The Annals of Probability
2012, Vol. 40, No. 6, 2299–2361
DOI: 10.1214/11-AOP680

NOVEL SCALING LIMITS FOR CRITICAL INHOMOGENEOUS RANDOM GRAPHS

BY SHANKAR BHAMIDI[1], REMCO VAN DER HOFSTAD[2] AND JOHAN S. H. VAN LEEUWAARDEN[2]

University of North Carolina, Eindhoven University of Technology and Eindhoven University of Technology

We find scaling limits for the sizes of the largest components at criticality for rank-1 inhomogeneous random graphs with power-law degrees with power-law exponent τ. We investigate the case where $\tau \in (3, 4)$, so that the degrees have finite variance but infinite third moment. The sizes of the largest clusters, rescaled by $n^{-(\tau-2)/(\tau-1)}$, converge to hitting times of a "thinned" Lévy process, a special case of the general multiplicative coalescents studied by Aldous [*Ann. Probab.* **25** (1997) 812–854] and Aldous and Limic [*Electron. J. Probab.* **3** (1998) 1–59].

Our results should be contrasted to the case $\tau > 4$, so that the third moment is finite. There, instead, the sizes of the components rescaled by $n^{-2/3}$ converge to the excursion lengths of an inhomogeneous Brownian motion, as proved in Aldous [*Ann. Probab.* **25** (1997) 812–854] for the Erdős–Rényi random graph and extended to the present setting in Bhamidi, van der Hofstad and van Leeuwaarden [*Electron. J. Probab.* **15** (2010) 1682–1703] and Turova [(2009) Preprint].

1. Introduction. The critical behavior of random graphs has received tremendous attention in the past decades. The simplest example of a random graph is the Erdős–Rényi random graph, whose critical regime has been intensely explored (see, e.g., [2, 5, 10, 19, 26] and the references therein). In the past few years, many examples of real-world networks have been found where the degrees are highly variable and heavy tailed, unlike the degrees in the Erdős–Rényi random graph, which instead are extremely light tailed. As a result, there has been a concerted effort to define and analyze models for such real-world networks. See, for example, [1, 18, 30] for major reviews of real-world networks and models for them.

In this paper, we study how *inhomogeneity* in the random graph model changes the critical regime of the random graph. In our model, the vertices have a *weight* associated to them, and the weight of a vertex moderates its degree. Therefore, by choosing these weights appropriately, we can generate random graphs with highly

Received October 2010; revised May 2011.

[1]Supported in part by PIMS and NSERC, Canada and NSF Grant DMS-11-05581.

[2]Supported in part by the Netherlands Organisation for Scientific Research (NWO).

MSC2010 subject classifications. 60C05, 05C80, 90B15.

Key words and phrases. Critical random graphs, phase transitions, inhomogeneous networks, thinned Lévy processes, multiplicative coalescent.

variable degrees. For our class of random graphs, it is shown in [34], Theorem 1.1, that when the weights do not vary too much, the critical behavior is similar to the one in the Erdős–Rényi random graph. See in particular the recent works [8, 33], where it was shown that if the degrees have finite *third* moment, then the scaling limit for the largest critical components in the critical window are essentially the same as for the Erdős–Rényi random graph, as identified by Aldous in [2].

Interestingly, in [34], Theorem 1.2, it was shown that when the degrees have *infinite* third moment, then the sizes of the largest critical clusters are quite *different*. See also [22] for a related result for the configuration model, another random graph model having flexible degrees. In this paper, we identify the *scaling limits* of the largest critical clusters in the critical window in the regime where the degrees have infinite third moments. As we shall see, this scaling limit is rather different compared to that for the Erdős–Rényi random graph. Let us first introduce the model that shall be the focus of our investigations for the rest of this article.

1.1. *Model.* In our random graph model, vertices have *weights*, and the edges are independent with the edge probability being approximately equal to the rescaled product of the weights of the two end vertices of the edge. While there are many different versions of such random graphs (see below), it will be convenient for us to work with the so-called *Poissonian random graph* or Norros–Reittu model [31]. To define the model, we consider the vertex set $[n] := \{1, 2, \dots, n\}$ and suppose each vertex is assigned a *weight*, vertex i having weight w_i. Now, attach an edge with probability p_{ij} between vertices i and j, where

$$p_{ij} = 1 - \exp(-w_i w_j/\ell_n) \tag{1.1}$$

with ℓ_n denoting the *total weight*

$$\ell_n = \sum_{i\in[n]} w_i. \tag{1.2}$$

Different edges are independent. In this model, the average degree of vertex i is close to w_i, which brings *inhomogeneity* into the model.

There are many adaptations of this model, for which equivalent results hold. Indeed, the model considered here is a special case of the so-called *rank-1 inhomogeneous random graph* introduced in great generality by Bollobás, Janson and Riordan [11]. It is asymptotically equivalent with many related models, such as the *random graph with given prescribed degrees* or the Chung–Lu model, where instead

$$p_{ij} = \max(w_i w_j/\ell_n, 1), \tag{1.3}$$

and which has been studied intensively by Chung and Lu (see [13–17]). A further adaptation is the *generalized random graph* introduced by Britton, Deijfen and Martin-Löf in [12], for which

$$p_{ij} = \frac{w_i w_j}{\ell_n + w_i w_j}. \tag{1.4}$$

See Janson [25] for conditions under which these random graphs are *asymptotically equivalent*, meaning that all events have asymptotically equal probabilities. As discussed in more detail in [34], Section 1.3, these conditions apply in the setting to be studied in this paper. Therefore, all results proved here also hold for these related rank-1 models.

Let us now specify how the weights are chosen. We let the weight sequence $\mathbf{w} = (w_i)_{i\in[n]}$ be defined by

$$w_i = [1 - F]^{-1}(i/n), \tag{1.5}$$

where F is a distribution function on $[0, \infty)$ for which we assume that there exists a $\tau \in (3, 4)$ and $0 < c_F < \infty$ such that

$$\lim_{x\to\infty} x^{\tau-1}[1 - F(x)] = c_F, \tag{1.6}$$

and where $[1 - F]^{-1}$ is the generalized inverse function of $1 - F$ defined, for $u \in (0, 1)$, by

$$[1 - F]^{-1}(u) = \inf\{s : [1 - F](s) \leq u\}. \tag{1.7}$$

By convention, we set $[1-F]^{-1}(1) = 0$. We often make use of the fact that, with U uniform on $[0, 1]$, the random variable $[1 - F]^{-1}(U)$ has distribution function F.

An interpretation of the choice in (1.5) is that the weight of a vertex V_n chosen uniformly in $[n]$ has distribution function F_n given by

$$\begin{aligned} F_n(x) &= \mathbb{P}(w_{V_n} \leq x) = \frac{1}{n}\sum_{j\in[n]} \mathbb{1}_{\{w_j\leq x\}} \\ &= \frac{1}{n}\sum_{j\in[n]} \mathbb{1}_{\{[1-F]^{-1}(j/n)\leq x\}} = \frac{1}{n}\sum_{i=0}^{n-1} \mathbb{1}_{\{[1-F]^{-1}(1-i/n)\leq x\}} \\ &= \frac{1}{n}\sum_{i=0}^{n-1} \mathbb{1}_{\{F^{-1}(i/n)\leq x\}} = \frac{1}{n}\sum_{i=0}^{n-1} \mathbb{1}_{\{i/n\leq F(x)\}} \\ &= \frac{1}{n}\big(\lfloor nF(x)\rfloor + 1\big) \wedge 1, \end{aligned} \tag{1.8}$$

where, throughout this paper and for $x, y \in \mathbb{R}$, we write $(x \vee y) = \max(x, y)$ and $(x \wedge y) = \min(x, y)$. By (1.8), $F_n \to F$ uniformly. As a result, a uniformly chosen vertex has a weight which is close in distributional sense to F.

For the setting in (1.1) and (1.5), by [11], Theorem 3.13, the number of vertices with degree k, denoted by N_k, satisfies

$$N_k/n \xrightarrow{\mathbb{P}} \mathbb{E}\bigg[\mathrm{e}^{-W}\frac{W^k}{k!}\bigg], \qquad k \geq 0, \tag{1.9}$$

where $\xrightarrow{\mathbb{P}}$ denotes convergence in probability, and where W has distribution function F appearing in (1.5). We recognize the limiting distribution as a *mixed Poisson distribution with mixing distribution* F; that is, conditionally on $W = w$, the distribution is Poisson with mean w. Equation (1.9) also implies that the distribution of the degree of a uniformly chosen vertex in $[n]$ converges to a mixed Poisson distribution with mixing distribution F. This can be understood by noting that the weight of a uniformly chosen vertex is, by (1.8), close in distribution to F. In turn, when a vertex has weight w, then, by (1.1), its degree is close to Poisson with parameter w. Since a Poisson random variable with large parameter w is closely concentrated around its mean w, we see that the tail behavior of the degrees in our random graph is close to that of the distribution F. As a result, when (1.6) holds, and with D_n the degree of a uniformly chosen vertex in $[n]$, $\limsup_{n\to\infty} \mathbb{E}[D_n^a] < \infty$ when $a < \tau - 1$ and $\limsup_{n\to\infty} \mathbb{E}[D_n^a] = \infty$ when $a \geq \tau - 1$. In particular, the degree of a uniformly chosen vertex in $[n]$ has finite second, but infinite third moment when (1.6) holds with $\tau \in (3, 4)$.

We shall frequently make use of the fact that (1.6) implies that, as $u \downarrow 0$,

$$[1 - F]^{-1}(u) = (c_F/u)^{1/(\tau-1)}\bigl(1 + o(1)\bigr). \tag{1.10}$$

Under the key assumption in (1.6), we have that the third moment of the degrees tends to infinity; that is, with $W \sim F$, $\mathbb{E}[W^3] = \infty$. Define

$$\nu = \mathbb{E}[W^2]/\mathbb{E}[W], \tag{1.11}$$

so that, again by (1.6), $\nu < \infty$. Then, by [11], Theorem 3.1 (see also [11], Section 16.4, for a detailed discussion on rank-1 inhomogeneous random graphs, of which our random graph is an example), when $\nu > 1$, there is one giant component of size proportional to n, while all other components are of smaller size $o(n)$, and when $\nu \leq 1$, the largest connected component contains a proportion of vertices that converges to zero in probability. Thus, the critical value of the model is $\nu = 1$. The main aim of this paper is to investigate what happens *close to* the critical point, that is, when $\nu = 1$.

A simple example of our model arises when we take

$$F(x) = \begin{cases} 0, & \text{for } x < a, \\ 1 - (a/x)^{\tau-1}, & \text{for } x \geq a, \end{cases} \tag{1.12}$$

in which case $[1 - F]^{-1}(u) = a(1/u)^{1/(\tau-1)}$, so that $w_j = a(n/j)^{1/(\tau-1)}$ and

$$\mathbb{E}[W] = \frac{a(\tau-1)}{\tau-2}, \qquad \mathbb{E}[W^2] = \frac{a^2(\tau-1)}{\tau-3}. \tag{1.13}$$

The critical case thus arises when

$$\nu = \mathbb{E}[W^2]/\mathbb{E}[W] = \frac{a(\tau-2)}{\tau-3} = 1, \tag{1.14}$$

that is, when $a = (\tau-3)/(\tau-2)$.

With the definition of the weights in (1.5), we shall write $\mathcal{G}_n^0(\mathbf{w})$ for the graph constructed with the probabilities in (1.1), while, for any fixed $\lambda \in \mathbb{R}$, we shall write $\mathcal{G}_n^\lambda(\mathbf{w})$ when we use the weight sequence $\mathbf{w}(\lambda) = (w_i(\lambda))_{i\in[n]}$ defined by

$$\mathbf{w}(\lambda) = \big(1 + \lambda n^{-(\tau-3)/(\tau-1)}\big)\mathbf{w}. \tag{1.15}$$

We shall assume that n is so large that $1 + \lambda n^{-(\tau-3)/(\tau-1)} \geq 0$, so that $w_i(\lambda) \geq 0$ for all $i \in [n]$. This setting was first explored in [34], where, for the largest connected component $\mathcal{C}_{\max}$ and all $\lambda \in \mathbb{R}$, it is proved that both $n^{-(\tau-2)/(\tau-1)}|\mathcal{C}_{\max}|$ and $n^{(\tau-2)/(\tau-1)}/|\mathcal{C}_{\max}|$ are *tight* sequences of random variables. In this paper, we bring the discussion of the critical behavior of such inhomogeneous random graphs substantially further, by identifying the scaling limit of $(n^{-(\tau-2)/(\tau-1)}|\mathcal{C}_{(i)}|)_{i\geq 1}$, where $(\mathcal{C}_{(i)})_{i\geq 1}$ denote the connected components ordered in size, that is, $|\mathcal{C}_{\max}| = |\mathcal{C}_{(1)}| \geq |\mathcal{C}_{(2)}| \geq \cdots$.

Interestingly, as proved in [8, 33, 34], when $\tau > 4$, so that $\mathbb{E}[W^3] < \infty$, the scaling limits of the random graphs studied here are (apart from a trivial scaling constant) *equal* to the scaling limit of the ordered connected components in the Erdős–Rényi random graph, as first identified by Aldous in [2]. This suggests that the *high-weight vertices* play a crucial role in our setting, a fact that shall feature extensively throughout our proof. The importance of the high-weight vertices also partly explains why we restrict our setting to (1.5) and (1.6), which give us sharp asymptotics of the weights of the high-weight vertices in the heavy-tailed setting we study here. We shall comment on extensions of our results in more detail in Section 1.5 below.

Before stating our main results, we introduce some notation. For a vertex $i \in [n]$, we write $\mathcal{C}(i)$ for the vertices in the connected component or *cluster* of i. Further, let

$$\mathcal{C}_{\leq}(i) = \begin{cases} \mathcal{C}(i), & \text{if } i \leq j\ \forall j \in \mathcal{C}(i), \\ \varnothing, & \text{otherwise}. \end{cases} \tag{1.16}$$

Then, clearly, $|\mathcal{C}_{\max}| = \max_{i\in[n]} |\mathcal{C}(i)| = \max_{i\in[n]} |\mathcal{C}_{\leq}(i)|$, and $(|\mathcal{C}_{(i)}|)_{i\geq 1}$ is equal to the sequence $(|\mathcal{C}_{\leq}(i)|)_{i\geq 1}$ ordered in size. We further define the *cluster weight* of vertex i to be

$$\mathcal{W}(i) = \sum_{j\in\mathcal{C}(i)} w_j, \tag{1.17}$$

and let $\mathcal{W}_{\leq}(i)$ be as in (1.17), where the sum is restricted to $\mathcal{C}_{\leq}(i)$. We again let $(\mathcal{W}_{(i)})_{i\geq 1}$ be equal to the sequence $(\mathcal{W}_{\leq}(i))_{i\geq 1}$ ordered in size.

Throughout this paper, we shall make use of the following standard notation. We let $\xrightarrow{d}$ denote convergence in distribution, and $\xrightarrow{\mathbb{P}}$ convergence in probability. For a sequence of random variables $(X_n)_{n\geq 1}$, we write $X_n = O_{\mathbb{P}}(b_n)$ when $|X_n|/b_n$ is a tight sequence of random variables as $n \to \infty$, and $X_n = o_{\mathbb{P}}(b_n)$ when $|X_n|/b_n \xrightarrow{\mathbb{P}} 0$ as $n \to \infty$. For a nonnegative function $n \mapsto g(n)$, we write

$f(n) = O(g(n))$ when $|f(n)|/g(n)$ is uniformly bounded, and $f(n) = o(g(n))$ when $\lim_{n\to\infty} f(n)/g(n) = 0$. Furthermore, we write $f(n) = \Theta(g(n))$ if $f(n) = O(g(n))$ and $g(n) = O(f(n))$. Finally, we write that a sequence of events $(E_n)_{n\geq 1}$ occurs *with high probability* (*whp*) when $\mathbb{P}(E_n) \to 1$.

Now we are ready to state our main results. We start in Section 1.2 by describing the scaling limit of the ordered clusters, and in Section 1.3 we discuss further properties of the scaling limit.

1.2. *The scaling limit for* $\tau \in (3, 4)$. In this section, we investigate the scaling limit of the connected components ordered in size. Our first main result is as follows:

THEOREM 1.1 [Weak convergence of the ordered critical clusters for $\tau \in (3, 4)$]. *Fix the Norros–Reittu random graph with weights* $\mathbf{w}(\lambda)$ *defined in* (1.5) *and* (1.15). *Assume that* $\nu = 1$ *and that* (1.6) *holds. Then, for all* $\lambda \in \mathbb{R}$,

$$(1.18)\qquad \big(n^{-(\tau-2)/(\tau-1)}\big|\mathcal{C}_{(i)}\big|\big)_{i\geq 1} \xrightarrow{d} (\gamma_i(\lambda))_{i\geq 1}$$

in the product topology, for some nondegenerate limit $(\gamma_i(\lambda))_{i\geq 1}$.

We next study the *joint* convergence of the clusters for different values of $\lambda \in \mathbb{R}$. By increasing λ, more and more edges are added to the system. These extra edges potentially create connections between *disjoint* clusters, thus merging them. As a result, we can interpret λ as a *time variable*, and as time increases, clusters are being merged. This resembles a *coalescence process*, as studied in [7]. We now make this connection precise. Before being able to do so, we introduce some necessary notation.

We first give a quick overview of Aldous's standard multiplicative coalescent and how it relates to the limiting random variables in Theorem 1.1, seen as functions of the parameter λ. It will not be possible to give a full description of the process and its many fascinating properties here, and we refer the interested reader to the paper [3], the survey paper [4] and the book [7].

Write $\ell^2_{\searrow}$ for the metric space of infinite real-valued sequences $\mathbf{x} = (x_1, x_2, \ldots)$ with $x_1 \geq x_2 \geq \cdots \geq 0$ and $\sum_{i=1}^{\infty} x_i^2 < \infty$, with the ℓ^2-norm as the metric. The standard multiplicative coalescent is described as the Markov process with states in $\ell^2_{\searrow}$ whose dynamics is as follows: for each pair of clusters (x, y), the pair merges at rate xy. Thus, the multiplicative coalescent is a continuous-time Markov process of the masses of an infinite number of particles, where two particles merge at a rate equal to the *product* of their masses.

In [2], Aldous showed that there is a Feller process on the space $\ell^2_{\searrow}$ defined for all times $-\infty < t < \infty$ starting from infinitesimally small masses at time $-\infty$, and following the above merging dynamics. The distribution of the coalescent process at any time t is the same as the limiting ordered cluster sizes of an Erdős–Rényi random graph with edge probabilities $p_n = (1 + tn^{-1/3})/n$.

Aldous and Limic [3] explicitly characterize the entrance boundary at $-\infty$ of the above Markov process, in the sense that they prove that every *extreme* version of the above Markov process is characterized by a diffusion parameter κ, a translation parameter β, and a vector of "limiting largest weights" $\mathbf{c} = (c_1, c_2, \dots)$ that describe the asymptotic decay of the masses of the particles at time $-\infty$. In this terminology, the multiplicative coalescent can be described as the ordered lengths of excursions beyond past minima of the process

$$W^{\kappa,\beta,\mathbf{c}}(s) = \kappa^{1/2} W(s) + \beta s - \tfrac{1}{2}\kappa s^2 + V^{\mathbf{c}}(s), \tag{1.19}$$

where $(W(s))_{s\geq 0}$ is a standard Brownian motion, while

$$V^{\mathbf{c}}(s) = \sum_{j=1}^{\infty} c_j \big(\mathbb{1}_{\{E_j \leq s\}} - c_j s\big) \tag{1.20}$$

with $(E_j)_{j\geq 1}$ independent exponential random variables, E_j having mean $1/c_j$. Then, the $(\kappa, \beta, \mathbf{c})$-multiplicative coalescent is the set of ordered lengths of excursions from zero of the reflected process

$$B^{\kappa,\beta,\mathbf{c}}(s) = W^{\kappa,\beta,\mathbf{c}}(s) - \min_{0\leq s' \leq s} W^{\kappa,\beta,\mathbf{c}}(s'). \tag{1.21}$$

Part of the proof in [3] is the fact that these ordered excursions can be defined properly.

The following theorem draws a connection between the components of the graph for a fixed λ and the sizes of clusters at the same time in a multiplicative coalescent with a particular entrance boundary, scale and translation parameter. For this, define the sequence

$$\mathbf{c} = \big(c i^{-1/(\tau-1)}\big)_{i\geq 1} \qquad \text{with } c = c_F^{1/(\tau-1)}. \tag{1.22}$$

Then, we have the following theorem:

THEOREM 1.2 (Relation to multiplicative coalescents). *Assume that the conditions in Theorem 1.1 hold. Consider the sequence-valued random variables $\mathbf{X}^*(\lambda) = (\gamma_1(\mathbb{E}[W]\lambda), \gamma_2(\mathbb{E}[W]\lambda), \dots)$ with $(\gamma_i(\lambda))_{i\geq 1}$ as in Theorem 1.1. Then $\mathbf{X}^*(\lambda)$ has the same distribution as a multiplicative coalescent at time λ with entrance boundary $\mathbf{c}/\mathbb{E}[W]$, diffusion constant $\kappa = 0$ and centering constant $\beta = -\zeta/\mathbb{E}[W]$, where ζ is identified explicitly in (2.18) below. More precisely, there exists a simultaneous coupling of the clusters $(|\mathcal{C}_{(i)}(\lambda)|)_{i\geq 1}$, where $|\mathcal{C}_{(i)}(\lambda)|$ is the ith largest cluster when the weights are equal to $\mathbf{w}(\lambda)$, such that, for every vector $(\lambda_1, \lambda_2, \dots, \lambda_k)$,*

$$\big(n^{-(\tau-2)/(\tau-1)}\big(|\mathcal{C}_{(i)}(\lambda_l)|\big)_{i\geq 1}\big)_{l=1}^k \xrightarrow{d} \big(\mathbf{X}^*(\lambda_l/\mathbb{E}[W])\big)_{l=1}^k. \tag{1.23}$$

In particular, with c_j defined as in (1.22),

$$|\lambda|\gamma_j(\lambda) \xrightarrow{\mathbb{P}} c_j \qquad \text{as } \lambda \to -\infty \text{ for each } j \geq 1. \tag{1.24}$$

Theorem 1.2 proves that the finite-dimensional distributions of the rescaled cluster sizes converge to those of a multiplicative coalescent. While we believe that also *process* convergence holds, viewing the processes as elements of an appropriate function space, we have no proof for this fact. See Section 7 for a full proof of Theorem 1.2. The setting in this paper is the first example where the multiplicative coalescent with $\kappa = 0$ arises in random graph theory. Indeed, all random graph examples in [3] have largest component sizes of the order $n^{2/3}$, like for the Erdős–Rényi random graph studied in [2]. Our example links the multiplicative coalescent also to random graphs with the largest critical connected components of the order $n^{(\tau-2)/(\tau-1)}$ instead of $n^{2/3}$.

A crucial part of the proof of Theorem 1.2 is the analysis of the *subcritical* phase of our model. The asymptotics of the rescaled ordered cluster sizes in the subcritical regime acts as the *entrance boundary* of the multiplicative coalescent, as explained in more detail in [3], Proposition 7. This entrance boundary is identified in the following theorem, which is of independent interest. In the statement of Theorem 1.3, the lower bound on λ_n appears only to ensure that $w_i(\lambda_n) = (1 + \lambda_n n^{-(\tau-3)/(\tau-1)})w_i \geq 0$ for every $i \in [n]$.

THEOREM 1.3 (Subcritical phase). *Assume that the conditions in Theorem 1.1 hold, but now take $\lambda = \lambda_n \to -\infty$ as $n \to \infty$ such that $\lambda_n \geq -n^{-(\tau-3)/(\tau-1)}$. Then, for each $j \in \mathbb{N}$, with c_j defined as in (1.22),*

$$|\lambda_n| n^{-(\tau-2)/(\tau-1)} |\mathcal{C}_{(j)}| \xrightarrow{\mathbb{P}} c_j. \tag{1.25}$$

Theorem 1.3 is proved in Section 6. Interestingly, the limit in (1.25) is *deterministic* [recall also (1.22)]. The rough idea for this is as follows. As $\lambda = \lambda_n \to -\infty$, the random graph becomes more and more subcritical. Now, if we look at $\mathcal{C}(j)$, the cluster of vertex j, then we can view it as the union of approximately w_j (which is roughly the degree of vertex j) almost *independent* clusters. These clusters are close to total progenies of branching processes having mean offspring $\nu_n(\lambda_n) \approx 1 + \lambda_n n^{-(\tau-3)/(\tau-1)}$. The expected total progeny of a branching process with mean offspring ν equals $1/(1-\nu)$. As a result, the expected cluster size of vertex j is close to

$$\frac{w_j}{1 - \nu_n(\lambda_n)} \approx \frac{w_j}{|\lambda_n| n^{-(\tau-3)/(\tau-1)}} = \frac{n^{(\tau-2)/(\tau-1)}}{|\lambda_n|} c_j \bigl(1 + o(1)\bigr). \tag{1.26}$$

In our setting, $c_j = c_F^{1/(\tau-1)} j^{-1/(\tau-1)}$, so that $j \mapsto c_j$ is strictly decreasing. Thus, we must also have that $|\mathcal{C}(j)| = |\mathcal{C}_{(j)}|$ *whp*. The proof of Theorem 1.3 makes this argument precise, by investigating the deviation from a branching process, a technique that is also crucially used in [34] to study tightness of the sequence of random variables $|\mathcal{C}_{\max}| n^{-(\tau-2)/(\tau-1)}$. A result similar to Theorem 1.3 is proved for the near-critical phase of the configuration model in [36], but the proof we give here is entirely different.

We also obtain that the ordered cluster *weights* as defined in (1.17) satisfy the same scaling results as described above.

THEOREM 1.4 (Scaling limit of cluster weights). *Theorems 1.1, 1.2 and 1.3 also hold for the ordered cluster weights $(n^{-(\tau-2)/(\tau-1)}\mathcal{W}_{(i)})_{i\geq 1}$, with identical scaling limits as in Theorems 1.1, 1.2 and 1.3.*

As explained in more detail in Section 2.1 below, Theorem 1.4 can be heuristically understood by noting that the average weight of a vertex in a cluster is close to $\nu = 1$, and therefore it contributes the same to the weight of the cluster as it does to the cluster size. In fact, the proof will show that $n^{-(\tau-2)/(\tau-1)}\mathcal{W}_{(i)}$ and $n^{-(\tau-2)/(\tau-1)}|\mathcal{C}_{(i)}|$ converge to the *same* limit. The proof of Theorem 1.4 shall be given simultaneously with the proofs of Theorems 1.1, 1.2 and 1.3, respectively, adapted so as to deal with cluster weights or cluster sizes. Sometimes, it is more convenient to study cluster sizes (e.g., since cluster explorations can more naturally be formulated in terms of the number of vertices than their weight), in some cases it is more convenient to work with cluster weights (e.g., since the cluster weights can be described in terms of multiplicative coalescents, a fact that is crucial in the proof of Theorem 1.2).

1.3. *Properties of large critical clusters.* We shall also derive some related interesting properties of the limiting largest clusters. In the following theorem, we consider the connectivity structure of the high-weight vertices:

THEOREM 1.5 (Connectivity of high-weight vertices). *Under the assumptions in Theorem 1.1, for every $i, j \geq 1$ fixed,*

$$\lim_{n\to\infty} \mathbb{P}\big(j \in \mathcal{C}(i)\big) = q_{ij}(\lambda) \in (0,1) \tag{1.27}$$

and

$$\lim_{n\to\infty} \mathbb{P}(i \in \mathcal{C}_{\max}) = q_i(\lambda) \in (0,1). \tag{1.28}$$

Theorem 1.5 states that the high-weight vertices play an essential role in the critical regime. Indeed, we shall see that in the subcritical regime, with high probability, $\mathcal{C}_{\max} = \mathcal{C}(1)$, so that $\mathbb{P}(1 \in \mathcal{C}_{\max}) = 1 - o(1)$, while $\mathbb{P}(i \in \mathcal{C}_{\max}) = o(1)$ for $i > 1$. In the supercritical regime, instead, $\mathbb{P}(i \in \mathcal{C}_{\max}) = 1 - o(1)$ for every $i \geq 1$ fixed. Thus, the critical regime is precisely the regime where the high-weight vertices start to form connections. Informally, this can be phrased as "power to the wealthy." Theorem 1.5 should be contrasted with the situation when $\mathbb{E}[W^3] < \infty$ studied in [8, 33], where the probability that any specific vertex is an element of $\mathcal{C}_{\max}$ is negligible, and, instead, the largest cluster is born out of many trials each having a small probability. This can be informally phrased as "power to the masses."

The following theorem, which is a crucial ingredient in the proof of Theorem 1.1, essentially says that, for each fixed λ, the maximal size components are those attached to the largest weight vertices:

THEOREM 1.6 (Large clusters contain a high-weight vertex). *Assume that the conditions in Theorem 1.1 hold. Then*:

(a) *for any* $\varepsilon \in (0,1)$, *there exists a* $K = K(\varepsilon) \geq 1$, *such that, for all* n,

$$\mathbb{P}\Big(\max_{i \geq K} |\mathcal{C}_{\leq}(i)| \geq \varepsilon n^{(\tau-2)/(\tau-1)}\Big) \leq \varepsilon; \tag{1.29}$$

(b) *for any* $m \geq 1$,

$$\lim_{\varepsilon \downarrow 0} \liminf_{n \to \infty} \mathbb{P}\big((|\mathcal{C}_{\leq}(i)|)_{i \in [n]} \text{ contains } m \text{ components of size } \geq \varepsilon n^{(\tau-2)/(\tau-1)}\big) = 1. \tag{1.30}$$

1.4. *Overview of the proofs*. In this section, we give an overview of the proofs of our main results. We start by explaining the proof of Theorem 1.1, along the way also explaining the key ideas behind Theorems 1.5 and 1.6. After this, we shall discuss the proofs of Theorem 1.2 and 1.3.

We note that, since $u \mapsto [1-F]^{-1}(u)$ is nonincreasing, $\mathbf{w}$ is ordered in size, that is, $w_1 \geq w_2 \geq w_3 \geq \cdots$. We start by *exploring* the clusters from the largest weight vertices onward. Here, by a cluster exploration, we mean the recursive investigation of the neighbors of the vertices already found to be in the cluster. This cluster exploration shall be described in detail in Section 2.1. The rough idea is as follows. We start with a vertex i, and wish to find all the vertices that are in its cluster. For this, we sequentially take each vertex in the cluster being currently explored and find its direct neighbors, that have not yet been found by the exploration process. Call a vertex *active* when it is found to be in the cluster, but has not yet been explored. A vertex is called *explored* when its neighbors have been investigated and *neutral* when it has not yet appeared in the exploration process. Then, in the exploration process at time t, we take a vertex, turn it from active to explored, and explore it, that is, see which neutral neighbors it has. Turn the status of its neutral neighbors to active. Let Z_l denote the number of active neighbors after the exploration of the lth active vertex. When $Z_l = 0$ for the first time, then there are no more active vertices, so all elements of the cluster have been found. (The description in Section 2.1 is slightly different than the one described here, as it studies *potential* elements of the cluster instead.)

We note that the high-weight vertices have weights of the order $w_j \sim (c_F n/j)^{1/(\tau-1)}$, so, when we start with a high-weight vertex, initially, the number of active vertices shall be of the order $n^{1/(\tau-1)}$. When our exploration process hits *another* high-weight vertex, then the number of active vertices gets a large push of the order $n^{1/(\tau-1)}$ upward. It is these upward pushes that change the number of active vertices in a substantial way, and, therefore, the high-weight vertices play a crucial role in the critical behavior of our random graph. In turn, this suggests that the largest clusters contain at least one high-weight vertex, as indicated by Theorem 1.6. Due to the critical nature of our random graph, it turns out that the *average* number and weight of active vertices being added in each exploration is close to one, so that, due to the removal of the vertex which is being explored, the exploration process has increments that have a mean close to zero.

In Section 2, we start by identifying the scaling limit of $n^{-(\tau-2)/(\tau-1)}|\mathcal{C}_{\leq}(1)| = n^{-(\tau-2)/(\tau-1)}|\mathcal{C}(1)|$. The weak limit of $n^{-(\tau-2)/(\tau-1)}|\mathcal{C}(1)|$ is given in terms of the hitting time of zero of an exploration process exploring the cluster of vertex 1 (the vertex with the highest weight). See Theorems 2.1 and 2.4. The scaling limit of the exploration process of a cluster exists (see Theorem 2.4), and can be viewed as a “thinned” Lévy process. Therefore, the convergence in distribution of $n^{-(\tau-2)/(\tau-1)}|\mathcal{C}(1)|$ in Theorem 2.1 is equivalent to the convergence of the first hitting time of zero of the exploration process to the one of this thinned Lévy process. In proving this, we perform a careful analysis of hitting times of a spectrally positive Lévy process that stochastically dominates the thinned Lévy process.

Following the proof of convergence of $n^{-(\tau-2)/(\tau-1)}|\mathcal{C}(1)|$ in Theorem 2.1, we shall prove the convergence in distribution of $(n^{-(\tau-2)/(\tau-1)}|\mathcal{C}_{\leq}(i)|)_{i\in[n]}$ in Theorem 4.1. This proof makes crucial use of the estimates in the proof of Theorem 2.1, and allows us to extend the result in Theorem 2.1 to the (joint) convergence of several rescaled clusters by an inductive argument. The largest m clusters are given by the largest m elements of the vector $(|\mathcal{C}_{\leq}(i)|)_{i\in[n]}$, so that this completes the proof of Theorem 1.1. The conclusion of this argument shall be carried out in Section 5.

In Section 6, we prove Theorem 1.3 by a second moment argument, using the fact that the *subcritical* phase of our random graph is closely related to (and even stochastically dominated by) a branching process. In Section 7, we use the results proved in Section 6, jointly with the results in [3], to prove Theorem 1.2. We now discuss in a bit more detail how one can understand the appearance of multiplicative coalescents in the random graphs we study here.

We make crucial use of [3], Proposition 7, whose application we now explain. Fix a sequence $\lambda_n \to -\infty$. For each fixed t, consider the construction of the inhomogeneous random graph as in (1.1) but with the weight sequence $\bar{\mathbf{w}}(t) = (\bar{w}_j(t))_{1\leq j\leq n}$ given by

$$\bar{w}_j(t) = w_j\big(1 + (t+\lambda_n)\ell_n n^{-2(\tau-2)/(\tau-1)}\big). \tag{1.31}$$

Let

$$\mathbf{X}^{(n)}(t) = \big(n^{-(\tau-2)/(\tau-1)}\mathcal{W}_{(i)}(t)\big)_{i\geq 1} \tag{1.32}$$

denote the ordered version of cluster weights when the vertex weights are given by $\bar{\mathbf{w}}(t)$.

Note that the above process, when taking $t = -\lambda_n + \lambda/\mathbb{E}[W]$, is closely related to the ordered cluster weights of our random graph with weights $w_j(\lambda) = w_j(1 + \lambda n^{-(\tau-3)/(\tau-1)})$, since $\ell_n = \mathbb{E}[W]n(1+o(1))$. We then note that $\mathbf{X}^{(n)}$ can be constructed so that, viewed as a function in t, it is a multiplicative coalescent.

LEMMA 1.7 (Discrete multiplicative coalescent). *We can construct the process $\mathbf{X}^{(n)} = (\mathbf{X}^{(n)}(t))_{t\geq 0}$ such that, for each fixed t, $\mathbf{X}^{(n)}(t)$ has the distribution of the ordered rescaled weighted component sizes of the random graph with weight sequence given by* (1.31) *and such that, for each fixed n, the process viewed as a process in t is a multiplicative coalescent. The initial state denoted by $\mathbf{x}^{(n)}(0)$ has the same distribution as the ordered cluster weights of a random graph with edge probabilities as in* (1.1) *and weight sequence*

$$\bar{w}_j(0) = w_j\big(1 + \lambda_n \ell_n n^{-2(\tau-2)/(\tau-1)}\big). \tag{1.33}$$

PROOF. For each unordered pair (i, j), let ξ_{ij} be an exponential random variable with rate $w_i w_j/\ell_n$, where $(\xi_{ij})_{(i,j)}$ are independent. For fixed t, define the graph $\bar{\mathcal{G}}_n^t$ to consist of all edges (i, j) for which

$$\xi_{ij} \leq 1 + \frac{(\lambda_n + t)\ell_n}{n^{2(\tau-2)/(\tau-1)}}. \tag{1.34}$$

Then, by construction, for all $t \geq 0$, the rescaled weighted component sizes of $\bar{\mathcal{G}}_n^t(\mathbf{w})$ have the same distribution as $\mathbf{X}^{(n)}(t)$. Further, for any time t we note that two *distinct* clusters $\mathcal{C}_1$ and $\mathcal{C}_2$ having weights $\mathcal{W}_{(i)}(t)$ and $\mathcal{W}_{(j)}(t)$, respectively, coalesce at rate

$$\begin{aligned} &\ell_n n^{-2(\tau-2)/(\tau-1)} \sum_{s_1\in\mathcal{C}_1, s_2\in\mathcal{C}_2} \frac{w_{s_1} w_{s_2}}{\ell_n} \\ &\quad = \big(n^{-(\tau-2)/(\tau-1)}\mathcal{W}_{(i)}(t)\big)\big(n^{-(\tau-2)/(\tau-1)}\mathcal{W}_{(j)}(t)\big) \end{aligned} \tag{1.35}$$

as required. □

In effect, Theorems 1.1 and 1.2 give us two *distinct* proofs of the statement that the ordered cluster weights converge, and we now discuss the advantages of these two different proofs. Theorem 1.1 proves that for any *fixed* λ, $(n^{-(\tau-2)/(\tau-1)}\mathcal{W}_{(i)})_{i\geq 1}$ converges in distribution. Further, by the fact that this vector is obtained by sequentially investigating the clusters of the high-weight vertices, it allows us to prove properties about the high-weight vertices that are part of the largest clusters, as in Theorems 1.5 and 1.6. Finally, it allows us to show that the ordered cluster *sizes* have the same scaling limit as the ordered cluster *weights* (see Theorem 1.4), a feature that is also crucial in the proofs of Theorems 1.2 and 1.3.

Theorem 1.2, instead, shows that the *process* of the ordered cluster sizes or weights converges in distribution. This means that there exists a *stochastic process* that describes the joint convergence of the ordered cluster sizes or weights for different values of λ simultaneously. Due to the fact that the proof of Theorem 1.2 relies on [3], Proposition 7, however, we obtain less information about the vertices that are part of the large critical clusters. The combination of the two proofs provides us with a detailed and full understanding of the scaling limit of the ordered cluster sizes or weights.

1.5. *Discussion.*

Comparison to the case of weights with finite third moments. In [2, 8, 33], the scaling limit was considered when $\mathbb{E}[W^3] < \infty$. In this case, the scaling limit turns out to be (a trivial rescaling of) the scaling limit for the Erdős–Rényi random graph as found by Aldous in [2]. Thus, the setting for $\tau \in (3,4)$ is fundamentally different. When $\mathbb{E}[W^3] < \infty$, the probability that $1 \in \mathcal{C}_{\max}$ is negligible, while in our setting this is not true, as shown in Theorem 1.5.

Other weights. Our proof reveals that the precise limits of $w_i n^{-1/(\tau-1)}$, for fixed $i \geq 1$, arise in the scaling limit. We make crucial use of the fact that, by (1.10) $c_i = \lim_{n\to\infty} w_i n^{-1/(\tau-1)} = (c_F/i)^{1/(\tau-1)}$. However, we believe that our results can be appropriately adapted to the situation that $\lim_{n\to\infty} w_i n^{-1/(\tau-1)}$ exists for every $i \geq 1$ and is *asymptotically* equal to $ai^{-1/(\tau-1)}$ for some $a > 0$. This suggests that, by varying the precise values of high weights, there are many possible scaling limits. It would be of interest to investigate this further.

Also, we restrict to tail distributions $1 - F(x)$ that are, for large $x \geq 0$, asymptotic to an inverse power of x; see (1.6). It would be of interest to investigate the scaling behavior when (1.6) is replaced with the assumption that $1 - F(x)$ is regularly varying with exponent $1 - \tau$, that is, $[1 - F](x) = x^{-(\tau-1)}\ell(x)$ for some $x \mapsto \ell(x)$ which is slowly varying at ∞. In this case, we believe that the asymptotic sizes of the largest critical clusters are given by $\ell^*(n)n^{(\tau-2)/(\tau-1)}$ for some suitable slowly varying function $n \mapsto \ell^*(n)$ that can be described in terms of $x \mapsto \ell(x)$. For more details, see [34], Section 1.3, where also the critical cases $\tau = 3$ and $\tau = 4$ are discussed.

I.i.d. weights. In our analysis, we make crucial use of the choice for w_i in (1.5). In the literature, also the setting where $(W_i)_{i\in[n]}$ are independent and identically distributed (i.i.d.) random variables with distribution function F has been considered. We expect the behavior in this model to be *different*. Indeed, let $w_i = W_{(i)}$, where $W_{(i)}$ are the order statistics of the i.i.d. sequence $(W_i)_{i\in[n]}$. It is well known that

$$n^{-1/(\tau-1)} W_{(i)} \xrightarrow{d} \xi_i \equiv (E'_1 + \cdots + E'_i)^{-1/(\tau-1)}, \tag{1.36}$$

where $(E'_i)_{i=1}^{\infty}$ are i.i.d. exponential random variables with mean 1. In particular, when $\tau \in (3,4)$, $\mathbb{E}[\xi_1^a] < \infty$ whenever $a < \tau - 1$. The extra randomness of the order statistics has an effect on the scaling limit, which is thus *different*. In most cases, the two settings have the *same* behavior (see, e.g., [8], where this is shown to hold for weights for which $\mathbb{E}[W^3] < \infty$, where W has distribution function F). See [27] for the identification of the scaling limit of the largest cluster sizes in the critical configuration model with i.i.d. degrees, which is markedly *different* from ours. We believe that the same applies to the Norros–Reittu model with i.i.d. weights.

High-weight vertices. The fact that the vertex i is in the largest connected component with nonvanishing probability as $n \to \infty$ (see Theorem 1.5) is remarkable. In our setting, a *uniformly* chosen vertex in $[n]$ is an element of $\mathcal{C}_{\max}$ with negligible probability. The point is that vertex i has weight w_i, which, for i fixed, is close to $(c_F/i)^{1/(\tau-1)} n^{1/(\tau-1)}$, while a uniformly chosen vertex has a bounded weight. Thus, Theorem 1.5 can be interpreted as saying that the highest-weight vertices characterize the largest components. In the subcritical case (see, e.g., the results by Janson in [24] or Theorem 1.3), the largest connected component is the one of the vertex with the highest weight, and the critical situation arises when the highest-weight vertices start connecting to each other.

Connection to the multiplicative coalescent. The mental picture associated with the entrance boundary of the coalescent here seems to be different from [3], where in spirit many of the component sizes are of order $n^{2/3}$. Here the entrance boundary describes the sizes of the maximal components rescaled by $n^{-(\tau-2)/(\tau-1)}$ in the $\lambda \to -\infty$ regime, whilst in [3] they arise as limits of random graphs similar to critical Erdős–Rényi random graphs, where, in addition to the random edges, there are initially a number of large "planted" components of sizes $\lfloor c_i n^{1/3} \rfloor$; see [3], Section 1.3. However, the results of [3] are crucial in identifying the distribution of the limiting component sizes for fixed λ. It would be interesting to see if the stochastic calculus techniques developed in [3] can be further modified to give useful information about the surplus of edges in the maximal components [the surplus of a component $\mathcal{C}$ with $E(\mathcal{C})$ edges and $V(\mathcal{C})$ vertices is equal to $E(\mathcal{C}) - (V(\mathcal{C}) - 1)$ and denotes the minimal number of edges that must be removed from the component to make it a tree].

2. The scaling limit of the cluster of vertex 1. In this section, we identify the scaling limit of $|\mathcal{C}(1)|$. We note from (1.5) that the weight of vertex 1 is *maximal*, that is, $w_1 \geq w_2 \geq \cdots \geq w_n$. When $\tau > 4$, the probability that vertex 1 belongs to $\mathcal{C}_{\max}$ is negligible. When $\tau \in (3, 4)$, instead, we shall see that vertex 1 is in $\mathcal{C}_{\max}$ with *positive* probability, so that it is quite reasonable to start exploring the cluster of vertex 1 first, since $|\mathcal{C}(1)|$ stochastically dominates $|\mathcal{C}(j)|$ for all $j \in [n]$. Theorem 2.1 below states that $|\mathcal{C}(1)|$ is of order $n^{(\tau-2)/(\tau-1)}$. By [34], Theorem 1.2, the same is valid for $|\mathcal{C}_{\max}|$, which confirms the above heuristic.

THEOREM 2.1 [Weak convergence of the cluster of vertex 1 for $\tau \in (3,4)$]. *Fix the Norros–Reittu random graph with weights* $\mathbf{w}(\lambda)$ *defined in* (1.15). *Assume that* $\nu = 1$ *and that* (1.6) *holds. Then, for all* $\lambda \in \mathbb{R}$,

$$n^{-(\tau-2)/(\tau-1)}|\mathcal{C}(1)| \xrightarrow{d} H_1(0) \tag{2.1}$$

for some nondegenerate limit $H_1(0)$.

Theorem 2.1 is proved in Section 3.2. We now start by discussing *cluster explorations* and their relation to branching processes, which play an essential role in our proofs.

2.1. *Cluster explorations and their relation to branching processes.* We fix the weight sequence to be $\mathbf{w}(\lambda)$ defined in (1.15), and we shall denote the weight of vertex i [or the ith coordinate of $\mathbf{w}(\lambda)$] by $w_i(\lambda)$.

In order to prove Theorem 2.1, we make heavy use of the *cluster exploration*, which is described in detail in [31] and [34]. The model in [31] is a *random multigraph*, that is, a random graph potentially having self-loops and multiple edges. Indeed, for each $i, j \in [n]$, we let the number of edges between vertex i and j be $\mathrm{Poi}(w_i(\lambda)w_j(\lambda)/\ell_n(\lambda))$, where, for $\mu \geq 0$, we let $\mathrm{Poi}(\mu)$ denote a Poisson random variable with mean μ, and we define

$$\ell_n(\lambda) = \sum_{i\in[n]} w_i(\lambda) = \ell_n\bigl(1+\lambda n^{-(\tau-3)/(\tau-1)}\bigr). \tag{2.2}$$

The number of edges between different pairs of vertices are *independent*. To retrieve our random graph model, we merge multiple edges and erase self-loops. Then, the probability that an edge exists between two vertices $i, j \in [n]$ is equal to

$$p_{ij} = \mathbb{P}\bigl(\mathrm{Poi}\bigl(w_i(\lambda)w_j(\lambda)/\ell_n(\lambda)\bigr) \geq 1\bigr) = 1 - \mathrm{e}^{-w_i(\lambda)w_j(\lambda)/\ell_n(\lambda)} \tag{2.3}$$

as required. Further, the number of potential edges from a vertex i has a Poisson distribution with mean $w_i(\lambda)$. We shall work with the above Poisson random graph instead, and we shall refer to the Poisson random variable $\mathrm{Poi}(w_i(\lambda))$ as the number of *potential neighbors* of vertex i. When we find what the vertices are that correspond to these $\mathrm{Poi}(w_i(\lambda))$ potential neighbors, that is, when we determine their *marks*, then we can see how many real neighbors there are. Here by a "mark" we mean a random variable M with distribution

$$\mathbb{P}(M = m) = w_m(\lambda)/\ell_n(\lambda) = w_m/\ell_n, \qquad 1 \leq m \leq n. \tag{2.4}$$

The variable M corresponds to the actual *vertex label* associated to the potential vertex. A potential vertex arising in our exploration process is an actual vertex when its mark has not arisen in the exploration up to that point. We now describe this cluster exploration in detail.

We denote by $(Z_l)_{l\geq 0}$ the exploration process in the breadth-first search, where $Z_0 = 1$ and where Z_1 denotes the number of potential neighbors of the initial vertex (which is in the case of Theorem 2.1 equal to vertex 1, and which we shall often take to be vertex i). The variable Z_l has the interpretation of the number of potential neighbors of the first l explored potential vertices in the cluster whose neighbors have not yet been explored. As a result, we explore by taking one vertex of the "stack" of size Z_l, drawing its mark and checking whether it is a real vertex, followed by drawing its number of potential neighbors. Thus, we set $Z_0 = 1$, $Z_1 = \mathrm{Poi}(w_i(\lambda))$, and note that, for $l \geq 2$, Z_l satisfies the recursion relation

$$Z_l = Z_{l-1} + X_l - 1, \tag{2.5}$$

where X_l denotes the number of potential neighbors of the lth potential vertex that is explored. More precisely, when we explore a potential vertex, we start by drawing its mark in an i.i.d. way with distribution (2.4). When we have already explored a vertex with the same mark as the one drawn, we turn the status of the vertex to be explored to inactive, the potential vertex does not become a real vertex, and proceed with the next potential vertex. When, instead, it receives a mark which we have not yet seen, then the potential vertex becomes a real vertex, its mark $M_l \in [n]$ indicating which vertex in $[n]$ the lth explored vertex corresponds to, so that $M_l \in \mathcal{C}(i)$. We then draw $X_l = \mathrm{Poi}(w_{M_l})$, and X_l denotes the number of potential vertices incident to the real vertex M_l. Again, upon exploration, these potential vertices might become real vertices, and this occurs precisely when their mark corresponds to a vertex in $[n]$ that has not appeared in the cluster exploration so far. We call the above procedure of drawing a mark for a potential vertex to investigate whether it corresponds to a real vertex a *vertex check*.

In [31], Proposition 3.1 (see also [34], Section 3.2, in particular, Proposition 3.4), the cluster exploration was described in terms of a thinned marked mixed Poisson branching process. This description implies that the distribution of X_l (for $2 \leq l \leq n$) is equal to $\mathrm{Poi}(w_{M_l}(\lambda))J_l$, where (a) the marks $(M_l)_{l=2}^{\infty}$ are i.i.d. random variables with distribution (2.4); and (b) $J_l = \mathbb{1}_{\{M_l \notin \{i\} \cup \{M_2,\dots,M_{l-1}\}\}}$ is the indicator that the mark M_l has not been found before and is not 1. Here, the mark M_l is the label of the potential element of the cluster that we are exploring, and, clearly, if a vertex has already been observed to be part of $\mathcal{C}(i)$, and its neighbors have been explored, then we should not do so again. We sometimes write $J_j^{(i)}$, $X_l^{(i)}$ and $Z_l^{(i)}$ to explicitly indicate the vertex whose cluster we are exploring, and omit the superscript when no confusion can arise.

We conclude that we arrive at, for $l \geq 2$,

$$\begin{aligned} &Z_l = Z_{l-1} + X_l - 1 \\ &\quad \text{where } X_l = \mathrm{Poi}(w_{M_l}(\lambda))J_l \quad \text{and} \quad J_l = \mathbb{1}_{\{M_l \notin \{i\} \cup \{M_2,\dots,M_{l-1}\}\}}. \end{aligned} \tag{2.6}$$

Then, the number of vertex checks that have been performed when exploring the cluster of vertex i equals $V(i)$, which is given by

$$V(i) = \inf\{l : Z_l = 0\} \tag{2.7}$$

since the first time at which there are no more potential vertices to be checked, all vertices in the cluster have been checked.

Further, the number of *real vertices* found to be part of $\mathcal{C}(i)$ after l vertex checks equals

$$|\mathcal{C}(i;l)| = 1 + \sum_{j=2}^{l} J_j, \tag{2.8}$$

that is, all the potential vertices, except for those that have a mark that has appeared previously. Therefore, we conclude that

$$|\mathcal{C}(i)| = 1 + \sum_{j=2}^{V(i)} J_j = V(i) - \sum_{j=2}^{V(i)} (1 - J_j). \tag{2.9}$$

It turns out that the second contribution is an error term (see Lemma 3.6 below), so that the cluster size of 1 asymptotically corresponds to the first hitting time of 0 of $l \mapsto Z_l$. We prove Theorem 2.1 by applying the above to $i = 1$.

Throughout the paper, we abbreviate

$$\alpha = 1/(\tau - 1), \qquad \rho = (\tau - 2)/(\tau - 1), \qquad \eta = (\tau - 3)/(\tau - 1). \tag{2.10}$$

2.2. *Branching process computations*. In this section, we discuss some useful facts about branching processes. Note that if, in the recursion arising in the exploration of the cluster in (2.6), we ignore the J_l's (i.e., we ignore the effect of marks that have already been used), then we arrive at the recursion

$$Z_l^{(\mathrm{BP})} = Z_{l-1}^{(\mathrm{BP})} + X_l^{(\mathrm{BP})} - 1, \tag{2.11}$$

where now

$$X_l^{(\mathrm{BP})} = \mathrm{Poi}(w_{M_l}(\lambda)), \tag{2.12}$$

and where $(\mathrm{Poi}(w_{M_l}(\lambda)))_{l \geq 2}$ are i.i.d. random variables, while $M_1 = i$. This recursion is the random walk description in the exploration of the total progeny of a branching process. Indeed, let

$$T(i) = \inf\{l : Z_l^{(\mathrm{BP})} = 0\} \tag{2.13}$$

be the first hitting time of 0 of the process $(Z_l^{(\mathrm{BP})})_{l \geq 0}$. Then, by the random walk description of a branching process (see, e.g., [35], Section 3.3), $T(i)$ has the same distribution as the total progeny of a branching process in which the root has offspring distribution $\mathrm{Poi}(w_i(\lambda))$, while the offspring of all other individuals is i.i.d. with mixed Poisson offspring distribution $\mathrm{Poi}(w_M(\lambda))$, where M is the mark distribution in (2.4). In the setting in Section 2.1, we have $i = M_1 = 1$, so that we start from the root having mark 1, but in this section, we shall generalize as well to $M_1 = i$, where $i \in [n]$. Further, we shall also denote the total progeny of the branching process with offspring distribution $\mathrm{Poi}(w_M(\lambda))$ by T. In this section, we investigate properties of such branching processes.

The connection to branching processes [in particular, the *stochastic domination* of the cluster sizes by branching processes due to (2.6)] plays a crucial role in [34], where this comparison was used in order to prove that $n^{-\rho}|\mathcal{C}_{\max}|$ and $n^{\rho}/|\mathcal{C}_{\max}|$ are tight sequences of random variables. There, only *bounds* on the maximal cluster size were shown, while, in this paper, we identify the *scaling limit* of all large clusters.

The difference between the branching process recursion relation in (2.11) and (2.12), and the corresponding one for the cluster exploration in (2.6) resides in the random variables $(J_l)_{l\geq 1}$. Indeed, when $J_l = 0$, then $X_l = \mathrm{Poi}(w_{M_l}(\lambda))J_l = 0$, while $X_l^{(\mathrm{BP})} = \mathrm{Poi}(w_{M_l}(\lambda))$ is unaffected. Therefore, we can think of this procedure as a *thinning* of our branching process. Indeed, when the mark of the lth potential vertex has been seen before, then, in the cluster exploration, we remove this vertex and *all of its offspring*. Thus, the recursions in (2.11), (2.12) and (2.6) give us a *simultaneous coupling* of the cluster exploration process and the branching process such that any deviation between the two arises from the thinning of the potential vertices and the subsequent removal of the branching process tree that is attached to the thinned potential vertices. This description shall prove to be crucial in the comparison of cluster sizes and branching process total progenies used in the proofs of Theorems 1.2 and 1.3.

We continue to investigate the critical behavior of the branching processes at hand. We denote

$$\nu_n(\lambda) = \frac{1}{\ell_n} \sum_{j \in [n]} w_j w_j(\lambda), \tag{2.14}$$

and we write $\nu_n = \nu_n(0)$. Then, we note that

$$\nu_n(\lambda) = \mathbb{E}[\mathrm{Poi}(w_M(\lambda))], \tag{2.15}$$

so that $\nu_n(\lambda)$ is the *mean offspring* of the branching process, and $\nu_n(\lambda) \to 1$ corresponds to our branching process being *critical*. Further,

$$\begin{aligned} &\mathbb{E}\big[\mathrm{Poi}(w_M(\lambda))\big(\mathrm{Poi}(w_M(\lambda)) - 1\big)\big] \\ &\quad = \mathbb{E}[w_M^2(\lambda)] = \frac{1}{\ell_n} \sum_{j \in [n]} w_j w_j(\lambda)^2 \to \infty, \end{aligned} \tag{2.16}$$

so that our branching process has asymptotically *infinite variance* in the setting in (1.6). We now give detailed asymptotics for the mean $\nu_n = \nu_n(0)$ of the above branching process. From this asymptotics, we can easily deduce the asymptotics of $\nu_n(\lambda) = \nu_n(1 + \lambda n^{-\eta})$ [recall (2.10), (2.14) and (1.15)].

LEMMA 2.2 (Sharp asymptotics of ν_n). *Let the distribution function F satisfy* (1.6), *and let $\nu_n = \nu_n(0)$ be given by* (2.14) *and ν by* (1.11). *Then, with η given in* (2.10),

$$\nu_n = \nu + \zeta n^{-\eta} + o(n^{-\eta}), \tag{2.17}$$

where

$$\zeta = -\frac{c_F^{2/(\tau-1)}}{\mathbb{E}[W]} \sum_{i=1}^{\infty} \left[\int_{i-1}^{i} u^{-2\alpha}\, du - i^{-2\alpha} \right] \in (-\infty, 0). \tag{2.18}$$

PROOF. By [34], Corollary 3.2, $\ell_n = \sum_{i\in[n]} w_i = n\mathbb{E}[W] + O(n^{\alpha})$, where it is also proved that $\nu_n - \nu = O(n^{-\eta})$. The sharper asymptotics for ν_n in (2.17) is obtained by a more careful analysis of the arising sum. We note that, by the remark below (1.7),

$$\nu = \frac{\int_0^1 [1-F]^{-1}(u)^2\, du}{\int_0^1 [1-F]^{-1}(u)\, du}. \tag{2.19}$$

By the asymptotics of ℓ_n above, we have that

$$\nu_n = \frac{\sum_{i\in[n]} w_i^2}{n\mathbb{E}[W]} + o(n^{-\eta}). \tag{2.20}$$

We shall make use of the fact that, when f is nonincreasing,

$$f(i) \le \int_{i-1}^{i} f(u)\, du \le f(i-1). \tag{2.21}$$

Applying this to $f(u) = [1-F]^{-1}(u)^2$, which is nonincreasing, we obtain in particular that, for any $K \ge 1$,

$$\begin{aligned} &\int_{K/n}^{1} [1-F]^{-1}(u)^2\, du - \frac{1}{n} w_{K/n}^2 \\ &\qquad \le \frac{1}{n} \sum_{i=K+1}^{n} w_i^2 \le \int_{K/n}^{1} [1-F]^{-1}(u)^2\, du. \end{aligned} \tag{2.22}$$

Now,

$$\frac{1}{n} w_{K/n}^2 = \frac{c_F}{n} (n/K)^{2\alpha} \bigl(1 + o(1)\bigr) = \Theta(K^{-2\alpha} n^{-\eta}). \tag{2.23}$$

Thus we conclude that

$$\begin{aligned} \nu - \nu_n = {}& \frac{1}{\mathbb{E}[W] n} \sum_{i=1}^{K} \int_{(i-1)/n}^{i/n} [1-F]^{-1}(u)^2\, du - \frac{1}{\mathbb{E}[W] n} \sum_{i=1}^{K} w_i^2 \\ & + \Theta(K^{-2\alpha} n^{-\eta}) + o(n^{-\eta}). \end{aligned} \tag{2.24}$$

Next, by (1.6), for every $K \geq 1$ fixed,

$$\frac{1}{n}\sum_{i=1}^{K} w_i^2 = n^{-\eta}\sum_{i=1}^{K}(c_F/i)^{2\alpha} + o(n^{-\eta}) \tag{2.25}$$

and

$$\frac{1}{n}\sum_{i=1}^{K}\int_{(i-1)/n}^{i/n}[1-F]^{-1}(u)^2\,du = n^{-\eta}\sum_{i=1}^{K}\int_{i-1}^{i}(c_F/u)^{2\alpha}\,du + o(n^{-\eta}). \tag{2.26}$$

Combining these two estimates yields

$$n^{\eta}[\nu - \nu_n] = \frac{c_F^{2\alpha}}{\mathbb{E}[W]}\sum_{i=1}^{K}\biggl[\int_{i-1}^{i} u^{-2\alpha}\,du - i^{-2\alpha}\biggr] + \Theta(K^{-2\alpha}) + o(1). \tag{2.27}$$

Letting first $n \to \infty$ followed by $K \to \infty$, we conclude that

$$\lim_{n\to\infty} n^{\eta}[\nu_n - \nu] = \zeta \tag{2.28}$$

as required. The fact that $\zeta > -\infty$ follows from the fact that, for $i \geq 2$,

$$0 \leq \int_{i-1}^{i} u^{-2\alpha}\,du - i^{-2\alpha} \leq (i-1)^{-2\alpha} - i^{-2\alpha}, \tag{2.29}$$

which is a summable sequence. □

We conclude that, in the critical regime where $\nu = 1$, we have

$$\nu_n(\lambda) = 1 + \theta n^{-\eta} + o(n^{-\eta}), \tag{2.30}$$

where $\theta = \lambda + \zeta$. The parameter $\theta \in \mathbb{R}$ indicates the location inside the critical window formed by the weights $\mathbf{w}(\lambda)$. Indeed, in the asymptotics for $\nu_n(\lambda)$ in (2.30), the fact that $\theta = \zeta + \lambda$ arises from $\nu_n(\lambda) = (1 + \lambda n^{-\eta})\nu_n$, together with the sharp asymptotics of ν_n in (2.17). The value of ζ is constant and does not depend on λ, while the value of λ indicates the location inside the scaling window, so we can, alternatively, measure the location inside the scaling window by $\theta \in \mathbb{R}$.

We continue by computing first and second moments of total progenies and their weights, where, for our marked mixed Poisson branching processes, we define the *weight of the branching process total progeny* to be

$$w_T = \sum_{l=1}^{T} w_{M_l} \tag{2.31}$$

and similar for $w_{T(i)}$. Then we can compute the following moments, the proof of which is standard and shall be omitted:

LEMMA 2.3 (Branching process characteristics).

(a)

$$\mathbb{E}[T]=\frac{1}{1-\nu_n},\qquad \mathbb{E}[T^2]=\frac{1+\nu_n}{(1-\nu_n)^2}+\frac{1}{(1-\nu_n)^3}\sum_{j\in[n]}\frac{w_j^3}{l_n}. \tag{2.32}$$

(b)

$$\mathbb{E}[w_T]=\frac{\nu_n}{1-\nu_n},\qquad \mathbb{E}[w_T^2]=\frac{1}{(1-\nu_n)^3}\sum_{j\in[n]}\frac{w_j^3}{l_n}. \tag{2.33}$$

(c)

$$\begin{aligned}&\mathbb{E}[T(i)]=1+\frac{w_i}{1-\nu_n},\\ &\mathbb{E}[T(i)^2]=\left(1+\frac{w_i}{1-\nu_n}\right)^2+\frac{w_i(1+\nu_n)}{(1-\nu_n)^2}+\frac{w_i}{(1-\nu_n)^3}\sum_{j\in[n]}\frac{w_j^3}{l_n}.\end{aligned} \tag{2.34}$$

(d)

$$\mathbb{E}[w_{T(i)}]=\frac{w_i}{1-\nu_n},\qquad \mathbb{E}[w_{T(i)}^2]=\left(\frac{w_i}{1-\nu_n}\right)^2+\frac{w_i}{(1-\nu_n)^3}\sum_{j\in[n]}\frac{w_j^3}{l_n}. \tag{2.35}$$

2.3. *Scaling limit of the cluster exploration process*. Theorem 2.1 will follow from the fact that we can identify the scaling limit of the process $(Z_l)_{l\geq 0}$. To do so, we let

$$\mathcal{Z}_t^{(n)}=n^{-1/(\tau-1)}Z_{tn^{(\tau-2)/(\tau-1)}}=n^{-\alpha}Z_{tn^{\rho}}, \tag{2.36}$$

where we recall the abbreviations in (2.10). By convention, for $t\geq 0$ and for a discrete-time process $(S_l)_{l\geq 0}$, we let $S_t=S_{\lfloor t\rfloor}$.

The intuition behind (2.36) is as follows. First, since the largest connected components are of order n^ρ as proved in [34], Theorem 1.2, and the successive elapsed times between hits of zero of the process $(Z_l)_{l\geq 0}$ correspond to the cluster sizes, the relevant time scale is tn^ρ. Further, by Theorem 1.6, we see that the large clusters correspond to the clusters of the high-weight vertices. The maximal weight is of the order n^α, so that this needs to be the relevant scale on which the process Z_l runs. The proof below makes this intuition precise.

In order to define the scaling limit, we introduce a nonnegative continuous-time process $(\mathcal{S}_t)_{t\geq 0}$. For some $a>0$, we let $(\mathcal{I}_i(t))_{i=1}^\infty$ denote independent increasing indicator processes defined by

$$\mathcal{I}_i(s)=\mathbb{1}_{\{\mathrm{Exp}(ai^{-\alpha})\in[0,s]\}},\qquad s\geq 0, \tag{2.37}$$

so that

$$\mathbb{P}\big(\mathcal{I}_i(s)=0\ \forall s\in[0,t]\big)=\mathrm{e}^{-ati^{-\alpha}}. \tag{2.38}$$

We further let, for some $b>0$ and $c\in\mathbb{R}$, and a as in (2.37),

$$\mathcal{S}_t=b-abt+ct+\sum_{i=2}^{\infty}bi^{-\alpha}[\mathcal{I}_i(t)-ati^{-\alpha}] \tag{2.39}$$

for all $t\geq 0$. We call $(\mathcal{S}_t)_{t\geq 0}$ a *thinned Lévy process*, a name we shall explain in more detail after the theorem. To make the dependence on (a,b,c) explicit, we now denote $\mathcal{S}_t=\mathcal{S}_t(a,b,c)$. Then, we have the obvious scaling relation

$$\mathcal{S}_t(a,b,c)=b\mathcal{S}_{at}\big(1,1,c/(ab)\big), \tag{2.40}$$

where

$$\begin{aligned}\mathcal{S}_t(1,1,\beta)&=1+(\beta-1)t+\sum_{i=2}^{\infty}i^{-\alpha}[\mathcal{I}_i(t)-ti^{-\alpha}],\\ \mathcal{I}_i(t)&=\mathbb{1}_{\{\mathrm{Exp}(i^{-\alpha})\in[0,t]\}}.\end{aligned} \tag{2.41}$$

The main result concerning the scaling limit of the exploration process is the following theorem:

THEOREM 2.4 (The scaling limit of Z_l). *As $n\to\infty$, under the conditions of Theorem 1.1,*

$$\big(\mathcal{Z}_t^{(n)}\big)_{t\geq 0}\xrightarrow{d}(\mathcal{S}_t)_{t\geq 0}, \tag{2.42}$$

where $a=c_F^{\alpha}/\mathbb{E}[W]$, $b=c_F^{\alpha}$, $c=\theta-ab$, in the sense of convergence in the J_1-Skorokhod topology on the space of càdlàg functions on $\mathbb{R}^+$.

It is worthwhile to note that while the convergence in Theorem 2.4 only has implications for our random graph for $t\leq H_1(0)$, which is the hitting time of zero of the process $(\mathcal{S}_t)_{t\geq 0}$, the processes $(\mathcal{Z}_t^{(n)})_{t\geq 0}$ and $(\mathcal{S}_t)_{t\geq 0}$ are well defined also for larger t, and convergence holds for *all* t. This is, in fact, useful in the proof.

The proof of Theorem 2.4 shall be given in Section 3 below. We now first discuss the limiting process $(\mathcal{S}_t)_{t\geq 0}$ and its connection to Lévy processes. To do this, we denote by $(\mathcal{R}_t)_{t\geq 0}$ the process given by

$$\mathcal{R}_t=b-abt+ct+\sum_{i=2}^{\infty}bi^{-\alpha}[N_i(t)-ati^{-\alpha}], \tag{2.43}$$

where $(N_i)_{t\geq 0}$ are independent Poisson processes with rates $ai^{-\alpha}$. Clearly, the process $(\mathcal{R}_t)_{t\geq 0}$ is a spectrally positive Lévy process, that is, $(\mathcal{R}_t)_{t\geq 0}$ has no negative jumps (see, e.g., [6, 29] for more information on Lévy processes), with exponent $\psi(\vartheta)$ [for which $\mathbb{E}(\mathrm{e}^{-\vartheta(\mathcal{R}_t-\mathcal{R}_0)}) = \mathrm{e}^{-t\psi(\vartheta)}$] given by

$$\psi(\vartheta) = (c - ab)\vartheta + \sum_{i=2}^{\infty} ai^{-\alpha}[1 - \mathrm{e}^{-\vartheta bi^{-\alpha}} - b\vartheta i^{-\alpha}]. \tag{2.44}$$

Alternatively, the exponent $\psi(\vartheta)$ can be expressed as

$$\begin{aligned}\psi(\vartheta) &= (c - ab)\vartheta - \vartheta \int_1^{\infty} x\Pi(dx) \\ &\quad + \int_0^{\infty} (1 - \mathrm{e}^{-\vartheta x} - \vartheta x \mathbb{1}_{\{x<1\}})\Pi(dx),\end{aligned} \tag{2.45}$$

where the Lévy measure Π is defined by

$$\Pi(dx) = \sum_{i=2}^{\infty} ai^{-\alpha}\delta_{x,bi^{-\alpha}}. \tag{2.46}$$

Since $\Pi(b, \infty) = 0$, the jumps of $(\mathcal{R}_t)_{t\geq 0}$ are bounded by b. Further,

$$\int_0^{\infty} (1 \wedge x^2)\Pi(dx) \leq \int_0^{\infty} x^2\Pi(dx) = a\sum_{i=2}^{\infty}\left(\frac{b}{i^{\alpha}}\right)^3 = ab^3\sum_{i=2}^{\infty} i^{-3\alpha} < \infty, \tag{2.47}$$

since $\tau \in (3, 4)$ so that $3\alpha = 3/(\tau - 1) > 1$. Therefore, the process $(\mathcal{R}_t)_{t\geq 0}$ is a well-defined Lévy process.

We may reformulate (2.39) as

$$\mathcal{S}_t = b - abt + ct + \sum_{i=2}^{\infty} bi^{-\alpha}[\mathbb{1}_{\{N_i(t)\geq 1\}} - ati^{-\alpha}], \tag{2.48}$$

so that the process $(\mathcal{S}_t)_{t\geq 0}$ does not include multiple counts of the independent processes $(N_i(t))_{t\geq 0}$. This is the reason that we call the process $(\mathcal{S}_t)_{t\geq 0}$ a *thinned* Lévy process. In [3], this process is called a Lévy process *without repetitions*. Naturally, we have that the descriptions in (2.43) and (2.48) satisfy that, a.s., for all $t \geq 0$,

$$\mathcal{S}_t \leq \mathcal{R}_t, \tag{2.49}$$

which allows us to make use of Lévy process methodology in our proofs. We do note that $\mathcal{R}_t$ is a rather poor approximation for $\mathcal{S}_t$, particularly on large time scales, because the *thinning* becomes more important as time progresses.

3. Proofs of Theorems 2.1 and 2.4. In this section, we prove Theorems 2.1 and 2.4. We start by proving Theorem 2.4 in Section 3.1, and make use of Theorem 2.4 to prove Theorem 2.1 in Section 3.2.

3.1. *Proof of Theorem* 2.4. Instead of $(Z_l)_{l\geq 0}$, it is convenient to work with a related process $(S_l)_{l\geq 0}$, which is defined as $S_0 = 1$, $S_1 = w_1(\lambda)$ and satisfies the recursion relation, for $l \geq 2$,

$$S_l = S_{l-1} + w_{M_l}(\lambda) J_l - 1, \tag{3.1}$$

that is, the Poisson random variables $\mathrm{Poi}(w_{M_l}(\lambda))$ appearing in the recursion for Z_l in (2.6) are replaced with their (random) weights $w_{M_l}(\lambda)$. We shall first show that S_l and Z_l are quite close:

LEMMA 3.1. *Uniformly in* $m \geq 0$,

$$\sup_{l\leq m} |Z_l - S_l| = O_{\mathbb{P}}(m^{1/2}). \tag{3.2}$$

PROOF. We have that $(Z_l - S_l)_{l\geq 0}$ is a martingale w.r.t. the filtration $\mathcal{F}_l = \sigma((M_i)_{i=1}^l)$. Therefore, by the Doob–Kolmogorov inequality ([21], Theorem (7.8.2), page 338) for any $M > 0$,

$$\mathbb{P}\Big(\sup_{l\leq m} |Z_l - S_l| > M\sqrt{m}\Big) \leq \frac{1}{mM^2}\mathbb{E}[|Z_m - S_m|^2]. \tag{3.3}$$

Now,

$$\begin{aligned}\mathbb{E}[|Z_m - S_m|^2] &= \mathbb{E}\big[\mathbb{E}[|Z_m - S_m|^2 \mid (M_i)_{i=1}^m]\big] = \mathbb{E}\Bigg[\sum_{l=1}^m w_{M_l}(\lambda) J_l\Bigg] \\ &\leq \mathbb{E}\Bigg[\sum_{l=1}^m w_{M_l}(\lambda)\Bigg] = m\nu_n(\lambda) = m\big(1 + o(1)\big)\end{aligned} \tag{3.4}$$

by (2.30). This proves the claim. □

We proceed by investigating the scaling limit of $(S_l)_{l\geq 1}$. For this, we define

$$\mathcal{S}_t^{(n)} = n^{-\alpha} S_{tn^\rho}, \tag{3.5}$$

where we recall the rounding convention right below (2.36).

We shall prove that, in the sense of convergence in the J_1-Skorokhod topology on the space of càdlàg functions on $\mathbb{R}^+$,

$$\big(\mathcal{S}_t^{(n)}\big)_{t\geq 0} \xrightarrow{d} (\mathcal{S}_t)_{t\geq 0}, \tag{3.6}$$

which shall be enough to prove Theorem 2.4. Indeed, to see that (3.6) implies Theorem 2.4, we note that by Lemma 3.1, for every $t = o(n^{(4-\tau)/(\tau-1)})$,

$$\sup_{s\leq t}\big|\mathcal{Z}_s^{(n)} - \mathcal{S}_s^{(n)}\big| = O_{\mathbb{P}}\big(\sqrt{t} n^{(\tau-4)/(2(\tau-1))}\big) = o_{\mathbb{P}}(1). \tag{3.7}$$

We continue with the proof of (3.6). We shall prove that, due to (2.9) and Lemma 3.1, the first hitting time of $\mathcal{S}_s^{(n)}$ of 0 is close to $n^{-\rho}|\mathcal{C}_{\le}(1)|$. We note that, by (3.1),

$$
\begin{aligned}
S_l &= w_1(\lambda) + \sum_{i\in\mathcal{V}_l^{(n)}} w_i(\lambda) - (l-1) \\
&= w_1(\lambda) + \sum_{i=2}^{n} w_i(\lambda)\mathcal{I}_i^{(n)}(l) - (l-1),
\end{aligned}
\tag{3.8}
$$

where

$$
\mathcal{I}_i^{(n)}(l) = \mathbb{1}_{\{i\in\mathcal{V}_l^{(n)}\}} \qquad \text{with} \qquad \mathcal{V}_l^{(n)} = \bigcup_{j=2}^{l}\{M_j\}.
\tag{3.9}
$$

Using that

$$
\nu_n(\lambda) = \sum_{i\in[n]} \frac{w_i(\lambda)w_i}{\ell_n},
\tag{3.10}
$$

we can rewrite S_l as

$$
\begin{aligned}
S_l &= w_1(\lambda) - \frac{(l-1)w_1(\lambda)w_1}{\ell_n} + \sum_{i=2}^{n} w_i(\lambda)\biggl[\mathcal{I}_i^{(n)}(l) - \frac{(l-1)w_i}{\ell_n}\biggr] \\
&\quad + \bigl(\nu_n(\lambda)-1\bigr)(l-1).
\end{aligned}
\tag{3.11}
$$

Now we take $l = tn^{\rho}$, use that $\nu_n(\lambda) - 1 = \theta n^{-\eta}\nu_n + o(n^{-\eta})$ [recall (2.30) and (2.10)], and we recall from (1.5) and (1.6) that, for i such that $n/i \to \infty$,

$$
w_i = [1-F]^{-1}(i/n) = b(n/i)^{\alpha}\bigl(1+o(1)\bigr),
\tag{3.12}
$$

where $b = c_F^{\alpha}$ and c_F is defined in (1.6). As a result, by (3.11),

$$
\begin{aligned}
\mathcal{S}_t^{(n)} &= n^{-\alpha} S_{tn^{\rho}} \\
&= b - \frac{b^2}{\mathbb{E}[W]}t + \sum_{i=2}^{n} n^{-\alpha} w_i(\lambda)\biggl[\mathcal{I}_i^{(n)}(tn^{\rho}) - n^{-\alpha}\frac{w_i t}{\mu_n}\biggr] \\
&\quad + \theta t + o(1),
\end{aligned}
\tag{3.13}
$$

where we write $\mu_n = \ell_n/n = \mathbb{E}[W] + o(1)$.

We proceed by showing that the sum in (3.13) is predominantly carried by the first few terms. Define

$$
M_l^{(n,K)} = \sum_{i=K}^{n} n^{-\alpha} w_i(\lambda)\biggl[\mathcal{I}_i^{(n)}(l) - \frac{(l-1)w_i}{\ell_n}\biggr].
\tag{3.14}
$$

We compute the mean and variance of $M_l^{(n,K)}$ for K large. For the mean, we compute

$$
\begin{aligned}
\mathbb{E}[M_l^{(n,K)}] &= \sum_{i=K}^{n} n^{-\alpha} w_i(\lambda)\Big[\mathbb{P}(\mathcal{I}_i^{(n)}(l)=1) - \frac{(l-1)w_i}{\ell_n}\Big] \\
&= \sum_{i=K}^{n} n^{-\alpha} w_i(\lambda)\Big[\Big(1-\frac{w_i}{\ell_n}\Big)^{l-1} - 1 + \frac{(l-1)w_i}{\ell_n}\Big].
\end{aligned}
\tag{3.15}
$$

Thus, since $0 \le 1-(1-x)^l - lx \le (lx)^2/2$, we have that $\mathbb{E}[M_l^{(n,K)}] \le 0$ and

$$
\begin{aligned}
\big|\mathbb{E}[M_l^{(n,K)}]\big| &= \sum_{i=K}^{n} n^{-\alpha} w_i(\lambda)\Big[1-\Big(1-\frac{w_i}{\ell_n}\Big)^{l-1} - \frac{(l-1)w_i}{\ell_n}\Big] \\
&\le \sum_{i=K}^{n} n^{-\alpha} w_i(\lambda)\Big(\frac{lw_i}{\ell_n}\Big)^2 \le C\frac{l^2 n^{2\alpha}}{\ell_n^2}\sum_{i=K}^{n} i^{-3\alpha} \\
&\le C\frac{l^2n^{2\alpha}}{\ell_n^2}K^{1-3\alpha},
\end{aligned}
\tag{3.16}
$$

where, here and in the sequel, $C > 0$ denotes a constant that can change from line to line. By (2.10) and the fact that $\ell_n = \Theta(n)$, we have that $ln^\alpha/\ell_n = \Theta(ln^{\alpha-1}) = \Theta(ln^{-\rho})$, so that, uniformly in $l \le tn^\rho$,

$$
\big|\mathbb{E}[M_l^{(n,K)}]\big| \le Ct^2K^{1-3\alpha}. \tag{3.17}
$$

To compute the variance of $M_l^{(n,K)}$, we start by noting that $\mathcal{I}_i^{(n)}(l)$ is the indicator that $i \in \mathcal{V}_l^{(n)}$, and $\mathcal{V}_l^{(n)}$ contains the first l marks drawn, where $M_1 = 1$ and the marks $(M_i)_{i=2}^l$ are i.i.d. with distribution given by (2.4). Therefore, $\mathcal{I}_i^{(n)}(l)$ and $\mathcal{I}_j^{(n)}(l)$ are, for different i, j, *negatively correlated*, so that

$$
\mathrm{Var}(M_l^{(n,K)}) \le \sum_{i=K}^{n} (n^{-\alpha}w_i(\lambda))^2\,\mathrm{Var}(\mathcal{I}_i^{(n)}(l)). \tag{3.18}
$$

Since $\mathcal{I}_i^{(n)}(l)$ is an indicator,

$$
\mathrm{Var}(\mathcal{I}_i^{(n)}(l)) \le \mathbb{E}[\mathcal{I}_i^{(n)}(l)] \le lw_i/\ell_n. \tag{3.19}
$$

Therefore, when $l = tn^\rho$, and using that $\rho + \alpha = 1$ [recall (2.10)]

$$
\begin{aligned}
\mathrm{Var}(M_l^{(n,K)}) &\le \sum_{i=K}^{n}(n^{-\alpha}w_i(\lambda))^2\frac{w_i tn^\rho}{\ell_n} \le Ct\sum_{i=K}^{n} i^{-3\alpha} \\
&\le CtK^{1-3\alpha} = o(1),
\end{aligned}
\tag{3.20}
$$

when $K \to \infty$, since $\tau \in (3,4)$, so that $\alpha = 1/(\tau-1) > 1/3$.

We next observe that $(M_l^{(n,K)})_{l\geq 1}$ is a supermartingale, since

$$\begin{aligned}
&\mathbb{E}[M_{l+1}^{(n,K)} - M_l^{(n,K)} \mid (\mathcal{I}_i^{(n)}(l))_{i\in[n]}] \\
&\quad = \mathbb{E}\Bigg[\sum_{i=K}^{n} n^{-\alpha} w_i(\lambda)\Big[\mathcal{I}_i^{(n)}(l+1) - \mathcal{I}_i^{(n)}(l) - \frac{w_i}{\ell_n}\Big] \,\Big|\, (\mathcal{I}_i^{(n)}(l))_{i\in[n]}\Bigg] \\
&\quad \leq \sum_{i=K}^{n} n^{-\alpha} w_i(\lambda)\big(1 - \mathcal{I}_i^{(n)}(l)\big)\Big(\mathbb{E}[\mathcal{I}_i^{(n)}(l+1) \mid (\mathcal{I}_i^{(n)}(l))_{i\in[n]}] - \frac{w_i}{\ell_n}\Big) \\
&\quad = 0.
\end{aligned} \tag{3.21}$$

Therefore, by the maximal inequality ([21], Theorem 12.6.1, page 496),

$$\mathbb{P}\Big(\max_{l\leq m}|M_l^{(n,K)}| \geq \varepsilon\Big) \leq \frac{-\mathbb{E}[M_0^{(n,K)}] + \mathbb{E}[|M_m^{(n,K)}|]}{\varepsilon}. \tag{3.22}$$

We further bound, using Cauchy–Schwarz,

$$\mathbb{E}[|M_m^{(n,K)}|] \leq |\mathbb{E}[M_m^{(n,K)}]| + \sqrt{\mathrm{Var}(M_l^{(n,K)})}. \tag{3.23}$$

Thus by (3.17) and (3.20), and uniformly in $m \leq tn^{\rho}$,

$$\mathbb{P}\Big(\max_{l\leq m}|M_l^{(n,K)}| \geq \varepsilon\Big) \leq Ct^2\varepsilon^{-1}K^{1-3\alpha} + \varepsilon^{-1}\sqrt{CtK^{1-3\alpha}}. \tag{3.24}$$

Since $\tau < 4$, we obtain that, uniformly in n, we can take $K = K(\varepsilon)$ so large that $\mathbb{P}(\max_{l\leq m}|M_l^{(n,K)}| \geq \varepsilon) \leq \varepsilon$.

We denote, with $\mu_n = \ell_n/n$,

$$\mathcal{S}_t^{(n,K)} = b - \frac{b^2}{\mathbb{E}[W]}t + \sum_{i=2}^{K} n^{-\alpha} w_i(\lambda)\Big[\mathcal{I}_i^{(n)}(tn^{\rho}) - n^{-\alpha}\frac{w_i t}{\mu_n}\Big] + \theta t. \tag{3.25}$$

Then we obtain the following corollary:

COROLLARY 3.2 (Finite sum approximation of $\mathcal{Z}^{(n)}$). *For every $\varepsilon, \delta, T > 0$, there exists $K > 0$ and $N \geq 1$ such that for all $n \geq N$,*

$$\mathbb{P}\Big(\sup_{t\leq T}|\mathcal{Z}_t^{(n)} - \mathcal{S}_t^{(n,K)}| \geq \delta\Big) \leq \varepsilon. \tag{3.26}$$

The above suggests that it suffices to investigate $(\mathcal{I}_i^{(n)}(tn^{\rho}))_{i\in[K]}$.

LEMMA 3.3 (Convergence of indicators). *As $n \to \infty$, for all $K \geq 1$,*

$$\big(\mathcal{I}_i^{(n)}(tn^{\rho})\big)_{i\in[K],t\geq 0} \xrightarrow{d} (\mathcal{I}_i(t))_{i\in[K],t\geq 0}. \tag{3.27}$$

As a consequence, for all $K \geq 1$,

$$(\mathcal{S}_t^{(n,K)})_{t\geq 0} \xrightarrow{d} (\mathcal{S}_t^{(\infty,K)})_{t\geq 0}, \tag{3.28}$$

where the limiting process $(\mathcal{S}_t^{(\infty,K)})_{t\geq 0}$ is defined as

$$\mathcal{S}_t^{(\infty,K)} = b - \frac{b^2}{\mathbb{E}[W]} t + \sum_{i=2}^{K} b i^{-\alpha} [\mathcal{I}_i(t) - a i^{-\alpha} t] + \theta t. \tag{3.29}$$

In both statements, $\xrightarrow{d}$ refers to convergence in the J_1-Skorokhod topology on the space of càdlàg functions on $\mathbb{R}^+$.

PROOF. Convergence of the process for $t \geq 0$ follows when the process converges for $t \in [0, T]$ for all $T > 0$ (see [9], Lemma 3, page 173).

Since $(\mathcal{I}_i^{(n)}(tn^{\rho}))_{t\geq 0}$ are all indicator processes of the form

$$\mathcal{I}_i^{(n)}(tn^{\rho}) = \mathbb{1}_{\{T_i \leq tn^{\rho}\}}, \tag{3.30}$$

where T_i is the first time that mark i is chosen, it suffices to prove that

$$(n^{-\rho} T_i)_{i\in[K]} \xrightarrow{d} (E_i)_{i\in[K]}, \tag{3.31}$$

where E_i are independent exponentials with rate $ai^{-\alpha}$. For this, in turn, it suffices to prove that, for every sequence $t_1, \ldots, t_K$,

$$\mathbb{P}(n^{-\rho} T_i > t_i \ \forall i \in [K]) \to \exp\left(-a \sum_{i=1}^{K} i^{-\alpha} t_i\right). \tag{3.32}$$

The latter is equivalent to

$$\begin{aligned} \mathbb{P}(\mathcal{I}_i^{(n)}(t_i n^{\rho}) = 0 \ \forall i \in [K]) &\to \mathbb{P}(\mathcal{I}_i(t_i) = 0 \ \forall i \in [K]) \\ &= \exp\left(-a \sum_{i=1}^{K} i^{-\alpha} t_i\right). \end{aligned} \tag{3.33}$$

Now, since the marks are i.i.d., we obtain that

$$\begin{aligned} \mathbb{P}(\mathcal{I}_i^{(n)}(m_i) = 0 \ \forall i \in [K]) &= \prod_{l=1}^{\infty} \mathbb{P}(M_l \notin \{i \in [K] : l \leq m_i\}) \\ &= \prod_{l=1}^{\infty} \left(1 - \sum_{i : l \leq m_i} \frac{w_i}{\ell_n}\right). \end{aligned} \tag{3.34}$$

A Taylor expansion gives that

$$\begin{aligned}\mathbb{P}\big(\mathcal{I}_i^{(n)}(m_i)=0\ \forall i\in[K]\big)&=\exp\biggl(-\sum_{l=1}^{n}\sum_{i:m_i\geq l}\frac{w_i}{\ell_n}+o(1)\biggr)\\&=\exp\biggl(-\sum_{i\in[K]}\frac{w_im_i}{\ell_n}+o(1)\biggr).\end{aligned}\tag{3.35}$$

Applying this to $m_i=t_in^{\rho}$, for which

$$\frac{m_iw_i}{\ell_n}=\frac{bi^{-\alpha}t_i}{\mathbb{E}[W]}(1+o(1)),\tag{3.36}$$

we arrive at the claim in (3.27) with $a=b/\mathbb{E}[W]$. The claim in (3.28) follows from the fact that, by (3.25), $\mathcal{S}_t^{(n,K)}$ is a weighted sum of the $(\mathcal{I}_i^{(n)}(tn^{\rho}))_{i\in[K]}$, and the (deterministic) weights converge. Thus, the continuous mapping theorem gives the claim. □

PROOF OF THEOREM 2.4. Again we use that convergence of the process for $t\geq 0$ follows when the process converges for $t\in[0,T]$ for all $T>0$ (see [9], Lemma 3, page 173). By (3.26), with probability $1-o(1)$ when first $n\to\infty$ and then $K\to\infty$, the process $(\mathcal{Z}_t^{(n)})_{t\in[0,T]}$ is uniformly close to $(\mathcal{S}_t^{(n,K)})_{t\in[0,T]}$. By Lemma 3.3, the process $(\mathcal{S}_t^{(n,K)})_{t\geq 0}$ converges to $(\mathcal{S}_t^{(\infty,K)})_{t\geq 0}$. Now,

$$\mathcal{S}_t-\mathcal{S}_t^{(\infty,K)}=\sum_{i\geq K+1}bi^{-\alpha}[\mathcal{I}_i(t)-ai^{-\alpha}],\tag{3.37}$$

and similar techniques as used to prove (3.24) can be used to prove that

$$\mathbb{P}\Bigl(\max_{t\leq T}|\mathcal{S}_t-\mathcal{S}_t^{(\infty,K)}|\geq\varepsilon\Bigr)\leq CT^2\varepsilon^{-1}K^{1-3\alpha}+\varepsilon^{-1}\sqrt{CTK^{1-3\alpha}},\tag{3.38}$$

so that again we can take $K=K(\varepsilon)$ so large that $\mathbb{P}(\max_{t\leq T}|\mathcal{S}_t-\mathcal{S}_t^{(\infty,K)}|\geq\varepsilon)\leq\varepsilon$. This proves the claim. □

3.2. *Proof of Theorem* 2.1. In this section, we give a proof of Theorem 2.1. We start by looking at the first hitting time of zero of the process $l\mapsto Z_l$, and use the fact that by (2.7), $V(1)=\inf\{l:Z_l=0\}$, where we recall that $V(1)$ denotes the number of vertex checks performed in exploring the cluster of vertex 1. Recall further that $\mathcal{C}(1)$ denotes the cluster of vertex 1, $|\mathcal{C}(1)|$ the number of vertices in it, and $\mathcal{W}(1)=\sum_{j\in\mathcal{C}(1)}w_j$ its weight.

The proof proceeds as follows. We shall first use Theorem 2.4 and Lemma 3.1 to prove that $V(1)n^{-\rho}$ converges in distribution to $H_{\mathcal{S}}(0)$, where $H_{\mathcal{S}}(0)$ denotes the first hitting time of 0 of the process $(\mathcal{S}_t)_{t\geq 0}$; see Corollary 3.4 below. We then prove that $V(1)n^{-\rho}$, $|\mathcal{C}(1)|n^{-\rho}$ and $\mathcal{W}(1)n^{-\rho}$ have identical scaling limits, by looking at the contribution due to the second term in (2.9) for $|\mathcal{C}(1)|n^{-\rho}$, and a similar computation for $\mathcal{W}(1)n^{-\rho}$; see Lemma 3.6 below. We then complete the proof of Theorem 2.1, both for $|\mathcal{C}(1)|n^{-\rho}$ and for $\mathcal{W}(1)n^{-\rho}$. Finally, in Proposition 3.7, we state and prove an auxiliary result concerning joint convergence of $|\mathcal{C}(1)|n^{-\rho}$ and the indicators $\mathbb{1}_{\{q\in\mathcal{C}(1)\}}$ for all q. This result is useful in the proofs of Theorems 1.1 and 1.5 and plays a crucial role in the proof of Theorem 4.1 in the next section, where we investigate the scaling limit of several clusters simultaneously.

By Theorem 2.4 and Lemma 3.1, the process $(\mathcal{Z}_t^{(n)})_{t\geq 0}$, where $\mathcal{Z}_t^{(n)} = n^{-\alpha} Z_{tn^\rho}$, converges in distribution to the process $(\mathcal{S}_t)_{t\geq 0}$. By (3.6), the same applies to $(\mathcal{S}_t^{(n)})_{t\geq 0}$. Note that

$$n^{-\rho} V(1) = \min\{t : \mathcal{Z}_t^{(n)} = 0\} \equiv H^{(n)}(0). \tag{3.39}$$

We next prove convergence in distribution of $n^{-\rho} V(1)$:

COROLLARY 3.4 (Convergence of hitting times). *As $n \to \infty$,*

$$n^{-\rho} V(1) \xrightarrow{d} H_{\mathcal{S}}(0), \tag{3.40}$$

where

$$H_{\mathcal{S}}(x) = \inf\{t : \mathcal{S}_t \leq x\} \tag{3.41}$$

is the first hitting time of level x of $(\mathcal{S}_t)_{t\geq 0}$.

PROOF. Since the process $(\mathcal{S}_t)_{t\geq 0}$ has only positive jumps ([23], Proposition 2.11 in Chapter 6) implies that the hitting time of zero is a continuous function a.s. under the probability measure of the limiting process on the space of càdlàg functions equipped with the J_1-Skorokhod topology. □

LEMMA 3.5 ($\mathcal{S}_t$ has a density). *For all $t > 0$, $\mathcal{S}_t$ has a density. As a result, the distribution of $H_{\mathcal{S}}(0)$ has no atoms.*

PROOF. We note that $\mathcal{S}_t$ has a density if and only if $\mathcal{S}_t'$ has, where

$$\mathcal{S}_t' = \sum_{j=2}^{\infty} j^{-\alpha}[\mathcal{I}_j'(t) - tj^{-\alpha}], \tag{3.42}$$

and $(\mathcal{I}_j'(t))_{j\geq 2}$ are independent indicator processes with rate $j^{-\alpha}$. This, in turn, follows when the characteristic function of $\mathcal{S}_t'$ is integrable; see, for example, [21], page 189.

The characteristic function of $\mathcal{S}_t'$ is given by

$$\hat{f}_{\mathcal{S}_t'}(\vartheta) = \mathbb{E}[\mathrm{e}^{\mathrm{i}\vartheta\mathcal{S}_t'}] = \prod_{j=2}^{\infty} \mathrm{e}^{-j^{-2\alpha}\mathrm{i}\vartheta}\bigl(1 + (\mathrm{e}^{-j^{-\alpha}\mathrm{i}\vartheta} - 1)\mathrm{e}^{-j^{-\alpha}t}\bigr). \tag{3.43}$$

Thus, for every $j_\vartheta \geq 2$,

$$|\hat{f}_{\mathcal{S}'_t}(\vartheta)| \leq \prod_{j \geq j_\vartheta}^{\infty} |1 + (\mathrm{e}^{-j^{-\alpha}\mathrm{i}\vartheta} - 1)\mathrm{e}^{-j^{-\alpha}t}|. \tag{3.44}$$

Next, note that

$$\begin{aligned}
&|1 + (\mathrm{e}^{-j^{-\alpha}\mathrm{i}\vartheta} - 1)\mathrm{e}^{-j^{-\alpha}t}|^2 \\
&\quad = \mathrm{e}^{-2j^{-\alpha}t} \sin(j^{-\alpha}\vartheta)^2 + \big(1 - \mathrm{e}^{-j^{-\alpha}t} + \cos(j^{-\alpha}\vartheta)\mathrm{e}^{-j^{-\alpha}t}\big)^2 \\
&\quad = 1 - 2(1 - \mathrm{e}^{-j^{-\alpha}t})\mathrm{e}^{-j^{-\alpha}t}[1 - \cos(j^{-\alpha}\vartheta)] \\
&\quad \leq \mathrm{e}^{-2(1-\mathrm{e}^{-j^{-\alpha}t})\mathrm{e}^{-j^{-\alpha}t}[1-\cos(j^{-\alpha}\vartheta)]} \\
&\quad \leq \mathrm{e}^{-j^{-\alpha}t[1-\cos(j^{-\alpha}\vartheta)]},
\end{aligned}$$

so that

$$|\hat{f}_{\mathcal{S}'_t}(\vartheta)| \leq \mathrm{e}^{-t\sum_{j \geq j_\vartheta} j^{-\alpha}t[1-\cos(j^{-\alpha}\vartheta)]} \equiv \mathrm{e}^{-t\Phi(\vartheta)}. \tag{3.45}$$

We choose

$$j_\vartheta = \max\{j \geq 2 : b\vartheta j^{-\alpha} \geq \pi/2\}, \tag{3.46}$$

so that

$$j_\vartheta = \lfloor (2b\vartheta/\pi)^{1/\alpha} \rfloor \vee 2 = \lfloor (2b\vartheta/\pi)^{\tau-1} \rfloor \vee 2. \tag{3.47}$$

Then we bound

$$\Phi(\vartheta) \geq \sum_{j=j_\vartheta}^{\infty} \frac{a}{j^\alpha}[1 - \cos(b\vartheta j^{-\alpha})]. \tag{3.48}$$

Next, we use that

$$1 - \cos(x) \geq \frac{2}{\pi}x^2, \qquad x \in \left[-\frac{1}{2}\pi, \frac{1}{2}\pi\right], \tag{3.49}$$

to arrive at

$$\Phi(\vartheta) \geq c\vartheta^2 \sum_{j=j_\vartheta}^{\infty} j^{-3\alpha}, \tag{3.50}$$

where $c > 0$ denotes a positive constant appearing in lower bounds that possibly changes from line to line. We arrive at the fact that

$$\Phi(\vartheta) \geq c\vartheta^2 j_\vartheta^{1-3\alpha} \geq c\vartheta^2 \vee \vartheta^{\tau-2}, \tag{3.51}$$

so that $|\hat{f}_{\mathcal{S}'_t}(\vartheta)|$ is integrable. To prove that $H_{\mathcal{S}}(0)$ has no atoms, note that when $\mathbb{P}(H_{\mathcal{S}}(0) = u) > 0$ for some $u \geq 0$; then, in particular, $\mathbb{P}(\mathcal{S}_u = 0) > 0$, which contradicts the fact that $\mathcal{S}_u$ has a density. $\square$

We proceed by showing that the scaling limits of the number of vertex checks of a cluster and the cluster size are identical. For this, we shall make use of the following lemma:

LEMMA 3.6 (Number of multiple hits is small). *As $n \to \infty$, for any $m \geq 1$,*

$$\mathbb{E}\left[\sum_{j=2}^{m}[1-J_j]\right] \leq \frac{mw_1}{\ell_n} + \frac{m(m-1)\nu_n}{2\ell_n}. \tag{3.52}$$

Consequently, there exists $t_n \to \infty$, such that

$$n^{-\rho}\sum_{j=2}^{t_n n^{\rho}}[1-J_j] \xrightarrow{\mathbb{P}} 0. \tag{3.53}$$

PROOF. We note that $J_j = 0$ precisely when $M_l = 1$ or when there exists an $l < j$ and $i \in [n]$ such that $M_l = M_j = i$. By independence and (2.4),

$$\mathbb{P}(M_l = M_j = i) = \mathbb{P}(M_l = i)\mathbb{P}(M_j = i) = w_i^2/\ell_n^2. \tag{3.54}$$

Therefore,

$$\mathbb{E}[1-J_j] \leq \frac{w_1}{\ell_n} + \sum_{l=2}^{j-1}\sum_{i=2}^{n}\frac{w_i^2}{\ell_n^2} \leq \frac{w_1}{\ell_n} + (j-1)\frac{\nu_n}{\ell_n}. \tag{3.55}$$

Summing the above inequality over $2 \leq j \leq m$ proves the claim in (3.52).

For (3.53), we use the Markov inequality to bound

$$\mathbb{P}\left(n^{-\rho}\sum_{j=2}^{t_n n^{\rho}}[1-J_j] \geq \varepsilon_n\right) \leq \varepsilon_n^{-1}n^{-\rho}\mathbb{E}\left[\sum_{j=2}^{t_n n^{\rho}}[1-J_j]\right] \leq \frac{t_n w_1}{\varepsilon_n \ell_n} + \frac{t_n^2 n^{\rho}\nu_n}{2\ell_n\varepsilon_n} = o(1),$$

whenever $t_n^2 n^{-\alpha}/\varepsilon_n = o(1)$. Choosing, for example, $t_n = \log n$ and $\varepsilon_n = 1/\log n$ does the trick. □

Now we are ready to complete the proof of Theorem 2.1.

PROOF OF THEOREM 2.1. By Corollary 3.4, $n^{-\rho}V(1) \xrightarrow{d} H_{\mathcal{S}}(0)$. In particular, this implies that $|\mathcal{C}(1)| \leq V(1) \leq n^{\rho}t_n$ for any $t_n \to \infty$. Therefore, by (2.9), and *whp*,

$$n^{-\rho}V(1) - n^{-\rho}\sum_{j=2}^{t_n n^{\rho}}[1-J_j] \leq n^{-\rho}|\mathcal{C}(1)| \leq n^{-\rho}V(1). \tag{3.56}$$

Now, by Lemma 3.6, the difference between the left-hand and right-hand sides of (3.56) converges to zero in probability, so that also

$$n^{-\rho}|\mathcal{C}(1)| \xrightarrow{d} H_{\mathcal{S}}(0). \tag{3.57}$$

This completes the proof of Theorem 2.1 and identifies $H_1(0) = H_{\mathcal{S}}(0)$. In the same vein,

$$\mathcal{W}(1) = \sum_{i \in \mathcal{C}(1)} w_i = \sum_{j=1}^{V(1)} w_{M_j} J_j. \tag{3.58}$$

Now, by (3.1), for any $l \geq 1$,

$$\sum_{j=1}^{l} w_{M_j} J_j = S_l + l. \tag{3.59}$$

As a result,

$$\mathcal{W}(1) = V(1) + S_{V(1)}, \tag{3.60}$$

so that

$$n^{-\rho}\mathcal{W}(1) = n^{-\rho}V(1) + n^{-\rho}S_{V(1)}. \tag{3.61}$$

Finally, $n^{-\rho}V(1) \xrightarrow{d} H_{\mathcal{S}}(0)$, and, since $\alpha < \rho$, $n^{-\rho}|S_{V(1)}| = o(1)n^{-\alpha}|S_{V(1)}| \xrightarrow{\mathbb{P}} 0$. This proves that $n^{-\rho}\mathcal{W}(1) \xrightarrow{d} H_{\mathcal{S}}(0)$ as well. □

In the next section, where we study the joint convergence of various clusters simultaneously, we shall also need the following joint convergence result:

PROPOSITION 3.7 (Weak convergence of functionals). *As $n \to \infty$,*

$$\left(n^{-\rho}|\mathcal{C}(1)|, \left(\mathbb{1}_{\{q \in \mathcal{C}(1)\}}\right)_{q \geq 1}\right) \xrightarrow{d} \left(H_{\mathcal{S}}(0), (\mathcal{I}_q(H_{\mathcal{S}}(0)))_{q \geq 1}\right) \tag{3.62}$$

in the product topology, where $\mathcal{I}_q(H_{\mathcal{S}}(0))$ denotes the indicator that $\mathcal{I}_q(t) = 1$ at the hitting time of 0 *of $(\mathcal{S}_t)_{t \geq 0}$. Moreover,* (i) *the random variable $H_{\mathcal{S}}(0)$ is nondegenerate; and* (ii) *the indicators $(\mathcal{I}_q(H_{\mathcal{S}}(0)))_{q \geq 2}$ are nontrivial in the sense that they take the values* 0 *and* 1 *each with positive probability.*

We note that, while the indicator processes $(\mathcal{I}_q(t))_{t \geq 0}$ are independent for different q, the random variables $(\mathcal{I}_q(H_{\mathcal{S}}(0)))_{q \geq 1}$ are *not independent* since $H_{\mathcal{S}}(0)$, the hitting time of 0 of the process $(\mathcal{S}_t)_{t \geq 0}$, depends sensitively on all of the indicator processes.

PROOF. We shall use a randomization trick. Indeed, let $(N_j^{(n)}(t))_{t \geq 0}$ be a sequence of independent Poisson processes with rate w_j/ℓ_n. Let

$$T_j = \inf\{t : N(t) = j\} \qquad \text{where} \qquad N(t) = \sum_{j \in [n]} N_j^{(n)}(t). \tag{3.63}$$

Then $t \mapsto N(t)$ is a rate 1 Poisson process, and we have that [recall (3.11)]

$$S_l = S'_{T_l}, \tag{3.64}$$

where the continuous-time process $(S'_t)_{t\geq 0}$ is defined by

$$\begin{aligned} S'_t = w_1(\lambda) - \frac{w_1 w_1(\lambda) N(t)}{\ell_n} + \sum_{i=2}^{n} w_i(\lambda)\left[\mathbb{1}_{\{N_i^{(n)}(t)\geq 1\}} - \frac{w_i N(t)}{\ell_n}\right] \\ + \big(\nu_n(\lambda) - 1\big)N(t). \end{aligned} \tag{3.65}$$

By construction, the processes $(\mathbb{1}_{\{N_q^{(n)}(n^\rho t)\geq 1\}})_{t\geq 0}$ are *independent*, and are characterized by the birth times

$$E_q^{(n)} = \inf\{t : N_q^{(n)}(n^\rho t) \geq 1\}. \tag{3.66}$$

Again by construction, these birth times are independent for different $q \geq 2$, and $E_q^{(n)}$ has an exponential distribution with parameter $n^\rho w_q/\ell_n$. The parameters of these exponential random variables converge to

$$n^\rho w_q/\ell_n \to a q^{-\alpha}, \tag{3.67}$$

where $a = c_F^\alpha/\mathbb{E}[W]$, and which are the parameters of the limiting exponential random variables in terms of which we can identify $\mathcal{I}_q(t) = \mathbb{1}_{\{N_q(t)\geq 1\}} = \mathbb{1}_{\{\mathrm{Exp}(aq^{-\alpha})\leq t\}}$; see (2.48). By the convergence of the parameters, we can *couple* $E_q^{(n)}$ with $E_q = \mathrm{Exp}(aq^{-\alpha})$ in such a way that, for every $q \geq 2$ fixed,

$$\mathbb{P}\big(E_q^{(n)} \neq E_q\big) = o(1). \tag{3.68}$$

Indeed, (3.68) follows by noting that, by (3.67), the density of $E_q^{(n)}$ converges pointwise to that of E_q, which, by [32], (7.3), implies that we can couple $(E_q^{(n)})_{n\geq 1}$ to E_q in such a way that (3.68) holds.

Equation (3.68), jointly with the *independence* of $(E_q^{(n)})_{n\geq 1}$ for different q's, immediately implies that, for each $K \geq 1$,

$$\mathbb{P}\big(\mathbb{1}_{\{N_q^{(n)}(n^\rho t)\geq 1\}} = \mathcal{I}_q(t)\ \forall t \geq 0, q \in [K]\big) = 1 - o(1), \tag{3.69}$$

so that we have also, *whp*, perfectly coupled the entire processes

$$\big(\mathbb{1}_{\{N_q^{(n)}(n^\rho t)\geq 1\}}\big)_{t\geq 0, q\in[K]} \quad \text{and} \quad (\mathcal{I}_q(t))_{t\geq 0, q\in[K]}.$$

In particular, this implies that, for every $K \geq 2$,

$$\mathbb{P}\big(\mathbb{1}_{\{N_q^{(n)}(T_l)\geq 1\}} = \mathcal{I}_q(T_l)\ \forall l \geq 1, q \in [K]\big) = 1 - o(1) \tag{3.70}$$

and, by construction, $\mathbb{1}_{\{N_q^{(n)}(T_l)\geq 1\}} = \mathcal{I}_q^{(n)}(l)$.

Applying the perfect coupling to $l = V(1)$, for which $\mathbb{1}_{\{N_q^{(n)}(T_l)\geq 1\}} = \mathbb{1}_{\{q\in\mathcal{C}(1)\}}$, this provides a perfect coupling between $\mathbb{1}_{\{q\in\mathcal{C}(1)\}}$ and $\mathcal{I}_q(T_{V(1)})$. We then note that

$$n^{-\rho}|\mathcal{C}(1)| \xrightarrow{d} H_{\mathcal{S}}(0) \tag{3.71}$$

and, since T_j is the birth time of the jth individual in a rate 1 Poisson process,

$$\sup_{t\leq u}|n^{-\rho}T_{tn^\rho} - t| \xrightarrow{\mathbb{P}} 0, \tag{3.72}$$

where, for noninteger tn^ρ, we recall the convention below (2.36).

Weak convergence of $(\mathbb{1}_{\{q\in\mathcal{C}(1)\}})_{q\geq 1}$ in the product topology is equivalent to the weak convergence of $(\mathbb{1}_{\{q\in\mathcal{C}(1)\}})_{q\in[m]}$ for any $m \geq 1$; see [28], Theorem 4.29. Therefore, together with the exact coupling in (3.70), this completes the proof of (3.62), since the processes $(\mathcal{I}_i(t))_{t\geq 0}$ have a.s. no jump close to $H_{\mathcal{S}}(0)$.

We continue to show the properties of the limiting variables. The random variable $H_{\mathcal{S}}(0)$ is nondegenerate, since its distribution does not have any atoms. We shall next show that $\mathbb{1}_{\{q\in\mathcal{C}(1)\}}$ is nontrivial. We shall show this only for $q = 2$, the proof for $q > 2$ being identical. For this, we use the fact that

$$\begin{aligned}\lim_{n\to\infty} \mathbb{P}\big(2 \in \mathcal{C}(1)\big) &\geq \mathbb{P}\big(H_{\mathcal{S}}(0) \geq \varepsilon, \mathcal{I}_2(\varepsilon) = 1\big)\\ &\geq \mathbb{P}\big(H_{\mathcal{S}}(0) \geq \varepsilon\big)\mathbb{P}\big(\mathcal{I}_2(\varepsilon) = 1\big) > 0\end{aligned} \tag{3.73}$$

by the Fortuin–Kasteleyn–Ginibre (FKG) inequality (see [20], Theorem 2.4) and the fact that both random variables $H_{\mathcal{S}}(0)$ and $\mathcal{I}_2(\varepsilon)$ are monotone in the independent exponential random variables that describe the first hit of q for all $q \geq 1$, so that both $\{H_{\mathcal{S}}(0) \geq \varepsilon\}$ and $\mathcal{I}_2(\varepsilon) = 1$ are increasing events.

Further,

$$\begin{aligned}\lim_{n\to\infty} \mathbb{P}\big(2 \notin \mathcal{C}(1)\big) &\geq \mathbb{P}\big(H_{\mathcal{S}}(0) \leq K, \mathcal{I}_2(K) = 0\big)\\ &\geq \mathbb{P}\big(H_{\mathcal{S}}(0) \leq K\big)\mathbb{P}\big(\mathcal{I}_2(K) = 0\big),\end{aligned} \tag{3.74}$$

again by FKG, now using that both $\{H_{\mathcal{S}}(0) \leq K\}$ and $\{\mathcal{I}_2(K) = 0\}$ are decreasing events. Thus

$$\lim_{n\to\infty} \mathbb{P}\big(2 \notin \mathcal{C}(1)\big) \geq \mathbb{P}\big(H_{\mathcal{S}}(0) \leq K\big)\mathbb{P}\big(\mathcal{I}_2(K) = 0\big) > 0, \tag{3.75}$$

which proves the claim. □

REMARK 3.8 (Convergence in the uniform topology). In fact, by the proof of Proposition 3.7, we even obtain that the weak convergence in Theorem 2.1 holds in the *uniform topology*. Indeed, the coupling obtained in the proof of Proposition 3.7 [see in particular (3.69)] shows that we can couple $(\mathcal{S}_t^{(n,K)})_{t\geq 0}$ and $(\mathcal{S}_t^{(\infty,K)})_{t\geq 0}$ such that these processes are *whp* equal *for all* $t \geq 0$. By (3.26) in Corollary 3.2,

$(\mathcal{Z}_t^{(n)})_{t\geq 0}$ is close to $(\mathcal{S}_t^{(n,K)})_{t\geq 0}$ in the uniform topology on $[0,T]$, while (3.38) shows that $(\mathcal{S}_t^{(\infty,K)})_{t\geq 0}$ is uniformly close to $(\mathcal{S}_t)_{t\geq 0}$. This proves the convergence in the uniform topology.

REMARK 3.9 (Convergence of cluster size of vertex i). We next remark on the scaling limits of $|\mathcal{C}(i)|$ and $|\mathcal{C}_{\leq}(i)|$. As in (2.39), define

$$\mathcal{S}_t^{(i)} = b - abti^{-\alpha} + ct + \sum_{j=1:\, j\neq i}^{\infty} bj^{-\alpha}[\mathcal{I}_j(t) - atj^{-\alpha}], \tag{3.76}$$

so that $(\mathcal{S}_t)_{t\geq 0} = (\mathcal{S}_t^{(i)})_{t\geq 0}$. Define

$$H^{(i)}(0) = \inf\{t : \mathcal{S}_t^{(i)} = 0\} \tag{3.77}$$

to be the first hitting time of zero of the process $(\mathcal{S}_t^{(i)})_{t\geq 0}$. Then, in an identical way as in the proof of Proposition 3.7, it follows that, as $n\to\infty$,

$$\bigl(n^{-\rho}|\mathcal{C}(i)|, (\mathbb{1}_{\{q\in\mathcal{C}(i)\}})_{q\geq 1}\bigr) \xrightarrow{d} \bigl(H^{(i)}(0), (\mathcal{I}_q(H^{(i)}(0)))_{q\geq 1}\bigr) \tag{3.78}$$

in the product topology. As a result,

$$n^{-\rho}|\mathcal{C}_{\leq}(i)| \xrightarrow{d} H^{(i)}(0)\prod_{j=1}^{i-1}\bigl(1 - \mathcal{I}_q(H^{(i)}(0))\bigr). \tag{3.79}$$

4. Convergence of multiple clusters. In this section, we extend the analysis of one cluster in Section 2 to multiple clusters. This sets the stage for the proof of Theorem 1.1, which is completed in the next section. The main result is as follows:

THEOREM 4.1 (Weak convergence of clusters of first vertices). *Fix the Norros–Reittu random graph with weights* $\mathbf{w}(\lambda)$ *defined in* (1.15). *Assume that* $\nu = 1$ *and that* (1.6) *holds. Then, for all* $\lambda\in\mathbb{R}$,

$$(n^{-\rho}|\mathcal{C}_{\leq}(i)|)_{i\geq 1} \xrightarrow{d} (H_i(0))_{i\geq 1} \tag{4.1}$$

for some nondegenerate limit $(H_i(0))_{i\geq 1}$.

In the remainder of this section, we shall prove Theorem 4.1 and use it to complete the proof of Theorem 1.1. We let $I_1^{(n)} = 1$, and let

$$I_2^{(n)} = \min[n]\setminus\mathcal{C}(1) \tag{4.2}$$

be the minimal element that is not part of $\mathcal{C}(1)$, where, for a set of indices $A\subseteq[n]$, we let $\min A$ denote the minimal element of A. To extend the above definitions further, we define, recursively,

$$\mathcal{D}_i^{(n)} = \mathcal{C}_{\leq}(I_i^{(n)}) \quad\text{and}\quad \mathcal{D}_{\leq i}^{(n)} = \bigcup_{j\leq i}\mathcal{D}_j^{(n)}. \tag{4.3}$$

Then we define $I_{i+1}^{(n)}$ by

$$I_{i+1}^{(n)} = \min[n] \setminus \mathcal{D}_{\leq i}^{(n)}, \tag{4.4}$$

which is the vertex with the smallest index of which we have not yet explored its cluster.

Obviously, $|\mathcal{C}_{\leq}(i)| = 0$ unless $i = I_j^{(n)}$ for some j. This prompts us to investigate the weak convergence of $n^{-\rho}|\mathcal{D}_i^{(n)}|$. This will be done by induction on i. The *induction hypothesis* is that

$$\big(n^{-\rho}|\mathcal{D}_j^{(n)}|, (\mathbb{1}_{\{q \in \mathcal{D}_{\leq j}^{(n)}\}})_{q \geq 1}\big)_{j \in [i]} \xrightarrow{d} \big(H_j(0), (\mathbb{1}_{\{q \in \mathcal{D}_{\leq j}\}})_{q \geq 1}\big)_{j \in [i]} \tag{4.5}$$

in the product topology, for some limiting random variables. Part of the induction hypothesis is that these limiting random variables satisfy the following facts: (1) the limiting random variables $(H_j(0))_{j \in [i]}$ are *nondegenerate*, in the sense that the essential support of the random vector $(H_j(0))_{j \in [i]}$ is i-dimensional, and (2) the random indicators $(\mathbb{1}_{\{q \in \mathcal{D}_{\leq j}\}})_{j \in [i], q > i}$ are all *nontrivial*, in the sense that they take the values zero and one, each with positive probability. By construction, $\mathbb{1}_{\{q \in \mathcal{D}_{\leq j}^{(n)}\}} = 1$ for $q \leq i$, so the restriction to $q > i$ in condition (2) is the most we can hope for.

We shall start by initializing the induction hypothesis for $j = 1$, which follows from Proposition 3.7, as we show now. Indeed, we have that $\mathcal{D}_1^{(n)} = \mathcal{D}_{\leq 1}^{(n)} = \mathcal{C}(1)$, so that (4.5) is identical to the statement in Proposition 3.7.

We next advance the induction hypothesis by verifying that (4.5) also holds for $j = i + 1$. We first intuitively explain our approach. The random variable $H_{i+1}(0)$ shall be the weak limit of $n^{-\rho}|\mathcal{D}_{i+1}^{(n)}|$. We shall show that $H_{i+1}(0)$ is the hitting time of zero of a process similar to $(\mathcal{S}_t)_{t \geq 0}$ in Section 2. We now start by explaining how this process arises.

Assume that the induction hypothesis (4.5) holds for i. By (4.5), the index set $\mathcal{D}_{\leq i}$ is the (random) set of indices for which

$$\big(\mathbb{1}_{\{q \in \mathcal{D}_{\leq i}^{(n)}\}}\big)_{q \geq 1} \xrightarrow{d} \big(\mathbb{1}_{\{q \in \mathcal{D}_{\leq i}\}}\big)_{q \geq 1}. \tag{4.6}$$

Then, we note that, by (4.5), we have that

$$I_{i+1}^{(n)} \equiv \min\{q : \mathbb{1}_{\{q \in \mathcal{D}_{\leq i}^{(n)}\}} = 0\} \quad \xrightarrow{d} \quad I_{i+1} \equiv \min\{q : \mathbb{1}_{\{q \in \mathcal{D}_{\leq i}\}} = 0\}, \tag{4.7}$$

and we see that $I_{i+1}^{(n)}$ and I_{i+1} are *deterministic* functions of the sets $\mathcal{D}_{\leq i}^{(n)}$ and $\mathcal{D}_{\leq i}$, respectively. The random variable I_{i+1} is *finite*, since, for $K, Q \geq 1$ large,

$$\mathbb{P}(I_{i+1}^{(n)} > K) \leq \mathbb{P}(|\mathcal{D}_{\leq j}^{(n)}| \geq Qn^{\rho}) + \mathbb{P}(I_{i+1}^{(n)} \geq K, |\mathcal{D}_{\leq j}^{(n)}| < Qn^{\rho}). \tag{4.8}$$

The first probability converges, by (4.5) and the continuous-mapping theorem, to $\mathbb{P}(H_1(0) + \cdots + H_i(0) \geq Q)$, which is small for $Q \geq 1$ large. For the second probability in (4.8), and for $i \leq K/2$, we can bound

$$\begin{aligned}\mathbb{P}\big(I_{i+1}^{(n)} > K, \big|\mathcal{D}_{\leq i}^{(n)}\big| < Qn^{\rho}\big) &\leq \mathbb{P}(\text{vertex } K \text{ drawn in } Qn^{\rho} \text{ vertex checks})\\ &\leq \mathbb{P}(\exists l \leq Qn^{\rho} : M_l = K) \leq \sum_{l=1}^{Qn^{\rho}} \frac{w_K}{\ell_n}\\ &\leq CQK^{-\alpha},\end{aligned} \tag{4.9}$$

which converges to zero as $K \to \infty$ when $Q = K^{\beta}$ with $\beta < \alpha$. As a result, we have that

$$\mathbb{P}(I_{i+1} > K) = \lim_{n\to\infty} \mathbb{P}\big(I_{i+1}^{(n)} > K\big) \tag{4.10}$$

is small for K large.

We conclude that, from the induction hypothesis in (4.5), we obtain the *joint* convergence

$$\begin{aligned}&\big(\big(n^{-\rho}\big|\mathcal{D}_i^{(n)}\big|, \big(\mathbb{1}_{\{q\in\mathcal{D}_{\leq j}^{(n)}\}}\big)_{q\geq 1}\big)_{j\in[i]}, I_{i+1}^{(n)}\big)\\ &\quad \xrightarrow{d} \big(\big(H_j(0), \big(\mathbb{1}_{\{q\in\mathcal{D}_{\leq j}\}}\big)_{q\geq 1}\big)_{j\in[i]}, I_{i+1}\big).\end{aligned} \tag{4.11}$$

We now start exploring the cluster of $I_{i+1}^{(n)}$, and we need to show that this cluster size, as well as the indices in it, converge in distribution. More precisely, the joint convergence in (4.5) for $i+1$ (and thus the advancement of the induction hypothesis) follows when we prove that, conditionally on $\mathcal{D}_{\leq i}^{(n)}$,

$$\begin{aligned}&\big(n^{-\rho}\big|\mathcal{D}_{i+1}^{(n)}\big|, I_{i+1}^{(n)}, \big(\mathbb{1}_{\{q\in\mathcal{D}_{\leq i+1}^{(n)}\}}\big)_{q\geq 1}\big)\\ &\quad \xrightarrow{d} \big(H_{i+1}(0), I_{i+1}, \big(\mathbb{1}_{\{q\in\mathcal{D}_{\leq i+1}\}}\big)_{q\geq 1}\big).\end{aligned} \tag{4.12}$$

To prove (4.12), we follow the approach in Section 2 as closely as possible. A crucial observation is that after the exploration of $\mathcal{D}_{\leq i}^{(n)}$ and conditionally on it, the remaining graph is again a rank-1 inhomogeneous random graph, with (a) vertex set $[n] \setminus \mathcal{D}_{\leq i}^{(n)}$, and (b) edge probabilities, for $u, v \in [n] \setminus \mathcal{D}_{\leq i}^{(n)}$, given by $p_{uv} = 1 - \mathrm{e}^{-w_u w_v/\ell_n}$.

We now extend the exploration process of clusters described in Section 2.1 to the setting above. As in Section 2, we set $Z_0(i) = 1$ and let $Z_1(i)$ denote the number of neighbors of the vertex $I_{i+1}^{(n)}$ outside $\mathcal{D}_{\leq i}^{(n)}$, that is,

$$Z_1(i) = \sum_{j\notin\mathcal{D}_{\leq i}^{(n)}} \mathrm{Poi}\big(w_{I_{i+1}^{(n)}}(\lambda)w_j/\ell_n\big) = \mathrm{Poi}\big(w_{I_{i+1}^{(n)}}(\lambda)\ell_n(i)/\ell_n\big), \tag{4.13}$$

where we let

$$\ell_n(i) = \sum_{j \notin \mathcal{D}_{\leq i}^{(n)}} w_j \tag{4.14}$$

be the total weight of vertices *outside* $\mathcal{D}_{\leq i}^{(n)}$. For $l \geq 2$, $(Z_l(i))_{l\geq 1}$ satisfies the recursion relation

$$Z_l(i) = Z_{l-1}(i) + X_l(i) - 1, \tag{4.15}$$

where $X_l(i)$ denotes the number of potential neighbors outside of $\mathcal{D}_{\leq i}^{(n)}$ of the lth vertex which is explored. As explained in more detail in Section 2, the distribution of $X_l(i)$ (for $2 \leq l \leq n$) is equal to $\mathrm{Poi}(w_{M_l(i)}\ell_n(i)/\ell_n)J_l(i)$, where now the marks $(M_l(i))_{l=1}^{\infty}$ are i.i.d. random variables with distribution $M(i)$ given by

$$\mathbb{P}\big(M(i) = m\big) = w_m/\ell_n(i), \qquad m \in [n] \setminus \mathcal{D}_{\leq i}^{(n)}, \tag{4.16}$$

and

$$J_l(i) = \mathbb{1}_{\{M_l(i) \notin \{I_{i+1}^{(n)}\} \cup \{M_2(i),\dots,M_{l-1}(i)\}\}} \tag{4.17}$$

is the indicator that the mark $M_l(i)$ has not been found up to time l and is not equal to vertex $I_{i+1}^{(n)}$.

Then the number of vertex checks $V(I_{i+1}^{(n)})$ in the exploration of $\mathcal{D}_{i+1}^{(n)} = \mathcal{C}_{\leq}(I_{i+1}^{(n)})$ equals

$$V\big(I_{i+1}^{(n)}\big) = \inf\{l : Z_l(i) = 0\} \tag{4.18}$$

and

$$\big(\mathbb{1}_{\{q \in \mathcal{D}_{i+1}^{(n)}\}}\big)_{q \neq I_{i+1}^{(n)}} = (\mathbb{1}_{\{\exists l \leq |\mathcal{D}_{i+1}^{(n)}| : M_l(i) = q\}})_{q \neq I_{i+1}^{(n)}}, \tag{4.19}$$

while $\mathbb{1}_{\{I_{i+1}^{(n)} \in \mathcal{D}_{i+1}^{(n)}\}} = 1$. We again note that

$$\big|\mathcal{D}_{i+1}^{(n)}\big| = \big|\mathcal{C}_{\leq}\big(I_{i+1}^{(n)}\big)\big| \leq V\big(I_{i+1}^{(n)}\big), \tag{4.20}$$

while

$$n^{-\rho}\big[V\big(I_{i+1}^{(n)}\big) - \big|\mathcal{D}_{i+1}^{(n)}\big|\big] \xrightarrow{\mathbb{P}} 0, \tag{4.21}$$

which can be proved along the lines of the proof of Lemma 3.6. This gives us a convenient description of all the random variables needed to advance the induction hypothesis.

In order to prove the weak convergence of $n^{-\rho}V(I_{i+1}^{(n)})$, we again investigate the scaling limit of the process $(Z_l(i))_{l\geq 0}$. For this, we define $S_0(i) = 1$, $S_1(i) = w_{I_{i+1}^{(n)}}(\lambda)\ell_n(i)/\ell_n$ and, for $l \geq 2$,

$$S_l(i) = S_{l-1}(i) + w_{M_l(i)}(\lambda)J_l(i)\ell_n(i)/\ell_n - 1. \tag{4.22}$$

Then, as in Lemma 3.1, it is easy to show that, conditionally on $\mathcal{D}_{\le i}^{(n)}$, the processes $(S_l(i))_{l\ge 0}$ and $(Z_l(i))_{l\ge 0}$ are uniformly close. Denote by $\mathcal{B}_i^{(n)} = \mathcal{D}_{\le i}^{(n)} \cup \{I_{i+1}^{(n)}\}$ the union of all vertices explored in the first i clusters and the minimal element not in the first i clusters. We rewrite

$$
\begin{aligned}
S_l(i) &= w_{I_{i+1}^{(n)}}(\lambda)\frac{\ell_n(i)}{\ell_n} + \sum_{j=2}^{l} w_{M_j(i)}\frac{\ell_n(i)}{\ell_n} J_j(i) - (l-1) \\
&= w_{I_{i+1}^{(n)}}(\lambda)\frac{\ell_n(i)}{\ell_n} + \sum_{q\in[n]\setminus\mathcal{B}_i^{(n)}} w_q(\lambda)\frac{\ell_n(i)}{\ell_n}\mathcal{I}_q^{(n)}(l;i) - (l-1),
\end{aligned} \tag{4.23}
$$

where

$$
\mathcal{I}_q^{(n)}(l;i) = \mathbb{1}_{\{\exists j\le l:\, M_j(i)=q\}}. \tag{4.24}
$$

We further rewrite the above as

$$
\begin{aligned}
S_l(i) &= w_{I_{i+1}^{(n)}}(\lambda)\frac{\ell_n(i)}{\ell_n} + \sum_{q\in[n]\setminus\mathcal{B}_i^{(n)}} w_q(\lambda)\frac{\ell_n(i)}{\ell_n}\biggl(\mathcal{I}_q^{(n)}(l;i) - \frac{l w_q}{\ell_n(i)}\biggr) \\
&\quad + l\biggl(\sum_{q\in[n]\setminus\mathcal{B}_i^{(n)}} \frac{w_q(\lambda) w_q}{\ell_n} - 1\biggr) + 1.
\end{aligned} \tag{4.25}
$$

We note that we can rewrite the last sum, using (2.30), as

$$
\begin{aligned}
(1+\lambda n^{-\eta}) \sum_{q\in[n]\setminus\mathcal{B}_i^{(n)}} \frac{w_q^2}{\ell_n} - 1 &= \bigl(\nu_n(\lambda) - 1\bigr) - (1+\lambda n^{-\eta}) \sum_{q\in\mathcal{B}_i^{(n)}} \frac{w_q^2}{\ell_n} \\
&= \theta n^{-\eta} - \sum_{q\in\mathcal{B}_i^{(n)}} \frac{w_q^2}{\ell_n} + o(n^{-\eta}).
\end{aligned} \tag{4.26}
$$

In turn, the sum can be approximated by

$$
\sum_{q\in\mathcal{B}_i^{(n)}} \frac{w_q^2}{\ell_n} = d n^{-\eta} \sum_{q\in\mathcal{B}_i^{(n)}} q^{-2\alpha}\bigl(1 + o_{\mathbb{P}}(1)\bigr), \tag{4.27}
$$

where $d = c_F^{2\alpha}/\mathbb{E}[W]$. Denoting

$$
D_i^{(n)} = d \sum_{q\in\mathcal{B}_i^{(n)}} q^{-2\alpha}, \tag{4.28}
$$

we therefore have that

$$
\begin{aligned}
S_l(i) &= w_{I_{i+1}^{(n)}}\frac{\ell_n(i)}{\ell_n} + \sum_{q\in[n]\setminus\mathcal{B}_i^{(n)}} w_q\frac{\ell_n(i)}{\ell_n}\biggl(\mathcal{I}_q^{(n)}(l;i) - \frac{l w_q}{\ell_n(i)}\biggr) \\
&\quad + l(\theta - D_i^{(n)}) n^{-\eta} + o_{\mathbb{P}}(l n^{-\eta}).
\end{aligned} \tag{4.29}
$$

We conclude that we arrive at a similar process as when exploring $\mathcal{C}(1)$, apart from the fact that: (i) fewer vertices are allowed to participate, (ii) a negative drift $-D_i^{(n)}$ is introduced and (iii) a factor $\frac{\ell_n(i)}{\ell_n} = 1 + o_{\mathbb{P}}(1)$ is introduced.

We proceed by investigating the convergence of $D_i^{(n)}$:

LEMMA 4.2 (Weak convergence of random drift). *As $n \to \infty$, and assuming* (4.11),

$$D_i^{(n)} \xrightarrow{d} D_i \equiv \sum_{q \in \mathcal{D}_{\leq i} \cup \{I_{i+1}\}} q^{-2\alpha}, \tag{4.30}$$

where $(\mathcal{D}_{\leq i}, I_{i+1})$ is the weak limit of $(\mathcal{D}_{\leq i}^{(n)}, I_{i+1}^{(n)})$ given in (4.11).

PROOF. We start by bounding $\mathbb{P}(q \in \mathcal{D}_{\leq i}^{(n)})$, for $q > 0$ large. We shall first prove that the probability that $|\mathcal{D}_{\leq i}^{(n)}| \leq n^\rho K$ is $1 - o(1)$ when $K > 0$ grows large. Indeed, by [34], Theorem 1.2, we have that, $|\mathcal{C}_{\max}| = \max_i |\mathcal{C}_{\leq i}| \leq \omega n^\rho$ with probability $1 - o(1)$, as $\omega \to \infty$. Thus, $|\mathcal{D}_{\leq i}^{(n)}| \leq n^\rho (i\omega) = n^\rho K$, with probability $1 - o(1)$ as $K \to \infty$, when we take $K = \omega i$. Denoting

$$\mathcal{E}_{i,K}^{(n)} = \{|\mathcal{D}_{\leq i}^{(n)}| \leq n^\rho K\}, \tag{4.31}$$

we have that

$$\mathbb{P}(\{q \in \mathcal{D}_{\leq i}^{(n)} \setminus \{I_j^{(n)}\}_{j=1}^i\} \cap \mathcal{E}_{i,K}^{(n)}) \leq n^\rho K \frac{w_q}{\sum_{j > K n^\rho} w_j}, \tag{4.32}$$

since, independently of the choices before, the probability of drawing q is at most $w_q / \sum_{j > K n^\rho} w_j$. Now,

$$\sum_{j > K n^\rho} w_j = \ell_n\big(1 + o(1)\big) = \mathbb{E}[W] n\big(1 + o(1)\big). \tag{4.33}$$

Thus, for some $C > 0$,

$$\mathbb{P}(\{q \in \mathcal{D}_{\leq i}^{(n)} \setminus \{I_j^{(n)}\}_{j=1}^i\} \cap \mathcal{E}_{i,K}^{(n)}) \leq C K q^{-\alpha}, \tag{4.34}$$

so that

$$\mathbb{E}\Bigg[\sum_{q \in \mathcal{B}_i^{(n)} : q > Q} q^{-2\alpha} \mathbb{1}_{\mathcal{E}_{i,K}^{(n)}}\Bigg] \leq i Q^{-2\alpha} + C K Q^{1-3\alpha}, \tag{4.35}$$

where the first contribution arises from the (at most i) values of $q = I_j^{(n)}$ for $j \in [i+1]$ for which $I_j^{(n)} > Q$, and the second contribution from the $q \notin \{I_j^{(n)}\}_{j \in [i+1]}$.

Equation (4.35) implies that the weak convergence of $D_i^{(n)}$ follows from the weak convergence of

$$\sum_{q \in \mathcal{B}_i^{(n)} : q \le Q} q^{-2\alpha}, \tag{4.36}$$

which, in turn, follows from (4.11) and the continuous mapping theorem. □

Now we are ready to complete the proof of Theorem 4.1.

PROOF OF THEOREM 4.1. We start by setting the stage for the weak convergence of processes needed to advance the induction hypothesis as formulated in (4.12). Define

$$\mathcal{Z}_t^{(n)}(i) = n^{-\alpha} Z_{tn^\rho}(i), \qquad \mathcal{S}_t^{(n)}(i) = n^{-\rho} S_{tn^\rho}(i) \tag{4.37}$$

and

$$\mathcal{S}_t(i) = bI_{i+1}^{-\alpha} + \sum_{q \in \mathcal{D}_{\le i} \cup \{I_{i+1}\}} aq^{-\alpha}\big(\mathcal{I}_q(t) - btq^{-\alpha}\big) + t(c - D_i). \tag{4.38}$$

Then, using Lemma 4.2, the proof of Theorem 2.1 can easily be adapted to prove that $n^{-\rho}|\mathcal{D}_{i+1}^{(n)}| \xrightarrow{d} H_{i+1}(0)$, where $H_{i+1}(0)$ is the hitting time of 0 of $(\mathcal{S}_t(i))_{t\ge 0}$, and where a, b, c are given by $a = c_F^\alpha/\mathbb{E}[W]$, $b = c_F^\alpha$ and $c = \theta$.

Indeed, in more detail, we shall work *conditionally* on $\mathcal{D}_{\le i}^{(n)}$. The proof of Theorem 2.1 reveals that the main contribution to $(\mathcal{S}_t(i))_{t\ge 0}$ and $(\mathcal{S}_t^{(n)}(i))_{t\ge 0}$ arises from the vertices $q \in [K]$. Now, since $(\mathbb{1}_{\{a \in \mathcal{D}_{\le i}^{(n)}\}})_{a\in[K]}$ is a sequence of *discrete* random variables taking a finite number of outcomes and that converge in distribution, we have that its probability mass function converges pointwise. By [32], (6.3) on page 16, this implies that we can *couple* $(\mathbb{1}_{\{a \in \mathcal{D}_{\le i+1}^{(n)}\}})_{a\in[K]}$ to $(\mathbb{1}_{\{a \in \mathcal{D}_{\le i+1}\}})_{a\in[K]}$ in such a way that

$$\mathbb{P}\big((\mathbb{1}_{\{a \in \mathcal{D}_{\le i+1}^{(n)}\}})_{a\in[K]} \ne (\mathbb{1}_{\{a \in \mathcal{D}_{\le i+1}\}})_{a\in[K]}\big) = o(1). \tag{4.39}$$

Therefore, *whp*, there is a perfect coupling between the elements of $\mathcal{D}_{\le i+1}^{(n)} \cap [K]$ and $\mathcal{D}_{\le i+1} \cap [K]$. When this is the case, we can basically think of the set of summands in (4.25) as being deterministic and follow the proof of Theorem 2.1 verbatim.

Further, the proof of Proposition 3.7 can be adapted to prove the joint convergence of

$$\big(n^{-\rho}|\mathcal{D}_{i+1}^{(n)}|, (\mathbb{1}_{\{q \in \mathcal{D}_{i+1}^{(n)}\}})_{q\ge 1}\big) \xrightarrow{d} (H_{i+1}(0), (\mathcal{I}_q(H_{i+1}(0)))_{q\ge 1}). \tag{4.40}$$

Together with the induction hypothesis, this proves that (4.5) also holds for all $j \le i+1$, and, thus, we have advanced the induction hypothesis. This, in particular, proves Theorem 4.1. The proof for cluster weights follows in an identical way as the convergence proof of $n^{-\rho}\mathcal{W}(1)$ in the proof of Theorem 2.1. □

5. Proofs of Theorems 1.1, 1.5 and 1.6. In this section, we prove Theorems 1.1, 1.5 and 1.6 using the results in Theorems 2.1 and 4.1, as well as Proposition 3.7. We start with a proof of Theorem 1.6, followed by those of Theorems 1.5 and 1.1. Note that, combining parts (a) and (b) in Theorem 1.6, we obtain that, with high probability as K becomes large, the largest m clusters are all among the first $(|\mathcal{C}_{\le}(i)|)_{i\in[K]}$. This explains why we start the cluster exploration from the vertices with the highest weights.

PROOF OF THEOREM 1.6. (a) For $\max_{i\ge K}|\mathcal{C}_{\le}(i)|\ge \varepsilon n^{\rho}$ to occur, we must have that there exists a cluster using the vertices in $[n]\setminus[K]$ such that (1) $|\mathcal{C}_{\le}(i)|\ge \varepsilon n^{\rho}$, and (2) the cluster $\mathcal{C}_{\le}(i)$ is not connected to any of the vertices in $[K]$.

By construction, the graph restricted to the vertices in $[n]\setminus[K]$ is again a Norros–Reittu model, with edge probabilities $p_{ij}=1-\mathrm{e}^{-w_iw_j/\ell_n}$, for all $i,j\in[n]\setminus[K]$. However, no vertex in $[n]\setminus[K]$ found to be in the cluster $\mathcal{C}(i)$ is allowed to have an edge to any of the vertices in $[K]$. We shall now bound this probability, making use of the results in [34].

With

$$Z_{\ge k}^{[K]}=\sum_{v=1}^{n}\mathbb{1}_{\{|\mathcal{C}(v)|\ge k,\mathcal{C}(v)\cap[K]=\varnothing\}}, \tag{5.1}$$

we have

$$\begin{aligned}\mathbb{P}\Bigl(\max_{i\ge K}|\mathcal{C}_{\le}(i)|\ge k\Bigr)&=\mathbb{P}\bigl(Z_{\ge k}^{[K]}\ge k\bigr)\le\frac{\mathbb{E}[Z_{\ge k}^{[K]}]}{k}\\&=\frac{1}{k}\sum_{v=K+1}^{n}\mathbb{P}\bigl(|\mathcal{C}(v)|\ge k,\mathcal{C}(v)\cap[K]=\varnothing\bigr).\end{aligned}\tag{5.2}$$

Denote by $\mathcal{C}^{[K]}(v)$ the cluster of v restricted to the vertices $[n]\setminus[K]$. Then, due to the independence of disjoint sets of edges, and the fact that $\mathcal{C}(v)\cap[K]=\varnothing$ only depends on edges between $[K]$ and $[n]\setminus[K]$, while $|\mathcal{C}^{[K]}(v)|\ge k$ depends only on edges between pairs of vertices in $[n]\setminus[K]$, we obtain

$$\begin{aligned}&\mathbb{P}\bigl(|\mathcal{C}(v)|\ge k,\mathcal{C}(v)\cap[K]=\varnothing\bigr)\\&\quad=\mathbb{E}\bigl[\mathbb{P}\bigl(\mathcal{C}(v)\cap[K]=\varnothing\mid\mathcal{C}^{[K]}(v)\bigr)\mathbb{1}_{\{|\mathcal{C}^{[K]}(v)|\ge k\}}\bigr]\\&\quad=\mathbb{E}\bigl[\mathrm{e}^{-\mathcal{W}_{[K]}\mathcal{W}^{[K]}(v)/\ell_n}\mathbb{1}_{\{|\mathcal{C}^{[K]}(v)|\ge k\}}\bigr],\end{aligned}\tag{5.3}$$

where, similarly to (1.17), we define

$$\mathcal{W}^{[K]}(v)=\sum_{a\in\mathcal{C}^{[K]}(v)}w_a\quad\text{and}\quad\mathcal{W}_{[K]}=\sum_{j=1}^{K}w_j.\tag{5.4}$$

We split depending on whether $\mathcal{W}^{[K]}(v) \geq k/2$ or not, to obtain

$$\mathbb{P}\Big(\max_{i\geq K} |\mathcal{C}_{\leq}(i)| \geq k\Big) \leq \frac{1}{k} \sum_{v=K+1}^{n} \mathrm{e}^{-\mathcal{W}_{[K]}k/(2\ell_n)} \mathbb{P}\big(\big|\mathcal{C}^{[K]}(v)\big| \geq k\big) \tag{5.5}$$

$$+ \frac{1}{k} \sum_{v=K+1}^{n} \mathbb{P}\big(\big|\mathcal{C}^{[K]}(v)\big| \geq k, \mathcal{W}^{[K]}(v) \leq k/2\big). \tag{5.6}$$

For the first term we compute that, for some $C > 0$,

$$\mathcal{W}_{[K]} \geq c_F \sum_{j=1}^{K} (n/j)^{\alpha}\big(1+o(1)\big) \geq C n^{\alpha} K^{\rho}. \tag{5.7}$$

Thus, when $k = k_n = \varepsilon n^{\rho}$, we obtain, for some $u > 0$, and using $\alpha + \rho = 1$ [see (2.10)],

$$\begin{aligned}
&\frac{1}{k_n} \mathrm{e}^{-\mathcal{W}_{[K]}k_n/(2\ell_n)} \sum_{v\in[n]} \mathbb{P}\big(\big|\mathcal{C}^{[K]}(v)\big| \geq k_n\big) \\
&\quad \leq \frac{1}{k_n} \mathrm{e}^{-u\varepsilon K^{\rho}} \sum_{v\in[n]} \mathbb{P}\big(\big|\mathcal{C}^{[K]}(v)\big| \geq k_n\big) \\
&\quad \leq \mathrm{e}^{-u\varepsilon K^{\rho}} \frac{1}{k_n} \sum_{v\in[n]} \mathbb{P}\big(|\mathcal{C}(v)| \geq k_n\big) \\
&\quad = \mathrm{e}^{-u\varepsilon K^{\rho}} \frac{n}{k_n} \mathbb{P}\big(|\mathcal{C}(V)| \geq k_n\big),
\end{aligned} \tag{5.8}$$

where $V \in [n]$ is a vertex chosen uniformly at random from $[n]$. By [34], Proposition 2.4(a), there exists a constant $a_1 < \infty$ such that

$$\begin{aligned}
\mathbb{P}\big(|\mathcal{C}(V)| \geq k_n\big) &\leq a_1\big(k_n^{-1/(\tau-2)} + \big(\varepsilon_n \vee n^{-(\tau-3)/(\tau-1)}\big)^{1/(\tau-3)}\big) \\
&\leq a_1\big(k_n^{-1/(\tau-2)} + n^{-1/(\tau-1)}\big),
\end{aligned} \tag{5.9}$$

so that, for $k = k_n = \varepsilon n^{\rho}$ with $\varepsilon < 1$ and with $a_1' = 2a_1$,

$$\frac{n}{k_n} \mathbb{P}\big(|\mathcal{C}(V)| \geq k_n\big) \leq a_1' \varepsilon^{-(\tau-1)/(\tau-2)} n^{-\rho}. \tag{5.10}$$

Therefore, the term in (5.5) is bounded by

$$\mathrm{e}^{-a\varepsilon K^{\rho}} a_1' \varepsilon^{-(\tau-1)/(\tau-2)}. \tag{5.11}$$

When we pick $K = K(\varepsilon)$ sufficiently large, we can make this as small as we wish.

We continue with the term in (5.6), for which we use a large deviation argument. We formulate this result in the following lemma:

LEMMA 5.1 (Large deviations for cluster weights). *For every* $k = o(n)$ *and* $K = o(n)$, *there exists a* $J > 0$ *such that*

$$\mathbb{P}\big(\exists v : \big|\mathcal{C}^{[K]}(v)\big| \geq k, \mathcal{W}^{[K]}(v) \leq k/2\big) \leq n\mathrm{e}^{-Jk}. \tag{5.12}$$

PROOF. When $|\mathcal{C}^{[K]}(v)| \geq k$, then $\mathcal{W}^{[K]}(v)$ is stochastically bounded from below by the sum $\sum_{i=1}^{k} w_{v(i)}$, where $(v(i))_{i=1}^{k}$ is the sized-biased ordering of $[n]$, that is, for every $j \notin (v(s))_{s\in[i-1]}$,

$$\mathbb{P}\big(v(i) = j \mid (v(s))_{s\in[i-1]}\big) = \frac{w_j}{\sum_{l\notin(v(s))_{s\in[i-1]}} w_l}. \tag{5.13}$$

See [8], Section 2, Lemma 2.1, for more details about the size-biased reordering. Indeed, each time we draw a random mark and, conditionally on this mark not being one that has been found earlier as well as on all the marks found so far, it will be equal to j with the probability in (5.13). When $|\mathcal{C}^{[K]}(v)| \geq k$, we must draw a vertex that we have not seen yet, a total of at least k times.

We apply the size-biased reordering to the vertex set $[n] \setminus [K]$. Then, for each i and conditionally on $(v(s))_{s\in[i-1]}$, the random variable $w_{v(i)}$ is stochastically bounded from above by the random variable W_i' with distribution

$$\mathbb{P}(W_i' = w_j) = \frac{w_j}{\ell_n - \sum_{s=1}^{i-1+K} w_s}, \qquad j \in [n] \setminus [i-1+K], \tag{5.14}$$

that is, we have removed the vertices with the largest $i - 1 + K$ weights. As a result, the random variables $(W_i')_{i\geq 1}$ are *independent*. Now take $\kappa > 0$ very small, and note that, whenever $k - 1 + K \leq \kappa n$ and for every $i \leq k$, W_i' is stochastically bounded from above by a random variable $W_i^{(n)}(\kappa)$ with distribution

$$\mathbb{P}\big(W_i^{(n)}(\kappa) = w_j\big) = \frac{w_j}{\ell_n - \sum_{s=1}^{\kappa n} w_s}, \qquad j \in [n] \setminus [\kappa n], \tag{5.15}$$

where the random variables $(W_i^{(n)}(\kappa))_{i=1}^{k}$ are i.i.d. Now take $\kappa > 0$ so small that

$$\mathbb{E}\big[W_i^{(n)}(\kappa)\big] = \sum_{j=\kappa n}^{n} \frac{w_j^2}{\ell_n - \sum_{s=1}^{\kappa n} w_s} \geq 3/4. \tag{5.16}$$

Then,

$$\begin{aligned} &\mathbb{P}\big(\exists v : \big|\mathcal{C}^{[K]}(v)\big| \geq k, \mathcal{W}^{[K]}(v) \leq k/2\big) \\ &\quad \leq \sum_{v=K+1}^{n} \mathbb{P}\big(\big|\mathcal{C}^{[K]}(v)\big| \geq k, \mathcal{W}^{[K]}(v) \leq k/2\big) \\ &\quad \leq n\mathbb{P}\left(\sum_{i=1}^{k} W_i^{(n)}(\kappa) \leq k/2\right). \end{aligned} \tag{5.17}$$

Intuitively, since $\mathbb{E}[W_i^{(n)}(\kappa)] \approx \nu_n \approx \nu = 1$ for $\kappa > 0$ small, the Chernoff bound proves that $\mathbb{P}(\sum_{i=1}^k W_i^{(n)}(\kappa) \le k/2)$ is exponentially small in k, so that the term in (5.6) is exponentially small. We now make this intuition precise.

By the Chernoff bound, for each $\vartheta \ge 0$, and by the fact that $(W_i^{(n)}(\kappa))_{i\in[k]}$ are i.i.d. random variables, we have

$$\mathbb{P}\left(\sum_{i=1}^k W_i^{(n)}(\kappa) \le k/2\right) \le \mathrm{e}^{\vartheta k/2}\mathbb{E}\big[\mathrm{e}^{-\vartheta \sum_{i=1}^k W_i^{(n)}(\kappa)}\big] = (\mathrm{e}^{\vartheta/2}\phi_{n,\kappa}(\vartheta))^k, \tag{5.18}$$

where

$$\phi_{n,\kappa}(\vartheta) = \mathbb{E}\big[\mathrm{e}^{-\vartheta W_1^{(n)}(\kappa)}\big] \tag{5.19}$$

denotes the Laplace transform of $W_1(\kappa)$. By (5.18), it suffices to prove that there exists a $\vartheta > 0$ such that, uniformly in n sufficiently large, $\vartheta/2 + \log\phi_{n,\kappa}(\vartheta) < 0$. This is what we shall show now. By dominated convergence, for each fixed $\vartheta > 0$,

$$\log\phi_{n,\kappa}(\vartheta) \to \log\phi_\kappa(\vartheta) = \log\mathbb{E}\big[\mathrm{e}^{-\vartheta W(\kappa)}\big], \tag{5.20}$$

where

$$\mathbb{P}\big(W(\kappa) \le x\big) = \mathbb{E}\big[[1-F]^{-1}(U) \mid U \ge \kappa\big], \tag{5.21}$$

and U is a uniform random variable on $[0,1]$. As a result, the distribution of U *conditionally on* $U \ge \kappa$ is uniform on $[\kappa, 1]$. Let U_κ denote a uniform random variable on $[\kappa, 1]$, so that $W(\kappa) \stackrel{d}{=} [1-F]^{-1}(U_\kappa)$. Then, $W(\kappa)$ has mean $\mathbb{E}[W(\kappa)] \ge 3/4$ and bounded variance σ_κ^2 (since $W(\kappa) \le [1-F]^{-1}(\kappa) < \infty$ a.s.). Therefore, a Taylor expansion yields that, for *fixed* $\kappa > 0$,

$$\log\phi_\kappa(\vartheta) \le -3\vartheta/4 + \sigma_\kappa^2\vartheta^2 + o(\vartheta^2). \tag{5.22}$$

Now, fix a $\vartheta > 0$ so small that

$$\vartheta/2 - 3\vartheta/4 + \sigma_\kappa^2\vartheta^2 \le -\vartheta/6, \tag{5.23}$$

and then N so large that, for all $n \ge N$,

$$\log\phi_{n,\kappa}(\vartheta) \le \log\phi_\kappa(\vartheta) + \vartheta/12. \tag{5.24}$$

Then, indeed, for $n \ge N$, since $\vartheta > 0$,

$$\vartheta/2 + \log\phi_{n,\kappa}(\vartheta) \le -\vartheta/6 + \vartheta/12 = -\vartheta/12 < 0, \tag{5.25}$$

so that

$$\mathrm{e}^{\vartheta/2}\phi_{n,\kappa}(\vartheta) \le \mathrm{e}^{-\vartheta/12}, \tag{5.26}$$

which, in turn, implies that

$$\sum_{v=1}^n \mathbb{P}\big(|\mathcal{C}^{[K]}(v)| \ge k, \mathcal{W}^{[K]}(v) \le k/2\big) \le n\mathrm{e}^{-k\vartheta/12}. \tag{5.27}$$

When $n \to \infty$, this proves the claim for $J = \vartheta/12$. $\square$

To prove Theorem 1.6(a), we apply Lemma 5.1 to the term in (5.6), which is then bounded by $e^{-\Theta(\varepsilon n^{\rho})}$ when we take $k = \varepsilon n^{\rho}$.

(b) We denote by

$$Z_{\geq k} = \sum_{v=1}^{n} \mathbb{1}_{\{|\mathcal{C}(v)| \geq k\}} \tag{5.28}$$

the number of vertices that are contained in connected components of size at least k. In [34], the random variable $Z_{\geq k}$ has been used in a crucial way to prove probabilistic bounds on $|\mathcal{C}_{\max}|$. We now slightly extend these results.

We shall prove that, for all $\varepsilon > 0$ sufficiently small, there exist constants b_2, C such that

$$\mathbb{P}\big(Z_{\geq \varepsilon n^{\rho}} \leq b_2 n^{\rho} \varepsilon^{-1/(\tau-2)}\big) \leq C\varepsilon^{2/(\tau-2)}. \tag{5.29}$$

We first note that it suffices to prove (5.29) when $\nu_n \leq 1 - Kn^{-\eta}$. Indeed, the random variable $Z_{\geq \varepsilon n^{\rho}}$ is increasing in the edge occupation statuses, and, therefore, we may take $\lambda < 0$ so that $-\lambda > K$ to achieve the claim.

We shall use a second moment method. By [34], Proposition 2.4(b), there exists $a_2 = a_2(K)$ such that

$$\mathbb{E}[Z_{\geq \varepsilon n^{\rho}}] \geq n\mathbb{P}\big(|\mathcal{C}(V)| \geq \varepsilon n^{\rho}\big) \geq a_2 n^{\rho} \varepsilon^{-1/(\tau-2)}, \tag{5.30}$$

where V is chosen uniformly from $[n]$. Therefore, when we take $b_2 = a_2/2$,

$$\mathbb{P}\big(Z_{\geq \varepsilon n^{\rho}} \leq b_2 n^{\rho} \varepsilon^{-1/(\tau-2)}\big) \leq \mathbb{P}(Z_{\geq \varepsilon n^{\rho}} \leq \mathbb{E}[Z_{\geq \varepsilon n^{\rho}}]/2). \tag{5.31}$$

We take $\varepsilon > 0$ small, and bound, by the Chebychev inequality,

$$\mathbb{P}(Z_{\geq \varepsilon n^{\rho}} \leq \mathbb{E}[Z_{\geq \varepsilon n^{\rho}}]/2) \leq \frac{4\,\mathrm{Var}(Z_{\geq \varepsilon n^{\rho}})}{\mathbb{E}[Z_{\geq \varepsilon n^{\rho}}]^2}. \tag{5.32}$$

By [34], Proposition 2.2, and [34], Proposition 2.5 and (2.22), uniformly in $k \geq 1$,

$$\mathrm{Var}(Z_{\geq k}) \leq n\mathbb{E}[|\mathcal{C}(V)|] \leq n^{1+\eta} = n^{\rho}. \tag{5.33}$$

As a result, we obtain

$$\mathbb{P}(Z_{\geq \varepsilon n^{\rho}} \leq \mathbb{E}[Z_{\geq \varepsilon n^{\rho}}]/2) \leq \frac{4n^{2\rho}}{a_2^2 \varepsilon^{-2/(\tau-2)} n^{2\rho}} = C\varepsilon^{2/(\tau-2)}, \tag{5.34}$$

which is small when $\varepsilon > 0$ is small. We conclude that, with probability at least $1 - o_{\varepsilon}(1)$, where $o_{\varepsilon}(1)$ denotes a function that is $o(1)$ uniformly in n as $\varepsilon \downarrow 0$,

$$Z_{\geq \varepsilon n^{\rho}} \geq \mathbb{E}[Z_{\geq \varepsilon n^{\rho}}]/2 \geq \frac{a_2}{2} \varepsilon^{-1/(\tau-2)} n^{\rho}. \tag{5.35}$$

Since, by [34], Theorem 1.2, $|\mathcal{C}_{\max}| \le \varepsilon^{-1/2} n^{\rho}$ with probability at least $1 - o_\varepsilon(1)$, there are, again with probability at least $1 - o_\varepsilon(1)$, at least

$$\frac{a_2}{2}\varepsilon^{-1/(\tau-2)} n^{\rho}/(\varepsilon^{-1/2} n^{\rho}) = C\varepsilon^{1/2-1/(\tau-2)} \tag{5.36}$$

clusters of size at least εn^{ρ}. Since $1/2 - 1/(\tau-2) < 0$, the number of clusters of size at least εn^{ρ} tends to infinity when $\varepsilon \downarrow 0$. By part (a), *whp* for $K \ge 1$ large, these clusters will be part of $(|\mathcal{C}_{\le}(i)|)_{i\in[K]}$ when $K = K(\varepsilon) \ge 1$ is sufficiently large. □

We now complete the proof of Theorem 1.5.

PROOF OF THEOREM 1.5. We use Proposition 3.7 and note that the limiting variables are all nontrivial (i.e., they are equal to 0 or 1 each with positive probability). This proves (1.27). The proof of (1.28) is similar, noting that $|\mathcal{C}_{\le}(i)|$ equals $|\mathcal{C}_{\max}|$ with strictly positive probability. □

We finally use Theorem 1.6 to complete the proof of Theorem 1.1:

PROOF OF THEOREM 1.1. Weak convergence of $(|\mathcal{C}_{(i)}|n^{-\rho})_{i\ge1}$ in the product topology is equivalent to the weak convergence of $(|\mathcal{C}_{(i)}|n^{-\rho})_{i\in[m]}$ for any $m \ge 1$; see [28], Theorem 4.29. In turn, by Theorem 1.6, this follows from the convergence in distribution of $(|\mathcal{C}_{\le}(i)|n^{-\rho})_{i\in[m]}$ for all m. The latter follows from Theorem 4.1. Since, *whp* for large K, again by Theorem 1.6, $(|\mathcal{C}_{(i)}|n^{-\rho})_{i\in[m]}$ is equal to the largest m components of $(|\mathcal{C}_{\le}(i)|n^{-\rho})_{i\in[K]}$, we have identified

$$(\gamma_i(\lambda))_{i\ge1} \stackrel{d}{=} \big(H_{(i)}(0)\big)_{i\ge1}, \tag{5.37}$$

where $(H_{(i)}(0))_{i\ge1}$ is $(H_i(0))_{i\ge1}$ ordered in size. This completes the proof of Theorem 1.1 and identifies the limiting random variables. □

6. Proof of Theorem 1.3. In this section, we shall prove Theorem 1.3 on the largest subcritical clusters. We shall extend the result also to the ordered weights of subcritical clusters as formulated in Theorem 1.4, which shall be a crucial ingredient in the proof of Theorem 1.2, which is given in Section 7 below.

We shall prove that Theorem 1.3 holds for $\mathcal{W}_{(j)}$ as well as for $|\mathcal{C}_{(j)}|$. Indeed, it shall also follow from the result that *whp*, $\mathcal{W}_{(j)} = \sum_{i\in\mathcal{C}_{(j)}} w_i$, that is, the jth largest cluster weight is the weight of the jth largest cluster, as claimed in Theorem 1.4.

To prove this scaling, we shall prove that, when the weights are equal to $\mathbf{w}(\lambda_n)$ as defined in (1.15), and when $\lambda_n \to -\infty$,

$$|\lambda_n|n^{-\rho}|\mathcal{C}(j)| \xrightarrow{\mathbb{P}} c_j, \qquad |\lambda_n|n^{-\rho}\mathcal{W}(j) \xrightarrow{\mathbb{P}} c_j, \tag{6.1}$$

where we recall that

$$c_j = c_F^{\alpha} j^{-\alpha} = \lim_{n\to\infty} n^{-\alpha} w_j. \tag{6.2}$$

Since $j \mapsto c_j$ is strictly decreasing, this means that, *whp*, $\mathcal{C}(j) = \mathcal{C}_{\leq}(j)$. Thus, this also implies that *whp*, $\mathcal{C}(j) = \mathcal{C}_{(j)}$ for all $j \leq m$. Then (6.1) proves the result for the ordered cluster sizes and weights.

Recall the definitions of T, $T(i)$ and their weights w_T and $w_{T(i)}$ introduced in Section 2.2, where also their moments are computed in Lemma 2.3. We make frequent use of these computations. The proof of Theorem 1.3 consists of four key steps, which we shall prove one by one.

Asymptotics of mean cluster size and weight of high-weight vertices. In the following lemma we investigate the means of $|\mathcal{C}(j)|$ and $\mathcal{W}(j)$:

LEMMA 6.1 (Mean cluster size and weights). *As $n \to \infty$, for every $j \in \mathbb{N}$ fixed, and when $\lambda_n \to -\infty$ such that $\nu_n(\lambda_n) \to 1$,*

$$\begin{aligned} \mathbb{E}[|\mathcal{C}(j)|] &= \frac{w_j}{1-\nu_n(\lambda_n)}(1+o(1)), \\ \mathbb{E}[\mathcal{W}(j)] &= \frac{w_j}{1-\nu_n(\lambda_n)}(1+o(1)). \end{aligned} \tag{6.3}$$

PROOF. By the fact that $|\mathcal{C}(j)|$ and $T(j)$ can be coupled so that $|\mathcal{C}(j)| \leq T(j)$ a.s., we obtain that

$$\mathbb{E}[|\mathcal{C}(j)|] \leq \mathbb{E}[T(j)] = \frac{w_j}{1-\nu_n(\lambda_n)}, \tag{6.4}$$

the latter equality following from Lemma 2.3(c). A similar upper bound follows for $\mathbb{E}[\mathcal{W}(j)]$ now using Lemma 2.3(d).

For the lower bound, we rewrite

$$\mathbb{E}[|\mathcal{C}(j)|] = \mathbb{E}[T(j)] - \mathbb{E}[T(j) - |\mathcal{C}(j)|]. \tag{6.5}$$

Now, for $a_n = n^\rho \gg \mathbb{E}[T(j)]$, we bound

$$\mathbb{E}[T(j) - |\mathcal{C}(j)|] \leq \mathbb{E}\big[T(j)\mathbb{1}_{\{T(j)>a_n\}}\big] + \mathbb{E}\big[[T(j) - |\mathcal{C}(j)|]\mathbb{1}_{\{T(j)\leq a_n\}}\big]. \tag{6.6}$$

By Lemma 2.3(c), the first term in (6.6) is bounded by

$$\begin{aligned} &\mathbb{E}\big[T(j)\mathbb{1}_{\{T(j)>a_n\}}\big] \\ &\quad \leq \frac{1}{a_n}\mathbb{E}[T(j)^2] \\ &\quad = \frac{1}{a_n}\biggl(\biggl(1+\frac{w_j}{1-\nu_n(\lambda_n)}\biggr)^2 + \frac{w_j(1+\nu_n(\lambda_n))}{(1-\nu_n(\lambda_n))^2} \\ &\qquad\qquad + \frac{w_j}{(1-\nu_n(\lambda_n))^3}\frac{1}{\ell_n}\sum_{i\in[n]} w_i^3\biggr). \end{aligned} \tag{6.7}$$

The first two terms in (6.7) are $o(w_j/(1-\nu_n(\lambda_n)))$ since $\nu_n(\lambda_n) = 1 + n^{-\eta}\lambda_n + o(n^{-\eta}|\lambda_n|)$ by (2.30) and the fact that $\lambda_n \to -\infty$, so that

$$w_j/(1-\nu_n(\lambda_n)) \le cn^{\alpha+\eta}|\lambda_n|^{-1} = cn^{\rho}|\lambda_n|^{-1} = o(n^{\rho}) = o(a_n), \tag{6.8}$$

since $\alpha + \eta = \rho$ [recall (2.10)]. The last term in (6.7) is bounded by

$$\frac{w_j}{1-\nu_n(\lambda_n)} \frac{cn^{3\alpha-1}}{a_n(1-\nu_n(\lambda_n))^2} = \frac{w_j}{1-\nu_n(\lambda_n)} cn^{3\alpha-1-\rho-2\eta}|\lambda_n|^{-2}. \tag{6.9}$$

By (2.10), $3\alpha - 1 - \rho - 2\eta = 3(\tau-4)/(\tau-1) < 0$, so that also this term is $o(w_j/(1-\nu_n(\lambda_n)))$.

For the second term in (6.6), we note that differences between $T(j)$ and $|\mathcal{C}(j)|$ arise due to vertices which have been used *at least twice* in $T(j)$. Indeed, as explained in more detail in Section 2.2, the law of $|\mathcal{C}(j)|$ can be obtained from the branching process by removing vertices (and their complete offspring) of which the mark has already been used (see the description of the cluster exploration in Section 2.1 and the relation to branching processes described in Sections 2.1 and 2.2). Thus, when we draw vertex i twice, then the second time we must thin the entire tree that is rooted at this vertex with mark i. The expected number of vertices in the tree equals $\mathbb{E}[T(i)]$, so that we arrive at

$$\begin{aligned}
&\mathbb{E}\big[[T(j) - |\mathcal{C}(j)|]\mathbb{1}_{\{T(j)\le a_n\}}\big] \\
&\quad \le \sum_{i\in[n]} \mathbb{E}\big[[T(j) - |\mathcal{C}(j)|]\mathbb{1}_{\{T(j)\le a_n\}}\mathbb{1}_{\{\text{mark } i \text{ drawn at least twice}\}}\big] \\
&\quad \le \sum_{i\in[n]} \mathbb{E}[T(i)] \sum_{s_1<s_2=1}^{a_n} \mathbb{P}(\text{mark } i \text{ drawn at times } s_1, s_2).
\end{aligned} \tag{6.10}$$

Now, i can only be chosen at time s_1 when $T(j) \ge s_1 - 1$, which is independent of the event that the mark i is chosen at times s_1, s_2. Therefore,

$$\begin{aligned}
\mathbb{E}\big[[T(j) - |\mathcal{C}(j)|]\mathbb{1}_{\{T(j)\le a_n\}}\big] &\le \sum_{i\in[n]} \mathbb{E}[T(i)] \sum_{s_1<s_2=1}^{a_n} \mathbb{P}\big(T(j) \ge s_1 - 1\big)\frac{w_i^2}{\ell_n^2} \\
&\le a_n \sum_{s_1=1}^{a_n} \mathbb{P}\big(T(j) \ge s_1 - 1\big) \sum_{i\in[n]} \mathbb{E}[T(i)]\frac{w_i^2}{\ell_n^2} \\
&\le a_n \mathbb{E}[T(j)] \sum_{i\in[n]} \mathbb{E}[T(i)]\frac{w_i^2}{\ell_n^2}.
\end{aligned} \tag{6.11}$$

This is $o(w_j/(1-\nu_n(\lambda_n)))$ when $\lambda_n \to -\infty$, since

$$\begin{aligned}
a_n \sum_{i\in[n]} \mathbb{E}[T(i)]\frac{w_i^2}{\ell_n^2} &= a_n \sum_{i\in[n]} \frac{w_i^3}{\ell_n^2(1-\nu_n(\lambda_n))} \le \frac{C}{|\lambda_n|} n^{\rho-2+3\alpha+\eta} \\
&= \frac{C}{|\lambda_n|} = o(1).
\end{aligned} \tag{6.12}$$

This completes the proof for $\mathbb{E}[|\mathcal{C}(j)|]$. The proof for $w_{T(j)}$ is similar. Indeed, we split

$$\begin{aligned}\mathbb{E}\big[w_{T(j)} - \mathcal{W}(j)\big] \leq \mathbb{E}\big[w_{T(j)}\mathbb{1}_{\{T(j)>a_n\}}\big] \\ + \mathbb{E}\big[\big[w_{T(j)} - \mathcal{W}(j)\big]\mathbb{1}_{\{T(j)\leq a_n\}}\big].\end{aligned} \tag{6.13}$$

The first term is now bounded by

$$\mathbb{E}\big[w_{T(j)}\mathbb{1}_{\{T(j)>a_n\}}\big] \leq \frac{1}{a_n}\mathbb{E}\big[w_{T(j)}T(j)\big], \tag{6.14}$$

which we can again bound using $\mathbb{E}[w_{T(j)}T(j)] \leq \mathbb{E}[w_{T(j)}^2] + \mathbb{E}[T(j)^2]$ together with Lemma 2.3(a) and (b). Further,

$$\begin{aligned}&\mathbb{E}\big[\big[w_{T(j)} - \mathcal{W}(j)\big]\mathbb{1}_{\{T(j)\leq a_n\}}\big] \\ &\quad \leq \sum_{i\in[n]} \mathbb{E}\big[\big[w_{T(j)} - \mathcal{W}(j)\big]\mathbb{1}_{\{T(j)\leq a_n\}}\mathbb{1}_{\{\text{mark } i \text{ drawn at least twice}\}}\big] \\ &\quad \leq \sum_{i\in[n]} \mathbb{E}\big[w_{T(i)}\big] \sum_{s_1,s_2=1}^{a_n} \mathbb{P}(\text{mark } i \text{ drawn at times } s_1, s_2) \\ &\quad \leq a_n\mathbb{E}[T(j)] \sum_{i\in[n]} \mathbb{E}\big[w_{T(i)}\big]\frac{w_i^2}{\ell_n^2} = a_n \frac{w_j \nu_n(\lambda_n)}{(1-\nu_n(\lambda_n))^2} \sum_{i\in[n]} \frac{w_i^3}{\ell_n^2},\end{aligned} \tag{6.15}$$

so that

$$\mathbb{E}[\mathcal{W}(j)] \geq \mathbb{E}\big[w_{T(j)}\big] - \frac{1}{a_n}\mathbb{E}\big[w_{T(j)}T(j)\big] - a_n \frac{w_j}{(1-\nu_n(\lambda_n))^2} \sum_{i\in[n]} \frac{w_i^3}{\ell_n^2}. \tag{6.16}$$

We bound $\mathbb{E}[w_{T(j)}T(j)] \leq \mathbb{E}[w_{T(j)}^2] + \mathbb{E}[T(j)^2]$. Now we can simply follow the argument for $\mathbb{E}[|\mathcal{C}(j)|]$. □

Cluster size and weight of high weight vertices are concentrated. We note that, by the stochastic domination and the fact that $\mathbb{E}[|\mathcal{C}(j)|] = \frac{w_j}{1-\nu_n(\lambda_n)}(1+o(1))$, we have

$$\operatorname{Var}(|\mathcal{C}(j)|) \leq \operatorname{Var}(T(j)) + o(\mathbb{E}[T(j)]^2). \tag{6.17}$$

By Lemma 2.3(a),

$$\begin{aligned}\operatorname{Var}(T(j)) &= \frac{w_j(1+\nu_n(\lambda_n))}{1-\nu_n(\lambda_n)} + \frac{w_j}{(1-\nu_n(\lambda_n))^3}\bigg(\frac{1}{\ell_n}\sum_{l\in[n]} w_l^3\bigg) \\ &= o\bigg(\frac{w_j^2}{(1-\nu_n(\lambda_n))^2}\bigg),\end{aligned} \tag{6.18}$$

since j is fixed and

$$\left(\frac{1}{\ell_n}\sum_{l\in[n]} w_l^3\right)\bigl(1-\nu_n(\lambda_n)\bigr)^{-1} = \frac{C}{|\lambda_n|} n^{3\alpha-1+\eta} = o(n^\alpha) = o(w_j). \tag{6.19}$$

For $w_{T(j)}$ the argument is identical. We conclude that, for j fixed, $\mathrm{Var}(|\mathcal{C}(j)|) = o(\mathbb{E}[|\mathcal{C}(j)|]^2)$ and $\mathrm{Var}(\mathcal{W}(j)) = o(\mathbb{E}[\mathcal{W}(j)]^2)$, so that

$$\frac{|\mathcal{C}(j)|}{\mathbb{E}[|\mathcal{C}(j)|]} \xrightarrow{\mathbb{P}} 1, \qquad \frac{\mathcal{W}(j)}{\mathbb{E}[\mathcal{W}(j)]} \xrightarrow{\mathbb{P}} 1, \tag{6.20}$$

and then Lemma 6.1 completes the proof of (6.1).

Cluster weight sums. We start by proving a convenient result relating the cluster weights $\mathcal{W}(j)$ and $\mathcal{W}_{\le}(j)$.

LEMMA 6.2 (Cluster weight properties). (a) *For every integer $m \ge 2$,*

$$\sum_{j\in[n]} \mathcal{W}_{\le}(j)^m = \sum_{j\in[n]} w_j \mathcal{W}(j)^{m-1}. \tag{6.21}$$

(b) *For every $i, j \in [n]$,*

$$\mathbb{E}\bigl[\mathcal{W}(i)\mathcal{W}(j)\mathbb{1}_{\{i \not\longleftrightarrow j\}}\bigr] \le \mathbb{E}[\mathcal{W}(i)]\mathbb{E}[\mathcal{W}(j)]. \tag{6.22}$$

PROOF. (a) We compute

$$\begin{aligned} \sum_{j\in[n]} \mathcal{W}_{\le}(j)^m &= \sum_{j\in[n]}\sum_{i_1,\dots,i_m}\prod_{s=1}^m w_{i_s}\mathbb{1}_{\{i_s\in\mathcal{C}(i_1)\ \forall s=2,\dots,m, \min\mathcal{C}(i_1)=j\}} \\ &= \sum_{i_1,\dots,i_m}\prod_{s=1}^m w_{i_s}\mathbb{1}_{\{i_s\in\mathcal{C}(i_1)\ \forall s=2,\dots,m\}} \\ &= \sum_{i_1\in[n]} w_{i_1}\mathcal{W}(i_1)^{m-1}. \end{aligned} \tag{6.23}$$

(b) We write out

$$\begin{aligned} \mathbb{E}\bigl[\mathcal{W}(i)\mathcal{W}(j)\mathbb{1}_{\{i \not\longleftrightarrow j\}}\bigr] &= \sum_{k,l} w_k w_l \mathbb{P}(i\longleftrightarrow k, j \longleftrightarrow l, i \not\longleftrightarrow j) \\ &\le \sum_{k,l} w_k w_l \mathbb{P}(i\longleftrightarrow k)\mathbb{P}(j\longleftrightarrow l) \\ &= \mathbb{E}[\mathcal{W}(i)]\mathbb{E}[\mathcal{W}(j)] \end{aligned} \tag{6.24}$$

by the BK-inequality; see [20], Section 2.3. □

Only high-weight vertices matter. We start by proving that the probability that, for $K \geq 1$, there exists a $j > K$ such that $\mathcal{W}_{\leq}(j) \geq \varepsilon n^{\rho}/|\lambda_n|$ is small. Since, for all $j \leq K$, we have that $|\lambda_n| n^{-\rho} \mathcal{W}(j) \xrightarrow{\mathbb{P}} c_j$, we have that, for all $i \leq m$ and m such that $c_m > \varepsilon$, $\mathcal{W}(j) = \mathcal{W}_{(j)}$.

Recall that $\mathcal{W}^{[K]}(j)$ is the weight of the cluster of j in the random graph only making use of the vertices in $[n] \setminus [K]$, and let $\mathcal{W}_{\leq}^{[K]}(j) = \mathcal{W}^{[K]}(j)$ when j is the minimal element in $\mathcal{C}^{[K]}(j)$. If there exists a $j > K$ such that $\mathcal{W}^{[K]}(j) \geq \varepsilon n^{\rho}/|\lambda_n|$, then

$$\sum_{j>K} \mathcal{W}_{\leq}^{[K]}(j)^3 \geq \frac{\varepsilon^3}{|\lambda_n|^3} n^{3\rho}. \tag{6.25}$$

Since

$$\ell_n \geq \sum_{j>K} w_j, \tag{6.26}$$

we see that this random graph is stochastically bounded by the random graph having weights $\mathbf{w}^{[K]}$, where $w_j^{[K]} = 0$ when $j \leq K$ and $w_j^{[K]} = w_j$ otherwise. By the Markov inequality,

$$\begin{aligned} &\mathbb{P}\big(\exists j > K : \mathcal{W}_{\leq}(j) \geq \varepsilon n^{\rho}/|\lambda_n|\big) \\ &\quad \leq \sum_{j>K} \mathcal{W}_{\leq}^{[K]}(j)^3 = \frac{|\lambda_n|^3}{\varepsilon^3} n^{-3\rho} \sum_{j>K} w_j \mathbb{E}\big[\mathcal{W}^{[K]}(j)^2\big], \end{aligned} \tag{6.27}$$

where we have used Lemma 6.2 for the equality. We note that we can again stochastically dominate $|\mathcal{C}^{[K]}(j)|$ by $T^{[K]}(j)$ and $\mathcal{W}^{[K]}(j)$ by $w_{T^{[K]}(j)}$, where now the offspring distribution is equal to $X_i^{[K]} = X_i^{(\mathrm{BP})} \mathbb{1}_{\{M_i > K\}}$ (recall Section 2.1). Therefore, by Lemma 2.3(d), we obtain that

$$\begin{aligned} \mathbb{E}\big[\mathcal{W}^{[K]}(j)^2\big] &\leq \mathbb{E}\big[w_{T^{[K]}(j)}^2\big] \\ &= \left(\frac{w_j^{[K]}}{1 - \nu_n^{[K]}}\right)^2 + \frac{w_j^{[K]}}{(1 - \nu_n^{[K]})^3} \left(\frac{1}{\ell_n} \sum_{i \in [n]} (w_i^{[K]})^3\right), \end{aligned} \tag{6.28}$$

where

$$w_j^{[K]} = w_j \mathbb{1}_{\{j>K\}}, \qquad \nu_n^{[K]} = \sum_{j \in [n]} (w_j^{[K]})^2/\ell_n. \tag{6.29}$$

It is not hard to see that, for each $K \geq 1$ fixed, as $n \to \infty$,

$$\mathbb{E}\big[w_{T^{[K]}(j)}^2\big] \leq \left(\frac{w_j}{1 - \nu_n(\lambda_n)}\right)^2 + (1 + o(1)) \frac{w_j}{(1 - \nu_n(\lambda_n))^3} \left(\frac{1}{\ell_n} \sum_{i>K}^{n} w_i^3\right). \tag{6.30}$$

Substitution of the bound (6.28) in the right-hand side of (6.27) and performing the sum over j gives that

$$\begin{aligned}\sum_{j>K} w_j \mathbb{E}[\mathcal{W}^{[K]}(j)^2] &\leq \frac{1}{(1-\nu_n(\lambda_n))^2}\left(1+\frac{1}{1-\nu_n(\lambda_n)}\right)\sum_{j>K} w_j^3 \\ &\leq CK^{1-3\alpha}(n^\rho/|\lambda_n|)^3,\end{aligned} \tag{6.31}$$

so that

$$\begin{aligned}\mathbb{P}(\exists j > K : \mathcal{W}_{\leq}(j) \geq \varepsilon n^\rho/|\lambda_n|) &\leq |\lambda_n|^3 \varepsilon^{-3} n^{-3\rho} CK^{1-3\alpha}(n^\rho/|\lambda_n|)^3 \\ &= CK^{1-3\alpha}\varepsilon^{-3},\end{aligned} \tag{6.32}$$

which can be made arbitrarily small by taking $K = K(\varepsilon)$ large.

We complete this section by proving that the probability that there exists a $j > K$ such that $|\mathcal{C}_{\leq}(j)| \geq \varepsilon n^\rho/|\lambda_n|$ is small. For this, we use Lemma 5.1, which proves that, *whp*, if $|\mathcal{C}_{\leq}(j)| \geq \varepsilon n^\rho/|\lambda_n|$, then also $\mathcal{W}_{\leq}(j) \geq \varepsilon n^\rho/(2|\lambda_n|)$. Thus, the result for cluster sizes follows from the proof for cluster weights. This completes the proof of Theorem 1.3.

7. Proof of Theorem 1.2.

In this section, we prove Theorem 1.2. We start by using [3], Proposition 7, to show that the random graph multiplicative coalescent converges (recall Lemma 1.7).

Convergence of the random graph multiplicative coalescent at fixed time. We apply [3], Proposition 7, which gives conditions to show that, for fixed $\lambda \in \mathbb{R}$, the random sequence $\mathbf{X}^{(n)}(|\lambda_n| + \lambda)$ converges in distribution to a random variable which has the same distribution as the $(0, \beta, \mathbf{d})$-multiplicative coalescent at time λ when three conditions are satisfied about the initial state $\mathbf{x}^{(n)} = \mathbf{X}^{(n)}(0)$. To state these conditions, we define, for $r = 2, 3$, with $\mathbf{x}^{(n)} = (x_j^{(n)})_{j\geq 1}$,

$$\sigma_r(\mathbf{x}^{(n)}) = \sum_j (x_j^{(n)})^r. \tag{7.1}$$

Then, the conditions in [3], Proposition 7, are that, as $\lambda_n \to -\infty$:

(a)

$$|\lambda_n|(|\lambda_n|\sigma_2(\mathbf{x}^{(n)}) - 1) \xrightarrow{\mathbb{P}} -\beta; \tag{7.2}$$

(b)

$$\frac{x_j^{(n)}}{\sigma_2(\mathbf{x}^{(n)})} \xrightarrow{\mathbb{P}} d_j; \tag{7.3}$$

(c)

$$|\lambda_n|^3\sigma_3(\mathbf{x}^{(n)}) \xrightarrow{\mathbb{P}} \sum_{j=1}^{\infty} d_j^3. \tag{7.4}$$

The conditions (a)–(c) above are not precisely what is in [3], Proposition 7, and we start by explaining how (a)–(c) imply the conditions for [3], Proposition 7. Indeed, in [3], Proposition 7, the condition in (a) is replaced by $\sigma_2(\mathbf{x}^{(n)}) \to 0$, and the process

$$\mathbf{X}^{(n)}\left(\frac{1}{\sigma_2(\mathbf{x}^{(n)})} + \lambda\right) \tag{7.5}$$

is proved to converge to the realization of a $(0, 0, \mathbf{d})$-multiplicative coalescent at time λ. Under condition (a) (and the fact that $\lambda_n \to -\infty$), (a) implies that $1/\sigma_2(\mathbf{x}^{(n)}) = |\lambda_n| - \beta + o(1)$. Since if $(\mathbf{X}(t))_t$ is a multiplicative coalescent with parameters $(0, 0, \mathbf{d})$, then $(\mathbf{X}(t - \beta))_t$ is a multiplicative coalescent with parameters $(0, \beta, \mathbf{d})$ (see [3], (13)), and using the continuity proved in [3], Lemma 27, this proves the fact that $\mathbf{X}^{(n)}(|\lambda_n| + \lambda)$ converges in distribution to a random variable which has the same distribution as a $(0, \beta, \mathbf{d})$-multiplicative coalescent at time λ. Also, in [3], Proposition 7, condition (c) is replaced by the condition that

$$\frac{\sigma_3(\mathbf{x}^{(n)})}{\sigma_2(\mathbf{x}^{(n)})^3} \xrightarrow{\mathbb{P}} \sum_{j=1}^{\infty} d_j^3, \tag{7.6}$$

which follows from a combination of (a) and (c). Further, in (a)–(c), we work with *convergence in probability* (as the initial state is a random variable), while in [3], Proposition 7, the initial state is considered to be deterministic. This is a minor change.

In the remainder of this section, we shall show that conditions (a)–(c) hold with $\beta = -\zeta/\mathbb{E}[W]$ and $d_j = c_j/\mathbb{E}[W]$.

Asymptotics of $\sigma_2(\mathbf{x}^{(n)})$. In the following lemma, we state the properties of $\sigma_2(\mathbf{x}^{(n)})$ that we shall rely on. In order to state the result, we recall that

$$\sigma_2(\mathbf{x}^{(n)}) = \sum_j (x_j^{(n)})^2, \tag{7.7}$$

where $x_j^{(n)} = n^{-\rho}\mathcal{W}_{(j)}$, and where the vertex weights are now given by

$$\bar{w}_j(0) = (1 + \lambda_n \ell_n n^{-2\rho}) w_j = w_j(\lambda_n \ell_n n^{-2\rho+\eta}) = w_j(\lambda_n \ell_n / n), \tag{7.8}$$

since $2\rho - \eta = 1$, so that

$$\bar{\mathbf{w}}(0) = \mathbf{w}(\lambda_n \ell_n / n) = \mathbf{w}(\mathbb{E}[W]\lambda_n)(1 + o(1)). \tag{7.9}$$

Now,

$$\sigma_2(\mathbf{x}^{(n)}) = n^{-2\rho} \sum_{j\geq 1} \mathcal{W}_{(j)}^2 = n^{-2\rho} \sum_{j\geq 1} \mathcal{W}_{\leq}(j)^2 \tag{7.10}$$

and, thus, by Lemma 6.2,

$$\sigma_2(\mathbf{x}^{(n)}) = n^{-2\rho} \sum_{j\in[n]} \mathcal{W}_{\leq}(j)^2 = n^{-2\rho} \sum_{j\in[n]} w_j \mathcal{W}(j). \tag{7.11}$$

We continue by investigating the mean and variance of the above sum:

LEMMA 7.1 [Mean and variance of $\sigma_2(\mathbf{x}^{(n)})$]. *When the weights* $\mathbf{w}(\lambda_n)$ *satisfy that* $\nu_n(\lambda_n) < 1 - \lambda_n n^{-\eta}$, *then*:

(i)

$$\mathbb{E}\Big[\sum_{i\in[n]} w_i \mathcal{W}(i)\Big] = \frac{\ell_n}{1-\nu_n(\lambda_n)} + o(n^{2\rho}\lambda_n^{-2}); \tag{7.12}$$

(ii)

$$\begin{aligned} &\operatorname{Var}\Big(\sum_{i\in[n]} w_i \mathcal{W}(i)\Big) \\ &\quad \leq \ell_n \mathbb{E}[w_T^3] \leq C\Big(\mathbb{E}[w_T]^4 \frac{1}{\ell_n} \sum_{i\in[n]} w_i^4 + \mathbb{E}[w_T]^2 \mathbb{E}[w_T^2] \frac{1}{\ell_n} \sum_{i\in[n]} w_i^3\Big). \end{aligned} \tag{7.13}$$

PROOF. (i) We bound

$$\mathbb{E}\Big[\sum_{i\in[n]} w_i \mathcal{W}(i)\Big] \leq \mathbb{E}\Big[\sum_{i\in[n]} w_i w_{T(i)}\Big] = \ell_n \mathbb{E}[w_T] = \frac{\ell_n \nu_n}{1-\nu_n}. \tag{7.14}$$

For the lower bound, we make use of the bound alike in (6.16),

$$\begin{aligned} \mathbb{E}[w_{T(i)} - \mathcal{W}(i)] &\leq \sum_{j\in[n]} \mathbb{E}\big[[w_{T(i)} - \mathcal{W}(i)] \mathbb{1}_{\{\text{mark } j \text{ drawn at least twice}\}}\big] \\ &\leq \sum_{j\in[n]} \mathbb{E}[w_{T(j)}] \sum_{s_1<s_2} \mathbb{P}(\text{mark } j \text{ drawn at times } s_1, s_2). \end{aligned} \tag{7.15}$$

Now, there are two contributions, depending on whether s_2 is in the family tree of s_1 or not. When it is not, then the events $\{\text{mark } j \text{ drawn at time } s_1\}$ and $\{\text{mark } j \text{ drawn at time } s_2\}$ are completely independent, and we arrive at

$$\begin{aligned} \sum_{s_1<s_2} \mathbb{P}(\text{mark } j \text{ drawn at times } s_1, s_2) &= \frac{w_j^2}{\ell_n^2} \sum_{s_1<s_2} \mathbb{P}(T(i) \geq s_1) \\ &\leq \frac{w_j^2}{\ell_n^2} \mathbb{E}[T(i)^2]. \end{aligned} \tag{7.16}$$

When s_2 is in the family tree of s_1, then we obtain the bound

$$\sum_{s_1<s_2} \mathbb{P}(j \text{ chosen at times } s_1, s_2) = \frac{w_j^2}{\ell_n^2} \sum_{s_1<s_2} \mathbb{P}\big(T(i) \geq s_1\big)\mathbb{P}(s_2 \in T_{s_1}), \tag{7.17}$$

where we denote the tree rooted at s_1 by T_{s_1}. Thus, denoting by $|T_{s_1}|$ the number of elements in T_{s_1},

$$\sum_{s_2} \mathbb{P}(s_2 \in T_{s_1}) \leq \mathbb{E}[|T_{s_1}|] = \mathbb{E}[T(j)], \tag{7.18}$$

and we arrive at a contribution of

$$\begin{aligned} \sum_{s_1<s_2} \mathbb{P}(j \text{ chosen at times } s_1, s_2) &\leq \frac{w_j^2}{\ell_n^2} \sum_{s_1<s_2} \mathbb{P}\big(T(i) \geq s_1\big)\mathbb{E}[T(j)] \\ &= \frac{w_j^2}{\ell_n^2}\mathbb{E}[T(i)]\mathbb{E}[T(j)]. \end{aligned} \tag{7.19}$$

Therefore,

$$\begin{aligned} \mathbb{E}\big[w_{T(i)} - \mathcal{W}(i)\big] &\leq \sum_{j\in[n]} \mathbb{E}\big[w_{T(j)}\big]\frac{w_j^2}{\ell_n^2}\big(\mathbb{E}[T(i)^2] + \mathbb{E}[T(i)]\mathbb{E}[T(j)]\big) \\ &= \mathbb{E}[T(i)^2]\frac{1}{1-\nu_n(\lambda_n)} \sum_{j\in[n]} \frac{w_j^3}{\ell_n^2} \\ &\quad + \mathbb{E}[T(i)]\frac{1}{(1-\nu_n(\lambda_n))^2} \sum_{j\in[n]} \frac{w_j^4}{\ell_n^2}. \end{aligned} \tag{7.20}$$

Thus we obtain

$$\begin{aligned} \sum_{i\in[n]} w_i\mathbb{E}[\mathcal{W}(i)] \geq \sum_{i\in[n]} w_i\mathbb{E}\big[w_{T(i)}\big] &- \sum_{i\in[n]} w_i\mathbb{E}[T(i)^2]\frac{1}{1-\nu_n(\lambda_n)} \sum_{j\in[n]} \frac{w_j^3}{\ell_n^2} \\ &- \sum_{i\in[n]} w_i^2 \frac{1}{(1-\nu_n(\lambda_n))^3} \sum_{j\in[n]} \frac{w_j^4}{\ell_n^2}. \end{aligned}$$

We bound

$$\begin{aligned} &\sum_{i\in[n]} w_i\mathbb{E}[T(i)^2] \sum_{j\in[n]} \frac{w_j^3}{\ell_n^2(1-\nu_n(\lambda_n))} \\ &\quad \leq \frac{C}{\ell_n^2(1-\nu_n(\lambda_n))^3}\Big(\sum_{i\in[n]} w_i^3\Big)^2 + C\sum_{i\in[n]} \frac{w_i^2}{(1-\nu_n(\lambda_n))^4} \sum_{j\in[n]} \frac{w_j^3}{\ell_n^2} \\ &\quad \leq C|\lambda_n|^{-3}n^{3\eta-2+6\alpha} + C|\lambda_n|^{-4}n^{4\eta-1+3\alpha}. \end{aligned} \tag{7.21}$$

Now, $3\eta - 2 + 6\alpha = 1 < 2\rho = (\tau - 2)/(\tau - 1)$, since $\tau > 3$, so that the first term is $o(n^{2\rho}/|\lambda_n|^2)$. For the second term $4\eta - 1 + 3\alpha = 2\rho + (\tau - 4)/(\tau - 1) < 2\rho$, so this terms is also $o(n^{2\rho}|\lambda_n|^{-2})$. Similarly,

$$\begin{aligned}\sum_{i\in[n]} w_i^2 \frac{1}{(1-\nu_n(\lambda_n))^3} \sum_{j\in[n]} \frac{w_j^4}{\ell_n^2} &= \nu_n(\lambda_n) \frac{1}{(1-\nu_n(\lambda_n))^3} \sum_{j\in[n]} \frac{w_j^4}{\ell_n} \\ &\le C|\lambda_n|^{-3} n^{3\eta-1+4\alpha}.\end{aligned} \tag{7.22}$$

Again, $3\eta - 1 + 4\alpha = 2(\tau - 3)/(\tau - 1) < 2\rho$, so also this contribution is $o(n^{2\rho}|\lambda_n|^{-2})$.

(ii) We shall start by bounding the second moment. For this, we rewrite

$$\mathbb{E}\biggl[\biggl(\sum_{i\in[n]} w_i \mathcal{W}(i)\biggr)^2\biggr] = \sum_{i_1,i_2} w_{i_1} w_{i_2} \mathbb{E}[\mathcal{W}(i_1)\mathcal{W}(i_2)]. \tag{7.23}$$

Now we split

$$\begin{aligned}\mathbb{E}[\mathcal{W}(i_1)\mathcal{W}(i_2)] &= \mathbb{E}\bigl[\mathcal{W}(i_1)\mathcal{W}(i_2)\mathbb{1}_{\{i_1 \longleftrightarrow i_2\}}\bigr] \\ &\quad + \mathbb{E}\bigl[\mathcal{W}(i_1)\mathcal{W}(i_2)\mathbb{1}_{\{i_1 \not\longleftrightarrow i_2\}}\bigr].\end{aligned} \tag{7.24}$$

By Lemma 6.2(b), the second term is bounded from above by $\mathbb{E}[\mathcal{W}(i_1)]\mathbb{E}[\mathcal{W}(i_2)]$. Therefore, summing over i_1, i_2, we obtain that

$$\begin{aligned}\mathbb{E}\biggl[\biggl(\sum_{i\in[n]} w_i \mathcal{W}(i)\biggr)^2\biggr] &\le \mathbb{E}\biggl[\sum_{i\in[n]} w_i \mathcal{W}(i)\biggr]^2 \\ &\quad + \sum_{i_1,i_2} w_{i_1} w_{i_2} \mathbb{E}\bigl[\mathcal{W}(i_1)\mathcal{W}(i_2)\mathbb{1}_{\{i_1 \longleftrightarrow i_2\}}\bigr],\end{aligned} \tag{7.25}$$

so that

$$\begin{aligned}\mathrm{Var}\biggl(\sum_{i\in[n]} w_i \mathcal{W}(i)\biggr) &= \mathbb{E}\biggl[\biggl(\sum_{i\in[n]} w_i \mathcal{W}(i)\biggr)^2\biggr] - \mathbb{E}\biggl[\sum_{i\in[n]} w_i \mathcal{W}(i)\biggr]^2 \\ &\le \sum_{i_1,i_2} w_{i_1} w_{i_2} \mathbb{E}\bigl[\mathcal{W}(i_1)\mathcal{W}(i_2)\mathbb{1}_{\{i_1 \longleftrightarrow i_2\}}\bigr] \\ &= \sum_{i\in[n]} w_i \mathbb{E}[\mathcal{W}(i)^3] \\ &\le \sum_{i\in[n]} w_i \mathbb{E}[w_{T(i)}^3] = \ell_n \mathbb{E}[w_T^3].\end{aligned} \tag{7.26}$$

The upper bound on $\mathbb{E}[w_T^3]$ follows as in the proof of Lemma 2.3. □

Check of convergence conditions. We conclude that we are left to prove that conditions (a), (b) and (c) in (7.2)–(7.4) hold. We shall prove these conditions in the order (b), (c) and (a), condition (a) being the most difficult one.

Condition (b) follows from (6.1) and condition (a), as we show now. Substituting (7.9) into (6.1), we obtain that

$$x_j^{(0)} = n^{-\rho}\mathcal{W}_{(j)} = \frac{c_j}{\mathbb{E}[W]|\lambda_n|}\big(1+o_{\mathbb{P}}(1)\big) = \big(1+o_{\mathbb{P}}(1)\big)d_j/|\lambda_n|, \tag{7.27}$$

where $d_j = c_j/\mathbb{E}[W]$. Further, the first-order asymptotics in condition (a) proves that $|\lambda_n|\sigma_2(\mathbf{x}^{(n)}) \xrightarrow{\mathbb{P}} 1$, so that the factor $1/\sigma_2(\mathbf{x}^{(n)})$ in condition (b) can be replaced by a multiplication by $|\lambda_n|$. We conclude that $|\lambda_n|x_j^{(0)} \xrightarrow{\mathbb{P}} d_j$, where $d_j = c_j/\mathbb{E}[W]$, as required.

For condition (c), we apply similar ideas and start with

$$|\lambda_n|^3\sigma_3\big(\mathbf{x}^{(n)}\big) = \sum_{j\in[n]}(|\lambda_n|n^{-\rho}\mathcal{W}_{\leq}(j))^3. \tag{7.28}$$

The summands for $j > K$ can be bounded using Lemma 6.2 by

$$\sum_{j>K}(|\lambda_n|n^{-\rho}\mathcal{W}_{\leq}(j))^3 \leq (|\lambda_n|n^{-\rho})^3\sum_{j>K}w_j\mathcal{W}^{[K]}(j)^2, \tag{7.29}$$

which is small in probability by the Markov inequality and (6.31). The summands for $j \leq K$ converge in probability by (6.1). Thus condition (c) follows from (6.1) and (6.31).

We continue with condition (a), which is equivalent to the statement that

$$\sigma_2\big(\mathbf{x}^{(n)}\big) = \frac{1}{|\lambda_n|} - \frac{\beta}{\lambda_n^2} + o_{\mathbb{P}}(|\lambda_n|^{-2}), \tag{7.30}$$

where $\beta = -\zeta/\mathbb{E}[W]$.

We shall prove (7.30) by a second moment method. We first identify, by Lemma 6.2(a),

$$\sigma_2\big(\mathbf{x}^{(n)}\big) = n^{-2\rho}\sum_{j\in[n]}\mathcal{W}_{\leq}(j)^2 = n^{-2\rho}\sum_{i\in[n]}w_i\mathcal{W}(i). \tag{7.31}$$

Thus, in order to prove (7.30), it suffices to show that

$$\mathbb{E}\Big[\sum_{i\in[n]}w_i\mathcal{W}(i)\Big] = n^{2\rho}\Big(|\lambda_n|^{-1} + \frac{\zeta}{\mathbb{E}[W]}|\lambda_n|^{-2} + o(|\lambda_n|^{-2})\Big) \tag{7.32}$$

and

$$\mathrm{Var}\Big(\sum_{i\in[n]}w_i\mathcal{W}(i)\Big) = o(n^{4\rho}|\lambda_n|^{-4}). \tag{7.33}$$

Indeed, by (7.32), we have that, for n sufficiently large,

$$\begin{aligned}\mathbb{P}\biggl(\biggl|\sigma_2(\mathbf{x}^{(n)}) - |\lambda_n|^{-1} - \frac{\zeta}{\mathbb{E}[W]}|\lambda_n|^{-2}\biggr| \geq \varepsilon|\lambda_n|^{-2}\biggr) \\ \leq \mathbb{P}\bigl(\bigl|\sigma_2(\mathbf{x}^{(n)}) - \mathbb{E}[\sigma_2(\mathbf{x}^{(n)})]\bigr| \geq \varepsilon|\lambda_n|^{-2}/2\bigr),\end{aligned} \tag{7.34}$$

which, by the Chebychev inequality is bounded by

$$\begin{aligned}\mathbb{P}\biggl(\biggl|\sigma_2(\mathbf{x}^{(n)}) - |\lambda_n|^{-1} + \frac{\zeta}{\mathbb{E}[W]}|\lambda_n|^{-2}\biggr| \geq \varepsilon|\lambda_n|^{-2}\biggr) &\leq \frac{4|\lambda_n|^4}{\varepsilon^2}\operatorname{Var}(\sigma_2(\mathbf{x}^{(n)})) \\ &= o(1)\end{aligned} \tag{7.35}$$

by (7.33). Thus, (7.30) follows from (7.32) and (7.33).

To prove (7.32), we apply Lemma 7.1, in the setting that

$$\nu_n(\lambda_n) = \nu_n(1 + \lambda_n \ell_n n^{-2\rho}) = 1 + \lambda_n \ell_n n^{-2\rho} + \zeta n^{-\eta} + o(n^{-\eta}), \tag{7.36}$$

so that, by Lemma 7.1(i),

$$\begin{aligned}\mathbb{E}\biggl[\sum_{i\in[n]} w_i \mathcal{W}(i)\biggr] &= \frac{\sum_{i\in[n]} w_i^2}{1-\nu_n(\lambda_n)} + o(n^{2\rho}|\lambda_n|^{-2}) \\ &= \nu_n(\lambda_n)\ell_n\bigl(|\lambda_n|\ell_n n^{-2\rho} - \zeta n^{-\eta} + o(n^{-\eta})\bigr)^{-1} \\ &\quad + o(n^{2\rho}|\lambda_n|^{-2}) \\ &= |\lambda_n|^{-1} n^{2\rho} + \frac{\zeta}{\mathbb{E}[W]} n^{2\rho}|\lambda_n|^{-2} \\ &\quad + o(|\lambda_n|^{-2} n^{2\rho}),\end{aligned} \tag{7.37}$$

which proves (7.32) with $\beta = -\zeta/\mathbb{E}[W]$.

By Lemma 7.1(ii),

$$\begin{aligned}\operatorname{Var}\biggl(\sum_{i\in[n]} w_i \mathcal{W}(i)\biggr) &\leq C\biggl(\mathbb{E}[w_T]^4 \frac{1}{\ell_n}\sum_{i\in[n]} w_i^4 + \mathbb{E}[w_T]^2 \mathbb{E}[w_T^2]\frac{1}{\ell_n}\sum_{i\in[n]} w_i^3\biggr) \\ &= o(n^{4\rho}\lambda_n^{-4}),\end{aligned} \tag{7.38}$$

precisely when both terms in the middle inequality satisfy this bound. We complete the proof by checking these estimates. The first contribution is bounded by

$$\frac{1}{\ell_n(1-\nu_n(\lambda_n))^4}\sum_{i\in[n]} w_i^4 \leq \frac{C}{|\lambda_n|^4} n^{4\alpha+3\eta-1} = o(n^{4\rho}|\lambda_n|^{-4}), \tag{7.39}$$

since $4\alpha + 3\eta - 1 = 2(\tau-2)/(\tau-1) = 2\rho < 4\rho$. The second contribution, instead, is bounded by

$$\frac{1}{\ell_n^2(1-\nu_n(\lambda_n))^5}\biggl(\sum_{i\in[n]} w_i^3\biggr)^2 \leq \frac{C}{|\lambda_n|^5} n^{6\alpha+5\eta-2} = o(n^{4\rho}|\lambda_n|^{-4}), \tag{7.40}$$

since $6\alpha + 5\eta - 2 = (3\tau - 7)/(\tau - 1) < 4\rho = 4(\tau - 2)/(\tau - 1)$. This proves the required concentration for $\sigma_2(\mathbf{x}^{(n)})$ and hence completes the proof of Theorem 1.2 for cluster weights and for any *fixed* λ.

Convergence of the finite-dimensional distributions random graph multiplicative coalescent. So far, we have proved the convergence of $\mathbf{X}^{(n)}(|\lambda_n| + \lambda)$ for a fixed time λ. By [3], Lemma 26, there exists an eternal multiplicative coalescent with the same marginal for every λ. By the strong Feller property of multiplicative coalescents proved in [2], as well as [3], Lemma 27, the convergence of $\mathbf{X}^{(n)}(|\lambda_n| + \lambda_1)$ implies that the future finite-dimensional distributions $(\mathbf{X}^{(n)}(|\lambda_n| + \lambda_l))_{l=1}^{k}$ converge in distribution to the finite-dimensional distributions of the eternal multiplicative coalescent. This completes the proof of the convergence of the finite-dimensional distributions in Theorem 1.2 for cluster weights.

Convergence of cluster sizes from cluster weights. By the adaptation of Theorem 1.1 to cluster weights in Theorem 1.4, we obtain that $\mathcal{W}_{\leq}(j) = |\mathcal{C}_{\leq}(j)|(1 + o_{\mathbb{P}}(1))$, so that the result immediately follows for the cluster sizes.

Acknowledgments. We thank Tom Kurtz for a discussion that helped us to simplify the proof of Corollary 3.4 substantially and to note the extension to the convergence in the uniform topology in Remark 3.8. We thank David Aldous and Vlada Limic for help on their results in [3], which in particular clarified the convergence of finite-dimensional distributions in Theorem 1.2. We thank Sandra Kliem and two anonymous referees for their valuable comments that helped us to substantially improve the presentation.

REFERENCES

[1] ALBERT, R. and BARABÁSI, A.-L. (2002). Statistical mechanics of complex networks. *Rev. Modern Phys.* **74** 47–97. MR1895096

[2] ALDOUS, D. (1997). Brownian excursions, critical random graphs and the multiplicative coalescent. *Ann. Probab.* **25** 812–854. MR1434128

[3] ALDOUS, D. and LIMIC, V. (1998). The entrance boundary of the multiplicative coalescent. *Electron. J. Probab.* **3** 59 pp. (electronic). MR1491528

[4] ALDOUS, D. J. (1999). Deterministic and stochastic models for coalescence (aggregation and coagulation): A review of the mean-field theory for probabilists. *Bernoulli* **5** 3–48. MR1673235

[5] ALON, N. and SPENCER, J. H. (2000). *The Probabilistic Method*, 2nd ed. Wiley, New York. With an appendix on the life and work of Paul Erdős. MR1885388

[6] BERTOIN, J. (1996). *Lévy Processes. Cambridge Tracts in Mathematics* **121**. Cambridge Univ. Press, Cambridge. MR1406564

[7] BERTOIN, J. (2006). *Random Fragmentation and Coagulation Processes. Cambridge Studies in Advanced Mathematics* **102**. Cambridge Univ. Press, Cambridge. MR2253162

[8] BHAMIDI, S., VAN DER HOFSTAD, R. and VAN LEEUWAARDEN, J. S. H. (2010). Scaling limits for critical inhomogeneous random graphs with finite third moments. *Electron. J. Probab.* **15** 1682–1703. MR2735378

[9] BILLINGSLEY, P. (1999). *Convergence of Probability Measures*, 2nd ed. Wiley, New York. MR1700749

[10] BOLLOBÁS, B. (2001). *Random Graphs*, 2nd ed. *Cambridge Studies in Advanced Mathematics* **73**. Cambridge Univ. Press, Cambridge. MR1864966

[11] BOLLOBÁS, B., JANSON, S. and RIORDAN, O. (2007). The phase transition in inhomogeneous random graphs. *Random Structures Algorithms* **31** 3–122. MR2337396

[12] BRITTON, T., DEIJFEN, M. and MARTIN-LÖF, A. (2006). Generating simple random graphs with prescribed degree distribution. *J. Stat. Phys.* **124** 1377–1397. MR2266448

[13] CHUNG, F. and LU, L. (2002). The average distances in random graphs with given expected degrees. *Proc. Natl. Acad. Sci. USA* **99** 15879–15882 (electronic). MR1944974

[14] CHUNG, F. and LU, L. (2002). Connected components in random graphs with given expected degree sequences. *Ann. Comb.* **6** 125–145. MR1955514

[15] CHUNG, F. and LU, L. (2003). The average distance in a random graph with given expected degrees. *Internet Math.* **1** 91–113. MR2076728

[16] CHUNG, F. and LU, L. (2006). *Complex Graphs and Networks*. *CBMS Regional Conference Series in Mathematics* **107**. Conference Board of the Mathematical Sciences, Washington, DC. MR2248695

[17] CHUNG, F. and LU, L. (2006). The volume of the giant component of a random graph with given expected degrees. *SIAM J. Discrete Math.* **20** 395–411 (electronic). MR2257269

[18] DOROGOVTSEV, S. N. and MENDES, J. F. F. (2002). Evolution of networks. *Adv. Phys.* **51** 1079–1187.

[19] DURRETT, R. (2007). *Random Graph Dynamics*. Cambridge Univ. Press, Cambridge. MR2271734

[20] GRIMMETT, G. (1999). *Percolation*, 2nd ed. *Grundlehren der Mathematischen Wissenschaften* [*Fundamental Principles of Mathematical Sciences*] **321**. Springer, Berlin. MR1707339

[21] GRIMMETT, G. R. and STIRZAKER, D. R. (2001). *Probability and Random Processes*, 3rd ed. Oxford Univ. Press, New York. MR2059709

[22] HATAMI, H. and MOLLOY, M. (2009). The scaling window for a random graph with a given degree sequence. Available at arXiv:0907.4211.

[23] JACOD, J. and SHIRYAEV, A. N. (2003). *Limit Theorems for Stochastic Processes*, 2nd ed. *Grundlehren der Mathematischen Wissenschaften* [*Fundamental Principles of Mathematical Sciences*] **288**. Springer, Berlin. MR1943877

[24] JANSON, S. (2008). The largest component in a subcritical random graph with a power law degree distribution. *Ann. Appl. Probab.* **18** 1651–1668. MR2434185

[25] JANSON, S. (2010). Asymptotic equivalence and contiguity of some random graphs. *Random Structures Algorithms* **36** 26–45. MR2591045

[26] JANSON, S., ŁUCZAK, T. and RUCINSKI, A. (2000). *Random Graphs*. Wiley, New York. MR1782847

[27] JOSEPH, A. (2010). The component sizes of a critical random graph with pre-described degree sequence. Preprint.

[28] KALLENBERG, O. (2002). *Foundations of Modern Probability*, 2nd ed. Springer, New York. MR1876169

[29] KYPRIANOU, A. E. (2006). *Introductory Lectures on Fluctuations of Lévy Processes with Applications*. Springer, Berlin. MR2250061

[30] NEWMAN, M. E. J. (2003). The structure and function of complex networks. *SIAM Rev.* **45** 167–256 (electronic). MR2010377

[31] NORROS, I. and REITTU, H. (2006). On a conditionally Poissonian graph process. *Adv. in Appl. Probab.* **38** 59–75. MR2213964

[32] THORISSON, H. (2000). *Coupling, Stationarity, and Regeneration*. Springer, New York. MR1741181

[33] TUROVA, T. S. (2009). Diffusion approximation for the components in critical inhomogeneous random graphs of rank 1. Preprint.

[34] VAN DER HOFSTAD, R. (2009). Critical behavior in inhomogeneous random graphs. Preprint.

[35] VAN DER HOFSTAD, R. (2011). Random graphs and complex networks. Unpublished manuscript. Available at http://www.win.tue.nl/~rhofstad/NotesRGCN.pdf.

[36] VAN DER HOFSTAD, R., JANSON, S. and LUCZAK, M. The near-critical behavior for the configuration model with finite-variance degrees. Unpublished manuscript.

S. BHAMIDI
DEPARTMENT OF STATISTICS
AND OPERATIONS RESEARCH
UNIVERSITY OF NORTH CAROLINA
304 HANES HALL
CHAPEL HILL, NORTH CAROLINA 27599
USA
E-MAIL: bhamidi@email.unc.edu

R. VAN DER HOFSTAD
J. S. H. VAN LEEUWAARDEN
DEPARTMENT OF MATHEMATICS
AND COMPUTER SCIENCE
EINDHOVEN UNIVERSITY OF TECHNOLOGY
P.O. BOX 513
5600 MB EINDHOVEN
THE NETHERLANDS
E-MAIL: rhofstad@win.tue.nl
j.s.h.v.leeuwaarden@tue.nl