}$$

with respect to the \mathbb{U}_\downarrow^0 topology, where $\mathbf{Z}(\lambda)$ is defined in (2.7).

Proof. Recall from Proposition 2 that the limiting process $\mathbf{S}_\infty^\lambda$ is good in the sense that all the conditions in Definition 1 are satisfied. Also, Proposition 5 ensures that the additional restriction on the pre-limit process in Proposition 1 is satisfied. Thus, using Theorem 5, an application of Proposition 1 yields the finite-dimensional convergence in (4.24). Finally, the convergence in the \mathbb{U}_\downarrow^0 -topology follows using the tightness in Lemma 10. \square

We now provide a proof of Theorem 1:

Proof of Theorem 1. Throughout the proof, we ignore the λ in a predefined notation to simplify writing. We will work under the coupling under which Proposition 3 holds, i.e., $\mathcal{G}_n(p_c(1 - \varepsilon_n)) \subset \text{CM}_n(\mathbf{d}, p_c) \subset \mathcal{G}_n(p_c(1 + \varepsilon_n))$, where $\varepsilon_n \rightarrow 0$. We write $\mathcal{C}_{(i)}^-$, $\mathcal{C}_{(i)}$ and $\mathcal{C}_{(i)}^+$ to denote the i -th largest component of $\mathcal{G}_n(p_c(1 - \varepsilon_n))$, $\text{CM}_n(\mathbf{d}, p_c)$ and $\mathcal{G}_n(p_c(1 + \varepsilon_n))$ respectively, and let \mathbf{Z}_n^- , \mathbf{Z}_n and \mathbf{Z}_n^+ be the corresponding vectors, rearranged as elements of \mathbb{U}_\downarrow^0 . Then,

\mathbf{Z}_n^+ and \mathbf{Z}_n^- have identical scaling limits as Proposition 6.

Let $d_{\mathbb{U}}$ denote the metric for the \mathbb{U}_\downarrow^0 topology defined in (2.1). The proof is complete if we can show that, as $n \rightarrow \infty$,

$$d_{\mathbb{U}}(\mathbf{Z}_n^+, \mathbf{Z}_n) \xrightarrow{\mathbb{P}} 0. \tag{4.25}$$

First, we prove that, for any $K \geq 1$,

$$\lim_{n \rightarrow \infty} \mathbb{P}(\mathcal{C}_{(i)}^- \subset \mathcal{C}_{(i)}^+, \forall i \leq K) = 1. \tag{4.26}$$

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

$$\lim_{n_{0k} \rightarrow \infty} \mathbb{P}(|\mathcal{C}_{(1)}^-| \leq |\mathcal{C}_{(2)}^+|) > 0. \tag{4.27}$$

If (4.27) yields a contradiction, then (4.26) is proved for $K = 1$. To this end, first note that $(n^{-\rho}(|\mathcal{C}_{(i)}^-|, |\mathcal{C}_{(i)}^+|)_{i \geq 1})_{n \geq 1}$ 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

$$n_k^{-\rho}(|\mathcal{C}_{(i)}^-|, |\mathcal{C}_{(i)}^+|)_{i \geq 1} \xrightarrow{d} (\gamma_i, \bar{\gamma}_i)_{i \geq 1} \quad \text{in } (\ell_{\downarrow}^2)^2,$$

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

$$\lim_{n_k \rightarrow \infty} \mathbb{P}(|\mathcal{C}_{(1)}^-| \leq |\mathcal{C}_{(2)}^+|) = \mathbb{P}(\gamma_1 \leq \bar{\gamma}_2). \tag{4.28}$$

□

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

Proof. Under the coupling in Proposition 3, $\sum_{j \leq i} |\mathcal{C}_{(j)}^-| \leq \sum_{j \leq i} |\mathcal{C}_{(j)}^+|$ 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, $\gamma_1 \leq \bar{\gamma}_1$ almost surely. But, since $\gamma_1, \bar{\gamma}_1$ 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, (4.28) reduces to

$$\lim_{n_k \rightarrow \infty} \mathbb{P}(|\mathcal{C}_{(1)}^-| \leq |\mathcal{C}_{(2)}^+|) = \mathbb{P}(\gamma_1 \leq \gamma_2) = \mathbb{P}(\gamma_1 = \gamma_2) = 0, \tag{4.29}$$

where the last equality follows from Definition 1(f) and Proposition 2. Note that (4.29) contradicts (4.27), and thus (4.26) follows for $K = 1$. For $K \geq 2$, we can use a similar argument to show that, with high probability, $\cup_{i \leq K} \mathcal{C}_{(i)}^- \subset \cup_{i \leq K} \mathcal{C}_{(i)}^+$. If both $\mathcal{C}_{(1)}^-$ and $\mathcal{C}_{(2)}^-$ are contained in $\mathcal{C}_{(1)}^+$, then $|\mathcal{C}_{(1)}^+| \geq |\mathcal{C}_{(1)}^-| + |\mathcal{C}_{(2)}^-|$, 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, $\mathcal{C}_{(2)}^- \subset \mathcal{C}_{(2)}^+$ with high probability and we can use similar arguments to conclude (4.26) for $i \leq K$.

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

$$\lim_{n \rightarrow \infty} \mathbb{P}(\mathcal{C}_{(i)}^- \subset \mathcal{C}_{(i)} \subset \mathcal{C}_{(i)}^+, \forall i \leq K) = 1. \tag{4.30}$$

If $\mathcal{C}_{(1)}$ is not contained in $\mathcal{C}_{(1)}^+$, then $|\mathcal{C}_{(1)}| \leq |\mathcal{C}_{(2)}^+|$. However, since $|\mathcal{C}_{(1)}^-| \leq |\mathcal{C}_{(1)}|$, it follows that $|\mathcal{C}_{(1)}^-| \leq |\mathcal{C}_{(2)}^+|$. Now, one can repeat identical argument as in (4.26) to prove that $\mathcal{C}_{(i)} \subset \mathcal{C}_{(i)}^+$ for all $i \leq K$ with high probability. Moreover, since $\mathcal{G}_n(p_c(1 - \varepsilon_n)) \subset \text{CM}_n(\mathbf{d})$ and $\mathcal{C}_{(i)}^- \subset \mathcal{C}_{(i)}^+$ for all $i \leq K$ with high probability, it must also be the case that $\mathcal{C}_{(i)}^- \subset \mathcal{C}_{(i)} \subset \mathcal{C}_{(i)}^+$ for all $i \leq K$ with high probability. Thus we conclude (4.30). Finally, since \mathbf{Z}_n^- and \mathbf{Z}_n^+ have the same distributional limit, it follows using (4.26) that, for all $i \leq K$,

$$|\mathcal{C}_{(i)}^+| - |\mathcal{C}_{(i)}^-| = o_{\mathbb{P}}(n^{\rho}) \quad \text{and} \quad \text{SP}(\mathcal{C}_{(i)}^+) - \text{SP}(\mathcal{C}_{(i)}^-) \xrightarrow{\mathbb{P}} 0.$$

Thus, (4.30) yields

$$||\mathcal{C}_{(i)}^+| - |\mathcal{C}_{(i)}|| = o_{\mathbb{P}}(n^{\rho}) \quad \text{and} \quad |\text{SP}(\mathcal{C}_{(i)}^+) - \text{SP}(\mathcal{C}_{(i)})| \xrightarrow{\mathbb{P}} 0.$$

Moreover, since both $(\mathbf{Z}_n^-)_{n \geq 1}$ and $(\mathbf{Z}_n^+)_{n \geq 1}$ are tight in $\mathbb{U}_{\downarrow}^0$, it also follows that $(\mathbf{Z}_n)_{n \geq 1}$ is tight in $\mathbb{U}_{\downarrow}^0$. Thus (4.25) follows and the proof of Theorem 1 is now complete. □

4.3. *Analysis of the diameter.* In this section, we investigate the asymptotics of the diameter of $\mathcal{G}_n(p_c(\lambda))$. As in the proof of Theorem 1, an application of Proposition 3 yields the diameter of $\text{CM}_n(\mathbf{d}, p_c(\lambda))$ and completes the proof.

Proof of Theorem 2. First let us fix $\lambda < 1$ and use path counting. Let P_l denote the number of paths of length l in $\mathcal{G}_n(p_c(\lambda))$. Since $\lambda < 1$, we have that $\tilde{v}_n = \lambda(1 + o_{\mathbb{P}}(1)) < 1$ with high probability. Now, an application of [40, Lemma 5.1] yields that for all $l \geq 1$, $\tilde{\mathbb{E}}[P_l] \leq \tilde{\ell}_n(\tilde{v}_n)^{l-1}$. Thus, on the event $\{\tilde{v}_n < 1\}$, for any $K \geq 1$,

$$\tilde{\mathbb{P}}(\text{diam}(\mathcal{G}_n(p_c(\lambda))) > K) \leq \sum_{l>K} \tilde{\mathbb{E}}[P_l] \leq \frac{\tilde{\ell}_n(\tilde{v}_n)^K}{1 - \tilde{v}_n} \tag{4.31}$$

Now, taking $K = C \log n$ for some large constant $C > 0$ gives the desired $\log n$ bound on the diameter of $\mathcal{G}_n(p_c(\lambda))$ with high probability for $\lambda < 1$.

To extend to the case $\lambda \geq 1$, we delete R highest-degree vertices to obtain a new graph \mathcal{G}_n^R . Using (4.17), \mathcal{G}_n^R is a configuration model with the criticality parameter $v_n^R < 1$ with high probability. Thus the above result applies for \mathcal{G}_n^R . However, after putting back the R deleted vertices, the diameter of $\mathcal{G}^{>R}$ can increase by at most a factor of R . This implies the $\log n$ bound on the diameter of $\mathcal{G}_n(p_c(\lambda))$ with high probability for $\lambda \geq 1$. Finally, as remarked in the beginning of this section, the proof of Theorem 2 follows by invoking Proposition 3. \square

5. Near-Critical Behavior

Finally we consider the near-critical behavior for $\text{CM}_n(\mathbf{d}, p)$ in this section. The analysis for the barely subcritical and supercritical regimes are given separately in Sects. 5.1 and 5.2 respectively.

5.1. *Barely-subcritical regime.* In this section, we analyze the barely-subcritical regime ($p_n \ll p_c$) for percolation and complete the proof of Theorem 3. Recall the exploration process from Algorithm 3 on the graph $\mathcal{G}_n(p_n)$, starting with vertex j . Let $\mathcal{C}(j, p_n)$ denote the connected component in $\mathcal{G}_n(p_n)$ containing vertex j . We will use the same notation for the quantities defined in Sect. 4.2.1, but the reader should keep in mind that we now deal with different p_n values. We avoid augmenting p_n in the notation for the sake of simplicity. Consider exploring the graph using Algorithm 3 but starting from vertex j . The exploration process \mathbf{S}_n^j is given by

$$S_n^j(0) = \tilde{d}_j, \quad S_n^j(l) = \tilde{d}_j + \sum_{i:i \neq j} \tilde{d}_i \mathcal{S}_i^n(l) - 2l.$$

Thus the exploration process starts from \tilde{d}_j now. Now, for any $u > 0$, as $n \rightarrow \infty$,

$$\sup_{u \leq t} (n^\alpha p_n)^{-1} \left| \sum_{i:i \neq j} \mathcal{S}_i^n(un^\alpha p_n) - un^\alpha p_n \right| \xrightarrow{\mathbb{P}} 0.$$

This follows using identical arguments as in Lemma 7, and thus is skipped here. Consider the re-scaled process \tilde{S}_n^j defined as $\tilde{S}_n^j(t) = (n^\alpha p_n)^{-1} S_n^j(\lfloor tn^\alpha p_n \rfloor)$. Then,

$$\begin{aligned} \tilde{S}_n^j(t) &= (n^\alpha p_n)^{-1} \tilde{d}_j + (n^\alpha p_n)^{-1} \sum_{i:i \neq j} \tilde{d}_i \mathcal{I}_i^n(tn^\alpha p_n) - 2t + o_{\mathbb{P}}(1) \\ &= \theta_j + (n^\alpha p_n)^{-1} \sum_{i:i \neq j} (\tilde{d}_i - 1) \mathcal{I}_i^n(tn^\alpha p_n) - t + o_{\mathbb{P}}(1). \end{aligned}$$

Recall that $\tilde{\mathbb{E}}$ is the conditional expectation conditionally on $(\tilde{d}_i)_{i \in [n]}$. Now, since the vertices are explored in a size-biased manner with the sizes being $(\tilde{d}_i/\tilde{\ell}_n)_{i \in [n]}$, for any $t \geq 0$,

$$\tilde{\mathbb{E}} \left[\frac{1}{n^\alpha p_n} \sum_{i:i \neq j} (\tilde{d}_i - 1) \mathcal{I}_i^n(\lfloor tn^\alpha p_n \rfloor) \right] \leq \frac{tn^\alpha p_n}{n^\alpha p_n \tilde{\ell}_n} \sum_{i \in [n]} \tilde{d}_i (\tilde{d}_i - 1) = o_{\mathbb{P}}(1),$$

where the first inequality uses (4.5), and the final step follows from Lemma 5. Consequently, \tilde{S}_n^j converges in probability to the deterministic process $(\theta_j - t)_{t \in [0, \theta_j]}$. Thus

$$\# \text{ edges in } \mathcal{C}(j, p_n) \xrightarrow{\mathbb{P}} \theta_j. \tag{5.1}$$

Next, the proof above shows that $\max_{l \leq \theta_j n^\alpha p_n} S_n^j(l) \leq 2\theta_j n^\alpha p_n$ with high probability. Thus, the probability of creating a surplus edge at each step is at most $2\theta_j n^\alpha p_n / \tilde{\ell}_n$. This implies that the probability of creating at least one surplus edge before $\theta_j n^\alpha p_n$ is at most $2\theta_j^2 n^{2\alpha} p_n^2 / \tilde{\ell}_n = O_{\mathbb{P}}(n^{2\alpha-1} p_n) = o_{\mathbb{P}}(1)$. Together with (5.1) yields

$$(n^\alpha p_n)^{-1} |\mathcal{C}(j, p_n)| \xrightarrow{\mathbb{P}} \theta_j, \quad \text{and} \quad \mathbb{P}(\text{SP}(\mathcal{C}(j, p_n)) = 0) \rightarrow 1.$$

From (5.1), we can also conclude that $\lim_{n \rightarrow \infty} \mathbb{P}(i \in \mathcal{C}(j)) = 0$ for all $i, j \geq 1$ and $i \neq j$, since, if $i \in \mathcal{C}(j)$, then the number of edges in $\mathcal{C}(j, p_n)$ is at least $\tilde{d}_i + \tilde{d}_j = n^\alpha p_n(\theta_i + \theta_j)$. Thus, $\mathcal{C}(i, p_n)$ and $\mathcal{C}(j, p_n)$ are disjoint with high probability.

To conclude Theorem 3, we show that the rescaled vector of ordered component sizes is tight in ℓ_1^2 . This tightness also yields that, for each fixed $j \geq 1$,

$$|\mathcal{C}(j, p_n)| = |\mathcal{C}_{(j)}(p_n)|, \quad \text{with high probability.}$$

To show ℓ_1^2 -tightness, it is enough to show that, for any $\varepsilon > 0$,

$$\lim_{K \rightarrow \infty} \limsup_{n \rightarrow \infty} \mathbb{P} \left(\sum_{i>K} |\mathcal{C}_{(i)}(p_n)|^2 > \varepsilon n^{2\alpha} p_n^2 \right) = 0.$$

This can be concluded using identical arguments as in the proof of Proposition 4 above. The proof of Theorem 3 is now complete.

5.2. *Barely-supercritical regime.* In this section, we provide the proof of Theorem 4. Let $p_n = \lambda_n n^{-\eta}$, where $\lambda_n \rightarrow \infty$ since $p_n \gg p_c(\lambda)$. Our main tool here is a general result [38, Theorem 5.4], that provides asymptotics of the component sizes, if one can verify certain properties of an associated exploration process. Using Proposition 3, it is enough to prove Theorem 4 for the graph $\mathcal{G}_n(p_n)$ generated by Algorithm 1. Let $\tilde{\mathbf{d}}$ denote the degree sequence obtained after performing Algorithm 1 (S1). Thus, $\mathcal{G}_n(p_n)$ is distributed as $\text{CM}_n(\tilde{\mathbf{d}})$. We will verify Assumptions (B1)–(B8) from [38] on the graph $\mathcal{G}_n(p_n)$, which allows us to conclude Theorem 4 from [38, Theorem 5.4]. We start by describing the following exploration process on $\mathcal{G}_n(p_n)$ from [38, Sect. 5.1]:

Algorithm 4. (S0) Associate an independent Exponential(1) clock ξ_e to each half-edge e . Any half-edge can be in one of the *states* among sleeping, active, and dead. Initially at time 0, all the half-edges are sleeping. Whenever the set of active half-edges is empty, select a sleeping half-edge e uniformly at random among all sleeping half-edges and declare it to be active. If e is incident to v , then declare all the other half-edges of v to be active as well. The process stops when there is no sleeping half-edge left; the remaining sleeping vertices are all isolated and we have explored all other components.

- (S1) Pick an active half-edge (which one does not matter) and kill it, i.e., change its status to dead.
- (S2) Wait until the next half-edge dies (spontaneously). This half-edge is paired to the one killed in the previous step (S1) to form an edge of the graph. If the vertex it belongs to is sleeping, then we declare this vertex awake and all of its other half-edges active. Repeat from (S1) if there is any active half-edge; otherwise from (S0).

Denote the number of living half-edges upto time t by $L_n(t)$. Let $\tilde{V}_{n,k}(t)$ denote the number of sleeping vertices of degree k such that all the k associated exponential clocks ring after time t . Define

$$\tilde{V}_n(t) = \sum_{k=1}^{\infty} \tilde{V}_{n,k}(t), \quad \tilde{S}_n(t) = \sum_{k=1}^{\infty} k \tilde{V}_{n,k}(t), \quad \tilde{A}_n(t) = L_n(t) - \tilde{S}_n(t). \quad (5.2)$$

We show that Assumptions (B1)–(B8) from [38] hold with

$$\zeta = \kappa^{\frac{1}{3-\tau}}, \quad \gamma_n = \beta_n = p_n^{\frac{\tau-2}{3-\tau}}, \quad \psi(t) = \kappa t^{\tau-2} - t, \quad \hat{g}(t) = t, \quad \hat{h}(t) = \kappa t^{\tau-2} + t,$$

where we recall the definition of κ from (2.10). The ζ in our notation corresponds to τ in the notation of [38, Theorem 5.4]. We have used ζ instead of τ , since in our paper τ denotes the power-law exponent.

We first find the number of vertices in $\mathcal{G}_n(p_n)$. Let $\tilde{n} := \#\{i : \tilde{d}_i \geq 1\}$. Recall that V_n is a vertex chosen uniformly at random from $[n]$ and let $D_n = d_{V_n}$ be the degree of V_n in $\text{CM}_n(\mathbf{d})$. Note that

$$\mathbb{E}[\tilde{n}] = \mathbb{E}\left[\sum_{i \in [n]} \mathbb{1}_{\{\tilde{d}_i \geq 1\}}\right] = \sum_{i \in [n]} (1 - (1 - p_n)^{d_i}) = n \mathbb{E}[1 - (1 - p_n)^{D_n}]. \quad (5.3)$$

Using that $1 - (1 - x)^k \leq kx$ for any $k \geq 1$ and $x \in (0, 1)$, we have $\mathbb{E}[\tilde{n}] \leq n\mathbb{E}[D_n]$. Also, using $1 - (1 - x)^k \geq kx - k^2x^2/2$ for any $kx < 1$, $k \geq 1$ and $x \in (0, 1)$,

$$\begin{aligned} \mathbb{E}[1 - (1 - p_n)^{D_n}] &\geq \mathbb{E}[1 - (1 - p_n)^{D_n} \mathbb{1}_{\{p_n D_n < 1\}}] \\ &\geq p_n \mathbb{E}[D_n \mathbb{1}_{\{p_n D_n < 1\}}] - \frac{p_n^2}{2} \mathbb{E}[D_n^2 \mathbb{1}_{\{p_n D_n < 1\}}] \\ &= p_n \mathbb{E}[D_n] - p_n \mathbb{E}[D_n \mathbb{1}_{\{p_n D_n \geq 1\}}] - \frac{p_n^2}{2} \mathbb{E}[D_n^2 \mathbb{1}_{\{p_n D_n < 1\}}]. \end{aligned}$$

Using Assumption 2 (ii), $(D_n)_{n \geq 1}$ is uniformly integrable and thus $\mathbb{E}[D_n \mathbb{1}_{\{p_n D_n \geq 1\}}] = o(1)$, where in the last step we have used that $p_n \ll 1$. For the third term, since $(D_n)_{n \geq 1}$ is uniformly integrable, we have that $(D_n)_{n \geq 1}$ is also tight. Thus, $p_n D_n \xrightarrow{\mathbb{P}} 0$. Using the uniform integrability of $(D_n)_{n \geq 1}$ again together with $p_n D_n \mathbb{1}_{\{p_n D_n < 1\}} \leq 1$ and $p_n D_n \xrightarrow{\mathbb{P}} 0$, we conclude that $\mathbb{E}[D_n \times (p_n D_n \mathbb{1}_{\{p_n D_n < 1\}})] \rightarrow 0$. From (5.3), and Assumption 2 (ii), we now conclude that

$$\mathbb{E}[\tilde{n}] = np_n(\mu + o(1)). \quad (5.4)$$

Further, using standard concentration inequalities for sums of independent Bernoulli random variables [42, (2.9), Theorem 2.8], it follows that

$$\mathbb{P}(|\tilde{n} - \mathbb{E}[\tilde{n}]| > \log n \sqrt{np_n}) \leq 2e^{-C(\log n)^2}, \quad (5.5)$$

for some constant $C > 0$. In what follows, we will often use (5.4) and (5.5) to replace \tilde{n} by $np_n\mu$.

Conditions (B1)–(B4) [38] are straightforward. (B8) follows using $\max_{i \in [n]} \tilde{d}_i = O_{\mathbb{P}}(n^\alpha p_n) = o_{\mathbb{P}}(\tilde{n}\gamma_n)$. To verify Conditions (B5)–(B7), we first obtain below the asymptotics of the mean-curve and then show that the processes \tilde{S}_n , \tilde{V}_n , \tilde{A}_n remain uniformly close to their expected curves. These are summarized in the following two propositions:

Proposition 7. *For any fixed $u > 0$, as $n \rightarrow \infty$,*

$$\sup_{t \leq u} \left| \frac{1}{np_n\mu\beta_n} (\mathbb{E}[\tilde{S}_n(0)] - \mathbb{E}[\tilde{S}_n(\beta_n t)]) - \hat{h}(t) \right| \rightarrow 0, \quad (5.6)$$

$$\sup_{t \leq u} \left| \frac{1}{np_n\mu\beta_n} (\mathbb{E}[\tilde{V}_n(0)] - \mathbb{E}[\tilde{V}_n(\beta_n t)]) - \hat{g}(t) \right| \rightarrow 0,$$

$$\sup_{t \leq u} \left| \frac{1}{np_n\mu\gamma_n} \mathbb{E}[\tilde{A}_n(\beta_n t)] - \psi(t) \right| \rightarrow 0. \quad (5.7)$$

Proposition 8. *For any fixed $u > 0$, as $n \rightarrow \infty$, all the terms $\sup_{t \leq u} |\tilde{S}_n(\beta_n t) - \mathbb{E}[\tilde{S}_n(\beta_n t)]|$, $\sup_{t \leq u} |\tilde{V}_n(\beta_n t) - \mathbb{E}[\tilde{V}_n(\beta_n t)]|$, and $\sup_{t \leq u} |\tilde{A}_n(\beta_n t) - \mathbb{E}[\tilde{A}_n(\beta_n t)]|$ are $o_{\mathbb{P}}(np_n\beta_n)$ (and thus $o_{\mathbb{P}}(np_n\gamma_n)$).*

To prove Propositions 7 and 8, we make crucial use of the following lemma:

Lemma 11. *For any $t > 0$, as $n \rightarrow \infty$,*

$$\mathbb{E} \left[\sum_{i \in [n]} \tilde{d}_i e^{-t\beta_n \tilde{d}_i} \right] = (1 + o(1)) p_n e^{-t\beta_n} \sum_{i \in [n]} d_i e^{-t\beta_n p_n d_i},$$

$$\mathbb{E}\left[\sum_{i \in [n]} e^{-t\beta_n \tilde{d}_i} \mathbb{1}_{\{\tilde{d}_i \geq 1\}}\right] = (1 + o(1)) \sum_{i \in [n]} (e^{-t\beta_n p_n d_i} - (1 - p_n)^{d_i}). \quad (5.8)$$

Proof. Note that if $X \sim \text{Bin}(m, p)$, then

$$\mathbb{E}[Xe^{-sX}] = mpe^{-s}(1 - p + pe^{-s})^{m-1}.$$

Putting $m = d_i$, $p = p_n$, and $s = t\beta_n$, it follows that

$$\begin{aligned} \mathbb{E}[\tilde{d}_i e^{-t\beta_n \tilde{d}_i}] &= d_i p_n e^{-t\beta_n} (1 - p_n(1 - e^{-t\beta_n}))^{d_i-1} = (1 + o(1)) d_i p_n e^{-t\beta_n} (1 - p_n t \beta_n)^{d_i} \\ &= (1 + o(1)) d_i p_n e^{-t\beta_n} e^{-t\beta_n p_n d_i}. \end{aligned} \quad (5.9)$$

To prove (5.8), note that $\mathbb{E}[e^{-sX} \mathbb{1}_{\{X \geq 1\}}] = \mathbb{E}[e^{-sX}] - \mathbb{P}(X = 0)$. The proof of (5.8) now follows similarly. \square

Proof of Proposition 7. Note that, by Lemma 11,

$$\begin{aligned} \mathbb{E}[\tilde{S}_n(\beta_n t)] &= \mathbb{E}\left[\sum_{i \in [n]} \tilde{d}_i e^{-t\beta_n \tilde{d}_i}\right] = (1 + o(1)) \ell_n p_n e^{-t\beta_n} \mathbb{E}[e^{-t\beta_n p_n D_n^*}], \\ \mathbb{E}[\tilde{V}_n(\beta_n t)] &= \mathbb{E}\left[\sum_{i \in [n]} e^{-t\beta_n \tilde{d}_i} \mathbb{1}_{\{\tilde{d}_i \geq 1\}}\right] = (1 + o(1)) n (\mathbb{E}[e^{-t\beta_n p_n D_n} - (1 - p_n)^{D_n}]), \end{aligned} \quad (5.10)$$

where D_n^* has a size-biased distribution with the sizes being $(d_i/\ell_n)_{i \in [n]}$, and D_n is the degree of a vertex chosen uniformly at random from $[n]$. By the convergence of $\mathbb{E}[D_n]$ in Assumption 1,

$$\mathbb{E}[\tilde{V}_n(0)] - \mathbb{E}[\tilde{V}_n(\beta_n t)] = (1 + o(1)) n \mathbb{E}[1 - e^{-t\beta_n p_n D_n}] = (1 + o(1)) t n \beta_n p_n \mu,$$

where the asymptotics of $n \mathbb{E}[1 - e^{-t\beta_n p_n D_n}]$ follows using identical arguments as (5.3). Further, by using (2.10),

$$\begin{aligned} \mathbb{E}[\tilde{S}_n(0)] - \mathbb{E}[\tilde{S}_n(\beta_n t)] &= (1 + o(1)) \ell_n p_n \mathbb{E}[1 - e^{-t\beta_n} e^{-t\beta_n p_n D_n^*}] \\ &= (1 + o(1)) \ell_n p_n \mathbb{E}[1 - (1 - t\beta_n + o(\beta_n)) e^{-t\beta_n p_n D_n^*}] \\ &= (1 + o(1)) \ell_n p_n (\mathbb{E}[1 - e^{-t\beta_n p_n D_n^*}] + t\beta_n + o(\beta_n)) \\ &= (1 + o(1)) n \mu p_n \beta_n (\kappa t^{\tau-2} + t + o(1)). \end{aligned} \quad (5.11)$$

Thus, (5.6) and (5.7) follows. Moreover, $L_n(t)$ is a pure death process, where $L_n(0) = \sum_{i \in [n]} \tilde{d}_i$, and the jumps occur at rate $L_n(t)$, and at each jump $L_n(t)$ decreases by 2. Therefore, $\mathbb{E}[L_n(t)] = \mathbb{E}[L_n(0)]e^{-2t}$ and consequently, by (5.2) and (5.9),

$$\begin{aligned} \mathbb{E}[\tilde{A}(\beta_n t)] &= \ell_n p_n (e^{-2\beta_n t} - e^{-\beta_n t} \mathbb{E}[e^{-t\beta_n p_n D_n^*}]) + o(n\beta_n p_n) \\ &= n \mu p_n \gamma_n (\kappa t^{\tau-2} - t) + o(n\beta_n p_n). \end{aligned}$$

Thus the proof follows. \square

Proof of Proposition 8. Let us consider \tilde{S}_n only; the other inequalities follow using identical arguments. We will show that

$$\mathbb{E}\left[\sup_{t \leq u\beta_n} |\tilde{S}_n(t) - \mathbb{E}[\tilde{S}_n(t)]|^2\right] = o((np_n\beta_n)^2), \tag{5.12}$$

then an application of Markov’s inequality completes the proof. To prove (5.12), we will use [38, Lemma 5.15], which says that

$$\mathbb{E}\left[\sup_{t \leq u\beta_n} |\tilde{S}_n(t) - \mathbb{E}[\tilde{S}_n(t)]|^2\right] \leq C\mathbb{E}\left[\sum_{i \in [n]} \tilde{d}_i^2 \min\{\tilde{d}_i u\beta_n, 1\}\right]. \tag{5.13}$$

Although, [38, Lemma 5.15] was stated under Assumptions (A1)–(A4) of this paper, this particular proof does not use this assumption. The proof only uses [38, Lemma 4.2]. Indeed, the deductions in (5.62)–(5.65) of [38] does not require any assumption on the degrees. We skip redoing the proof of (5.13) here. Using the fact that $1 - e^{-x} \geq (1 \wedge x)/3$ in (5.13), it follows that

$$\mathbb{E}\left[\sup_{t \leq u\beta_n} |\tilde{S}_n(t) - \mathbb{E}[\tilde{S}_n(t)]|^2\right] \leq C\mathbb{E}\left[\sum_{i \in [n]} \tilde{d}_i^2 (1 - e^{-u\beta_n \tilde{d}_i})\right].$$

Now, using standard concentration inequalities for tails of binomial distributions [42, Theorem 2.1], for any $i \in [n]$,

$$\mathbb{P}(\tilde{d}_i > 2d_1 p_n) \leq C e^{-C d_1 p_n} = C e^{-C n^\rho \lambda_n},$$

where $\lambda_n = p_n n^\rho \rightarrow \infty$. Therefore $\max_{i \in [n]} \tilde{d}_i \leq 2d_1 p_n$, almost surely. Thus,

$$\begin{aligned} \frac{1}{(\ell_n p_n \beta_n)^2} \mathbb{E}\left[\sup_{t \leq u\beta_n} |\tilde{S}_n(t) - \mathbb{E}[\tilde{S}_n(t)]|^2\right] &\leq \frac{C 2d_1 p_n}{(\ell_n p_n \beta_n)^2} \mathbb{E}\left[\sum_{i \in [n]} \tilde{d}_i (1 - e^{-u\beta_n \tilde{d}_i})\right] \\ &\leq \frac{C 2d_1 p_n \ell_n p_n}{(\ell_n p_n \beta_n)^2} \mathbb{E}[1 - e^{-u\beta_n p_n D_n^*}], \end{aligned} \tag{5.14}$$

where the last step follows using (5.10). The final term in (5.14) can be shown to be $O(\beta_n)$ using identical computations as (5.11). Thus,

$$\frac{1}{(\ell_n p_n \beta_n)^2} \mathbb{E}\left[\sup_{t \leq u\beta_n} |\tilde{S}_n(t) - \mathbb{E}[\tilde{S}_n(t)]|^2\right] \leq \frac{C 2d_1 p_n \ell_n p_n \beta_n}{(\ell_n p_n \beta_n)^2} = O(d_1/n\beta_n) = O(\lambda_n^{-\frac{\rho-2}{3-\rho}}) = o(1),$$

since $\lambda_n \rightarrow \infty$, as $n \rightarrow \infty$. Thus the proof follows. \square

Proof of Theorem 4. The proof follows by applying [38, Theorem 5.4]. Propositions 7, 8 verify conditions (B5)–(B7) in [38], and the rest of the conditions are straightforward to verify. \square

Acknowledgements. We sincerely thank the referee for an extremely thorough review, and in particular for pointing out an error in the properties of the scaling limit. Also, we sincerely thank Shankar Bhamidi and Debankur Mukherjee for carefully reading the revised proof in Sect. 3.

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

A. Appendix

A.1. Path counting. Recall the notation from in Sect. 4.2.2. We complete the proof of (4.18) using path-counting techniques for configuration models from [40, Lemma 5.1]. Let $\mathcal{A}_l(v, k)$ denote the event that there exists a path of length l from v to k in the graph \mathcal{G}_n^K . Also, let P_l denote the number of paths of length l . Notice that

$$\begin{aligned} & \tilde{\mathbb{E}} \left[\sum_{k \in [n]} (\tilde{d}_k - 1) \mathbb{1}_{\{V_n^{*,K} \rightsquigarrow k\}} \middle| V_n^{*,K} = v \right] \\ & \leq \tilde{d}_v - 1 + \sum_{l=1}^{(\log n)^2} \sum_{k \in [n]} (\tilde{d}_k - 1) \tilde{\mathbb{P}}(\mathcal{A}_l(v, k)) + \max_{k \in [n]} (\tilde{d}_k - 1) \times n \sum_{l \geq (\log n)^2} \tilde{\mathbb{E}}[P_l]. \end{aligned} \quad (\text{A.1})$$

Let $\mathfrak{J}_l(v, k)$ denote the collection of $\mathbf{x} = (x_i)_{0 \leq i \leq l}$ such that $x_0 = v, x_l = k$ and the x_i 's are distinct. Then, an identical argument to the proof of [40, Lemma 5.1] shows that, for $l = o(n^{2\rho})$, the expected number of paths of length exactly l starting from vertex v and ending at k is given by

$$\sum_{\mathbf{x} \in \mathfrak{J}_l(v, k)} \frac{d'_{x_0} d'_{x_l} \prod_{i=1}^{l-1} d'_{x_i} (d'_{x_i} - 1)}{(\ell'_n - 1) \cdots (\ell'_n - 2l + 1)} \leq \frac{d'_v \ell'_n}{\ell'_n - 2l + 3} (v_n^K)^{l-1} = \left(1 + o_{\mathbb{P}}\left(\frac{l}{\tilde{\ell}_n}\right)\right) d'_v (v_n^K)^{l-1},$$

where $\ell'_n = \sum_{i \in [n]} d'_i$. Recall that $\ell'_n = \tilde{\ell}_n(1 + o_{\mathbb{P}}(1))$. Thus, the second term in (A.1) is at most

$$\begin{aligned} & \sum_{l=1}^{(\log n)^2} \sum_{k \in [n]} (\tilde{d}_k - 1) \sum_{x_i \neq x_j, \forall i \neq j} \frac{d'_v d'_k \prod_{i=1}^{l-1} d'_{x_i} (d'_{x_i} - 1)}{(\ell'_n - 1) \cdots (\ell'_n - 2l + 1)} \\ & \leq (1 + o_{\mathbb{P}}(1)) \tilde{d}_v \left(\frac{1}{\tilde{\ell}_n} \sum_{k \in [n]} d'_k (\tilde{d}_k - 1) \right) \sum_{l=1}^{\infty} (v_n^K)^{l-1} \\ & \leq (1 + o_{\mathbb{P}}(1)) \tilde{d}_v \left(\frac{1}{\tilde{\ell}_n} \sum_{k > K} \tilde{d}_k (\tilde{d}_k - 1) \right) \sum_{l=1}^{\infty} (v_n^K)^{l-1} \leq (1 + o_{\mathbb{P}}(1)) \frac{\tilde{d}_v \mathbb{E}[\tilde{d}_{V_n^{*,K}} - 1]}{1 - v_n^K}, \end{aligned}$$

where in the one-but-last step we have used $d'_i = 0$ for $i \leq K$, $d'_i \leq \tilde{d}_i$ for $i > K$ and $v_n^K < 1$. The third term in (A.1) is $o_{\mathbb{P}}(1)$ uniformly over v by (4.31). Thus the proof of (4.18) follows. \square

A.2. Convergence of process tracking surplus. In this section, we complete the proof of Lemma 9. We first argue that, for any fixed $u > 0$,

$$(\tilde{N}_n^\lambda(u))_{n \geq 1} \text{ is tight in } \mathbb{R}_+. \quad (\text{A.2})$$

Fix $\varepsilon > 0$. Recall the asymptotics from Lemma 5 which will be used throughout the proof. Also, recall that $\tilde{\mathbb{P}}$ and $\tilde{\mathbb{E}}$ respectively denote the conditional probability and expectation conditionally on $(\tilde{d}_i)_{i \in [n]}$. To simplify writing, when we write bounds on the

conditionals probabilities $\tilde{\mathbb{P}}$ and $\tilde{\mathbb{E}}$, we always implicitly assume that the bounds hold with high probability. Recall from (4.23) that the compensator of $\tilde{\mathbf{N}}_n^\lambda$ is approximately proportional to $\text{refl}(\tilde{\mathbf{S}}_n) \xrightarrow{d} \text{refl}(\mathbf{S}_\infty^\lambda)$, where the distributional convergence follows using Theorem 5 and the continuity of the reflection map (see [56, Lemma 13.5.1]). We write A_i denote the number of active half-edges after stage i while implementing Algorithm 3. Thus $n^{-\rho} A_{\lfloor tn^\rho \rfloor} = \text{refl}(\tilde{S}_n(t))$. Using the fact that the supremum of a process is continuous with respect to the Skorohod J_1 -topology [56, Theorem 13.4.1], we can choose $K \geq 1$ large enough so that for all sufficiently large n

$$\tilde{\mathbb{P}}\left(\sup_{i \leq \lfloor un^\rho \rfloor} A_i > Kn^\rho\right) < \varepsilon. \tag{A.3}$$

Fix times $0 < l_1 < \dots < l_m \leq \lfloor un^\rho \rfloor$, 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 $A_{l_j-1} \leq Kn^\rho$ for all $j \in [m]$. Then,

$$\begin{aligned} \tilde{\mathbb{P}}\left(\sum_{i=2}^{\lfloor un^\rho \rfloor} \xi_i \geq m, \text{ and } \sup_{i \leq \lfloor un^\rho \rfloor} A_i \leq Kn^\rho\right) &\leq \sum_{0 < l_1 < \dots < l_m \leq \lfloor un^\rho \rfloor} \tilde{\mathbb{P}}(\mathcal{A}(l_1, \dots, l_m)) \\ &\leq \sum_{0 < l_1 < \dots < l_m \leq \lfloor un^\rho \rfloor} \tilde{\mathbb{E}}[\tilde{\mathbb{P}}(\text{surplus created at } l_m | \mathcal{F}_{l_m-1}) \mathbb{1}_{\{A_{l_m-1} \leq Kn^\rho\}} \mathbb{1}_{\mathcal{A}(l_1, \dots, l_{m-1})}] \\ &\leq \frac{Kn^\rho}{\tilde{\ell}_n - 2 \lfloor un^\rho \rfloor + 1} \sum_{0 < l_1 < \dots < l_m \leq \lfloor un^\rho \rfloor} \tilde{\mathbb{P}}(\mathcal{A}(l_1, \dots, l_{m-1})). \end{aligned}$$

Continuing the iteration in the last step, it follows that with high probability

$$\tilde{\mathbb{P}}\left(\sum_{i=2}^{\lfloor un^\rho \rfloor} \xi_i \geq m, \text{ and } \sup_{i \leq \lfloor un^\rho \rfloor} A_i \leq Kn^\rho\right) \leq (1 + o(1)) \left(\frac{Kn^\rho}{\tilde{\ell}_n}\right)^m \frac{(\lfloor un^\rho \rfloor)_m}{m!}, \tag{A.4}$$

where $(n)_m = n(n-1) \dots (n-m+1)$. The last term in (A.4) tends to zero in the iterated limit $\lim_{m \rightarrow \infty} \limsup_{n \rightarrow \infty}$. An application of (A.3) now yields (A.2).

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

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

Let $\tilde{S}'_n(t) = n^{-\rho} S'_n(\lfloor tn^\rho \rfloor)$. 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 $\varepsilon > 0$,

$$\lim_{n \rightarrow \infty} \mathbb{P}(d_{J_1, T}(\tilde{S}'_n, \tilde{\mathbf{S}}_n) > \varepsilon) = 0. \tag{A.5}$$

First, let $1 \leq l_1 < \dots < l_K \leq \lfloor Tn^\rho \rfloor$ 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

$$\tilde{\mathbb{P}}\left(\mathcal{A}^c \cap \left\{\sup_{i \leq \lfloor Tn^\rho \rfloor} A_i \leq Kn^\rho\right\}\right) \leq Tn^\rho \left(\frac{Kn^\rho}{\tilde{\ell}_n}\right)^2 = O(n^{-\rho}), \tag{A.6}$$

and thus using (A.3), $\mathbb{P}(\mathcal{A}^c) \rightarrow 0$. We now restrict ourselves on \mathcal{A} . Putting $l_0 = 0$ and $l_{K+1} = \lfloor Tn^\rho \rfloor + 1$, let

$$\Lambda_n(l) = \begin{cases} l + j - 1 & \text{for } l_{j-1} < l < l_j, \\ l_j + j - 1 & \text{for } l = l_j - 0.5, \\ l_j + j & \text{for } l = l_j. \end{cases} \quad (\text{A.7})$$

$\Lambda_n(t)$ is obtained by linearly interpolating between the values specified by (A.7). 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 (A.2) and (A.6), it immediately follows that

$$\sup_{l \leq Tn^\rho} |\Lambda_n(l) - l| = o_{\mathbb{P}}(n^\rho). \quad (\text{A.8})$$

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

$$\sup_{l \leq Tn^\rho} |S'_n(l) - S_n(\Lambda_n(l))| = o_{\mathbb{P}}(n^\rho). \quad (\text{A.9})$$

Now, (A.5) follows by combining (A.8) and (A.9). We now proceed to complete the proof of Lemma 9. Let set up some notation for the rest of the proof. Fix $T > 0$, $k \geq 0$ and let $\text{Surp}_T = \{l_1, \dots, l_k\}$, where $1 \leq l_1 < l_2 < \dots < l_k \leq \lfloor Tn^\rho \rfloor + k$. Let $(z_l)_{l \leq \lfloor Tn^\rho \rfloor + k}$ be a sequence of integers such that $z_{l_i} = -2$ and $z_l \geq -1$ for $l \notin \{l_1, \dots, l_k\}$. Thus $(z_l)_{l \leq \lfloor Tn^\rho \rfloor + k}$ represents 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, $(z'_l)_{l \leq \lfloor Tn^\rho \rfloor}$ denote the sequence obtained from $(z_l)_{l \leq \lfloor Tn^\rho \rfloor + k}$ by deleting the -2 's. Thus, $(z'_l)_{l \leq \lfloor Tn^\rho \rfloor}$ corresponds to a sample path of S'_n . Recall that $\zeta_l = S_n(l) - S_n(l-1)$. Let $\omega_n \rightarrow \infty$ sufficiently slowly. Thus,

$$\begin{aligned} \tilde{\mathbb{P}}(N_n^\lambda(\lfloor Tn^\rho \rfloor + k) = k | (S'_n(l))_{l \leq \lfloor Tn^\rho \rfloor} = (z'_l)_{l \leq \lfloor Tn^\rho \rfloor}, N_n^\lambda(\lfloor Tn^\rho \rfloor + k) \leq \omega_n) \\ = \sum_{1 \leq l_1 < \dots < l_k \leq Tn^\rho} \mathbb{P} \left(\text{surplus occurs only at times } l_1, \dots, l_k \middle| \begin{matrix} (S'_n(l))_{l \leq Tn^\rho} = (z'_l)_{l \leq Tn^\rho}, \\ N_n^\lambda(\lfloor Tn^\rho \rfloor + k) \leq \omega_n \end{matrix} \right) \\ = \sum_{1 \leq l_1 < \dots < l_k \leq Tn^\rho} \frac{\tilde{\mathbb{P}}(\zeta_l = z_l, \text{ for all } 1 \leq l \leq \lfloor Tn^\rho \rfloor + k)}{\tilde{\mathbb{P}}((S'_n(l))_{l \leq Tn^\rho} = (z'_l)_{l \leq Tn^\rho}, N_n^\lambda(\lfloor Tn^\rho \rfloor + k) \leq \omega_n)}. \end{aligned} \quad (\text{A.10})$$

Define $m_1 = \{i \in [n] : d_i = z_1 + 2\}$, and for $l \notin \text{Surp}_T$, we denote $m_l = \#\{i \in [n] : d_i = z_l + 2\} - \#\{j < l : z_j = z_l\}$. Next, let a_l denote the number of active half-edges at time l when the exploration process takes the path $(z_l)_{l \leq \lfloor Tn^\rho \rfloor + k}$, and $a'_l = S'_n(l) - \min_{j < l} S'_n(j)$. Now,

$$\begin{aligned} \tilde{\mathbb{P}}(\zeta_l = z_l, \forall l \leq \lfloor Tn^\rho \rfloor + k) &= \frac{\prod_{l \notin \text{Surp}_T} m_l \times \prod_{j=1}^k (a_{l_{j-1}} - 1)}{(\tilde{\ell}_n - 1)(\tilde{\ell}_n - 3) \dots (\tilde{\ell}_n - 2 \lfloor Tn^\rho \rfloor - 2k + 1)} \\ &= \frac{\prod_{l \notin \text{Surp}_T} m_l \times \prod_{j=1}^k (a_{l_{j-1}} - 1)}{(\tilde{\ell}_n - 1) \dots (\tilde{\ell}_n - 2 \lfloor Tn^\rho \rfloor + 1)} \times (1 + o_{\mathbb{P}}(1)) \prod_{j=1}^k \frac{a'_{l_{j-1}}}{\tilde{\ell}_n^k}, \end{aligned} \quad (\text{A.11})$$

where the $o_{\mathbb{P}}(1)$ term above is uniform over $k \leq \omega_n = \log n$. Thus,

$$(A.10) = (1 + o(1)) \frac{\sum_{1 \leq l_1 < \dots < l_k \leq \lfloor Tn^\rho \rfloor + k} \prod_{j=1}^k \frac{a'_{l_j-1}}{\tilde{\ell}_n^k}}{\sum_{r=0}^{\omega_n} \sum_{1 \leq l_1 < \dots < l_r \leq \lfloor Tn^\rho \rfloor + r} \prod_{j=1}^r \frac{a'_{l_j-1}}{\tilde{\ell}_n^r}} =: (1 + o(1)) \frac{\beta_{n,k}}{\sum_{r=0}^{\infty} \beta_{n,r}},$$

where $\beta_{n,r} = 0$ for $r > \omega_n$. We write $\tilde{\mu} = \lambda \mu^2 / \sum_i \theta_i^2$, so that $\tilde{\ell}_n = \tilde{\mu} n^{2\rho} (1 + o_{\mathbb{P}}(1))$. Now, using $\text{refl}(\tilde{S}'_n) \xrightarrow{d} \text{refl}(S^\lambda_\infty)$, it follows that

$$\left((\beta_{n,r})_{r \geq 0}, (\tilde{S}'_n(u))_{u \leq T} \right) \xrightarrow{d} \left(\left(\frac{1}{r!} \left(\frac{1}{\tilde{\mu}} \int_0^T \text{refl}(S^\lambda_\infty(u)) du \right)^r \right)_{r \geq 0}, (S^\lambda_\infty(u))_{u \leq T} \right), \quad (A.12)$$

where the convergence of $(\beta_{n,r})_{r \geq 0}$ holds with respect to the product topology on \mathbb{R}^∞ . Next, let us ensure that $\sum_{r=0}^\infty \beta_{n,r}$ in (A.10) converges to the desired quantity. To this end, consider a probability space where the convergence of (A.12) holds almost surely. On this space, $\sup_{l \leq Tn^\rho + k} \text{refl}(S'_n(l)) \leq 2(\sup_{l \leq Tn^\rho + k} S'_n(l) + \omega_n) =: X_n(T)$, and thus

$$\beta_{n,r} \leq \frac{(Tn^\rho + \omega_n)^r X_n(T)^r}{r! \tilde{\ell}_n^r}.$$

Since $n^{-\rho} \sup_{l \leq Tn^\rho + k} S'_n(l)$ converges, an application of Dominated Convergence Theorem yields that

$$\sum_{r \geq 0} \beta_{n,r} \xrightarrow{\text{a.s.}} \sum_{r \geq 0} \frac{1}{r!} \left(\frac{1}{\tilde{\mu}} \int_0^T \text{refl}(S^\lambda_\infty(u)) du \right)^r = \exp \left(\frac{1}{\tilde{\mu}} \int_0^T \text{refl}(S^\lambda_\infty(u)) du \right). \quad (A.13)$$

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

$$\begin{aligned} & \mathbb{E}[\phi_1((\tilde{S}'_n(u))_{u \leq T}) \phi_2(\tilde{N}_n^\lambda(T))] \\ &= \mathbb{E}[\phi_1((\tilde{S}'_n(u))_{u \leq T}) \phi_2(\tilde{N}_n^\lambda(T)) \mathbb{1}_{\{N_n^\lambda(\lfloor Tn^\rho \rfloor + k) \leq \omega_n\}}] + o(1) \\ &= o(1) + \mathbb{E} \left[\phi_1((\tilde{S}'_n(u))_{u \leq T}) \mathbb{1}_{\{N_n^\lambda(\lfloor Tn^\rho \rfloor + k) \leq \omega_n\}} \times (1 + o(1)) \frac{\sum_{k \geq 0} \phi_2(k) \beta_{n,k}}{\sum_{r \geq 0} \beta_{n,r}} \right] \\ &= o(1) + \mathbb{E} \left[\phi_1((\tilde{S}'_n(u))_{u \leq T}) \times \frac{\sum_{k \geq 0} \phi_2(k) \beta_{n,k}}{\sum_{r \geq 0} \beta_{n,r}} \right] \rightarrow \mathbb{E}[\phi_1((S^\lambda_\infty(u))_{u \leq T}) \phi_2(N^\lambda(T))], \end{aligned}$$

where $N^\lambda(T)$, conditionally on $(S^\lambda_\infty(u))_{u \leq T}$, is distributed as Poisson $(\frac{1}{\tilde{\mu}} \int_0^T \text{refl}(S^\lambda_\infty(u)) du)$. We have used (A.2) in the third step, and the final step follows by combining (A.12) and (A.13). Hence, we have shown that, for any $T > 0$,

$$\left((\tilde{S}'_n(u))_{u \leq T}, \tilde{N}_n^\lambda(T) \right) \xrightarrow{d} \left((S^\lambda_\infty(u))_{u \leq T}, N^\lambda(T) \right). \quad (A.14)$$

Next, let $U_1^n < U_2^n < \dots$ denote the location of surplus edges in the process S_n . Then, using (A.11) yields

$$\begin{aligned} & \tilde{\mathbb{P}} \left(U_j^n = l_j, \text{ for all } j \in [k] \mid (\tilde{S}'_n(u))_{u \leq T}, \tilde{N}_n^\lambda(T) = k \right) \\ &= (1 + o(1)) \frac{\frac{1}{\tilde{\ell}_n^k} \prod_{j=1}^k (A_{l_j} - 1)}{\sum_{1 \leq l'_1 < \dots < l'_k \leq \lfloor Tn^\rho \rfloor + k} \frac{1}{\tilde{\ell}_n^k} \prod_{j=1}^k (A_{l'_j} - 1)}. \end{aligned} \quad (A.15)$$

From this, it can be seen that the law of $n^{-\rho}(U_j^n)_{j \in [k]}$, 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 $\frac{\mathbb{1}_{\{u \in [0, T]\}} \text{refl}(S_\infty^\lambda(u))}{\int_0^T \text{refl}(S_\infty^\lambda(u)) du}$. This shows that the location of the occurrence of surplus edges, conditionally on $(\bar{S}'_n(u))_{u \leq T}$, converges in distribution to the location of the points of the Poisson process (2.6) on $[0, T]$ conditionally on $(S_\infty^\lambda(u))_{u \leq T}$. Convergence of the total number of surplus edges created, conditionally on $(\bar{S}'_n(u))_{u \leq T}$, is given by (A.14). Thus combining (A.14) and (A.15), it follows that

$$\left((\bar{S}'_n(u))_{u \leq T}, (\bar{N}_n^\lambda(u))_{u \leq T} \right) \xrightarrow{d} \left((S_\infty^\lambda(u))_{u \leq T}, (N^\lambda(u))_{u \leq T} \right).$$

Now, an application of (A.5) completes the proof of Lemma 9. \square

A.3. Tightness of component sizes and surplus. In this section, we prove Lemma 10. Let V_n^* denote a vertex chosen in a size-biased manner with sizes being $(\tilde{d}_i)_{i \in [n]}$, independently of the graph $\text{CM}_n(\mathbf{d})$. Let $\mathcal{C}(V_n^*)$ denote the component containing V_n^* , $D(V_n^*) = \sum_{k \in \mathcal{C}(V_n^*)} \tilde{d}_k$, and $D_i = \sum_{k \in \mathcal{C}_i} \tilde{d}_k$. Since component sizes corresponding to the components having one vertex and no edges is zero by our convention, $|\mathcal{C}_i| \leq D_i$ for all i . Thus, it is enough to show that, for any $\varepsilon > 0$,

$$\tilde{\mathbb{P}} \left(\sum_{i: D_i \leq \delta n^\rho} D_i \times \text{SP}(\mathcal{C}_i) > \varepsilon n^\rho \right) \xrightarrow{\mathbb{P}} 0,$$

in the iterated limit $\lim_{\delta \rightarrow 0} \limsup_{n \rightarrow \infty}$. The following estimate will be our crucial ingredient. We first prove Lemma 10 using Lemma 12, and the proof of Lemma 12 will come after that.

Lemma 12. *Assume that $\lambda < 1$. Let $\delta_k = \delta k^{-0.12}$. Then, for $\delta > 0$ sufficiently small, with high probability,*

$$\tilde{\mathbb{P}} \left(\text{SP}(\mathcal{C}(V_n^*)) \geq K, D(V_n^*) \in (\delta_K n^\rho, 2\delta_K n^\rho) \right) \leq \frac{C\sqrt{\delta}}{n^\rho K^{1.1}},$$

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

Proof of Theorem 4. First, let us consider the case $\lambda < 1$. Fix any $\varepsilon, \delta > 0$. Note that

$$\begin{aligned} \tilde{\mathbb{P}} \left(\sum_{D_i \leq \delta n^\rho} D_i \text{SP}(\mathcal{C}_i) > \varepsilon n^\rho \right) &\leq \frac{1}{\varepsilon n^\rho} \tilde{\mathbb{E}} \left[\sum_{i=1}^{\infty} D_i \text{SP}(\mathcal{C}_i) \mathbb{1}_{\{D_i \leq \delta n^\rho\}} \right] \\ &= \frac{\tilde{\ell}_n}{\varepsilon n^\rho} \tilde{\mathbb{E}} \left[\text{SP}(\mathcal{C}(V_n^*)) \mathbb{1}_{\{|\mathcal{C}(V_n^*)| \leq \delta n^\rho\}} \right] \\ &= \frac{\tilde{\ell}_n}{\varepsilon n^\rho} \sum_{k=1}^{\infty} \sum_{i \geq \log_2(1/(k^{0.12}\delta))} \tilde{\mathbb{P}} \left(\text{SP}(\mathcal{C}(V_n^*)) \geq k, |\mathcal{C}(V_n^*)| \in (2^{-(i+1)} k^{-0.12} n^\rho, 2^{-i} k^{-0.12} n^\rho) \right) \\ &\leq \frac{C}{\varepsilon} \sum_{k=1}^{\infty} \frac{1}{k^{1.1}} \sum_{i \geq \log_2(1/(k^{0.12}\delta))} 2^{-i/2} \leq \frac{C}{\varepsilon} \sum_{k=1}^{\infty} \frac{\sqrt{\delta}}{k^{1.04}} = O(\sqrt{\delta}/\varepsilon), \end{aligned}$$

where the last-but-second step follows from Lemma 12, and the inequality holds with high probability. The proof of Lemma 10 now follows for the $\lambda < 1$ case.

Now consider the case $\lambda > 1$. Fix a large integer $R \geq 1$ such that $\lambda \sum_{i>R} \theta_i^2 < 1$. This can be done because $\theta \in \ell_{\downarrow}^2$. Using (4.22), for any $\delta_0 > 0$, it is possible to choose $T > 0$ such that

$$\limsup_{n \rightarrow \infty} \mathbb{P}(\text{all the vertices } 1, \dots, R \text{ are explored within time } Tn^\rho) > 1 - \delta_0.$$

Let T_e denote the first time after Tn^ρ when we finish exploring a component. By Theorem 5, $(n^{-\rho} T_e)_{n \geq 1}$ 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}^* . Then using (4.17) and the fact that $\lambda \sum_{i>R} \theta_i^2 < 1$, $\nu_n^* < 1 - \varepsilon_0$ with high probability for some $\varepsilon_0 > 0$. Thus, if $\mathcal{C}_{(i)}^*$ denotes the i -th largest component of \mathcal{G}_T^* , then the argument for $\lambda < 1$ yields

$$\lim_{T \rightarrow \infty} \lim_{\delta \rightarrow 0} \limsup_{n \rightarrow \infty} \mathbb{P}\left(\sum_{i: |\mathcal{C}_{(i)}^*| \leq \delta n^\rho} |\mathcal{C}_{(i)}^*| \times \text{SP}(\mathcal{C}_{(i)}^*) > \varepsilon n^\rho\right) = 0. \tag{A.16}$$

To conclude the proof for the whole graph (with $\lambda > 1$), let

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

Note that

$$\begin{aligned} \sum_{i \in \mathcal{K}_n^T} |\mathcal{C}_{(i)}| \times \text{SP}(\mathcal{C}_{(i)}) &\leq \left(\sum_{i \in \mathcal{K}_n^T} |\mathcal{C}_{(i)}|^2\right)^{1/2} \times \left(\sum_{i \in \mathcal{K}_n^T} \text{SP}(\mathcal{C}_{(i)})^2\right)^{1/2} \\ &\leq \left(\sum_{|\mathcal{C}_{(i)}| \leq \delta n^\rho} |\mathcal{C}_{(i)}|^2\right)^{1/2} \times \text{SP}(T_e), \end{aligned}$$

where $\text{SP}(t)$ is the number of surplus edges explored up to time tn^ρ and we have used the fact that $\sum_{i \in \mathcal{K}_n^T} \text{SP}(\mathcal{C}_{(i)})^2 \leq (\sum_{i \in \mathcal{K}_n^T} \text{SP}(\mathcal{C}_{(i)}))^2 \leq \text{SP}(T_e)^2$. From Lemma 9 and Proposition 4 we can conclude that for any $T > 0$,

$$\lim_{\delta \rightarrow 0} \limsup_{n \rightarrow \infty} \mathbb{P}\left(\sum_{i \in \mathcal{K}_n^T} |\mathcal{C}_{(i)}| \times \text{SP}(\mathcal{C}_{(i)}) > \varepsilon n^\rho\right) = 0. \tag{A.17}$$

The proof is now complete for the case $\lambda > 1$ by combining (A.16) and (A.17). \square

Proof of Lemma 12. We use a generic constant C to denote a positive constant independent of n, δ, K . Consider the graph exploration described in Algorithm 3, 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 (4.3) with $S_n(0) = \tilde{d}_{V_n^*}$. Note that $\mathcal{C}(V_n^*)$ is explored when S_n hits zero, and the hitting time at zero gives $D(V_n^*)/2$. For $H > 0$, let

$$\gamma := \inf\{l \geq 1 : S_n(l) \geq H \text{ or } S_n(l) = 0\} \wedge 2\delta_K n^\rho.$$

Here, we let \mathcal{A} be the intersection of all the events described in Lemma 5, which are shown to hold with high probability. Recall that we write $\mathcal{F}_l = \sigma(\mathcal{S}_i(l) : i \in [n]) \cap \mathcal{A}$. Note that

$$\begin{aligned} \tilde{\mathbb{E}}[S_n(l+1) - S_n(l) \mid \mathcal{F}_l] &= \sum_{i \in [n]} \tilde{d}_i \tilde{\mathbb{P}}(i \notin \mathcal{V}_l, i \in \mathcal{V}_{l+1} \mid (\mathcal{S}_i^n(l))_{i=1}^n) - 2 \\ &= \frac{\sum_{i \notin \mathcal{V}_l} \tilde{d}_i^2}{\tilde{\ell}_n - 2l - 1} - 2 \leq \frac{\sum_{i \in [n]} \tilde{d}_i^2}{\tilde{\ell}_n - 2l - 1} - 2 \\ &:= (\lambda - 1) + \frac{2l + 1}{\tilde{\ell}_n - 2l - 1} \times \frac{\sum_{i \in [n]} \tilde{d}_i^2}{\tilde{\ell}_n} \leq 0, \end{aligned}$$

uniformly over $l \leq 2\delta_K n^\rho$ for all small $\delta > 0$ and large n , where the last step uses that $\lambda < 1$. Therefore, $\{S_n(l)\}_{l=1}^{2\delta_K n^\rho}$ is a super-martingale. The optional stopping theorem now implies

$$\tilde{\mathbb{E}}[d_{V_n^*}] \geq \tilde{\mathbb{E}}[S_n(\gamma)] \geq H \tilde{\mathbb{P}}(S_n(\gamma) \geq H).$$

Thus,

$$\tilde{\mathbb{P}}(S_n(\gamma) \geq H) \leq \frac{\tilde{\mathbb{E}}[d_{V_n^*}]}{H}.$$

Put $H = n^\rho K^{1.1} / \sqrt{\delta}$. 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

$$\begin{aligned} &\tilde{\mathbb{P}}(\text{SP}(\mathcal{C}(V_n^*)) \geq K, D(V_n^*) \in (\delta_K n^\rho, 2\delta_K n^\rho)) \\ &\leq \tilde{\mathbb{P}}(S_n(\gamma) \geq H) + \tilde{\mathbb{P}}(\text{SP}(\mathcal{C}(V_n^*)) \geq K, S_n[0, 2\delta_K n^\rho] < H, S_n[0, \delta_K n^\rho] > 0). \end{aligned} \tag{A.18}$$

Now,

$$\begin{aligned} &\tilde{\mathbb{P}}(\text{SP}(\mathcal{C}(V_n^*)) \geq K, S_n[0, 2\delta_K n^\rho] < H, S_n[0, \delta_K n^\rho] > 0) \\ &\leq \sum_{1 \leq l_1 < \dots < l_K \leq 2\delta_K n^\rho} \tilde{\mathbb{P}}(\text{surpluses occur at times } l_1, \dots, l_K, S_n[0, 2\delta_K n^\rho] < H, S_n[0, \delta_K n^\rho] > 0) \\ &= \sum_{1 \leq l_1 < \dots < l_K \leq 2\delta_K n^\rho} \tilde{\mathbb{E}}[\mathbb{1}_{\{0 < S_n[0, l_{K-1}] < H, \text{SP}(l_{K-1}) = K-1\}} Y], \end{aligned}$$

where

$$\begin{aligned} Y &= \tilde{\mathbb{P}}\left(K^{\text{th}} \text{ surplus occurs at time } l_K, S_n[l_K, 2\delta_K n^\rho] < H, S_n[l_K, \gamma] > 0 \mid \mathcal{F}_{l_{K-1}}\right) \\ &\leq \frac{CK^{1.1}n^\rho}{\tilde{\ell}_n\sqrt{\delta}} \leq \frac{CK^{1.1}}{n^\rho\sqrt{\delta}}. \end{aligned}$$

Therefore, using induction, (A.18) yields

$$\begin{aligned} &\tilde{\mathbb{P}}(\text{SP}(\mathcal{C}(V_n^*)) \geq K, S_n[0, 2\delta_K n^\rho] < H, S_n[0, \delta_K n^\rho] > 0) \\ &\leq C \left(\frac{K^{1.1}}{\sqrt{\delta}n^\rho} \right)^K \frac{(2\delta n^\rho)^{K-1}}{K^{0.12(K-1)}(K-1)!} \sum_{l_1=1}^{2\delta_K n^\rho} \tilde{\mathbb{P}}(D(V_n^*) \geq l_1) \leq C \frac{\delta^{K/2}}{K^{1.1}n^\rho} \tilde{\mathbb{E}}[D(V_n^*)], \end{aligned}$$

where we have used the fact that $\#\{1 \leq l_2, \dots, l_K \leq 2\delta n^\rho\} = (2\delta n^\rho)^{K-1}/(K-1)!$ and Stirling's approximation for $(K-1)!$ in the last step. Since $\lambda < 1$, we can use (4.18) to conclude that, for all sufficiently large n ,

$$\tilde{\mathbb{E}}[D(V_n^*) - 1] \leq C,$$

with high probability for some constant $C > 0$. Thus, we get the desired bound for (A.18). The proof of Lemma 12 is now complete. \square

References

- [1] Addario-Berry, L., Broutin, N., Goldschmidt, C.: The continuum limit of critical random graphs. *Probab. Theory Relat. Fields* **152**(3), 367–406 (2012)
- [2] Albert, R., Barabási, A.L.: Statistical mechanics of complex networks. *Rev. Mod. Phys.* **74**(1), 47–97 (2002)
- [3] Albert, R., Jeong, H., Barabási, A.L.: Error and attack tolerance of complex networks. *Nature* **406**, 378 (2000)
- [4] Aldous, D.: Brownian excursions, critical random graphs and the multiplicative coalescent. *Ann. Probab.* **25**(2), 812–854 (1997)
- [5] Aldous, D., Limic, V.: The entrance boundary of the multiplicative coalescent. *Electron. J. Probab.* **3**(3), 1–59 (1998)
- [6] Aldous, D., Pittel, B.: On a random graph with immigrating vertices: emergence of the giant component. *Random Struct. Algor.* **17**(2), 79–102 (2000)
- [7] Barabási, A.L.: *Network Science*, 1st edn. Cambridge University Press, Cambridge (2016)
- [8] Bender, E.A., Canfield, E.R.: The asymptotic number of labeled graphs with given degree sequences. *J. Combin. Theory Ser. A* **24**(3), 296–307 (1978)
- [9] Bertoin, J.: Eternal additive coalescents and certain bridges with exchangeable increments. *Ann. Probab.* **29**(1), 344–360 (2001)
- [10] Bhamidi, S., Dhara, S., van der Hofstad, R.: Birth of the tiny giant for percolation on scale-free random graphs (in preparation) (2021)
- [11] Bhamidi, S., van der Hofstad, R., van Leeuwen, J.S.H.: Scaling limits for critical inhomogeneous random graphs with finite third moments. *Electron. J. Probab.* **15**(6), 1682–1702 (2010)
- [12] Bhamidi, S., van der Hofstad, R., van Leeuwen, J.S.H.: Novel scaling limits for critical inhomogeneous random graphs. *Ann. Probab.* **40**(6), 2299–2361 (2012)
- [13] Bhamidi, S., van der Hofstad, R., Sen, S.: The multiplicative coalescent, inhomogeneous continuum random trees, and new universality classes for critical random graphs. *Probab. Theory Relat. Fields* **170**(1), 387–474 (2018)
- [14] Billingsley, P.: *Convergence of Probability Measures*. Wiley, London (1999)
- [15] Bingham, N.H., Goldie, C.M., Teugels, J.L.: *Regular Variation*. No. 1 in *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge (1989)
- [16] Bollobás, B.: A probabilistic proof of an asymptotic formula for the number of labelled regular graphs. *Eur. J. Combin.* **1**(4), 311–316 (1980)
- [17] Bollobás, B., Riordan, O.: Robustness and vulnerability of scale-free random graphs. *Internet Math.* **1**(1), 1–35 (2003)
- [18] Borgs, C., Chayes, J.T., van der Hofstad, R., Slade, G., Spencer, J.: Random subgraphs of finite graphs: I. The scaling window under the triangle condition. *Random Struct. Algor.* **27**(2), 137–184 (2005)
- [19] Braunstein, L.A., Buldyrev, S.V., Cohen, R., Havlin, S., Stanley, H.E.: Optimal paths in disordered complex networks. *Phys. Rev. Lett.* **91**(16), 168701 (2003)
- [20] Braunstein, L.A., Wu, Z., Chen, Y., Buldyrev, S.V., Kalisky, T., Sreenivasan, S., Cohen, R., López, E., Havlin, S., Stanley, H.E.: Optimal path and minimal spanning trees in random weighted networks. *Int. J. Bifurc. Chaos* **17**(07), 2215–2255 (2007)
- [21] Breiman, L.: *Probability*. *Classics in Applied Mathematics*. SIAM: Society for Industrial and Applied Mathematics, New York (1968)
- [22] Brown, T.C.: Compensators and Cox convergence. *Math. Proc. Camb. Philos. Soc.* **90**(02), 305 (1981)
- [23] Callaway, D.S., Newman, M.E.J., Strogatz, S.H., Watts, D.J.: Network robustness and fragility: percolation on random graphs. *Phys. Rev. Lett.* **85**(25), 5468–5471 (2000)
- [24] Cohen, R., Ben-Avraham, D., Havlin, S.: Percolation critical exponents in scale-free networks. *Phys. Rev. E* **66**(3), 36113 (2002)

- [25] Cohen, R., Erez, K., Ben-Avraham, D., Havlin, S.: Resilience of the internet to random breakdowns. *Phys. Rev. Lett.* **85**(21), 4626–4628 (2000)
- [26] Dhara, S.: Critical percolation on random networks with prescribed degrees. Ph.D. Thesis, Technische Universiteit Eindhoven, [arXiv:1809.03634](https://arxiv.org/abs/1809.03634) (2018)
- [27] Dhara, S., van der Hofstad, R., van Leeuwaarden, J.S.H., Sen, S.: Heavy-tailed configuration models at criticality. To appear *Ann. Inst. H. Poincaré (B) Probab. Statist.* [arXiv:1612.00650](https://arxiv.org/abs/1612.00650) (2016)
- [28] Dhara, S., van der Hofstad, R., van Leeuwaarden, J.S.H., Sen, S.: Critical window for the configuration model: finite third moment degrees. *Electron. J. Probab.* **22**(16), 1–33 (2017)
- [29] Dorogovtsev, S.N., Goltsev, A.V., Mendes, J.F.F.: Critical phenomena in complex networks. *Rev. Mod. Phys.* **80**(4), 1275–1335 (2008)
- [30] Durrett, R.: *Random Graph Dynamics*. Cambridge University Press, Cambridge (2010)
- [31] Feller, W.: *An Introduction to Probability Theory and Its Applications: Volume 2*. Wiley, London (1991)
- [32] Fountoulakis, N.: Percolation on sparse random graphs with given degree sequence. *Internet Math.* **4**(1), 329–356 (2007)
- [33] Grimmett, G.R., Stirzaker, D.R.: *Probability and Random Processes*. Oxford University Press, Oxford (2001)
- [34] Havlin, S., Braunstein, L.A., Buldyrev, S.V., Cohen, R., Kalisky, T., Sreenivasan, S., Eugene Stanley, H.: Optimal path in random networks with disorder: a mini review. *Physica A* **346**(1–2), 82–92 (2005)
- [35] Heydenreich, M., van der Hofstad, R.: *Progress in High-Dimensional Percolation and Random Graphs*. Springer, London (2017)
- [36] van der Hofstad, R.: *Random Graphs and Complex Networks*, vol. I. Cambridge University Press, Cambridge (2017)
- [37] van der Hofstad, R.: *Stochastic Processes on Random Graphs*. Lecture notes for the 47th Summer School in Probability Saint-Flour 2017 (2017)
- [38] van der Hofstad, R., Janson, S., Łuczak, M.: Component structure of the configuration model: barely supercritical case. *Random Struct. Algor.* **55**(1), 3–55 (2019)
- [39] Janson, S.: On percolation in random graphs with given vertex degrees. *Electron. J. Probab.* **14**, 87–118 (2009)
- [40] Janson, S.: Susceptibility of random graphs with given vertex degrees. *J. Combin.* **1**(3–4), 357–387 (2010)
- [41] Janson, S., Łuczak, M.J.: A new approach to the giant component problem. *Random Struct. Algor.* **34**(2), 197–216 (2009)
- [42] Janson, S., Łuczak, T., Rucinski, A.: *Random Graphs*. Wiley, New York (2000)
- [43] Janson, S., Warnke, L.: On the critical probability in percolation. *Electron. J. Probab.* **23**, 25 (2018)
- [44] Joseph, A.: The component sizes of a critical random graph with given degree sequence. *Ann. Appl. Probab.* **24**(6), 2560–2594 (2014)
- [45] Lipster, R.S., Shiryaev, A.N.: *Theory of Martingales*. Springer, Dordrecht (1989)
- [46] Molloy, M., Reed, B.: A critical-point for random graphs with a given degree sequence. *Random Struct. Algor.* **6**(2–3), 161–179 (1995)
- [47] Nachmias, A., Peres, Y.: Critical random graphs: Diameter and mixing time. *Ann. Probab.* **36**(4), 1267–1286 (2008)
- [48] Nachmias, A., Peres, Y.: Critical percolation on random regular graphs. *Random Struct. Algor.* **36**(2), 111–148 (2010)
- [49] Nachmias, A., Peres, Y.: The critical random graph, with martingales. *Isr. J. Math.* **176**(1), 29–41 (2010)
- [50] Newman, M.E.J.: *Networks: An Introduction*. Oxford University Press, Oxford (2010)
- [51] Norros, I., Reittu, H.: On a conditionally Poissonian graph process. *Adv. Appl. Probab.* **38**(1), 59–75 (2006)
- [52] Pang, G., Talreja, R., Whitt, W.: Martingale proofs of many-server heavy-traffic limits for Markovian queues. *Probab. Surv.* **4**, 193–267 (2007)
- [53] Riordan, O.: The phase transition in the configuration model. *Comb. Probab. Comp.* **21**, 265–299 (2012)
- [54] Rogers, L.C.G., Williams, D.: *Diffusions, Markov processes, and Martingales*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics., vol. 1, 2nd edn. Wiley, Chichester (1994)
- [55] Shorack, G.R., Wellner, J.A.: *Empirical Processes with Applications to Statistics*. Wiley, London (1986)
- [56] Whitt, W.: *Stochastic-Process Limits: An Introduction to Stochastic-Process Limits and Their Application to Queues*. Springer, New York (2002)