Check for updates

Geometry of the minimal spanning tree in the heavy-tailed regime: new universality classes

Shankar Bhamidi[1] · Sanchayan Sen[2]

Abstract

A well-known open problem on the behavior of optimal paths in random graphs in the strong disorder regime, formulated by statistical physicists, and supported by a large amount of numerical evidence over the last decade (Braunstein et al. in Phys Rev Lett 91(16):168701, 2003; Braunstein et al. in Int J Bifurc Chaos 17(07):2215–2255, 2007; Chen et al. in Phys Rev Lett 96(6):068702, 2006; Wu et al. in Phys Rev Lett 96(14):148702, 2006) is as follows: for a large class of random graph models with degree exponent $\tau \in (3, 4)$, distances in the minimal spanning tree (MST) on the giant component in the supercritical regime scale like $n^{(\tau-3)/(\tau-1)}$. The aim of this paper is to make progress towards a proof of this conjecture. We consider a supercritical inhomogeneous random graph model with degree exponent $\tau \in (3, 4)$ that is closely related to Aldous's multiplicative coalescent, and show that the MST constructed by assigning i.i.d. continuous weights to the edges in its giant component, endowed with the tree distance scaled by $n^{-(\tau-3)/(\tau-1)}$, converges in distribution with respect to the Gromov–Hausdorff topology to a random compact real tree. Further, almost surely, every point in this limiting space either has degree one (leaf), or two, or infinity (hub), both the set of leaves and the set of hubs are dense in this space, and the Minkowski dimension of this space equals $(\tau - 1)/(\tau - 3)$. The multiplicative coalescent, in an asymptotic sense, describes the evolution of the component sizes of various near-critical random graph processes. We expect the limiting spaces in this paper to be the candidates for the scaling limit of the MST constructed for a wide array of other heavy-tailed random graph models.

✉ Sanchayan Sen
sanchayan.sen1@gmail.com

Shankar Bhamidi
bhamidi@email.unc.edu

[1] Department of Statistics and Operations Research, University of North Carolina, Chapel Hill, USA

[2] Department of Mathematics, Indian Institute of Science, Bengaluru, India

Contents

1 Introduction

Consider a finite, connected, and weighted graph (V, E, b), where (V, E) is the underlying graph and $b : E \to [0, \infty)$ is the weight function. A spanning tree of (V, E) is a tree that is a subgraph of (V, E) with vertex set V. A minimal spanning tree (MST) T of (V, E, b) satisfies

$$\sum_{e \in T} b(e) = \min \left\{ \sum_{e \in T'} b(e) : T' \text{ is a spanning tree of } (V, E) \right\}. \qquad (1)$$

The MST is one of the most studied functionals in combinatorial optimization. Studying this object when the edge weights and potentially the underlying graph is random has stimulated an enormous body of work in probabilistic combinatorics and geometric probability. The papers [9, 13, 15, 17, 19, 20, 22, 38, 49–52, 56, 59, 66–69, 72] and the references therein give a non-exhaustive account of the enormous literature on the probabilistic study of MSTs.

We are interested in the global geometric properties of the MST, e.g., the diameter and the typical distance. In the early 2000s, several major conjectures were made about the intrinsic geometry of MSTs in the statistical physics community where this object arises as models of disordered networks, but until very recently, there were few rigorous mathematical results on this problem. Consider a finite connected graph $G = (V, E)$, and assign costs $\exp(\beta \varepsilon_e)$ to the edges $e \in E$ where $\beta > 0$. An optimal path P in G between $u, v \in V$ minimizes the total cost $\sum_{e \in P'} \exp(\beta \varepsilon_e)$ among all paths P' between u and v. This model interpolates between the first passage percolation regime (the weak disorder regime) and the minimal spanning tree regime (strong disorder regime). Assuming that ε_e, $e \in E$, are pairwise distinct, it is easy to see that for sufficiently large values of β, the optimal path between any two vertices u and v is the path P that minimizes the maximal edge weight $\max_{e \in P'} \varepsilon_e$ among all paths P' connecting u and v in G. It is well-known (see Lemma 4.3) that this is simply the path connecting u and v in the MST of G constructed using the weights ε_e, $e \in E$. Thus, the number of edges or the hopcount l_{opt} in the optimal path between two typical vertices in the presence of strong disorder is simply the length of the path in the MST connecting two typical vertices.

Motivated by the availability of data on a host of real-world networks as well as the impact of complex networks in our daily lives, the last few years have witnessed an explosion in the formulation and study of mathematical models of networks. These models try to capture properties of networks observed empirically. Of relevance to us is the heavy-tailed nature of the empirical degree distribution. To study this phenomenon, a plethora of random graph models have been proposed that have heavy-tailed degree distributions with some degree exponent $\tau \in (1, \infty]$. A precise definition of the degree exponent will not be needed in the sequel, so we instead refer the reader to [45, 54, 55] for a detailed discussion on the random graph models now available to practitioners.

Coming back to disordered networks, in the 2000s, statistical physicists predicted [32, 33, 39, 73] that if the underlying graph possesses a heavy-tailed degree distribution with exponent τ, then in the presence of strong disorder, l_{opt} exhibits the following scaling behavior:

$$l_{opt} \sim \begin{cases} n^{1/3}, & \text{if } \tau > 4, \\ n^{(\tau-3)/(\tau-1)}, & \text{if } \tau \in (3, 4), \end{cases} \tag{2}$$

where n denotes the number of vertices in the underlying graph, and further, such behavior should be universal, i.e., in principle should apply to a wide array of random graph models.

The above conjecture is related to the universality of the intrinsic geometry of the MST, proving which in full generality remains open to date. Only recently, there has been some progress in the $\tau > 4$ regime. In [6], it was shown that the MST

of the complete graph on n vertices constructed using i.i.d. Unif[0, 1] edge weights, endowed with the tree distance scaled by $n^{-1/3}$ and the uniform probability measure on the vertices, converges in distribution to a random compact $\mathbb{R}$-tree $\mathscr{M}$. Further, almost surely, $\mathscr{M}$ is binary and the Minkowski dimension of $\mathscr{M}$ equals 3. This result can be seen as a confirmation of (2) for the case $\tau = \infty$. The limiting space $\mathscr{M}$ is expected to be the scaling limit of the MST of a variety of random discrete structures including random graphs with degree exponent $\tau > 4$. A first step in this broader program of establishing universality of $\mathscr{M}$ was taken in [8], where it was shown that the scaling limit of the MST of the random 3-regular (simple) graph as well as the 3-regular configuration model, with the tree distance scaled by $n^{-1/3}$, is $6^{1/3} \cdot \mathscr{M}$.

On the other hand, the problem has stayed completely open in the $\tau \in (3, 4)$ regime. The aim of this work is to study the MST in this regime and prove the existence of the scaling limit of the MST. We will consider a supercritical inhomogeneous random graph (IRG) model that corresponds to the rank-1 case of the general class of IRGs studied in [30]. We will show that under certain assumptions, the MST on the giant component of this random graph, endowed with the tree distance scaled by $n^{-(\tau-3)/(\tau-1)}$, converges in distribution with respect to the Gromov–Hausdorff topology to a random compact $\mathbb{R}$-tree. We will then study the topological properties of the scaling limit. We will show that almost surely, every point in this limiting space either has degree one (leaf), or two, or infinity (hub), and that both the set of leaves and the set of hubs are dense in this space. Further, the Minkowski dimension of this space equals $(\tau - 1)/(\tau - 3)$ almost surely. Note the contrasting characteristics of such a space in comparison to the space $\mathscr{M}$.

The inhomogeneous random graph model considered in this paper is of special interest as it is closely related to the multiplicative coalescent [10, 11]. The evolution of a large class of dynamic random graph models around the point of phase transition can be well-approximated [24, 26] by the multiplicative coalescent, and we expect that this fact can be leveraged to establish universality of the scaling limits obtained in this paper. We defer further discussion regarding the general program to establish universality to Sect. 8. Our main result proves convergence of the MST viewed as a metric space, which in particular implies distributional convergence of the diameter of the MST rescaled by $n^{-(\tau-3)/(\tau-1)}$. This presents some rigorous evidence supporting the prediction in (2), although the asymptotic behavior of the typical distance claimed in (2) does not follow from this alone. The scaling limit of the typical distance can be deduced if one were able to establish distributional convergence of the MST viewed as a metric measure space. This strengthening of our result can be achieved if one additional estimate is proved. This will also be discussed in Sect. 8.

1.1 Organization of the paper

A reasonable amount of notation regarding notions of convergence of metric space-valued random variables as well as $\mathbb{R}$-trees is required. To quickly get to the main result, we first define the random graph models of interest in Sect. 2, and then describe the main result in Sect. 3. Various definitions and preparatory results are then given in Sect. 4. In Sect. 5, we show that it is equivalent to work with a modified random graph

model; this will make many of our calculations easier. A key step in the proof of our main result is relating the MST in the supercritical graph to the MST in the critical window; this is accomplished in Sect. 6. Then we complete the proof of the main result in Sect. 7. In Sect. 8 we discuss possible extensions and how the scaling limit obtained here can be shown to be universal. We also discuss how our main result can be strengthened to convergence with respect to the Gromov–Haussdorff–Prokhorov topology. A few technical proofs are relegated to Appendix A.

2 Random graph models

We start by describing the class of random graph models of interest for this paper. Start with the vertex set $[n] := \{1, 2, \ldots, n\}$, and suppose each vertex $i \in [n]$ has a weight $w_i^{(n)} \geq 0$ attached to it; intuitively this measures the propensity or attractiveness of this vertex in the formation of links. We assume that the vertices are labeled so that $w_1^{(n)} \geq w_2^{(n)} \geq \cdots \geq w_n^{(n)}$. Write $\boldsymbol{w}^{(n)} = (w_1^{(n)}, \ldots, w_n^{(n)})$, and let

$$\bar{\boldsymbol{q}}_{ij} := \bar{\boldsymbol{q}}_{ij}(\boldsymbol{w}^{(n)}) = \frac{w_i^{(n)} w_j^{(n)}}{L_n} \wedge 1, \qquad 1 \leq i \neq j \leq n, \tag{3}$$

where L_n is the total weight given by

$$L_n := \sum_{i \in [n]} w_i^{(n)}. \tag{4}$$

Now construct a random graph on $[n]$ by placing an edge between i and j independently for each $i < j \in [n]$ with probability $\bar{\boldsymbol{q}}_{ij}$. This random graph model corresponds to the rank-1 case of the general class of IRGs studied by Bollobás, Janson, and Riordan [30]. We will denote this random graph by $\bar{\boldsymbol{G}}_n$. A closely related model [30, 65] is obtained by using the edge connection probabilities

$$\boldsymbol{q}_{ij} = \boldsymbol{q}_{ij}(\boldsymbol{w}^{(n)}) := 1 - \exp(-w_i^{(n)} w_j^{(n)}/L_n), \tag{5}$$

and will be denoted by $\boldsymbol{G}_n$. In the regime of interest for this paper, as shown in [58], this model is equivalent to the Chung-Lu model [40–43] and the Britton–Deijfen–Martin-Löf model [34]. Probabilists will be more familiar with the above random graph via its connection to one of the most famous stochastic coalescent models–the multiplicative coalescent. We describe only a special case and refer the interested reader to [10, 11, 14, 21] for more general constructions of this specific model as well as other coalescent models.

Definition 2.1 *(Finite state multiplicative coalescent)* Fix a finite vertex set $\mathscr{V}$ and a collection of nonnegative vertex weights $\boldsymbol{x} = (x_v \,;\, v \in \mathscr{V})$. Start at time zero with each vertex $v \in \mathscr{V}$ being in a separate cluster of size x_v. At any time $t \geq 0$ we will have a collection of connected clusters with the weight of a cluster $\mathscr{C}$ given by $\mathscr{W}_{\boldsymbol{x}}(\mathscr{C}) = \sum_{v \in \mathscr{C}} x_v$. Consider the continuous time Markov chain where two existing
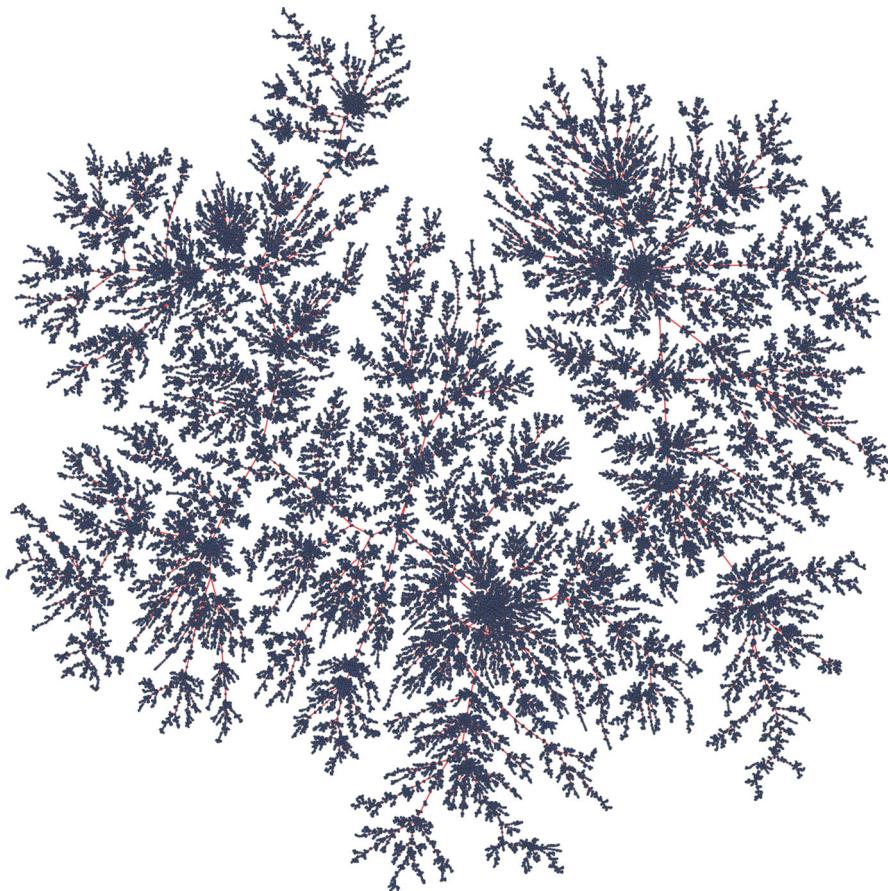
Fig. 1 The MST on the component of the vertex 1 in $\overline{G}_n$ with $w_i = 3(n/i)^\alpha$, where $\tau = 3.05$ and $n = 80000$

distinct clusters $\mathscr{C}_a$ and $\mathscr{C}_b$ merge at rate $\mathscr{W}_{\boldsymbol{x}}(\mathscr{C}_a) \cdot \mathscr{W}_{\boldsymbol{x}}(\mathscr{C}_b)$ into a single cluster of size $\mathscr{W}_{\boldsymbol{x}}(\mathscr{C}_a) + \mathscr{W}_{\boldsymbol{x}}(\mathscr{C}_b)$. Write $\big(\mathrm{MC}\,\big((\mathscr{V}, \boldsymbol{x}), t\big);\ t \geq 0 \big)$ for this process.

Associated with the above dynamics is the following random graph related to $\boldsymbol{G}_n$.

Definition 2.2 *(Random graph* $\mathscr{G}\big((\mathscr{V}, \boldsymbol{x}), t\big)$*)* Consider a finite set $\mathscr{V}$, nonnegative weights $\big(x_v;\ v \in \mathscr{V}\big)$, and $t \geq 0$. Let $\mathscr{G}\big((\mathscr{V}, \boldsymbol{x}), t\big)$ be the random graph on vertex set $\mathscr{V}$ obtained by independently placing edges between pairs of vertices $u, v \in \mathscr{V}$ with probability $1 - \exp(-t x_v x_u)$.

The following lemma is easy to check from the above description of the dynamics.

Lemma 2.3 *For any $t \geq 0$, the ordered sequence of weights of the connected components of $\mathscr{G}((\mathscr{V}, \boldsymbol{x}), t)$ has the same distribution as the ordered sequence of weights of clusters of* $\mathrm{MC}((\mathscr{V}, \boldsymbol{x}), t)$.
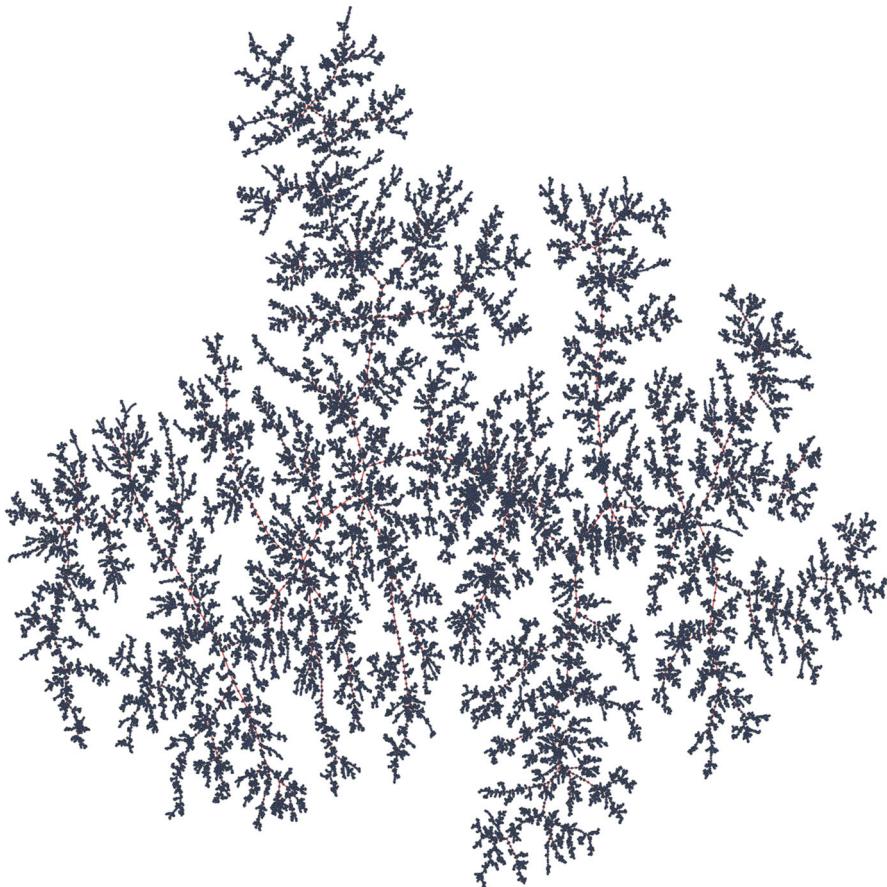
Fig. 2 The MST on the component of the vertex 1 in $\overline{G}_n$ with $w_i = 3(n/i)^\alpha$, where $\tau = 3.95$ and $n = 80000$

The random graph $\mathcal{G}$ is essentially the same as G_n with the parameters expressed in a different way. This connection between the random graph G_n and the multiplicative coalescent will play a major role in our proofs.

We now specify how the vertex weights $\boldsymbol{w}^{(n)}$ are chosen. For the rest of this paper, we will work with a fixed exponent $\tau \in (3, 4)$. We will use the following notation for constants associated to this exponent:

$$\alpha := \frac{1}{\tau - 1}, \qquad \rho := \frac{\tau - 2}{\tau - 1}, \quad \text{and} \quad \eta := \frac{\tau - 3}{\tau - 1}. \tag{6}$$

Assumption 2.4 Consider a sequence $\big(\boldsymbol{w}^{(n)}; n \geq 1\big)$ of weight sequences, where $\boldsymbol{w}^{(n)} = \big(w_1^{(n)}, w_2^{(n)}, \ldots, w_n^{(n)}\big)$ with $w_1^{(n)} \geq w_2^{(n)} \geq \ldots \geq w_n^{(n)} > 0$. Then we assume the following:

(i) **Supercriticality condition:** Let L_n be as in (4). Then

$$\nu := \liminf_{n \to \infty} \frac{\sum_{j=1}^n (w_j^{(n)})^2}{L_n} > 1 \, .$$

(ii) For each $i \geq 1$, there exists $\theta_i^* > 0$ such that

$$\lim_{n \to \infty} \frac{w_i^{(n)}}{n^\alpha} \cdot \left(\frac{n}{\sum_{j=1}^n (w_j^{(n)})^2} \right)^{1/2} = \theta_i^* \, .$$

(iii) There exist constants $A_1, A_2 \in (0, \infty)$ such that for all $n \geq 2$ and $1 \leq i \leq n/2$,

$$A_1 \left(\frac{n}{i} \right)^\alpha \leq w_i^{(n)} \leq A_2 \left(\frac{n}{i} \right)^\alpha \, .$$

(iv) For all $n \geq 2$, $w_n^{(n)} \geq A_1 \left(\log n \right)^{\frac{3}{2}} \cdot n^{-\frac{\eta}{4}} \, .$

Assumption 2.4 (i) ensures supercriticality of the random graph model, and Assumption 2.4 (ii) corresponds to the condition in [11, Display (19)]. It should be possible to relax Assumption 2.4 (iii) and (iv) with a more intricate analysis; however, we do not pursue it here. We will discuss this briefly in Sect. 8.

Note that Assumption 2.4 implies that for all $n \geq 2$ and $n/2 \leq i \leq n$,

$$w_i^{(n)} \leq w_{n/2}^{(n)} \leq 2^\alpha A_2 \leq 2^\alpha A_2 \left(\frac{n}{i} \right)^\alpha \, . \tag{7}$$

Write

$$\mathcal{P}mtr := \left(A_1, A_2, \tau, \nu \right) \tag{8}$$

for the parameters in Assumption 2.4. The following lemma is easy to verify and gives two natural settings that give rise to weights satisfying the above condition. The proof is omitted.

Lemma 2.5 *Suppose F is a cumulative distribution function (cdf) with support in $[0, \infty)$ such that for some $\tau \in (3, 4)$, $\beta_F > 4/\eta$, and $c_F \in (0, \infty)$,*

$$\limsup_{x \downarrow 0} \left(x^{-\beta_F} F(x) \right) < \infty, \qquad \lim_{x \to \infty} x^{\tau - 1} [1 - F(x)] = c_F \, , \quad and$$

$$\int_0^\infty x^2 F(dx) > \int_0^\infty x F(dx) \, .$$

Consider $\mathbf{w}^{(n)} = \left(w_i^{(n)}, \, 1 \leq i \leq n \right)$ obtained via one of the following:

(a) *Let $w_i^{(n)} = [1 - F]^{-1} \left(i/(n + 1) \right)$ for $1 \leq i \leq n$.*
(b) *Let $W_1, \ldots, W_n$ be i.i.d. random variables with cdf F, and let $w_1^{(n)} \geq \ldots \geq w_n^{(n)}$ be the corresponding ordered values.*

Then under (a), $(\boldsymbol{w}^{(n)}, n \geq 1)$ satisfies Assumption 2.4. Under (b), we can construct $\boldsymbol{w}^{(n)}, n \geq 1$, on the same probability space such that Assumption 2.4 is satisfied almost surely in this space with a deterministic v, and random A_1, A_2, and $\theta_i^*, i \geq 1$.

Write $\mathscr{C}(1, \boldsymbol{G}_n)$ (resp. $\mathscr{C}(1, \bar{\boldsymbol{G}}_n)$) for the component of vertex 1 in the graph $\boldsymbol{G}_n$ (resp. $\bar{\boldsymbol{G}}_n$). For any component $\mathscr{C}$ of $\boldsymbol{G}_n$ or $\bar{\boldsymbol{G}}_n$, write $\mathscr{W}(\mathscr{C}) = \sum_{i \in \mathscr{C}} w_i^{(n)}$. The following result states some basic properties of these two random graphs. The notations $O_P(\cdot)$ and $\Theta_P(\cdot)$ used in the sequel are as explained at the end of Sect. 4.1.

Proposition 2.6 *The following hold under Assumption 2.4:*

(a) *We have, $\mathscr{W}(\mathscr{C}(1, \boldsymbol{G}_n)) = \Theta_P(n)$, and*

$$\max \left\{ \mathscr{W}(\mathscr{C}) : \mathscr{C} \text{ component in } \boldsymbol{G}_n \text{ and } \mathscr{C} \neq \mathscr{C}(1, \boldsymbol{G}_n) \right\} = O_P(\log n). \quad (9)$$

We call $\mathscr{C}(1, \boldsymbol{G}_n)$ the giant component in $\boldsymbol{G}_n$. Further, an analogous result holds for the random graph $\bar{\boldsymbol{G}}_n$.

(b) *By [30, Theorem 3.13], if further the empirical distribution $n^{-1} \sum_{i \in [n]} \delta\{w_i^{(n)}\}$ converges to a cdf with tail exponent $\tau \in (3, 4)$, then the degree distribution of $\boldsymbol{G}_n$ (resp. $\bar{\boldsymbol{G}}_n$) converges in probability to a deterministic distribution with tail exponent τ.*

The claim that $\mathscr{W}(\mathscr{C}(1, \boldsymbol{G}_n)) = \Theta_P(n)$ follows from Proposition 6.8, (42), and Lemma 5.1. The proof of (9) is similar to that of Lemma 6.26. Proposition 2.6 (b) is not important for this study, but it gives some justification as to why these models are used to understand real world systems. We now define the central object of interest in this paper.

Definition 2.7 *(Minimal spanning tree)* Let $U = (U_{ij} ; 1 \leq i < j \leq n)$ be a collection of i.i.d. Unif [0, 1] random variables, and let $U_{ji} = U_{ij}$ for $1 \leq i < j \leq n$. Let $\boldsymbol{M}^n$ denote the minimal spanning tree on the giant component of $\boldsymbol{G}_n$ using the edge weights $(U_{ij} ; \{i, j\}$ is an edge in $\mathscr{C}(1, \boldsymbol{G}_n))$. Define $\bar{\boldsymbol{M}}^n$ in analogous way.

We make a convention here that we will follow throughout this paper:

When a finite connected graph H is viewed as a metric space, the underlying set will be the collection of vertices in H joined by line segments of unit length that represent the edges in H, and the distance between two points will be the minimum of the lengths of paths connecting the two points. For any metric space (X, d) and $a > 0$, $a \cdot X$ will denote the metric space $(X, a \cdot d)$, i.e, the space where the distance is scaled by a.

Using this convention, we can view $\boldsymbol{M}^n$ and $\bar{\boldsymbol{M}}^n$ as (random) compact metric spaces.

3 Main result

Recall that the lower and upper box counting dimensions of a compact metric space X are given by

$$\underline{\dim}(X) := \liminf_{\delta \downarrow 0} \frac{\log [\mathcal{N}(X, \delta)]}{\log(1/\delta)}, \quad \text{and}$$

$$\overline{\dim}(X) := \limsup_{\delta \downarrow 0} \frac{\log [\mathcal{N}(X, \delta)]}{\log(1/\delta)}$$

respectively, where $\mathcal{N}(X, \delta)$ is the minimum number of closed balls with radius δ required to cover X. If $\underline{\dim}(X) = \overline{\dim}(X)$, then the box-counting dimension or the Minkowski dimension of X exists and equals this common value. For an $\mathbb{R}$-graph X and any point $x \in X$, let $\deg(x ; X)$ denote the degree of x in X; see Sect. 4.3 for the relevant definitions. We write

$$\mathcal{L}(X) := \{x \in X : \deg(x ; X) = 1\}, \quad \text{and}$$
$$\mathcal{H}(X) := \{x \in X : \deg(x ; X) = \infty\} \tag{10}$$

for the set of leaves and the set of hubs in X respectively. Recall the notation from (6).

Theorem 3.1 *Under Assumption* 2.4 *on the weight sequence, there exists a random compact $\mathbb{R}$-tree $\mathscr{M}^{\boldsymbol{\theta}^*}$ whose law depends only on $\boldsymbol{\theta}^* := (\theta_i^* ; i \geq 1)$ such that*

$$n^{-\eta} \cdot \boldsymbol{M}^n \xrightarrow{d} \mathscr{M}^{\boldsymbol{\theta}^*}, \quad \text{as} \quad n \to \infty, \tag{11}$$

with respect to the Gromov–Hausdorff topology. Further, almost surely,

(a) $\deg\left(x; \mathscr{M}^{\boldsymbol{\theta}^*}\right) \in \{1, 2, \infty\}$ *for all $x \in \mathscr{M}^{\boldsymbol{\theta}^*}$;*
(b) *both the set of leaves $\mathcal{L}(\mathscr{M}^{\boldsymbol{\theta}^*})$ and the set of hubs $\mathcal{H}(\mathscr{M}^{\boldsymbol{\theta}^*})$ are dense in $\mathscr{M}^{\boldsymbol{\theta}^*}$; and*
(c) *the Minkowski dimension of $\mathscr{M}^{\boldsymbol{\theta}^*}$ satisfies*

$$\dim\left(\mathscr{M}^{\boldsymbol{\theta}^*}\right) = \frac{1}{\eta} = \frac{\tau - 1}{\tau - 3}.$$

Moreover, (11) *continues to hold under Assumption* 2.4 *if we replace $\boldsymbol{M}^n$ by $\bar{\boldsymbol{M}}^n$.*

It should be possible to lift the Gromov–Hausdorff convergence in (11) to convergence with respect to the Gromov–Hausdorff–Prokhorov topology. For this, it would suffice to prove an additional technical condition. We will discuss this further in Sect. 8.

4 Definitions and preliminary results

4.1 Notation

Throughout this paper, C, C' etc. will denote constants that depend only on $\mathcal{P}mtr$ as defined in (8), and their values may change from line to line. Special constants will be indexed by the relevant equations, e.g., C_{45} etc., and their values will depend only on $\mathcal{P}mtr$ unless specified otherwise. If a constant depends on any parameter other than $\mathcal{P}mtr$, then that will be explicitly mentioned when the constant is first introduced. For example, $C_{21} = C_{21}(\Delta)$ appearing in the statement of Theorem 6.1 below depends on $\mathcal{P}mtr$ and Δ.

A claim holds 'for all large n' or '$\bar{\forall}n$' will imply that there exists $n_0 \geq 1$ depending only on $\mathcal{P}mtr$ such that the claim holds for all $n \geq n_0$. If the threshold depends on any other parameter, then that will be explicitly mentioned, and it will be assumed, without explicit mention, that such a threshold is chosen bigger than all thresholds involving n that depend only on $\mathcal{P}mtr$ and were previously introduced in the proof. For example, $n_{46} = n_{46}(\kappa)$ appearing around (46) depends on $\mathcal{P}mtr$ and κ, and although not explicitly mentioned, is chosen so that it is bigger than the threshold involving n above which (41) and (42) hold.

Similarly, all thresholds involving λ or ε will depend only on $\mathcal{P}mtr$ unless specified otherwise when the threshold is first introduced.

A relation of the form $a \asymp b$ will mean that there exist C, $C' > 0$ depending only on $\mathcal{P}mtr$ such that $Ca \leq b \leq C'b$. Here, a, b could be elements of two sequences, or two functions defined on the same domain. As an example, consider the following claim made around (36) below: "...for all large n, $i_\lambda(u) \asymp u^{1/\alpha}$ for $(\lambda, u) \in I^{(n),2}$." This statement can be rewritten as follows: There exist C, $C' > 0$, and $n_0 \geq 1$ depending only on $\mathcal{P}mtr$ such that $Cu^{1/\alpha} \leq i_\lambda(u) \leq C'u^{1/\alpha}$ for all $n \geq n_0$ and $(\lambda, u) \in I^{(n),2}$.

For a set S, we use $|S|$ or $\#S$ to denote the number of elements in S. For any graph H, we write $V(H)$ and $E(H)$ for the set of vertices and the set of edges of H respectively. We write $|H|$ for the number of vertices in H, i.e., $|H| = |V(H)|$. For any finite connected graph $H = (V, E)$, we write $\mathrm{sp}(H)$ for the number of surplus edges in H, i.e., $\mathrm{sp}(H) := |E| - |V| + 1$. For a finite (not necessarily connected) graph H, we let

$$\max \mathrm{sp}(H) = \max \left\{ \mathrm{sp}(\mathscr{C}) \; : \; \mathscr{C} \text{ connected component of } H \right\}. \tag{12}$$

Similarly, we write $\mathrm{diam}(H)$ for the maximum of the diameters (with respect to graph distance) of all the components in H. We will write $\mathrm{LP}(H)$ to denote the length of the longest self-avoiding path in H. For a rooted tree T, $\mathrm{ht}(T)$ will denote the height of T, i.e., the distance to the farthest leaf from the root of T.

For a graph $H = (V, E)$ and $v \in V$, we write $\mathscr{C}(v, H)$ to denote the component of v in H. For $V' \subseteq V$, we write $H \setminus V'$ to denote the restriction of H to the vertex set $V \setminus V'$. If the vertices in H have weights $\boldsymbol{y} = (y_v; \, v \in V)$ associated to them, then for any subgraph H_0 in H, we write $\mathscr{W}_{\boldsymbol{y}}(H_0) = \sum_{v \in V(H_0)} y_v$ for the weight or mass of H_0 as measured by the prescribed vertex weights. When the weight sequence is $\boldsymbol{w}$, we omit the subscript and simply write $\mathscr{W}(\cdot)$.

For two real valued random variables X, Y we write $X \preceq_{\mathsf{sd}} Y$ for the stochastic domination relation between the distributions of these random variables. We use $\mathrm{Poi}(\cdot)$, $\mathrm{Unif}(\cdot)$ and $\mathrm{Bernoulli}(\cdot)$ to respectively denote Poisson, uniform, and Bernoulli distributions with parameters in $\cdot$ that will be specified in the setting of interest.

For a non-negative 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 $g(n) = O(f(n))$. We let $\xrightarrow{\mathrm{d}}$, $\xrightarrow{\mathrm{P}}$, and $\xrightarrow{\mathrm{a.s.}}$ respectively denote convergence in distribution, convergence in probability, and almost sure convergence. For a sequence of random variables $(X_n; \, n \geq 1)$ and a sequence of positive real numbers $(b_n; \, n \geq 1)$, we write $X_n = O_P(b_n)$ if the sequence of laws of $(X_n/b_n; \, n \geq 1)$ is tight, and $X_n = o_{\mathrm{P}}(b_n)$ if $X_n/b_n \xrightarrow{\mathrm{d}} 0$ as $n \to \infty$. Further, we write $X_n = \Theta_P(b_n)$ if $X_n = O_P(b_n)$ and $1/X_n = O_P(1/b_n)$. For simplicity, we will freely omit ceilings and floors; this will not affect the argument.

4.2 Convergence of metric spaces

The following five topologies will be relevant to us: (i) the Hausdorff topology on closed subsets of a compact metric space, (ii) the Gromov–Hausdorff (GH) topology on $\mathfrak{S}_{\mathrm{GH}}$–the isometry equivalence classes of compact metric spaces, (iii) the marked Gromov–Hausdorff topology on the isometry equivalence classes of triples of the form (X, C, d), where (X, d) is a compact metric space and $C \subseteq X$ is closed, (iv) the Gromov–Hausdorff–Prokhorov (GHP) topology on the isometry equivalence classes of compact metric measure spaces, and (v) the Gromov-weak topology on the isometry equivalence classes of metric measure spaces. The Hausdorff distance and the GHP distance will be denoted by $d_{\mathrm{H}}(\cdot, \cdot)$ and $d_{\mathrm{GHP}}(\cdot, \cdot)$ respectively. We refer the reader to [36] for background on the topologies in (i) and (ii). For the topologies in (iii) and (iv), we will primarily follow [64] and [1, 6] respectively. For the Gromov-weak topology, we refer the reader to [18, 53]. The definitions and results related to the topologies in (i)–(iv) needed in this paper can be found in one place in [8, Section 3.2].

4.3 $\mathbb{R}$-trees and $\mathbb{R}$-graphs

For any metric space (X, d), a geodesic between $x_1, x_2 \in X$ is an isometric embedding $f : [0, d(x_1, x_2)] \to X$ such that $f(0) = x_1$ and $f(d(x_1, x_2)) = x_2$. (X, d) is a geodesic space if there is a geodesic between any two points in X. An embedded cycle in X is a subset of X that is a homeomorphic image of the unit circle S^1.

Definition 4.1 *(Real trees [47, 62])* A compact geodesic metric space (X, d) is called a real tree or $\mathbb{R}$-tree if it has no embedded cycles.

For a metric space (X, d), $x \in X$, and $\varepsilon > 0$, let $B(x, \varepsilon; X) := \{ y \in X \, : \, d(y, x) \leq \varepsilon \}$. We next recall some definitions and constructs from [6]. We refer the reader to [6, Section 2.3] for a detailed treatment.

Definition 4.2 *(ℝ-graphs [6])* A compact geodesic metric space (X, d) is called an ℝ-graph if for every $x \in X$, there exists $\varepsilon > 0$ such that $\big(B(x, \varepsilon\,;\,X), d|_{B(x,\varepsilon\,;\,X)}\big)$ is an ℝ-tree.

The core of an ℝ-graph (X, d), denoted by $\mathrm{Core}(X)$, is the union of all the simple arcs having both endpoints in embedded cycles of X. If it is non-empty, then $(\mathrm{Core}(X), d)$ is an ℝ-graph with no leaves. We define $\mathrm{Conn}(X)$ to be the set of all $x \in X$ such that x belongs to an embedded cycle in X.

Clearly, $\mathrm{Conn}(X) \subseteq \mathrm{Core}(X)$. By [6, Theorem 2.7], if X is an ℝ-graph with a non-empty core, then $(\mathrm{Core}(X), d)$ can be represented as $(k(X), e(X), l)$, where $(k(X), e(X))$ is a finite connected multigraph in which all vertices have degree at least 3 and $l : e(X) \to (0, \infty)$ gives the edge lengths of this multigraph. We denote by $\mathrm{sp}(X)$ the number of surplus edges in $(k(X), e(X))$. On any ℝ-graph (X, d) there exists a unique σ-finite Borel measure len called the length measure such that if $x_1, x_2 \in X$ and $[x_1, x_2]$ is a geodesic path between x_1 and x_2 then $\mathrm{len}\big([x_1, x_2]\big) = d(x_1, x_2)$. Note that

$$\sum_{e \in e(X)} l(e) = \mathrm{len}\big(\mathrm{Core}(X)\big). \tag{13}$$

Clearly, $\mathrm{len}\big(\mathrm{Conn}(X)\big) \leq \mathrm{len}\big(\mathrm{Core}(X)\big) < \infty$. If $\mathrm{Conn}(X) \neq \emptyset$ (in which case $\mathrm{len}\big(\mathrm{Conn}(X)\big) > 0$), we write $\mathrm{len}_{\mathrm{Conn}(X)}$ for the restriction of the length measure to $\mathrm{Conn}(X)$ normalized to be a probability measure, i.e.,

$$\mathrm{len}_{\mathrm{Conn}(X)}(\cdot) = \frac{\mathrm{len}(\cdot)}{\mathrm{len}(\mathrm{Conn}(X))}\,.$$

For an ℝ-graph (X, d) and $x \in X$, choose $\varepsilon > 0$ such that $B(x, \varepsilon\,;\,X)$ is an ℝ-tree, and define the degree of x as

$$\deg(x\,;\,X) := \big|\{\text{connected components of } B(x, \varepsilon\,;\,X) \setminus \{x\}\}\big|.$$

Note that the value of $\deg(x\,;\,X)$ is independent of the choice of ε.

Since any finite connected graph, viewed as a metric space, is an ℝ-graph, the above definitions make sense for any finite connected graph H. Note the difference between $e(H)$ defined above and $E(H)$-the set of edges in H. Note also that in this case, the graph theoretic 2-core of H, viewed as a metric space, coincides with the space $\mathrm{Core}(H)$ as defined above, and $\mathrm{len}(\mathrm{Core}(H))$ equals the number of edges in the graph theoretic 2-core of H.

4.4 Some properties of MSTs

Suppose $H = (V, E, b)$ is a weighted, connected, and labeled graph. Assume that $b(e) \neq b(e')$ whenever $e \neq e'$. We now state a useful property of the MST.

Lemma 4.3 (Minimax paths property) *Let $H = (V, E, b)$ be as above. Then the MST T of H is unique. Further, T has the following property: Any path $(x_0, \ldots, x_n)$ with $x_i \in V$ and $\{x_i, x_{i+1}\} \in E(T)$ satisfies*

$$\max_{1 \leq i \leq n} b\big(\{x_i, x_{i+1}\}\big) \leq \max_{1 \leq j \leq m} b\big(\{x'_j, x'_{j+1}\}\big)$$

for any path $(x'_0, \ldots, x'_m)$ with $\{x'_j, x'_{j+1}\} \in E$ and $x_0 = x'_0$ and $x_n = x'_m$. In words, the maximum edge weight in the path in the MST connecting two given vertices is smallest among all paths in G connecting those two vertices.

Moreover, T is the only spanning tree of G with the above property.

The above lemma is just a restatement of [59, Lemma 2]; see also [16, Proposition 2.1]. We record the following useful observations that follow directly from Lemma 4.3:

Observation 4.1 The MST can be constructed just from the ranks of the different edge weights. This fact is not needed in the sequel, but it shows that the laws of M^n and $\bar{M}^n$ will remain unchanged if we used a set of exchangeable and pairwise distinct edge weights instead of U as in Definition 2.7.

Observation 4.2 Let $H = (V, E, b)$ be a connected and labeled graph with pairwise distinct edge weights. Let $u \in [0, \infty)$ and $\mathscr{C}$ be a component of the graph $G^u = (V, E^u)$, where $E^u \subseteq E$ contains only those edges e for which $b(e) \leq u$. Then the restriction of the MST of (V, E, b) to $\mathscr{C}$ is the MST of $\big(V(\mathscr{C}), E(\mathscr{C}), b|_{E(\mathscr{C})}\big)$.

This fact is extremely useful as it can be used to connect the structure of the MST to the geometry of components of the graph under percolation.

4.5 Cycle-breaking

In this section we recall two procedures from [6] that can be applied to $\mathbb{R}$-graphs and combinatorial graphs. We refer the reader to [6, Sections 3.1 and 3.2] for a detailed treatment. Recall the notation $k(X)$, $e(X)$, $(l(e), \ e \in e(X))$, and $\mathrm{sp}(X)$ introduced below Definition 4.2.

Definition 4.4 *(Cycle-breaking (*CB*), [6, Section 3.2])* Let X be an $\mathbb{R}$-graph. If X has no embedded cycles, then set $\mathrm{CB}(X) = X$. Otherwise, sample $x \in X$ using the measure $\mathrm{len}_{\mathrm{Conn}(X)}$. Endow $X \setminus \{x\}$ with the intrinsic metric: the distance between two points is the minimum of the lengths of paths in $X \setminus \{x\}$ that connect the two points. Set $\mathrm{CB}(X)$ to be the completion of $X \setminus \{x\}$ with respect to the intrinsic metric. (Thus, $\mathrm{CB}(X)$ is also an $\mathbb{R}$-graph.)

For $k \geq 2$, we inductively define $\mathrm{CB}^k(X)$ to be the space $\mathrm{CB}\big(\mathrm{CB}^{k-1}(X)\big)$. (Thus, at the k-th step, if $\mathrm{CB}^{k-1}(X)$ has an embedded cycle, then we are using the measure to $\mathrm{len}_{\mathrm{Conn}(\mathrm{CB}^{k-1}(X))}$ to sample a point.)

Note that $\mathrm{CB}^k(X) = \mathrm{CB}^{\mathrm{sp}(X)}(X)$ for all $k \geq \mathrm{sp}(X)$, i.e., the spaces $\mathrm{CB}^k(X)$ remain the same after all cycles have been cut open. We denote this final space (which is a real tree) by $\mathrm{CB}^\infty(X)$.

Next we define a cycle-breaking process for discrete graphs.

Definition 4.5 *(Cycle-breaking for discrete graphs (*CBD*), [6, Section 3.1])* Let $H = (V, E)$ be a finite connected graph. Sample $e \in E$ uniformly. If $(V, E \setminus \{e\})$ is connected, set $\mathrm{CBD}(H) = (V, E \setminus \{e\})$. Otherwise, set $\mathrm{CBD}(H) = H$. Inductively set $\mathrm{CBD}^{k+1}(H) = \mathrm{CBD}(\mathrm{CBD}^k(H))$, $k \geq 1$.

Almost surely, the graphs $\mathrm{CBD}^k(H)$ are the same (and are all trees) for all large values of k. We denote this tree by $\mathrm{CBD}^\infty(H)$.

Suppose H is a finite connected graph. Let $f_1, \ldots, f_s$ be the edges of H that get removed in the process $(\mathrm{CBD}^k(H), k \geq 1)$. Clearly, $s = \mathrm{sp}(H)$. For $1 \leq i \leq s$, let y_i be a uniformly sampled point on f_i. It is easy to see that viewing H as an $\mathbb{R}$-graph, the completion of the space $H \setminus \{y_1, \ldots, y_s\}$ with repect to the intrinsic metric has the same distribution as $\mathrm{CB}^\infty(H)$. In this coupling, $\mathrm{CBD}^\infty(H)$ is a subspace of $\mathrm{CB}^\infty(H)$, and

$$d_{\mathrm{H}}\big(\mathrm{CBD}^\infty(H), \ \mathrm{CB}^\infty(H)\big) \leq 1 . \tag{14}$$

We now state a lemma that connects cycle-breaking to MSTs.

Lemma 4.6 *Suppose H is a finite connected graph. Then $\mathrm{CBD}^\infty(H)$ has the same law as the MST of H constructed by assigning exchangeable pairwise distinct weights to the edges in H.*

Lemma 4.6 follows easily from Lemma 4.3. A proof can be found in [6, Proposition 3.5]. For $r \in (0, 1)$ define $\mathscr{A}_r$ to be the set of all $\mathbb{R}$-graphs X that satisfy

$$\mathrm{sp}(X) + \mathrm{len}(\mathrm{Core}(X)) \leq 1/r, \quad \text{and} \quad \min_{e \in e(X)} \mathrm{len}(e) \geq r . \tag{15}$$

The following theorem will allow us to prove convergence of MSTs from GHP convergence of the underlying graphs.

Theorem 4.7 *Fix $r \in (0, 1)$. Suppose (X, d) and (X_n, d_n), $n \geq 1$, are $\mathbb{R}$-graphs in $\mathscr{A}_r$ such that $(X_n, d_n) \to (X, d)$ as $n \to \infty$ w.r.t. GH topology. Further, suppose for each $n \geq 1$, (X_n, d_n) is isometric to $\varepsilon_n \cdot H_n$, where H_n, $n \geq 1$, are finite connected graphs and $\varepsilon_n \to 0$. Then as $n \to \infty$, $\varepsilon_n \cdot \mathrm{CBD}^\infty(H_n) \xrightarrow{\mathrm{d}} \mathrm{CB}^\infty(X)$ w.r.t. GH topology.*

Theorem 4.7 follows from [6, Theorem 3.3] and (14).

5 A slightly different model

Recall the definition of L_n from (4), and let

$$\ell_n := \sum_{i=2}^n w_i^{(n)} = L_n - w_1^{(n)} . \tag{16}$$

Analogous to (3) and (5), define

$$\bar{q}_{ij} = \bar{q}_{ij}(\boldsymbol{w}^{(n)}) = 1 \wedge \left(w_i^{(n)} w_j^{(n)}/\ell_n\right), \text{ and } q_{ij} = q_{ij}(\boldsymbol{w}^{(n)})$$
$$= 1 - \exp\left(-w_i^{(n)} w_j^{(n)}/\ell_n\right) \tag{17}$$

for $1 \le i \neq j \le n$. Note that under Assumption 2.4, for all large n, $\bar{q}_{ij} = w_i^{(n)} w_j^{(n)}/\ell_n$ for all $1 \le i \neq j \le n$. Let $\bar{G}_n$ (resp. G_n) be the random graph on $[n]$ obtained by placing an edge between i and j independently for each $i < j \in [n]$ with probability $\bar{q}_{ij}$ (resp. q_{ij}).

Lemma 5.1 *Under Assumption 2.4, there exists a coupling of $\bar{G}_n$ and $\tilde{G}_n$ such that $\mathbb{P}(\bar{G}_n \neq \tilde{G}_n) \to 0$ as $n \to \infty$. Similar assertions hold for the pairs $(G_n, \tilde{G}_n)$, and $(\bar{G}_n, G_n)$.*

Proof We will prove the assertion for $\bar{G}_n$ and $\tilde{G}_n$; a similar argument works for the other pairs. Under Assumption 2.4, $\max_{i,j}\left(w_i^{(n)} w_j^{(n)}/\ell_n\right) = O(n^{-\eta})$. By [58, Corollary 2.12], it is enough to prove that $\sum_{i \neq j} (\bar{q}_{ij} - \overline{q}_{ij})^2/\overline{q}_{ij} \to 0$, as $n \to \infty$. Now, under Assumption 2.4, for all large n,

$$\sum_{i \neq j} \frac{(\bar{q}_{ij} - \overline{q}_{ij})^2}{\overline{q}_{ij}} = \sum_{i \neq j} \frac{w_i w_j}{\ell_n}\left(1 - \frac{\ell_n}{L_n}\right)^2 \le \frac{L_n^2}{\ell_n} \cdot \left(\frac{w_1}{L_n}\right)^2 \le C n^{-\eta}.$$

This completes the proof. $\square$

Similar to Definition 2.7, let M^n denote the minimal spanning tree on $\mathscr{C}(1, G_n)$ using the edge weights $\left(U_{ij}; \{i, j\} \in E\left(\mathscr{C}(1, G_n)\right)\right)$. Define $\bar{M}^n$ as the MST on $\mathscr{C}(1, \bar{G}_n)$ in an analogous way. In view of Lemma 5.1, it is enough to prove Theorem 3.1 for M^n and $\bar{M}^n$. Hence, from now on we will only work with the random graphs G_n and $\bar{G}_n$. This will make some of the computations simpler.

6 Relating the MST with the components in the critical window

We will now need some notation. Recall that $\ell_n = \sum_{i=2}^n w_i^{(n)}$, and define

$$\sigma_2^{(n)} = n^{-1} \sum_{i=2}^n (w_i^{(n)})^2, \text{ and } v_n := n\sigma_2^{(n)}/\ell_n. \tag{18}$$

(Note that once again we have omitted the weight of vertex 1 in the formulae.) Using Assumption 2.4, it is easy to see that $\liminf_{n\to\infty} v_n = v > 1$. Without loss of generality, we can assume the following:

Assumption 6.1 For all $n \ge 1$, $v_n \ge v' := 1 + (v - 1)/2$.

For $\lambda \geq 0$, define

$$p_\lambda^n := \left(1 + \frac{\lambda}{n^\eta}\right)\frac{1}{\nu_n}.$$ (19)

Let U be as in Definition 2.7. For any subgraph H of $\bar{G}_n$, we write

$$U\big|_H := \left(U_{ij}\,;\,\{i,j\} \in E(H)\right).$$ (20)

In words, $U\big|_H$ is the collection of edge weights corresponding to the edges in H.

Definition 6.2 (The graph $\bar{G}_n(\lambda)$) Fix $\lambda \geq 0$. Let $\bar{G}_n(\lambda)$ be the subgraph of $\bar{G}_n$ with edge set

$$\left\{\{i,j\}\,:\,\{i,j\} \in E(\bar{G}_n) \text{ and } U_{ij} \leq p_\lambda^n\right\}.$$

For $v \in [n]$, let $\bar{\mathscr{C}}_v^n(\lambda) = \mathscr{C}\big(v, \bar{G}_n(\lambda)\big)$. Let $\bar{M}_\lambda^n$ denote the MST on $\bar{\mathscr{C}}_1^n(\lambda)$ constructed using $U\big|_{\bar{\mathscr{C}}_1^n(\lambda)}$.

Thus, $\bar{M}^n = \bar{M}_\lambda^n$ for all $\lambda \geq (\nu_n-1)n^\eta$. Note that if $p_\lambda^n \in [0,1]$ and $w_i^{(n)}w_j^{(n)}/\ell_n \leq 1$ for all $i \neq j \in [n]$, then $\bar{G}_n(\lambda)$ is a random graph on $[n]$ with independent edge connection probabilities $\big(p_\lambda^n w_i^{(n)}w_j^{(n)}/\ell_n\ 1 \leq i < j \leq n\big)$. Note also that by Observation 4.2, $\bar{M}_{\lambda_1}^n$ is a subtree of $\bar{M}_{\lambda_2}^n$ whenever $0 \leq \lambda_1 \leq \lambda_2$. In particular, $\bar{M}_\lambda^n$ is a subtree of $\bar{M}^n$ for all $\lambda \geq 0$.

Definition 6.3 (The graph $G_n(\lambda)$) Fix $\lambda \geq 0$. Let $G_n(\lambda)$ be the random graph on $[n]$ obtained by placing an edge between i and j independently for each $i \neq j \in [n]$ with probability $1 - \exp\big(-p_\lambda^n w_i^{(n)}w_j^{(n)}/\ell_n\big)$ for $1 \leq i < j \leq n$. Let $\mathscr{C}_1^n(\lambda) = \mathscr{C}\big(1, G_n(\lambda)\big)$.

We are now ready to state the result that connects $\bar{M}^n$ to $\bar{M}_\lambda^n$.

Theorem 6.1 *Under Assumptions 2.4 and 6.1, for every $\Delta \in (0, 1/2]$, there exist $\lambda_{21} = \lambda_{21}(\Delta) \geq 1$, $n_{21} = n_{21}(\Delta) \geq 2$, and $C_{21} = C_{21}(\Delta) > 0$ such that for $n \geq n_{21}$ and $\lambda \in [\lambda_{21}, (\nu_n - 1)n^\eta]$,*

$$\mathbb{P}\left(d_{\mathrm{H}}\big(\bar{M}_\lambda^n,\,\bar{M}^n\big) \geq \frac{n^\eta}{\lambda^{1-\Delta}} + \frac{n^\eta}{(\log n)^{1/6}}\right) \leq \frac{C_{21}}{\sqrt{\lambda}}.$$ (21)

Remark 1 Note that $\mathscr{W}\big(\mathscr{C}(1, \bar{G}_n)\big) = \Theta_p(n)$, while it turns out (see (43), (42), and Lemma 7.4) that for large (and fixed) λ, $\mathscr{W}(\bar{\mathscr{C}}_1^n(\lambda)) = \Theta_P\big(\lambda^{1/(\tau-3)}n^\rho\big) = o_P(n)$. Thus, the above result shows that despite this major gap in their respective masses, $n^{-\eta} \cdot \bar{M}_\lambda^n$ approximates $n^{-\eta} \cdot \bar{M}^n$ quite well when λ is large. However, as we will see later (see (145) and [29, Theorem 1.2(c)]), if λ is kept fixed and $n \to \infty$, $n^{-\eta} \cdot \bar{M}_\lambda^n$ converges in distribution in the GH sense to a limiting compact $\mathbb{R}$-tree whose Minkowski dimension is $(\tau - 2)/(\tau - 3)$, whereas $\dim(\mathscr{M}^{\theta^*}) = (\tau - 1)/(\tau - 3)$.

Thus, the intrinsic geometry of $\bar{M}_\lambda^n$ has features significantly different from that of $\bar{M}^n$. Similar results were proved in [6] in the context of the MST of the complete graph.

Theorem 6.1 follows upon combining the next two propositions.

Proposition 6.4 *Under Assumptions* 2.4 *and* 6.1, *there exists* $\varepsilon_{22} \in (0, v' - 1)$, *and for every* $\Delta \in (0, 1/2)$, *there exist* $\lambda_{21} = \lambda_{21}(\Delta) \geq 1$, $n_{22} = n_{22}(\Delta) \geq 1$, *and* $C_{22} = C_{22}(\Delta) > 0$ *such that for* $n \geq n_{22}$ *and* $\lambda \in [\lambda_{21}, \varepsilon_{22}n^\eta]$,

$$\mathbb{P}\left(d_{\mathrm{H}}\big(\bar{M}_\lambda^n, \bar{M}_{\varepsilon_{22}n^\eta}^n\big) \geq \frac{n^\eta}{\lambda^{1-\Delta}}\right) \leq \frac{C_{22}}{\sqrt{\lambda}}. \tag{22}$$

Proposition 6.5 *Under Assumptions* 2.4 *and* 6.1, *for all large n,*

$$\mathbb{P}\left(d_{\mathrm{H}}\big(\bar{M}_{\varepsilon_{22}n^\eta}^n, \bar{M}^n\big) \geq \big(\log n\big)^{-1/6} n^\eta\right) \leq n^{-1}.$$

The rest of Sect. 6 is devoted to the proofs of Propositions 6.4 and 6.5. In Sect. 6.1 below, we first explain the rationale behind proving these two results, and then the proofs are completed in several steps in the following sections. From now on, all results will be proved under Assumptions 2.4 and 6.1, and we will not mention this explicitly.

6.1 The general strategy

The scaling limit of the MST of the complete graph viewed as a metric measure space was established in [6] relying on the results of [2, 3, 7]. This was an important breakthrough, and the limiting space is–quoting the authors of [6]–"one of the first scaling limits to be identified for any problem from combinatorial optimisation." This proof has four key ingredients: (i) One of them is deriving the critical scaling limit of the Erdős-Rényi random graph. The scaling limit of the maximal components of the Erdős-Rényi random graph inside the critcal window was established in [3] with respect to the GH topology. This was strengthened to convergence with respect to the GHP topology in [6]. (ii) Consider the MST $\mathfrak{M}_n$ on the complete graph K_n on n vertices constructed using i.i.d. Uniform[0, 1] edge weights U_{ij}, $1 \leq i < j \leq n$. Now, for $\lambda > 0$, consider the subgraph of K_n with vertex set $[n]$ and edge set $\big\{\{i, j\} : 1 \leq i < j \leq n, U_{ij} \leq n^{-1} + \lambda n^{-4/3}\big\}$; let $\mathfrak{M}_{n,\lambda}$ denote the restriction of $\mathfrak{M}_n$ to the maximal component in this graph. An important step in the proof is getting a tail bound on the Hausdorff distance between $\mathfrak{M}_{n,\lambda}$ and $\mathfrak{M}_n$ for large fixed λ. This bound was obtained in [7]. In words, this result states that $\mathfrak{M}_{n,\lambda}$ is quite close to $\mathfrak{M}_n$ in the GH sense if λ is sufficiently large, and thus, the structure of $\mathfrak{M}_n$ viewed as a metric space is essentially determined in the late stages of the critical window. (Note that the authors of [7] prove their results in the slightly different setting of a random graph process evolving through the addition of edges in discrete time. However, this result translates to the setting mentioned above in a straighforward way.) (iii) The third ingredient is proving a tail bound on the maximal number of vertices in the trees

obtained by removing the edges of $\mathfrak{M}_{n,\lambda}$ from $\mathfrak{M}_n$ for large λ. This was established in [6, Lemma 4.11]. (iv) Finally, certain topological properties of the scaling limit were established in [6]. This included showing that the Minkowski dimension of the limiting space is 3 almost surely; the proof of this result also made use of the results in [2].

The scaling limit of $\mathfrak{M}_n$ with respect to the GH topology was established in [6] by combining the critical scaling limit of the Erdős-Rényi random graph mentioned in **(i)** with the tail bound in **(ii)** via the results of [6, Section 3]. This can be strengthened to GHP convergence by using the bound in **(iii)** and using the fact that the number of vertices in the trees obtained by removing the edges of $\mathfrak{M}_{n,\lambda}$ from $\mathfrak{M}_n$ satisfy a certain exchangeability propoerty; see the proof of [6, Proposition 4.8] and [8, Lemma 6.19 and Remark 5].

In the context of the multiplicative coalescent in the regime of interest in this paper, the critical metric scaling limit of the maximal components was obtained in [29, 35] (we will need slightly tweaked versions of these results in our proof as will be discussed in Sect. 7.1 below). The missing ingredient for proving the convergence in (11) was the analogue of the result in **(ii)** mentioned above in the present setting. Theorem 6.1 provides this tail bound. Once we have this bound, it can be combined with the critical scaling limit in a manner similar to [6] to deduce the claimed GH convergence in (11). The techniques used in the proof of Proposition 6.4 to analyze the graph $G_n(\lambda)$ will also be useful in Sects. 7.4 and 7.5 where we prove the claimed Minkowski dimension in Theorem 3.1 (c).

Now, the tail bound mentioned in **(ii)** above was established in [7] in two stages: (a) First, a bound on the Hausdorff distance between $\mathfrak{M}_{n,n^{1/3}(\log n)^{-1}}$ and $\mathfrak{M}_n$ (going from the barely supercritical regime to the purely supercritical regime) is proved using a variation of Prim's algorithm [71]; this approach used in [7] is explained in Algorithm 1 in Sect. 6.8. (b) Next, the Hausdorff distance between $\mathfrak{M}_{n,\lambda}$ and $\mathfrak{M}_{n,n^{1/3}(\log n)^{-1}}$ (from the critical window to the barely supercritical regime) is bounded for fixed $\lambda > 0$. This step makes use of [63, Theorem 7] which, in words, says that the identity of the maximal component in the Erdős-Rényi random graph process gets fixed in the late critical window. The desired bound is then established by considering time points in a suitably chosen geometric progression in the interval $[\lambda, n^{1/3}(\log n)^{-1}]$, and estimating the Hausdorff distance between the MSTs of the maximal components at consecutive time points in this progression. This technique of proving a property of a random graph process by considering time points where the consecutive points are neither too close nor too far away was previously used in [63, Sections 4 and 6] where it was termed the "scanning method."

Proposition 6.5 stated above gives a bound on the Hausdorff distance between $\bar{M}^n$ and $\bar{M}^n_{\varepsilon_{22}n^\eta}$–the MST on the component of the vertex 1 in the purely supercritical regime where the graph is only slightly supercritical. This gives a result analogous to the one derived in [7] (mentioned in **(a)** above) for the model of interest in this paper. In our setting, applying Algorithm 1 directly would not yield the desired bound; rather we have to use a modification of this approach (explained in Algorithm 2 in Sect. 6.8). Proposition 6.4 proves the complementary bound by connecting $\bar{M}^n_\lambda$ to $\bar{M}^n_{\varepsilon_{22}n^\eta}$. In the proof of Proposition 6.4, we use the scanning method as in [7, 63]. Here, the bulk of the work lies in choosing an appropriate geometric progression to which

the scanning method can be applied, and bounding the Hausdorff distance between the MSTs at successive time points in this progression. This requires several new techniques. In particular, we rely on a new method of getting tail bounds on heights of branching processes recently developed in [4], concentration inequalities proved in [5], and results on the relation between connected components of inhomogeneous rank-1 graphs and **p**-trees derived in [28, 29].

In Sect. 6.2 below, we will define two processes called 'breadth-first walks' that will help us analyze the random graph models of interest. In Sect. 6.3, we use a concentration inequality for the suprema of the centered breadth-first walk (which relies on a similar inequality derived in [5]) to prove bounds on the lower tails of the total weight of the component of 1 in $G_n(\lambda)$ as well as the sum of the squares of the vertex weights in the component of 1 in $G_n(\lambda)$. These bounds hold for a range of values of λ where $G_n(\lambda)$ passes from the late critcal window to the purely supercritical regime. We also show that when λ is in this range, the probability that many of the high-weight vertices are contained in the component of 1 is lower bounded by an appropriately chosen function of λ. In particular, these results will allow us to show that for λ in this range, if the component of 1 is removed from $G_n(\lambda)$, then the rest of the graph is sufficiently subcritical. In Sect. 6.4, we obtain a tail bound on the height of a branching process closely related to the random graph $G_n(\lambda)$. Here, we use a technique developed in [4]. Now, as mentioned above, we aim to use the scanning method to prove the claimed bound in Proposition 6.4. With this in mind and building on the results of Sects. 6.2, 6.3, and 6.4, we achieve the following in Sects. 6.5 and 6.6: (1) We define $\delta_1 > 0$ such that the scanning method can be applied to a geometric progression of time points where the common ratio is $(1 + \delta_1/2)$. We also define a random graph $H_n(\lambda, \delta_1)$ such that $d_H(\bar{M}_\lambda^n, \bar{M}_{\lambda(1+\delta_1/2)}^n)$ can be stochastically bounded in terms of the longest self-avoiding path in $H_n(\lambda, \delta_1)$. (2) We prove a tail bound on the diameter of $H_n(\lambda, \delta_1)$. This is done by establishing height bounds for a random tree that has (potentially) three layers, each of which resembles a multitype branching process. The offspring distributions in these three layers and the depths of the different layers are chosen in a suitable way to obtain the desired bound. (3) We obtain a lower bound for the probability that each component of $H_n(\lambda, \delta_1)$ is either a tree or is unicyclic. Here, we make use of a construction of a connected component of a rank-1 inhomogeneous random graph using **p**-trees [28, 29]. We then use these results and apply the scanning method to complete the proof of Proposition 6.4 in Sect. 6.7. Finally, the proof of Proposition 6.5 is given in Sect. 6.8. As mentioned above, here we use a modification of the approach used in the proof of [7, Lemma 4].

6.2 An exploration process

Fix $\lambda \geq 0$. It will be useful in the proof to express $G_n(\lambda)$ in a reparametrized form. Define

$$
x_i^{(n)} := \frac{w_i^{(n)}}{n^\rho \big(\sigma_2^{(n)}\big)^{1/2}}, \quad \text{and} \quad \theta_{i,\lambda}^{(n)} := \left(1 + \frac{\lambda}{n^\eta}\right) \frac{w_i^{(n)}}{n^\alpha \big(\sigma_2^{(n)}\big)^{1/2}}, \qquad 1 \leq i \leq n.
$$

(23)

Write $\mathbf{x}^{(n)} = (x_i^{(n)}, \ i \in [n])$, and $\sigma_2(\mathbf{x}^{(n)}) := \sum_{i=2}^{n}(x_i^{(n)})^2$. (Note that similar to (18), $x_1^{(n)}$ is not included in the sum.) Then

$$\sigma_2(\mathbf{x}^{(n)}) = n^{-\eta}, \quad \text{and} \quad (\lambda + n^{\eta})x_j^{(n)} = \theta_{j,\lambda}^{(n)} \quad \text{for} \quad j \in [n]. \tag{24}$$

Further, by Assumption 2.4, for each $i \geq 1$,

$$(\sigma_2(\mathbf{x}^{(n)}))^{-1}x_i^{(n)} = n^{\eta} \cdot x_i^{(n)} \to \theta_i^*, \quad \text{as} \quad n \to \infty. \tag{25}$$

Now note that for $2 \leq i < j \leq n$,

$$p_\lambda^n\left(\frac{w_i^{(n)}w_j^{(n)}}{\ell_n}\right) = \left(\lambda + \frac{1}{\sigma_2(\mathbf{x}^{(n)})}\right)x_i^{(n)}x_j^{(n)}, \tag{26}$$

and consequently,

$$G_n(\lambda) \stackrel{\mathrm{d}}{=} \mathscr{G}_n(\lambda) := \mathscr{G}\left(([n], \mathbf{x}^{(n)}), \ \lambda + (\sigma_2(\mathbf{x}^{(n)}))^{-1}\right), \tag{27}$$

where the latter is as in Definition 2.2.

Let $\mathscr{G}_n^-(\lambda) = \mathscr{G}_n(\lambda) \setminus [1]$. A useful tool in the study of the random graph $\mathscr{G}_n^-(\lambda)$ is the "breadth-first walk" process $(Z_\lambda^{n,-}(u), u \geq 0)$ associated with a breadth-first exploration of the random graph $\mathscr{G}_n^-(\lambda)$, which we describe next. This is very much in the spirit of [11], although our breadth-first walk (defined in (29)) is slightly different from the one considered in [11], as it is easier to analyze.

For $2 \leq j \leq n$, let the size of vertex j be $x_j^{(n)}$. Choose $v(1)$ from $[n] \setminus \{1\}$ in a size-biased way, i.e., $\mathbb{P}(v(1) = v) \propto x_v^{(n)}, v \in [n]\setminus\{1\}$. Explore the component of $v(1)$ in $\mathscr{G}_n^-(\lambda)$ in a breadth-first fashion. Let h^1 be the height of the breadth-first tree, and for $0 \leq i \leq h^1$, let Gen_i^1 be the set of vertices in the i-th generation of the breadth-first tree. For $t \geq 2$, having explored the component of $v(1), \ldots, v(t-1)$, choose $v(t)$ in a size-biased way from the remaining vertices, explore its component in $\mathscr{G}_n^-(\lambda)$ in a breadth-first fashion, let h^t be the height of its breadth-first tree, and for $0 \leq i \leq h^t$, let Gen_i^t be the set of vertices in the i-th generation of the breadth-first tree. Stop when all vertices $j \in [n] \setminus \{1\}$ have been found.

Using properties of exponential random variables, it can be easily checked that the above collection of random variables can be constructed in the following way: Recall the relations from (24). Let $\xi_j^n, 2 \leq j \leq n$, be independent random variables such that

$$\xi_{j,\lambda}^n \sim \mathrm{Exp}\left((\lambda + (\sigma_2(\mathbf{x}^{(n)}))^{-1})x_j^{(n)}\right), \quad \text{or equivalently,} \quad \xi_{j,\lambda}^n \sim \mathrm{Exp}(\theta_{j,\lambda}^{(n)}). \tag{28}$$

To simplify notation we will write ξ_j^n instead of $\xi_{j,\lambda}^n$. Let $v(1)$ be such that $\xi_{v(1)}^n = \min\left\{\xi_j^n : 2 \le j \le n\right\}$, and set $\text{Gen}_0^1 = \{v(1)\}$. Inductively define

$$\text{Gen}_i^1 = \left\{ j \in [n] \setminus \{1\} : \right.$$

$$\left. \xi_j^n \in \left[\xi_{v(1)}^n + \sum_{k=0}^{i-2} \sum_{v \in \text{Gen}_k^1} x_v^{(n)} , \ \xi_{v(1)}^n + \sum_{k=0}^{i-1} \sum_{v \in \text{Gen}_k^1} x_v^{(n)} \right] \right\} \text{ for } 1 \le i \le h^1 + 1 ,$$

where

$$h^1 + 1 = \min \left\{ i \ge 1 : \text{Gen}_i^1 = \emptyset \right\} .$$

For $t \ge 2$, let $v(t)$ be such that

$$\xi_{v(t)}^n = \min \left\{ \xi_j^n : 2 \le j \le n \text{ and } j \notin \bigcup_{k=1}^{t-1} \bigcup_{i=0}^{h^k} \text{Gen}_i^k \right\} .$$

Set $\text{Gen}_0^t = \{v(t)\}$, and define Gen_i^t, $1 \le i \le h^t + 1$, in a manner analogous to the case $t = 1$. Stop when all vertices $j \in [n] \setminus \{1\}$ have been found.

We define the breadth-first walk as

$$Z_\lambda^{n,-}(u) = -u + \sum_{2 \le j \le n} x_j^{(n)} \mathbb{1}\{\xi_j^n \le u\}, \quad u \ge 0. \tag{29}$$

The correspondence described above allows one to prove various properties of $\mathscr{G}_n^-(\lambda)$ by studying the process $Z_\lambda^{n,-}$. Here we make note of an elementary property of $Z_\lambda^{n,-}$ that will be useful to us: Suppose $\mathscr{G}_n^-(\lambda)$ and ξ_j^n, $2 \le j \le n$, are coupled by means of the correspondence described above. Then in this coupling, for $1 \le t \le m$,

$$Z_\lambda^{n,-}\left(\xi_{v(t)}^n -\right) = Z_\lambda^{n,-}\left(\xi_{v(t)}^n + \mathscr{W}_{\mathbf{x}^{(n)}}\left(\mathscr{C}\left(v_{(t)}, \mathscr{G}_n^-(\lambda)\right)\right)\right)$$

$$= -\xi_{v(t)}^n + \sum_{k=1}^{t-1} \mathscr{W}_{\mathbf{x}^{(n)}}\left(\mathscr{C}\left(v_{(k)}, \mathscr{G}_n^-(\lambda)\right)\right) < 0 ,$$

where m denotes the number of components in $\mathscr{G}_n^-(\lambda)$. This leads to the following:

Lemma 6.6 *In the above coupling, if for some $0 < u_1 < u_2$, $Z_\lambda^{n,-}(u) > 0$ for $u \in [u_1, u_2]$, then there exists a component $\mathscr{C}^\star$ in $\mathscr{G}_n^-(\lambda)$ such that $\mathscr{W}_{\mathbf{x}^{(n)}}(\mathscr{C}^\star) \ge u_2 - u_1$.*

Note that we can explore $\mathscr{G}_n(\lambda)$ starting from the vertex 1 in a manner similar to the exploration of $\mathscr{G}_n^-(\lambda)$. In this case, we define the breadth-first walk as

$$Z_\lambda^{n,(1)}(u) = x_1^{(n)} - u + \sum_{2 \le j \le n} x_j^{(n)} \mathbb{1}\{\xi_j^n \le u\}, \quad u \ge 0, \tag{30}$$

where ξ_j^n, $2 \le j \le n$, are as in (28). (Here we append a "(1)" to specify that the process starts at vertex 1.) As in the case of $\mathscr{G}_n^-(\lambda)$, there is a natural coupling between $\mathscr{G}_n(\lambda)$ and ξ_j^n, $2 \le j \le n$.

Lemma 6.7 *In the above coupling,*

$$\mathscr{W}_{\boldsymbol{x}^{(n)}}\big(\mathscr{C}\big(1, \mathscr{G}_n(\lambda)\big)\big) = \inf \big\{u \ge 0 \ : \ Z_\lambda^{n,(1)}(u) = 0\big\}.$$

Let us get back to the process $Z_\lambda^{n,-}(\cdot)$. Using (28) and the second identity in (24), it can be directly checked that for any $u \ge 0$,

$$\mathbb{E}\big[n^\eta Z_\lambda^{n,-}(u)\big] = \lambda u - \sum_{j=2}^n \frac{\theta_{j,\lambda}^{(n)}}{\big(1 + \lambda n^{-\eta}\big)} \Big(u\theta_{j,\lambda}^{(n)} + \exp(-u\theta_{j,\lambda}^{(n)}) - 1\Big) =: \Phi_\lambda^{(n)}(u). \tag{31}$$

Note that $\Phi_\lambda^{(n)}(\cdot)$ is a strictly concave function (in u). In particular, for any $\lambda > 0$, $\Phi_\lambda^{(n)}(\cdot)$ has a unique positive zero, which we will denote by $s^{(n)}(\lambda)$. Define

$$\varphi_\lambda^{(n)}(u) := \sum_{j=2}^n \big(\theta_{j,\lambda}^{(n)}\big)^2 \frac{\Big(u\theta_{j,\lambda}^{(n)} + \exp(-u\theta_{j,\lambda}^{(n)}) - 1\Big)}{u\theta_{j,\lambda}^{(n)}} \tag{32}$$

for $u \ge 0$ (the value at $u = 0$ is understood to be the limit of $\varphi_\lambda^{(n)}(u)$ as $u \downarrow 0$), so that

$$\Phi_\lambda^{(n)}(u) = \lambda u - u\varphi_\lambda^{(n)}(u)\big/\big(1 + \lambda n^{-\eta}\big). \tag{33}$$

From now on, we will drop the dependence on n in the superscripts and simply write $\theta_{j,\lambda}, x_j, \boldsymbol{x}$, and w_j to ease notation. Recall the constants A_1, A_2 from Assumption 2.4. Consider the interval

$$I^{(n)} := \left[\frac{2 \cdot \big(\sigma_2^{(n)}\big)^{1/2}}{A_1}, \ \frac{n^\alpha \big(\sigma_2^{(n)}\big)^{1/2}}{A_2 2^{\alpha+1}}\right]. \tag{34}$$

For $u \in I^{(n)}$, define $i_\lambda(u) = \min \big\{i \ge 1 : \theta_{i,\lambda} u < 1\big\}$. From the definition of $I^{(n)}$ and Assumption 2.4 (iii), it follows that $3 \le i_\lambda(u) \le n/2$ for $u \in I^{(n)}$. Writing

$$I^{(n),2} = \big\{(\lambda, u) \ : \ 0 \le \lambda \le n^\eta/10, \ u \in I^{(n)}\big\}, \tag{35}$$

we have, for all large n,

$$i_\lambda(u) \asymp u^{1/\alpha} \quad \text{for} \quad (\lambda, u) \in I^{(n),2}. \tag{36}$$

Define the function $g(s) := s + e^{-s} - 1$ for $s \geq 0$. Since $g(s) \asymp s^2$ on $[0, 1]$ while $g(s) \asymp s$ on $[1, \infty)$, using Assumption 2.4 we have, $\bar{\forall}n$ and for all $(\lambda, u) \in I^{(n),2}$,

$$u\varphi_\lambda^{(n)}(u) \asymp \sum_{j=2}^{i_\lambda(u)-1} \theta_{j,\lambda}(u\theta_{j,\lambda}) + \sum_{j=i_\lambda(u)}^{n} \theta_{j,\lambda}(u\theta_{j,\lambda})^2$$

$$\asymp u \sum_{j=2}^{i_\lambda(u)-1} \frac{1}{j^{2\alpha}} + u^2 \sum_{j=i_\lambda(u)}^{n} (\theta_{j,\lambda})^3. \tag{37}$$

Using (36), we see that $\bar{\forall}n$ and for all $(\lambda, u) \in I^{(n),2}$,

$$u \sum_{j=2}^{i_\lambda(u)-1} \frac{1}{j^{2\alpha}} \asymp u \cdot (i_\lambda(u))^{1-2\alpha} \asymp u \cdot u^{(1-2\alpha)/\alpha} \asymp u^{\tau-2}, \tag{38}$$

whereas Assumption 2.4 (iii) and (7) yield

$$0 \leq u^2 \sum_{j=i_\lambda(u)}^{n} (\theta_{j,\lambda})^3 \leq C u^2 \sum_{j=i_\lambda(u)}^{n} \frac{1}{j^{3\alpha}} \leq C' u^2 (i_\lambda(u))^{1-3\alpha} \leq C'' u^{\tau-2}. \tag{39}$$

Combining (37), (38), and (39), we get, $\bar{\forall}n$ and for all $(\lambda, u) \in I^{(n),2}$,

$$u\varphi_\lambda^{(n)}(u) \asymp u^{\tau-2}. \tag{40}$$

Now let us switch back to $\Phi_\lambda^{(n)}(\cdot)$. Using the asymptotics for $\varphi_\lambda^{(n)}(\cdot)$ in (40), choose $\lambda_{41} \geq 1$ large enough and $\varepsilon_{41} \in (0, \frac{1}{10} \wedge (\nu' - 1))$ small so that $\bar{\forall}n$,

$$\Phi_{\lambda_{41}}^{(n)}\left(\frac{2(\sigma_2^{(n)})^{1/2}}{A_1}\right) > 0, \quad \text{and} \quad \Phi_{\varepsilon_{41}n^\eta}^{(n)}\left(\frac{n^\alpha(\sigma_2^{(n)})^{1/2}}{A_2 2^{\alpha+1}}\right) < 0,$$

and consequently, the unique positive zero $s^{(n)}(\lambda)$ of $\Phi_\lambda^{(n)}(\cdot)$ satisfies, $\bar{\forall}n$,

$$s^{(n)}(\lambda) \in I^{(n)} \quad \text{for} \quad \lambda \in [\lambda_{41}, \varepsilon_{41}n^\eta]. \tag{41}$$

(Note that $\varepsilon_{41} \in (0, \frac{1}{10} \wedge (\nu' - 1))$ implies that $\bar{\forall}n$, $\varepsilon_{41} < (\nu_n - 1)$.) Thus, (40), (33), and the relation $\Phi_\lambda^{(n)}(s^{(n)}(\lambda)) = 0$ implies, $\bar{\forall}n$,

$$s^{(n)}(\lambda) \asymp \lambda^{1/(\tau-3)} \quad \text{for} \quad \lambda \in [\lambda_{41}, \varepsilon_{41}n^\eta]. \tag{42}$$

6.3 The component of the vertex 1 in $G_n(\lambda)$

Recall the notation $\mathscr{C}_1^n(\lambda)$ from Definition 6.3. In this section we will study properties of $\mathscr{C}_1^n(\lambda)$. We start with a lower bound on $\mathscr{W}(\mathscr{C}_1^n(\lambda))$.

Proposition 6.8 *There exists $\kappa_0 \in (0, 1/4)$ such that the following holds for all $\kappa \in (0, \kappa_0]$: there exists $n_{43} = n_{43}(\kappa)$ and $\lambda_{43} = \lambda_{43}(\kappa) \geq \lambda_{41}$ such that for all $n \geq n_{43}$ and $\lambda \in [\lambda_{43}, \varepsilon_{41} n^\eta]$,*

$$\mathbb{P}\left(\mathscr{W}_x\big(\mathscr{C}\big(1, \mathscr{G}_n(\lambda)\big)\big) \geq (1 - 2\kappa)s^{(n)}(\lambda) \right) \geq 1 - \exp\big(-C\lambda^{1/(\tau-3)}\big), \qquad (43)$$

and consequently, $\mathbb{P}\left(\mathscr{W}\big(\mathscr{C}_1^n(\lambda)\big) \geq (1 - 2\kappa)s^{(n)}(\lambda)\big(\sigma_2^{(n)}\big)^{1/2} n^\rho \right) \geq 1 - \exp\big(- C\lambda^{1/(\tau-3)}\big).$

We will make use of the next two results in the proof of Proposition 6.8. Recall the independent exponential random variables ξ_j^n, $2 \leq j \leq n$, from (28).

Lemma 6.9 *([5, Lemma 6.3]) There exists a constant $C_{44} > 0$ such that for all $s \geq 1/C_{44}$, $y \geq 1/C_{44}$, for all large n, and $\lambda \in [1, n^\eta]$,*

$$\mathbb{P}\left(\sup_{u \leq s} \left| \sum_{j=2}^n \theta_{j,\lambda}\left(\mathbb{1}\{\xi_j^n \leq u\} - \mathbb{P}\big(\xi_j^n \leq u\big)\right) \right| \geq y s^{\frac{\tau-3}{2}} \right) \leq \exp\big(- C_{44} y \log\log y\big). \tag{44}$$

Remark 2 The result in [5, Lemma 6.3] is given in a slightly different setting. However, the key ingredient in its proof is the Klein-Rio bound [60, Theorem 1.1]. Applying [60, Theorem 1.1], the proof of [5, Lemma 6.3] boils down to establishing a uniform upper bound on the variance of a certain collection of functionals, and an upper bound on the expectation of the supremum of the said collection of functionals. This is achieved in the proof of [5, Lemma 6.3] and in [5, Lemma 6.4]. Now, using Assumption 2.4 (iii) and (7), those same arguments can be used to prove Lemma 6.9. Further, examining the proofs of [5, Lemma 6.3 and Lemma 6.4] will reveal that the constant C_{44} can be chosen so that it depends only on A_1, A_2, and τ. We omit the proof of Lemma 6.9 as no new idea is involved.

The next lemma describes technical properties of the function $\varphi_\lambda^{(n)}$ defined in (32). For fixed $\kappa > 0$, define analogous to the set $I^{(n),2}$,

$$I_\kappa^{(n),2} := \left\{ (\lambda, u) : 0 \leq \lambda \leq n^\eta/10, \ u \in \left[\frac{2 \cdot \big(\sigma_2^{(n)}\big)^{1/2}}{A_1 \kappa}, \ \frac{n^\alpha \cdot \big(\sigma_2^{(n)}\big)^{1/2}}{A_2 2^{\alpha+1}} \right] \right\}.$$

Lemma 6.10 *There exist $C_{45} > 0$ and $\kappa_0 \in (0, 1/4)$ such that for all large n, for all $\kappa \in (0, \kappa_0]$, and $(\lambda, u) \in I_\kappa^{(n),2}$,*

$$\frac{\varphi_\lambda^{(n)}(\kappa u)}{\varphi_\lambda^{(n)}(u)} \leq \frac{\varphi_\lambda^{(n)}\big((1-\kappa)u\big)}{\varphi_\lambda^{(n)}(u)} \leq \frac{1}{1 + C_{45}\kappa}. \tag{45}$$

Proof of Proposition 6.8 assuming Lemma 6.10: Recall the function $\Phi_\lambda^{(n)}$ from (31) and its connection to $\varphi_\lambda^{(n)}$ from (33). By (41) and (42), for any $\kappa \in (0, \kappa_0]$, there exist $\lambda_{46} = \lambda_{46}(\kappa) \geq \lambda_{41}$ and $n_{46} = n_{46}(\kappa)$ such that $(\lambda, s^{(n)}(\lambda)) \in I_\kappa^{(n),2}$ for all $\lambda \in [\lambda_{46}, \varepsilon_{41} n^\eta]$ and $n \geq n_{46}$. Then for $n \geq n_{46}$ and $\lambda \in [\lambda_{46}, \varepsilon_{41} n^\eta]$,

$$
\Phi_\lambda^{(n)}\big((1-\kappa)s^{(n)}(\lambda)\big) \geq (1-\kappa)s^{(n)}(\lambda)\left(\lambda - \frac{1}{1+\lambda n^{-\eta}} \cdot \frac{\varphi_\lambda^{(n)}(s^{(n)}(\lambda))}{1+C_{45}\kappa}\right)
$$
$$
= (1-\kappa)s^{(n)}(\lambda) \cdot \lambda \cdot \left(\frac{\kappa C_{45}}{1+\kappa C_{45}}\right) \geq C_{46}\kappa \lambda^{\frac{\tau-2}{\tau-3}}, \qquad (46)
$$

where the first inequality uses (33) and Lemma 6.10, the second step follows from the definition of $s^{(n)}(\lambda)$ which implies that $\varphi_\lambda^{(n)}(s^{(n)}(\lambda)) = \lambda \cdot (1 + \lambda n^{-\eta})$, and the final step follows from (42). Similarly, for all $n \geq n_{46}$ and $\lambda \in [\lambda_{46}, \varepsilon_{41} n^\eta]$,

$$
\Phi_\lambda^{(n)}\big(\kappa s^{(n)}(\lambda)\big) \geq C_{46}\kappa^2 \lambda^{(\tau-2)/(\tau-3)}. \qquad (47)
$$

Now recall the process $Z_\lambda^{n,-}$ from (29), and note that for any $u \geq 0$,

$$
n^\eta \left(Z_\lambda^{n,-}(u) - \mathbb{E}[Z_\lambda^{n,-}(u)]\right) = \sum_{j=2}^{n} \frac{\theta_{j,\lambda}}{(1+\lambda n^{-\eta})}\left(\mathbb{1}\{\xi_j^n \leq u\} - \mathbb{P}(\xi_j^n \leq u)\right). \qquad (48)
$$

Thus, by Lemma 6.9, there exists $\lambda_{49} \geq \lambda_{41}$ such that $\bar{\forall} n$ and $\lambda \in [\lambda_{49}, \varepsilon_{41} n^\eta]$,

$$
\mathbb{P}\Bigg(\sup_{u \leq s^{(n)}(\lambda)} (\lambda + n^\eta)\left|Z_\lambda^{n,-}(u) - \mathbb{E}[Z_\lambda^{n,-}(u)]\right|
$$
$$
\geq \lambda^{\frac{1}{\tau-3}}\left(s^{(n)}(\lambda)\right)^{\frac{\tau-3}{2}}\Bigg) \leq \exp\left(-C\lambda^{\frac{1}{\tau-3}}\log\log\lambda^{\frac{1}{\tau-3}}\right). \qquad (49)
$$

Using (42), $\bar{\forall} n$ and $\lambda \in [\lambda_{49}, \varepsilon_{41} n^\eta]$, $\lambda^{\frac{1}{\tau-3}}\left(s^{(n)}(\lambda)\right)^{\frac{\tau-3}{2}} \leq C\lambda^{\frac{\tau-1}{2(\tau-3)}}$. Now using (46), (47), and (49) together with the concavity of $\Phi_\lambda^{(n)}$ and the fact that $(\tau-1)/(2(\tau-3)) < (\tau-2)/(\tau-3)$, we can find $\lambda_{50} = \lambda_{50}(\kappa) \geq \lambda_{41}$ and $n_{50} = n_{50}(\kappa)$ such that for all $n \geq n_{50}$ and $\lambda \in [\lambda_{50}, \varepsilon_{41} n^\eta]$,

$$
\mathbb{P}\Bigg(Z_\lambda^{n,-}(u) > 0 \text{ for every } u \in \big[\kappa s^{(n)}(\lambda), (1-\kappa)s^{(n)}(\lambda)\big]\Bigg) \geq 1 - \exp\left(-C\lambda^{\frac{1}{\tau-3}}\right). \qquad (50)
$$

By Lemma 6.6, on the event in (50), there exists a component $\mathscr{C}^\star$ in $\mathscr{G}_n^-(\lambda)$ with $\mathscr{W}_x(\mathscr{C}^\star) \geq (1-2\kappa)s^{(n)}(\lambda)$. Now we can generate $\mathscr{G}_n(\lambda)$ by first generating $\mathscr{G}_n^-(\lambda)$,

and then independently sampling the edges from vertex 1 to the vertex set $[n] \setminus \{1\}$. Thus, for all $n \geq n_{50}$ and $\lambda \in [\lambda_{50}, \, \varepsilon_{41}n^{\eta}]$,

$$\mathbb{P}\left(\mathscr{W}_{\boldsymbol{x}}\big(\mathscr{C}\big(1, \mathscr{G}_n(\lambda)\big)\big) \leq (1-2\kappa)s^{(n)}(\lambda)\right)$$

$$\leq \exp\left(-C\lambda^{\frac{1}{\tau-3}}\right) + \exp\left(-(\lambda + n^{\eta})x_1 \cdot (1-2\kappa)s^{(n)}(\lambda)\right),$$

where the second term is an upper bound for the probability that vertex 1 is **not** connected to the component $\mathscr{C}^{\star}$ of mass at least $(1-2\kappa)s^{(n)}(\lambda)$. Now an application of (42) and (25) completes the proof of Proposition 6.8. $\qquad\square$

Let us now turn to the proof of Lemma 6.10.

Proof of Lemma 6.10: Recall the definition of $I^{(n),2}$ from (35). Let $f : [0, \infty) \to [0, 1]$ be given by $f(u) := 1 - e^{-u} - ue^{-u}$. Note that $f(u) \asymp 1$ for $u \geq 1$ while $f(u) \asymp u^2$ for $u \in [0, 1]$. Thus, $\bar{\forall}n$ and for $(\lambda, u) \in I^{(n),2}$,

$$\sum_{j=2}^{n} \theta_{j,\lambda}\big(1 - e^{-u\theta_{j,\lambda}} - u\theta_{j,\lambda}e^{-u\theta_{j,\lambda}}\big) \asymp \sum_{j=2}^{i_{\lambda}(u)-1} \theta_{j,\lambda} + \sum_{j=i_{\lambda}(u)}^{n} \theta_{j,\lambda}\big(u\theta_{j,\lambda}\big)^2 \asymp u^{\tau-2},$$

(51)

where the last step uses arguments similar to the ones leading to (40). Now for any $u \geq 0$, $\kappa \in (0, 1)$, and $s \in [1 - \kappa, 1]$,

$$\Big|\sum_{j=2}^{n} \theta_{j,\lambda}\big(u\theta_{j,\lambda}e^{-us\theta_{j,\lambda}} - u\theta_{j,\lambda}e^{-u\theta_{j,\lambda}}\big)\Big| \leq \sum_{j=2}^{n} u\theta_{j,\lambda}^2\big(1 - e^{-\kappa u\theta_{j,\lambda}}\big). \qquad (52)$$

For all large n and $(\lambda, \kappa u) \in I^{(n),2}$,

$$\sum_{j=2}^{n} \theta_{j,\lambda}^2 u\big(1 - e^{-\kappa u\theta_{j,\lambda}}\big) \asymp \sum_{j=2}^{i_{\lambda}(\kappa u)-1} \theta_{j,\lambda}^2 u + \sum_{j=i_{\lambda}(\kappa u)}^{n} \theta_{j,\lambda}^2 u\big(\kappa u\theta_{j,\lambda}\big)$$

$$\asymp u\big(i_{\lambda}(\kappa u)\big)^{1-2\alpha} + \kappa u^2 \sum_{j=i_{\lambda}(\kappa u)}^{n} \big(\theta_{j,\lambda}\big)^3 \asymp \kappa^{\tau-3}u^{\tau-2}, \qquad (53)$$

where we have used Assumption 2.4 (iii), (7), and (36). Combining (51), (52), and (53), we see that there exists $\kappa_0 \in (0, 1/4)$ small such that $\bar{\forall}n$ and for all $\kappa \in (0, \kappa_0]$,

$$\sum_{j=2}^{n} \theta_{j,\lambda}\big(1 - e^{-u\theta_{j,\lambda}} - u\theta_{j,\lambda}e^{-us\theta_{j,\lambda}}\big) \asymp u^{\tau-2} \quad \text{for} \quad (\lambda, u) \in I_{\kappa}^{(n),2} \qquad (54)$$

uniformly over $s \in [1 - \kappa, 1]$. Hence, $\bar{\forall} n$, for any $\kappa \in (0, \kappa_0]$, and $(\lambda, u) \in I_\kappa^{(n),2}$,

$$
\begin{aligned}
\frac{\varphi_\lambda^{(n)}(u)}{\varphi_\lambda^{(n)}((1 - \kappa)u)} - 1 &= \frac{\sum_{j=2}^n \theta_{j,\lambda}\left(\kappa\left(1 - e^{-u\theta_{j,\lambda}}\right) + e^{-u\theta_{j,\lambda}} - e^{-(1-\kappa)u\theta_{j,\lambda}}\right)}{(1 - \kappa)u \cdot \varphi_\lambda^{(n)}((1 - \kappa)u)} \\
&\geq \frac{1}{Cu^{\tau-2}} \cdot \left(\sum_{j=2}^n \theta_{j,\lambda}\left(\kappa\left(1 - e^{-u\theta_{j,\lambda}}\right) + e^{-u\theta_{j,\lambda}} - e^{-(1-\kappa)u\theta_{j,\lambda}}\right)\right) \\
&= \frac{1}{Cu^{\tau-2}} \cdot \left(\sum_{j=2}^n \theta_{j,\lambda} \int_{1-\kappa}^1 \left(1 - e^{-u\theta_{j,\lambda}} - u\theta_{j,\lambda}e^{-u\theta_{j,\lambda}s}\right)ds\right) \geq C_{45}\kappa,
\end{aligned}
$$

where the second step uses (40), and the last step uses (54). It thus follows that $\bar{\forall} n$, for all $\kappa \in (0, \kappa_0]$ and $(\lambda, u) \in I_\kappa^{(n),2}$,

$$
\frac{\varphi_\lambda^{(n)}(\kappa u)}{\varphi_\lambda^{(n)}(u)} \leq \frac{\varphi_\lambda^{(n)}((1 - \kappa)u)}{\varphi_\lambda^{(n)}(u)} \leq \frac{1}{1 + C_{45}\kappa},
$$

where the first inequality follows since $\varphi_\lambda^{(n)}(\cdot)$ is an increasing function. This completes the proof of Lemma 6.10. □

Next, we study the sum of squares of weights in the component of vertex 1, as well as the inclusion of maximal weight vertices within this component.

Proposition 6.11 *There exist $\delta_0 > 0$ and $\lambda_{55} \geq \lambda_{41}$ such that the following hold for all large n and for $\lambda \in [\lambda_{55}, \varepsilon_{41}n^\eta]$:*

$$
\mathbb{P}\left(\sum_{j \in \mathscr{C}(1,\mathscr{G}_n(\lambda)), \, j \neq 1} \theta_{j,\lambda}^2 \leq (1 + \delta_0)\lambda\right) \leq \exp(-C\lambda), \quad \text{and} \quad (55)
$$

$$
\mathbb{P}\left(j \notin \mathscr{C}\left(1, \mathscr{G}_n(\lambda)\right) \text{ for some } 1 \leq j \leq \frac{\lambda^{1/\eta}}{\log^3 \lambda}\right) \leq \exp\left(-C\log^{3\alpha}\lambda\right). \quad (56)
$$

Proof of (55): Define $\psi_\lambda^{(n)} : [0, \infty) \to [0, \infty)$ by

$$
\psi_\lambda^{(n)}(u) := \sum_{j=2}^n \theta_{j,\lambda}^2\left(1 - e^{-\theta_{j,\lambda}u}\right).
$$

Since $(1 - e^{-s}) \geq (s - 1 + e^{-s})/s$ for $s \in (0, \infty)$, we have $\psi_\lambda^{(n)}(u) \geq \varphi_\lambda^{(n)}(u)$ for $u \geq 0$. Further, $\bar{\forall} n$ and for $(\lambda, u) \in I^{(n),2}$,

$$
u\left(\psi_\lambda^{(n)}(u) - \varphi_\lambda^{(n)}(u)\right) = \sum_{j=2}^n \theta_{j,\lambda}\left(1 - e^{-u\theta_{j,\lambda}} - u\theta_{j,\lambda}e^{-u\theta_{j,\lambda}}\right) \asymp u^{\tau-2} \asymp u\varphi_\lambda^{(n)}(u),
$$

where the penultimate step uses (51), and the last step uses (40). Hence, there exists $\delta_0 > 0$ such that $\bar{\forall} n$ and for $(\lambda, u) \in I^{(n),2}$, $\psi_\lambda^{(n)}(u) \geq (1 + 3\delta_0)\varphi_\lambda^{(n)}(u)$. Combined with (41), we see that $\bar{\forall} n$ and $\lambda \in [\lambda_{41}, \varepsilon_{41}n^\eta]$,

$$\psi_\lambda^{(n)}\big(s^{(n)}(\lambda)\big) \geq (1 + 3\delta_0)\varphi_\lambda^{(n)}\big(s^{(n)}(\lambda)\big). \tag{57}$$

Let κ_0 be as in Lemma 6.10, and choose $\kappa_1 \in (0, \kappa_0]$ small so that

$$(1 - 2\kappa_1)(1 + 3\delta_0) \geq (1 + 2\delta_0). \tag{58}$$

Recall the process $Z_\lambda^{n,(1)}$ from (30). By Proposition 6.8 and Lemma 6.7, writing $n_{59} = n_{43}(\kappa_1)$ and $\lambda_{59} = \lambda_{43}(\kappa_1)$, we have, for all $n \geq n_{59}$ and $\lambda \in [\lambda_{59},\ \varepsilon_{41}n^\eta]$,

$$\mathbb{P}\big(Z_\lambda^{n,(1)}(\cdot) \text{ hits zero before } (1 - 2\kappa_1)s^{(n)}(\lambda)\big) \leq \exp\big(-C\lambda^{1/(\tau-3)}\big). \tag{59}$$

Consider the coupling between $\mathscr{G}_n(\lambda)$ and the random variables ξ_j^n, $2 \leq j \leq n$, as mentioned below (30). Then note that in this coupling, on the complement of the event in (59), $j \in \mathscr{C}\big(1, \mathscr{G}_n(\lambda)\big)$ whenever $\xi_j^n \leq (1 - 2\kappa_1)s^{(n)}(\lambda)$, and consequently,

$$\sum_{j \in \mathscr{C}(1, \mathscr{G}_n(\lambda)),\, j \neq 1} \theta_{j,\lambda}^2 \geq \sum_{j=2}^n \theta_{j,\lambda}^2 \cdot \mathbb{1}\big\{\xi_j^n \leq (1 - 2\kappa_1)s^{(n)}(\lambda)\big\}$$

$$= \sum_{j=2}^n \theta_{j,\lambda}^2 \bigg(\mathbb{1}\big\{\xi_j^n \leq (1 - 2\kappa_1)s^{(n)}(\lambda)\big\} - \mathbb{P}\big(\xi_j^n \leq (1 - 2\kappa_1)s^{(n)}(\lambda)\big)\bigg)$$

$$+ \psi_\lambda^{(n)}\big((1 - 2\kappa_1)s^{(n)}(\lambda)\big) := \mathfrak{T}_1^{(n)} + \mathfrak{T}_2^{(n)}.$$

To lower bound $\mathfrak{T}_2^{(n)}$, note that $1 - e^{-su} \geq s(1 - e^{-u})$ for all $s \in (0, 1)$ and $u \geq 0$. Thus, $\psi_\lambda^{(n)}\big((1-2\kappa_1)s^{(n)}(\lambda)\big) \geq (1-2\kappa_1)\psi_\lambda^{(n)}\big(s^{(n)}(\lambda)\big)$. Thus, $\bar\forall n$ and $\lambda \in [\lambda_{41},\ \varepsilon_{41}n^\eta]$,

$$\psi_\lambda^{(n)}\big((1 - 2\kappa_1)s^{(n)}(\lambda)\big) \geq (1 - 2\kappa_1)(1 + 3\delta_0)\varphi_\lambda^{(n)}\big(s^{(n)}(\lambda)\big) \geq (1 + 2\delta_0)\lambda, \tag{60}$$

where the first step uses (57), and the second step uses (58) and the definition of $s^{(n)}(\lambda)$.

Turning to $\mathfrak{T}_1^{(n)}$, an application of the bounded difference inequality shows that $\bar\forall n$ and $\lambda \in [1,\ \varepsilon_{41}n^\eta]$,

$$\mathbb{P}\big(\mathfrak{T}_1^{(n)} \leq -\delta_0\lambda\big) \leq \exp\big(-C\lambda^2\big). \tag{61}$$

Combining (59), (60), and (61), it follows that there exists $\lambda_{62} \geq \lambda_{59}$ such that $\bar\forall n$ and $\lambda \in [\lambda_{62},\ \varepsilon_{41}n^\eta]$,

$$\mathbb{P}\bigg(\sum_{j \in \mathscr{C}(1, \mathscr{G}_n(\lambda)),\, j \neq 1} \theta_{j,\lambda}^2 \leq (1 + \delta_0)\lambda\bigg) \leq \exp\big(-C\lambda^{2 \wedge \frac{1}{\tau-3}}\big) \leq \exp\big(-C\lambda\big). \tag{62}$$

This completes the proof of (55). □

Proof of (56): Write $\sum_*$ for $\sum_{j=2}^{\frac{\lambda^{1/\eta}}{\log^3 \lambda}}$, and note that $\bar{\forall} n$ and $\lambda \in [\lambda_{41}, \ \varepsilon_{41} n^\eta]$,

$$\sum_*\mathbb{P}\big(\xi_j^n \geq (1 - 2\kappa_1)s^{(n)}(\lambda)\big) = \sum_* \exp\big(-\theta_{j,\lambda}(1 - 2\kappa_1)s^{(n)}(\lambda)\big)$$

$$\leq \sum_* \exp\Big(-\frac{C}{j^\alpha}\lambda^{\frac{1}{\tau-3}}\Big) \leq \frac{\lambda^{1/\eta}}{\log^3 \lambda} \exp\big(-C\log^{3\alpha}\lambda\big) \leq C' \exp\big(-C''\log^{3\alpha}\lambda\big), \tag{63}$$

where the second step uses (42), and the last step uses the fact $3\alpha > 1$.

Now consider again the coupling between $\mathscr{G}_n(\lambda)$ and the random variables ξ_j^n, $2 \leq j \leq n$, as mentioned below (30). As already noted below (59), in this coupling, on the complement of the event in (59), $j \in \mathscr{C}\big(1, \mathscr{G}_n(\lambda)\big)$ whenever $\xi_j^n \leq (1 - 2\kappa_1)s^{(n)}(\lambda)$. Thus, combining (63) and (59) completes the proof. $\qquad\square$

6.4 Height bounds for a branching process

Recall the definition of ν_n from (18), and define $v_i = w_i/\nu_n$ for $2 \leq i \leq n$. Let V_n be a random variable with distribution

$$\frac{1}{\sum_{i=2}^n v_i} \sum_{i=2}^n v_i \cdot \delta\{v_i\} = \frac{1}{\ell_n} \sum_{i=2}^n w_i \cdot \delta\{v_i\}. \tag{64}$$

Let $\mathrm{Poi}(V_n)$ denote a random variable that conditional on V_n is distributed as a Poisson random variable with mean V_n. We will need the following property of the random variable $\mathrm{Poi}(V_n)$.

Lemma 6.12 *There exist $C'_{65} > C_{65} > 0$ such that for all $n \geq 1$,*

$$C'_{65}/u^{\tau-2} \geq \mathbb{P}\big(\mathrm{Poi}(V_n) \geq u\big) \geq C_{65}/u^{\tau-2} \quad for \quad 1 \leq u \leq v_2/2. \tag{65}$$

The proof of Lemma 6.12 is given in Sect. 1. The main result of this section, stated in the next proposition, describes height asymptotics of a branching process T_n with offspring distribution $\mathrm{Poi}(V_n)$. Note that $\mathbb{E}\big[\mathrm{Poi}(V_n)\big] = 1$, so that T_n is finite almost surely.

Proposition 6.13 *There exists $\gamma_* > 0$ and $C > 0$ such that for $\gamma \in (0, \gamma_*]$ and all $n \geq 1$,*

$$\mathbb{P}\big(\mathrm{ht}(T_n) \geq \gamma n^\eta\big) \leq \frac{C}{n^\alpha \gamma^{1/(\tau-3)}}.$$

The proof of this result uses a technique recently developed in [4]. For a random variable X, define Lévy's concentration function as

$$Q(X, u) := \sup_{x \in \mathbb{R}} \mathbb{P}\big(x \leq X \leq x + u\big), \quad u > 0.$$

Let $(X_i \; ; \; i \geq 1)$ be an i.i.d. sequence with $X_1 \overset{\mathrm{d}}{=} \mathrm{Poi}(V_n) - 1$. Let $(S_k \; ; \; k \geq 0)$ be a random walk with $S_k - S_{k-1} = X_k, k \geq 1$, and for $x \in \mathbb{Z}$, write P_x for the probability distribution of this random walk when $S_0 = x$. By [46, Theorem 3.1] there exists a (universal) constant $C > 0$ such that for any $x \in \mathbb{Z}$, under P_x,

$$Q(S_k, u) \leq \frac{Cu}{\sqrt{k \cdot \mathbb{E}\big[(X_1 - X_2)^2 \mathbb{1}\{|X_1 - X_2| \leq u\}\big]}}, \quad k \geq 1. \tag{66}$$

We start with the following lemma.

Lemma 6.14 *Let* $Y = X_1 - X_2$. *Let* $\beta_* > 0$ *be such that* $\bar{\forall} n$, *one has* $\beta_* n^\alpha \leq v_2/2 - 1$. *Then* $\bar{\forall} n$ *and for any* $u \in [1, \beta_* n^\alpha]$,

$$\mathbb{E}\big[Y^2 \mathbb{1}\{|Y| \leq u\}\big] \geq Cu^{4-\tau}.$$

Proof Choose $\Delta \in (0, 1)$ such that $C_{65} = 2C'_{65}\Delta^{\tau-2}$. Then an application of Lemma 6.12 shows that $\bar{\forall} n$, for any $u \in [2/\Delta, \beta_* n^\alpha]$ and $y \in [1, u\Delta]$,

$$\mathbb{P}\big(\mathrm{Poi}(V_n) \in [y, u]\big) = \mathbb{P}\big(\mathrm{Poi}(V_n) \geq y\big) - \mathbb{P}\big(\mathrm{Poi}(V_n) > u\big) \geq \frac{C_{65}}{y^{\tau-2}} - \frac{C'_{65}}{u^{\tau-2}}$$

$$\geq \frac{C_{65}}{(u\Delta)^{\tau-2}} - \frac{C'_{65}}{u^{\tau-2}} = \frac{C'_{65}}{u^{\tau-2}}. \tag{67}$$

Hence, $\bar{\forall} n$ and for any $u \in [2/\Delta, \beta_* n^\alpha]$,

$$\mathbb{E}\big[Y^2 \mathbb{1}\{|Y| \leq u\}\big] = 2 \int_0^u y\mathbb{P}\big(u \geq |Y| \geq y\big)dy$$

$$\geq 2\mathbb{P}\big(X_2 = -1\big) \int_1^{u\Delta} y\mathbb{P}\big(u \geq X_1 + 1 \geq y\big)dy$$

$$\geq 2\mathbb{P}\big(X_2 = -1\big) \int_1^{u\Delta} \frac{C'_{65}y}{u^{\tau-2}}dy \geq C\mathbb{P}\big(X_2 = -1\big)u^{4-\tau}, \tag{68}$$

where the penultimate step uses (67). Now, using Assumption 2.4,

$$\mathbb{P}\big(X_2 = -1\big) = \mathbb{E}\big[e^{-V_n}\big] = \frac{1}{\ell_n}\sum_{i=2}^n w_i e^{-w_i/v_n} \geq \frac{C}{n}\sum_{i=n/4}^{n/2}\Big(\frac{n}{i}\Big)^\alpha e^{-C'(\frac{n}{i})^\alpha} \geq C'' > 0. \tag{69}$$

This yields the desired result for $u \in [2/\Delta, \beta_* n^\alpha]$. Now, an argument similar to the one used in (69) will show that $\mathbb{P}(X_1 = 0) \geq C > 0$, which would in turn imply that for $u \in [1, 2/\Delta]$,

$$\mathbb{E}\big[Y^2 \mathbb{1}\{|Y| \leq u\}\big] \geq \mathbb{P}(X_1 = 0)\mathbb{P}(X_2 = -1) \geq C' \geq C'u^{4-\tau}(\Delta/2)^{4-\tau}.$$

Using this last observation we extend the lower bound to the interval $u \in [1, \beta_* n^\alpha]$. ☐

Proof of Proposition 6.13: Consider the intervals $I_l = [2^{l-1}, 2^{l+2})$ for $l \geq 1$. Define

$$r_{nl} := \min\left\{k \geq 1 : \sup_{x \in I_l} P_x\left(\bigcap_{j=1}^k \{S_j \in I_l\}\right) \leq 1/2\right\}.$$

By (66) and Lemma 6.14, when $2^{l+2} \leq \beta_* n^\alpha$, we have $\bar\forall n$ and a large constant $C_{70} > 0$,

$$Q\left(S_{C_{70}2^{l(\tau-2)}}, 2^{l+2}\right) \leq \frac{C2^{l+2}}{\sqrt{C_{70}2^{l(\tau-2)}(2^{l+2})^{(4-\tau)}}} \leq 1/2, \tag{70}$$

Hence, $\bar\forall n$,

$$r_{nl} \leq C_{70}2^{l(\tau-2)}, \qquad \text{for } 1 \leq l \leq m(\beta_*), \tag{71}$$

where $m(\beta) := \max\left\{k \in \mathbb{Z}_{>0} : 2^k \leq \lfloor \beta n^\alpha \rfloor\right\}$ for $\beta > 0$.

For the rest of this proof, we will work with the random walk $(S_k ; k \geq 0)$ started at $S_0 = 1$, i.e., we will work under the measure P_1. Let $\xi = \inf\left\{k \geq 0 : S_k = 0\right\}$. By [4, Proposition 1.7],

$$\text{ht}(T_n) \preceq_{\text{sd}} 3\sum_{k=0}^{\xi-1} \frac{1}{S_k} =: 3 \cdot J_n(\xi). \tag{72}$$

Following [4] we derive tail bounds for $J_n(\xi)$ by decomposing the trajectory of the random walk into various "scales" which we now define. Let $\zeta_0 = 0$ and $R_0 = 1$. For $i \geq 0$, define the stopping times $\zeta_{i+1} = \min\left\{t \geq \zeta_i : S_t \notin [2^{R_i-1}, 2^{R_i+2})\right\}$, and let $R_{i+1} = \max\left\{l : S_{\zeta_{i+1}} \geq 2^l\right\}$.

Next, for $0 \leq k \leq \xi$, let $\Lambda(k)$ denote the scale of S at time k. Precisely, let $j = \max\left\{i : \zeta_i \leq k\right\}$ be the most recent epoch for a change in scale, and let $\Lambda(k) = R_j$. Finally, for $l \geq 1$, define

$$J_{nl}(\xi) := \sum_{k=0}^{\xi-1} \frac{1}{S_k} \mathbb{1}\left\{\Lambda(k) = l\right\}.$$

Now consider $\beta \in [4n^{-\alpha}, \beta_*]$, and note that $\bar\forall n$ and for any such choice of $\beta, m(\beta) \geq 2$, where $m(\cdot)$ is as defined below (71). Define

$$b_l := 18\left(m(\beta) - l + 1 + 2\log_2\left(m(\beta) - l + 1\right)\right), \qquad 1 \leq l \leq m(\beta).$$

Then $\bar{\forall}n$,

$$\sum_{l=1}^{m(\beta)} \frac{b_l r_{nl}}{2^{l-1}} \le 18 C_{70} \sum_{l=1}^{m(\beta)} \frac{1}{2^{l-1}} \left(m(\beta) - l + 1 + 2 \log_2 \left(m(\beta) - l + 1 \right) \right) \cdot 2^{l(\tau-2)}$$

$$\le 36 C_{70} 2^{(m(\beta)+1)(\tau-3)} \sum_{j=1}^{\infty} \frac{(j + 2\log_2 j)}{2^{j(\tau-3)}} \le C_{73}(\beta n^\alpha)^{\tau-3}, \qquad (73)$$

where the first step uses (71), and in the second step we have made the substitution $j = m(\beta) - l + 1$.

Note that $\max_{0 \le k \le \xi - 1} \Lambda(k) \le m(\beta)$ on the event $\{ \max_{0 \le k \le \xi} S_k < \lfloor \beta n^\alpha \rfloor \}$. Hence, $\bar{\forall}n$ and for $\beta \in [4n^{-\alpha}, \beta_*]$,

$$P_1 \left(J_n(\xi) \ge C_{73}(\beta n^\alpha)^{\tau-3} \right)$$

$$\le P_1 \left(\max_{0 \le k \le \xi} S_k \ge \lfloor \beta n^\alpha \rfloor \right) + P_1 \left(\sum_{l=1}^{m(\beta)} J_{nl}(\xi) \ge \sum_{l=1}^{m(\beta)} \frac{b_l r_{nl}}{2^{l-1}} \right)$$

$$\le \frac{1}{\lfloor \beta n^\alpha \rfloor} + \sum_{l=1}^{m(\beta)} P_1 \left(J_{nl}(\xi) \ge \frac{b_l r_{nl}}{2^{l-1}} \right) \le \frac{1}{\lfloor \beta n^\alpha \rfloor} + \sum_{l=1}^{m(\beta)} \frac{1}{2^l} \cdot \frac{2}{2^{b_l/18}},$$

where the first step uses (73), the second step follows from a simple application of the optional stopping theorem, and the third step follows from [4, Theorem 3.6]. Using the expression for b_l, the above bound yields, $\bar{\forall}n$ and for $\beta \in [4n^{-\alpha}, \beta_*]$,

$$P_1 \left(J_n(\xi) \ge C_{73}(\beta n^\alpha)^{\tau-3} \right) \le \frac{1}{\lfloor \beta n^\alpha \rfloor} + \sum_{l=1}^{m(\beta)} \frac{2}{2^{m(\beta)+1} \left(m(\beta) - l + 1 \right)^2} \le \frac{C}{\beta n^\alpha}. \qquad (74)$$

To reparametrize from β to γ as in the statement of Proposition 6.13, write $\gamma = 3C_{73}\beta^{\tau-3}$, and $\gamma_* = 3C_{73}\beta_*^{\tau-3}$. Then (72) and (74) give the desired result for all large n and for $\gamma \in [3C_{73}4^{\tau-3}n^{-\eta}, \gamma_*]$. Now choose a larger constant to make the bound work for $\gamma \in (0, 3C_{73}4^{\tau-3}n^{-\eta}]$ and for all $n \ge 1$. This completes the proof. $\square$

6.5 Diameter outside the component of the vertex 1

For $\lambda \ge 0$ and $\delta > 0$, let $H_n(\lambda, \delta)$ be the random graph constructed in the following way: Let the vertex set be

$$\mathscr{V}_\lambda' := [n] \setminus V\left(\mathscr{C}_1^n(\lambda) \right), \qquad (75)$$

and place edges independently between $i, j \in \mathscr{V}_\lambda'$ with probability $1 - \exp\left(- p_{(1+\delta)\lambda}^n w_i w_j / \ell_n \right)$.

Proposition 6.15 *There exist $\delta_1 > 0$, $\varepsilon_{76} \in (0, \varepsilon_{41}]$ and $\lambda_{76} \geq \lambda_{55}$ such that for all large n, for $\lambda \in [\lambda_{76}, \varepsilon_{76} n^\eta]$, and for every $\Delta \in (0, 1/2]$,*

$$\mathbb{P}\left(\operatorname{diam}\left(H_n(\lambda, \delta_1)\right) \geq \frac{n^\eta}{\lambda^{1-\Delta}} \right) \leq C\lambda^{\frac{\tau+1}{\tau-3}} \exp\left(-C'\lambda^\Delta\right). \tag{76}$$

Let us specify here how we choose the thresholds. Choose $\lambda_{76} \geq \lambda_{55}$, $\varepsilon_{76} \in (0, \varepsilon_{41}]$, $\varrho \in (0, 1/2)$, and $\delta_1 > 0$ such that

$$\varrho \cdot \left(\lambda_{76}\right)^{-1/2} \leq \gamma_*, \quad \frac{(1+\delta_0)(1-2\varrho)}{(1+\varepsilon_{76})^2} - (1+\delta_1) \geq \frac{\delta_0}{2}, \quad \text{and} \quad \varepsilon_{76} \leq \frac{\varrho}{10}, \tag{77}$$

where γ_* is as in Proposition 6.13, and δ_0 is as in Proposition 6.11.

Let $i_1 < i_2 < \cdots$ be the vertices in $\mathscr{V}'_\lambda$. For $1 \leq k \leq |\mathscr{V}'_\lambda|$, write $\mathscr{C}_{\mathrm{res}}(i_k; \lambda)$ for the component of i_k in $H_n(\lambda, \delta_1) \backslash \{i_1, \ldots, i_{k-1}\}$. Write

$$\mathbb{P}_1(\cdot) = \mathbb{P}\left(\cdot \mid \mathscr{C}_1^n(\lambda)\right), \quad \text{and} \quad \mathbb{E}_1[\cdot] = \mathbb{E}\left[\cdot \mid \mathscr{C}_1^n(\lambda)\right]. \tag{78}$$

Now,

$$\mathbb{P}_1\left(\operatorname{diam}\left(H_n(\lambda, \delta_1)\right) \geq \frac{n^\eta}{\lambda^{1-\Delta}} \right) \leq \sum_k \mathbb{P}_1\left(\operatorname{diam}\left(\mathscr{C}_{\mathrm{res}}(i_k; \lambda)\right) \geq \frac{n^\eta}{\lambda^{1-\Delta}} \right). \tag{79}$$

For $1 \leq k \leq |\mathscr{V}'_\lambda|$, let

$$\mathscr{V}'_{\lambda,k} := \mathscr{V}'_\lambda \backslash \{i_1, i_2, \ldots, i_{k-1}\}. \tag{80}$$

Note that for any $s > 0$,

$$\operatorname{Bernoulli}\left(1 - e^{-s}\right) \preceq_{\mathrm{sd}} \operatorname{Poi}(s), \tag{81}$$

and consequently, the breadth-first exploration tree of $\mathscr{C}_{\mathrm{res}}(i_k; \lambda)$ starting from i_k is upper bounded by a multitype branching process with state space $\mathscr{V}'_{\lambda,k}$ in which the type of the root is i_k, and any vertex of type i has $\operatorname{Poi}\left(p^n_{(1+\delta_1)\lambda} w_i w_j / \ell_n\right)$ many type j children for $i, j \in \mathscr{V}'_{\lambda,k}$. This leads us to the following definition which will be useful in the proof.

Definition 6.16 For $\lambda \geq 0$, $D \subseteq [n]$, and $i \in D$, let $\operatorname{MTBP}_n^{\lambda,i}(D)$ be a multitype branching process tree with type space D that is rooted at a vertex of type i, and in which a vertex of type j has $\operatorname{Poi}\left(p^n_{(1+\delta_1)\lambda} w_j w_k / \ell_n\right)$ many children of type k for each $j, k \in D$. Let $\operatorname{MTBP}_n^i(D) = \operatorname{MTBP}_n^{0,i}(D)$, and note that in this case a vertex of type j has $\operatorname{Poi}(w_j w_k / \nu_n \ell_n)$ many children of type k.

The bounds in the next lemma will be crucial for dealing with $\operatorname{diam}\left(\mathscr{C}_{\mathrm{res}}(i_k; \lambda)\right)$.

Lemma 6.17 *For any* $\lambda \geq 0$, $1 \leq k \leq |\mathscr{V}'_\lambda|$, *and* $h > 0$,

$$\mathbb{P}_1\big(\mathrm{diam}\left(\mathscr{C}_{\mathrm{res}}(i_k; \lambda)\right) \geq h\big) \leq \mathbb{P}_1\big(\mathrm{ht}\left(\mathrm{MTBP}_n^{\lambda, i_k}\left(\mathscr{V}'_{\lambda, k}\right)\right) \geq h/2\big).$$

Further, for any $i \in D \subseteq [n]$, *and* $h \in \mathbb{Z}_{>0}$,

$$\mathbb{P}\bigg(\mathrm{ht}\left(\mathrm{MTBP}_n^{\lambda, i}(D)\right) \geq h\bigg) \leq \exp\big((1 + \delta_1)h\lambda n^{-\eta}\big) \cdot \mathbb{P}\bigg(\mathrm{ht}\left(\mathrm{MTBP}_n^{i}(D)\right) \geq h\bigg).$$

Proof The first assertion follows from the discussion above Definition 6.16. For the second assertion, write $\zeta = 1 + (1 + \delta_1)\lambda n^{-\eta}$, and note that $\mathrm{MTBP}_n^i(D)$ can be obtained as a subtree of $\mathrm{MTBP}_n^{\lambda, i}(D)$ by killing every child independently with probability $1 - \zeta^{-1}$. Write $\mathscr{A}$ for the event in which $\mathrm{ht}\left(\mathrm{MTBP}_n^{\lambda, i}(D)\right) \geq h$, and no vertex in the leftmost path of length h starting from the root in $\mathrm{MTBP}_n^{\lambda, i}(D)$ is killed. Then

$$\mathbb{P}(\mathscr{A}) = \zeta^{-h}\mathbb{P}\bigg(\mathrm{ht}\left(\mathrm{MTBP}_n^{\lambda, i}(D)\right) \geq h\bigg)$$
$$\geq \exp\big(-(1 + \delta_1)h\lambda n^{-\eta}\big)\mathbb{P}\bigg(\mathrm{ht}\left(\mathrm{MTBP}_n^{\lambda, i}(D)\right) \geq h\bigg).$$

To finish the proof, note that $\mathscr{A}$ implies $\mathrm{ht}\left(\mathrm{MTBP}_n^i(D)\right) \geq h$. □

We next record two properties of the above branching process. Let V_n be as defined around (64).

Lemma 6.18 *Fix* $i \in [n] \setminus \{1\}$. *Consider* $\mathrm{MTBP}_n^i\big([n]\setminus\{1\}\big)$, *and erase the types of all vertices. Then this tree has the same distribution as a branching process tree where the root has* $\mathrm{Poi}\big(w_i/v_n\big)$ *many children, and every other vertex has* $\mathrm{Poi}(V_n)$ *many children.*

This result was noted in the discussion above [65, Proposition 3.2]. We will briefly include the proof.

Proof of Lemma 6.18: Note that $\mathrm{MTBP}_n^i\big([n] \setminus \{1\}\big)$ can be constructed by starting from the root and inductively continuing through the generations as follows: to each vertex of type j assign $\mathrm{Poi}(w_j/v_n)$ many children, and conditional on this step, declare the type of every child independently to be $k \in \{2, \ldots, n\}$ with probability w_k/ℓ_n. Thus, in the tree obtained by erasing the types of every vertex in $\mathrm{MTBP}_n^i\big([n]\setminus\{1\}\big)$, the root has $\mathrm{Poi}(w_i/v_n)$ many children, and every other vertex has $\mathrm{Poi}(w_Y/v_n)$ many children, where $\mathbb{P}\big(Y = j\big) = w_j/\ell_n$, $j = 2, \ldots, n$. The proof is complete upon noting that $w_Y/v_n \overset{\mathrm{d}}{=} V_n$. □

The next lemma follows easily from Definition 6.16.

Lemma 6.19 *Consider* $i \in D \subseteq D' \subseteq [n] \setminus \{1\}$. *Then* $\mathrm{MTBP}_n^i(D)$ *can be coupled with* $\mathrm{MTBP}_n^i(D')$ *so that the former is a subtree of the latter.*

Given $\mathscr{V}'_\lambda$, we will now construct a three layer branching process (L3BP), which we will use to obtain tail bounds on ht $\big(\text{MTBP}_n^{ik}(\mathscr{V}'_{\lambda,k})\big)$. Let ϱ be as in (77), and $\Delta \in (0, 1/2]$ be as in the statement of Proposition 6.15.

Definition 6.20 For $\lambda > 0$ and $1 \le k \le |\mathscr{V}'_\lambda|$, consider the following (potentially) three layer process L3BP$_n^{ik}(\lambda)$:

(a) **Layer 1:** Start MTBP$_n^{ik}(\mathscr{V}'_\lambda)$, and run this process up to generation $(1 - 2\varrho)n^\eta/(2\lambda^{1-\Delta})$. Call this the first layer. If there is at least one vertex in generation $(1 - 2\varrho)n^\eta/(2\lambda^{1-\Delta})$, then we say that the first layer has been fully activated.

(b) **Layer 2:** If the first layer is fully activated, then starting from every vertex v in generation $(1-2\varrho)n^\eta/(2\lambda^{1-\Delta})$, run independent MTBP$_n^{\text{type}(v)}\big([n]\backslash[k]\big)$ processes up to generation $\varrho n^\eta/(2\lambda^{1-\Delta})$, where type$(v) \in [n]$ denotes the type of the vertex v. Call this the second layer. If any of these branching processes survives up to generation $\varrho n^\eta/(2\lambda^{1-\Delta})$, then we say that the second layer has been fully activated; in this case, there is at least one vertex in generation $(1 - \varrho)n^\eta/(2\lambda^{1-\Delta})$ of L3BP$_n^{ik}(\lambda)$.

(c) **Layer 3:** If the second layer is fully activated, then starting from every vertex v in generation $(1 - \varrho)n^\eta/(2\lambda^{1-\Delta})$, run independent MTBP$_n^{\text{type}(v)}\big([n] \backslash \{1\}\big)$ processes. Call this the third layer. If any of these branching processes survives up to generation $\varrho n^\eta/(2\lambda^{1-\Delta})$, then we say that the third layer has been fully activated.

Write $\mathfrak{F}_{n,\lambda}^{ik}$ for the event that all three layers have been fully activated, and note that $\mathfrak{F}_{n,\lambda}^{ik} = \big\{ \text{ht}\big(\text{L3BP}_n^{ik}(\lambda)\big) \ge n^\eta/(2\lambda^{1-\Delta})\big\}$. Since $\mathscr{V}'_{\lambda,k} \subseteq \mathscr{V}'_\lambda$ and $\mathscr{V}'_{\lambda,k} \subseteq [n]\backslash[k] \subseteq [n]\backslash\{1\}$, using Lemma 6.19, ht $\big(\text{MTBP}_n^{ik}(\mathscr{V}'_{\lambda,k})\big) \preceq_{\text{sd}}$ ht $\big(\text{L3BP}_n^{ik}(\lambda)\big)$. In particular,

$$\mathbb{P}_1\left(\text{ht}\big(\text{MTBP}_n^{ik}(\mathscr{V}'_{\lambda,k})\big) \ge \frac{n^\eta}{2\lambda^{1-\Delta}} \right) \le \mathbb{P}_1\big(\mathfrak{F}_{n,\lambda}^{ik}\big). \qquad (82)$$

For $\lambda > 0$, define $\mathfrak{E}_{n,\lambda} := \big\{ \sum_{j=2}^n \theta_{j,\lambda}^2 \mathbb{1}\big\{j \in \mathscr{C}_1^n(\lambda)\big\} \ge (1 + \delta_0)\lambda\big\}$. Note that by (55) and (27), $\bar{\forall}n$ and for $\lambda \in [\lambda_{55}, \ \varepsilon_{41}n^\eta]$,

$$\mathbb{P}\big(\mathfrak{E}_{n,\lambda}^c\big) \le \exp\big(-C\lambda\big). \qquad (83)$$

Lemma 6.21 *We have, for all $\Delta \in (0, 1/2]$, for all large n, and for $\lambda \in [\lambda_{76}, \ \varepsilon_{76}n^\eta]$, on the event $\mathfrak{E}_{n,\lambda}$,*

$$\mathbb{P}_1\big(\mathfrak{F}_{n,\lambda}^{ik}\big) \le C\lambda^{\frac{\tau+1}{\tau-3}} k^{-\frac{\tau+1}{\tau-1}} \exp\left(-\frac{(1 + \delta_0)(1 - 2\varrho)\lambda^\Delta}{2(1 + \varepsilon_{76})^2} \right),$$

for $1 \le k \le |\mathscr{V}'_\lambda|$.

Proof Fix $k \in \{1, \ldots, |\mathscr{V}'_\lambda|\}$ and $\Delta \in (0, 1/2]$. Write $\Gamma_{2 \rightsquigarrow 3}$ for the number of vertices in generation $(1 - \varrho)n^\eta/(2\lambda^{1-\Delta})$ of L3BP$_n^{i_k}(\lambda)$ (i.e., vertices at the end of the second layer), whose subtree in the third layer has height at least $\varrho n^\eta/(2\lambda^{1-\Delta})$. Clearly,

$$\mathbb{P}_1\left(\mathfrak{F}_{n,\lambda}^{i_k}\right) \leq \mathbb{E}_1\left[\Gamma_{2 \rightsquigarrow 3}\right]. \tag{84}$$

Hence, it is enough to find an upper bound for $\mathbb{E}_1\left[\Gamma_{2 \rightsquigarrow 3}\right]$.

For $j \geq 1$, let Gen(j) be the set of vertices in the j-th generation of L3BP$^{i_k}(\lambda)$, and let $Y(j) = \sum_{v \in \text{Gen}(j)} w_{\text{type}(v)}$ denote the sum of weights of vertices in the j-th generation. Conditioning on the first two layers, using Lemma 6.18 for the branching processes in the third layer, and then using Proposition 6.13 for the height of such branching processes, we get, $\bar{\forall} n$ and for all $\lambda \in [\lambda_{76}, \varepsilon_{76}n^\eta]$,

$$\mathbb{E}_1\left[\Gamma_{2 \rightsquigarrow 3}\right] \leq \frac{C}{n^\alpha \left(\varrho/\lambda^{1-\Delta}\right)^{\frac{1}{\tau-3}}} \cdot \mathbb{E}_1\left[Y\left(\frac{(1-\varrho)n^\eta}{2\lambda^{1-\Delta}}\right)\right]. \tag{85}$$

It can be checked by a direct computation that in the second layer,

$$\mathbb{E}_1\left[Y\left(\frac{(1-\varrho)n^\eta}{2\lambda^{1-\Delta}}\right)\right] = R_k \mathbb{E}_1\left[Y\left(\frac{(1-\varrho)n^\eta}{2\lambda^{1-\Delta}} - 1\right)\right]$$
$$= \ldots = (R_k)^{\frac{\varrho n^\eta}{2\lambda^{1-\Delta}}} \mathbb{E}_1\left[Y\left(\frac{(1-2\varrho)n^\eta}{2\lambda^{1-\Delta}}\right)\right], \tag{86}$$

where $R_k = \sum_{j=k+1}^n w_j^2/(v_n \ell_n)$. Similarly, in the first layer,

$$\mathbb{E}_1\left[Y\left(\frac{(1-2\varrho)n^\eta}{2\lambda^{1-\Delta}}\right)\right] = \left(\sum_{j \in \mathscr{V}'_\lambda} \frac{w_j^2}{v_n \ell_n}\right)\mathbb{E}_1\left[Y\left(\frac{(1-2\varrho)n^\eta}{2\lambda^{1-\Delta}} - 1\right)\right]$$
$$= \ldots = \left(\sum_{j \in \mathscr{V}'_\lambda} \frac{w_j^2}{v_n \ell_n}\right)^{\frac{(1-2\varrho)n^\eta}{2\lambda^{1-\Delta}}} \cdot w_{i_k},$$

which combined with (86) and (85) yields, $\bar{\forall} n$ and for all $\lambda \in [\lambda_{76}, \varepsilon_{76}n^\eta]$,

$$\mathbb{E}_1\left[\Gamma_{2 \rightsquigarrow 3}\right] \leq C\lambda^{\frac{1}{\tau-3}} \times \left(\frac{w_k}{n^\alpha}\right) \times \left(\sum_{j \in \mathscr{V}'_\lambda} \frac{w_j^2}{v_n \ell_n}\right)^{\frac{(1-2\varrho)n^\eta}{2\lambda^{1-\Delta}}} \times \left(\sum_{j=k+1}^n \frac{w_j^2}{v_n \ell_n}\right)^{\frac{\varrho n^\eta}{2\lambda^{1-\Delta}}}, \tag{87}$$

where we have used the relation $w_{i_k} \leq w_k$. Writing $\wedge k$ for $k \wedge (n/2)$, we have

$$\sum_{j=k+1}^{n} \frac{w_j^2}{v_n \ell_n} = 1 - \sum_{j=2}^{k} \frac{w_j^2}{v_n \ell_n} \leq 1 - \frac{C}{n^{1-2\alpha}} \sum_{j=2}^{\wedge k} \frac{1}{j^{2\alpha}}$$
$$\leq 1 - \frac{C'(\wedge k)^{\eta}}{n^{\eta}} \leq 1 - \frac{C'(k/2)^{\eta}}{n^{\eta}}, \tag{88}$$

where the first step uses the relation

$$\sum_{j=2}^{n} w_j^2 / (v_n \ell_n) = 1, \tag{89}$$

and the second step uses Assumption 2.4. Further, on the event $\mathfrak{E}_{n,\lambda}$,

$$\sum_{j \in \mathscr{C}_1^n(\lambda), \, j \neq 1} w_j^2 \geq \frac{n^{2\alpha} \sigma_2^{(n)}}{(1 + \lambda n^{-\eta})^2} \cdot (1 + \delta_0)\lambda = \frac{v_n \ell_n}{n^{\eta}(1 + \lambda n^{-\eta})^2} \cdot (1 + \delta_0)\lambda.$$

Thus, for $n \geq 2$ and $\lambda \in (0, \, \varepsilon_{76} n^{\eta}]$, on $\mathfrak{E}_{n,\lambda}$,

$$\sum_{j \in \mathscr{V}_{\lambda}'} \frac{w_j^2}{v_n \ell_n} = 1 - \sum_{j \in \mathscr{C}_1^n(\lambda), \, j \neq 1} \frac{w_j^2}{v_n \ell_n} \leq 1 - \frac{(1 + \delta_0)\lambda}{(1 + \lambda n^{-\eta})^2 n^{\eta}} \leq 1 - \frac{(1 + \delta_0)\lambda}{(1 + \varepsilon_{76})^2 n^{\eta}}. \tag{90}$$

Using (87), (88), (90), and the inequality $1 - u \leq e^{-u}$, we get, $\forall n$ and for $\lambda \in [\lambda_{76}, \, \varepsilon_{76} n^{\eta}]$,

$$\mathbb{E}_1[\Gamma_{2 \rightsquigarrow 3}] \leq C\lambda^{\frac{1}{\tau-3}} \times k^{-\alpha} \times \exp\left(-\frac{(1 + \delta_0)(1 - 2\varrho)\lambda^{\Delta}}{2(1 + \varepsilon_{76})^2}\right) \times \exp\left(-\frac{C_{91} k^{\eta}}{\lambda^{1-\Delta}}\right) \tag{91}$$

on the event $\mathfrak{E}_{n,\lambda}$. Since $\sup_{u \geq 0} u^{\frac{\tau}{\tau-3}} e^{-u} < \infty$, and $\lambda_{76} \geq \lambda_{41} \geq 1$,

$$\exp\left(-\frac{C_{91} k^{\eta}}{\lambda^{1-\Delta}}\right) \leq \exp\left(-\frac{C_{91} k^{\eta}}{\lambda}\right) \leq C\left(\frac{\lambda}{k^{\eta}}\right)^{\frac{\tau}{\tau-3}}$$

for $\lambda \geq \lambda_{76}$, which combined with (91) yields the desired result. $\qquad \square$

Proof of Proposition 6.15: Combining Lemma 6.17, (82), and Lemma 6.21, we see that $\bar{\forall} n$, for all $\lambda \in [\lambda_{76}, \varepsilon_{76} n^\eta]$, $\Delta \in (0, 1/2]$, and $1 \le k \le |\mathcal{V}'_\lambda|$, on the event $\mathfrak{E}_{n,\lambda}$,

$$
\mathbb{P}_1\left(\text{diam}\left(\mathscr{C}_{\text{res}}(i_k; \lambda) \right) \ge \frac{n^\eta}{\lambda^{1-\Delta}} \right)
$$

$$
\le C\lambda^{\frac{\tau+1}{\tau-3}} k^{-\frac{\tau+1}{\tau-1}} \exp\left(-\frac{(1+\delta_0)(1-2\varrho)\lambda^\Delta}{2(1+\varepsilon_{76})^2} \right) \exp\left(\frac{(1+\delta_1)\lambda^\Delta}{2} \right)
$$

$$
\le C\lambda^{\frac{\tau+1}{\tau-3}} k^{-\frac{\tau+1}{\tau-1}} \exp\left(-\frac{\delta_0 \lambda^\Delta}{4} \right), \tag{92}
$$

where the last step uses the second relation in (77). Combining (92) with (79) and (83) completes the proof. □

6.6 Maximum surplus outside the component of the vertex 1

Let $H_n(\lambda, \delta_1)$ and ε_{76} be as in the setting of Proposition 6.15. Our aim in this section is to prove the following result.

Proposition 6.22 *There exists* $\lambda_{93} \ge \lambda_{55}$ *such that* $\bar{\forall} n$ *and* $\lambda \in [\lambda_{93}, \varepsilon_{76} n^\eta]$,

$$
\mathbb{P}\left(\text{max sp}\left(H_n(\lambda, \delta_1) \right) \ge 2 \right) \le C/\sqrt{\lambda}. \tag{93}
$$

The proof of this proposition will require some results from [28, 29], which we will now recall briefly. Fix a finite vertex set $\mathcal{V}$ and write $\mathbb{G}_{\mathcal{V}}^{\text{con}}$ for the space of all simple connected graphs with vertex set $\mathcal{V}$. For fixed $a > 0$ and probability mass function $\boldsymbol{q} = (q_v, v \in \mathcal{V})$, define a probability distribution $\mathbb{P}_{\text{con}}(\cdot; \boldsymbol{q}, a, \mathcal{V})$ on $\mathbb{G}_{\mathcal{V}}^{\text{con}}$ as follows:

$$
\mathbb{P}_{\text{con}}(G; \boldsymbol{q}, a, \mathcal{V}) := \frac{1}{Z(\boldsymbol{q}, a)} \prod_{(i,j) \in E(G)} \left(1 - \exp(-aq_i q_j) \right)
$$

$$
\prod_{(i,j) \notin E(G)} \exp(-aq_i q_j), \quad \text{for } G \in \mathbb{G}_{\mathcal{V}}^{\text{con}}, \tag{94}
$$

where $Z(\boldsymbol{q}, a)$ is a normalizing constant so that $\mathbb{P}_{\text{con}}(\mathbb{G}_{\mathcal{V}}^{\text{con}}; \boldsymbol{q}, a, \mathcal{V}) = 1$.

For $t \ge 0$, consider the random graph $\mathscr{G}\left([n], \boldsymbol{w}^{(n)} \right), t)$ from Definition 2.2, and write $(\mathscr{C}_i, i \ge 1)$ for its components in decreasing order of their masses. Let $\mathcal{V}^{(i)} := V(\mathscr{C}_i)$ be the vertex set of $\mathscr{C}_i, i \ge 1$, and note that $(\mathcal{V}^{(i)}; i \ge 1)$ is a random partition of $[n]$.

Proposition 6.23 ([28, Proposition 6.1]) *Conditional on the partition* $(\mathcal{V}^{(i)}; i \ge 1)$, *define*

$$
\boldsymbol{q}^{(i)} := \left(\frac{w_v}{\sum_{v \in \mathcal{V}^{(i)}} w_v}; v \in \mathcal{V}^{(i)} \right), \quad and \quad a^{(i)} := t \left(\sum_{v \in \mathcal{V}_{(i)}} w_v \right)^2, \quad i \ge 1.
$$

For each fixed $i \geq 1$, let $G_i \in \mathbb{G}_{\mathscr{V}^{(i)}}^{\mathrm{con}}$ be a connected simple graph with vertex set $\mathscr{V}^{(i)}$. Then

$$\mathbb{P}\left(\mathscr{C}_i = G_i \ \forall i \geq 1 \mid \left(\mathscr{V}^{(i)}; \ i \geq 1\right)\right) = \prod_{i \geq 1} \mathbb{P}_{\mathrm{con}}\left(G_i \ ; \ \boldsymbol{q}^{(i)}, a^{(i)}, \mathscr{V}^{(i)}\right).$$

Thus, the random graph $\mathscr{G}\left(([n], \boldsymbol{w}^{(n)}), t\right)$ can be generated in two stages: **(i) Stage I:** Generate the partition of the vertices into different components, i.e., generate $\left(\mathscr{V}^{(i)}; \ i \geq 1\right)$. **(ii) Stage II:** Conditional on the partition, generate the internal structure of each component according to the law $\mathbb{P}_{\mathrm{con}}\left(\cdot \ ; \ \boldsymbol{q}^{(i)}, a^{(i)}, \mathscr{V}^{(i)}\right)$ independently across different components.

In Proposition 6.24 given below we will describe an algorithm to generate such connected components. To state this result, we need some definitions.

For fixed $m \geq 1$, write $\mathbb{T}_m$ for the set of all rooted trees with vertex set $[m]$. Let $\mathbb{T}_m^{\mathrm{ord}}$ be the collection of all plane trees with m vertices where the vertices are labeled by elements of $[m]$. Thus, an element of $\mathbb{T}_m^{\mathrm{ord}}$ is a rooted tree with vertex set $[m]$ where the children of each vertex are arranged from right to left. For $\mathbf{t} \in \mathbb{T}_m^{\mathrm{ord}}$ and $i \in [m]$, let $\mathcal{P}(i; \mathbf{t})$ be the following set of vertices: $j \in \mathcal{P}(i; \mathbf{t})$ if and only if the parent of j is a strict ancestor of i in $\mathbf{t}$, and j lies on the right of the path connecting i to the root of $\mathbf{t}$. Let

$$\mathcal{P}(\mathbf{t}) := \left\{(i, j) \ : \ i \in [m], \ j \in \mathcal{P}(i; \mathbf{t})\right\}.$$

For a probability mass function $\boldsymbol{q} = (q_1, \ldots, q_m)$ with $q_i > 0$ for all $i \in [m]$ and $a > 0$, define

$$L(\mathbf{t}) = L(\mathbf{t}; a, \boldsymbol{q}) := \prod_{\{i,j\} \in E(\mathbf{t})} \left[\frac{\exp(a q_i q_j) - 1}{a q_i q_j}\right] \exp\left(\sum_{(i,j) \in \mathcal{P}(\mathbf{t})} a q_i q_j\right), \quad \mathbf{t} \in \mathbb{T}_m^{\mathrm{ord}}. \tag{95}$$

Consider $\mathbf{t} \in \mathbb{T}_m^{\mathrm{ord}}$, and suppose the vertices of $\mathbf{t}$, arranged in a depth-first order, are $v(1), \ldots, v(m)$, with $v(1)$ being the root. Define the function $f_{\mathbf{t}}(\cdot) = f_{\mathbf{t}}(\cdot \ ; a, \boldsymbol{q})$ on $[0, 1]$ as follows:

$$f_{\mathbf{t}}(s) = \sum_{j \in \mathcal{P}(v(i); \mathbf{t})} q_j, \quad \text{if} \quad \sum_{k \leq i-1} q_{v(k)} \leq s < \sum_{k \leq i} q_{v(k)},$$

and $f_{\mathbf{t}}(1) = 0$. Clearly,

$$\sum_{(i,j) \in \mathcal{P}(\mathbf{t})} q_i q_j = \int_0^1 f_{\mathbf{t}}(s) ds. \tag{96}$$

Now, since $1 + s \leq e^s \leq 1 + se^s$ for any $s \geq 0$,

$$1 \leq \prod_{\{i,j\}\in E(\mathbf{t})} \left[\frac{\exp(aq_i q_j) - 1}{aq_i q_j} \right] \leq \prod_{\{i,j\}\in E(\mathbf{t})} \exp\left(aq_i q_j\right)$$

$$= \exp\left(a \sum_{\{i,j\}\in E(\mathbf{t})} q_i q_j \right) \leq \exp\left(aq_{\max}\right),$$

where $q_{\max} = \max_{j\in[m]} q_j$. Hence,

$$1 \leq L(\mathbf{t}) \leq \exp\left(aq_{\max}\right) \cdot \exp\left(a \int_0^1 f_{\mathbf{t}}(s)ds \right). \tag{97}$$

Associated to the probability mass function $\mathbf{q}$ there is a random tree model called a $\mathbf{q}$-tree [37, 70] which we now define. (This random tree is usually referred to as a $\mathbf{p}$-tree, but we instead use $\mathbf{q}$ to avoid confusion with p_λ^n as defined in (19).) For $\mathbf{t} \in \mathbb{T}_m$ and $v \in [m]$, write $d_v(\mathbf{t})$ for the number of children of v in tree $\mathbf{t}$. Then the law of the $\mathbf{q}$-tree, denoted by P_{tree}, is defined as follows:

$$P_{\text{tree}}(\mathbf{t}) = P_{\text{tree}}(\mathbf{t}; \mathbf{q}) = \prod_{v\in[m]} q_v^{d_v(\mathbf{t})}, \qquad \mathbf{t} \in \mathbb{T}_m. \tag{98}$$

Generating a random $\mathbf{q}$-tree having distribution P_{tree}, and then assigning a uniform random order to the children of every vertex in this tree gives a random element in $\mathbb{T}_m^{\text{ord}}$ with law $P_{\text{ord}}(\cdot\,; \mathbf{q})$ given by

$$P_{\text{ord}}(\mathbf{t}) = P_{\text{ord}}(\mathbf{t}; \mathbf{q}) = \prod_{v\in[m]} \frac{q_v^{d_v(\mathbf{t})}}{(d_v(\mathbf{t}))!}, \qquad \mathbf{t} \in \mathbb{T}_m^{\text{ord}}. \tag{99}$$

Using $L(\cdot)$ to tilt P_{ord} results in the following distribution:

$$P_{\text{ord}}^\star(\mathbf{t}) := P_{\text{ord}}(\mathbf{t}) \cdot \frac{L(\mathbf{t})}{\mathbb{E}[L(T_{\mathbf{q}})]}, \qquad \mathbf{t} \in \mathbb{T}_m^{\text{ord}}, \tag{100}$$

where $T_{\mathbf{q}} \sim P_{\text{ord}}$.

Proposition 6.24 ([28, Proposition 7.4]) *Fix $m \geq 1$, $a > 0$, and a probability mass function $\mathbf{q}$ on $[m]$ with $\min_i q_i > 0$. Construct a random connected graph on $[m]$ as follows:*

(a) *First generate $T_{\mathbf{q}}^\star$ having distribution $P_{\text{ord}}^\star$ as in (100).*
(b) *Conditional on $T_{\mathbf{q}}^\star$, add the edge $\{i, j\}$ with probability $1 - \exp\left(- aq_i q_j\right)$ independently for $(i, j) \in \mathcal{P}(T_{\mathbf{q}}^\star)$.*

Then the resulting random graph is distributed as $\mathbb{P}_{\text{con}}(\cdot\,; \mathbf{q}, a, [m])$ on $\mathbb{G}_{[m]}^{\text{con}}$.

Note that in the above construction, the number of surplus edges in the resulting graph equals the number of edges added in part (b) of the construction. Using (81) and (96), we see that the number of surplus edges in the random graph resulting in the above construction is stochastically dominated by Y, where

$$\left(Y \mid T_{\boldsymbol{q}}^{\star}\right) \sim \mathrm{Poi}\left(a \int_0^1 f_{T_{\boldsymbol{q}}^{\star}}(s)ds\right).$$

Hence, if $\mathscr{G}_{\mathrm{con}}$ is distributed as $\mathbb{P}_{\mathrm{con}}\left(\,\cdot\,; \boldsymbol{q}, a, [m]\right)$, then by Proposition 6.24,

$$\mathbb{P}\left(\mathrm{sp}\left(\mathscr{G}_{\mathrm{con}}\right) \geq 2\right) \leq \mathbb{P}\left(Y \geq 2\right) \leq a^2 \cdot \mathbb{E}\left[\|f_{T_{\boldsymbol{q}}^{\star}}\|_{\infty}^2\right] = a^2 \cdot \frac{\mathbb{E}\left[\|f_{T_{\boldsymbol{q}}}\|_{\infty}^2 \cdot L(T_{\boldsymbol{q}})\right]}{\mathbb{E}\left[L(T_{\boldsymbol{q}})\right]}$$

$$\leq a^2 \exp\left(aq_{\max}\right) \cdot \mathbb{E}\left[\|f_{T_{\boldsymbol{q}}}\|_{\infty}^2 \cdot \exp\left(a\|f_{T_{\boldsymbol{q}}}\|_{\infty}\right)\right], \tag{101}$$

where $T_{\boldsymbol{q}} \sim P_{\mathrm{ord}}$. Here, the second step uses the fact that $\mathbb{P}\left(\mathrm{Poi}(s) \geq 2\right) \leq s^2$ for all $s > 0$, and the final step uses (97). The next result gives a tail bound on $\|f_{T_{\boldsymbol{q}}}\|_{\infty}$.

Lemma 6.25 *There exists a universal constant $C > 0$ such that for every $m \geq 1$, probability mass function $\boldsymbol{q}$ on $[m]$ with $\min_i q_i > 0$, and $x \geq \mathrm{e}$,*

$$\mathbb{P}\left(\|f_{T_{\boldsymbol{q}}}\|_{\infty} \geq x\|\boldsymbol{q}\|_2\right) \leq \exp\left(-Cx\log(\log x)\right).$$

Lemma 6.25 follows by combining [28, Lemma 7.9] and [29, Lemma 4.9]. We are now ready to prove Proposition 6.22.

Proof of Proposition 6.22: Let $\mathbb{P}_1(\cdot)$ and $\mathbb{E}_1(\cdot)$ be as in (78). We will write $\sum_1$ for $\sum_{j\in\mathscr{V}_{\lambda}'}$, where $\mathscr{V}_{\lambda}'$ is as in (75). Let $\mathfrak{E}_{n,\lambda}$ be as defined before (83). Then, for $n \geq 2$ and $\lambda \in (0, \varepsilon_{76}n^{\eta}]$, on the event $\mathfrak{E}_{n,\lambda}$,

$$p_{(1+\delta_1)\lambda}^n \cdot \sum_1 \frac{w_j^2}{\ell_n} = \left(1 + \frac{(1+\delta_1)\lambda}{n^{\eta}}\right) \sum_1 \frac{w_j^2}{v_n\ell_n}$$

$$\leq \left(1 + \frac{(1+\delta_1)\lambda}{n^{\eta}}\right)\left(1 - \frac{(1+\delta_0)\lambda}{(1+\varepsilon_{76})^2 n^{\eta}}\right)$$

$$\leq 1 + \frac{\lambda}{n^{\eta}}\left((1+\delta_1) - \frac{(1+\delta_0)}{(1+\varepsilon_{76})^2}\right) \leq 1 - \frac{\lambda\delta_0}{2n^{\eta}}, \tag{102}$$

where the second step uses (90), and the final step uses the second relation in (77), Consequently,

$$
\mathbb{P}_1\left(\sum_{\substack{\mathscr{C} \text{ component} \\ \text{of } H_n(\lambda, \delta_1)}} \mathscr{W}(\mathscr{C})^2 \geq \frac{n^{2\rho}}{\sqrt{\lambda}}\right)
$$

$$
\leq \frac{\sqrt{\lambda}}{n^{2\rho}} \cdot \mathbb{E}_1\left[\sum {}_1 w_v \cdot \mathscr{W}\left(\mathscr{C}(v, H_n(\lambda, \delta_1))\right)\right]
$$

$$
\leq \frac{\sqrt{\lambda}}{n^{2\rho}} \cdot \sum {}_1 w_v \cdot \left(\frac{w_v}{1 - p^n_{(1+\delta_1)\lambda} \cdot \sum {}_1 w_j^2/\ell_n}\right)
$$

$$
\leq \frac{C\sqrt{\lambda}}{n^{2\rho}} \cdot \sum {}_1 w_v^2 \cdot \left(\frac{n^\eta}{\lambda}\right) \leq \frac{C'}{\sqrt{\lambda}}, \tag{103}
$$

where the first step uses Markov's inequality, the second step uses [5, Lemma 4.5], the third step uses (102), and the final step uses the relation $\sum_{j=1}^n w_j^2 = O(n)$.

For $\lambda \geq \lambda_{55}$, define

$$
\mathfrak{E}'_{n,\lambda} = \mathfrak{E}_{n,\lambda} \cap \left\{j \in \mathscr{C}_1^n(\lambda) \text{ for all } 1 \leq j \leq \frac{\lambda^{1/\eta}}{\log^3 \lambda}\right\}. \tag{104}
$$

For $\lambda > 0$, let

$$
\mathfrak{B}_{n,\lambda} := \left\{\sum_{\substack{\mathscr{C} \text{ component} \\ \text{of } H_n(\lambda, \delta_1)}} \mathscr{W}(\mathscr{C})^2 < n^{2\rho}\lambda^{-1/2}\right\}, \tag{105}
$$

and note that $\mathfrak{B}_{n,\lambda}$ is the complement of the event studied in (103). On the event $\mathfrak{B}_{n,\lambda}$,

$$
\mathscr{W}(\mathscr{C}) \leq n^\rho \lambda^{-1/4} \quad \text{for all components } \mathscr{C} \text{ of } H_n(\lambda, \delta_1). \tag{106}
$$

Let $i_1 < i_2 < \dots$ and $\mathscr{C}_{\text{res}}(i_k; \lambda)$ be as defined before (78). Then

$$
\mathbb{P}_1\left(\sum_{j \in \mathscr{C}} \theta_{j,\lambda}^2 \geq \theta_{1,\lambda}^2 + 2(1+\delta_1)\sum_{j=2}^n \theta_{j,\lambda}^3 \quad \text{for some component } \mathscr{C} \text{ of } H_n(\lambda, \delta_1)\right)
$$

$$
\leq \mathbb{P}_1(\mathfrak{B}_{n,\lambda}^c) + \sum_k \mathbb{P}_1\left(\left\{\sum_{j \in \mathscr{C}_{\text{res}}(i_k; \lambda)} \theta_{j,\lambda}^2 \geq \theta_{1,\lambda}^2 + 2(1+\delta_1)\sum_{j=2}^n \theta_{j,\lambda}^3\right\} \bigcap \mathfrak{B}_{n,\lambda}\right). \tag{107}
$$

Using (106), we can choose $\lambda_{93} \geq \lambda_{55}$ such that $\bar{\forall} n$ and for $\lambda \geq \lambda_{93}$, on the event $\mathfrak{B}_{n,\lambda}$,

$$
\mathscr{W}_{\boldsymbol{x}}\left(\mathscr{C}_{\text{res}}(i_k; \lambda)\right) \leq 1 \quad \text{for} \quad 1 \leq k \leq |\mathscr{V}'_\lambda|. \tag{108}
$$

Analogous to the exploration mentioned before Lemma 6.7, conditional on $\mathscr{C}_1^n(\lambda)$, $\mathscr{C}_{\mathrm{res}}(i_k; \lambda)$ can be explored in a breadth-first manner starting from i_k, and $\mathscr{W}_x\big(\mathscr{C}_{\mathrm{res}}(i_k; \lambda)\big)$ will have the same distribution as the hitting time of zero by the corresponding breadth-first walk. Let $\xi_{j,(1+\delta_1)\lambda}^n$, $2 \leq j \leq n$, be as in (28) with λ replaced by $(1+\delta_1)\lambda$. Using (108), $\bar{\forall}n$ and for $\lambda \in [\lambda_{93}, \ \varepsilon_{76}n^\eta]$, for every $1 \leq k \leq |\mathscr{V}_\lambda'|$,

$$
\mathbb{P}_1\Bigg(\Bigg\{\sum_{j \in \mathscr{C}_{\mathrm{res}}(i_k;\lambda)} \theta_{j,\lambda}^2 \geq \theta_{1,\lambda}^2 + 2(1+\delta_1)\sum_{j=2}^n \theta_{j,\lambda}^3\Bigg\} \bigcap \mathfrak{B}_{n,\lambda}\Bigg)
$$

$$
\leq \mathbb{P}_1\Bigg(\theta_{i_k,\lambda}^2 + \sum_{j=i_k}^n \theta_{j,\lambda}^2 \mathbb{1}\{\xi_{j,(1+\delta_1)\lambda}^n \leq 1\} \geq \theta_{1,\lambda}^2 + 2(1+\delta_1)\sum_{j=2}^n \theta_{j,\lambda}^3\Bigg)
$$

$$
\leq \exp\Bigg(-\frac{C}{\sum_{j=i_k}^n \theta_{j,\lambda}^4}\Bigg) \leq \exp\Bigg(-\frac{C'}{\sum_{j \geq i_k} j^{-4\alpha}}\Bigg) \leq \exp\Big(-C'' i_k^{4\alpha-1}\Big), \tag{109}
$$

where the second step follows from the bounded difference inequality. Combining (109) with (107) and (103), we see that $\bar{\forall}n$ and for $\lambda \in [\lambda_{93}, \ \varepsilon_{76}n^\eta]$, on the event $\mathfrak{E}_{n,\lambda}'$,

$$
\mathbb{P}_1\Bigg(\sum_{j \in \mathscr{C}} \theta_{j,\lambda}^2 \geq \theta_{1,\lambda}^2 + 2(1+\delta_1)\sum_{j=2}^n \theta_{j,\lambda}^3 \ \ \text{for some component } \mathscr{C} \text{ of } H_n(\lambda, \delta_1)\Bigg)
$$

$$
\leq \frac{C}{\sqrt{\lambda}} + \sum_{k \geq \lambda^{1/\eta}/\log^3 \lambda} \exp\Big(-C' k^{4\alpha-1}\Big) \leq \frac{C''}{\sqrt{\lambda}}. \tag{110}
$$

Now, using Proposition 6.23, conditional on $\mathscr{C}_1^n(\lambda)$, we can generate the components of $H_n(\lambda, \delta_1)$ in two steps: first generate the random partition $\big(\mathscr{V}^{(i)}; \ i \geq 1\big)$ of $\mathscr{V}_\lambda'$ into component vertex sets, and then conditionally generate the internal structure of each component. Here, the relevant parameters are

$$
\boldsymbol{q}^{(i)} = \big(q_v^{(i)}; \ v \in \mathscr{V}^{(i)}\big) := \Bigg(\frac{w_v}{\mathscr{W}(\mathscr{V}^{(i)})}; \ v \in \mathscr{V}^{(i)}\Bigg), \quad \text{and}
$$

$$
a^{(i)} = p_{(1+\delta_1)\lambda}^n \big(\mathscr{W}(\mathscr{V}^{(i)})\big)^2 / \ell_n. \tag{111}
$$

We will write

$$
\mathbb{P}_{\mathrm{ptn}}(\cdot) := \mathbb{P}\big(\ \cdot \ \big| \mathscr{C}_1^n(\lambda), \ (\mathscr{V}^{(i)}; \ i \geq 1)\big), \quad \text{and}
$$

$$
\mathbb{E}_{\mathrm{ptn}}(\cdot) := \mathbb{E}\big(\ \cdot \ \big| \mathscr{C}_1^n(\lambda), \ (\mathscr{V}^{(i)}; \ i \geq 1)\big).
$$

For $\lambda > 0$, let $\mathfrak{B}'_{n,\lambda} := \mathfrak{B}_{n,\lambda} \cap \{ \sum_{j \in \mathscr{C}} w_j^2 \le C_{112} n^{2\alpha}$ for each component $\mathscr{C}$ of $H_n(\lambda, \delta_1) \}$, where

$$C_{112} := \sup \left\{ \left(\theta_{1,\varepsilon_{76}n^\eta}^2 + 2(1 + \delta_1) \sum_{j=2}^n \theta_{j,\varepsilon_{76}n^\eta}^3 \right) \cdot \sigma_2^{(n)} : n \ge 1 \right\}. \tag{112}$$

Writing $q_{\max}^{(i)} := \max \{ q_v^{(i)} : v \in \mathscr{V}^{(i)} \}$, we get, $\bar{\forall} n$ and for $\lambda \in (0, \varepsilon_{76}n^\eta]$, on the event $\mathfrak{B}'_{n,\lambda}$, for $i \ge 1$,

$$a^{(i)} q_{\max}^{(i)} \le a^{(i)} \| \mathbf{q}^{(i)} \|_2 = p_{(1+\delta_1)\lambda}^n \cdot \frac{(\mathscr{W}(\mathscr{V}^{(i)}))^2}{\ell_n} \cdot \frac{(\sum_{v \in \mathscr{V}^{(i)}} w_v^2)^{1/2}}{\mathscr{W}(\mathscr{V}^{(i)})}$$

$$\le C \cdot \frac{\mathscr{W}(\mathscr{V}^{(i)})}{n} \cdot n^\alpha = C \cdot \frac{\mathscr{W}(\mathscr{V}^{(i)})}{n^\rho} \tag{113}$$

$$\le \frac{C}{\lambda^{1/4}}, \tag{114}$$

where the last step uses (106). Hence, $\bar{\forall} n$ and for $\lambda \in [1, \varepsilon_{76}n^\eta]$, on the event $\mathfrak{B}'_{n,\lambda}$, for $i \ge 1$,

$$\mathbb{P}_{\mathrm{ptn}} \left(\mathrm{sp}(\mathscr{C}_i) \ge 2 \right) \le \left(a^{(i)} \right)^2 \cdot \exp \left(a^{(i)} q_{\max}^{(i)} \right)$$
$$\cdot \mathbb{E} \left[\| f_{T_{q^{(i)}}} \|_\infty^2 \cdot \exp \left(a^{(i)} \| f_{T_{q^{(i)}}} \|_\infty \right) \right]$$
$$= \left(a^{(i)} \| \mathbf{q}^{(i)} \|_2 \right)^2 \cdot \exp \left(a^{(i)} q_{\max}^{(i)} \right)$$
$$\cdot \mathbb{E} \left[\frac{\| f_{T_{q^{(i)}}} \|_\infty^2}{\| \mathbf{q}^{(i)} \|_2^2} \cdot \exp \left(a^{(i)} \| \mathbf{q}^{(i)} \|_2 \cdot \frac{\| f_{T_{q^{(i)}}} \|_\infty}{\| \mathbf{q}^{(i)} \|_2} \right) \right]$$
$$\le C \left(\mathscr{W}(\mathscr{V}^{(i)}) \right)^2 n^{-2\rho}, \tag{115}$$

where the first step uses (101), and the final step uses Lemma 6.25 together with the bounds in both (113) and (114). Consequently, $\bar{\forall} n$ and for $\lambda \in [1, \varepsilon_{76}n^\eta]$, on the event $\mathfrak{B}'_{n,\lambda}$,

$$\mathbb{P}_{\mathrm{ptn}} \left(\max \mathrm{sp} \left(H_n(\lambda, \delta_1) \right) \ge 2 \right) \le C \sum_i \frac{(\mathscr{W}(\mathscr{V}^{(i)}))^2}{n^{2\rho}} \le \frac{C}{\sqrt{\lambda}}.$$

Thus, to complete the proof it is enough to show that $\bar{\forall} n$ and $\lambda \in [\lambda_{93}, \varepsilon_{76}n^\eta]$,

$$\mathbb{P}\left((\mathfrak{B}'_{n,\lambda})^c \right) \le C\lambda^{-1/2}. \tag{116}$$

To this end, observe that $\bar{\forall} n$ and for $\lambda \in [\lambda_{93}, \varepsilon_{76}n^\eta]$,

$$\mathbb{P}\left((\mathfrak{B}'_{n,\lambda})^c \right) \le \mathbb{P}\left((\mathfrak{E}'_{n,\lambda})^c \right) + \mathbb{E}\left[\mathbb{1}_{\mathfrak{E}'_{n,\lambda}} \cdot \mathbb{P}_1 \left((\mathfrak{B}'_{n,\lambda})^c \right) \right]$$
$$\le \mathbb{P}\left((\mathfrak{E}'_{n,\lambda})^c \right) + C\lambda^{-1/2} \le \exp \left(- C' \log^{3\alpha} \lambda \right) + \exp(-C'\lambda) + C\lambda^{-1/2},$$

where the second inequality uses (103) and (110), and the final inequality uses Proposition 6.11. This completes the proof of (116), and hence of Proposition 6.22. □

6.7 Proof of Proposition 6.4

Let δ_1 be as chosen in (77). Let δ_2 and ε_{22} be given by

$$\delta_2 := \delta_1/2, \quad \text{and} \quad \varepsilon_{22} := \varepsilon_{76}/2.$$

For $\lambda \geq 0$, let $\bar{H}_n(\lambda, \delta_2)$ be the random graph on the vertex set $\bar{\mathscr{V}}'_\lambda := [n] \backslash V(\bar{\mathscr{C}}^n_1(\lambda))$ obtained by placing edges independently between $i, j \in \bar{\mathscr{V}}'_\lambda$ with probability $1 \wedge (p^n_{(1+\delta_2)\lambda} w_i w_j / \ell_n)$.

Since $1 - e^{-u} \leq u \wedge 1$ for all $u \geq 0$, $\mathscr{C}^n_1(\lambda)$ can be coupled with $\bar{\mathscr{C}}^n_1(\lambda)$ so that the former is a subgraph of the latter. Further, we can choose $\lambda_{117} \geq \lambda_{76} \vee \lambda_{93}$ such that $\bar{\forall} n$ and $\lambda \in [\lambda_{117}, \varepsilon_{22} n^\eta]$,

$$p^n_{(1+\delta_2)\lambda} w_i w_j / \ell_n \leq 1 - \exp\left(- p^n_{(1+\delta_1)\lambda} w_i w_j / \ell_n \right) \tag{117}$$

for all $i, j \in [n]$. Then it follows that $\bar{\forall} n$ and $\lambda \in [\lambda_{117}, \varepsilon_{22} n^\eta]$,

$$d_{\mathrm{H}}\left(\bar{M}^n_\lambda, \bar{M}^n_{(1+\delta_2)\lambda} \right) \preceq_{\mathsf{sd}} 1 + \mathrm{LP}(\bar{H}_n(\lambda, \delta_2)) \preceq_{\mathsf{sd}} 1 + \mathrm{LP}(H_n(\lambda, \delta_1)), \tag{118}$$

where $H_n(\lambda, \delta_1)$ is as in the setting of Proposition 6.15, and $\mathrm{LP}(\cdot)$ is as defined below (12). It is easy to check that for any finite graph H,

$$\begin{aligned} 1 + \mathrm{LP}(H) &\leq 1 + 8 \cdot \mathrm{diam}(H) \cdot \left(\max \mathrm{sp}(H) + 1 \right) \\ &\leq 8\left(\mathrm{diam}(H) + 1 \right)\left(\max \mathrm{sp}(H) + 1 \right). \end{aligned} \tag{119}$$

Combining (118) and (119), we see that for any $\Delta \in (0, 1/2]$, $\bar{\forall} n$ and $\lambda \in [\lambda_{117}, \varepsilon_{22} n^\eta]$,

$$\begin{aligned} \mathbb{P}\left(d_{\mathrm{H}}\left(\bar{M}^n_\lambda, \bar{M}^n_{(1+\delta_2)\lambda} \right) \geq \frac{24 n^\eta}{\lambda^{1-\Delta/2}} \right) \\ \leq \mathbb{P}\left(1 + \mathrm{diam}\left(H_n(\lambda, \delta_1) \right) \geq \frac{n^\eta}{\lambda^{1-\Delta/2}} \right) + \mathbb{P}\left(\max \mathrm{sp}\left(H_n(\lambda, \delta_1) \right) \geq 2 \right) \\ \leq \mathbb{P}\left(\mathrm{diam}\left(H_n(\lambda, \delta_1) \right) \geq \frac{n^\eta}{(2\lambda)^{1-\Delta/2}} \right) + \mathbb{P}\left(\max \mathrm{sp}\left(H_n(\lambda, \delta_1) \right) \geq 2 \right) \\ \leq C\left(\lambda^{\frac{\tau+1}{\tau-3}} \exp\left(- C' \lambda^{\Delta/2} \right) + \lambda^{-1/2} \right) \\ \leq C\lambda^{-1/2} \sup_{u \geq 1}\left[u^{\frac{1}{2} + \frac{\tau+1}{\tau-3}} \exp\left(- C' u^{\Delta/2} \right) + 1 \right], \end{aligned} \tag{120}$$

where the third inequality follows from Proposition 6.15 and Proposition 6.22. Now for any $\lambda_{117} \leq \lambda \leq \varepsilon_{22} n^{\eta}$,

$$d_{\mathrm{H}}\left(\bar{M}_{\lambda}^{n}, \bar{M}_{\varepsilon_{22} n^{\eta}}^{n}\right) \leq \sum_{j=0}^{k_{0}} d_{\mathrm{H}}\left(\bar{M}_{(1+\delta_{2})^{j}\lambda}^{n}, \bar{M}_{(1+\delta_{2})^{j+1}\lambda}^{n}\right), \tag{121}$$

where $k_0 = k_0(n, \lambda) = \min \left\{ j \geq 0 : (1+\delta_2)^{j+1}\lambda \geq \varepsilon_{22} n^{\eta} \right\}$. Using (121) and (120), the proof of Proposition 6.4 can be completed via a simple union bound. Note that

$$\sum_{j=0}^{k_{0}} \frac{24n^{\eta}}{\left((1+\delta_{2})^{j}\lambda\right)^{1-\Delta/2}} \leq \frac{24n^{\eta}}{\lambda^{1-\Delta/2}} \sum_{j=0}^{\infty} \frac{1}{(1+\delta_{2})^{3j/4}} = \frac{C_{122}n^{\eta}}{\lambda^{1-\Delta/2}} \leq \frac{n^{\eta}}{\lambda^{1-\Delta}}, \tag{122}$$

where the last step is true for $\lambda \geq \lambda_{21} = \lambda_{21}(\Delta)$, and $\lambda_{21} \geq \lambda_{117}$ is chosen in a way so that $C_{122} \leq \lambda_{21}^{\Delta/2}$. We then choose $n_{22}(\Delta) \geq 1$ so that the interval $[\lambda_{21}(\Delta), \varepsilon_{22} n^{\eta}]$ is nonempty for $n \geq n_{22}(\Delta)$. The rest of the argument is routine. $\qquad \square$

6.8 Proof of Proposition 6.5

To simplify notation, we will write

$$\gamma_n = \varepsilon_{22} n^{\eta} .$$

Before starting the proof of Proposition 6.5, we need two elementary lemmas.

Lemma 6.26 *For $C > 0$, write $\mathfrak{C}_n(C)$ for the event that $\mathscr{W}(\mathscr{C}) \leq C \log n$ for all components $\mathscr{C}$ of $\bar{G}_n(\gamma_n)$ other than $\mathscr{C}_1^n(\gamma_n)$. Then for every $\kappa > 0$, there exists $C_{123} = C_{123}(\kappa) < \infty$ and $n_{123} = n_{123}(\kappa)$ such that for $n \geq n_{123}$,*

$$\mathbb{P}\big(\mathfrak{C}_n(C_{123})\big) \geq 1 - n^{-\kappa} . \tag{123}$$

The proof of Lemma 6.26 will be given in Sect. 2.

Lemma 6.27 (a) *Suppose $k, m_0 \in \mathbb{Z}_{>0}$, and $x_0, x_1, \ldots, x_k \in (0, 1)$ with $x_0 \leq \min \left\{ x_j : 1 \leq j \leq k \right\}$. Further, assume that $Z_i^{(0)}$, $1 \leq i \leq m_0$, and $Z^{(1)}, \ldots, Z^{(k)}$ are independent random variables such that $Z_i^{(0)} \sim \mathrm{Unif} [x_0, 1]$, $i = 1, \ldots, m_0$, and $Z^{(j)} \sim \mathrm{Unif} [x_j, 1]$, $j = 1, \ldots, k$. Then*

$$\mathbb{P}\left(\min_{1 \leq i \leq m_0} Z_i^{(0)} < \min_{1 \leq j \leq k} Z^{(j)} \right) \geq \frac{m_0}{m_0 + k} .$$

(b) *Suppose in addition to the parameters and random variables in (a), we have, for some $\bar{k}, \bar{m}_0 \in \mathbb{Z}_{>0}$, another collection of numbers $\bar{x}_0, \bar{x}_1, \ldots, \bar{x}_{\bar{k}} \in (0, 1)$ with $\bar{x}_0 \leq \min \left\{ \bar{x}_l : 1 \leq l \leq \bar{k} \right\}$. Further, assume that $\bar{Z}_i^{(0)}$, $1 \leq i \leq \bar{m}_0$, and $\bar{Z}^{(1)}, \ldots, \bar{Z}^{(\bar{k})}$ are independent random variables that are also independent of the collection of random*

variables in (a) *such that* $\bar{Z}_i^{(0)} \sim$ Unif $[\bar{x}_0, 1]$, $i = 1, \ldots, \bar{m}_0$, *and* $\bar{Z}^{(l)} \sim$ Unif $[\bar{x}_l, 1]$, $l = 1, \ldots, \bar{k}$. *Then*

$$\mathbb{P}\left(\min_{1 \le i \le m_0} Z_i^{(0)} \wedge \min_{1 \le i \le \bar{m}_0} \bar{Z}_i^{(0)} < \min_{1 \le j \le k} Z^{(j)} \wedge \min_{1 \le l \le \bar{k}} \bar{Z}^{(l)} \right) \ge \frac{m_0}{m_0 + k} \wedge \frac{\bar{m}_0}{\bar{m}_0 + \bar{k}}.$$

Proof We only prove (b). We start with a specific construction of the random variables in (b). First generate $m_0 + k$ i.i.d. Unif $[x_0, 1]$ random variables $Y_1^{(0)}, \ldots, Y_{m_0}^{(0)}, Y^{(1)}, \ldots, Y^{(k)}$, and independent of this collection, generate $\bar{m}_0 + \bar{k}$ i.i.d. Unif $[\bar{x}_0, 1]$ random variables $\bar{Y}_1^{(0)}, \ldots, \bar{Y}_{\bar{m}_0}^{(0)}, \bar{Y}^{(1)}, \ldots, \bar{Y}^{(\bar{k})}$. Since $x_0 \le \min \{x_j : 1 \le j \le k\}$ and $\bar{x}_0 \le \min \{\bar{x}_l : 1 \le l \le \bar{k}\}$, we can construct the random variables in (b) on the same probability space as the above collection of random variables such that

$$Z_i^{(0)} = Y_i^{(0)} \quad \text{for} \quad i = 1, \ldots, m_0, \quad \bar{Z}_i^{(0)} = \bar{Y}_i^{(0)} \quad \text{for} \quad i = 1, \ldots, \bar{m}_0,$$
$$Z^{(j)} \ge Y^{(j)} \quad \text{for} \quad j = 1, \ldots, k, \quad \text{and} \quad \bar{Z}^{(l)} \ge \bar{Y}^{(l)} \quad \text{for} \quad l = 1, \ldots, \bar{k}.$$

Then note that the probability in Lemma 6.27 (b) is lower bounded by

$$\mathbb{P}\left(\min_{1 \le i \le m_0} Y_i^{(0)} \wedge \min_{1 \le i \le \bar{m}_0} \bar{Y}_i^{(0)} < \min_{1 \le j \le k} Y^{(j)} \wedge \min_{1 \le l \le \bar{k}} \bar{Y}^{(l)} \right) =: \mathbb{P}(\mathfrak{M}),$$

say. Let $\mathcal{M}in := \min \{Y_1^{(0)}, \ldots, Y_{m_0}^{(0)}, Y^{(1)}, \ldots, Y^{(k)}\}$, and similarly define $\overline{\mathcal{M}in}$. By symmetry,

$$\mathbb{P}(\mathfrak{M} \,|\, \mathcal{M}in < \overline{\mathcal{M}in}) = \frac{m_0}{m_0 + k}, \quad \text{and} \quad \mathbb{P}(\mathfrak{M} \,|\, \overline{\mathcal{M}in} < \mathcal{M}in) = \frac{\bar{m}_0}{\bar{m}_0 + \bar{k}}.$$

Hence the claim follows.   □

Given $\bar{G}_n(\gamma_n)$ and the edge weights $U|_{\bar{G}_n(\gamma_n)}$, we can construct the restriction of $\bar{M}^n$ to $\bar{G}_n(\gamma_n)$. For any vertex $v \notin V(\bar{\mathscr{C}}_1^n(\gamma_n))$, in order to find the path in $\bar{M}^n$ that connects $\bar{M}_{\gamma_n}^n$ to the restriction of $\bar{M}^n$ to $\bar{\mathscr{C}}_v^n(\gamma_n)$, we can proceed via the following algorithm.

Algorithm 1 In this algorithm, we will join connected components of $\bar{G}_n(\gamma_n)$ sequentially using edges from $E(\bar{G}_n) \setminus E(\bar{G}_n(\gamma_n))$. We will refer to a collection of connected components joined by such edges as a "cluster."

(a) Look at the edges in $E(\bar{G}_n) \setminus E(\bar{G}_n(\gamma_n))$ going out of $\bar{\mathscr{C}}_v^n(\gamma_n)$. Add the edge with the minimum weight to $\bar{\mathscr{C}}_v^n(\gamma_n)$, thereby connecting it to another component of $\bar{G}_n(\gamma_n)$.
(b) Repeat sequentially with the current cluster of v. At the $k + 1$-th step, we look at the edges (if any) that are in $E(\bar{G}_n) \setminus E(\bar{G}_n(\gamma_n))$, and have one endpoint in the current cluster of v (i.e., the cluster after the addition of the k-th edge) and one

endpoint outside of this cluster. We then add the edge with the minimum weight among all the outgoing edges from the cluster, i.e., including the ones considered in the first k steps.

(c) Stop if v gets connected to $\bar{\mathscr{C}}_1^n(\gamma_n)$, or if there are no outgoing edges from the current cluster of v. In the latter case, $v \notin \mathscr{C}(1, \bar{G}_n)$.

Using Lemma 4.3, it is easy to see that if $v \in \mathscr{C}(1, \bar{G}_n)$, then every edge added in Algorithm 1 is an edge in $\bar{M}^n$. Now, the two-layered randomness present in the problem (presence/absence of edges and edge weights) makes Algorithm 1 hard to analyze directly. Loosely speaking, the problem arises from the fact that if $\mathscr{W}(\mathscr{C})$ is small for some component $\mathscr{C}$ of $\bar{G}_n(\gamma_n)$ other than $\bar{\mathscr{C}}_1^n(\gamma_n)$, then we cannot say that there are enough edges in $E(\bar{G}_n) \setminus E(\bar{G}_n(\gamma_n))$ that connect $\mathscr{C}$ to $\bar{\mathscr{C}}_1^n(\gamma_n)$ with high probability– something we need for the argument to work. So instead we will work with the following modified algorithm. Let

$$t_n := \frac{n^{\eta/4}}{(\log n)^{1/3}}. \tag{124}$$

Algorithm 2 The word "cluster" will have the same meaning as in Algorithm 1.

(a) Look at the edges in $E(\bar{G}_n) \setminus E(\bar{G}_n(\gamma_n))$ that connect $\bar{\mathscr{C}}_v^n(\gamma_n)$ to $[n] \setminus (V(\bar{\mathscr{C}}_v^n(\gamma_n)) \cup V(\bar{\mathscr{C}}_1^n(\gamma_n)))$, i.e., we do not check for edges that connect $\bar{\mathscr{C}}_v^n(\gamma_n)$ to $\bar{\mathscr{C}}_1^n(\gamma_n)$. Add the edge with the minimum weight to $\bar{\mathscr{C}}_v^n(\gamma_n)$, thereby connecting it to another component of $\bar{G}_n(\gamma_n)$.

(b) Repeat for another $(t_n - 1)$ many steps in a manner similar to Algorithm 1, but without checking for edges that connect the current cluster to $\bar{\mathscr{C}}_1^n(\gamma_n)$, or until there are no outgoing edges from the current cluster to vertices that are not in $\bar{\mathscr{C}}_1^n(\gamma_n)$.

(c) After the t_n-th edge has been added, if the algorithm has not terminated already, look at the edges (if any) that are in $E(\bar{G}_n) \setminus E(\bar{G}_n(\gamma_n))$, and have one endpoint in the current cluster and one endpoint outside of this cluster. Let us emphasize that at this step, we are checking for edges that connect the current cluster to $\bar{\mathscr{C}}_1^n(\gamma_n)$. We then add the edge with the minimum weight among all the outgoing edges from the current cluster, thereby connecting it to another component of $\bar{G}_n(\gamma_n)$. Stop if this component is $\bar{\mathscr{C}}_1^n(\gamma_n)$.

(d) Else, ignore the edges (if any) found at the $(t_n + 1)$-th step between $\bar{\mathscr{C}}_1^n(\gamma_n)$ and the cluster of v, and continue as in (b) for another t_n many steps. Thus, for $j = t_n + 2, \ldots, 2t_n + 1$, the j-th edge added will be between the current cluster of v and a vertex that is not in $\bar{\mathscr{C}}_1^n(\gamma_n)$. At step $2(t_n + 1)$, again check for new edges that connect the current cluster of v to its complement (including edges that connect the cluster of v to $\bar{\mathscr{C}}_1^n(\gamma_n)$), and at this step the edges found in step $(t_n + 1)$ that connect to $\bar{\mathscr{C}}_1^n(\gamma_n)$ will again be considered.

(e) Continue while checking for possible new edges that connect the current cluster of v to $\bar{\mathscr{C}}_1^n(\gamma_n)$ every $(t_n + 1)$ steps. Stop if we connect to $\bar{\mathscr{C}}_1^n(\gamma_n)$ (this is only possible at step $j(t_n + 1)$ for some $j \geq 1$), or if no new edge can be added to the current cluster.

The next lemma states a simple property relating the two algorithms.

Lemma 6.28 *For* $v \notin \bar{\mathscr{C}}_1^n(\gamma_n)$ *and any* $i_0 \geq 1$, *if* v *gets connected to* $\bar{\mathscr{C}}_1^n(\gamma_n)$ *after the addition of* $i_0 + 1$ *many edges in Algorithm* 1, *then Algorithm* 1 *and Algorithm* 2 *coincide up to the addition of* i_0 *edges.*

Define the event

$$
\mathfrak{B}_n := \left\{ \mathscr{W}\left(\bar{\mathscr{C}}_1^n(\gamma_n)\right) \geq (1 - 2\kappa_0)s^{(n)}(\gamma_n) \cdot n^\rho \cdot \left(\sigma_2^{(n)}\right)^{1/2}, \right.
$$

$$
w_n|\mathscr{C}| \leq \mathscr{W}(\mathscr{C}) \leq C_{123}(2) \cdot \log n \text{ for all other components } \mathscr{C} \text{ of } \bar{G}_n(\gamma_n), \text{ and}
$$

$$
\left. \text{degree of all } v \in [n] \setminus V\left(\bar{\mathscr{C}}_1^n(\gamma_n)\right) \text{ in } \bar{G}_n \text{ is } \leq (\log n)^{3/2} \right\}. \quad (125)
$$

Here, the first line corresponds to the event of interest in Proposition 6.8, and the second line corresponds to the event in Lemma 6.26 with $\kappa = 2$. By (56), $\bar{\forall} n$,

$$
\mathbb{P}\left(j \in \bar{\mathscr{C}}_1^n(\gamma_n) \text{ for } 1 \leq j \leq \frac{\gamma_n^{1/\eta}}{(\log \gamma_n)^3} \right) \geq 1 - \exp\left(-C \log^{3\alpha} n\right). \quad (126)
$$

By Assumption 2.4 (iii) and (7), for any $j > \gamma_n^{1/\eta}(\log \gamma_n)^{-3}$, $w_j \leq 2^\alpha A_2(n/j)^\alpha \leq C \log^{3\alpha} n$ for all large n. Now the degree of a vertex v in $\bar{G}_n$ is distributed as $\sum_{j \neq v} \text{Bernoulli}\left(w_v w_j / \ell_n\right)$, where the summands are independent random variables. Thus, using (126), a union bound and Bennett's inequality [31] shows that $\bar{\forall} n$,

$$
\mathbb{P}\left(\text{degree of some } v \in [n] \setminus V\left(\bar{\mathscr{C}}_1^n(\gamma_n)\right) \text{ in } \bar{G}_n \text{ is } > (\log n)^{3/2} \right)
$$

$$
\leq n \exp(-C \log^{3\alpha} n). \quad (127)
$$

Combining (127), Proposition 6.8, and Lemma 6.26 yields, $\bar{\forall} n$,

$$
\mathbb{P}\left(\mathfrak{B}_n^c\right) \leq 2/n^2. \quad (128)
$$

Write

$$
\mathbb{P}_2(\cdot) = \mathbb{P}\left(\cdot \mid \left(\bar{G}_n(\gamma_n), \ \bar{G}_n \text{ restricted to } [n] \setminus V\left(\bar{\mathscr{C}}_1^n(\gamma_n)\right)\right)\right),
$$

and let $\mathbb{E}_2[\cdot]$ denote the corresponding expectation. Under $\mathbb{P}_2$, for any realization of $\left(\bar{G}_n(\gamma_n), \ \bar{G}_n \text{ restricted to } [n] \setminus V\left(\bar{\mathscr{C}}_1^n(\gamma_n)\right)\right)$, the rest of the edges in $\bar{G}_n$ and the weights associated to the edges in $E(\bar{G}_n) \setminus E(\bar{G}_n(\gamma_n))$ can be generated as follows: Let

$$
E_\star = \left\{ \{i, j\} \ : \ 1 \leq i \neq j \leq n, \text{ at least one of } i \text{ or } j \text{ is in } V\left(\bar{\mathscr{C}}_1^n(\gamma_n)\right), \right.
$$

$$
\left. \text{and } \{i, j\} \notin E\left(\bar{\mathscr{C}}_1^n(\gamma_n)\right) \right\}.
$$

Let $\bar{q}_{ij}$ be as in (17). Place the edges $\{i, j\} \in E_\star$ independently with respective probabilities

$$\frac{(1 - p_{\gamma_n}^n)\bar{q}_{ij}}{(1 - p_{\gamma_n}^n)\bar{q}_{ij} + (1 - \bar{q}_{ij})} = (1 + O(n^{-\eta})) \cdot (1 - p_{\gamma_n}^n)\frac{w_i w_j}{\ell_n}, \qquad (129)$$

where the $O(n^{-\eta})$ term is uniform over $\{i, j\} \in E_\star$, and $p_\bullet^n$ is as in (19). Adding these edges would generate the complete set of edges $E(\bar{G}_n)$. Now assign i.i.d. Unif $[p_{\gamma_n}^n, 1]$ weights to the edges in $E(\bar{G}_n) \setminus E(\bar{G}_n(\gamma_n))$.

Fix a vertex $v \notin \bar{\mathscr{C}}_1^n(\gamma_n)$. Suppose Algorithm 2 started from $\bar{\mathscr{C}}_1^n(\gamma_n)$ terminates after the addition of K edges; let $j_0 \geq 0$ be such that $j_0(t_n + 1) \leq K < (j_0 + 1)(t_n + 1)$. Let $T_j := (j \wedge j_0)(t_n + 1)$ for $j \geq 0$. If the path in $\bar{M}^n$ connecting v to $\bar{M}_{\gamma_n}^n$ has length $\geq (\log n)^{-1/6} n^\eta$, then $\bar{\forall} n$, on the event $\mathfrak{B}_n$, Algorithm 1 started from $\bar{\mathscr{C}}_v^n(\gamma_n)$ will run for at least

$$\frac{n^\eta}{(\log n)^{1/6}} \times \frac{w_n}{w_n + C_{123}(2)\log n} \geq \frac{Cw_n n^\eta}{(\log n)^{7/6}} + 1$$

many steps. Using Lemma 6.28, we see that Algorithm 2 started from $\bar{\mathscr{C}}_v^n(\gamma_n)$ will run for at least $Cw_n n^\eta (\log n)^{-7/6}$ many steps. Using (124) and Assumption 2.4(iv), we see that in this case, for all large n, the events $\mathfrak{A}_j$, $j = 1, 2, \dots, n^{\eta/2}(\log n)^{7/12}$, will take place, where

$$\mathfrak{A}_j = \{T_j = j(t_n + 1) \text{ and the } T_j\text{-th edge in Algorithm 2 does not connect to } \bar{\mathscr{C}}_1^n(\gamma_n)\}.$$

Both T_j and $\mathfrak{A}_j$ depend on v, but we will suppress this dependence to simplify notation.

Lemma 6.29 *For all large n, on the event $\mathfrak{B}_n$, $\mathbb{P}_2(\mathfrak{A}_{g(n)}) \leq 1/n^3$ for any $v \notin \bar{\mathscr{C}}_1^n(\gamma_n)$, where $g(n) = n^{\eta/2}(\log n)^{7/12}$.*

Proof The proof recursively analyzes $\mathfrak{A}_j$, $j \geq 1$. We will discuss how to analyze $\mathfrak{A}_1$ and $\mathfrak{A}_2$ in detail; the argument for a general j is similar. Let S_1 be the sequence of the first $(T_1 - 1) \vee 0$ edges added to the cluster of v in Algorithm 2 (arranged in the order they were added). To bound $\mathbb{P}_2(\mathfrak{A}_1)$, we will actually prove an upper bound on $\mathbb{P}_2(\mathfrak{A}_1 \mid S_1 = s_1)$ that is uniform over the choice of s_1 with $\mathbb{P}_2(S_1 = s_1) > 0$ and length$(s_1) = t_n$, where length(s_1) denotes the number of edges in the sequence s_1. Call such a choice of s_1 *tenable*.

Since the status of the edges (presence or absence) connecting to $\bar{\mathscr{C}}_1^n(\gamma_n)$ are not checked in the first t_n steps of Algorithm 2, the probability of these edges being present under $\mathbb{P}_2(\cdot | S_1 = s_1)$ is the same as in (129) for any tenable choice of edges s_1. Let $\mathscr{E}_1 \subseteq E(\bar{G}_n)$ be the set of edges found between the cluster of v and $\bar{\mathscr{C}}_1^n(\gamma_n)$ in the

T_1-th step in Algorithm 2. For any tenable s_1,

$$\mathbb{E}_2\big[|\mathscr{E}_1|\,\big|\,S_1 = s_1\big] = \big(1 + O(n^{-\eta})\big) \cdot \big(1 - p_{\gamma_n}^n\big) \sum_{i \in \bar{\mathscr{C}}_1^n(\gamma_n)} \sum_{\substack{j \text{ in cluster of} \\ v \text{ at time } t_n}} \frac{w_i w_j}{\ell_n}.$$

Now, the weight of the cluster of v after the t_n-th edge has been added in Algorithm 2 is at least $t_n w_n$. Further, for all large n, on the event $\mathfrak{B}_n$, $\mathscr{W}\big(\bar{\mathscr{C}}_1^n(\gamma_n)\big) \geq Cn$ by virtue of (42), and consequently, for any tenable s_1,

$$\mathbb{E}_2\big[|\mathscr{E}_1|\big|S_1 = s_1\big] \geq 2C_{130} t_n w_n. \tag{130}$$

Letting $\mathfrak{E}_1 = \big\{|\mathscr{E}_1| \geq C_{130} t_n w_n\big\}$, using Bennet's inequality [31] and Assumption 2.4 (iv), we see that there exist $n_{131} \geq 1$ and $C_{131} > 0$ depending only on $\mathcal{P}mtr$ such that for all $n \geq n_{131}$, on the event $\mathfrak{B}_n$, for any tenable s_1,

$$\mathbb{P}_2\big(\mathfrak{E}_1^c\,\big|\,S_1 = s_1\big) \leq \exp\big(-C_{131}\big(\log n\big)^{7/6}\big). \tag{131}$$

Hence, for all $n \geq n_{131}$, on the event $\mathfrak{B}_n$, for any tenable s_1,

$$\begin{aligned}
\mathbb{P}_2\big(\mathfrak{A}_1\,\big|\,S_1 = s_1\big) &\leq \mathbb{P}_2\big(\mathfrak{A}_1 \cap \mathfrak{E}_1\,\big|\,S_1 = s_1\big) + \mathbb{P}_2\big(\mathfrak{E}_1^c\,\big|\,S_1 = s_1\big) \\
&\leq \mathbb{P}_2\big(\mathfrak{A}_1\,\big|\,\mathfrak{E}_1 \cap \{S_1 = s_1\}\big) + \exp\big(-C_{131}\big(\log n\big)^{7/6}\big). \tag{132}
\end{aligned}$$

We will bound the first term on the right side of (132) with the help of Lemma 6.27 (a). To this end, note that on $\mathfrak{E}_1 \cap \{S_1 = s_1\}$, in the T_1-th step, we have found $m_0 \geq C_{130} t_n w_n$ many edges connecting the cluster of v to $\bar{\mathscr{C}}_1^n(\gamma_n)$. The weights associated with these edges are i.i.d. Unif$[x_0, 1]$ random variables, where $x_0 = p_{\gamma_n}^n$. Next, on $\mathfrak{B}_n$, the number of vertices in the cluster of v after the addition of the t_n-th edge is at most $C_{123}(2) \cdot (1 + t_n) \log n / w_n$, and the degree of each of these vertices in $\bar{G}_n$ is at most $(\log n)^{3/2}$. Hence, the number of outgoing edges from the cluster of v at this point that do not connect to $\bar{\mathscr{C}}_1^n(\gamma_n)$ is $k \leq C_{123}(2) \cdot (1 + t_n)(\log n)^{5/2}/w_n$. Conditional on the values of the weights $U_{ij}, \{i, j\} \in s_1$, the weights of these k edges are independent Unif $[x_j, 1]$ random variables for some $x_j \geq x_0$, $j = 1, \ldots, k$. Thus, using Lemma 6.27 (a), we get, on the event $\mathfrak{B}_n$, for any tenable s_1,

$$\mathbb{P}_2\big(\mathfrak{A}_1\,\big|\,\mathfrak{E}_1 \cap \{S_1 = s_1\}\big) \leq 1 - \frac{C_{130} t_n w_n}{C_{130} t_n w_n + C_{123}(2) \cdot (1 + t_n)(\log n)^{5/2}/w_n}. \tag{133}$$

Combining (132) and (133), we see that there exists $n_{134} \geq n_{131}$ such that on the event $\mathfrak{B}_n$,

$$\mathbb{P}_2\big(\mathfrak{A}_1\big) \leq \max_{\text{s}_1 \text{ tenable}} \mathbb{P}_2\big(\mathfrak{A}_1 \mid S_1 = \text{s}_1\big)$$

$$\leq 1 - \left(\frac{C_{130}\, t_n w_n}{C_{130}\, t_n w_n + C_{123}(2) \cdot (1 + t_n)(\log n)^{5/2}/w_n} \right)$$

$$+ \exp\big(- C_{131}\big(\log n\big)^{7/6}\big) \leq 1 - \frac{(\log n)^{1/2}}{C_{134} n^{\eta/2}}, \qquad (134)$$

where the last step uses Assumption 2.4 (iv), and holds for $n \geq n_{134}$.

Now let us move to $\mathfrak{A}_2$. We first note that $\mathbb{P}_2\big(\mathfrak{A}_2\big) = \mathbb{P}_2\big(\mathfrak{A}_2 \cap \mathfrak{A}_1\big)$. Thus,

$$\mathbb{P}_2\big(\mathfrak{A}_2\big) \leq \mathbb{P}_2\big(\mathfrak{A}_2 \mid \mathfrak{A}_1 \cap \mathfrak{E}_1\big) \cdot \mathbb{P}_2\big(\mathfrak{A}_1\big) + \mathbb{P}_2\big(\mathfrak{E}_1^c\big). \qquad (135)$$

Let $S_2^{(1)}$ be the sequence of the first $(T_2 - 1) \vee 0$ edges added in Algorithm 2 (arranged in the order they were added), and let $S_2^{(2)}$ be the collection of edges found between $\bar{\mathscr{C}}_1^n(\gamma_n)$ and the cluster of v in the T_1-th step in Algorithm 2. Write $S_2 = \big(S_2^{(1)}, S_2^{(2)}\big)$. Call $\text{s}_2 = \big(\text{s}_2^{(1)}, \text{s}_2^{(2)}\big)$ tenable if $\text{length}(\text{s}_2^{(1)}) = 2t_n + 1$, $|\text{s}_2^{(2)}| \geq C_{130}\, t_n w_n$, and $\mathbb{P}\big(S_2 = \text{s}_2\big) > 0$. To bound $\mathbb{P}_2\big(\mathfrak{A}_2 \mid \mathfrak{A}_1 \cap \mathfrak{E}_1\big)$ appearing on the right side of (135), it is enough to prove a bound on $\mathbb{P}_2\big(\mathfrak{A}_2 \mid S_2 = \text{s}_2\big)$ that is uniform over all tenable s_2.

Let $\mathscr{E}_2$ be the collection of **new** edges found between $\bar{\mathscr{C}}_1^n(\gamma_n)$ and the cluster of v in the T_2-th step; we emphasize that these edges were not present in $\mathscr{E}_1$. Let $\mathfrak{E}_2 = \big\{|\mathscr{E}_2| \geq C_{130}\, t_n w_n\big\}$. Then, for all $n \geq n_{131}$, on the event $\mathfrak{B}_n$, for any tenable s_2

$$\mathbb{P}_2\big(\mathfrak{A}_2 \mid S_2 = \text{s}_2\big) \leq \mathbb{P}_2\big(\mathfrak{A}_2 \cap \mathfrak{E}_2 \mid S_2 = \text{s}_2\big) + \mathbb{P}_2\big(\mathfrak{E}_2^c \mid S_2 = \text{s}_2\big)$$

$$\leq \mathbb{P}_2\big(\mathfrak{A}_2 \mid \mathfrak{E}_2 \cap \{S_2 = \text{s}_2\}\big) + \exp\big(- C_{131}\big(\log n\big)^{7/6}\big), \quad (136)$$

where the last step follows from an argument similar to the one leading to (131). We will bound the first term on the right side of (136) with the help of Lemma 6.27 (b). To this end, note that on $\mathfrak{E}_2 \cap \{S_2 = \text{s}_2\}$, in the T_2-th step, we have found $\bar{m}_0 \geq C_{130}\, t_n w_n$ many new edges connecting the cluster of v to $\bar{\mathscr{C}}_1^n(\gamma_n)$. The weights associated with these edges are i.i.d. Unif $[\bar{x}_0, 1]$ random variables, where $\bar{x}_0 = p_{\gamma_n}^n$. By an argument similar to the one used while analyzing $\mathfrak{A}_1$, on the event $\mathfrak{B}_n$, the number of outgoing edges from the cluster of v that do not connect to $\bar{\mathscr{C}}_1^n(\gamma_n)$ and were found after the T_1-th step is $\bar{k} \leq C_{123}(2) \cdot (1 + t_n)(\log n)^{5/2}/w_n$. Conditional on the values of the weights $U_{ij}, \{i, j\} \in \text{s}_2^{(1)}$, the weights of these $\bar{k}$ edges are independent Unif $[\bar{x}_l, 1]$ random variables for some $\bar{x}_l \geq \bar{x}_0$, $l = 1, \ldots, \bar{k}$. Note also that conditional on the values of the weights $U_{ij}, \{i, j\} \in \text{s}_2^{(1)}$, the weights associated with the $m_0 \geq C_{130}\, t_n w_n$ edges found between $\bar{\mathscr{C}}_1^n(\gamma_n)$ and the cluster of v in the T_1-th step are i.i.d. Unif $[x_0', 1]$ random variables, where x_0' is the weight of the $(t_n + 1)$-th edge in $\text{s}_2^{(2)}$, and the weights associated with the k' (say) outgoing edges from the cluster of v that do not connect to $\bar{\mathscr{C}}_1^n(\gamma_n)$ and were found in the first T_1 steps are independent Unif $[x_j', 1]$ random variables for some $x_j' \geq x_0'$, $j = 1, \ldots, k'$. Thus, using Lemma 6.27 (b), we get, on

the event $\mathfrak{B}_n$, for any tenable s_2,

$$\mathbb{P}_2\big(\mathfrak{A}_2 \,\big|\, \mathfrak{E}_2 \cap \{S_2 = s_2\}\big) \le 1 - \frac{C_{130}\, t_n w_n}{C_{130}\, t_n w_n + C_{123}(2) \cdot (1 + t_n)(\log n)^{5/2}/w_n},$$

which combined with (136) shows that for $n \ge n_{134}$, on the event $\mathfrak{B}_n$,

$$\mathbb{P}_2\big(\mathfrak{A}_2 \,\big|\, \mathfrak{A}_1 \cap \mathfrak{E}_1\big) \le \max_{s_2 \text{ tenable}} \mathbb{P}_2\big(\mathfrak{A}_2 \,\big|\, S_2 = s_2\big) \le 1 - \frac{(\log n)^{1/2}}{C_{134} n^{\eta/2}}. \qquad (137)$$

Now going back to (135) and using (131), (134), and (137), we see that for $n \ge n_{134}$, on the event $\mathfrak{B}_n$,

$$\mathbb{P}_2\big(\mathfrak{A}_2\big) \le \left(1 - \frac{(\log n)^{1/2}}{C_{134} n^{\eta/2}}\right)^2 + \exp\big(-C_{131}\big(\log n\big)^{7/6}\big).$$

Turning to the events $\mathfrak{A}_j$ for $3 \le j \le g(n)$, we define $\mathfrak{E}_j$ in a manner analogous to $\mathfrak{E}_1$ and $\mathfrak{E}_2$. Proceeding similarly, we can show that for $n \ge n_{134}$ and $2 \le j \le g(n)$, on the event $\mathfrak{B}_n$,

$$\begin{aligned}
\mathbb{P}_2\big(\mathfrak{A}_{j+1}\big) = \mathbb{P}_2\bigg(\bigcap_{i=1}^{j+1} \mathfrak{A}_i\bigg) &\le \mathbb{P}_2\bigg(\mathfrak{A}_{j+1} \,\bigg|\, \bigcap_{i=1}^{j} (\mathfrak{A}_i \cap \mathfrak{E}_i)\bigg) \cdot \mathbb{P}_2\big(\mathfrak{A}_j\big) + \sum_{i=1}^{j} \mathbb{P}_2\big(\mathfrak{E}_i^c\big) \\
&\le \left(1 - \frac{(\log n)^{1/2}}{C_{134} n^{\eta/2}}\right) \cdot \mathbb{P}_2\big(\mathfrak{A}_j\big) + j \exp\big(-C_{131}\big(\log n\big)^{7/6}\big) \\
&\le \left(1 - \frac{(\log n)^{1/2}}{C_{134} n^{\eta/2}}\right)^{j+1} + \frac{j(j+1)}{2} \exp\big(-C_{131}\big(\log n\big)^{7/6}\big).
\end{aligned}$$

For the third step, we need to use the analogue of Lemma 6.27 for $(j+1)$ collections of independent uniform random variables. However, this generalization is straightforward. This completes the proof of Lemma 6.29. □

Completing the Proof of Proposition 6.5: Combining (128) with Lemma 6.29 and using a union bound over $v \notin \overline{\mathscr{C}}_1^n(\gamma_n)$, we get, for all large n,

$$\mathbb{P}\bigg(d_{\mathrm{H}}\big(\bar{M}_{\gamma_n}^n,\, \bar{M}^n\big) \ge \big(\log n\big)^{-1/6} n^\eta\bigg) \le 3/n^2.$$

This completes the proof. □

7 Proof of Theorem 3.1

We will complete the proof in five steps. As observed in Sect. 5, it is enough prove the result for M^n and $\bar{M}^n$.

7.1 Existence of the scaling limit

Our aim in this section is to show that there exists a random compact $\mathbb{R}$-tree $\mathscr{M}^{\boldsymbol{\theta}^*}$ whose law depends only on $\boldsymbol{\theta}^*$ such that (11) is satisfied if we replace $\boldsymbol{M}^n$ by M^n or $\bar{M}^n$.

Note that (25), the first relation in (24), Assumption 2.4 (iii), and (7) imply that $\boldsymbol{x}^{(n)}$, $n \geq 1$, satisfies [29, Assumption 1.6] with the limiting sequence $\boldsymbol{\theta}^*$. Now consider the random graph $\mathscr{G}_n(\lambda)$ as in (27). Then [29, Theorem 1.8] shows that the largest components (arranged according to their masses) of $\mathscr{G}_n(\lambda)$, endowed with the graph distance scaled by $n^{-\eta}$ and the probability measure that assigns mass proportional to x_v to each vertex v in a component, converges in distribution to a sequence of random metric measure spaces with respect to the product topology induced by the Gromov-weak topology on each coordinate. We claim that in a similar manner, we can show that for every $\lambda \in \mathbb{R}$ there exists a random metric measure space $\mathcal{S}_\lambda^{\boldsymbol{\theta}^*}$ whose law depends only on λ and $\boldsymbol{\theta}^*$ such that as $n \to \infty$,

$$n^{-\eta} \cdot \mathscr{C}\big(1, \mathscr{G}_n(\lambda)\big) \xrightarrow{\mathrm{d}} \mathcal{S}_\lambda^{\boldsymbol{\theta}^*} \quad \text{w.r.t. the Gromov-weak topology}, \tag{138}$$

where the measure on $n^{-\eta} \cdot \mathscr{C}\big(1, \mathscr{G}_n(\lambda)\big)$ assigns mass $x_v \big[\mathscr{W}_x\big(\mathscr{C}\big(1, \mathscr{G}_n(\lambda)\big)\big)\big]^{-1}$ to each vertex v in $\mathscr{C}\big(1, \mathscr{G}_n(\lambda)\big)$. We briefly explain how this can be done. The proof of [29, Theorem 1.8], applied to the special case $\mathscr{G}_n(\lambda)$, can be divided into two steps: (i) In [29, Theorem 4.5], it is proved that under some assumptions, a related connected random graph has a scaling limit in the Gromov-weak topology. (ii) Then the arguments of [29, Section 5.1] show that the maximal connected components of $\mathscr{G}_n(\lambda)$ satisfy the assumptions in [29, Theorem 4.5]. A key result needed to complete step (ii) is [11, Proposition 9] (restated in [29, Proposition 5.2]), which shows that a breadth-first walk of $\mathscr{G}_n(\lambda)$ defined in [11] converges in distribution w.r.t. the Skorohod J_1 topology to a limiting process. In a similar way (see also the proof of [23, Theorem 2.4]), one can prove that for any $\lambda \geq 0$, the breadth-first walk of $\mathscr{G}_n(\lambda)$ started from the vertex 1 as given by (30) satisfies

$$n^\eta \cdot Z_\lambda^{n,(1)}(u) \xrightarrow{\mathrm{d}} Z_\lambda^{(1)}(u) := \theta_1^* + \lambda u + \sum_{j=2}^{\infty} \theta_j^* \big(\mathbb{1}_{\{\xi_j \leq u\}} - u\theta_j^*\big), \quad u \geq 0 \tag{139}$$

with respect to the Skorohod J_1 topology on $D[0, \infty)$, where $\xi_j \sim \mathrm{Exp}(\theta_j^*)$, $j \geq 1$, are independent random variables. Now using (139), the rest of the arguments in [29, Section 5.1] repeated verbatim would show that the vertex weights in $\mathscr{C}\big(1, \mathscr{G}_n(\lambda)\big)$ satisfy the assumptions of [29, Theorem 4.5], which would then yield (138). For concreteness, we describe the construction of $\mathcal{S}_\lambda^{\boldsymbol{\theta}^*}$.

Construction 7.1 *Define* $\mathscr{Z}_\lambda := \inf \big\{u \geq 0 : Z_\lambda^{(1)}(u) \leq 0\big\}$, *and* $\mathscr{B}_\lambda := \{1\} \cup \big\{j \geq 2 : \xi_j \leq \mathscr{Z}_\lambda\big\}$. *Let*

$$\bar{\gamma} := \mathscr{Z}_\lambda \cdot \bigg(\sum_{j \in \mathscr{B}_\lambda} \big(\theta_j^*\big)^2\bigg)^{1/2}, \quad \text{and} \quad \bar{\boldsymbol{\theta}} := \bigg(\theta_j^* \cdot \bigg(\sum_{i \in \mathscr{B}_\lambda} \big(\theta_i^*\big)^2\bigg)^{-1/2}; \ j \in \mathscr{B}_\lambda\bigg).$$

Set

$$S_\lambda^{\theta^*} = \mathscr{Z}_\lambda \cdot \left(\sum_{i \in \mathscr{B}_\lambda} (\theta_i^*)^2 \right)^{-1/2} \cdot \mathscr{G}_\infty(\bar{\boldsymbol{\theta}}, \bar{\gamma}),$$

where $\mathscr{G}_\infty(\cdot, \cdot)$ is as given in [29, Definition 2.2].

Next, we claim that under Assumption 2.4, $S_\lambda^{\theta^*}$ is compact almost surely, and further, as $n \to \infty$,

$$n^{-\eta} \cdot \mathscr{C}\big(1, \mathscr{G}_n(\lambda)\big) \xrightarrow{\mathrm{d}} S_\lambda^{\theta^*} \quad \text{w.r.t. the GHP topology.} \tag{140}$$

To deduce (140) from (138), we need to prove that $n^{-\eta} \cdot \mathscr{C}\big(1, \mathscr{G}_n(\lambda)\big)$ satisfies a global lower mass-bound; see [18, Theorem 6.1] and [29, Lemma 6.1]. The global lower mass-bound was established in [29] under the stronger [29, Assumption 1.1]. This stronger assumption was needed to prove [29, Lemma 6.7], which relied on a tail bound on the heights of branching processes established in [61, Theorem 2]. The result in [61, Theorem 2] was proved for a (conditioned) branching process with a given offspring distribution, whereas in the random graph setting, one needs to consider branching processes with varying offspring distributions for different n; e.g., the offspring distributions of interest in this paper are Poi(V_n), $n \geq 1$, where V_n is as defined around (64). To get around this difficulty, [29, Assumption 1.1] was used in the proof of [29, Lemma 6.6] to stochastically upper bound the offspring distributions for different n by a single offspring distribution to which [61, Theorem 2] is applicable. However, instead of [29, Lemma 6.7], we can now appeal to Proposition 6.13 (which relies on the techniques developed in the more recent work [4]), and avoid the use of the stronger [29, Assumption 1.1]. Then the rest of the argument from [29] carries over in an identical way under Assumption 2.4 to establish the desired global lower mass-bound for $n^{-\eta} \cdot \mathscr{C}\big(1, \mathscr{G}_n(\lambda)\big)$, which then yields (140). Let us also note that the construction of $S_\lambda^{\theta^*}$ given above together with the almost sure compactness of $S_\lambda^{\theta^*}$ shows that $S_\lambda^{\theta^*}$ is an $\mathbb{R}$-graph almost surely. Further, (27) and (140) imply

$$n^{-\eta} \cdot \mathscr{C}_1^n(\lambda) \xrightarrow{\mathrm{d}} S_\lambda^{\theta^*} \quad \text{w.r.t. the GH topology.} \tag{141}$$

The next result shows that Theorem 4.7 can be applied to the sequence $n^{-\eta} \cdot \mathscr{C}_1^n(\lambda)$, $n \geq 1$. Let $\mathscr{A}_r$ be as defined around (15).

Lemma 7.2 *Fix $\lambda \geq 0$. Then we can construct $S_\lambda^{\theta^*}$, $n^{-\eta} \cdot \mathscr{C}_1^n(\lambda)$, $n \geq 1$, and a positive random variable R on the same probability space such that $n^{-\eta} \cdot \mathscr{C}_1^n(\lambda) \xrightarrow{\mathrm{a.s.}} S_\lambda^{\theta^*}$ w.r.t. the GH topology as $n \to \infty$, and $\mathbb{P}\big(n^{-\eta} \cdot \mathscr{C}_1^n(\lambda) \in \mathscr{A}_R \ \text{eventually}\big) = 1$.*

Remark 3 We will omit the proof of Lemma 7.2, as this result is essentially contained in [29]. We only make a brief comment on how this result follows. Recall the notation $\big(k(X), e(X)\big)$ for an $\mathbb{R}$-graph X from Sect. 4.3 and the probability distribution $\mathbb{P}_{\mathrm{con}}$ given by (94). For $m \geq 1$, $a > 0$, and a probability mass function $\mathbf{p} = (p_1, \ldots, p_m)$

on $[m]$ with $p_1 \geq p_2 \geq \cdots \geq p_m > 0$, write $\tilde{\mathscr{G}}_m(\mathbf{p}, a)$ for a random graph with distribution $\mathbb{P}_{\mathrm{con}}(\cdot, \mathbf{p}, a, [m])$. Let $\mathscr{G}_\infty(\cdot, \cdot)$ be as given in [29, Definition 2.2]. Now, as mentioned at the beginning of this section, the key ingredient in the proof of (138) is [29, Theorem 4.5], which shows that under [29, Assumption 4.4],

$$\|\mathbf{p}\|_2 \cdot \tilde{\mathscr{G}}_m(\mathbf{p}, a) \xrightarrow{\mathrm{d}} \mathscr{G}_\infty(\boldsymbol{\theta}, \gamma) \tag{142}$$

with respect to Gromov-weak topology as $m \to \infty$. It also follows from the proof of [29, Lemma 4.12] and [29, (4.23)] that under [29, Assumption 4.4],

$$\mathrm{sp}\left(\tilde{\mathscr{G}}_m(\mathbf{p}, a)\right) \xrightarrow{\mathrm{d}} \mathrm{sp}\left(\mathscr{G}_\infty(\boldsymbol{\theta}, \gamma)\right) \tag{143}$$

jointly with the convergence in (142). Now, from the proof of [29, Theorem 4.5], it can be easily deduced that

$$\|\mathbf{p}\|_2 \cdot \left(\mathrm{len}\left(\mathrm{Core}\left(\tilde{\mathscr{G}}_m(\mathbf{p}, a) \right) \right), \min_{e \in e(\tilde{\mathscr{G}}_m(\mathbf{p}, a))} \mathrm{len}(e) \right)$$
$$\xrightarrow{\mathrm{d}} \left(\mathrm{len}\left(\mathrm{Core}\left(\mathscr{G}_\infty(\boldsymbol{\theta}, \gamma) \right) \right), \min_{e \in e(\mathscr{G}_\infty(\boldsymbol{\theta}, \gamma))} \mathrm{len}(e) \right)$$

jointly with the convergences in (142) and (143). These results directly translate into the corresponding convergences for $\mathscr{C}_1^n(\lambda)$, which then yield the claim in Lemma 7.2.

We will now deduce the existence of the MST scaling limit from Lemma 7.2 and Theorem 6.1. As mentioned in Sect. 6.1, the argument for combining these two results to get the GH scaling limit of $\bar{M}^n$ is similar to the one used in [6]. Using Lemma 7.2 and Theorem 4.7, we get

$$n^{-\eta} \cdot \mathrm{CBD}^\infty\left(\mathscr{C}_1^n(\lambda) \right) \xrightarrow{\mathrm{d}} \mathrm{CB}^\infty\left(\mathcal{S}_\lambda^{\boldsymbol{\theta}^*} \right) =: \mathscr{M}_\lambda^{\boldsymbol{\theta}^*} \quad \text{w.r.t. the GH topology} \tag{144}$$

as $n \to \infty$. Similar to Lemma 5.1, for any $\lambda \geq 0$, there exist couplings of $G_n(\lambda)$ and $\bar{G}_n(\lambda)$ such that $\mathbb{P}\left(G_n(\lambda) \neq \bar{G}_n(\lambda) \right) \to 0$ as $n \to \infty$. This fact, together with (144) and Lemma 4.6, gives, for any $\lambda \geq 0$,

$$n^{-\eta} \cdot \bar{M}_\lambda^n \xrightarrow{\mathrm{d}} \mathscr{M}_\lambda^{\boldsymbol{\theta}^*} \quad \text{w.r.t. the GH topology} \tag{145}$$

as $n \to \infty$. As discussed below Definition 6.2, $\bar{M}_\lambda^n$ is a subtree of $\bar{M}_{\lambda'}^n$ whenever $0 \leq \lambda \leq \lambda'$. For $0 \leq \lambda \leq \lambda'$, consider the sequence of marked spaces $\left(n^{-\eta} \cdot \bar{M}_{\lambda'}^n, n^{-\eta} \cdot \bar{M}_\lambda^n \right)$, $n \geq 1$. Since $\mathscr{M}_{\lambda'}^{\boldsymbol{\theta}^*}$ is compact almost surely, [64, Proposition 9] implies that this sequence is tight in the marked GH topology. Thus, by passing to a subsequence, we can assume that

$$\left(n^{-\eta} \cdot \bar{M}_{\lambda'}^n, n^{-\eta} \cdot \bar{M}_\lambda^n \right) \xrightarrow{\mathrm{d}} \left(\mathscr{M}_{\lambda'}^{\boldsymbol{\theta}^*}, \mathscr{M}_\lambda^{\boldsymbol{\theta}^*} \right) \tag{146}$$

with respect to the marked GH topology. On the right side of (146), we have a coupling of $\mathscr{M}_{\lambda'}^{\boldsymbol{\theta}^*}$ and $\mathscr{M}_{\lambda}^{\boldsymbol{\theta}^*}$ in which the latter is a subspace of the former. Since $d_{\mathrm{H}}\bigl(\bar{M}_{\lambda}^n, \bar{M}_{\lambda'}^n\bigr) \le d_{\mathrm{H}}\bigl(\bar{M}_{\lambda}^n, \bar{M}^n\bigr)$, an application of Theorem 6.1 shows that for all $\Delta \in (0, 1/2]$, in the coupling of (146),

$$\mathbb{P}\bigl(d_{\mathrm{H}}\bigl(\mathscr{M}_{\lambda}^{\boldsymbol{\theta}^*}, \mathscr{M}_{\lambda'}^{\boldsymbol{\theta}^*}\bigr) > \lambda^{-1+\Delta}\bigr) \le C_{21}(\Delta) \cdot \lambda^{-1/2} \tag{147}$$

for $\lambda' \ge \lambda \ge \lambda_{21}(\Delta)$. This implies that for $\lambda' \ge \lambda \ge \lambda_{21}(\Delta)$,

$$d_{\mathrm{Pr}}\left(\mathcal{L}aw\bigl(\mathscr{M}_{\lambda}^{\boldsymbol{\theta}^*}\bigr), \ \mathcal{L}aw\bigl(\mathscr{M}_{\lambda'}^{\boldsymbol{\theta}^*}\bigr)\right) \le \lambda^{-1+\Delta} + C_{21}(\Delta) \cdot \lambda^{-1/2}, \tag{148}$$

where $\mathcal{L}aw(\cdot)$ denotes the law of a random object, and $d_{\mathrm{Pr}}(\cdot, \cdot)$ denotes the Prokhorov distance. Thus, $\bigl(\mathcal{L}aw(\mathscr{M}_{\lambda}^{\boldsymbol{\theta}^*}), \ \lambda \ge 0\bigr)$ is a Cauchy sequence in the space of probability measures on $\mathfrak{S}_{\mathrm{GH}}$. Since $\mathfrak{S}_{\mathrm{GH}}$ is Polish, the space of probability measures on $\mathfrak{S}_{\mathrm{GH}}$ is complete under $d_{\mathrm{Pr}}(\cdot, \cdot)$. Hence, there exists a random compact metric space $\mathscr{M}^{\boldsymbol{\theta}^*}$ whose law depends only on $\boldsymbol{\theta}^*$ such that $\mathscr{M}_{\lambda}^{\boldsymbol{\theta}^*} \xrightarrow{\mathrm{d}} \mathscr{M}^{\boldsymbol{\theta}^*}$ w.r.t. the GH topology as $\lambda \to \infty$. Since $\mathscr{M}_{\lambda}^{\boldsymbol{\theta}^*}$ is an $\mathbb{R}$-tree for any λ, [48, Lemma 2.1] shows that $\mathscr{M}^{\boldsymbol{\theta}^*}$ is an $\mathbb{R}$-tree almost surely. From (148), we further get

$$d_{\mathrm{Pr}}\left(\mathcal{L}aw\bigl(\mathscr{M}_{\lambda}^{\boldsymbol{\theta}^*}\bigr), \ \mathcal{L}aw\bigl(\mathscr{M}^{\boldsymbol{\theta}^*}\bigr)\right) \le \lambda^{-1+\Delta} + C_{21}(\Delta) \cdot \lambda^{-1/2} \tag{149}$$

for $\lambda \ge \lambda_{21}(\Delta)$. Now, Theorem 6.1 and (149) coupled with an application of the triangle inequality yield

$$\limsup_{n \to \infty} d_{\mathrm{Pr}}\left(\mathcal{L}aw\bigl(n^{-\eta} \cdot \bar{M}^n\bigr), \ \mathcal{L}aw\bigl(\mathscr{M}^{\boldsymbol{\theta}^*}\bigr)\right)$$
$$\le 2\bigl(\lambda^{-1+\Delta} + C_{21}(\Delta) \cdot \lambda^{-1/2}\bigr) + \limsup_{n \to \infty} d_{\mathrm{Pr}}\left(\mathcal{L}aw\bigl(n^{-\eta} \cdot \bar{M}_{\lambda}^n\bigr), \ \mathcal{L}aw\bigl(\mathscr{M}_{\lambda}^{\boldsymbol{\theta}^*}\bigr)\right)$$

for any $\lambda \ge \lambda_{21}(\Delta)$. Now using (145) and letting $\lambda \to \infty$ gives

$$n^{-\eta} \cdot \bar{M}^n \xrightarrow{\mathrm{d}} \mathscr{M}^{\boldsymbol{\theta}^*} \quad \text{w.r.t. the GH topology} \tag{150}$$

as $n \to \infty$. Finally, Lemma 5.1 shows that (150) continues to hold if we replace $\bar{M}^n$ by M^n.

7.2 The degrees of points in $\mathscr{M}^{\boldsymbol{\theta}^*}$

In this section we will prove Theorem 3.1 (a). Recall the definition of degree from Sect. 4.3 and the notation $\mathscr{L}(\cdot)$ and $\mathscr{H}(\cdot)$ from (10). From Construction 7.1 and the construction of the space $\mathscr{G}_{\infty}(\cdot, \cdot)$ using an inhomogeneous continuum random tree (ICRT) given in [29, Section 2.3.1], it follows that the set of points in $\mathcal{S}_{\lambda}^{\boldsymbol{\theta}^*}$ with infinite

degree is countably infinite, and all other points in $\mathcal{S}_\lambda^{\boldsymbol{\theta}^*}$ either have degree 1 or 2. Since $\mathcal{M}_\lambda^{\boldsymbol{\theta}^*} = \mathrm{CB}^\infty(\mathcal{S}_\lambda^{\boldsymbol{\theta}^*})$, the same is true of $\mathcal{M}_\lambda^{\boldsymbol{\theta}^*}$, i.e., the set $\mathcal{H}(\mathcal{M}_\lambda^{\boldsymbol{\theta}^*})$ is countably infinite and $\deg(x\,;\,\mathcal{M}_\lambda^{\boldsymbol{\theta}^*}) \in \{1, 2\}$ for all $x \in \mathcal{M}_\lambda^{\boldsymbol{\theta}^*} \setminus \mathcal{H}(\mathcal{M}_\lambda^{\boldsymbol{\theta}^*})$.

By an argument similar to the one leading to (146), we can assume that for any $\lambda \geq 1$,

$$\left(n^{-\eta} \cdot \bar{M}^n \,,\, n^{-\eta} \cdot \bar{M}_\lambda^n\right) \overset{\mathrm{d}}{\longrightarrow} \left(\mathcal{M}^{\boldsymbol{\theta}^*} ,\, \mathcal{M}_\lambda^{\boldsymbol{\theta}^*}\right) \tag{151}$$

with respect to the marked GH topology along a suitable subsequence (which may depend on λ). In the coupling between $\mathcal{M}_\lambda^{\boldsymbol{\theta}^*}$ and $\mathcal{M}^{\boldsymbol{\theta}^*}$ on the right side of (151), the former is a closed subset of the latter. Hence, $\mathcal{M}^{\boldsymbol{\theta}^*}$ has infinitely many points of infinite degree almost surely.

Now, Theorem 6.1 and (151) imply that for all $\lambda \geq 1$,

$$\mathbb{P}\left(d_{\mathrm{H}}\left(\mathcal{M}_\lambda^{\boldsymbol{\theta}^*},\, \mathcal{M}^{\boldsymbol{\theta}^*}\right) > \lambda^{-1/2}\right) \leq C_{152}\lambda^{-1/2}. \tag{152}$$

in the coupling of (151). In this coupling, on the event

$$\mathfrak{D}_\lambda := \left\{d_{\mathrm{H}}\left(\mathcal{M}_\lambda^{\boldsymbol{\theta}^*},\, \mathcal{M}^{\boldsymbol{\theta}^*}\right) \leq \lambda^{-1/2}\right\}, \tag{153}$$

$\mathcal{M}^{\boldsymbol{\theta}^*}$ can be obtained by attaching countably many $\mathbb{R}$-trees each having diameter at most $2\lambda^{-1/2}$ to $\mathcal{M}_\lambda^{\boldsymbol{\theta}^*}$. Hence, on $\mathfrak{D}_\lambda$, any $x \in \mathcal{M}^{\boldsymbol{\theta}^*}$ that satisfies

(A) $3 \leq \deg(x\,;\,\mathcal{M}^{\boldsymbol{\theta}^*}) < \infty$, and

(B$_\lambda$) all $\deg(x\,;\,\mathcal{M}^{\boldsymbol{\theta}^*})$ of the components in $\mathcal{M}^{\boldsymbol{\theta}^*} \setminus \{x\}$ have diameter $> 2\lambda^{-1/2}$

must also satisfy $x \in \mathcal{M}_\lambda^{\boldsymbol{\theta}^*}$, and each of the $\deg(x\,;\,\mathcal{M}^{\boldsymbol{\theta}^*})$ components in $\mathcal{M}^{\boldsymbol{\theta}^*} \setminus \{x\}$ must have a non-empty intersection with $\mathcal{M}_\lambda^{\boldsymbol{\theta}^*}$. But this implies that $\deg(x\,;\,\mathcal{M}_\lambda^{\boldsymbol{\theta}^*}) \geq 3$, which in turn implies that $\deg(x\,;\,\mathcal{M}_\lambda^{\boldsymbol{\theta}^*}) = \infty$, and consequently, $\deg(x\,;\,\mathcal{M}^{\boldsymbol{\theta}^*}) = \infty$ – a contradiction. Hence, on the event $\mathfrak{D}_\lambda$, there does not exist $x \in \mathcal{M}^{\boldsymbol{\theta}^*}$ satisfying (A) and (B$_\lambda$) as above. This observation combined with (152) shows that for all $\lambda \geq 1$,

$$\mathbb{P}\left(\mathcal{Z}^{\boldsymbol{\theta}^*} > 2\lambda^{-1/2}\right) \leq C_{152}\lambda^{-1/2}, \tag{154}$$

where

$$\mathcal{Z}^{\boldsymbol{\theta}^*} := \sup\left\{\min\left\{\operatorname{diam}(\mathcal{T}) \,:\, \mathcal{T} \text{ component of } \mathcal{M}^{\boldsymbol{\theta}^*} \setminus \{x\}\right\} \,:\right.$$
$$\left. x \in \mathcal{M}^{\boldsymbol{\theta}^*},\, 3 \leq \deg(x;\,\mathcal{M}^{\boldsymbol{\theta}^*}) < \infty\right\}.$$

Here supremum of an empty set is defined as zero. Then $\mathcal{Z}^{\boldsymbol{\theta}^*} = 0$ if and only if there does not exist $x \in \mathcal{M}^{\boldsymbol{\theta}^*}$ with $3 \leq \deg(x;\,\mathcal{M}^{\boldsymbol{\theta}^*}) < \infty$. Now, letting $\lambda \to \infty$ in (154) shows that $\mathcal{Z}^{\boldsymbol{\theta}^*} = 0$ almost surely, which concludes the proof of Theorem 3.1 (a).

7.3 Leaves and hubs of $\mathcal{M}^{\boldsymbol{\theta}^*}$

In this section we will prove Theorem 3.1 (b). Consider $\bar{\boldsymbol{\theta}}$ as in Construction 7.1. Then in the ICRT corresponding to $\bar{\boldsymbol{\theta}}$, both the set of leaves and the set of hubs are dense almost surely. (We refer the reader to [12, 37] for background on ICRTs.) From the construction of the space $\mathscr{G}_\infty(\cdot,\ \cdot)$ given in [29, Section 2.3.1], the same is true for $\mathscr{G}_\infty(\bar{\boldsymbol{\theta}}, \bar{\gamma})$, and consequently for $\mathcal{S}_\lambda^{\boldsymbol{\theta}^*}$. Since $\mathcal{M}_\lambda^{\boldsymbol{\theta}^*} = \mathrm{CB}^\infty\big(\mathcal{S}_\lambda^{\boldsymbol{\theta}^*}\big)$, both $\mathscr{L}\big(\mathcal{M}_\lambda^{\boldsymbol{\theta}^*}\big)$ and $\mathscr{H}\big(\mathcal{M}_\lambda^{\boldsymbol{\theta}^*}\big)$ are dense in $\mathcal{M}_\lambda^{\boldsymbol{\theta}^*}$.

Fix $k \geq 1$, and consider $\lambda \geq 1$ large so that $3\lambda^{-1/2} < 1/k$. Consider the coupling between $\mathcal{M}_\lambda^{\boldsymbol{\theta}^*}$ and $\mathcal{M}^{\boldsymbol{\theta}^*}$ as in (151). On the event $\mathfrak{D}_\lambda$ defined in (153), any point in $\mathcal{M}^{\boldsymbol{\theta}^*}$ is within distance $\lambda^{-1/2}$ from a point in $\mathcal{M}_\lambda^{\boldsymbol{\theta}^*}$, and consequently within distance $2\lambda^{-1/2}$ from a point in $\mathscr{H}\big(\mathcal{M}_\lambda^{\boldsymbol{\theta}^*}\big)$. Since $\mathscr{H}\big(\mathcal{M}_\lambda^{\boldsymbol{\theta}^*}\big) \subseteq \mathscr{H}\big(\mathcal{M}^{\boldsymbol{\theta}^*}\big)$, using (152), we see that

$$\mathbb{P}\bigg(\mathcal{M}^{\boldsymbol{\theta}^*} = \bigcup_{x \in \mathscr{H}(\mathcal{M}^{\boldsymbol{\theta}^*})} B\big(x, k^{-1}\,;\, \mathcal{M}^{\boldsymbol{\theta}^*}\big)\bigg) \geq 1 - C_{152}\lambda^{-1/2}\,.$$

Letting $\lambda \to \infty$, we get

$$\mathbb{P}\bigg(\forall k \geq 1\,,\ \ \mathcal{M}^{\boldsymbol{\theta}^*} = \bigcup_{x \in \mathscr{H}(\mathcal{M}^{\boldsymbol{\theta}^*})} B\big(x, k^{-1}\,;\, \mathcal{M}^{\boldsymbol{\theta}^*}\big)\bigg) = 1\,,$$

which shows that $\mathscr{H}(\mathcal{M}^{\boldsymbol{\theta}^*})$ is dense in $\mathcal{M}^{\boldsymbol{\theta}^*}$ almost surely.

We now turn to the leaves of $\mathcal{M}^{\boldsymbol{\theta}^*}$. Once again, fix $k \geq 1$, and consider $\lambda \geq 1$ large so that $3\lambda^{-1/2} < 1/k$. Arguing as before, on the event $\mathfrak{D}_\lambda$, any $x \in \mathcal{M}^{\boldsymbol{\theta}^*}$ is within distance $2\lambda^{-1/2}$ from a point $x' \in \mathscr{L}\big(\mathcal{M}_\lambda^{\boldsymbol{\theta}^*}\big)$. If $x' \in \mathscr{L}\big(\mathcal{M}^{\boldsymbol{\theta}^*}\big)$, then we are done. Otherwise, $\mathcal{M}^{\boldsymbol{\theta}^*} \setminus \{x'\}$ has a component $\mathscr{T}$ that is a subset of $\mathcal{M}^{\boldsymbol{\theta}^*} \setminus \mathcal{M}_\lambda^{\boldsymbol{\theta}^*}$. Further, $\mathscr{T}$ contains a point $x'' \in \mathscr{L}\big(\mathcal{M}^{\boldsymbol{\theta}^*}\big)$ such that x'' is within distance $3\lambda^{-1/2}$ of x. The rest is routine.

7.4 The upper Minkowski dimension

The aim of this section is to prove

$$\overline{\dim}\big(\mathcal{M}^{\boldsymbol{\theta}^*}\big) \leq 1/\eta \quad \text{almost surely.} \tag{155}$$

We will make use of the following result in our proof.

Proposition 7.3 *There exist $\lambda_{156} \geq \lambda_{41}$ and $\varepsilon_{156} \in (0, \varepsilon_{41})$ such that for all large n and $\lambda \in [\lambda_{156},\ \varepsilon_{156}n^\eta]$,*

$$\mathbb{P}\bigg(\mathrm{sp}\big(\mathscr{C}_1^n(\lambda)\big) \geq C_{156}\lambda^{1/\eta}\bigg) \leq \exp\big(-C\lambda\big)\,. \tag{156}$$

The proof of Proposition 7.3 relies on the following result which will also be useful in Sect. 7.5.

Lemma 7.4 *There exist* $C_{157}, \lambda_{157} \geq 1$ *and* $\varepsilon_{157} > 0$ *such that for all large* n, $\lambda \in [\lambda_{157}, \varepsilon_{157}n^\eta]$ *and* $u \geq C_{157}\lambda^{1/(\tau-3)}$,

$$\mathbb{P}\left(\mathscr{W}_{\boldsymbol{x}}\big(\mathscr{C}\big(1, \mathscr{G}_n(\lambda)\big)\big) \geq u\right) \leq \exp\left(-Cu\right). \tag{157}$$

The proof of Lemma 7.4 will be given at the end of this section.

Proof of Proposition 7.3: In view of (27), it is enough to prove the claimed bound for $\mathrm{sp}\left(\mathscr{C}\big(1, \mathscr{G}_n(\lambda)\big)\right)$. Using (33) and (40), $\bar{\forall}n$ and for $(\lambda, u) \in I^{(n),2}$,

$$\Phi_\lambda^{(n)}(u) \leq \lambda u - Cu^{\tau-2} \leq C'\lambda^{\frac{\tau-2}{\tau-3}},$$

which together with (41) and the strict concavity of $\Phi_\lambda^{(n)}(\cdot)$ implies that $\bar{\forall}n$, and $\lambda \in [\lambda_{41}, \varepsilon_{41}n^\eta]$,

$$\sup_{u \geq 0} \Phi_\lambda^{(n)}(u) = \sup_{0 \leq u \leq s^{(n)}(\lambda)} \Phi_\lambda^{(n)}(u) \leq C_{158}\lambda^{\frac{\tau-2}{\tau-3}}. \tag{158}$$

Recall from (30) the breadth-first walk $Z_\lambda^{n,(1)}(\cdot)$ of $\mathscr{C}(1, \mathscr{G}_n(\lambda))$ started from the vertex 1. Let $\mathrm{Gen}_i^{(1)}$ be the set of vertices in the i-th generation of the breadth-first tree of $\mathscr{C}(1, \mathscr{G}_n(\lambda))$ rooted at the vertex 1. Then on the event $\big\{\mathscr{W}_{\boldsymbol{x}}\big(\mathscr{C}\big(1, \mathscr{G}_n(\lambda)\big)\big) \leq C_{157}\lambda^{1/(\tau-3)}\big\}$, in the coupling of Lemma 6.7,

$$\mathrm{MaxGen}^{(1)} := \max\left\{\sum_{v \in \mathrm{Gen}_j^{(1)}} n^\eta x_v : j \geq 0\right\} \leq \sup_{0 \leq u \leq C_{157}\lambda^{1/(\tau-3)}} n^\eta \cdot Z_\lambda^{n,(1)}(u). \tag{159}$$

Finally, note that (31) and (48) imply, for any $u \geq 0$,

$$n^\eta Z_\lambda^{n,(1)}(u) = n^\eta x_1 + \sum_{j=2}^n \frac{\theta_{j,\lambda}}{(1+\lambda n^{-\eta})}\left(\mathbb{1}\{\xi_j^n \leq u\} - \mathbb{P}\big(\xi_j^n \leq u\big)\right) + \Phi_\lambda^{(n)}(u). \tag{160}$$

Combining (158), (159), (160) with Lemma 7.4, and applying Lemma 6.9 with $s = C_{157}\lambda^{1/(\tau-3)}$ and $y = C_{158}\big(C_{157}\big)^{-\frac{\tau-3}{2}}\lambda$, we see that there exists $\lambda_{161} \geq \lambda_{41} \vee \lambda_{157}$ such that $\bar{\forall}n$ and $\lambda \in [\lambda_{161}, \varepsilon_{156}n^\eta]$,

$$\mathbb{P}\left(\left\{\mathscr{W}_{\boldsymbol{x}}\left(\big(\mathscr{C}(1, \mathscr{G}_n(\lambda))\big)\right) \geq C_{157}\lambda^{\frac{1}{\tau-3}}\right\} \bigcup \left\{\mathrm{MaxGen}^{(1)} \geq 2C_{158}\lambda^{\frac{\tau-2}{\tau-3}}\right\}\right)$$
$$\leq \exp\left(-C\lambda\right), \tag{161}$$

where $\varepsilon_{156} := \varepsilon_{41} \wedge \varepsilon_{157}$. Conditional on the breadth-first tree T (note that T is a plane tree with vertex labels) of $\mathscr{C}(1, \mathscr{G}_n(\lambda))$, the rest of the edges in $\mathscr{C}(1, \mathscr{G}_n(\lambda))$ can be generated by placing edges independently with probability $1 - \exp\left(-(\lambda + n^\eta)x_u x_v\right)$ for every $u, v \in V(T)$, where $u \neq v$, and either (i) $u, v \in \mathrm{Gen}_j^{(1)}$ for some $j \geq 1$, or (ii) $u \in \mathrm{Gen}_{j+1}^{(1)}$ and $v \in \mathrm{Gen}_j^{(1)}$ for some $j \geq 1$, and v lies on the right side of the ancestral line of u. Using (81), we see that conditional on T, on the complement of the event in (161), for any $\lambda \in [\lambda_{161}, \varepsilon_{156} n^\eta]$,

$$\mathrm{sp}\left(\mathscr{C}(1, \mathscr{G}_n(\lambda))\right) \preceq_{\mathsf{sd}} \mathrm{Poi}\left((\varepsilon_{156} + 1) \sum_{u \in V(T)} x_u \cdot 2 \cdot \mathrm{MaxGen}^{(1)}\right) \preceq_{\mathsf{sd}} \mathrm{Poi}\left(C\lambda^{1/\eta}\right).$$

Now the proof of Proposition 7.3 can be completed by combining (161) with standard tail bounds for a Poisson random variable. □

We are now ready to prove the claimed upper bound on $\overline{\dim}(\mathscr{M}^{\theta^*})$. Recall the process MTBP from Definition 6.16. Fix $\Delta \in (0, 1/2]$. Then for any $\lambda \geq 2^\eta$,

$$\mathbb{P}\left(\mathrm{diam}\left(G_n(\lambda) \setminus [\lambda^{\frac{1}{\eta}(1+\Delta)}]\right) \geq n^\eta/\lambda\right)$$

$$\leq \sum_{\lambda^{(1+\Delta)/\eta} \leq i \leq n} \mathbb{P}\left(\mathrm{diam}\left(\mathscr{C}(i, G_n(\lambda) \setminus [i-1])\right) \geq n^\eta/\lambda\right)$$

$$\leq \sqrt{e} \sum_{\lambda^{(1+\Delta)/\eta} \leq i \leq n} \mathbb{P}\left(\mathrm{ht}\left(\mathrm{MTBP}_n^i([n] \setminus [i-1])\right) \geq n^\eta/(2\lambda)\right), \qquad (162)$$

where the last step uses arguments similar to the ones used in the proof of Lemma 6.17. To bound the summands in the last step in (162), we can use an argument similar to the one used in Sect. 6.5.

Definition 7.5 For $\lambda \geq 1$ and $2 \leq i \leq n$, consider the following (potentially) two layer process $\mathrm{L2BP}_n^i(\lambda)$:

(a) **Layer 1:** Start $\mathrm{MTBP}_n^i([n] \setminus [i-1])$, and run this process up to generation $n^\eta/(4\lambda)$. Call this the first layer. If there is at least one vertex in generation $n^\eta/(4\lambda)$, then we say that the first layer has been fully activated.
(b) **Layer 2:** If the first layer is fully activated, then starting from every vertex v in generation $n^\eta/(4\lambda)$, run independent $\mathrm{MTBP}_n^{\mathrm{type}(v)}([n] \setminus \{1\})$ processes. Call this the second layer. If any of these branching processes survives up to generation $n^\eta/(4\lambda)$, then we say that the second layer has been fully activated.

Let γ_* be as in Proposition 6.13. Then $\bar{\forall}n$, for any $2 \leq i \leq n$ and $\lambda \in [1/\gamma_* , n^\eta/10]$,

$$
\mathbb{P}\left(\text{ht}\left(\text{MTBP}^i_n([n] \setminus [i-1]) \right) \geq n^\eta/(2\lambda) \right) \leq \mathbb{P}\left(\text{ht}\left(\text{L2BP}^i_n(\lambda) \right) \geq n^\eta/(2\lambda) \right)
$$

$$
\leq Cw_i\left(\sum_{j \geq i} w^2_j/(\nu_n \ell_n) \right)^{\frac{n^\eta}{4\lambda}} (4\lambda)^{\frac{1}{\tau-3}} \cdot n^{-\alpha} \leq C'i^{-\alpha} \cdot \lambda^{\frac{1}{\tau-3}} \cdot \exp\left(-C''i^\eta/\lambda \right),
$$

(163)

where we have used Proposition 6.13 and arguments similar to the ones used in the proof of Lemma 6.21. Combining (163) with (162), and calculations similar to the ones in the proof of Lemma 6.21 show that $\bar{\forall}n$ and for $\lambda \in [2^\eta \vee (\gamma_*)^{-1}, n^\eta/10]$,

$$
\mathbb{P}\left(\text{diam}\left(G_n(\lambda) \setminus [\lambda^{\frac{1}{\eta}(1+\Delta)}] \right) \geq n^\eta/\lambda \right) \leq C_{164}(\Delta) \cdot \exp\left(-C\lambda^\Delta \right).
$$

(164)

Recall the notation $\mathcal{N}(\cdot, \cdot)$ introduced at the beginning of Sect. 3. Now note that on the complement of the event in (164),

$$
\mathcal{N}\left(\mathcal{C}^n_1(\lambda), 2n^\eta/\lambda \right) \leq \lambda^{(1+\Delta)/\eta}.
$$

(165)

By (144), for any $k \geq 1$ and $\lambda \geq 1$,

$$
\mathbb{P}\left(\mathcal{N}\left(\mathcal{M}^{\theta^*}_\lambda, 3/\lambda \right) \geq k \right) \leq \limsup_{n \to \infty} \mathbb{P}\left(\mathcal{N}\left(\text{CBD}^\infty(\mathcal{C}^n_1(\lambda)), 2n^\eta/\lambda \right) \geq k \right).
$$

(166)

Since $\text{CBD}^\infty(\mathcal{C}^n_1(\lambda))$ can be covered by $\mathcal{N}(\mathcal{C}^n_1(\lambda), 2n^\eta/\lambda) + \text{sp}(\mathcal{C}^n_1(\lambda))$ many $2n^\eta/\lambda$ balls, combining (164), (165), and (166) with Proposition 7.3, we see that there exists $\lambda_{167} \geq 1$ such that for all $\lambda \geq \lambda_{167}$,

$$
\mathbb{P}\left(\mathcal{N}\left(\mathcal{M}^{\theta^*}_\lambda, 3/\lambda \right) > 2\lambda^{(1+\Delta)/\eta} \right) \leq \left(1 + C_{164}(\Delta) \right) \exp\left(-C\lambda^\Delta \right).
$$

(167)

In the coupling on the right side of (151),

$$
\mathcal{N}\left(\mathcal{M}^{\theta^*}, 3\lambda^{-1} + d_\text{H}(\mathcal{M}^{\theta^*}, \mathcal{M}^{\theta^*}_\lambda) \right) \leq \mathcal{N}\left(\mathcal{M}^{\theta^*}_\lambda, 3\lambda^{-1} \right),
$$

which combined with (167) and Theorem 6.1 shows that for all $\lambda \geq \lambda_{21}(\Delta) \vee \lambda_{167}$,

$$
\mathbb{P}\left(\mathcal{N}\left(\mathcal{M}^{\theta^*}, 4\lambda^{-1+\Delta} \right) > 2\lambda^{(1+\Delta)/\eta} \right) \leq C_{168}(\Delta) \cdot \lambda^{-1/2}.
$$

(168)

Replacing λ by k^4 in (168) for large integer values of k, an application of the Borel-Cantelli lemma gives

$$
\limsup_{k \to \infty} \frac{\log \mathcal{N}\left(\mathcal{M}^{\theta^*}, 4k^{-4(1-\Delta)} \right)}{\log\left(k^{4(1-\Delta)}/4 \right)} \leq \frac{1+\Delta}{\eta(1-\Delta)} \quad \text{almost surely.}
$$

(169)

Sandwiching δ between $4k^{-4(1-\Delta)}$ and $4(k+1)^{-4(1-\Delta)}$ and letting $\delta \downarrow 0$, we can show using (169) that $\overline{\dim}\big(\mathcal{M}^{\theta^*}\big) \leq (1+\Delta)/(\eta(1-\Delta))$ almost surely. We complete the proof of (155) by letting $\Delta \downarrow 0$.

Proof of Lemma 7.4: By (40), $\bar{\forall} n$ and for $(\lambda, u) \in I^{(n),2}$,

$$\varphi_\lambda^{(n)}(u) \geq 2C_{170}u^{\tau-3} \,. \tag{170}$$

Thus, using (33) and (170), we see that $\bar{\forall} n$,

$$\Phi_\lambda^{(n)}(u) \leq \lambda u - 10 \cdot u\varphi_\lambda^{(n)}(u)/11 \leq u\big(\lambda - 20C_{170}u^{\tau-3}/11\big) \leq -C_{170}u^{\tau-2} \,,$$

whenever $(\lambda, u) \in I^{(n),2}$ and $9C_{170}u^{\tau-3} \geq 11\lambda$, and consequently,

$$\mathbb{P}\bigg(\mathscr{W}_x\big(\mathscr{C}\big(1, \mathscr{G}_n(\lambda)\big)\big) \geq u\bigg) \leq \mathbb{P}\big(Z_\lambda^{n,(1)}(u) \geq 0\big)$$

$$\leq \mathbb{P}\bigg(n^\eta x_1 + \sum_{j=2}^n \theta_{j,\lambda}\bigg(\mathbb{1}_{\{\xi_j^n \leq u\}} - \mathbb{P}\big(\xi_j^n \leq u\big)\bigg) \geq C_{170}u^{\tau-2}\bigg) \,,$$

where the first step uses Lemma 6.7, and the last step uses (160). Now an application of Lemma 6.9 will yield the bound in (157), and this bound will be valid for all large n, for $\lambda \in [\lambda_{157}, \varepsilon_{157}n^\eta]$ and $u \in \big[C_{157}\lambda^{1/(\tau-3)}, n^\alpha\big(\sigma_2^{(n)}\big)^{1/2}/(A_2 2^{\alpha+1})\big]$, where $C_{157} = \big(11/(9C_{170})\big)^{1/(\tau-3)}$, and λ_{157} and ε_{157} are appropriately chosen constants. Finally, this bound can be extended to all $u \geq C_{157}\lambda^{1/(\tau-3)}$ by simply noting that the probability on the left side of (157) is zero if $u > \sum_i x_i \asymp n^\alpha$. $\qquad \square$

7.5 The lower Minkowski dimension

In this section we will prove

$$\underline{\dim}\big(\mathcal{M}^{\theta^*}\big) \geq 1/\eta \quad \text{almost surely,} \tag{171}$$

which combined with (155) will complete the proof of Theorem 3.1 (c). To that end, let us first introduce some notation. For disjoint $A, B \subseteq [n]$ and $r \geq 1$, we write $A \xleftrightarrow{r} B$ to mean there exist $1 \leq t \leq r$, $v_0 \in A$, $v_t \in B$, and $v_1, \ldots, v_{t-1} \in [n]\backslash\{1\}$ such that the edges $\{v_i, v_{i+1}\}$, $i = 0, \ldots, t-1$, are present in $G_n(\lambda)$. If $A = \{i\}$, then we simply write $i \xleftrightarrow{r} B$ instead of $\{i\} \xleftrightarrow{r} B$.

Next, note that if $\mathcal{N}(X, u) \leq k$ for some metric space (X, d), $u \geq 0$, and $k \geq 1$, then for any $x_1, \ldots, x_{k+1} \in X$, there exist $1 \leq i < j \leq k+1$ such that $d(x_i, x_j) \leq 2u$. Hence, for any $x_1, \ldots, x_{2k} \in X$,

$$\#\big\{2 \leq i \leq 2k : \min_{j \in [i-1]} d(x_i, x_j) \leq 2u\big\} \geq k \,, \tag{172}$$

since otherwise, we would be able to find $1 = i_1 < i_2 < \ldots < i_{k+1} \leq 2k$ such that the minimum of the pairwise distances between these $k+1$ points is more than $2u$.

For $\Delta \in (0, 1/2]$, choose $\lambda_{174}(\Delta) \geq \lambda_{55} \vee \lambda_{157}$ such that $\lambda^{1/\eta}(\log \lambda)^{-3} > \lambda^{(1-\Delta)/\eta}$ for all $\lambda \geq \lambda_{174}(\Delta)$. For $\Delta \in (0, 1/2]$ and $\lambda \geq \lambda_{174}(\Delta)$, choose $n_{174}(\lambda) \geq 1$ such that $\lambda < (\varepsilon_{41} \wedge \varepsilon_{157}) n^\eta$ for all $n \geq n_{174}(\lambda)$. (Thus, these thresholds are chose in a way so that the bounds in (56) and (157) are applicable whenever $\lambda \geq \lambda_{174}$ and $n \geq n_{174}$.) Letting

$$\mathscr{D}_i := \left\{ [i] \subseteq \mathscr{C}_1^n(\lambda), \text{ and } i \xleftarrow{n^\eta/\lambda^{1+\hbar}} [i-1] \right\},$$

we see that for $\lambda \geq \lambda_{174}$, $n \geq n_{174}$, and

$$\hbar := 2\Delta/(\tau - 3), \tag{173}$$

$$\mathbb{P}\left(\mathscr{N}\left(n^{-\eta} \cdot \mathscr{C}_1^n(\lambda), \left(2\lambda^{1+\hbar}\right)^{-1} \right) \leq 2^{-1}\lambda^{\frac{1}{\eta}(1-\Delta)} \right)$$

$$\leq \mathbb{P}\left(\left\{ \#\{2 \leq i \leq \lambda^{\frac{1}{\eta}(1-\Delta)} : i \xleftarrow{n^\eta/\lambda^{1+\hbar}} [i-1]\} \right.\right.$$

$$\left.\left. \geq 2^{-1}\lambda^{\frac{1}{\eta}(1-\Delta)} \right\} \bigcap \left\{ i \in \mathscr{C}_1^n(\lambda) \; \forall \; i \leq \lambda^{\frac{1}{\eta}(1-\Delta)} \right\} \right)$$

$$+ \mathbb{P}\left(\exists i \leq \lambda^{\frac{1}{\eta}} (\log \lambda)^{-3} \text{ such that } i \notin \mathscr{C}_1^n(\lambda) \right)$$

$$\leq \mathbb{P}\left(\left\{ \#\{2 \leq i \leq \lambda^{\frac{1}{\eta}(1-\Delta)} : \mathscr{D}_i \text{ holds}\} \geq 2^{-1}\lambda^{\frac{1}{\eta}(1-\Delta)} \right\} \right) + \exp\left(-C\log^{3\alpha}\lambda \right)$$

$$\leq 2\lambda^{-\frac{1}{\eta}(1-\Delta)} \cdot \sum_{i=2}^{\lambda^{\frac{1}{\eta}(1-\Delta)}} \mathbb{P}(\mathscr{D}_i) + \exp\left(-C\log^{3\alpha}\lambda \right), \tag{174}$$

where the first step uses the observation around (172), the second step uses (56), and the last step uses Markov's inequality.

On the event $\mathscr{D}_i$, consider a shortest path γ from i to $[i-1]$. Then depending on where the shortest path from 1 to γ meets γ, we may consider three possibilities as in Fig. 3. Hence,

$$\mathscr{D}_i \subseteq \left(\bigcup_{j=1}^{i-1} \left(\{j \in \mathscr{C}_1^n(\lambda)\} \circ \{j \xleftarrow{n^\eta/\lambda^{1+\hbar}} i\} \right) \right)$$

$$\bigcup \left(\bigcup_{j=2}^{i-1} \left(\{i \in \mathscr{C}_1^n(\lambda)\} \circ \{j \xleftarrow{n^\eta/\lambda^{1+\hbar}} i\} \right) \right)$$

$$\bigcup \left(\left(\bigcup_{j=2}^{i-1} \bigcup_{k=i+1}^{n} \{k \in \mathscr{C}_1^n(\lambda)\} \circ \{k \xleftarrow{n^\eta/\lambda^{1+\hbar}} i\} \circ \{k \xleftarrow{n^\eta/\lambda^{1+\hbar}} j\} \right) \right), \tag{175}$$

where $\circ$ denotes disjoint occurrence of events. Writing $\sum_\square$ for sum over $t = 0, 1, \ldots, -1 + n^\eta/\lambda^{1+\hbar}$, we have, for any $i \neq j$,

$$
\begin{aligned}
&\mathbb{P}\big(i \xleftrightarrow{n^\eta/\lambda^{1+\hbar}} j\big) \\
&\leq \sum_\square \sum_{2 \leq v_1, \ldots, v_t \leq n} \left(p_\lambda^n \frac{w_i w_{v_1}}{\ell_n}\right)\left(p_\lambda^n \frac{w_{v_1} w_{v_2}}{\ell_n}\right) \cdots \left(p_\lambda^n \frac{w_{v_t} w_j}{\ell_n}\right) \\
&\leq \sum_\square \left(1 + \frac{\lambda}{n^\eta}\right)^{t+1} \left(\frac{w_i w_j}{v_n \ell_n}\right) \sum_{2 \leq v_1, \ldots, v_t \leq n} \left(\frac{w_{v_1}^2}{v_n \ell_n}\right) \cdots \left(\frac{w_{v_t}^2}{v_n \ell_n}\right) \\
&\leq \sum_\square \left(1 + \frac{\lambda}{n^\eta}\right)^{\frac{n^\eta}{\lambda^{1+\hbar}}} \left(\frac{w_i w_j}{v_n \ell_n}\right) \leq \frac{n^\eta}{\lambda^{1+\hbar}} \cdot \mathrm{e} \cdot \left(\frac{w_i w_j}{v_n \ell_n}\right) \leq \frac{C}{\lambda^{1+\hbar} i^\alpha j^\alpha}, \quad (176)
\end{aligned}
$$

where the third step uses (89), and the last step uses Assumption 2.4 (iii) and (7). Note also that for $2 \leq j \leq n$,

$$
\begin{aligned}
\mathbb{P}\big(j \in \mathscr{C}_1^n(\lambda)\big) &\leq \sum_{j' \in [n]} \mathbb{P}\big(j' \in \mathscr{C}_1^n(\lambda)\big) \mathbb{P}\big(\{j, j'\} \in E(G_n(\lambda))\big) \\
&\leq \sum_{j' \in [n]} p_\lambda^n \frac{w_j w_{j'}}{\ell_n} \mathbb{P}\big(j' \in \mathscr{C}_1^n(\lambda)\big) \\
&= p_\lambda^n \frac{w_j}{\ell_n} \cdot \mathbb{E}\big[\mathscr{W}\big(\mathscr{C}_1^n(\lambda)\big)\big] \leq C p_\lambda^n \frac{w_j}{\ell_n} \cdot n^\rho \lambda^{1/(\tau-3)} \\
&\leq C' \frac{\lambda^{1/(\tau-3)}}{j^\alpha}, \quad (177)
\end{aligned}
$$

where the penultimate uses Lemma 7.4 and is valid for $\lambda \geq \lambda_{174}$ and $n \geq n_{174}$. Thus, combining (175), (176), and (177), an application of the BK inequality shows, for $\lambda \geq \lambda_{174}$, $n \geq n_{174}$, and $2 \leq i \leq n$,

$$
\begin{aligned}
\mathbb{P}\big(\mathscr{D}_i\big) &\leq C \sum_{j=1}^{i-1} \left(\mathbb{1}_{\{j=1\}} + \frac{\lambda^{1/(\tau-3)}}{j^\alpha}\right) \frac{1}{\lambda^{1+\hbar} i^\alpha j^\alpha} \\
&\quad + C \sum_{j=2}^{i-1} \sum_{k=i+1}^{n} \left(\frac{\lambda^{1/(\tau-3)}}{k^\alpha} \cdot \frac{1}{\lambda^{1+\hbar} i^\alpha k^\alpha} \cdot \frac{1}{\lambda^{1+\hbar} j^\alpha k^\alpha}\right) \\
&\leq C' \left(\lambda^{\frac{4-\tau}{\tau-3} - \hbar} \cdot i^{\frac{\tau-4}{\tau-1}} + \lambda^{\frac{7-2\tau}{\tau-3} - 2\hbar} \cdot i^{\frac{2\tau-7}{\tau-1}}\right), \quad (178)
\end{aligned}
$$

where the last step follows from some routine calculations. Combining (178) with (174) shows that for $\lambda \geq \lambda_{174}$ and $n \geq n_{174}$,
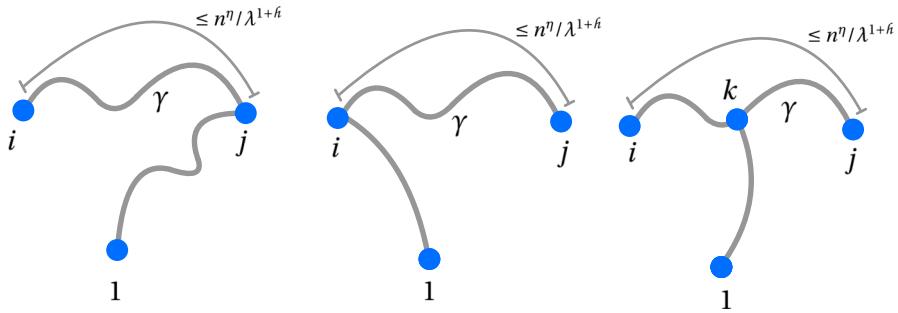
Fig. 3 The vertex $j \in [i-1]$ satisfies $\mathrm{dist}(i, j) = \min_{j' \in [i-1]} \mathrm{dist}(i, j')$, where dist denotes the graph distance in $G_n(\lambda)$. In the leftmost figure, $j = 1$ is possible, and $2 \leq j \leq i-1$ in the other two figures. In the rightmost figure, $i + 1 \leq k \leq n$

$$\mathbb{P}\left(\mathcal{N}\left(n^{-\eta} \cdot \mathscr{C}_1^n(\lambda), \left(2\lambda^{1+\hbar}\right)^{-1}\right) \leq 2^{-1}\lambda^{\frac{1}{\eta}(1-\Delta)}\right)$$

$$\leq \exp\left(-C \log^{3\alpha} \lambda\right)$$

$$+ C'\left[\lambda^{\frac{4-\tau}{\tau-3} - \hbar} \cdot \left(\lambda^{\frac{1}{\eta}(1-\Delta)}\right)^{\frac{\tau-4}{\tau-1}} + \lambda^{\frac{7-2\tau}{\tau-3} - 2\hbar} \cdot \left(\lambda^{\frac{1}{\eta}(1-\Delta)}\right)^{\frac{2\tau-7}{\tau-1}}\right]$$

$$= \exp\left(-C \log^{3\alpha} \lambda\right) + C'\left[\lambda^{\Delta\left(\frac{4-\tau}{\tau-3}\right) - \hbar} + \lambda^{\Delta\left(\frac{7-2\tau}{\tau-3}\right) - 2\hbar}\right]$$

$$\leq \exp\left(-C \log^{3\alpha} \lambda\right) + C''\lambda^{-\Delta/(\tau-3)}, \tag{179}$$

where the last step uses (173).

By (141), for any $k \geq 1$ and $\lambda \geq 1$,

$$\mathbb{P}\left(\mathcal{N}\left(\mathcal{S}_\lambda^{\boldsymbol{\theta}^*}, \left(4\lambda^{1+\hbar}\right)^{-1}\right) \leq k\right) \leq \limsup_{n\to\infty} \mathbb{P}\left(\mathcal{N}\left(n^{-\eta} \cdot \mathscr{C}_1^n(\lambda), \left(2\lambda^{1+\hbar}\right)^{-1}\right) \leq k\right),$$

which together with (179) yields, for $\lambda \geq \lambda_{174}$,

$$\mathbb{P}\left(\mathcal{N}\left(\mathcal{S}_\lambda^{\boldsymbol{\theta}^*}, \left(4\lambda^{1+\hbar}\right)^{-1}\right) \leq 2^{-1}\lambda^{\frac{1}{\eta}(1-\Delta)}\right) \leq \exp\left(-C \log^{3\alpha} \lambda\right) + C'\lambda^{-\Delta/(\tau-3)}. \tag{180}$$

Note that $\mathcal{N}\left(\mathcal{S}_\lambda^{\boldsymbol{\theta}^*}, r\right) \preceq_{\mathrm{sd}} \mathcal{N}\left(\mathcal{M}_\lambda^{\boldsymbol{\theta}^*}, r\right) \preceq_{\mathrm{sd}} \mathcal{N}\left(\mathcal{M}^{\boldsymbol{\theta}^*}, r\right)$ for any $r > 0$, where the first relation follows since $\mathcal{M}_\lambda^{\boldsymbol{\theta}^*} = \mathrm{CB}^\infty\left(\mathcal{S}_\lambda^{\boldsymbol{\theta}^*}\right)$, and the second relation follows from the existence of the coupling between $\mathcal{M}_\lambda^{\boldsymbol{\theta}^*}$ and $\mathcal{M}^{\boldsymbol{\theta}^*}$ as appears on the right side of (151). Hence, (180) continues to hold if we replace $\mathcal{S}_\lambda^{\boldsymbol{\theta}^*}$ by $\mathcal{M}^{\boldsymbol{\theta}^*}$. Let $g = g(\Delta) := 2(\tau - 3)/\Delta$. Using (180) with $\mathcal{M}^{\boldsymbol{\theta}^*}$ replacing $\mathcal{S}_\lambda^{\boldsymbol{\theta}^*}$, and letting $\lambda \to \infty$ along the sequence $\left(k^g, k \geq 1\right)$, we get, via an application of the Borel-Cantelli lemma,

$$\liminf_{k \to \infty} \frac{\log \mathcal{N}\left(\mathcal{M}^{\boldsymbol{\theta}^*}, \left(4k^{g(1+\hbar)}\right)^{-1}\right)}{\log \left(4k^{g(1+\hbar)}\right)} \geq \frac{1 - \Delta}{\eta(1 + \hbar)} \quad \text{almost surely}$$

for any $\Delta \in (0, 1/2]$. Sandwiching δ between $\left(4k^{g(1+\hbar)}\right)^{-1}$ and $\left(4(k+1)^{g(1+\hbar)}\right)^{-1}$ and letting $\delta \downarrow 0$, we conclude that for any $\Delta \in (0, 1/2]$,

$$\underline{\dim}\left(\mathcal{M}^{\boldsymbol{\theta}^*}\right) \geq \frac{1 - \Delta}{\eta(1 + \hbar)} \quad \text{almost surely.}$$

We complete the proof of (171) by letting $\Delta \downarrow 0$, and using (173).

8 Discussion

The random graph models considered in this paper are closely related to the multiplicative coalescent–a fact that was crucial in our proof. It is easy to argue heuristically that around the point of phase transition, many standard models of dynamic random graphs evolve roughly like the multiplicative coalescent. This intuition was formalized in [24], where it was shown that the continuum scaling limits for maximal components in the critical regime for a number of random graph models including the configuration model (under appropriate moment assumptions), the stochastic block model, and bounded size rules all have the same limit (up to constant scaling factors) as that of the Erdős-Rényi random graph established in [3]. In the heavy-tailed regime as considered in this paper, continuum limits for maximal components of critical inhomogeneous random graphs were first established in [29]. These results were then leveraged in [26] to establish the scaling limit of the critical configuration model.

We expect a similar program to be carried out building on the results of this paper for establishing universality of the MST for a host of random graph models. The two most important models for which this problem remains open are (i) the configuration model and (simple) random graphs with given degree sequence with tail exponent $\tau \in (3, 4)$, and (ii) a sequence of edge weighted graphs converging to a graphon whose leading eigenfunction is an element of $L^p[0, 1] \setminus L^3[0, 1]$ for some $p \in (2, 3)$. We expect that for both these models, the scaling limit of the MST on the giant components of these models, under suitable assumptions, will be the same as the ones obtained in this paper.

We now briefly remark on the assumptions in this paper. As mentioned before, Assumption 2.4 (i) implies supercriticality of the random graph model, whereas Assumption 2.4 (ii) corresponds to the condition in [11, Display (19)]. It should be possible to relax Assumption 2.4 (iii) and Assumption 2.4 (iv). For example, our proof of Proposition 6.13–a key ingredient used to establish tail bounds on the diameter outside the component of the vertex 1–does not require the full force of Assumption 2.4 (iii). This proof only uses the relation

$$\mathbb{P}\left(V_n \geq u\right) \asymp 1/u^{\tau - 2} \quad \text{for all} \quad n \geq 2, \quad \text{and} \quad u \in [1, v_2].$$

Similarly, it should be possible to relax Assumption 2.4 (iii) and Assumption 2.4 (iv) and still carry out the other steps in our proof at the cost of a more intricate analysis. In the context of critical inhomogeneous random graphs, [35] establishes convergence of the maximal components as well as compactness of the scaling limit under a weaker integrability condition [35, Display (9)]. For the critical heavy-tailed configuration model, [25] proves GHP convergence of the maximal components under [25, Assumptions 1 and 2], and [44] works with i.i.d. degrees having a power-law. However, all these papers deal only with the critical regime. It is not clear what the necessary and sufficient conditions are for results such as Theorem 3.1 to hold, as proving such a result requires careful control over the complement of the maximal component in the critical window as well as the barely supercritical regime. Establishing such a necessary and sufficient condition remains a challenging open problem.

We close this section with a discussion on the possible extension of the convergence in (11) with respect to the stronger GHP topology. Let $\boldsymbol{\mu}^n$ (resp. $\bar{\boldsymbol{\mu}}^n$) denote the uniform probability measure on the vertices of $\boldsymbol{M}^n$ (resp. $\bar{\boldsymbol{M}}^n$). View $\left(n^{-\eta} \cdot \boldsymbol{M}^n, \boldsymbol{\mu}^n\right)$ and $\left(n^{-\eta} \cdot \bar{\boldsymbol{M}}^n, \bar{\boldsymbol{\mu}}^n\right)$ as random metric measure spaces.

Conjecture 8.1 *Under Assumption 2.4 on the weight sequence, there exists a random compact measured $\mathbb{R}$-tree $\left(\mathscr{M}^{\boldsymbol{\theta}^*}, \mu\right)$ whose law depends only on $\boldsymbol{\theta}^*$ such that*

$$\left(n^{-\eta} \cdot \boldsymbol{M}^n, \boldsymbol{\mu}^n\right) \xrightarrow{\mathrm{d}} \left(\mathscr{M}^{\boldsymbol{\theta}^*}, \mu\right), \quad as \quad n \to \infty, \tag{181}$$

with respect to the GHP topology. Further, almost surely, the measure μ is nonatomic, i.e.,

$$\mathbb{P}\left(\mu(\{x\}) = 0 \text{ for all } x \in \mathscr{M}^{\boldsymbol{\theta}^*}\right) = 1, \tag{182}$$

and μ is concentrated on the set of leaves of $\mathscr{M}^{\boldsymbol{\theta}^*}$, i.e.,

$$\mathbb{P}\left(\mu\left(\mathscr{L}(\mathscr{M}^{\boldsymbol{\theta}^*})\right) = 1\right) = 1. \tag{183}$$

Moreover, (181) continues to hold under Assumption 2.4 if we replace the left side by $\left(n^{-\eta} \cdot \bar{\boldsymbol{M}}^n, \bar{\boldsymbol{\mu}}^n\right)$.

In the context of the complete graph, the analogues of (181) and (182) were proved in [6] and [8] respectively. In order to prove (181), it is enough to establish the following: Let $\bar{T}^n_{\lambda;i}$, $i = 1, \ldots, \bar{k}^n_\lambda$, be the trees in the forest obtained by removing from $\bar{M}^n$ the vertices in $\bar{M}^n_\lambda$ and all edges incident to the vertices in $\bar{M}^n_\lambda$. Let $\bar{Y}^n_\lambda := \max_{1 \leq i \leq \bar{k}^n_\lambda} \left|\bar{T}^n_{\lambda;i}\right|$. Then for all $\varepsilon > 0$,

$$\lim_{\lambda \to \infty} \limsup_{n \to \infty} \mathbb{P}\left(\bar{Y}^n_\lambda > \varepsilon n\right) = 0. \tag{184}$$

As mentioned in Sect. 6.1, in the setting of the complete graph, the analogue of (184) was established in [6, Lemma 4.11]. We briefly describe why proving (184) suffices.

Consider a degree sequence $\mathbf{d} = (d_1, \ldots, d_n)$, and let $\mathbb{CM}^{\mathbf{d}}$ (resp. $\mathbb{US}^{\mathbf{d}}$) be a configuration model (resp. uniform simple random graph) with degree sequence $\mathbf{d}$. (The reader can consult [8, Section 1.2 and Display (6.48)] for a quick overview of the definitions and the relevant properties of these models. We refer the reader to [54, 55] for a more detailed treatment.) Let U_e, $e \in E(\mathbb{CM}^{\mathbf{d}})$, be i.i.d. Uniform[0, 1] random variables conditional on $\mathbb{CM}^{\mathbf{d}}$. Let $M^{\mathbf{d}}$ be the MST of the component of the vertex 1 constructed using these edge weights. Fix $p \in (0, 1)$, and let $\mathbb{CM}_p^{\mathbf{d}}$ the graph with vertex set $[n]$ and edge set $\{e \in E(\mathbb{CM}^{\mathbf{d}}) : U_e \leq p\}$. Let $\mathscr{C}_p^{\mathbf{d}}(1)$ denote the component of the vertex 1 in $\mathbb{CM}_p^{\mathbf{d}}$, and $M_p^{\mathbf{d}}$ be the restriction of $M^{\mathbf{d}}$ to $\mathscr{C}_p^{\mathbf{d}}(1)$. For each vertex v in $\mathscr{C}_p^{\mathbf{d}}(1)$, let $d_{v,p}$ denote the degree of v in $\mathscr{C}_p^{\mathbf{d}}(1)$, and define $d_{v,p}^{\mathrm{avail}} := d_v - d_{v,p}$. Note that $M^{\mathbf{d}}$ can be viewed as $M_p^{\mathbf{d}}$ together with a collection of trees each of which is attached to a vertex of $M_p^{\mathbf{d}}$ via an edge; for every $v \in V(M_p^{\mathbf{d}})$, let $\mathfrak{T}_{v,p}^{(i)}$, $1 \leq i \leq r_{v,p}$, denote the trees that are attached to v (arranged following some deterministic rule). For every $v \in V(M_p^{\mathbf{d}})$, append $(d_{v,p}^{\mathrm{avail}} - r_{v,p})$ many zeros to the sequence $\left(|\mathfrak{T}_{v,p}^{(i)}|, \ 1 \leq i \leq r_{v,p}\right)$ and let $\left(\Delta_{v,p}^{(i)}, \ 1 \leq i \leq d_{v,p}^{\mathrm{avail}}\right)$ be a uniform permutation of the resulting sequence; use independent permutations for different $v \in V(M_p^{\mathbf{d}})$ that are also independent of all the other random variables being considered. Then conditional on $\mathbb{CM}_p^{\mathbf{d}}$ and $M_p^{\mathbf{d}}$, the family

$$\left(\Delta_{v,p}^{(i)}; \ 1 \leq i \leq d_{v,p}^{\mathrm{avail}}, v \in \mathscr{C}_p^{\mathbf{d}}(1)\right) \ \text{of random variables is exchangeable}. \quad (185)$$

The proof of (185) is similar to the proof of [8, Display (7.4)].

Now, using [58, Corollary 2.12], it is enough to prove (181) for an IRG with an equivalent kernel. Let $\widehat{G}_n$ be the random graph on $[n]$ obtained by placing an edge between i and j independently for each $i < j \in [n]$ with probability

$$\widehat{q}_{ij} := \frac{w_i w_j}{w_i w_j + \sum_{k=1}^n w_k} .$$

This is referred to as the Britton–Deijfen–Martin-Löf model [34]. This model has the following nice property: For $v \in [n]$, let D_v denote the degree of v of $\widehat{G}_n$, and define $\mathbf{D}_n := (D_1, \ldots, D_n)$. Then

$$\left(\widehat{G}_n \mid \mathbf{D}_n = \mathbf{d}\right) \ \overset{\mathrm{d}}{=} \ \mathbb{US}^{\mathbf{d}} \ \overset{\mathrm{d}}{=} \ \left(\mathbb{CM}^{\mathbf{d}} \mid \mathbb{CM}^{\mathbf{d}} \text{ is simple}\right) . \quad (186)$$

See [55, Theorem 7.18] for a proof of (186). For any degree sequence $\mathbf{d}$, let $g(\mathbf{d}) := \mathbb{P}\left(\mathbb{CM}^{\mathbf{d}} \text{ is simple}\right)$. Then using [57, Theorem 1.4], it can be shown that there exists $c > 0$ such that

$$\lim_{n \to \infty} \ \mathbb{P}\left(g(\mathbf{D}_n) \geq c\right) = 1 . \quad (187)$$

Define $\widehat{M}^n$, $\widehat{G}_n(\lambda)$, $\widehat{M}_\lambda^n$, and $\widehat{Y}_\lambda^n$ in a manner analogous to $\bar{M}^n$, $\bar{G}_n(\lambda)$, $\bar{M}_\lambda^n$, and $\bar{Y}_\lambda^n$ respectively. Once again, we can use [58, Corollary 2.12] to transfer the results proved

for the models considered earlier in this paper to the random graph $\widehat{G}_n$. Then (184) and Theorem 6.1 imply that for any $\varepsilon > 0$,

$$\lim_{\lambda \to \infty} \limsup_{n \to \infty} \left(\mathbb{P}\big(\widehat{Y}_\lambda^n > \varepsilon n\big) + \mathbb{P}\big(d_{\mathrm{H}}\big(\widehat{M}^n,\ \widehat{M}_\lambda^n\big) > \varepsilon n^\eta\big) \right) = 0 . \tag{188}$$

Further, by (140), Lemma 4.6, and [6, Theorem 3.3], for every $\lambda \geq 0$,

$$\big(n^{-\eta} \cdot \widehat{M}_\lambda^n,\ \widehat{\mu}_\lambda^{n,\boldsymbol{w}}\big) \xrightarrow{\ \mathrm{d}\ } \mathscr{M}_\lambda^{\boldsymbol{\theta}^*} \quad \text{w.r.t. the GHP topology} \tag{189}$$

as $n \to \infty$, where $\widehat{\mu}_\lambda^{n,\boldsymbol{w}}$ is the probability measure that assigns mass proportional to w_v to each vertex v in $\widehat{M}_\lambda^n$, and $\mathscr{M}_\lambda^{\boldsymbol{\theta}^*} = \mathrm{CB}^\infty\big(\mathcal{S}_\lambda^{\boldsymbol{\theta}^*}\big)$ is endowed with the measure inherited from $\mathcal{S}_\lambda^{\boldsymbol{\theta}^*}$.

For $v \in [n]$, let $D_{v,\lambda}$ denote the degree of v in $\widehat{G}_n(\lambda)$, and let $D_{v,\lambda}^{\mathrm{avail}} := D_v - D_{v,\lambda}$. Using (189), it is not too difficult to show that

$$\big(n^{-\eta} \cdot \widehat{M}_\lambda^n,\ \widehat{\mu}_\lambda^{n,\mathrm{avail}}\big) \xrightarrow{\ \mathrm{d}\ } \mathscr{M}_\lambda^{\boldsymbol{\theta}^*} \quad \text{w.r.t. the GHP topology}$$

as $n \to \infty$, where $\widehat{\mu}_\lambda^{n,\mathrm{avail}}$ is the probability measure that assigns mass proportional to $D_{v,\lambda}^{\mathrm{avail}}$ to each vertex v in $\widehat{M}_\lambda^n$. Let $\widehat{\mu}^n$ denote the uniform probability measure on the vertices of $\widehat{M}^n$. Then using (188), (189), (186), (185), (187), and [27, Lemma 7.5], it can be shown that for all $\varepsilon > 0$,

$$\lim_{\lambda \to \infty} \limsup_{n \to \infty} \mathbb{P}\left(d_{\mathrm{GHP}}\left(\big(n^{-\eta} \cdot \widehat{M}^n,\ \widehat{\mu}^n\big),\ \big(n^{-\eta} \cdot \widehat{M}_\lambda^n,\ \widehat{\mu}_\lambda^{n,\mathrm{avail}}\big) \right) > \varepsilon \right) = 0 .$$

The rest is routine.

Appendix A.

Proof of Lemma 6.12

By Assumption 2.4, for all $n \geq 4$ and $2 \leq j \leq n/2$,

$$\frac{\sum_{i=2}^j v_i}{\sum_{i=2}^n v_i} = \frac{1}{\ell_n} \sum_{i=2}^j w_i \asymp \frac{1}{n} \sum_{i=2}^j \left(\frac{n}{i}\right)^\alpha \asymp \left(\frac{j}{n}\right)^{1-\alpha} \asymp \frac{1}{v_j^{\tau-2}} ,$$

and hence, $\mathbb{P}(V_n \geq u) \asymp 1/u^{\tau-2}$ for all $n \geq 4$ and $u \in (v_{n/2}, v_2]$. Further, if $v_{n/2} \geq 1$, then for $u \in [1, v_{n/2}]$,

$$\frac{v_{n/2}^{\tau-2}}{u^{\tau-2}} \geq \mathbb{P}(V_n \geq u) \geq \mathbb{P}(V_n \geq v_{n/2}) \geq \frac{C}{v_{n/2}^{\tau-2}} \geq \frac{C}{\left(A_2 \cdot 2^\alpha\right)^{\tau-2}} \geq \frac{C}{\left(2^\alpha A_2 u\right)^{\tau-2}}.$$

Thus, $\mathbb{P}(V_n \geq u) \asymp 1/u^{\tau-2}$ for all $n \geq 4$ and $u \in [1, v_2]$.

Let us first prove the claimed lower bound in (65). For $n \geq 4$ and $k \geq 2$, we have

$$\begin{aligned}
\mathbb{P}(\text{Poi}(V_n) \geq k) &= \mathbb{E}\left[\int_0^{V_n} e^{-u} \frac{u^{k-1}}{(k-1)!} du\right] \\
&= \int_0^{v_2} e^{-u} \frac{u^{k-1}}{(k-1)!} \mathbb{P}(V_n \geq u) du \geq \int_1^{v_2} e^{-u} \frac{u^{k-1}}{(k-1)!} \mathbb{P}(V_n \geq u) du \\
&\geq C \int_1^{v_2} e^{-u} \frac{u^{k-1}}{(k-1)!} \frac{1}{u^{\tau-2}} du =: C(\mathfrak{T}_1 - \mathfrak{T}_2 - \mathfrak{T}_3),
\end{aligned} \tag{190}$$

where $\mathfrak{T}_1$, $\mathfrak{T}_2$, and $\mathfrak{T}_3$ are respectively the integrals $\int_0^\infty$, $\int_0^1$, and $\int_{v_2}^\infty$ of the integrand in the penultimate step. Then for $k \geq 2$,

$$\mathfrak{T}_1 = \frac{\Gamma(k-\tau+2)}{\Gamma(k)} \asymp \frac{k^{k-\tau+3/2}}{k^{k-1/2}} = \frac{1}{k^{\tau-2}}, \tag{191}$$

and

$$\mathfrak{T}_2 = \int_0^1 e^{-u} \frac{u^{k-1}}{(k-1)!} \frac{1}{u^{\tau-2}} du \leq \frac{1}{(k-1)! \times (k-\tau+2)}. \tag{192}$$

Next, note that for any $\delta > 0$, $u \mapsto u^\delta e^{-u/2}$ is decreasing on $[2\delta, \infty)$, and consequently, for $3 \leq k \leq v_2/2$,

$$\mathfrak{T}_3 \leq \frac{e^{-v_2/2} v_2^{k-\tau+1}}{(k-1)!} \int_{v_2}^\infty e^{-u/2} du = 2\frac{e^{-v_2} v_2^{k-\tau+1}}{(k-1)!} \asymp \frac{e^{-v_2}(ev_2)^k}{k^k} \cdot \frac{\sqrt{k}}{v_2^{\tau-1}},$$

where the last step follows from Stirling's approximation for $(k-1)!$. Using the fact that $\sup_{x \geq 1}(ev_2)^x/x^x = e^{v_2}$, we get, for $3 \leq k \leq v_2/2$,

$$\mathfrak{T}_3 \leq C\sqrt{k}/v_2^{\tau-1}. \tag{193}$$

From (190), (191), (192), and (193), we see that there exists $k_0 \geq 3$ such that for all large n and for $k_0 \leq k \leq v_2/2$,

$$\mathbb{P}(\text{Poi}(V_n) \geq k) \geq \frac{C}{k^{\tau-2}} - \frac{C'\sqrt{k}}{v_2^{\tau-1}} \geq \frac{C}{2k^{\tau-2}},$$

where we have used the scaling asymptotics of v_2 from Assumption 2.4. Choosing a smaller constant C if necessary, the above bound can be extended to $k \in \{1, \ldots, k_0\}$ and to all $n \geq 1$.

Now we prove the claimed upper bound in (65). Proceeding as in (190), we see that for any $n \geq 4$ and $k \geq 2$,

$$
\begin{aligned}
\mathbb{P}\big(\mathrm{Poi}(V_n) \geq k\big) &= \int_0^{v_2} e^{-u} \frac{u^{k-1}}{(k-1)!} \mathbb{P}(V_n \geq u) du \\
&\leq \int_0^1 \frac{u^{k-1}}{(k-1)!} du + \int_1^{v_2} e^{-u} \frac{u^{k-1}}{(k-1)!} \mathbb{P}(V_n \geq u) du \\
&\leq \frac{1}{k!} + C \int_1^{v_2} e^{-u} \frac{u^{k-1}}{(k-1)!} \frac{1}{u^{\tau-2}} du \leq \frac{1}{k!} + C \mathfrak{I}_1 \leq \frac{C'}{k^{\tau-2}},
\end{aligned}
$$

where the last step uses (191). This completes the proof. $\qquad\square$

Proof of Lemma 6.26

We can construct $\bar{G}_n(\gamma_n)$ in the following steps:

(a) Generate $\bar{\mathscr{C}}_1^n(\gamma_n/2)$.
(b) Place edges independently between $i, j \in [n] \backslash V\big(\bar{\mathscr{C}}_1^n(\gamma_n/2)\big)$, $i \neq j$, with respective probabilities $p_{\gamma_n}^n \bar{q}_{ij}$, where $\bar{q}_{ij}$ is as in (17).
(c) Place edges independently between $i \in V\big(\bar{\mathscr{C}}_1^n(\gamma_n/2)\big)$ and $j \in [n] \backslash V\big(\bar{\mathscr{C}}_1^n(\gamma_n/2)\big)$ with respective probabilities

$$
\frac{\big(p_{\gamma_n}^n - p_{\gamma_n/2}^n\big)\bar{q}_{ij}}{1 - p_{\gamma_n/2}^n \bar{q}_{ij}} \geq \frac{\varepsilon_{22} w_i w_j}{2 v_n \ell_n},
$$

where the last step holds for all large n.

Since $1 - e^{-u} \leq u \wedge 1$ for all $u \geq 0$, $\mathscr{W}\big(\mathscr{C}_1^n(\gamma_n/2)\big) \preceq_{\mathrm{sd}} \mathscr{W}\big(\bar{\mathscr{C}}_1^n(\gamma_n/2)\big)$. Further, $\varepsilon_{22} < \varepsilon_{41}$. Hence, an application of Proposition 6.15 and (42) shows that

$$
\mathbb{P}\big(\mathscr{W}\big(\bar{\mathscr{C}}_1^n(\gamma_n/2)\big) \geq C_{194} n\big) \geq 1 - \exp\big(-C n^\alpha\big) \tag{194}
$$

for all large n. Suppose for some $K_\star > 0$, $\mathscr{W}(\mathscr{C}_\star) > K_\star \log n$ for some component $\mathscr{C}_\star$ of $\bar{G}_n(\gamma_n)$, and $\mathscr{C}_\star \neq \bar{\mathscr{C}}_1^n(\gamma_n)$. Then the component $\mathscr{C}_\star$ will be constructed in step (b) above, and there will be no edges between $\mathscr{C}_\star$ and $\bar{\mathscr{C}}_1^n(\gamma_n/2)$ in step (c). Now, conditional on steps (a) and (b), on the event $\big\{\mathscr{W}\big(\bar{\mathscr{C}}_1^n(\gamma_n/2)\big) \geq C_{194} n\big\}$, the expected number of edges between $\mathscr{C}_\star$ and $\bar{\mathscr{C}}_1^n(\gamma_n/2)$ in step (c) is

$$
\geq \frac{\varepsilon_{22}}{2 v_n \ell_n} \times C_{194} n \times K_\star \log n \geq C_{195} K_\star \log n \tag{195}
$$

for all large n. Now the proof can be completed by combining (194) with an application of Bennett's inequality [31].

References

1. Abraham, R., Delmas, J.-F., Hoscheit, P.: A note on the Gromov–Hausdorff–Prokhorov distance between (locally) compact metric measure spaces. Electron. J. Probab. **18**(14), 1–21 (2013)
2. Addario-Berry, L., Broutin, N., Goldschmidt, C.: Critical random graphs: limiting constructions and distributional properties. Electron. J. Probab. **15**(25), 741–775 (2010)
3. Addario-Berry, L., Broutin, N., Goldschmidt, C.: The continuum limit of critical random graphs. Probab. Theory Relat. F. **152**(3–4), 367–406 (2012)
4. Addario-Berry, L.: Most trees are short and fat. Probab. Theory Relat. F. **173**(1–2), 1–26 (2019)
5. Addario-Berry, L., Bhamidi, S., Sen, S.: A probabilistic approach to the leader problem in random graphs. Random Struct. Algorithms **58**(1), 34–67 (2021)
6. Addario-Berry, L., Broutin, N., Goldschmidt, C., Miermont, G.: The scaling limit of the minimum spanning tree of the complete graph. Ann. Probab. **45**(5), 3075–3144 (2017)
7. Addario-Berry, L., Broutin, N., Reed, B.: Critical random graphs and the structure of a minimum spanning tree. Random Struct. Algorithms **35**(3), 323–347 (2009)
8. Addario-Berry, L., Sen, S.: Geometry of the minimal spanning tree of a random 3-regular graph. Probab. Theory Relat. F. **180**(3–4), 553–620 (2021)
9. Aldous, D.: A random tree model associated with random graphs. Random Struct. Algorithms **1**(4), 383–402 (1990)
10. Aldous, D.: Brownian excursions, critical random graphs and the multiplicative coalescent. Ann. Probab. **25**(2), 812–854 (1997)
11. Aldous, D., Limic, V.: The entrance boundary of the multiplicative coalescent. Electron. J. Probab. **3**(3), 59 (1998). ((**electronic**))
12. Aldous, D., Pitman, J.: Inhomogeneous continuum random trees and the entrance boundary of the additive coalescent. Probab. Theory Relat. F. **118**(4), 455–482 (2000)
13. Aldous, D., Steele, J.M.: Asymptotics for Euclidean minimal spanning trees on random points. Probab. Theory Relat. F. **92**(2), 247–258 (1992)
14. Aldous, D.J.: Deterministic and stochastic models for coalescence (aggregation and coagulation): a review of the mean-field theory for probabilists. Bernoulli **5**(1), 3–48 (1999)
15. Alexander, K.S.: Rates of convergence of means for distance-minimizing subadditive Euclidean functionals. Ann. Appl. Probab. **4**(3), 902–922 (1994)
16. Alexander, K.S.: Percolation and minimal spanning forests in infinite graphs. Ann. Probab. **23**(1), 87–104 (1995)
17. Alexander, K.S.: The RSW theorem for continuum percolation and the CLT for Euclidean minimal spanning trees. Ann. Appl. Probab. **6**(2), 466–494 (1996)
18. Athreya, S., Löhr, W., Winter, A.: The gap between Gromov-vague and Gromov–Hausdorff-vague topology. arXiv preprint arXiv:1407.6309 (2014)
19. Avram, F., Bertsimas, D.: The minimum spanning tree constant in geometrical probability and under the independent model: a unified approach. Ann. Appl. Probab. **2**(1), 113–130 (1992)
20. Beardwood, J., Halton, J.H., Hammersley, J.M.: The shortest path through many points. Proc. Camb. Philos. Soc. **55**, 299–327 (1959)
21. Bertoin, J.: Random Fragmentation and Coagulation Processes, Cambridge Studies in Advanced Mathematics, vol. 102. Cambridge University Press, Cambridge (2006)
22. Beveridge, A., Frieze, A., McDiarmid, C.: Random minimum length spanning trees in regular graphs. Combinatorica **18**(3), 311–333 (1998)
23. Bhamidi, S., van der Hofstad, R., van Leeuwaarden, J.S.H.: Novel scaling limits for critical inhomogeneous random graphs. Ann. Probab. **40**(6), 2299–2361 (2012)
24. S. Bhamidi, N. Broutin, S. Sen, and X. Wang, Scaling Limits of Random Graph Models at Criticality: Universality and the Basin of Attraction of the Erdős-Rényi Random Graph. arXiv preprint arXiv:1411.3417 (2014)
25. Bhamidi, S., Dhara, S., van der Hofstad, R., Sen, S.: Global Lower Mass-Bound for Critical Configuration Models in the Heavy-Tailed Regime. arXiv preprint arXiv:2005.02566 (2020)
26. Bhamidi, S., Dhara, S., van der Hofstad, R., Sen, S.: Universality for critical heavy-tailed network models: metric structure of maximal components. Electron. J. Probab. (2020). https://doi.org/10.1214/19-EJP408Processingraph

27. Bhamidi, S., Sen, S.: Geometry of the vacant set left by random walk on random graphs, Wright's constants, and critical random graphs with prescribed degrees. Random Struct. Algorithms **56**(3), 676–721 (2020)
28. Bhamidi, S., Sen, S., Wang, X.: Continuum limit of critical inhomogeneous random graphs. Probab. Theory Relat. F. **169**(1–2), 565–641 (2017)
29. Bhamidi, S., van der Hofstad, R., Sen, S.: The multiplicative coalescent, inhomogeneous continuum random trees, and new universality classes for critical random graphs. Probab. Theory Relat. F. **170**(1–2), 387–474 (2018)
30. Bollobás, B., Janson, S., Riordan, O.: The phase transition in inhomogeneous random graphs. Random Struct. Algorithms **31**(1), 3–122 (2007)
31. Boucheron, S., Lugosi, G., Massart, P.: Concentration Inequalities: A Nonasymptotic Theory of Independence. Oxford University Press, Oxford (2013)
32. Braunstein, L.A., Buldyrev, S.V., Cohen, R., Havlin, S., Stanley, H.E.: Optimal paths in disordered complex networks. Phys. Rev. Lett. **91**(16), 168701 (2003)
33. Braunstein, L.A., Wu, Z., Chen, Y., Buldyrev, S.V., Kalisky, T., Sreenivasan, S., Cohen, R., Lopez, E., Havlin, S., Stanley, H.E.: Optimal path and minimal spanning trees in random weighted networks,. Int. J. Bifurc. Chaos **17**(07), 2215–2255 (2007)
34. Britton, T., Deijfen, M., Martin-Lof, A.: Generating simple random graphs with prescribed degree distribution. J. Stat. Phys. **124**(6), 1377–1397 (2006)
35. Broutin, N., Duquesne, T., Wang, M.: Limits of Multiplicative Inhomogeneous Random Graphs and Levy Trees: Limit Theorems. arXiv preprint arXiv:2002.02769 (2020)
36. Burago, D., Burago, Y., Ivanov, S.: A Course in Metric Geometry, Graduate Studies in Mathematics, vol. 33. American Mathematical Society, Providence (2001)
37. Camarri, M., Pitman, J.: Limit distributions and random trees derived from the birthday problem with unequal probabilities. Electron. J. Probab. (Electron.) **5**(2), 18 (2000)
38. Chatterjee, S., Sen, S.: Minimal spanning trees and Stein's method. Ann. Appl. Probab. **27**(3), 1588–1645 (2017)
39. Chen, Y., Lopez, E., Havlin, S., Stanley, H.E.: Universal behavior of optimal paths in weighted networks with general disorder. Phys. Rev. Lett. **96**(6), 068702 (2006)
40. Chung, F., Lu, L.: The average distances in random graphs with given expected degrees. Proc. Natl. Acad. Sci. **99**(25), 15879–15882 (2002). (**electronic**))
41. Chung, F., Lu, L.: Connected components in random graphs with given expected degree sequences. Ann. Comb. **6**(2), 125–145 (2002)
42. Chung, F., Lu, L.: The average distance in a random graph with given expected degrees. Internet Math. **1**(1), 91–113 (2003)
43. Chung, F., Lu, L.: Complex graphs and networks, CBMS Regional Conference Series in Mathematics, vol. 107, Published for the Conference Board of the Mathematical Sciences, Washington, DC; by the AmericanMathematical Society, Providence, RI (2006)
44. Conchon-Kerjan, G., Goldschmidt, C.: The Stable Graph: The Metric Space Scaling Limit of a Critical Random Graph with i.i.d. Power-Law Degrees. arXiv:2002.04954 (2020)
45. Durrett, R.: Random graph dynamics. In: Lindsey, J.K. (ed.) Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge (2007)
46. Esseen, C.: On the concentration function of a sum of independent random variables. Zeitschrift fur Wahrscheinlichkeitstheorie und Verwandte Gebiete **9**(4), 290–308 (1968)
47. Evans, S.N.: Probability and real trees, Lecture Notes in Mathematics, vol. 1920, Springer, Berlin, 2008. Lectures from the 35th Summer School on Probability Theory held in Saint-Flour, July 6–23 (2005)
48. Evans, S.N., Pitman, J., Winter, A.: Rayleigh processes, real trees, and root growth with re-grafting. Probab. Theory Relat. F. **134**(1), 81–126 (2006)
49. Frieze, A., Ruszinko, M., Thoma, L.: A note on random minimum length spanning trees. Electron. J. Comb. **7**(1), R41 (2000)
50. MFrieze, A.: On the value of a random minimum spanning tree problem. Discrete Appl. Math. **10**(1), 47–56 (1985)
51. Frieze, A.M., McDiarmid, C.J.: On random minimum length spanning trees. Combinatorica **9**(4), 363–374 (1989)
52. Garban, C., Pete, G., Schramm, O.: The scaling limits of the minimal spanning tree and invasion percolation in the plane. Ann. Probab. **46**(6), 3501–3557 (2018)

53. Greven, A., Pfaffelhuber, P., Winter, A.: Convergence in distribution of random metric measure spaces (Λ-coalescent measure trees). Probab. Theory Relat. F. **145**(1–2), 285–322 (2009)
54. Van Der Hofstad, R.: Random graphs and complex networks. Volume 2, (2020+). Available at http://www.win.tue.nl/~rhofstad/
55. Van Der Hofstad, R.: Random graphs and complex networks. In: Lindsey, J.K. (ed.) Cambridge Series in Statistical and Probabilistic Mathematics, vol. 1. Cambridge University Press, Cambridge (2017)
56. Janson, S.: The minimal spanning tree in a complete graph and a functional limit theorem for trees in a random graph. Random Struct. Algorithms **7**(4), 337–355 (1995)
57. Janson, S.: The probability that a random multigraph is simple. Comb. Probab. Comput. **18**(1–2), 205–225 (2009)
58. Janson, S.: Asymptotic equivalence and contiguity of some random graphs. Random Struct. Algorithms **36**(1), 26–45 (2010)
59. Kesten, H., Lee, S.: The central limit theorem for weighted minimal spanning trees on random points. Ann. Appl. Probab. **6**(2), 495–527 (1996)
60. Klein, T., Rio, E.: Concentration around the mean for maxima of empirical processes. Ann. Probab. **33**(3), 1060–1077 (2005)
61. Kortchemski, I.: Sub-exponential tail bounds for conditioned stable Bienaymé–Galton–Watson trees. Probab. Theory Relat. F. **168**(1–2), 1–40 (2017)
62. Le Gall, J.-F.: Random trees and applications. Probab. Surv. **2**, 245–311 (2005)
63. Łuczak, T.: Component behavior near the critical point of the random graph process. Random Struct. Algorithms **1**(3), 287–310 (1990)
64. Miermont, G.: Tessellations of random maps of arbitrary genus. Ann. Sci. Éc. Norm. Supér. (4) **42**(5), 725–781 (2009)
65. Norros, I., Reittu, H.: On a conditionally Poissonian graph process. Adv. Appl. Probab. **38**(1), 59–75 (2006)
66. Penrose, M.D.: The random minimal spanning tree in high dimensions. Ann. Probab. **24**(4), 1903–1925 (1996)
67. Penrose, M.D.: The longest edge of the random minimal spanning tree. Ann. Appl. Probab. **7**(2), 340–361 (1997)
68. Penrose, M.D.: Random minimal spanning tree and percolation on the n-cube. Random Struct. Algorithms **12**(1), 63–82 (1998)
69. Penrose, M.D., Wade, A.R.: Random minimal directed spanning trees and Dickman-type distributions. Adv. Appl. Probab. **36**(3), 691–714 (2004)
70. Pitman, J.: Random mappings, forests, and subsets associated with Abel–Cayley–Hurwitz multinomial expansions, Sésm. Lothar. Combin. (electronic) **46**, B46h–45 (2001/2002)
71. Prim, R.C.: Shortest connection networks and some generalizations. Bell Syst. Tech. J. **36**(6), 1389–1401 (1957)
72. Steele, J.M.: Growth rates of Euclidean minimal spanning trees with power weighted edges. Ann. Probab. **16**(4), 1767–1787 (1988)
73. Wu, Z., Braunstein, L.A., Havlin, S., Stanley, H.E.: Transport in weighted networks: partition into superhighways and roads. Phys. Rev. Lett. **96**(14), 148702 (2006)