

EURANDOM PREPRINT SERIES

2015-020

August, 2015

**The multiplicative coalescent, inhomogeneous continuum random trees,  
and new universality classes for critical random graphs**

S. Bhamidi, R. van der Hofstad, S. Sen  
ISSN 1389-2355

# THE MULTIPLICATIVE COALESCENT, INHOMOGENEOUS CONTINUUM RANDOM TREES, AND NEW UNIVERSALITY CLASSES FOR CRITICAL RANDOM GRAPHS

SHANKAR BHAMIDI<sup>1</sup>, REMCO VAN DER HOFSTAD<sup>2</sup>, AND SANCHAYAN SEN<sup>2</sup>

**ABSTRACT.** One major open conjecture in the area of critical random graphs, formulated by statistical physicists, and supported by a large amount of numerical evidence over the last decade [23, 24, 28, 63] is as follows: for a wide array of random graph models with degree exponent  $\tau \in (3, 4)$ , distances between typical points both within maximal components in the critical regime as well as on the minimal spanning tree on the giant component in the supercritical regime scale like  $n^{(\tau-3)/(\tau-1)}$ .

In this paper we study the metric space structure of maximal components of the multiplicative coalescent, in the regime where the sizes converge to excursions of Lévy processes “without replacement” [10], yielding a completely new class of limiting random metric spaces. A by-product of the analysis yields the continuum scaling limit of one fundamental class of random graph models with degree exponent  $\tau \in (3, 4)$  where edges are rescaled by  $n^{-(\tau-3)/(\tau-1)}$  yielding the first rigorous proof of the above conjecture. The limits in this case are compact “tree-like” random fractals with finite fractal dimensions and with a dense collection of hubs (infinite degree vertices) a finite number of which are identified with leaves to form shortcuts. It is generally believed that dynamic versions of a number of fundamental random graph models, as one moves from the barely subcritical to the critical regime can be approximated by the multiplicative coalescent. In work in progress, the general theory developed in this paper is used to prove analogous limit results for other random graph models with degree exponent  $\tau \in (3, 4)$ .

Our proof makes crucial use of inhomogeneous continuum random trees (ICRT), which have previously arisen in the study of the entrance boundary of the *additive* coalescent. We show that *tilted* versions of the same objects using the associated mass measure, describe connectivity properties of the *multiplicative* coalescent. Since convergence of height processes of corresponding approximating **p**-trees is not known, we use general methodology in [14] and develop novel techniques relying on first showing convergence in the Gromov-weak topology and then extending this to Gromov-Hausdorff-Prokhorov convergence by proving a global lower mass-bound.

## CONTENTS

1. Introduction and results	2
1.1. Rank-one random graph	4
1.2. Connectivity asymptotics for the multiplicative coalescent	6
2. Definitions and limit objects	9
2.1. Convergence of metric spaces	9
2.2. Random <b>p</b> -trees and inhomogeneous continuum random trees (ICRTs)	12

<sup>1</sup>DEPARTMENT OF STATISTICS AND OPERATIONS RESEARCH, 304 HANES HALL, UNIVERSITY OF NORTH CAROLINA, CHAPEL HILL, NC 27599

<sup>2</sup>DEPARTMENT OF MATHEMATICS AND COMPUTER SCIENCE, EINDHOVEN UNIVERSITY OF TECHNOLOGY, THE NETHERLANDS

*E-mail addresses:* bhamidi@email.unc.edu, r.w.v.d.hofstad@tue.nl, sanchayan.sen1@gmail.com.

*2010 Mathematics Subject Classification.* Primary: 60C05, 05C80.

*Key words and phrases.* Multiplicative coalescent, **p**-trees, inhomogeneous continuum random trees, critical random graphs, Gromov-Hausdorff distance, Gromov-weak topology.

2.3. Continuum limits of components	15
3. Discussion	17
3.1. Universality and domains of attraction of critical random graph models	17
3.2. Minimal spanning tree on inhomogeneous random graphs	18
3.3. Inhomogeneous continuum random trees	18
3.4. Overview of the proof	19
4. Proofs: Asymptotics conditional on being connected	19
4.1. Tilted $\mathbf{p}$ -trees and connected components of $\mathcal{G}(\mathbf{x}, t)$	19
4.2. Convergence of connected components under weight assumptions	21
4.3. Uniform integrability of the tilt	24
4.4. Another construction of $\tilde{\mathcal{G}}_m(\mathbf{p}, a)$ and a modification	26
4.5. Completing the proof of Theorem 4.5.	29
4.6. Proof of Proposition 4.20.	35
5. Proofs: Convergence in Gromov-weak topology	43
5.1. Verification of weight assumptions in maximal components	44
5.2. Gromov-weak convergence in Theorem 1.1.	46
6. Proofs: Convergence in Gromov-Hausdorff-Prokhorov topology	47
6.1. Bound on size of $\varepsilon n^{\frac{\tau-3}{\tau-1}}$ -nets for the largest components	48
6.2. Proof of global lower-mass bound	52
7. Proofs: Fractal dimension	55
8. Open problems	57
Acknowledgements	59
References	59

## 1. INTRODUCTION AND RESULTS

The last two decades have witnessed a plethora of results regarding scaling limits of large discrete random objects to continuum analogs, ranging from Aldous's continuum random tree [7, 8, 51], Schramm-Loewner evolution and critical planar systems [61], to what is most closely related to this paper: scaling limits of maximal components in the critical regime for random graphs as well as the minimal spanning tree on the giant component in the supercritical regime [3–5].

Motivated by empirical observations on real world networks, in the last decade, researchers from a wide array of fields including computer science, the social sciences and statistical physics have proposed an enormous number of random graph models to explain various functionals of real world systems including power law degree distributions and small world scaling of distances between nodes in the network [6, 21, 32, 33, 35, 42, 55, 56]. Many of these models have a parameter  $t$  related to the edge density and a model dependent critical point  $t_c$ . Writing  $n$  for the number of vertices in the network, if  $t < t_c$  then the maximal connected component  $\mathcal{C}_1 = o_P(n)$  while if  $t > t_c$  one has a giant component  $\mathcal{C}_1(n) \sim f(t)n$  for some positive model dependent function  $f(t) > 0$  for  $t > t_c$ . The “ $t = t_c$ ” regime is often referred to as the critical regime. Just as a study of the classical critical Erdős-Rényi random graph spurred enormous activity in probabilistic combinatorics in the 90s [9, 21, 47, 52, 53], the study of the critical regime of these new random graph models and new phenomena such as *explosive percolation* [2, 60] have motivated a concerted effort to understand the critical regime of these new random graph models.

In this context, for more than a decade [23, 24, 28, 63], one of the fundamental open conjectures in this area (loosely stated) is as follows. Consider distances between typical points in the maximal component either in the critical regime or the minimal spanning tree on the giant component in the supercritical regime scale

- (a) If the random graph model has asymptotic degree distribution with finite third moments then distances scale like  $n^{1/3}$ .
- (b) If the random graph model has a limiting degree distribution  $\{p_k : k \geq 1\}$  with tail  $p_k \sim C/k^\tau$  for  $\tau \in (3, 4)$ , then distances scale like  $n^{(\tau-3)/(\tau-1)}$ .

**Contributions of the paper:** Since we will need to setup some notation before getting to the main results, let us give a general overview of the contributions of this paper.

- (i) **General theory:** The fundamental aim of the paper is development of general theory one can use to prove (b) in the conjecture above for a wide class of random graphs and in particular derive a new class of continuum scaling limits. To do so, we consider the multiplicative coalescent with entrance boundary in the space  $l_0$  as in [10] (see (1.13)). Viewing the maximal components as measured metric spaces (using graph distance and vertex weights), we show that these components with edges and associated measures properly rescaled converge to continuum random objects in the Gromov weak sense. These resulting objects are obtained via appropriate tilts and vertex identifications of inhomogeneous continuum random trees; *untitled* versions of the same objects have been used to describe the entrance boundary of the *additive* coalescent [13]. These resulting random objects are “tree-like” but with a dense collection of “hubs” (infinite degree vertices).
- (ii) **Proof techniques:** The standard technique in proving such results is studying height processes of certain spanning trees of the components and showing that these processes converge to limiting excursions that code the limiting random real trees. In our context, the convergence of height processes of corresponding approximating  $\mathbf{p}$ -trees is not known. In [11], the height processes of  $\mathbf{p}$ -trees were shown to converge to limiting excursions in certain regimes, but these results are not applicable to our situation.  
Owing to this reason, we develop new techniques relying on first showing convergence in Gromov-weak topology via a careful analysis of the tree spanning a finite collection of “typical” points in random “tilted”  $\mathbf{p}$ -trees. In one fundamental class of random graph models, we then extend Gromov-weak convergence to Gromov-Hausdorff-Prokhorov convergence by proving a global lower mass-bound.
- (iii) **Special case:** As an example of the general theory, we study the special case of the Norros-Reittu model [57] (which in the regime of interest has been proven [46] to be equivalent to the Chung-Lu model [30] and the rank-one random graph [22]). In this case, we show that the limiting spaces are compact. We also show that the upper box-counting dimension is bounded by  $(\tau - 2)/(\tau - 3)$  a.s. which we believe to be the correct value (see Conjecture 1.2).

In work in progress [19], we use the general theory in this paper to analyze another fundamental random graph model, the configuration model with degree distribution with exponent  $\tau \in (3, 4)$ , and derive the continuum analogs of the maximal components of this model. We defer a more detailed discussion of related work and the relevance of the current study to Section 3.

**Organization of the paper:** A reasonable amount of notation regarding the entrance boundary of the multiplicative coalescent is required to describe the main results (Theorems 1.7, 1.8). To ease the reader into the paper, we start in Section 1.1 with the special case of the Norros-Reittu

model and in Theorem 1.1 describe what the main results imply for this model. Then in Section 1.2 we define the multiplicative coalescent as well as the class of entrance boundaries of importance for the paper and then describe the two main results. The results use two notions of convergence of metric spaces, these are given a precise formulation in Section 2.1. Section 2.2 describes an important class of random trees called **p**-trees and the corresponding inhomogeneous continuum random trees that arise as scaling limits of these objects. These are then used in Section 2.3 to give a precise description of the scaling limits of maximal components. We discuss the relevance of the main results, relate these to existing work and give an overview of the proof in Section 3. The proofs of the main results are contained in Sections 4 - 7.

### 1.1. Rank-one random graph.

**1.1.1. Model formulation.** We start by describing a particular class of random graph models called the Poissonian random graph or the Norros-Reittu model [22, 57], sometimes also referred to as the rank-one random graph model [22]. In the regime of interest for this paper, as shown in [46], this model is equivalent to the Chung-Lu model [29–32] and the Britton-Deijfen-Martin-Löf model [25]. Start with vertex set  $[n] := \{1, 2, \dots, n\}$  and suppose each vertex  $i \in [n]$  has a weight  $w_i \geq 0$  attached to it; intuitively this measures the propensity or attractiveness of this vertex in the formation of links. Writing  $\mathbf{w} = (w_1, \dots, w_n)$ , place an edge between  $i$  and  $j$  independently for each  $i \neq j \in [n]$  with probability

$$q_{ij} = q_{ij}(\mathbf{w}) := 1 - \exp(-w_i w_j / \ell_n), \quad (1.1)$$

where  $\ell_n$  is the total weight given by

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

To complete the formulation, we need to specify how these vertex weights are chosen. Essentially we want the empirical distribution of weights  $n^{-1} \sum_{i \in [n]} \delta\{w_i\}$  to converge to a fixed pre-specified distribution  $F$  as  $n \rightarrow \infty$ . There are a number of ways to do this but for this paper, the following choice turns out to be convenient for a clear statement of the results. Let  $(w_i)_{i \in [n]}$  be constructed by

$$w_i := [1 - F]^{-1}(i/n), \quad i \in [n], \quad (1.2)$$

where  $F$  is a cumulative distribution function on  $[0, \infty)$  and  $[1 - F]^{-1}$  is the generalized inverse

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

We assume there exists  $\tau \in (3, 4)$  and  $c_F > 0$  such that

$$\lim_{x \rightarrow \infty} x^{\tau-1} [1 - F(x)] = c_F. \quad (1.3)$$

We will use  $W$  for a random variable with distribution  $F$ . We will use  $\text{NR}_n(\mathbf{w})$  to denote the corresponding random graph.

**1.1.2. Motivation and known results.** As described in the introduction, one impetus for the formulation of a wide array of network models, is to capture the heterogeneous and heavy-tailed nature of the degree distribution of empirical networks. Write  $N_k$  for the number of vertices with degree  $k$  in  $\text{NR}_n(\mathbf{w})$ . Under the assumptions in the previous section, one can show [22, Theorem 3.13] that

$$\frac{N_k}{n} \xrightarrow{P} \mathbb{E} \left( e^{-W} \frac{W^k}{k!} \right), \quad k \geq 0, \quad (1.4)$$

where  $W \sim F$ . In particular, the degree distribution also has tail exponent  $\tau$ . More important in the context of this paper is the connectivity threshold. For  $i \geq 1$  write  $\mathcal{C}_i$  for the  $i$ th largest connected component and let  $|\mathcal{C}_i|$  denote the number of vertices. Now define the parameter

$$\nu := \frac{\mathbb{E}(W^2)}{\mathbb{E}(W)}, \quad (1.5)$$

and note that by (1.3),  $\nu < \infty$ . Then by [22, Theorem 3.1 and Section 16.4], we have the following criterion for phase transition for the largest component:

- (a) **Supercritical regime:** If  $\nu > 1$ , then there exists  $\rho \in (0, 1)$  such that  $|\mathcal{C}_1|/n \xrightarrow{P} \rho$  whilst  $|\mathcal{C}_2|/n \xrightarrow{P} 0$ ;
- (b) **Subcritical regime:** If  $\nu < 1$  then  $|\mathcal{C}_1|/n \xrightarrow{P} 0$ .

The main aim of this paper is to understand the critical regime  $\nu = 1$ . In this setting, there are different universality classes. In the Erdős-Rényi universality class, critical clusters have size of order  $n^{2/3}$  and their metric space structure was discovered by Addario-Berry, Broutin and Goldschmidt [4]. Interestingly, when  $\mathbb{E}(W^3) < \infty$ , component sizes still scale like  $n^{2/3}$  [16] while assuming finite  $6 + \varepsilon$ -moments the metric space structure of rank-1 inhomogeneous random graphs is (apart from a trivial rescaling of size and time) the *same* [20]. However, when  $\mathbb{E}(W^3) = \infty$ , the scaling limits of critical clusters are dramatically different in the sense that their sizes are given by  $n^{(\tau-2)/(\tau-1)}$ , where  $\tau$  is the degree power-law exponent given by (1.3). In this paper, we focus on their *metric space structure*, obtained after rescaling edges by  $n^{-(\tau-3)/(\tau-1)}$  and taking the limit as  $n \rightarrow \infty$ . We show that this limiting metric space is compact and its upper box-counting dimension is bounded by  $(\tau-2)/(\tau-3)$ . This is expected to be the correct value of the Minkowski dimension whereas the Erdős-Rényi scaling limit has Minkowski dimension 2.

Further, in this paper, we analyze the entire *critical scaling window*. More precisely let  $\mathbf{w}$  denote the weight sequence as in (1.2) and fix  $\lambda \in \mathbb{R}$ . Now consider the weight sequence  $\mathbf{w}(\lambda) := (w_i(\lambda))_{i \in [n]}$  defined by

$$\mathbf{w}(\lambda) := \left(1 + \frac{\lambda}{n^{(\tau-3)/(\tau-1)}}\right) \mathbf{w}. \quad (1.6)$$

Write  $\text{NR}_n(\mathbf{w}(\lambda))$  for the corresponding random graph and let  $\mathcal{C}_i(\lambda)$  denote the corresponding  $i$ th largest component. Then this critical scaling window was first identified and studied in [62] where it was shown that for every fixed  $\lambda \in \mathbb{R}$ ,  $|\mathcal{C}_1|/n^{(\tau-2)/(\tau-1)}$  as well as  $n^{(\tau-2)/(\tau-1)}/|\mathcal{C}_1|$  are tight. The entire distributional asymptotics of component sizes were derived in [17] where it was shown that in the product topology on  $\mathbb{R}^{\mathbb{N}}$ ,

$$\left( \frac{|\mathcal{C}_i(\lambda)|}{n^{(\tau-2)/(\tau-1)}} : i \geq 1 \right) \xrightarrow{w} (Z_i(\lambda) : i \geq 1), \quad (1.7)$$

where  $(Z_i(\lambda) : i \geq 1)$  are excursions away from zero of a special stochastic process described in Section 1.2.

### 1.1.3. Our results. We make the following convention:

For any metric measure space  $(\mathbb{S}, d, \mu)$  and  $a > 0$ ,  $a\mathbb{S}$  will denote the metric measure space  $(\mathbb{S}, ad, \mu)$ , i.e, the space where the distance is scaled by  $a$  and the measure remains unchanged.

Consider the random graph  $\text{NR}_n(\mathbf{w}(\lambda))$  and view each connected component  $\mathcal{C}$  as a connected metric space via the usual graph distance where each edge has length one. Further, we can view each connected component  $\mathcal{C}$  as a measured metric space by assigning weight  $w_i/(\sum_{j \in \mathcal{C}} w_j)$  to

vertex  $i \in \mathcal{C}$ . Note that the normalization yields a probability measure on each connected component. Let  $\mathcal{S}$  denote the space of (equivalence classes) of **compact** measured metric spaces equipped with the Gromov-Hausdorff-Prokhorov metric (see Section 2.1.1 for definition). View

$$\mathbf{M}_n^{\text{nr}}(\lambda) := (\mathcal{C}_i(\lambda) : i \geq 1) \quad (1.8)$$

as a random element of  $\mathcal{S}^{\mathbb{N}}$ .

Next recall that the lower and upper box counting dimensions of a compact metric space  $\mathcal{M}$  are given by

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

respectively, where  $\mathcal{N}(\mathcal{M}, \delta)$  is the minimal number of open balls with radius  $\delta$  required to cover  $\mathcal{M}$ . Also let  $\dim_h(\mathcal{M})$  denote the Hausdorff dimension of  $\mathcal{M}$ .

The main result of this section is as follows:

**Theorem 1.1** (Scaling limits with degree exponent  $\tau \in (3, 4)$ ). *Fix  $\lambda \in \mathbb{R}$  and consider the critical Norros-Reittu model  $\text{NR}_n(\mathbf{w}(\lambda))$ , i.e, assume that  $v = 1$  where  $v$  is as in (1.5).*

*Then, there exists an appropriate limiting sequence of random compact metric measure spaces  $\mathbf{M}_\infty^{\text{nr}}(\lambda) := (M_i^{\text{nr}}(\lambda))_{i \geq 1}$  such that the components in the critical regime satisfy*

$$\frac{1}{n^{(\tau-3)/(\tau-1)}} \mathbf{M}_n^{\text{nr}}(\lambda) \xrightarrow{w} \mathbf{M}_\infty^{\text{nr}}(\lambda), \quad \text{as } n \rightarrow \infty. \quad (1.9)$$

Here convergence is with respect to the product topology on  $\mathcal{S}^{\mathbb{N}}$  induced by the Gromov-Hausdorff-Prokhorov metric on each coordinate  $\mathcal{S}$ . For each  $i \geq 1$ , the limiting metric spaces have the following properties:

- (a)  $M_i^{\text{nr}}(\lambda)$  is random compact metric measure space obtained by taking a random **real tree**  $\mathcal{T}_i(\lambda)$  and identifying a random (finite) number of pairs of points (thus creating shortcuts).
- (b) Call a point  $u \in \mathcal{T}_i(\lambda)$  a **hub** point if deleting the  $u$  results in infinitely many disconnected components of  $\mathcal{T}_i(\lambda)$ . Then  $\mathcal{T}_i(\lambda)$  has infinitely many hub points which are **everywhere dense** on the tree  $\mathcal{T}_i(\lambda)$ .
- (c) The upper box-counting dimension of  $M_i^{\text{nr}}(\lambda)$  satisfies

$$\overline{\dim}(M_i^{\text{nr}}(\lambda)) \leq \frac{\tau - 2}{\tau - 3} \quad \text{a.s.} \quad (1.10)$$

Consequently, the Hausdorff dimension satisfies the bound  $\dim_h(M_i^{\text{nr}}(\lambda)) \leq (\tau - 2)/(\tau - 3)$  a.s.

**Conjecture 1.2.** *We strongly believe that the lower box-counting dimension, the Hausdorff dimension, and the packing dimension of  $M_i^{\text{nr}}(\lambda)$  are all equal to  $(\tau - 2)/(\tau - 3)$  a.s. See Section 8 for a discussion.*

**1.2. Connectivity asymptotics for the multiplicative coalescent.** In this section we will consider a slightly more general setting than in Section 1.1. The motivation is as follows: recall that for the rank-one model, two vertices were connected with essentially probability proportional to the product of the weight between these two vertices. For probabilists, this connectivity pattern is quite reminiscent of the famous multiplicative coalescent [9, 10, 15]. Whilst interesting in its own right, its fundamental importance in the context of random graphs is as follows: A wide array of random graph models can be constructed in a dynamic fashion where as time progresses new edges are created between pre-existing clusters. Even though the merging dynamics between connected components tend to be quite different from that specified by the

multiplicative coalescent, the mergers from the barely subcritical regime through the critical scaling window can be approximated by the multiplicative coalescent. This idea was exploited in [18] to prove universality of scaling limits in the critical regime for several random graphs models.

Thus components at criticality of a wide array of random graph models can be thought of consisting of two major parts:

- (a) “Blobs” which are components formed in the barely subcritical regime.
- (b) Edges formed between such blobs as the system proceeds from the barely subcritical regime through the critical scaling window.

The results below (in particular Theorem 1.7) specify how to handle the second aspect. In a companion paper we show how one can use macroscopic averaging of distances within blobs in random graph models such as the configuration model to show that these models also have the same scaling limit in the critical regime as Theorem 1.1 in the setting where degrees obey power-laws with exponents  $\tau \in (3, 4)$ . Further, it will follow from Theorem 1.7 that the convergence in (1.9) holds with respect to the product topology induced by Gromov-weak topology on each coordinate. Therefore, Theorem 1.1 can be recovered partially from the more general Theorem 1.7 at the expense of working with a weaker topology.

Before stating the result we will need to define the multiplicative coalescent. The natural domain of this Markov process is the space

$$\ell_{\downarrow}^2 := \left\{ \mathbf{x} = (x_1, x_2, \dots) : x_1 \geq x_2 \geq \dots \geq 0, \sum_i x_i^2 < \infty \right\}, \quad (1.11)$$

equipped with the metric  $d(\mathbf{x}, \mathbf{y}) := \sqrt{\sum_{i \geq 1} (x_i - y_i)^2}$ . We will work in the simpler setup where the Markov process starts with a finite number of clusters, i.e, the process starts with  $\mathbf{x} \in \ell_{\downarrow}^2$  such that  $\exists n < \infty$  such that  $x_i = 0$  for  $i > n$ . Write  $\ell_{\downarrow}^2(n)$  for the collection of such vectors. Now the Markov process  $(\mathbf{X}(t))_{t \geq 0}$  with initial state  $\mathbf{X}(0) = \mathbf{x}$  evolves as follows. Write  $\mathbf{X}(t) = (X_i(t))_{i \geq 1}$ . Then for  $i \neq j$ , clusters  $i$  and  $j$  merge at rate  $X_i(t) \cdot X_j(t)$  into a single cluster of size  $X_i(t) + X_j(t)$ .

Note that for any fixed time  $t > 0$ , it is easy to find the distribution of masses  $\mathbf{X}(t)$  via the following random graph:

**Definition 1.3** (Random graph  $\mathcal{G}_n(\mathbf{x}, t)$ ). *Consider the vertex set  $[n] := \{1, 2, \dots, n\}$  and assign weight  $x_i$  to vertex  $i$ . Now connect each pair of vertices  $i, j$  with  $i \neq j$  independently with probability*

$$q_{ij} := 1 - \exp(-tx_i x_j). \quad (1.12)$$

*Call this random graph  $\mathcal{G}_n(\mathbf{x}, t)$ . For a connected component  $\mathcal{C} \subseteq \mathcal{G}_n(\mathbf{x}, t)$ , let  $\text{mass}(\mathcal{C}) := \sum_{i \in \mathcal{C}} x_i$ . Let  $(\mathcal{C}_i(t))_{i \geq 1}$  denote the connected components arranged in decreasing order of their masses.*

The following is obvious from the definition of the multiplicative coalescent:

**Lemma 1.4.** *For each fixed  $t \geq 0$ , the masses of the multiplicative coalescent at time  $t$  started with finite number of initial clusters with masses  $\mathbf{x}$  satisfies*

$$(X_i(t) : i \geq 1) \stackrel{d}{=} (\text{mass}(\mathcal{C}_i(t)) : i \geq 1).$$

Analogous to (1.11), consider the two spaces

$$\ell_{\downarrow}^3 := \left\{ \mathbf{c} := (c_1, c_2, \dots) : c_1 \geq c_2 \geq \dots \geq 0, \sum_{i \geq 1} c_i^3 < \infty \right\}, \quad l_0 := \ell_{\downarrow}^3 \setminus \ell_{\downarrow}^2. \quad (1.13)$$



These spaces turn out to be crucial in describing the entrance boundary of the eternal multiplicative coalescent in [10]. In the context of this paper, we are interested in studying scaling limits of connected components of the random graph  $\mathcal{G}_n(\mathbf{x}, t)$  when (suitably normalized) asymptotics of the weight vector  $\mathbf{x}$  are described by a vector  $\mathbf{c} \in l_0$ . Let

$$\sigma_r(\mathbf{x}) := \sum_i x_i^r, \quad 1 \leq r \leq 3. \quad (1.14)$$

We will make the following assumptions about the weight vector  $\mathbf{x} := \mathbf{x}(n)$  used to form the graph  $\mathcal{G}_n(\mathbf{x}, t)$ . These place the associated graph in a particular entrance boundary of the associated eternal multiplicative coalescent [10, Proposition 7].

**Assumption 1.5.** *For each  $n \geq 1$ , let  $\mathbf{x}^{(n)} = (x_i^{(n)} : 1 \leq i \leq n)$  be an initial finite-length vector belonging to  $\ell_1^2(n)$ . Suppose that as  $n \rightarrow \infty$  there exists  $\mathbf{c} \in l_0$  such that*

$$\frac{\sigma_3(\mathbf{x}^{(n)})}{(\sigma_2(\mathbf{x}^{(n)}))^3} \rightarrow \sum_j c_j^3, \quad (1.15)$$

$$\frac{x_j^{(n)}}{\sigma_2(\mathbf{x}^{(n)})} \rightarrow c_j \text{ for } j \geq 1, \text{ and} \quad (1.16)$$

$$\sigma_2(\mathbf{x}^{(n)}) \rightarrow 0. \quad (1.17)$$

Now let  $\{\xi_j : j \geq 1\}$  be a sequence of independent exponential random variables where  $\xi_j$  has rate  $c_j$  for each  $j \geq 1$ . For a fixed  $\lambda \in \mathbb{R}$ , consider the process

$$V_\lambda^{\mathbf{c}}(s) := \lambda s + \sum_j (c_j \mathbb{1}\{\xi_j \leq s\} - c_j^2 s), \quad s \geq 0. \quad (1.18)$$

It turns out that this process is well defined precisely if  $\mathbf{c} \in \ell_1^3$  [10]. Consider the “reflected at zero” process

$$\tilde{V}_\lambda^{\mathbf{c}}(s) := V_\lambda^{\mathbf{c}}(s) - \min_{0 \leq s' \leq s} \tilde{V}_\lambda^{\mathbf{c}}(s'). \quad (1.19)$$

Consider the excursions of  $\tilde{V}_\lambda^{\mathbf{c}}(\cdot)$  from zero. Then Aldous and Limic [10] showed that the lengths of these excursions are a.s. in  $l^2$  precisely when  $\mathbf{c} \in l_0$ , and thus can be arranged in decreasing order. Write

$$\mathcal{Z}(\lambda) := (\mathcal{Z}_1(\lambda), \mathcal{Z}_2(\lambda), \dots) \quad (1.20)$$

for these excursions in decreasing order of their length. Let  $Z_i(\lambda) := |\mathcal{Z}_i(\lambda)|$  denote the length of the  $i$ th largest excursion and let

$$\mathbf{Z}(\lambda) := (Z_1(\lambda), Z_2(\lambda), \dots) \in \ell_1^2 \quad \text{a.s.} \quad (1.21)$$

Then Aldous and Limic [10] proved the following result:

**Theorem 1.6.** *[10, Proposition 7] Fix  $\lambda \in \mathbb{R}$  and consider the time scale  $t_n := \lambda + [\sigma_2(\mathbf{x}^{(n)})]^{-1}$ . Under Assumptions (1.15), (1.16), (1.17), the masses of the connected components of the graph  $\mathcal{G}_n(\mathbf{x}, t_n)$  satisfy*

$$\left( \text{mass} \left[ \mathcal{C}_i \left( \lambda + \frac{1}{\sigma_2(\mathbf{x}^{(n)})} \right) : i \geq 1 \right] \right) \xrightarrow{w} \mathbf{Z}(\lambda), \quad \text{as } n \rightarrow \infty,$$

with respect to the topology in  $\ell_1^2$ , where  $\mathbf{Z}(\lambda)$  is as in (1.21).

Now consider the connected components in  $\mathcal{G}_n(\mathbf{x}, t)$ , and as before, view each component  $\mathcal{C}$  as a connected metric space via the usual graph distance where each edge has length one. Further, view each component  $\mathcal{C}$  as a measured metric space by assigning mass  $x_i / \text{mass}(\mathcal{C})$  to each vertex  $i \in \mathcal{C}$ . Let  $\mathcal{S}_*$  denote the space of (equivalence classes) of measured metric spaces equipped with Gromov-weak topology (see Section 2.1.2 for definition) and view

$$\mathbf{M}_n(\lambda) := \left( \mathcal{C}_i \left( \lambda + \frac{1}{\sigma_2(\mathbf{x}^{(n)})} \right) : i \geq 1 \right)$$

as a random element in  $\mathcal{S}_*^{\mathbb{N}}$ . Then our next result is about Gromov-weak convergence of  $\mathbf{M}_n(\lambda)$ .

**Theorem 1.7.** *Fix  $\lambda \in \mathbb{R}$ . Then under Assumption 1.5, there exist an appropriate limiting sequence of metric spaces  $\mathbf{M}_\infty^{\mathbf{c}}(\lambda) := (M_i^{\mathbf{c}}(\lambda) : i \geq 1)$  such that*

$$\sigma_2(\mathbf{x}^{(n)}) \mathbf{M}_n(\lambda) \xrightarrow{w} \mathbf{M}_\infty^{\mathbf{c}}(\lambda), \quad \text{as } n \rightarrow \infty.$$

Here weak convergence is on  $\mathcal{S}_*^{\mathbb{N}}$  which is equipped with the natural product topology induced by the Gromov-weak topology on each coordinate  $\mathcal{S}_*$ .

**Remark 1.** A full description of the limit objects is given in Section 2.3. The limit objects use tilted versions of inhomogeneous continuum random trees and checking compactness even of the original versions at this level of generality turns out to be quite intractable. However as the next theorem shows, in the special case of relevance to the rank one model, one can prove much more.

Consider the special sequence  $\mathbf{c} = \mathbf{c}(\alpha, \tau) := (c_i(\alpha, \tau) : i \geq 1) \in l_0$  with  $\tau \in (3, 4)$  and  $\alpha > 0$ , where

$$c_i(\alpha, \tau) := \frac{\alpha}{i^{1/(\tau-1)}}, \quad i \geq 1. \quad (1.22)$$

Then we have the following result about the limiting metric spaces.

**Theorem 1.8.** *Fix  $\alpha > 0$ ,  $\tau \in (3, 4)$  and let  $\mathbf{c} = \mathbf{c}(\alpha, \tau)$  as in (1.22). Consider the limiting metric spaces  $\mathbf{M}_\infty^{\mathbf{c}}(\lambda) := (M_i^{\mathbf{c}}(\lambda) : i \geq 1)$ .*

*Then almost surely  $M_i^{\mathbf{c}}(\lambda)$  is compact for every  $i \geq 1$ . Further, the upper box-counting dimension of  $M_i^{\mathbf{c}}(\lambda)$  satisfies*

$$\overline{\dim}(M_i^{\mathbf{c}}(\lambda)) \leq \frac{\tau-2}{\tau-3} \quad \text{a.s.} \quad (1.23)$$

*Consequently, the Hausdorff dimension satisfies the bound  $\dim_h(M_i^{\mathbf{c}}(\lambda)) \leq (\tau-2)/(\tau-3)$  a.s.*

**Remark 2.** Since we are dealing with equivalence classes of metric spaces (see Sections 2.1.1 and 2.1.2), Theorem 1.8 should be understood as claiming the existence of representative spaces  $M_i^{\mathbf{c}}(\lambda)$  that are compact, and satisfy the said conditions about the fractal dimensions. We will only work with these representative spaces throughout this paper.

## 2. DEFINITIONS AND LIMIT OBJECTS

**2.1. Convergence of metric spaces.** Proper notions of convergence of (measured) metric spaces is one of the central themes in this paper. Here we define the two topologies used in the statement of our results. We mainly follow [1, 26, 38, 39].

**2.1.1. Gromov-Hausdorff-Prokhorov metric.** In this section, all metric spaces under consideration will be compact metric spaces with associated probability measures. Let us first recall the Gromov-Hausdorff distance  $d_{\text{GH}}$  between metric spaces. Fix two metric spaces  $(X_1, d_1)$  and  $(X_2, d_2)$ . For a subset  $C \subseteq X_1 \times X_2$ , the *distortion* of  $C$  is defined as

$$\text{dis}(C) := \sup \{|d_1(x_1, y_1) - d_2(x_2, y_2)| : (x_1, x_2), (y_1, y_2) \in C\}. \quad (2.1)$$

A *correspondence*  $C$  between  $X_1$  and  $X_2$  is a measurable subset of  $X_1 \times X_2$  such that for every  $x_1 \in X_1$  there exists at least one  $x_2 \in X_2$  such that  $(x_1, x_2) \in C$  and vice-versa. The *Gromov-Hausdorff distance* between the two metric spaces  $(X_1, d_1)$  and  $(X_2, d_2)$  is defined as

$$d_{\text{GH}}(X_1, X_2) = \frac{1}{2} \inf \{\text{dis}(C) : C \text{ is a correspondence between } X_1 \text{ and } X_2\}. \quad (2.2)$$

Suppose  $(X_1, d_1)$  and  $(X_2, d_2)$  are two metric spaces and  $p_1 \in X_1$ , and  $p_2 \in X_2$ . Then the *pointed Gromov-Hausdorff distance* between  $\mathbf{X}_1 := (X_1, d_1, p_1)$  and  $\mathbf{X}_2 := (X_2, d_2, p_2)$  is given by

$$d_{\text{GH}}^{\text{pt}}(\mathbf{X}_1, \mathbf{X}_2) = \frac{1}{2} \inf \{\text{dis}(C) : C \text{ is a correspondence between } X_1 \text{ and } X_2 \text{ and } (p_1, p_2) \in C\}. \quad (2.3)$$

We will need a metric that also keeps track of associated measures on the corresponding spaces. A compact measured metric space  $(X, d, \mu)$  is a compact metric space  $(X, d)$  with an associated probability measure  $\mu$  on the Borel sigma algebra  $\mathcal{B}(X)$ . Given two compact measured metric spaces  $(X_1, d_1, \mu_1)$  and  $(X_2, d_2, \mu_2)$  and a measure  $\pi$  on the product space  $X_1 \times X_2$ , the *discrepancy* of  $\pi$  with respect to  $\mu_1$  and  $\mu_2$  is defined as

$$D(\pi; \mu_1, \mu_2) := \|\mu_1 - \pi_1\| + \|\mu_2 - \pi_2\|, \quad (2.4)$$

where  $\pi_1, \pi_2$  are the marginals of  $\pi$  and  $\|\cdot\|$  denotes the total variation distance between probability measures. Then the *Gromov-Hausdorff-Prokhorov distance* between  $X_1$  and  $X_2$  is defined as

$$d_{\text{GHP}}(X_1, X_2) := \inf \left\{ \max \left( \frac{1}{2} \text{dis}(C), D(\pi; \mu_1, \mu_2), \pi(C^c) \right) \right\}, \quad (2.5)$$

where the infimum is taken over all correspondences  $C$  and measures  $\pi$  on  $X_1 \times X_2$ .

Similar to (2.3), we can define a “*pointed Gromov-Hausdorff-Prokhorov distance*”,  $d_{\text{GHP}}^{\text{pt}}$  between two metric measure spaces  $X_1$  and  $X_2$  having two distinguished points  $p_1$  and  $p_2$  respectively by taking the infimum in (2.5) over all correspondences  $C$  and measures  $\pi$  on  $X_1 \times X_2$  such that  $(p_1, p_2) \in C$ .

Write  $\mathcal{S}$  for the collection of all measured metric spaces  $(X, d, \mu)$ . The function  $d_{\text{GHP}}$  is a pseudometric on  $\mathcal{S}$ , and defines an equivalence relation  $X \sim Y \Leftrightarrow d_{\text{GHP}}(X, Y) = 0$  on  $\mathcal{S}$ . Let  $\bar{\mathcal{S}} := \mathcal{S} / \sim$  be the space of isometry equivalent classes of measured compact metric spaces and  $\bar{d}_{\text{GHP}}$  the induced metric. Then by [1],  $(\bar{\mathcal{S}}, \bar{d}_{\text{GHP}})$  is a complete separable metric space. To ease notation, we will continue to use  $(\mathcal{S}, d_{\text{GHP}})$  instead of  $(\bar{\mathcal{S}}, \bar{d}_{\text{GHP}})$  and  $X = (X, d, \mu)$  to denote both the metric space and the corresponding equivalence class.

**2.1.2. Gromov-weak topology.** Here we mainly follow [38]. Introduce an equivalence relation on the space of complete and separable metric spaces that are equipped with a probability measure on the associated Borel  $\sigma$ -algebra by declaring two such spaces  $(X_1, d_1, \mu_1)$  and  $(X_2, d_2, \mu_2)$  to be equivalent when there exists an isometry  $\psi : \text{support}(\mu_1) \rightarrow \text{support}(\mu_2)$  such that  $\mu_2 = \psi_* \mu_1 := \mu_1 \circ \psi^{-1}$ , i.e., the push-forward of  $\mu_1$  under  $\psi$  is  $\mu_2$ . Write  $\mathcal{S}_*$  for the associated space of equivalence classes. As before, we will often ease notation by not distinguishing between a metric space and its equivalence class.

Fix  $n \geq 2$ , and a complete separable metric space  $(X, d)$ . Then given a collection of points  $\mathbf{x} := (x_1, x_2, \dots, x_n) \in X^n$ , let  $\mathbf{D}(\mathbf{x}) := (d(x_i, x_j))_{i,j \in [n]}$  denote the symmetric matrix of pairwise distances between the collection of points. A function  $\Phi : \mathcal{S}_* \rightarrow \mathbb{R}$  is called a polynomial of degree  $n$  if there exists a bounded continuous function  $\phi : \mathbb{R}_+^{n^2} \rightarrow \mathbb{R}$  such that

$$\Phi((X, d, \mu)) := \int \phi(\mathbf{D}(\mathbf{x})) \mu^{\otimes n}(d(\mathbf{x})). \quad (2.6)$$

Here  $\mu^{\otimes n}$  is the  $n$ -fold product measure of  $\mu$ . Let  $\Pi$  denote the space of all polynomials on  $\mathcal{S}_*$ .

**Definition 2.1** (Gromov-weak topology). *A sequence  $(X_n, d_n, \mu_n)_{n \geq 1} \in \mathcal{S}_*$  is said to converge to  $(X, d, \mu) \in \mathcal{S}_*$  in the Gromov-weak topology if and only if  $\Phi((X_n, d_n, \mu_n)) \rightarrow \Phi((X, d, \mu))$  for all  $\Phi \in \Pi$ .*

In [38, Theorem 1] it is shown that  $\mathcal{S}_*$  is a Polish space under the Gromov-weak topology. It is also shown that, in fact, this topology can be completely metrized using the so-called Gromov-Prokhorov metric.

**2.1.3. Spaces of trees with edge lengths, leaf weights and root-to-leaf measures.** In the proof of the main results we will need the following two spaces built on top of the space of discrete trees. The first space  $\mathbf{T}_{IJ}$  was formulated in [12, 13] where it was used to study trees spanning a finite number of random points sampled from an inhomogeneous continuum random tree (described in the next section). We will use the same notation in this paper.

**The space  $\mathbf{T}_{IJ}$ :** Fix  $I \geq 0$  and  $J \geq 1$ . Let  $\mathbf{T}_{IJ}$  be the space of trees having the following properties:

- (a) There are exactly  $J$  leaves labeled  $1+, \dots, J+$ , and the tree is rooted at another labeled vertex  $0+$ .
- (b) There may be extra labeled vertices (called hubs) with distinct labels in  $\{1, 2, \dots, I\}$ . (It is possible that only some, and not all labels in  $\{1, 2, \dots, I\}$  are used.)
- (c) Every edge  $e$  has a strictly positive edge length  $l_e$ .

A tree  $\mathbf{t} \in \mathbf{T}_{IJ}$  can be viewed as being composed of two parts: (a)  $\text{shape}(\mathbf{t})$  describing the shape of the tree (including the labels of leaves and hubs) but ignoring edge lengths. The set of all possible shapes  $\mathbf{T}_{IJ}^{\text{shape}}$  is obviously finite for fixed  $I, J$ . (b) The edge lengths  $\mathbf{l}(\mathbf{t}) := (l_e : e \in \mathbf{t})$ . Consider the product topology on  $\mathbf{T}_{IJ}$  consisting of the discrete topology on  $\mathbf{T}_{IJ}^{\text{shape}}$  and the product topology on  $\mathbb{R}^d$ .

**The space  $\mathbf{T}_{IJ}^*$ :** We will need a slightly more general space. Along with the three attributes above in  $\mathbf{T}_{IJ}$ , the trees in this space have the following two additional properties. Let  $\mathcal{L}(\mathbf{t}) := \{1+, \dots, J+\}$  denote the collection of non-root leaves in  $\mathbf{t}$ . Then every leaf  $v \in \mathcal{L}(\mathbf{t})$  has the following attributes:

- (d) **Leaf weights:** A strictly positive number  $A(v)$ . Write  $\mathbf{A}(\mathbf{t}) := (A(v) : v \in \mathcal{L}(\mathbf{t}))$ .
- (e) **Root-to-leaf measures:** A probability measure  $\nu_{\mathbf{t}, v}$  on the path  $[0+, v]$  connecting the root and the leaf  $v$ . Here the path is viewed as a line segment pointed at  $0+$  and has the usual Euclidean topology. Write  $\mathbf{v}(\mathbf{t}) := (\nu_{\mathbf{t}, v} : v \in \mathcal{L}(\mathbf{t}))$  for this collection of probability measures.

In addition to the topology on  $\mathbf{T}_{IJ}$ , the space  $\mathbf{T}_{IJ}^*$  with these additional two attributes inherits the product topology on  $\mathbb{R}^J$  owing to leaf weights and  $(d_{\text{GHP}}^{\text{pt}})^J$  owing to the the root to leaf measures.

For consistency, we add to the spaces  $\mathbf{T}_{IJ}$  and  $\mathbf{T}_{IJ}^*$  a conventional state  $\partial$ . Its use will be clear later.

**2.2. Random  $\mathbf{p}$ -trees and inhomogeneous continuum random trees (ICRTs).** For fixed  $m \geq 1$ , write  $\mathbb{T}_m$  and  $\mathbb{T}_m^{\text{ord}}$  for the collection of all rooted trees with vertex set  $[m]$  and rooted ordered trees with vertex set  $[m]$  respectively. Here we will view a rooted tree as being directed with the root being the original progenitor and each edge being directed from child to parent. An ordered rooted tree is a tree where children of each individual are assigned an order (meant to describe for example orientation in a planar embedding, e.g., right to left or some notion of age, e.g., oldest to youngest).

In this section, we define a family of random tree models called  $\mathbf{p}$ -trees [27, 59], and their corresponding limits, the so-called inhomogeneous continuum random trees, which play a key role in describing the limit metric spaces as well as in the proof. Fix  $m \geq 1$ , and a probability mass function  $\mathbf{p} = (p_1, p_2, \dots, p_m)$  with  $p_i > 0$  for all  $i \in [m]$ . A  $\mathbf{p}$ -tree is a random tree in  $\mathbb{T}_m$ , with law as follows. For any fixed  $\mathbf{t} \in \mathbb{T}_m$  and  $v \in \mathbf{t}$ , write  $d_v(\mathbf{t})$  for the number of children of  $v$  in the tree  $\mathbf{t}$ . Then the law of the  $\mathbf{p}$ -tree, denoted by  $\mathbb{P}_{\text{tree}}$ , is defined as:

$$\mathbb{P}_{\text{tree}}(\mathbf{t}) = \mathbb{P}_{\text{tree}}(\mathbf{t}; \mathbf{p}) = \prod_{v \in [m]} p_v^{d_v(\mathbf{t})}, \quad \mathbf{t} \in \mathbb{T}_m. \quad (2.7)$$

Generating a random  $\mathbf{p}$ -tree  $\mathcal{T} \sim \mathbb{P}_{\text{tree}}$  and then assigning a uniform random order on the children of every vertex  $v \in \mathcal{T}$  gives a random element with law  $\mathbb{P}_{\text{ord}}(\cdot; \mathbf{p})$  given by

$$\mathbb{P}_{\text{ord}}(\mathbf{t}) = \mathbb{P}_{\text{ord}}(\mathbf{t}; \mathbf{p}) = \prod_{v \in [m]} \frac{p_v^{d_v(\mathbf{t})}}{(d_v(\mathbf{t}))!}, \quad \mathbf{t} \in \mathbb{T}_m^{\text{ord}}. \quad (2.8)$$

Obviously a  $\mathbf{p}$ -tree can be constructed by first generating an ordered  $\mathbf{p}$ -tree with the above distribution and then forgetting about the order.

In a series of papers [11–13] it was shown that  $\mathbf{p}$ -trees, under various assumptions, converge to inhomogeneous continuum random trees which we now describe. Recall the space  $\ell_{\downarrow}^2$  in (1.11). Consider the subset  $\Theta \subset \ell_{\downarrow}^2$  given by

$$\Theta := \left\{ \boldsymbol{\theta} := (\theta_i : i \geq 1) \in \ell_{\downarrow}^2 : \sum_{i=1}^{\infty} \theta_i = \infty, \sum_{i=1}^{\infty} \theta_i^2 = 1 \right\}. \quad (2.9)$$

Now recall from [37, 51] that a real tree is a metric space  $(\mathcal{T}, d)$  that satisfies the following for every pair  $a, b \in \mathcal{T}$ :

- (a) There is a **unique** isometric map  $f_{a,b} : [0, d(a, b)] \rightarrow \mathcal{T}$  such that  $f_{a,b}(0) = a$ ,  $f_{a,b}(d(a, b)) = b$ .
- (b) For any continuous one-to-one map  $g : [0, 1] \rightarrow \mathcal{T}$  with  $g(0) = a$  and  $g(1) = b$ , we have  $g([0, 1]) = f_{a,b}([0, d(a, b)])$ .

**Construction of the ICRT:** Given  $\boldsymbol{\theta} \in \Theta$ , we will now define the inhomogeneous continuum random tree  $\mathcal{T}^{\boldsymbol{\theta}}$ . We mainly follow the notation in [13]. Assume that we are working on a probability space  $(\Omega, \mathcal{F}, \mathbb{P}_{\boldsymbol{\theta}})$  rich enough to support the following:

- (a) For each  $i \geq 1$ , let  $\mathcal{P}_i := (\xi_{i,1}, \xi_{i,2}, \dots)$  be a rate  $\theta_i$  Poisson process, independent for different  $i$ . The first point of each process  $\xi_{i,1}$  is special and is called a *joinpoint*, whilst the remaining points  $\xi_{i,j}$  with  $j \geq 2$  will be called *i-cutpoints* [13].
- (b) Independent of the above, let  $\mathbf{U} = (U_j^{(i)} : j \geq 1, i \geq 1)$  be a collection of i.i.d. uniform  $(0, 1)$  random variables. These are not required to construct the tree but will be used to define a certain function on the tree.

The **random** real tree (with marked vertices)  $\mathcal{T}_{(\infty)}^{\boldsymbol{\theta}}$  is then constructed as follows:

- (i) Arrange the cutpoints  $\{\xi_{i,j} : i \geq 1, j \geq 2\}$  in increasing order as  $0 < \eta_1 < \eta_2 < \dots$ . The assumption that  $\sum_i \theta_i^2 < \infty$  implies that this is possible. For every cutpoint  $\eta_k = \xi_{i,j}$ , let  $\eta_k^* := \xi_{i,1}$  be the corresponding joinpoint.
- (ii) Next, build the tree inductively. Start with the branch  $[0, \eta_1]$ . Inductively assuming we have completed step  $k$ , attach the branch  $(\eta_k, \eta_{k+1}]$  to the joinpoint  $\eta_k^*$  corresponding to  $\eta_k$ .

Write  $\mathcal{T}_0^\theta$  for the corresponding tree after one has used up all the branches  $[0, \eta_1], \{(\eta_k, \eta_{k+1}] : k \geq 1\}$ . Note that for every  $i \geq 1$ , the joinpoint  $\xi_{i,1}$  corresponds to a vertex with infinite degree. Label this vertex  $i$ . The ICRT  $\mathcal{T}_{(\infty)}^\theta$  is the completion of the marked metric tree  $\mathcal{T}_0^\theta$ . As argued in [13, Section 2], this is a real-tree as defined above which can be viewed as rooted at the vertex corresponding to zero. We call the vertex corresponding to joinpoint  $\xi_{i,1}$  **hub**  $i$ . Since  $\sum_i \theta_i = \infty$ , one can check that hubs are almost everywhere dense on  $\mathcal{T}_{(\infty)}^\theta$ .

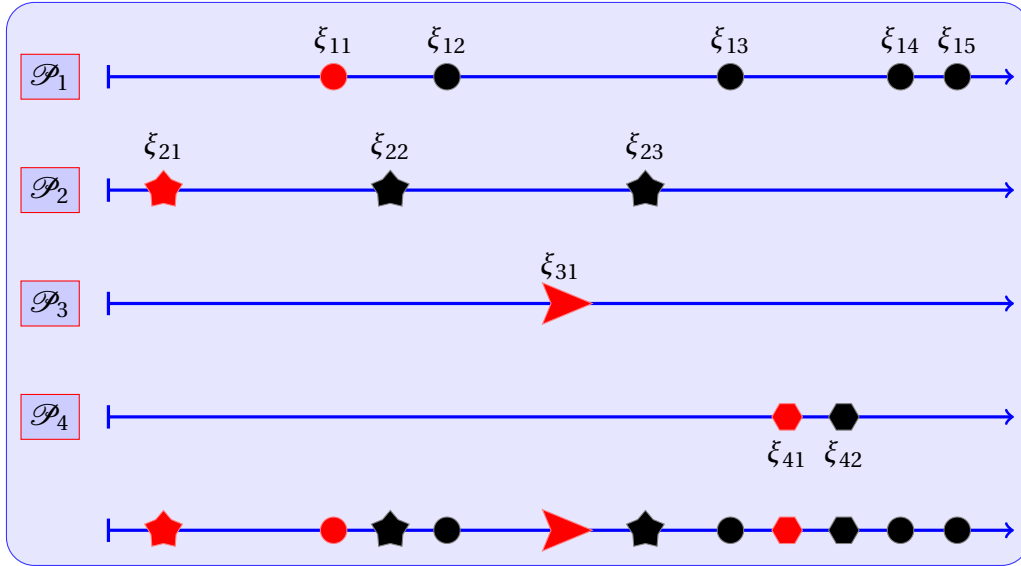


FIGURE 2.1. An illustration of the ICRT construction with four point process  $\{\mathcal{P}_i : 1 \leq i \leq 4\}$ . The red points represent the *joinpoint* of the corresponding point process and the blue points the corresponding cutpoints. The last line contains the union of the four point processes. See Figure 2.2 for the corresponding tree.

**Remark 3.** The uniform random variables  $(U_j^{(i)} : j \geq 1, i \geq 1)$  give rise to a natural ordering on  $\mathcal{T}_{(\infty)}^\theta$  (or a planar embedding of  $\mathcal{T}_{(\infty)}^\theta$ ) as follows. For  $i \geq 1$ , let  $(\mathcal{T}_j^{(i)} : j \geq 1)$  be the collection of subtrees hanging off of the  $i$ th hub. Associate  $U_j^{(i)}$  with the subtree  $\mathcal{T}_j^{(i)}$ , and think of  $\mathcal{T}_{j_1}^{(i)}$  appearing “to the right of”  $\mathcal{T}_{j_2}^{(i)}$  if  $U_{j_1}^{(i)} < U_{j_2}^{(i)}$ . This is the natural ordering on  $\mathcal{T}_{(\infty)}^\theta$  when it is being viewed as a limit of ordered  $\mathbf{p}$ -trees. We can think of the pair  $(\mathcal{T}_{(\infty)}^\theta, \mathbf{U})$  as the **ordered ICRT**.

**Reduced tree  $r_{IJ}^{(\infty)}$ :** Fix  $I \geq 0$  and  $J \geq 1$ . Now let  $\eta_0 = 0$  and for  $j \geq 0$  call vertex  $\eta_j$  the  $j$ th sampled leaf and label this as  $j+$  to differentiate this from hub  $j$ . Note that the subtree of  $\mathcal{T}_{(\infty)}^\theta$  spanned by  $\{0+, 1+, \dots, J+\}$  (namely the part of the tree constructed from the interval  $[0, \eta_J]$ ) is a tree in the usual sense with random edge lengths. For all hubs  $i$ , if  $i \leq I$ , retain its label and remove the

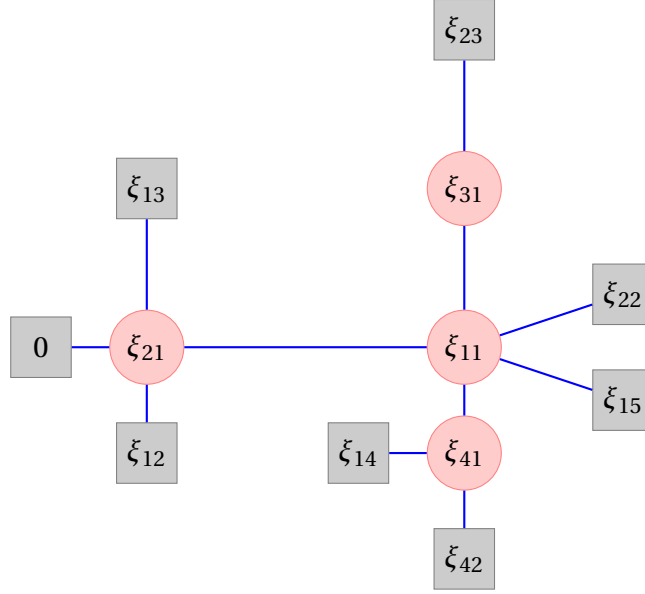
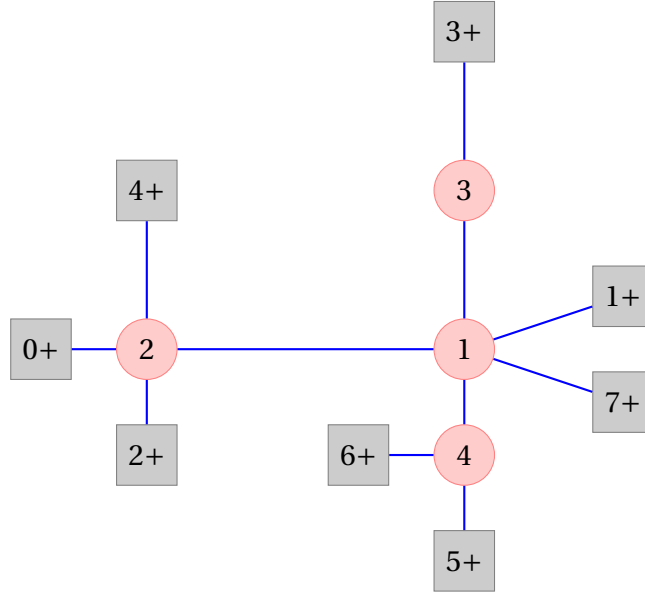


FIGURE 2.2. The tree constructed via the stick-breaking construction from Figure 2.1.

label otherwise. This gives a random element of  $\mathbf{T}_{IJ}$  (recall the definition Section 2.1.3), which we denote by  $r_{IJ}^{(\infty)}$ . See Figure 2.3 corresponding to the stick-breaking construction in Figures 2.1 and 2.2.

FIGURE 2.3. Reduced tree  $r_{46}$  corresponding to the tree in Figure 2.2.

**Mass measure:** For every vertex  $v \in \mathcal{T}_{(\infty)}^{\theta}$ , define the *degree* of  $v$  to be the number of connected components of  $\mathcal{T}_{(\infty)}^{\theta} \setminus \{v\}$ . Vertices with degree one are called *leaves* of  $\mathcal{T}_{(\infty)}^{\theta}$  and all other vertices form the *skeleton* of the tree. Let  $\mathcal{L}(\mathcal{T}_{(\infty)}^{\theta})$  denote the set of leaves of  $\mathcal{T}_{(\infty)}^{\theta}$ . In [13], it was shown that one can associate to  $\mathcal{T}_{(\infty)}^{\theta}$ , a natural probability measure called the **mass measure** satisfying  $\mu(\mathcal{L}(\mathcal{T}_{(\infty)}^{\theta})) = 1$ .

**Root-to-vertex path measures:** Now using the collection of uniform random variables above, we will define a function  $\mathfrak{G}_{(\infty)}$  on the tree as well as a collection of measures on paths emanating from the root. Recall that the hubs in  $\mathcal{T}_{(\infty)}^\theta$  have infinite degrees. Let  $(\mathcal{T}_j^{(i)} : j \geq 1)$  be the collection of subtrees of hub  $i$  in  $\mathcal{T}_{(\infty)}^\theta$  (labeled in some fashion). For each  $y \in \mathcal{T}_{(\infty)}^\theta$ , let

$$\mathfrak{G}_{(\infty)}(y) = \sum_{i \geq 1} \theta_i \left[ \sum_{j \geq 1} U_j^{(i)} \times \mathbb{1}\{y \in \mathcal{T}_j^{(i)}\} \right]. \quad (2.10)$$

We will show in our proof that  $\mathfrak{G}_{(\infty)}(y)$  is finite for almost every realization of  $\mathcal{T}_{(\infty)}^\theta$  and for  $\mu$ -almost every  $y \in \mathcal{T}_{(\infty)}^\theta$  (see Lemma 4.9 and Theorem 4.15). For  $y \in \mathcal{T}_{(\infty)}^\theta$ , let  $[\rho, y]$  denote the path from the root  $\rho$  to  $y$ . For every  $y$ , define a probability measure on  $[\rho, y]$  as

$$Q_y^{(\infty)}(v) := \frac{\theta_i U_j^{(i)}}{\mathfrak{G}_{(\infty)}(y)}, \text{ if } v \text{ is the } i\text{th hub and } y \in \mathcal{T}_j^{(i)} \text{ for some } j. \quad (2.11)$$

Thus, this probability measure is concentrated on the hubs on the path from  $y$  to the root.

**Remark 4.** Note that both  $\mathfrak{G}_{(\infty)}(\cdot)$  and  $Q_y^{(\infty)}(\cdot)$  depend on the realization of the pair  $(\mathcal{T}_{(\infty)}^\theta, \mathbf{U})$ , but we chose to suppress them to avoid cumbersome notation.

**Random tree  $\mathcal{R}_{IJ}^{(\infty)}$ :** Recall the tree  $r_{IJ}^{(\infty)}$  above. Recall that  $\eta_j$  is the vertex in the tree  $\mathcal{T}_{(\infty)}^\theta$  corresponding to leaf  $j+$  for  $1 \leq j \leq J$ . To each of these  $J$  leaves, associate the value  $\mathfrak{G}_{(\infty)}(\eta_j)$ , and associate the probability measure  $Q_{\eta_j}^{(\infty)}$  to the path  $[0+, j+]$ . This tree is a random element of the space  $\mathbf{T}_{IJ}^*$  (see Section 2.1.3), which we denote by  $\mathcal{R}_{IJ}^{(\infty)}$ .

**2.3. Continuum limits of components.** The aim of this section is to give an explicit description of the limiting (random) metric spaces in Theorem 1.7. We start by constructing a specific tilted version of the ICRT in Section 2.3.1. Then in Section 2.3.2 we describe the limits of maximal components.

**2.3.1. Tilted ICRTs and vertex identification.** Let  $(\Omega, \mathcal{F}, \mathbb{P}_\theta)$  and  $\mathcal{T}_{(\infty)}^\theta$  be as in Section 2.2 and let  $\gamma > 0$  a constant. Informally, the construction goes as follows: We will first tilt the distribution of the original ICRT  $\mathcal{T}_{(\infty)}^\theta$  using the functional

$$L_{(\infty)}(\mathcal{T}_{(\infty)}^\theta, \mathbf{U}) := \exp \left( \gamma \int_{y \in \mathcal{T}_{(\infty)}^\theta} \mathfrak{G}_{(\infty)}(y) \mu(dy) \right) \quad (2.12)$$

to get a tilted tree  $\mathcal{T}_{(\infty)}^{\theta, \star}$ . We then generate a random but finite number  $N_{(\infty)}^\star$  of pairs of points  $\{(x_k, y_k) : 1 \leq k \leq N_{(\infty)}^\star\}$ . The final metric space is obtained by creating “shortcuts” by identifying the points  $x_k$  and  $y_k$ . Formally the construction proceeds in four steps:

(a) **Tilted ICRT:** Define  $\mathbb{P}_\theta^\star$  on  $\Omega$  by

$$\frac{d\mathbb{P}_\theta^\star}{d\mathbb{P}_\theta} = \frac{\exp \left( \gamma \int_{y \in \mathcal{T}_{(\infty)}^\theta} \mathfrak{G}_{(\infty)}(y) \mu(dy) \right)}{\mathbb{E} \left[ \exp \left( \gamma \int_{x \in \mathcal{T}_{(\infty)}^\theta} \mathfrak{G}_{(\infty)}(x) \mu(dx) \right) \right]}.$$

The expectation in the denominator is with respect to the original measure  $\mathbb{P}_\theta$ . In our proof we will show that this object is finite. Write  $(\mathcal{T}_{(\infty)}^{\theta, \star}, \mu^\star)$  and  $\mathbf{U}^\star = (U_j^{(i), \star} : i, j \geq 1)$  for the tree and the mass measure on it, and the associated random variables under this change of measure.



- (b) **Poisson number of identification points:** Conditional on  $((\mathcal{T}_{(\infty)}^{\theta, \star}, \mu^\star), \mathbf{U}^\star)$ , generate  $N_{(\infty)}^\star$  having a  $\text{Poisson}(\Lambda_{(\infty)}^\star)$  distribution, where

$$\Lambda_{(\infty)}^\star := \gamma \int_{y \in \mathcal{T}_{(\infty)}^{\theta, \star}} \mathfrak{G}_{(\infty)}(y) \mu^\star(dy) = \gamma \sum_{i \geq 1} \theta_i \left[ \sum_{j \geq 1} U_j^{(i), \star} \mu^\star(\mathcal{T}_j^{(i), \star}) \right].$$

Here,  $(\mathcal{T}_j^{(i), \star} : j \geq 1)$  denotes the collection of subtrees of hub  $i$  in  $\mathcal{T}_{(\infty)}^{\theta, \star}$ . (As mentioned before in Remark 4,  $\mathfrak{G}_{(\infty)}(\cdot)$  depends on the realization of the ordered ICRT.  $U_j^{(i), \star}$  appears in the expression above as the function  $\mathfrak{G}_{(\infty)}$  acts on  $y \in \mathcal{T}_{(\infty)}^{\theta, \star}$  for which the associated order is described by  $\mathbf{U}^\star$ .)

- (c) **"First" endpoints (of shortcuts):** Conditional on (a) and (b), sample  $x_k$  from  $\mathcal{T}_{(\infty)}^{\theta, \star}$  with density proportional to  $\mathfrak{G}_{(\infty)}(x) \mu^\star(dx)$  for  $1 \leq k \leq N_{(\infty)}^\star$ .
- (d) **"Second" endpoints (of shortcuts) and identification:** Having chosen  $x_k$ , choose  $y_k$  from the path  $[\rho, x_k]$  joining the root  $\rho$  and  $x_k$  according to the probability measure  $Q_{x_k}^{(\infty)}$  as in (2.11) but with  $U_j^{(i), \star}$  replacing  $U_j^{(i)}$ . (Note that  $y_k$  is always a **hub** on  $[\rho, x_k]$ .) Identify  $x_k$  and  $y_k$ , i.e., form the quotient space by introducing the equivalence relation  $x_k \sim y_k$  for  $1 \leq k \leq N_{(\infty)}^\star$ .

**Definition 2.2.** Fix  $\gamma \geq 0$  and  $\theta \in \Theta$  as in (2.9). Let  $\mathcal{G}_\infty(\theta, \gamma)$  be the metric measure space constructed via the four steps above equipped with the measure inherited from the mass measure on  $\mathcal{T}_{(\infty)}^{\theta, \star}$ .

In our proofs, we will always think of the leaf end (of a shortcut or a surplus edge) as the first endpoint, and the second endpoint will be selected from the skeleton.

**2.3.2. Limits of the components.** Fix  $\lambda \in \mathbb{R}$  and  $\mathbf{c} \in l_0$  as in (1.13) and consider the setting of Theorem 1.7. We will need 2 main objects:

- (a) The process  $\tilde{V}_\lambda^{\mathbf{c}}(\cdot)$  in (1.19). Recall that the excursions of this process from zero could be arranged in increasing order of lengths as  $\mathcal{Z}(\lambda)$ . Let  $\Xi^{(i)} = (c_j : \xi_j \in \mathcal{Z}_i)$  denote the point process of jumps of the process  $\tilde{V}_\lambda^{\mathbf{c}}(\cdot)$  corresponding to the excursion  $\mathcal{Z}_i(\lambda)$ . Abusing notation we will write  $\Xi^{(i)} = (c_j : j \in \mathcal{Z}_i)$ .
- (b) The actual lengths of these excursions  $(Z_i(\lambda) : i \geq 1)$  as in (1.21).

From these objects, for each fixed  $i \geq 1$ , define the random variable  $\tilde{\gamma}^{(i)}$  and the point process  $\theta^{(i)} = (\theta_j^{(i)} : j \in \mathcal{Z}_i(\lambda))$  as

$$\tilde{\gamma}^{(i)} := Z_i(\lambda) \sqrt{\sum_{v \in \mathcal{Z}_i(\lambda)} c_v^2}, \quad \theta^{(i)} := \left( \frac{c_j}{\sqrt{\sum_{v \in \mathcal{Z}_i(\lambda)} c_v^2}} : j \in \mathcal{Z}_i(\lambda) \right). \quad (2.13)$$

Our proof (see Proposition 5.1) will imply that  $\theta^{(i)} \in \Theta$  as in (2.9) a.s. Define

$$\Gamma_i(\lambda) := Z_i(\lambda) \left( \sum_{v \in \mathcal{Z}_i(\lambda)} c_v^2 \right)^{-1/2}, \quad (2.14)$$

and generate the random metric measure spaces

$$M_i^{\mathbf{c}}(\lambda) := \Gamma_i(\lambda) \cdot \mathcal{G}_\infty(\theta^{(i)}, \tilde{\gamma}^{(i)}), \quad (2.15)$$

where  $\mathcal{G}_\infty(\boldsymbol{\theta}, \tilde{\gamma})$  is as described in Section 2.3.1 and the metric spaces are *conditionally independent* across  $i$  given the driving parameters in (2.13). Let  $\mathbf{M}_\infty^{\mathbf{c}}(\lambda) = (M_i^{\mathbf{c}}(\lambda) : i \geq 1)$ . Then this is the limiting collection of metric spaces in Theorem 1.7.

To describe the sequence of spaces  $\mathbf{M}_\infty^{\text{nr}}(\lambda)$  appearing in Theorem 1.1, define

$$\mathbf{c}^{\text{nr}} := (c_j^{\text{nr}} : j \geq 1), \text{ where } c_j^{\text{nr}} = \frac{1}{\mathbb{E} W} \left( \frac{c_F}{j} \right)^{1/(\tau-1)}, \quad (2.16)$$

$$\zeta := - \left( \frac{c_F^{2/(\tau-1)}}{\mathbb{E} W} \right) \sum_{i=1}^{\infty} \left[ \int_{i-1}^i \frac{du}{u^{2/(\tau-1)}} - \frac{1}{i^{2/(\tau-1)}} \right], \text{ and } t_\lambda^{\text{nr}} := \frac{(\lambda + \zeta)}{\mathbb{E} W}. \quad (2.17)$$

Here  $W$  is a random variable with distribution  $F$  as in (1.3). Then

$$\mathbf{M}_\infty^{\text{nr}}(\lambda) = \frac{1}{\mathbb{E} W} \cdot \mathbf{M}_\infty^{\mathbf{c}}(t_\lambda^{\text{nr}}). \quad (2.18)$$

### 3. DISCUSSION

We describe the two major motivations for developing the general theory of this paper in Sections 3.1 and 3.2. In Sections 3.3 and 3.4, we include a very brief discussion about ICRTs as well as give an overview of the order in which the proofs are carried out.

**3.1. Universality and domains of attraction of critical random graph models.** One natural question the reader might ask at this point is why the general theory in Section 1.2, why not just stick to the rank one random graph model as in Section 1.1. As we described in the introduction, the aim of this paper is the development of general theory applicable to a wide array of models. What does one mean by this? It turns out that many different random graph models can be constructed in a dynamic fashion as graph valued process  $\{\mathcal{G}_n(t) : t \geq 0\}$  where edges are added as time advances thus resulting in mergers of components as  $t \uparrow t_c$ . In this construction, there is a critical time  $t_c$  (model dependent) such that the giant component emerges after time  $t_c$ .

Now for most random graph models (including the configuration model) the dynamics of mergers of components starting at time zero **do not** look like the multiplicative coalescent. However if one were to zoom in at the critical time  $t_c$ , for many models,  $\exists \varepsilon_n \downarrow 0$  such that if one were to look at the interval  $[t_c - \varepsilon_n, t_c + \varepsilon_n]$ , then mergers of components **can** be approximated by the multiplicative coalescent. Here  $t_c - \varepsilon_n$  often corresponds to the *barely subcritical* regime of the random graph. Thus if one had good control over component functionals at the barely subcritical time  $t_c - \varepsilon_n$  and in particular if one was able to show that component sizes appropriately normalized satisfied the assumptions 1.5, then one can use Theorem 1.7 to derive convergence at the critical time  $t_c$  of the maximal components. Note that one does not expect component sizes at time  $t_c - \varepsilon_n$  to satisfy assumptions of the Norros-Reittu model in (1.4). Rather in most cases, at time  $t_c - \varepsilon_n$ , the size of the component of a randomly selected vertex  $V_n$  would scale like  $n^{\delta_1}$  while the maximal component would scale like  $n^{\delta_2}$  (ignoring logarithmic corrections) where  $\delta_1 < \delta_2$  are related to various scaling exponents of the system. In work in progress [19], Theorem 1.8 coupled with delicate estimates of various scaling exponents for the configuration model in the barely subcritical regime, proves analogous results for the configuration model with degree exponent  $\tau \in (3, 4)$ . Sizes of maximal components in the critical regime including the heavy tailed regime for this model was previously analyzed in [48]. Further as was done in [18], where a number of sufficient conditions for the domain of attraction of the critical Erdős-Rényi scaling

limits were derived, we hope to derive similar general conditions for a random graph model to belong to the same domain of attraction as the rank-one model, established in this paper.

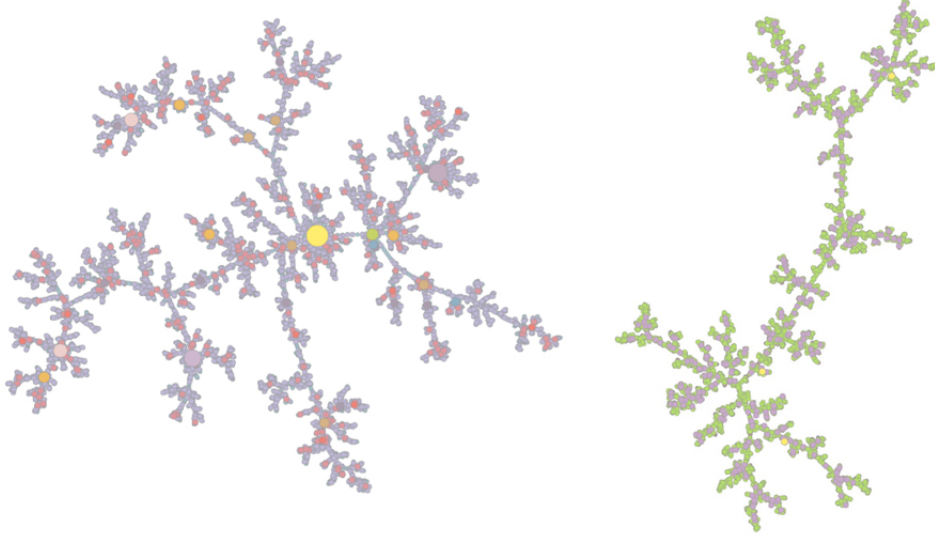


FIGURE 3.1. On the left, an approximation of an ICRT (using **p**-trees on approximately 20000 vertices) corresponding to  $\theta_i \propto i^{-1/(\tau-2)}$  where  $\tau = 3.01$ . The reason behind this choice of  $\theta_i$  is explained in Section 8. On the right, an approximation of a Brownian CRT (using a uniform random tree on the same number of vertices). Vertex sizes are proportional to the degree of the vertex.

**3.2. Minimal spanning tree on inhomogeneous random graphs.** As described in the introduction, a second major motivation for the technical analysis in this paper is the minimal spanning tree. To fix ideas, consider the Norros-Reittu model in the supercritical regime (the parameter in (1.5)  $\nu > 1$ ). To each edge attach a random edge weight *iid* across edges, assumed to be derived from a continuous distribution. Consider the minimal spanning tree (MST) of the giant component. A large amount of simulation based evidence from statistical physics [23, 24, 28, 63] suggests that when the degree exponent  $\tau \in (3, 4)$  then the distances in this object scale like  $n^{(\tau-3)/(\tau-1)}$ , the same distance scaling shown in this paper for the maximal components in the *critical* regime (Theorem 1.1).

Now this is not a coincidence. As has been shown in a series of fundamental papers [3–5] for the complete graph and the supercritical Erdős-Rényi random graph, a major component in the analysis of the MST problem is first the scaling of maximal components in the critical regime which then provides crucial input for the scaling limit of the MST. Till date we have no rigorous results for the scaling of the MST on any “inhomogeneous” random graph model. This paper provides the first step in answering this question in the heavy tailed regime. Further this program should enable one to analyze the MST for random graph models other than the rank one model which belong to the same “domain of attraction” in the critical regime.

**3.3. Inhomogeneous continuum random trees.** As evident from Section 2.2, ICRTs play a major role in the description of the limit objects. Despite a lot of work on these objects in the last decade [11, 13, 27], a number of questions regarding these continuum objects are still open,

ranging from sufficient conditions for compactness to the dependence of the fractal properties of this object on the driving parameter  $\theta$ . Our proof shows that in some special cases, ICRTs are compact metric spaces when  $\theta$  is sampled according to an appropriate size-biased distribution. This can be seen as an *annealed* result on compactness of the ICRT. Whether compactness is true for non-random sequences  $\theta \in \Theta$  has been open problem for more than a decade [11]. Similar questions hold for its fractal dimensions. See Section 8 for a more detailed account of these problems.

**3.4. Overview of the proof.** In Section 4, we study the random graph  $\mathcal{G}_n(\mathbf{x}, t)$  as in Definition 1.3. We start with the simple observation that conditional on the vertex set of components of  $\mathcal{G}_n(\mathbf{x}, t)$ , a fixed component  $\mathcal{C}$  has the same distribution as  $\mathcal{G}_n(\mathbf{x}, t)$  *conditional* on being connected. This section studies asymptotics for such distributions assuming specific regularity properties of vertex weights in the component in the large network limit, showing Gromov-weak convergence of the associated graph under proper normalization of edge lengths and vertex weights. Section 5 uses the size-biased exploration of the process  $\mathcal{G}_n(\mathbf{x}, t)$  [9] to show that maximal connected components satisfy the hypothesis required in Section 4. Section 6 studies the special entrance boundary in (1.22) proving both compactness of the limiting objects as well as strengthening the convergence in the Gromov-weak topology to convergence in  $d_{\text{GHP}}$ . In Section 7, we derive the upper bound on the box-counting dimension. In Section 8, we conclude by describing a number of open problems.

#### 4. PROOFS: ASYMPTOTICS CONDITIONAL ON BEING CONNECTED

The aim of this Section is to study large connected components of  $\mathcal{G}_n(\mathbf{x}, t)$  assuming vertex weights satisfy a few regularity properties.

**4.1. Tilted  $\mathbf{p}$ -trees and connected components of  $\mathcal{G}(\mathbf{x}, t)$ .** Recall the random graph  $\mathcal{G}(\mathbf{x}, t)$  from Definition 1.3. Here for any  $t \geq 0$ ,  $(\mathcal{C}_i(t) : i \geq 1)$  denoted the components in decreasing order of their mass sizes. The aim of this section is to describe results from [20] which gave a method of constructing **connected** components of  $\mathcal{G}(\mathbf{x}, t)$  conditional on the vertices of the components. This construction involved tilted versions of  $\mathbf{p}$ -trees introduced in Section 2.2. Since these trees are parametrized via a driving pmf  $\mathbf{p}$ , it will be easy to parametrize various random graph constructions in terms of pmfs as opposed to vertex weights  $\mathbf{x}$ . Proposition 4.1 will relate vertex weights to pmfs.

Fix  $n \geq 1$  and  $\mathcal{V} \subset [n]$  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  $\mathbf{p} = (p_v : v \in \mathcal{V})$ , define probability distributions  $\mathbb{P}_{\text{con}}(\cdot; \mathbf{p}, a, \mathcal{V})$  on  $\mathbb{G}_{\mathcal{V}}^{\text{con}}$  as follows: Define for  $i, j \in \mathcal{V}$ ,

$$q_{ij} := 1 - \exp(-ap_i p_j). \quad (4.1)$$

Then

$$\mathbb{P}_{\text{con}}(G; \mathbf{p}, a, \mathcal{V}) := \frac{1}{Z(\mathbf{p}, a)} \prod_{(i,j) \in E(G)} q_{ij} \prod_{(i,j) \notin E(G)} (1 - q_{ij}), \text{ for } G \in \mathbb{G}_{\mathcal{V}}^{\text{con}}, \quad (4.2)$$

where  $Z(\mathbf{p}, a)$  is the normalizing constant

$$Z(\mathbf{p}, a) := \sum_{G \in \mathbb{G}_{\mathcal{V}}^{\text{con}}} \prod_{(i,j) \in E(G)} q_{ij} \prod_{(i,j) \notin E(G)} (1 - q_{ij}).$$

Now let  $\mathcal{V}^{(i)} := V(\mathcal{C}_i(t))$  be the vertex set of  $\mathcal{C}_i(t)$  for  $i \geq 1$  and note that  $\{\mathcal{V}^{(i)} : i \geq 1\}$  denotes a random finite partition of the full vertex set  $[n]$ . The following result is obvious from the construction of  $\mathcal{G}(\mathbf{x}, t)$ :

**Proposition 4.1.** [20, Proposition 6.1] *Conditional on the partition  $\{\mathcal{V}^{(i)} : i \geq 1\}$  define*

$$\mathbf{p}_n^{(i)} := \left( \frac{x_v}{\sum_{v \in \mathcal{V}^{(i)}} x_v} : v \in \mathcal{V}^{(i)} \right), \quad a_n^{(i)} := t \left( \sum_{v \in \mathcal{V}^{(i)}} x_v \right)^2, \quad i \geq 1.$$

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

$$\mathbb{P}(\mathcal{G}_i(t) = G_i, \forall i \geq 1 \mid \{\mathcal{V}^{(i)} : i \geq 1\}) = \prod_{i \geq 1} \mathbb{P}_{\text{con}}(G_i; \mathbf{p}_n^{(i)}, a_n^{(i)}, \mathcal{V}^{(i)}).$$

Thus the random graph  $\mathcal{G}(\mathbf{x}, t)$  can be generated in two stages:

- (i) **Stage I:** Generate the partition of the vertices into different components, i.e., generate  $\{\mathcal{V}^{(i)} : i \geq 1\}$ .
- (ii) **Stage II:** Conditional on the partition, generate the internal structure of each component following the law of  $\mathbb{P}_{\text{con}}(\cdot; \mathbf{p}^{(i)}, a^{(i)}, \mathcal{V}^{(i)})$ , independently across different components.

Let us now describe an algorithm to generate such **connected** components using distribution (4.2). To ease notation, let  $\mathcal{V} = [m]$  for some  $m \geq 1$  and fix a probability mass function  $\mathbf{p}$  on  $[m]$  and a constant  $a > 0$  and write  $\mathbb{P}_{\text{con}}(\cdot) := \mathbb{P}_{\text{con}}(\cdot; \mathbf{p}, a, [m])$  on  $\mathbb{G}_m^{\text{con}} := \mathbb{G}_{[m]}^{\text{con}}$ . We will first need to set up some notation before describing this result.

**Depth-first exploration of ordered trees.** Recall that we used  $\mathbb{T}_m^{\text{ord}}$  for the space of ordered (or planar) trees with vertex set  $[m]$ . Given a tree  $\mathbf{t} \in \mathbb{T}_m^{\text{ord}}$ , one can use the associated order to explore the tree in a depth-first manner. More precisely we start with  $v(1)$  being the root of  $\mathbf{t}$ . At each stage  $1 \leq i \leq m$ , we will keep track of three types of vertices: the set of *active* vertices— $\mathcal{A}(i)$ , the set of *explored* vertices— $\mathcal{O}(i)$ , and the set of *unexplored* vertices— $\mathcal{U}(i)$ . The set of active vertices will in fact be viewed as a vertical stack (not just a set) with  $\mathcal{A}(i)$  representing the state of this stack at the end of step  $\mathcal{A}(i)$ . Initialize the process with  $\mathcal{A}(1) = \{v(1)\}$  (the root of  $\mathbf{t}$ ),  $\mathcal{O}(1) = \emptyset$  and  $\mathcal{U}(1) = [m] \setminus \{v(1)\}$ . At step  $i \geq 1$ , we let

- (i)  $v(i)$  denote the vertex at the top of the stack  $\mathcal{A}(i)$  and let  $\mathcal{D}(i) \subset \mathcal{U}(i)$  denote the set of children of  $v(i)$ . Delete  $v(i)$  from  $\mathcal{A}(i)$  and arrange the vertices of  $\mathcal{D}(i)$  from oldest to youngest at the *top* of the stack to form  $\mathcal{A}(i+1)$ ;
- (ii)  $\mathcal{O}(i+1) = \mathcal{O}(i) \cup \{v(i)\}$ ;
- (iii)  $\mathcal{U}(i+1) = \mathcal{U}(i) \setminus \mathcal{D}(i)$ .

Write  $\mathfrak{P}(\mathbf{t})$  for set of pairs of vertices  $\{u, v\}$  such that  $u, v \in \mathcal{A}(i)$  for some  $1 \leq i \leq m$ ; namely both vertices are active but have not yet been explored. Using terminology from [4], call this collection *the set of permitted edges*. Thus,

$$\mathfrak{P}(\mathbf{t}) := \{(v(i), u) \mid 2 \leq i \leq m, u \in \mathcal{A}(i-1) \setminus \{v(i)\}\}. \quad (4.3)$$

Write  $E(\mathbf{t})$  for the edge set of  $\mathbf{t}$ . Now define the function  $L : \mathbb{T}_m^{\text{ord}} \rightarrow \mathbb{R}_+$  by

$$L(\mathbf{t}) = L_{(m)}(\mathbf{t}) := \prod_{(k, \ell) \in E(\mathbf{t})} \left[ \frac{\exp(ap_k p_\ell) - 1}{ap_k p_\ell} \right] \exp \left( \sum_{(k, \ell) \in \mathfrak{P}(\mathbf{t})} ap_k p_\ell \right), \quad \mathbf{t} \in \mathbb{T}_m^{\text{ord}}. \quad (4.4)$$

Recall the (ordered)  $\mathbf{p}$ -tree distribution from (2.8). Using  $L(\cdot)$  to tilt this distribution results in the distribution

$$\mathbb{P}_{\text{ord}}^\star(\mathbf{t}) := \mathbb{P}_{\text{ord}}(\mathbf{t}) \cdot \frac{L(\mathbf{t})}{\mathbb{E}_{\text{ord}}[L(\mathcal{T}_m^\mathbf{p})]}, \quad \mathbf{t} \in \mathbb{T}_m^{\text{ord}}. \quad (4.5)$$

For future reference we fix notation for the various objects required in the proof below.

**Definition 4.2.** Fix  $m \geq 1$ ,  $a > 0$ , and a probability mass function  $\mathbf{p}$  on  $[m]$ . We will write  $\tilde{\mathcal{G}}_m(\mathbf{p}, a)$  to denote a random graph with distribution  $\mathbb{P}_{\text{con}}(\cdot, \mathbf{p}, a, [m])$ .  $\mathcal{T}_m^{\mathbf{p}, \star}$  will denote a random planar tree with the tilted  $\mathbf{p}$ -tree distribution (4.5), and  $\mathcal{T}_m^{\mathbf{p}}$  will denote a random tree with the original  $\mathbf{p}$  tree distribution (2.8).

**Proposition 4.3** ([20, Proposition 7.4]). Fix  $m \geq 1$ , a probability mass function  $\mathbf{p}$  on  $[m]$ , and  $a > 0$ . Consider a random connected graph on  $[m]$  constructed as follows:

- (a) First generate a rooted planar random tree  $\mathcal{T}_m^{\mathbf{p}, \star}$  with distribution  $\tilde{\mathbb{P}}_{\text{ord}}(\cdot)$  as in (4.5).
- (b) Let  $\mathfrak{P}(\mathcal{T}_m^{\mathbf{p}, \star})$  denote the permitted edge set of this random tree. Add each such edge  $\{u, v\} \in \mathfrak{P}(\mathcal{T}_m^{\mathbf{p}, \star})$  with probability  $q_{uv}$  as in (4.1), independent across permitted edges.

Then, the resulting random graph has distribution  $\mathbb{P}_{\text{con}}$  on  $\mathbb{G}_m^{\text{con}}$ , i.e, has the same distribution as  $\tilde{\mathcal{G}}_m(\mathbf{p}, a)$ .

**4.2. Convergence of connected components under weight assumptions.** The aim of this section is to prove Gromov-weak convergence for the connected graph  $\tilde{\mathcal{G}}_m(\mathbf{p}, a)$  under regularity conditions on  $a$  and  $\mathbf{p}$  as  $m \rightarrow \infty$ . We will assume that we have ordered the index set  $[m]$  so that  $p_1 \geq p_2 \geq \dots \geq p_m > 0$ . Let

$$\sigma(\mathbf{p}) := \sqrt{\sum_i p_i^2}. \quad (4.6)$$

**Assumption 4.4.** As  $m \rightarrow \infty$ , the following hold:

- (i)  $\sigma(\mathbf{p}) \rightarrow 0$  and further for each fixed  $i \geq 1$ ,  $p_i/\sigma(\mathbf{p}) \rightarrow \theta_i$  where  $\boldsymbol{\theta} := (\theta_1, \theta_2, \dots)$  is an element of  $\Theta$  as in (2.9).
- (ii) There is a constant  $\gamma > 0$  such that  $a\sigma(\mathbf{p}) \rightarrow \gamma$ .

The following theorem is the main result of this section.

**Theorem 4.5.** Consider the connected random graph  $\tilde{\mathcal{G}}_m(\mathbf{p}, a)$  viewed as a metric measure space via the graph distance where each vertex  $v$  is assigned measure  $p_v$ . Under Assumption 4.4,

$$\sigma(\mathbf{p})\tilde{\mathcal{G}}_m(\mathbf{p}, a) \xrightarrow{w} \mathcal{G}_\infty(\boldsymbol{\theta}, \gamma),$$

where  $\mathcal{G}_\infty(\boldsymbol{\theta}, \gamma)$  is the random metric space defined in Definition 2.2 and convergence is in the Gromov-weak topology on metric spaces.

The rest of this section proves this result. We will throughout assume that  $\tilde{\mathcal{G}}_m(\mathbf{p}, a)$  has been constructed using Proposition 4.3.

**4.2.1. Two constructions of  $\mathbf{p}$ -trees: Exploration process and the birthday construction.** We start by describing an explicit construction of the (untilted)  $\mathbf{p}$ -tree  $\mathcal{T}_m^{\mathbf{p}}$  first developed in [11]. At the end of this section we describe a second construction used later in the paper.

**Exploration process construction:** The first construction is initiated by setting up a map  $\psi_{\mathbf{p}} : [0, 1]^m \rightarrow \mathbb{T}^{\text{ord}}$  as follows. Let  $\mathbf{u} := (u_v : v \in [m])$  be a collection of distinct points in  $(0, 1)$ . Define

$$F^{\mathbf{p}}(s) := -s + \sum_{v=1}^m p_v \mathbb{1}\{u_v \leq s\}, \quad s \in [0, 1]. \quad (4.7)$$

Assume that there exists a unique point  $v^* \in [m]$  such that  $F^{\mathbf{p}}(u_{v^*} -) = \min_{u \in [0, 1]} F^{\mathbf{p}}(s)$ . Set  $v^*$  to be the root of the tree  $\psi_{\mathbf{p}}(\mathbf{u})$ . Define  $y_i := u_i - u_{v^*}$  for  $i \in [m]$ , and

$$F^{\text{exc}, \mathbf{p}}(s) := F^{\mathbf{p}}(u_{v^*} + s \bmod 1) - F^{\mathbf{p}}(u_{v^*} -), \quad 0 \leq s < 1.$$

Then  $F^{\text{exc}, \mathbf{p}}(1-) = 0$  and  $F^{\text{exc}, \mathbf{p}}(s) > 0$  for  $s \in [0, 1)$ . Extend the definition of  $F^{\text{exc}, \mathbf{p}}$  to  $s \in [0, 1]$  by define  $F^{\text{exc}, \mathbf{p}}(1) = 0$ . We use  $F^{\text{exc}, \mathbf{p}}$  to construct a depth-first-search of an ordered tree whose exploration in this depth-first manner is encoded by the function  $F^{\text{exc}, \mathbf{p}}$ . This in turn defines the tree  $\psi_{\mathbf{p}}(\mathbf{u})$ . As before, in this construction we carry along a set of explored vertices  $\mathcal{O}(i)$ , active vertices  $\mathcal{A}(i)$  and unexplored vertices  $\mathcal{U}(i) = [m] \setminus (\mathcal{A}(i) \cup \mathcal{O}(i))$ , for  $0 \leq i \leq m$ . We view  $\mathcal{A}(i)$  as the state of a vertical stack  $\mathcal{A}$  after the  $i$ th step in the depth-first-search. Initialize with  $\mathcal{O}(0) = \emptyset$ ,  $\mathcal{A}(0) = \{v^*\}$ ,  $\mathcal{U}(0) = [m] \setminus \{v(1)\}$ , and define  $y^*(0) = 0$ . At step  $i \in [m]$ , let  $v(i)$  be the value that is on the top of the stack  $\mathcal{A}(i-1)$  and define  $y^*(i) := y^*(i-1) + p_{v(i)}$ . Define  $\mathcal{D}(i) := \{i \in [m] : y^*(i-1) < y_i < y^*(i)\}$ . Suppose  $\mathcal{D}(i) = \{u(j) : 1 \leq j \leq k\}$  where we have ordered these vertices in the sequence that they are found in this interval, i.e.,

$$y^*(i-1) < y_{u(1)} < \dots < y_{u(k)} < y^*(i).$$

Update the stack  $\mathcal{A}$  as follows:

- (i) Delete  $v(i)$  from  $\mathcal{A}$ .
- (ii) Push  $u(j)$ ,  $1 \leq j \leq k$ , to the top of  $\mathcal{A}$  sequentially (so that  $u(k)$  will be on the **top** of the stack at the end).

Let  $\mathcal{A}(i)$  be the state of the stack after the above operations. Update  $\mathcal{O}(i) := \mathcal{O}(i-1) \cup \{v(i)\}$  and  $\mathcal{U} := \mathcal{U}(i-1) \setminus \mathcal{D}(i)$ . See Figure 4.1 for a pictorial description of this construction.

The tree  $\psi_{\mathbf{p}}(\mathbf{u}) \in \mathbb{T}_m^{\text{ord}}$  is constructed by putting the edges  $\{(v(i), v) : i \in [m], v \in \mathcal{D}(i)\}$  and using the order prescribed in the above exploration to make the tree an ordered tree. The fact that this procedure actually produces a tree is proved in [11].

**Lemma 4.6** ([11, Section 3.2]). *Consider the map  $\psi_{\mathbf{p}}$ . Let  $\mathbf{X} := (X_v : v \in [m])$  be i.i.d. random variables distributed uniformly on  $(0, 1)$ . Then the random tree  $\psi_{\mathbf{p}}(\mathbf{X})$  has distribution (2.8), i.e.,  $\psi_{\mathbf{p}}(\mathbf{X}) \stackrel{d}{=} \mathcal{T}^{\mathbf{p}}$ .*

For future reference, coupled with the above construction, define  $\mathcal{S}(i) := \mathcal{A}(i-1) \setminus \{v(i)\}$  for  $i \in [m]$ . Define the function  $A_m(\cdot)$  on  $[0, 1]$  via

$$A_m(u) := \sum_{v \in \mathcal{S}(i)} p_v, \quad \text{for } u \in (y^*(i-1), y^*(i)], i \in [m]. \quad (4.8)$$

Further let  $\bar{A}_m(u) := aA_m(u)$ ,  $u \in [0, 1]$ , where  $a$  is the scaling constant in (4.1).

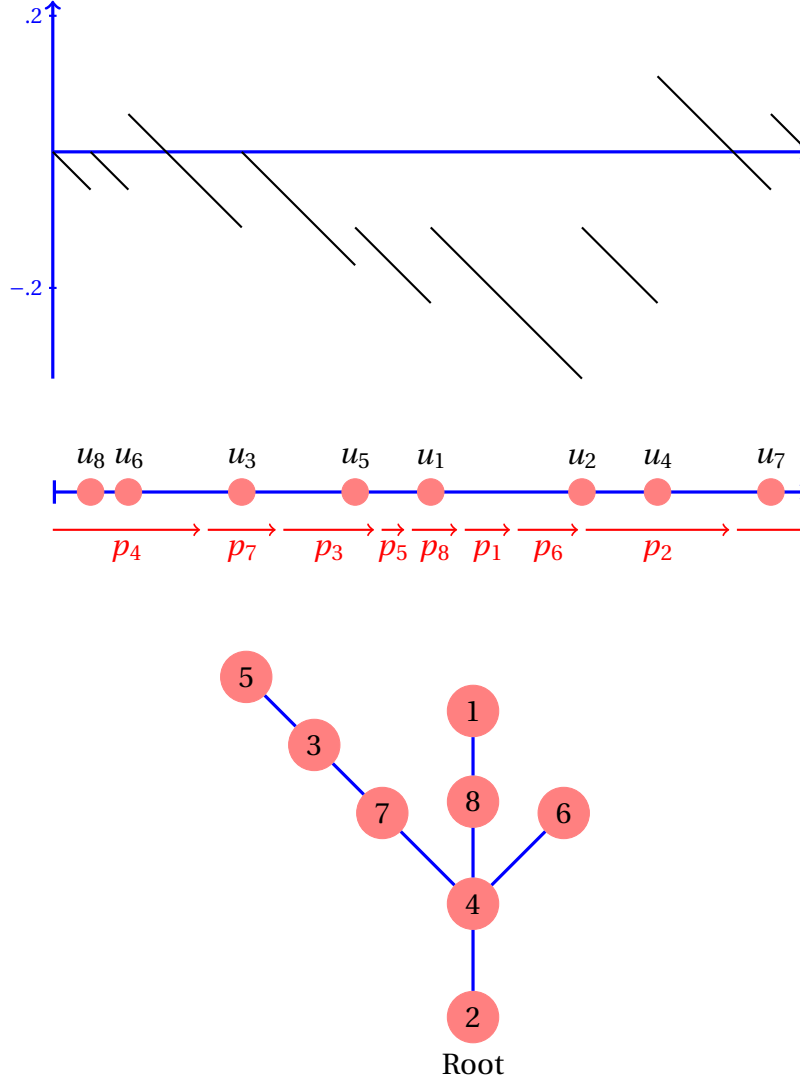
**Birthday construction:** We now describe a second construction of  $\mathbf{p}$ -trees, first formulated in [27]. We urge the reader to skim this portion and return to it once she has reached Section 4.5. Let  $\mathbf{Y} := (Y_0, Y_1, \dots)$  be an infinite sequence of i.i.d. random variables with distribution  $\mathbf{p}$ . Let  $R_0 = 0$  and for  $l \geq 1$ , let  $R_l$  denote the  $l$ -th repeat time, i.e.,

$$R_l = \min \left\{ k > R_{l-1} : Y_k \in \{Y_0, \dots, Y_{k-1}\} \right\}.$$

Now consider the directed graph formed via the edges

$$\mathcal{T}(\mathbf{Y}) := \{(Y_{j-1}, Y_j) : Y_j \notin \{Y_0, \dots, Y_{j-1}\}, j \geq 1\}.$$

It is easy to check that this gives a tree which we view as rooted at  $Y_0$ . Intuitively the process of constructing a tree is as follows: the tree “grows” via the addition of new vertices sampled using  $\mathbf{p}$  till it stumbles across a “repeat” (a vertex already found) when it goes back to the first occurrence of this “repeat” and starts growing from that position. The following striking result was shown in [27].

FIGURE 4.1. The function  $F^{\mathbf{p}}$  and the corresponding tree  $\psi_{\mathbf{p}}$ .

**Theorem 4.7** ([27, Lemma 1 and Theorem 2]). *The random tree  $\mathcal{T}(\mathbf{Y})$  viewed as an object in  $\mathbb{T}_m$  is distributed as a  $\mathbf{p}$ -tree with distribution (2.7) **independently** of  $Y_{R_1-1}, Y_{R_2-1}, \dots$  which are i.i.d with distribution  $\mathbf{p}$ .*

**Remark 5.** The independence between the sequence  $Y_{R_1-1}, Y_{R_2-1}, \dots$  and the constructed  $\mathbf{p}$  tree  $\mathcal{T}(\mathbf{Y})$  is truly remarkable. In particular, suppose  $\mathcal{S}$  is a  $\mathbf{p}$ -tree with distribution as in (2.7) and for fixed  $r \geq 1$ , let  $\tilde{Y}_1, \tilde{Y}_2, \dots, \tilde{Y}_r$  be i.i.d. with distribution  $\mathbf{p}$ . Write  $\mathcal{S}_r \subset \mathcal{S}$  for the tree spanned by these vertices and the root. Let  $\mathcal{T}_r^{\mathcal{B}} \subset \mathcal{T}(\mathbf{Y})$  denote the subtree with vertex set  $\{Y_0, Y_1, \dots, Y_{R_r-1}\}$ , namely the tree constructed in the first  $R_r$  steps. Here  $\mathcal{B}$  is a mnemonic for “birthday tree” and also to distinguish this construction from a generic random tree model with  $r$  vertices. Then the above result (formalized as [27, Corollary 3]) implies that these can be jointly constructed as

$$(\tilde{Y}_1, \tilde{Y}_2, \dots, \tilde{Y}_r; \mathcal{S}_r) \stackrel{d}{=} (Y_{R_1-1}, Y_{R_2-1}, \dots, Y_{R_r-1}; \mathcal{T}_r^{\mathcal{B}}). \quad (4.9)$$

We use this fact often in Section 4.5.



**4.3. Uniform integrability of the tilt.** The first use of the above construction of the  $\mathbf{p}$ -tree is to prove the following:

**Proposition 4.8.** *Fix  $s \geq 1$  and consider the tilt  $L(\cdot)$  as in (4.4). Under Assumptions 4.4, there is a constant  $K := K(s) < \infty$  such that*

$$\sup_{m \geq 1} \mathbb{E}_{\text{ord}} \left( [L(\mathcal{T}_m^{\mathbf{p}})]^s \right) \leq K.$$

*In particular, the collection of random variables  $\{L(\mathcal{T}_m^{\mathbf{p}}) : m \geq 1\}$  is uniformly integrable.*

**Proof:** Writing out the tilt  $L(\cdot)$  explicitly we have

$$L(\mathbf{t}) := \prod_{(k, \ell) \in E(\mathbf{t})} \left[ \frac{\exp(ap_k p_\ell) - 1}{ap_k p_\ell} \right] \exp \left( \sum_{(k, \ell) \in \mathfrak{P}(\mathbf{t})} ap_k p_\ell \right) = \mathbb{I}(\mathbf{t}) \bar{L}(\mathbf{t}), \quad (4.10)$$

say, where,

$$\mathbb{I}(\mathbf{t}) := \prod_{(i, j) \in E(\mathbf{t})} \frac{\exp(ap_i p_j) - 1}{ap_i p_j} \leq \exp \left( a \sum_{(i, j) \in E(\mathbf{t})} p_i p_j \right) \leq \exp(ap_{\max}). \quad (4.11)$$

Here we have used  $(e^x - 1)/x \leq e^x$  for  $x > 0$  for the first inequality and the second inequality follows using the fact that  $\mathbf{t}$  is a tree, so that for each  $(i, j) \in E(\mathbf{t})$  such that  $i$  is the parent of  $j$ , we have  $p_i p_j \leq p_{\max} p_j$ . By Assumption 4.4, we have  $ap_{\max} \rightarrow \gamma \theta_1$ . In particular, there is a constant  $C > 0$  such that for all  $m \geq 1$ , and  $\mathbf{t} \in \mathbb{T}_m^{\text{ord}}$ ,

$$\mathbb{I}(\mathbf{t}) \leq C \text{ and } L(\mathbf{t}) \leq C \exp \left( \sum_{(k, \ell) \in \mathfrak{P}(\mathbf{t})} ap_k p_\ell \right). \quad (4.12)$$

Now recall the functions  $A_m$  and  $\bar{A}_m := aA_m$  from (4.8). Using the equivalent characterization of the permitted edge set from (4.3) and comparing this with (4.8), it is easy to check that

$$\sum_{(i, j) \in \mathfrak{P}(\mathcal{T}_m^{\mathbf{p}})} ap_i p_j = a \sum_{i \in [m]} \sum_{j \in \mathcal{S}(i)} p_i p_j = \int_0^1 \bar{A}_m(s) ds.$$

Now by the definition of  $F^{\text{exc}, \mathbf{p}}$ ,

$$F^{\text{exc}, \mathbf{p}}(y^*(i)) = \sum_{v \in \mathcal{A}(i)} p_v, \text{ for } i \in [m]. \quad (4.13)$$

By (4.8),

$$A_m(t) = \sum_{v \in \mathcal{S}(i)} p_v = \sum_{v \in \mathcal{A}(i-1)} p_v - p_{v(i)}, \text{ for } t \in (y^*(i-1), y^*(i)].$$

Thus

$$\|A_m\|_\infty \leq \|F^{\text{exc}, \mathbf{p}}\|_\infty. \quad (4.14)$$

By Assumption 4.4(ii) and (4.12), for any  $s \geq 0$ , there exists  $K = K(s) < \infty$  such that

$$[L(\mathcal{T}_m^{\mathbf{p}})]^s \leq C^s \exp \left( K \frac{\|F^{\text{exc}, \mathbf{p}}\|_\infty}{\sigma(\mathbf{p})} \right). \quad (4.15)$$

Now the following lemma completes the proof of Proposition 4.8. ■

**Lemma 4.9.** *There exists a positive constant  $c > 0$  such that for every  $m \geq 1$  and  $x \geq e$ ,*

$$\mathbb{P} \left( \|F^{\text{exc}, \mathbf{p}}\|_\infty \geq x \sigma(\mathbf{p}) \right) \leq \exp(-cx \log(\log x)).$$

**Proof:** Write  $\mathcal{R}(m) := \|F^{\text{exc}, \mathbf{p}}\|_\infty / \sigma(\mathbf{p})$  and as before, let  $\mathbf{X} = (X_\nu : \nu \in [m])$  be the collection of uniform random variables used to construct  $F^{\mathbf{p}}$ . Write  $\mathbb{Q}[0, 1]$  for the set of rationals in  $[0, 1]$ . Then note that

$$\mathcal{R}(m) = \sup_{q \in \mathbb{Q}[0, 1]} \frac{F^{\mathbf{p}}(q)}{\sigma(\mathbf{p})} - \inf_{q \in \mathbb{Q}[0, 1]} \frac{F^{\mathbf{p}}(q)}{\sigma(\mathbf{p})} := \mathcal{R}_1(m) + \mathcal{R}_2(m). \quad (4.16)$$

We start by analyzing  $\mathcal{R}_1(m)$ . For fixed  $q \in \mathbb{Q}[0, 1]$ , define the collection of  $m$  functions

$$s_q^j(x) := \frac{p_j}{\sigma(\mathbf{p})} (\mathbb{1}\{x \leq q\} - q), \quad 1 \leq j \leq m.$$

Note that for all  $j \in [m]$ ,  $s_q^j : [0, 1] \rightarrow [-1, 1]$ , with  $\mathbb{E}(s_q^j(X_j)) = 0$  and further

$$\mathcal{R}_1(m) = \sup_{q \in \mathbb{Q}[0, 1]} \left( s_q^1(X_1) + \cdots + s_q^m(X_m) \right).$$

Also note that

$$\sup_{q \in \mathbb{Q}[0, 1]} \text{Var} \left( s_q^1(X_1) + \cdots + s_q^m(X_m) \right) = \sup_{q \in \mathbb{Q}[0, 1]} q(1 - q) = \frac{1}{4}.$$

If we can show that

$$\kappa := \sup_{m \geq 1} \mathbb{E}(\mathcal{R}_1(m)) < \infty, \quad (4.17)$$

then standard concentration inequalities for the maxima in empirical processes [49, Theorem 1.1(b)] will imply the existence of a constant  $c_1 > 0$  such that for all  $m \geq 1$  and  $x > 0$ ,

$$\mathbb{P}(\mathcal{R}_1(m) \geq \mathbb{E}(\mathcal{R}_1(m)) + x) \leq \exp \left( -\frac{x}{4} \log \left[ 1 + 2 \log \left( 1 + \frac{x}{2\kappa + \frac{1}{4}} \right) \right] \right). \quad (4.18)$$

Let us now prove (4.17). In fact we will show the stronger result:

$$\sup_{m \geq 1} \mathbb{E} \left( \sup_{q \in \mathbb{Q}[0, 1]} \left| \sum_{j=1}^m s_q^j(X_j) \right| \right) < \infty.$$

Let  $X_{(1)} < X_{(2)} < \cdots < X_{(m)}$  denote the order statistics of  $\mathbf{X}$  and let  $\pi$  denote the corresponding permutation of  $[m]$ , namely  $X_{(i)} = X_{\pi(i)}$ . Note that

$$\sup_{q \in \mathbb{Q}[0, 1]} \left| \sum_{j=1}^m s_q^j(X_j) \right| := \max_{1 \leq i \leq m} |\vartheta_i|, \quad \text{where} \quad \vartheta_i := \frac{-X_{(i)} + \sum_{j=1}^i p_{\pi(j)}}{\sigma(\mathbf{p})}.$$

Hence

$$\begin{aligned} \max_{i \in [m]} |\vartheta_i| &\leq \max_{i \in [m]} [\sigma(\mathbf{p})]^{-1} \left| -X_{(i)} + \frac{i}{m} \right| + \max_{i \in [m]} [\sigma(\mathbf{p})]^{-1} \left| \sum_{j=1}^i p_{\pi(j)} - \frac{i}{m} \right| \\ &:= \mathcal{R}_{11}(m) + \mathcal{R}_{12}(m). \end{aligned}$$

We first analyze  $\mathcal{R}_{11}(m)$ . By the DKW inequality [54],

$$\mathbb{P} \left( \max_{i \in [m]} \left| -X_{(i)} + \frac{i}{m} \right| \geq \sigma(\mathbf{p})x \right) \leq 2 \exp \left( -2m \cdot (\sigma(\mathbf{p})x)^2 \right)$$

By Cauchy-Schwartz,  $m\sigma^2(\mathbf{p}) \geq (\sum_i p_i)^2 = 1$ . Thus  $\sup_{m \geq 1} \mathbb{E}(\mathcal{R}_{11}(m)) < \infty$ . We now analyze  $\mathcal{R}_{12}(m)$ . Since

$$\mathbb{E}(p_{\pi(j)}) = \frac{1}{m}, \quad \text{and} \quad \mathbb{E}(p_{\pi(i)} p_{\pi(j)}) = \frac{\sum_{k \neq \ell \in [m]} p_k p_\ell}{m(m-1)} = \frac{1 - \sigma^2(\mathbf{p})}{m(m-1)} \quad \text{for } i \neq j \in [m],$$

for any  $i \in [m]$  we have

$$\mathbb{E} \left( \left( \sum_{j=1}^i p_{\pi(j)} - \frac{i}{m} \right)^2 \right) = \frac{i\sigma^2(\mathbf{p})}{m} + \frac{i(i-1)}{m(m-1)}(1 - \sigma^2(\mathbf{p})) - \frac{i^2}{m^2} \leq \frac{i}{m}\sigma^2(\mathbf{p}) \quad (4.19)$$

by simply expanding the square. Now note that since  $\pi$  is a uniform random permutation of the vertex set  $[m]$ , for any fixed  $i \geq 1$  we also have

$$\sum_{j=1}^i p_{\pi(j)} - \frac{i}{m} \stackrel{d}{=} \sum_{j=0}^{i-1} p_{\pi(m-j)} - \frac{i}{m} = \left( \frac{m-i}{m} - \sum_{j=1}^{m-i} p_{\pi(j)} \right).$$

Thus

$$\mathbb{E}(\mathcal{R}_{12}(m)) \leq 2\mathbb{E} \left( \max_{i \in [m/2]} [\sigma(\mathbf{p})]^{-1} \left| \sum_{j=1}^i p_{\pi(j)} - \frac{i}{m} \right| \right). \quad (4.20)$$

Now assuming that we construct  $\pi$  by sequentially sampling without replacement from  $[m]$ , let  $\mathcal{F}_k$  denote the  $\sigma$ -field generated by  $(\pi(1), \pi(2), \dots, \pi(k))$  for  $0 \leq k \leq m-1$ . Let  $M_0 = 0$  and consider the sequence

$$M_k := \frac{\sum_{j=1}^k p_{\pi(j)} - k/m}{m-k}, \quad 0 \leq k \leq m-1.$$

It is easy to check that  $\{M_k : 0 \leq k \leq m-1\}$  is a martingale with respect to the filtration  $\{\mathcal{F}_k : 0 \leq k \leq m-1\}$ . Then (4.20) and Doob's  $\mathbb{L}^2$ -maximal inequality yield

$$\mathbb{E}(\mathcal{R}_{12}(m)) \leq \frac{2m}{\sigma(\mathbf{p})} \sqrt{\mathbb{E}([M_{m/2}]^2)}$$

Using (4.19) with  $i = m/2$  then gives  $\mathbb{E}(\mathcal{R}_{12}(m)) \leq 16$  for all  $m \geq 1$ . Thus we have shown that  $\sup_{m \geq 1} \max(\mathbb{E}(\mathcal{R}_{11}(m)), \mathbb{E}(\mathcal{R}_{12}(m))) < \infty$ . This proves (4.17) and thus (4.18).

To complete the proof of the lemma, we need to get a tail bound on  $\mathcal{R}_2(m)$  appearing in (4.16). As before, using [49], it is enough to show  $\sup_{m \geq 1} \mathbb{E}(\mathcal{R}_2(m)) < \infty$ . However, note that

$$\mathcal{R}_2(m) = \max_{i \in [m]} \frac{\left| \sum_{j=1}^{i-1} p_{\pi(j)} - X_{(i)} \right|}{\sigma(\mathbf{p})} \leq \mathcal{R}_1(m) + \frac{p_{\max}}{\sigma(\mathbf{p})}.$$

We now use (4.17) together with Assumption 4.4 to complete the proof. ■

**4.4. Another construction of  $\tilde{\mathcal{G}}_m(\mathbf{p}, a)$  and a modification.** In this section, we start by giving a more explicit description of the algorithm described in Proposition 4.3 via adding permitted edges to a tilted  $\mathbf{p}$ -tree. We first set up some notation. As a matter of convention, we will view ordered rooted trees via their planar embedding, using the associated ordering to determine the relative locations of siblings of an individual. We think of the left most sibling as the “oldest”. Further, in a depth-first exploration, we explore the tree from left to right. Now given a planar rooted tree  $\mathbf{t} \in \mathbb{T}_m$ , let  $\rho$  denote the root and for every vertex  $v \in [m]$ , let  $[\rho, v]$  denote the path connecting  $\rho$  to  $v$  in the tree. Given this path and a vertex  $i \in [\rho, v]$ , write  $\mathcal{RC}(i, [\rho, v])$  for the set of all children of  $i$  which fall to the right of  $[\rho, v]$ . Thus in the depth-first exploration of the tree, when we get to  $v$ ,

$$\mathfrak{P}(v, \mathbf{t}) := \cup_{i \in [m]} \mathcal{RC}(i, [\rho, v]) \quad (4.21)$$

denotes the set of endpoints of all permitted edges emanating from  $v$ . Define

$$\mathfrak{G}_{(m)}(v) := \sum_{i \in [\rho, v]} \sum_{j \in [m]} p_j \mathbb{1}\{j \in \mathcal{RC}(i, [\rho, v])\}. \quad (4.22)$$

The function  $A_m(\cdot)$  defined in (4.8) is intimately connected to  $\mathfrak{G}_{(m)}(\cdot)$ . More precisely, let  $(v(1), v(2), \dots, v(m))$  denote the order in the depth-first exploration of the tree. Let  $y^*(0) = 0$  and  $y^*(i) = y^*(i-1) + p_{v(i)}$ . Define

$$A_{(m)}(u) = \mathfrak{G}_{(m)}(u), \quad \text{for } u \in (y^*(i-1), y^*(i)], \quad \text{and} \quad \bar{A}_{(m)}(\cdot) := aA_{(m)}(\cdot). \quad (4.23)$$

Then the function  $A_{(m)}(\cdot)$  associated with an ordered  $\mathbf{p}$ -tree has the same distribution as the function  $A_m(\cdot)$  associated with the tree  $\psi_{\mathbf{p}}(\mathbf{X})$ , where  $\mathbf{X} = (X_v : v \in [m])$  are i.i.d. random variables uniformly distributed on  $(0, 1)$ .

Finally, define the function

$$\Lambda_{(m)}(\mathbf{t}) := a \sum_{v \in [m]} p_v \mathfrak{G}_{(m)}(v). \quad (4.24)$$

While all of these objects depend on the tree  $\mathbf{t}$ , we suppress this dependence to ease notation. Now Proposition 4.3 implies we can construct  $\tilde{\mathcal{G}}_m(\mathbf{p}, a)$  via the following five steps:

- (i) **Tilted  $\mathbf{p}$ -tree:** Generate a tilted ordered  $\mathbf{p}$ -tree  $\mathcal{T}_m^{\mathbf{p}, \star}$  with distribution (4.5). Now consider the (random) objects  $\mathfrak{P}(v, \mathcal{T}_m^{\mathbf{p}, \star})$  for  $v \in [m]$  and the corresponding (random) functions  $\mathfrak{G}_{(m)}(\cdot)$  on  $[m]$  and  $A_{(m)}(\cdot)$  on  $[0, 1]$ .
- (ii) **Poisson number of possible surplus edges:** Let  $\mathcal{P}$  denote a rate one Poisson process on  $\mathbb{R}_+^2$  and define

$$\bar{A}_{(m)} \cap \mathcal{P} := \{(s, t) \in \mathcal{P} : s \in [0, 1], t \leq \bar{A}_{(m)}(s)\}. \quad (4.25)$$

Write  $\bar{A}_{(m)} \cap \mathcal{P} := \{(s_j, t_j) : 1 \leq j \leq N_{(m)}^{\star}\}$  where  $N_{(m)}^{\star} = |\bar{A}_{(m)} \cap \mathcal{P}|$ .

We will now use the set  $\{(s_j, t_j) : 1 \leq j \leq N_{(m)}^{\star}\}$  to generate pairs of points  $\{(\mathcal{L}_j, \mathcal{R}_j) : 1 \leq j \leq N_{(m)}^{\star}\}$  in the tree that will be joined to form the surplus edges.

- (iii) **"First" endpoints:** Fix  $j$  and suppose  $s_j \in (y^*(i-1), y^*(i)]$  for some  $i \geq 1$ , where  $y^*(i)$  is as given right above (4.23). Then the *first endpoint* of the surplus edge corresponding to  $(s_j, t_j)$  is  $\mathcal{L}_j := v(i)$ .
- (iv) **"Second" endpoints:** Note that in the interval  $(y^*(i-1), y^*(i)]$ , the function  $\bar{A}_{(m)}$  is of constant height  $a\mathfrak{G}_{(m)}(v(i))$ . We will view this height as being partitioned into sub-intervals of length  $ap_u$  for each  $u \in \mathfrak{P}(v(i), \mathcal{T}_m^{\mathbf{p}, \star})$ , the collection of endpoints of permitted edges emanating from  $\mathcal{L}_k$ . (Assume that this partitioning is done according to some preassigned rule, e.g., using the order of the vertices in  $\mathfrak{P}(v(i), \mathcal{T}_m^{\mathbf{p}, \star})$ .) Suppose  $t_j$  belongs to the interval corresponding to  $u$ . Then the *second endpoint* is  $\mathcal{R}_j = u$ . Form an edge between  $(\mathcal{L}_j, \mathcal{R}_j)$ .
- (v) In this construction, it is possible that one created more than one surplus edge between two vertices. Remove any multiple surplus edges.

**Lemma 4.10.** *The above construction gives a random graph with distribution  $\tilde{\mathcal{G}}_m(\mathbf{p}, a)$  as in Definition 4.2. Further, conditional on  $\mathcal{T}_m^{\mathbf{p}, \star}$ :*

- (a)  $N_{(m)}^{\star}$  has Poisson distribution with mean  $\Lambda_{(m)}(\mathcal{T}_m^{\mathbf{p}, \star})$  where  $\Lambda_{(m)}$  is as in (4.24).
- (b) Conditional on  $\mathcal{T}_m^{\mathbf{p}, \star}$  and  $N_{(m)}^{\star} = k$ , the first endpoints  $(\mathcal{L}_j : 1 \leq j \leq k)$  can be generated in an i.i.d. fashion by sampling from the vertex set  $[m]$  with probability distribution

$$\mathcal{J}^{(m)}(v) \propto p_v \mathfrak{G}_{(m)}(v), \quad v \in [m].$$

- (c) Conditional on  $\mathcal{T}_m^{\mathbf{p}, \star}$ ,  $N_{(m)}^{\star} = k$  and the first endpoints  $(\mathcal{L}_j : 1 \leq j \leq k)$ , the second endpoints can be generated in an i.i.d. fashion where the probability that  $\mathcal{R}_j = u$  is proportional to  $p_u$  if  $u$  is a right child of some individual  $y \in [\rho, \mathcal{L}_j]$ .

**Proof:** The assertions follow from Proposition 4.3 and standard properties of Poisson processes.  $\blacksquare$

**The modified space  $\mathcal{G}_m^{\text{mod}}(\mathbf{p}, a)$ :** We construct a modified graph  $\mathcal{G}_m^{\text{mod}}(\mathbf{p}, a)$  as follows:

- (i') Generate a tilted ordered  $\mathbf{p}$ -tree  $\mathcal{T}_m^{\mathbf{p}, \star}$  with distribution (4.5).
- (ii') Conditional on  $\mathcal{T}_m^{\mathbf{p}, \star} = k$ , generate  $N_{(m)}^{\star} \sim \text{Poi}(\Lambda_{(m)}(\mathcal{T}_m^{\mathbf{p}, \star}))$ .
- (iii') Conditional on  $\mathcal{T}_m^{\mathbf{p}, \star}$  and  $N_{(m)}^{\star} = k$ , generate the first endpoints  $(\mathcal{L}_j : 1 \leq j \leq k)$  in an i.i.d. fashion by sampling from the vertex set  $[m]$  with probability distribution

$$\mathcal{J}^{(m)}(v) \propto p_v \mathfrak{G}_{(m)}(v), \quad v \in [m].$$

- (iv') Conditional on  $\mathcal{T}_m^{\mathbf{p}, \star}$ ,  $N_{(m)}^{\star} = k$  and the first endpoints  $(\mathcal{L}_j : 1 \leq j \leq k)$ , generate the second endpoints in an i.i.d. fashion where conditional on  $\mathcal{L}_j = v$ , the probability distribution of  $\mathcal{R}_j$  is given by

$$Q_v^{(m)}(y) := \begin{cases} \sum_j p_u \mathbb{1}\{u \in \mathcal{RC}(y, [\rho, v])\} / \mathfrak{G}_{(m)}(v) & \text{if } y \in [\rho, v], \\ 0 & \text{otherwise.} \end{cases} \quad (4.26)$$

Identify  $\mathcal{L}_j$  and  $\mathcal{R}_j$  for  $1 \leq j \leq k$ .

Thus, instead of adding an edge between  $\mathcal{L}_j$  and one of the right children on the path  $[\rho, \mathcal{L}_j]$  as in Lemma 4.10(c), we identify it to the parent of this vertex which is on  $[\rho, \mathcal{L}_j]$ . Also, we do not remove any multiple surplus edges. This construction turns out to be easier to work with.  $\mathcal{G}_m^{\text{mod}}(\mathbf{p}, a)$  will be viewed as a metric measure space via the graph distance where vertex  $v$  has mass  $\sum p_u$  where the sum is taken over all  $u \in [m]$  which have been identified with  $v$ . Intuitively it is clear that  $\sigma(\mathbf{p})\tilde{\mathcal{G}}_m(\mathbf{p}, a)$  and  $\sigma(\mathbf{p})\mathcal{G}_m^{\text{mod}}(\mathbf{p}, a)$  are “close”. This is formalized in Lemma 4.12.

**Remark 6.** At this point we urge the reader to go back to Section 2.3.1 and remind themselves of the four steps in the construction of the limit metric space  $\mathcal{G}_{\infty}(\boldsymbol{\theta}, \gamma)$ , and note the similarities to the construction above. In particular, we make note of the following:

- (a) For finite  $m$ , we essentially tilt the  $\mathbf{p}$ -tree distribution via the functional  $\bar{L}(\mathcal{T}_m^{\mathbf{p}}) = \exp(a\mathbb{E}[\mathfrak{G}_{(m)}(V_1) | \mathcal{T}_m^{\mathbf{p}}])$  (the term  $\mathbb{I}(\mathcal{T}_m^{\mathbf{p}})$  as in (4.10) can be ignored as we will see in Lemma 4.14), and the number of shortcut points selected, namely  $N_{(m)}^{\star}$ , has a Poisson distribution with mean  $a\mathbb{E}[\mathfrak{G}_{(m)}(V_1) | \mathcal{T}_m^{\mathbf{p}, \star}]$ . Here  $V_1$  has distribution  $\mathbf{p}$ .
- (b) For the limit object, we tilt the measure using the functional  $L_{(\infty)}(\mathcal{T}_{(\infty)}^{\boldsymbol{\theta}}, \mathbf{U}) = \exp(\gamma\mathbb{E}[\mathfrak{G}_{(\infty)}(V_1) | \mathcal{T}_{(\infty)}^{\boldsymbol{\theta}}, \mathbf{U}])$ , and the number of shortcuts, namely  $N_{(\infty)}^{\star}$ , follows a Poisson distribution with mean  $\gamma\mathbb{E}[\mathfrak{G}_{(\infty)}(V_1) | \mathcal{T}_{(\infty)}^{\boldsymbol{\theta}, \star}, \mathbf{U}^{\star}]$ . Here  $V_1$  is distributed according to the mass measure  $\mu^{\star}$  on  $\mathcal{T}_{(\infty)}^{\boldsymbol{\theta}, \star}$ .

As a brief warm-up to the kind of calculations in the next section, we now prove a simple lemma on tightness of the number of surplus edges. We will prove distributional convergence of this object in the next section.

**Lemma 4.11.** *Under Assumption 4.4, the sequence  $\{N_{(m)}^{\star} : m \geq 1\}$  is tight, where  $N_{(m)}^{\star}$  is as given below (4.25).*

**Proof:** Fix  $r > 1$ . First note that conditional on  $\mathcal{T}_m^{\mathbf{p}, \star} = \mathbf{t}$ ,  $N_{(m)}^{\star}$  has a Poisson distribution with mean  $\Lambda_{(m)}(\mathbf{t})$ . Thus, there exists a constant  $C = C(r)$  such that

$$\mathbb{E}([N_{(m)}^{\star}]^r | \mathcal{T}_m^{\mathbf{p}, \star} = \mathbf{t}) \leq C[\Lambda_{(m)}(\mathbf{t})]^r.$$

Further, note that the tilt  $L(\mathbf{t})$  in (4.4) satisfies

$$L(\mathbf{t}) := \mathbb{I}(\mathbf{t}) \exp \left( \sum_{(k,\ell) \in \mathfrak{P}(\mathbf{t})} a p_k p_\ell \right) = \mathbb{I}(\mathbf{t}) \exp(\Lambda_{(m)}(\mathbf{t})),$$

where  $1 \leq \mathbb{I}(\mathbf{t}) \leq C'$  for a fixed constant  $C'$  independent of  $m$  by (4.11). Thus, Proposition 4.8 shows that

$$\sup_{m \geq 1} \mathbb{E}(\exp(\gamma \Lambda_{(m)}(\mathcal{T}_m^{\mathbf{p}}))) < \infty$$

for any  $\gamma > 0$ . In particular,

$$\sup_{m \geq 1} \mathbb{E}([N_{(m)}^*]^r) \leq \sup_{m \geq 1} C \mathbb{E}([\Lambda_{(m)}(\mathcal{T}_m^{\mathbf{p},*})]^r) = C \sup_{m \geq 1} \frac{\mathbb{E}([\Lambda_{(m)}(\mathcal{T}_m^{\mathbf{p}})]^r L(\mathcal{T}_m^{\mathbf{p}}))}{\mathbb{E}(L(\mathcal{T}_m^{\mathbf{p}}))} < \infty,$$

which proves tightness of  $\{N_{(m)}^* : m \geq 1\}$ .  $\blacksquare$

We conclude this section by proving a lemma which essentially says that it is enough to work with the modified space  $\mathcal{G}_m^{\text{mod}}(\mathbf{p}, a)$ .

**Lemma 4.12.** *Under Assumption 4.4, we have*

$$d_{\text{GHP}} \left( \sigma(\mathbf{p}) \tilde{\mathcal{G}}_m(\mathbf{p}, a), \sigma(\mathbf{p}) \mathcal{G}_m^{\text{mod}}(\mathbf{p}, a) \right) \xrightarrow{\mathbb{P}} 0.$$

**Proof:** Recall the five-step construction of  $\tilde{\mathcal{G}}_m(\mathbf{p}, a)$ . Construct  $\mathcal{G}_m^{\text{mod}}(\mathbf{p}, a)$  on the same space by coupling it with  $\tilde{\mathcal{G}}_m(\mathbf{p}, a)$  in the obvious way. Define the event

$$F := \{N_{(m)}^* \text{ equals the number of surplus edges in } \tilde{\mathcal{G}}_m(\mathbf{p}, a)\}.$$

In other words,  $F$  describes the event in which  $\tilde{\mathcal{G}}_m(\mathbf{p}, a)$  does not have multiple surplus edges. It is easy to check that

$$d_{\text{GHP}} \left( \tilde{\mathcal{G}}_m(\mathbf{p}, a), \mathcal{G}_m^{\text{mod}}(\mathbf{p}, a) \right) \leq N_{(m)}^* \text{ on the set } F.$$

Thus, Lemma 4.11 combined with the assumption  $\sigma(\mathbf{p}) \rightarrow 0$  yields the result provided we show that  $\mathbb{P}(F^c) \rightarrow 0$ . To this end, note that

$$\begin{aligned} \mathbb{P} \left( \exists \text{ multiple surplus edges between } u \text{ and } v \mid \mathcal{T}_m^{\mathbf{p},*} = \mathbf{t} \right) \\ = \mathbb{P} \left( \text{Poi}(a p_u p_v) \geq 2 \right) \leq c(a p_u p_v)^2 \end{aligned}$$

for every  $u \in [m]$ ,  $v \in \mathfrak{P}(u, \mathbf{t})$ , and some universal positive constant  $c$ . Hence

$$\begin{aligned} \mathbb{P} \left( F^c \mid \mathcal{T}_m^{\mathbf{p},*} = \mathbf{t} \right) &\leq c a^2 \sigma(\mathbf{p})^2 \sum_{u \in [m]} p_u^2 \sum_{v \in \mathfrak{P}(u, \mathbf{t})} (p_v / \sigma(\mathbf{p}))^2 \\ &\leq c (a \sigma(\mathbf{p}))^2 \sum_{u \in [m]} p_u^2 = c (a \sigma(\mathbf{p}))^2 \sigma(\mathbf{p})^2. \end{aligned}$$

Since  $\sigma(\mathbf{p}) \rightarrow 0$  and  $a \sigma(\mathbf{p}) \rightarrow \gamma$ ,  $\mathbb{P}(F^c) \rightarrow 0$  as desired.  $\blacksquare$

**4.5. Completing the proof of Theorem 4.5.** At this point we urge the reader to remind themselves of (a) the four steps in the construction of the limit object in Section 2.3, (b) the birthday construction of  $\mathbf{p}$ -trees at the end of Section 4.2.1 and (c) the definition of Gromov-weak topology in Section 2.1.2 of complete separable measured metric spaces  $\mathcal{S}_*$ . Fix  $\ell \geq 1$  and a bounded continuous function  $\phi : \mathbb{R}_+^{\ell^2} \rightarrow \mathbb{R}$ . Let  $\Phi$  be as in (2.6). To simplify notation, we will write  $\Phi(X)$

instead of  $\Phi(X, d, \mu)$ . To prove Theorem 4.5, we need to show that for every fixed  $\ell \geq 1$  and functions  $\phi$  and  $\Phi$  as above,

$$\mathbb{E}[\Phi(\sigma(\mathbf{p}) \cdot \tilde{\mathcal{G}}_m(\mathbf{p}, a))] \rightarrow \mathbb{E}[\Phi(\mathcal{G}_\infty(\boldsymbol{\theta}, \gamma))] \quad \text{as } m \rightarrow \infty,$$

where we sample  $\ell$  points according to  $\mathbf{p}$  in  $\tilde{\mathcal{G}}_m(\mathbf{p}, a)$  while we sample  $\ell$  points according to the measure on  $\mathcal{G}_\infty(\boldsymbol{\theta}, \gamma)$  inherited from the mass measure. Now recall the explicit five step construction of  $\tilde{\mathcal{G}}_m(\mathbf{p}, a)$  in Section 4.4 starting from the tilted  $\mathbf{p}$ -tree  $\mathcal{T}_m^{\mathbf{p}, \star}$  and the Poisson number of surplus edges  $N_{(m)}^\star$ . Fix  $K \geq 1$  and note that

$$\left| \mathbb{E}[\Phi(\sigma(\mathbf{p}) \tilde{\mathcal{G}}_m(\mathbf{p}, a))] - \sum_{k=0}^K \mathbb{E}[\Phi(\sigma(\mathbf{p}) \tilde{\mathcal{G}}_m(\mathbf{p}, a)) \mathbb{1}\{N_{(m)}^\star = k\}] \right| \leq \|\phi\|_\infty \mathbb{P}(N_{(m)}^\star \geq K+1).$$

Using Lemma 4.11, we can choose  $K$  large (independent of  $m$ ) to make the bound on the right arbitrarily small. Further, in view of Lemma 4.12, we can work with  $\mathcal{G}_m^{\text{mod}}(\mathbf{p}, a)$  instead of  $\tilde{\mathcal{G}}_m(\mathbf{p}, a)$ . Hence it suffices to prove the following convergence for every fixed  $k \geq 0$ :

$$\mathbb{E}[\Phi(\sigma(\mathbf{p}) \mathcal{G}_m^{\text{mod}}(\mathbf{p}, a)) \mathbb{1}\{N_{(m)}^\star = k\}] \rightarrow \mathbb{E}[\Phi(\mathcal{G}_\infty(\boldsymbol{\theta}, \gamma)) \mathbb{1}\{N_{(\infty)}^\star = k\}] \quad \text{as } m \rightarrow \infty. \quad (4.27)$$

To analyze this term, we first need to setup some notation.

Note that both the finite  $m$  and the limit object are obtained by starting with a discrete tree for finite  $m$  and a *real tree* in the limit, and sampling a random number of pairs to create “shortcuts”. Recall the space  $\mathbf{T}_{IJ}^*$  in Section 2.1.3. Fix  $k \geq 0$  and let  $\mathbf{t}$  be an element in  $\mathbf{T}_{I, (k+\ell)}^*$  for some  $I \geq 0$ . “ $I$ ” will not play a role in the definition below. Write  $\rho$  for the root and denote the leaves by

$$\mathbf{x}_{k, k+\ell} := (x_1, x_2, \dots, x_k, x_{k+1}, \dots, x_{k+\ell}). \quad (4.28)$$

Also recall that for each  $i$ , there is a probability measure  $\nu_{\mathbf{t}, i}(\cdot)$  on the path  $[\rho, x_i]$  for  $1 \leq i \leq k+\ell$ . For  $1 \leq i \leq k$ , sample  $y_i$  according to the distribution  $\nu_{\mathbf{t}, i}(\cdot)$  independently for different  $i$  and connect  $x_i$  and  $y_i$ . Let  $\mathbf{t}'$  denote the (random) tree thus obtained and let  $d_{\mathbf{t}'}$  denote the graph distance on  $\mathbf{t}'$ . Define the function  $g_\phi^{(k)} : \mathbf{T}_{I, (k+\ell)}^* \rightarrow \mathbb{R}$  by

$$g_\phi^{(k)}(\mathbf{t}) := \begin{cases} \mathbb{E}[\phi(d_{\mathbf{t}'}(x_i, x_j) : k+1 \leq i \leq k+\ell)], & \text{if } \mathbf{t} \neq \partial, \\ 0, & \text{if } \mathbf{t} = \partial. \end{cases} \quad (4.29)$$

In words, we look at the expectation of  $\phi$  applied to the pairwise distances between the last  $\ell$  leaves after sampling  $y_i$  on the path  $[\rho, x_i]$  for  $1 \leq i \leq k$  and connecting  $x_i$  and  $y_i$ . Note that here the expectation is only taken over the choices of  $y_i$ .

Next, given  $\mathbf{t} \in \mathbb{T}_m^{\text{ord}}$  and  $\mathbf{v} := (v_1, \dots, v_r)$  with  $v_i \in [m]$ , set  $\mathbf{t}(\mathbf{v})$  to be the subtree of  $\mathbf{t}$  spanning the vertices  $\mathbf{v}$  and the root provided  $v_1, \dots, v_r$  are all distinct and none of them is an ancestor of another vertex in  $\mathbf{v}$ . When this condition fails, set  $\mathbf{t}(\mathbf{v}) = \partial$ .

Now, conditional on  $\mathcal{T}_m^{\mathbf{p}, \star}$ , construct a tree  $\mathcal{T}_m^{\mathbf{p}, \star}(\tilde{\mathbf{V}}_{k, k+\ell}^{(m)})$  where

- (i)  $\tilde{\mathbf{V}}_{k, k+\ell}^{(m)} := (\bar{V}_1^{(m)}, \dots, \bar{V}_k^{(m)}, V_{k+1}^{(m)}, \dots, V_{k+\ell}^{(m)})$ ;
- (ii)  $\bar{V}_i^{(m)}$ ,  $1 \leq i \leq k$  are i.i.d. with the distribution  $\mathcal{J}^{(m)}(\cdot)$  as in Lemma 4.10(b); and
- (iii)  $V_{k+1}^{(m)}, \dots, V_{k+\ell}^{(m)}$  are i.i.d. with distribution  $\mathbf{p}$ . Further,  $\bar{V}_1^{(m)}, \dots, \bar{V}_k^{(m)}, V_{k+1}^{(m)}, \dots, V_{k+\ell}^{(m)}$  are jointly independent.

We will drop the superscript and simply write  $V_i^{(m)}$ ,  $\bar{V}_i^{(m)}$  etc. when there is no scope of confusion. Note that  $\mathcal{T}_m^{\mathbf{p}, \star}(\tilde{\mathbf{V}}_{k, k+\ell}) = \partial$  whenever  $\bar{V}_1, \dots, \bar{V}_k, V_{k+1}, \dots, V_{k+\ell}$  are not all distinct or one of them is an ancestor of another vertex in  $\tilde{\mathbf{V}}_{k, k+\ell}$ . In either of these two case, the subtree spanned by the root and  $\tilde{\mathbf{V}}_{k, k+\ell}$  will have less than  $k+\ell$  leaves. We made the convention of

setting  $\mathcal{T}_m^{\mathbf{p},*}(\tilde{\mathbf{V}}_{k,k+\ell}) = \partial$  to make sure that we are always working with a bona fide element in  $\mathbf{T}_{L,(k+\ell)}^*$ . However, this makes no difference at all since by [27, Corollary 15],

$$\lim_m \mathbb{P}(\mathcal{T}_m^{\mathbf{p}}(V_1, \dots, V_{k+\ell}) = \partial) = 0$$

where  $V_1, \dots, V_{k+\ell}$  are i.i.d.  $\mathbf{p}$  random variables. Now  $\mathcal{T}_m^{\mathbf{p},*}$  is obtained by tilting the distribution of  $\mathcal{T}_m^{\mathbf{p}}$ , where the tilt  $L(\cdot)$  is uniformly integrable (Proposition 4.8). Further,  $\tilde{V}_i$ ,  $1 \leq i \leq k$  are i.i.d. with the distribution  $\mathcal{J}^{(m)}(v) \propto p_v \mathfrak{G}_{(m)}(v)$  where  $\max_v \mathfrak{G}_{(m)}(v)$  is stochastically dominated by  $\|F^{\text{exc}, \mathbf{p}}\|_\infty$  (see (4.14) and the discussion below (4.23)). It thus follows that

$$\lim_m \mathbb{P}(\mathcal{T}_m^{\mathbf{p},*}(\tilde{\mathbf{V}}_{k,k+\ell}) = \partial) = 0. \quad (4.30)$$

Using (4.30), we see that

$$\begin{aligned} & \mathbb{E} \left[ \Phi \left( \sigma(\mathbf{p}) \mathcal{G}_m^{\text{mod}}(\mathbf{p}, a) \right) \mathbb{1} \{N_{(m)}^* = k\} \right] \\ &= \mathbb{E} \left\{ \mathbb{E}_{\mathbf{p},*} \left[ g_\phi^{(k)} \left( \sigma(\mathbf{p}) \mathcal{T}_m^{\mathbf{p},*}(\tilde{\mathbf{V}}_{k,k+\ell}) \right) \right] \mathbb{1} \{N_{(m)}^* = k\} \right\} + o(1), \end{aligned} \quad (4.31)$$

where  $\mathbb{E}_{\mathbf{p},*}(\cdot) := \mathbb{E}(\cdot | \mathcal{T}_m^{\mathbf{p},*})$ . At this point, we also define  $\mathbb{E}_{\mathbf{p}}(\cdot) := \mathbb{E}(\cdot | \mathcal{T}_m^{\mathbf{p}})$  where  $\mathcal{T}_m^{\mathbf{p}}$  has the original ordered  $\mathbf{p}$ -tree distribution (2.8).

Now since  $\mathcal{J}^{(m)}(v) \propto p_v \mathfrak{G}_{(m)}(v)$ , we see that the inner expectation in (4.31) can be simplified as

$$\mathbb{E}_{\mathbf{p},*} \left[ g_\phi^{(k)} \left( \sigma(\mathbf{p}) \mathcal{T}_m^{\mathbf{p},*}(\tilde{\mathbf{V}}_{k,k+\ell}) \right) \right] = \frac{\mathbb{E}_{\mathbf{p},*} \left[ \prod_{i=1}^k \mathfrak{G}_{(m)}(V_i) g_\phi^{(k)} \left( \sigma(\mathbf{p}) \mathcal{T}_m^{\mathbf{p},*}(\mathbf{V}_{k,k+\ell}) \right) \right]}{[\mathbb{E}_{\mathbf{p},*}(\mathfrak{G}_{(m)}(V_1))]^k}, \quad (4.32)$$

where  $\mathbf{V}_{k,k+\ell} = (V_1, V_2, \dots, V_{k+\ell})$ , and  $V_i$  are i.i.d. with distribution  $\mathbf{p}$ . Since  $\mathcal{T}_m^{\mathbf{p},*}$  is sampled according to a tilted  $\mathbf{p}$ -tree distribution, combining (4.31), (4.32), we get the following result:

**Lemma 4.13.** *Fix  $k \geq 0$ . Then*

$$\begin{aligned} & \mathbb{E} \left[ \Phi \left( \sigma(\mathbf{p}) \mathcal{G}_m^{\text{mod}}(\mathbf{p}, a) \right) \mathbb{1} \{N_{(m)}^* = k\} \right] \\ &= C_m \mathbb{E} \left[ \frac{\mathbb{E}_{\mathbf{p}} \left[ \left( \prod_{i=1}^k \mathfrak{G}_{(m)}(V_i) \right) g_\phi^{(k)} \left( \sigma(\mathbf{p}) \mathcal{T}_m^{\mathbf{p}}(\mathbf{V}_{k,k+\ell}) \right) \right]}{[\mathbb{E}_{\mathbf{p}}(\mathfrak{G}_{(m)}(V_1))]^k} L(\mathcal{T}_m^{\mathbf{p}}) \mathbb{1} \{N_{(m)} = k\} \right] + o(1), \end{aligned} \quad (4.33)$$

where  $C_m = \{\mathbb{E}(L(\mathcal{T}_m^{\mathbf{p}}))\}^{-1}$ , and  $L$  is the tilt as in (4.4). Further, conditional on  $\mathcal{T}_m^{\mathbf{p}}$ ,  $N_{(m)}$  has a Poisson distribution with mean  $\Lambda_{(m)}(\mathcal{T}_m^{\mathbf{p}}) = a \mathbb{E}_{\mathbf{p}}(\mathfrak{G}_{(m)}(V))$  as in (4.24), where  $V$  has distribution  $\mathbf{p}$  independent of  $\mathcal{T}_m^{\mathbf{p}}$ .

This formula will be the starting point to prove (4.27). Recall from (4.10) that the tilt  $L(\cdot) = I(\cdot) \bar{L}(\cdot)$ , where  $I(\cdot)$  has a messy form given by (4.11). We have already seen in (4.12) that under Assumption 4.4,  $\mathbb{I}(\cdot) \leq C$  for a constant  $C$  all  $m \geq 1$ . The following lemma coupled with dominated convergence theorem will now imply that we can replace  $L$  with  $\bar{L}$  in Lemma 4.13 and in all the subsequent analysis below:

**Lemma 4.14.** *Under Assumption 4.4,  $\mathbb{I}(\mathcal{T}_m^{\mathbf{p}}) \xrightarrow{\mathbf{P}} 1$  as  $m \rightarrow \infty$ .*

**Proof:** By (4.11) we have  $1 \leq \mathbb{I}(\mathcal{T}_m^{\mathbf{p}}) \leq \exp(a \sum_{(k,l) \in E(\mathcal{T}_m^{\mathbf{p}})} p_k p_l)$ . Thus it is enough to show that  $a \mathbb{E}(\sum_{(k,l) \in E(\mathcal{T}_m^{\mathbf{p}})} p_k p_l) \rightarrow 0$ . Now for  $k \neq l \in [m]$ , write  $\{k \rightsquigarrow l\}$  for the event in which  $l$  is a child of  $k$  in  $\mathcal{T}_m^{\mathbf{p}}$ . Then standard properties of  $\mathbf{p}$ -trees [59, Section 6.2] implies that for  $k \neq l_1 \neq l_2 \in [m]$

$$\mathbb{P}(k \rightsquigarrow l_1) = p_k, \quad \mathbb{P}(k \rightsquigarrow l_1 \text{ and } k \rightsquigarrow l_2) = p_k^2. \quad (4.34)$$



Thus

$$a \mathbb{E} \left( \sum_{(k,l) \in E(\mathcal{T}_m^{\mathbf{p}})} p_k p_l \right) = a \sum_{k=1}^m p_k \sum_{l \neq k} p_l p_k \leq a \sum_{k=1}^m p_k^2 = a[\sigma(\mathbf{p})]^2 \rightarrow 0,$$

as  $m \rightarrow \infty$  by Assumption 4.4.  $\blacksquare$

Write  $\mathbb{E}_{\theta}$  for expectation conditional on  $\mathcal{T}_{(\infty)}^{\theta}$  and the random variables  $U_j^{(i)}$  which encode the order on  $\mathcal{T}_{(\infty)}^{\theta}$ , i.e.,

$$\mathbb{E}_{\theta}(\cdot) := \mathbb{E}(\cdot \mid \mathcal{T}_{(\infty)}^{\theta}, \mathbf{U}),$$

and note that  $\mathbb{E}[\Phi(\mathcal{G}_{\infty}(\theta, \gamma)) \mathbb{1}\{N_{(\infty)}^{\star} = k\}]$  has an expression similar to (4.33). Indeed, from the construction of  $\mathcal{G}_{\infty}(\theta, \gamma)$  given in Section 2.3.1, it follows that

$$\begin{aligned} & \mathbb{E}[\Phi(\mathcal{G}_{\infty}(\theta, \gamma)) \mathbb{1}\{N_{(\infty)}^{\star} = k\}] \\ &= C_{\infty} \mathbb{E} \left[ \frac{\mathbb{E}_{\theta} \left[ \left( \prod_{i=1}^k \mathfrak{G}_{(\infty)}(V_i^{(\infty)}) \right) g_{\phi}^{(k)} \left( \mathcal{T}_{(\infty)}^{\theta}(\mathbf{V}_{k,k+\ell}^{(\infty)}) \right) \right]}{\left[ \mathbb{E}_{\theta}(\mathfrak{G}_{(\infty)}(V_1^{(\infty)})) \right]^k} L_{(\infty)}(\mathcal{T}_{(\infty)}^{\theta}, \mathbf{U}) \mathbb{1}\{N_{(\infty)} = k\} \right], \end{aligned} \quad (4.35)$$

where **(a)**  $\mathfrak{G}_{(\infty)}(\cdot)$  is as defined in (2.10) **(b)**  $L_{(\infty)}(\mathcal{T}_{(\infty)}^{\theta}, \mathbf{U})$  is as in (2.12), **(c)**  $C_{\infty} = [\mathbb{E} L_{(\infty)}(\mathcal{T}_{(\infty)}^{\theta}, \mathbf{U})]^{-1}$ , **(d)**  $V_i^{(\infty)}$  are i.i.d. random variables sampled from  $\mathcal{T}_{(\infty)}^{\theta}$  using the mass measure  $\mu$ , **(e)**  $\mathbf{V}_{k,k+\ell}^{(\infty)} = (V_1^{(\infty)}, \dots, V_{k+\ell}^{(\infty)})$ , **(f)**  $\mathcal{T}_{(\infty)}^{\theta}(\mathbf{V}_{k,k+\ell}^{(\infty)})$  is the tree spanned by the root of  $\mathcal{T}_{(\infty)}^{\theta}$  and  $\mathbf{V}_{k,k+\ell}^{(\infty)}$ , viewed as an element of  $\mathbf{T}_{0,k+\ell}^{\star}$  by declaring the leaf values to be  $\mathfrak{G}_{(\infty)}(V_j^{(\infty)})$  and the root-to-leaf measures to be  $Q_{V_j}^{(\infty)}(\cdot)$  as in (2.11), and **(g)** conditional on  $(\mathcal{T}_{(\infty)}^{\theta}, \mathbf{U})$ ,  $N_{(\infty)}$  has a Poisson distribution with mean

$$\Lambda_{(\infty)} := \gamma \int_{y \in \mathcal{T}_{(\infty)}^{\theta}} \mathfrak{G}_{(\infty)}(y) \mu(dy) = \mathbb{E}_{\theta}[\mathfrak{G}_{(\infty)}(V_1^{(\infty)})].$$

Finally, observe that  $L_{(m)}(\cdot) = \mathbb{I}_{(m)}(\cdot) \bar{L}_{(m)}(\cdot)$  where  $\bar{L}_{(m)}(\mathbf{t}) = \exp(a \mathbb{E}_{\mathbf{p}}[\mathfrak{G}_{(m)}(V_1^{(m)})])$ , and recall that  $a\sigma(\mathbf{p}) \rightarrow \gamma$  (Assumption 4.4) and  $L_{(m)}(\mathcal{T}_m^{\mathbf{p}})$  is uniformly integrable (Proposition 4.8). Therefore, combining Lemma 4.14, Lemma 4.13 and (4.35) with Theorem 4.15 stated below yields (4.27) and thus completes the proof of Theorem 4.5.

**Theorem 4.15.** *For each  $k \geq 0$ ,*

$$\begin{aligned} & \left( \mathbb{E}_{\mathbf{p}} \left[ \frac{\mathfrak{G}_{(m)}(V_1^{(m)})}{\sigma(\mathbf{p})} \right], \mathbb{E}_{\mathbf{p}} \left[ \left( \prod_{i=1}^k \frac{\mathfrak{G}_{(m)}(V_i^{(m)})}{\sigma(\mathbf{p})} \right) g_{\phi}^{(k)} \left( \sigma(\mathbf{p}) \mathcal{T}_m^{\mathbf{p}}(\mathbf{V}_{k,k+\ell}^{(m)}) \right) \right] \right) \xrightarrow{d} \\ & \left( \mathbb{E}_{\theta}[\mathfrak{G}_{(\infty)}(V_1^{(\infty)})], \mathbb{E}_{\theta} \left[ \left( \prod_{i=1}^k \mathfrak{G}_{(\infty)}(V_i^{(\infty)}) \right) g_{\phi}^{(k)} \left( \mathcal{T}_{(\infty)}^{\theta}(\mathbf{V}_{k,k+\ell}^{(\infty)}) \right) \right] \right). \end{aligned} \quad (4.36)$$

The proof of this theorem is accomplished via the following two theorems for which we need to set up some notation. Fix  $I \geq 0$  and  $J \geq 1$ . We will assume that  $\mathcal{T}_m^{\mathbf{p}}$  has been constructed via the birthday construction (see Section 4.2.1). This construction gives rise to an unordered  $\mathbf{p}$ -tree. To obtain an ordered  $\mathbf{p}$ -tree from this, let  $\mathcal{D}_{(m)}(i)$  denote the set of children of  $i$  in the  $\mathbf{p}$ -tree for every vertex  $i$ . Generate i.i.d. uniform random variables  $\mathbf{U}_{(m)}(i) := \{U_{(m),i}(v) : v \in \mathcal{D}_{(m)}(i)\}$ , independent across  $v \in \mathcal{T}_m^{\mathbf{p}}$ . Think of these as “ages” of the children and arrange the children from left to right in decreasing order of their ages. We can construct the function  $\mathfrak{G}_{(m)}(\cdot)$  as in (4.22) once this ordering has been defined.

Now recall that the right hand side of (4.9) tells us how to sample  $J$  i.i.d. points  $(V_1^{(m)}, \dots, V_J^{(m)})$  from distribution  $\mathbf{p}$  and the corresponding spanning subtree  $\mathcal{T}_J^{\mathcal{B}}$  from the tree using the repeat time sequence  $\{R_k^{(m)} : k \geq 1\}$ . Thus, by the  $J$ th repeat time  $R_J$ , we would have sampled all  $J$  vertices  $V_i^{(m)} = Y_{R_i-1}$ . View  $\mathcal{T}_J^{\mathcal{B}}$  as a tree with edge lengths and marked vertices as follows: **(a)** rescale every edge to have length  $\sigma(\mathbf{p})$ ; **(b)** relabel  $V_j$  as  $j+$  and the root as  $0+$ ; **(c)** mark only those vertices  $i \leq I$  which occur in  $\mathcal{T}_J^{\mathcal{B}}$ ; **(d)** for all  $1 \leq j \leq J$ , set the leaf values to be  $\mathfrak{G}_{(m)}(V_j)/\sigma(\mathbf{p})$ , and assign the measure  $\nu_j^{(m)} := Q_{V_j}^{(m)}$  as defined in (4.26) to the path connecting the root to  $V_j$ , i.e, to the path  $[0+, j+]$ .

**Definition 4.16.** Fix  $I \geq 0, J \geq 1$  and consider the tree constructed as above. Set  $r_{IJ}^{(m)} = \mathcal{R}_{IJ}^{(m)} = \partial$  if some  $j+$  is not a leaf or if some leaf has been multiply labeled. Otherwise, write  $r_{IJ}^{(m)} \in \mathbf{T}_{IJ}$  for the tree with edge lengths and at most  $I$  labelled hubs, namely where we retain information in (a) and (b) above. Write  $\mathcal{R}_{IJ}^{(m)} \in \mathbf{T}_{IJ}^*$  for the tree where we retain all information (a)-(d) above, namely the leaf values  $\mathfrak{G}_{(m)}(V_j)$  and the root-to-leaf probability measures  $Q_{V_j}^{(m)}(\cdot)$  in addition to (a) and (b).

Now recall the tree  $\mathcal{R}_{IJ}^{(\infty)}$  defined in Section 2.2 using the limit ICRT  $\mathcal{T}_{(\infty)}^{\theta}$ . The main ingredients in the proof of Theorem 4.15 are the following two theorems:

**Theorem 4.17.** Under Assumption 4.4,  $\mathcal{R}_{IJ}^{(m)} \xrightarrow{d} \mathcal{R}_{IJ}^{(\infty)}$  as  $m \rightarrow \infty$  for every fixed  $I \geq 0$  and  $J \geq 1$ . This convergence is with respect to the topology defined on  $\mathbf{T}_{IJ}^*$  in Section 2.1.3.

The second result we will need is as follows. Recall the function  $g_{\phi}^{(k)}$  on  $\mathbf{T}_{I,(k+\ell)}^*$  as in (4.29).

**Theorem 4.18.** Fix  $I \geq 0, k \geq 0, \ell \geq 2$  and a bounded continuous function  $\phi$  on  $\mathbb{R}^{\ell^2}$ . Then the function  $g_{\phi}^{(k)}$  is continuous on  $\mathbf{T}_{I,(k+\ell)}^*$ .

**Proof of Theorem 4.15:** Assuming Theorem 4.17 and 4.18, let us now show how this completes the proof. Getting a handle directly on the conditional expectations as required in Theorem 4.15 is a little tricky. Naturally, conditional on  $\mathcal{T}_m^{\mathbf{p}}$ , repeated sampling of vertices and calculating sample averages should give a good idea of the conditional expectations (and the same for the limit object  $\mathcal{T}_{(\infty)}^{\theta}$ ). This is made precise in the following simple lemma whose proof we leave to the reader.

**Lemma 4.19.** Suppose  $\mathbf{X}^{(m)} := (X^{(m),1}, X^{(m),2})$  with  $m \in \{1, 2, \dots\} \cup \{\infty\}$  is a sequence of  $\mathbb{R}^2$ -valued random variables such that for each fixed  $r \geq 1$ , there exist random variables  $\mathbf{X}_r^{(m)} := (X_r^{(m),1}, X_r^{(m),2})$  such that the following hold:

- (i) There exists a constant  $C < \infty$  such that for any  $m \in \{1, 2, \dots\} \cup \{\infty\}$ ,  $r \geq 1$  and  $\varepsilon > 0$ ,

$$\max_{s=1,2} \mathbb{P}(|X^{(m),s} - X_r^{(m),s}| > \varepsilon) \leq \frac{C}{\varepsilon^2 r}.$$

- (ii) For each fixed  $r \geq 1$ ,  $\mathbf{X}_r^{(m)} \xrightarrow{d} \mathbf{X}_r^{(\infty)}$ .

Then  $\mathbf{X}^{(m)} \xrightarrow{d} \mathbf{X}^{(\infty)}$ .

We will apply this lemma with the random variables that arise in Theorem 4.15. That is, we set

$$X^{(m),1} := \mathbb{E}_{\mathbf{p}} \left[ \frac{\mathfrak{G}_{(m)}(V^{(m)})}{\sigma(\mathbf{p})} \right], \quad X^{(\infty),1} := \mathbb{E}_{\theta} [\mathfrak{G}_{(\infty)}(V^{(\infty)})], \quad (4.37)$$

and similarly define  $X^{(m),2}$  and  $X^{(\infty),2}$  to be the second coordinates in the display (4.36). To define  $\mathbf{X}_r^{(m)}$ , we proceed as follows. For each fixed  $r \geq 1$ , sample a collection of  $J_r := [r + (k + \ell)r]$  points all i.i.d.  $\mathbf{p}$  from  $\mathcal{T}_m^{\mathbf{p}}$  and think of them as  $r$  individuals points- $(V_1^{(m)}, V_2^{(m)}, \dots, V_r^{(m)})$ , and  $r(k + \ell)$  dimensional vectors- $\mathbf{V}_{k,k+\ell}^{(m),i} := (V_{i1}^{(m)}, \dots, V_{i(k+\ell)}^{(m)})$  for  $1 \leq i \leq r$ . Define

$$H_\phi^{(m)}(i) := \prod_{j=1}^k \frac{\mathfrak{G}_{(m)}(V_{ij}^{(m)})}{\sigma(\mathbf{p})} g_\phi^{(k)}\left(\sigma(\mathbf{p}) \mathcal{T}_m^{\mathbf{p}}(\mathbf{V}_{k,k+\ell}^{(m),i})\right), \quad \text{for } 1 \leq i \leq r.$$

For  $m = \infty$ , sample as above  $J_r$  points using the mass measure  $\mu$  from  $\mathcal{T}_{(\infty)}^\theta$  and define

$$H_\phi^{(\infty)}(i) := \prod_{j=1}^k \mathfrak{G}_{(\infty)}(V_{ij}^{(\infty)}) g_\phi^{(k)}\left(\mathcal{T}_{(\infty)}^\theta(\mathbf{V}_{k,k+\ell}^{(\infty),i})\right), \quad \text{for } 1 \leq i \leq r.$$

Now define

$$\begin{aligned} X_r^{(m),1} &:= \frac{\sum_{i=1}^r \mathfrak{G}_{(m)}(V_i^{(m)})}{r \sigma(\mathbf{p})} \text{ for } m \in \{1, 2, \dots\}, & X_r^{(\infty),1} &:= \frac{\sum_{i=1}^r \mathfrak{G}_{(\infty)}(V_i^{(\infty)})}{r}, \text{ and} \\ X_r^{(m),2} &:= \frac{\sum_{i=1}^r H_\phi^{(m)}(i)}{r} \text{ for } m \in \{1, 2, \dots\} \cup \{\infty\}. \end{aligned} \quad (4.38)$$

Let  $\mathbf{X}_r^{(m)} := (X_r^{(m),1}, X_r^{(m),2})$  for  $m \in \{1, 2, \dots\} \cup \{\infty\}$ . To complete the proof of the theorem, we have to check the two conditions of Lemma 4.19. Let us check condition (i) of Lemma 4.19 for the first coordinate. The second coordinate can be handled in an identical fashion.

Applying Chebyshev's inequality conditional on  $\mathcal{T}_m^{\mathbf{p}}$  and then taking expectations, we get

$$\mathbb{P}(|X^{(m),1} - X_r^{(m),1}| > \varepsilon) \leq (\varepsilon^2 r)^{-1} \mathbb{E}(\text{Var}_{\mathbf{p}}(\mathfrak{G}_{(m)}(V_1)/\sigma(\mathbf{p}))) =: (\varepsilon^2 r)^{-1} C_{(m)}, \text{ say,}$$

where  $\text{Var}_{\mathbf{p}}$  defined analogously to  $\mathbb{E}_{\mathbf{p}}$  is the conditional variance operator. Obviously

$$\mathbb{E}\left(\text{Var}_{\mathbf{p}}\left(\frac{\mathfrak{G}_{(m)}(V_1)}{\sigma(\mathbf{p})}\right)\right) \leq \text{Var}\left(\frac{\mathfrak{G}_{(m)}(V_1)}{\sigma(\mathbf{p})}\right) \leq \mathbb{E}\left(\left(\frac{\mathfrak{G}_{(m)}(V_1)}{\sigma(\mathbf{p})}\right)^2\right).$$

From the argument given below (4.13), it follows that  $\|\mathfrak{G}_{(m)}\|_\infty \leq \|F^{\text{exc}, \mathbf{p}}\|_\infty$ . Hence Lemma 4.9 implies that  $\sup_m C_{(m)} < \infty$ . This verifies (i) of the lemma.

Let us now verify condition (ii) of the lemma. Writing this out explicitly, we have to show for each fixed  $r \geq 1$ ,

$$\left(\frac{\sum_{i=1}^r \mathfrak{G}_{(m)}(V_i^{(m)})}{r \sigma(\mathbf{p})}, \frac{\sum_{i=1}^r H_\phi^{(m)}(i)}{r}\right) \xrightarrow{d} \left(\frac{\sum_{i=1}^r \mathfrak{G}_{(\infty)}(V_i^{(\infty)})}{r}, \frac{\sum_{i=1}^r H_\phi^{(\infty)}(i)}{r}\right). \quad (4.39)$$

To this end, for each  $m \in \{1, 2, \dots\} \cup \{\infty\}$ , consider the subtree spanning the  $J_r$  points  $(V_i^{(m)})_{1 \leq i \leq r}, (\mathbf{V}_{k,k+\ell}^{(m),i})_{1 \leq i \leq r}$ , viewed as an element of  $\mathbf{T}_{JJ}^*$  as in Definition 4.16. Using Theorem 4.17 and continuity of the function  $g_\phi^{(k)}$  from Theorem 4.18, we get

$$\left(\left(\frac{\mathfrak{G}_{(m)}(V_i)}{\sigma(\mathbf{p})}\right)_{1 \leq i \leq r}, \left(H_\phi^{(m)}(i)\right)_{1 \leq i \leq r}\right) \xrightarrow{d} \left(\left(\mathfrak{G}_{(\infty)}(V_i)\right)_{1 \leq i \leq r}, \left(H_\phi^{(\infty)}(i)\right)_{1 \leq i \leq r}\right)$$

with respect to weak convergence on  $\mathbb{R}^{2r}$ , which in turn implies (4.39). This completes the verification of the conditions of Lemma 4.19 and thus the proof of Theorem 4.15.  $\blacksquare$

The rest of this section proves Theorems 4.17 and 4.18.

**Proof of Theorem 4.17:** The proof will rely on a truncation argument that is qualitatively similar to Lemma 4.19. Fix a truncation level  $R \geq 1$ . Recall the definition of  $\mathfrak{G}_{(m)}(v)$  from (4.22) which

kept track of the contribution of **all** right children of individuals  $i$  on the path  $[\rho, v]$ . We will look at a truncated version of this object where we keep track of the potential contributions of only the first  $R$  vertices. More precisely let

$$\mathfrak{G}_{(m)}^R(v) := \sum_{\substack{i \in [\rho, v] \\ i \leq R}} \sum_{j \in [m]} p_j \mathbb{1}\{j \in \mathcal{RC}(i, [\rho, v])\}. \quad (4.40)$$

Let  $\mathfrak{G}_{(\infty)}^R(\cdot)$  be the analogous modification of  $\mathfrak{G}_{(\infty)}(\cdot)$  defined in (2.10), i.e.,

$$\mathfrak{G}_{(\infty)}^R(v) = \sum_{i \leq R} \theta_i \left[ \sum_{j \geq 1} U_j^{(i)} \times \mathbb{1}\{v \in \mathcal{T}_j^{(i)}\} \right]. \quad (4.41)$$

Similarly modify the “second endpoint” measure in (4.26) to keep track of only ancestors with labels  $\leq R$ , namely

$$Q_v^{(m),R}(y) := \begin{cases} \sum_u p_u \mathbb{1}\{u \in \mathcal{RC}(y, [\rho, v])\} / \mathfrak{G}_{(m)}^R(v), & \text{if } y \in [\rho, v] \text{ and } y \leq R, \\ 0, & \text{otherwise.} \end{cases} \quad (4.42)$$

Note that this does not make sense if  $\mathfrak{G}_{(m)}^R(v) = 0$ , i.e., when there is **no** vertex with label  $\leq R$  on the path from the root to  $v$ . In this case we follow the convention of defining the measure to be the uniform probability measure on the line  $[\rho, v]$ . Define  $Q_v^{(\infty),R}(\cdot)$  on  $\mathcal{T}_{(\infty)}^\theta$  in an analogous fashion.

Consider the tree  $r_{IJ}^{(m)}$  as in Definition 4.16, and assign to leaf  $V_j$  the truncated measure  $Q_{V_j}^{(m),R}(\cdot)$  and leaf value  $\mathfrak{G}_{(m)}^R(V_j)$  (instead of  $Q_{V_j}^{(m)}(\cdot)$  and  $\mathfrak{G}_{(m)}(V_j)/\sigma(\mathbf{p})$ ). We denote the resulting object (which is an element of  $\mathbf{T}_{IJ}^*$ ) by  $\mathcal{R}_{IJ}^{(m),R}$ . Similarly construct  $\mathcal{R}_{IJ}^{(\infty),R}$ .

**Proposition 4.20.** *The following hold:*

- (a) For all  $R \geq 1$ ,  $\mathcal{R}_{IJ}^{(m),R} \xrightarrow{d} \mathcal{R}_{IJ}^{(\infty),R}$ .
- (b)  $\mathcal{R}_{IJ}^{(\infty),R} \xrightarrow{d} \mathcal{R}_{IJ}^{(\infty)}$  as  $R \rightarrow \infty$ .
- (c) For any bounded continuous function  $f : \mathbf{T}_{IJ}^* \rightarrow \mathbb{R}$ ,

$$\limsup_{R \rightarrow \infty} \limsup_{m \rightarrow \infty} \left| \mathbb{E}(f(\mathcal{R}_{IJ}^{(m),R})) - \mathbb{E}(f(\mathcal{R}_{IJ}^{(m)})) \right| = 0.$$

Assuming this proposition, we now complete the proof of Theorem 4.17. Note that for any fixed bounded continuous function  $f$  on  $\mathbf{T}_{IJ}^*$  and any truncation level  $R \geq 1$ , we have

$$\begin{aligned} |\mathbb{E}(f(\mathcal{R}_{IJ}^{(\infty)})) - \mathbb{E}(f(\mathcal{R}_{IJ}^{(m)}))| &\leq |\mathbb{E}(f(\mathcal{R}_{IJ}^{(\infty)})) - \mathbb{E}(f(\mathcal{R}_{IJ}^{(\infty),R}))| + |\mathbb{E}(f(\mathcal{R}_{IJ}^{(\infty),R})) - \mathbb{E}(f(\mathcal{R}_{IJ}^{(m),R}))| \\ &\quad + |\mathbb{E}(f(\mathcal{R}_{IJ}^{(m),R})) - \mathbb{E}(f(\mathcal{R}_{IJ}^{(m)}))|. \end{aligned}$$

Now letting  $m \rightarrow \infty$  and then letting  $R \rightarrow \infty$  and using Proposition 4.20 completes the proof. ■

We next prove Proposition 4.20.

**4.6. Proof of Proposition 4.20.** We start with three preliminary lemmas. Recall that  $\{i \rightsquigarrow j\}$  denotes the event that  $j$  is a child of  $i$  in  $\mathcal{T}_m^{\mathbf{p}}$ .

**Lemma 4.21.** *Under Assumption 4.4, for each fixed  $i \geq 1$ ,*

$$\left| \frac{\sum_{j \in [m]} p_j \mathbb{1}\{i \rightsquigarrow j\}}{\sigma(\mathbf{p})} - \frac{p_i}{\sigma(\mathbf{p})} \right| \xrightarrow{\mathbf{p}} 0 \quad \text{as } m \rightarrow \infty.$$

**Proof:** Recall from (4.34) that for fixed  $i$ , the collection of events  $\{\{i \rightsquigarrow j\} : j \neq i\}$  are pairwise independent and have the same probability  $p_i$ . Thus

$$\mathbb{E} \left( \frac{\sum_{j \in [m]} p_j \mathbb{1}\{i \rightsquigarrow j\}}{\sigma(\mathbf{p})} \right) = \frac{p_i}{\sigma(\mathbf{p})} \left( \sum_{j \neq i} p_j \right) = (1 - p_i) \frac{p_i}{\sigma(\mathbf{p})},$$

and

$$\text{Var} \left( \frac{\sum_{j \in [m]} p_j \mathbb{1}\{i \rightsquigarrow j\}}{\sigma(\mathbf{p})} \right) = \sum_{j \in [m]} \frac{p_j^2}{\sigma^2(\mathbf{p})} \text{Var}(\mathbb{1}\{i \rightsquigarrow j\}) \leq p_i.$$

This completes the proof as  $p_{\max} = p_1 \rightarrow 0$  and  $p_i/\sigma(\mathbf{p}) \rightarrow \theta_i$  under Assumption 4.4.  $\blacksquare$

**Lemma 4.22.** *Under Assumption 4.4, for each fixed  $i \geq 1$ ,*

$$\max_{j: i \rightsquigarrow j} \frac{p_j}{\sigma(\mathbf{p})} \xrightarrow{\mathbf{P}} 0, \quad \text{as } m \rightarrow \infty.$$

**Proof:** Fix  $\varepsilon > 0$  and write

$$\mathcal{N}_\varepsilon(m) := \{j : p_j \geq \sigma(\mathbf{p})\varepsilon\} \text{ and } n_\varepsilon(m) = |\mathcal{N}_\varepsilon(m)|.$$

Note that by Assumption 4.4, for every  $\varepsilon > 0$ ,  $\{n_\varepsilon(m) : m \geq 1\}$  is a bounded sequence. Further, (4.34) and Markov's inequality yield

$$\mathbb{P} \left( \max_{j: i \rightsquigarrow j} \frac{p_j}{\sigma(\mathbf{p})} > \varepsilon \right) \leq \sum_{j \in \mathcal{N}_\varepsilon(m)} p_i = n_\varepsilon(m) p_i \rightarrow 0,$$

as  $p_{\max} = p_1 \rightarrow 0$ .  $\blacksquare$

Recall that  $\mathcal{D}_m(i)$  is the set of children of vertex  $i$  in  $\mathcal{T}_m^{\mathbf{p}}$ . For later use let  $d_m(i) := |\mathcal{D}_m(i)|$  denote the degree of  $i$  in  $\mathcal{T}_m^{\mathbf{p}}$ . Note that Lemma 4.21 together with the lemma just proven gives

$$d_m(i) \xrightarrow{\mathbf{P}} \infty, \quad \text{as } m \rightarrow \infty. \quad (4.43)$$

**Lemma 4.23.** *For each fixed  $m$ , let  $\mathbf{q}(m) := (q_1, q_2, \dots, q_d)$  be a probability mass function with  $q_i > 0$  for all  $i, m$  where  $d = d(m) \rightarrow \infty$  as  $m \uparrow \infty$ . Assume further that  $q_{\max} := \max_{i \in [d]} q_i \rightarrow 0$  as  $m \rightarrow \infty$ . Let  $\{U_i^{(m)} : 1 \leq i \leq d\}$  be i.i.d. uniform random variables and consider the function*

$$W_m(t) := \sum_{i=1}^d q_i \mathbb{1}\{U_i^{(m)} \leq t\} - t, \quad t \in [0, 1].$$

*Then  $\sup_{t \in [0, 1]} |W_m(t)| \xrightarrow{\mathbf{P}} 0$  as  $m \rightarrow \infty$ .*

**Proof:** Recall the proof of Lemma 4.9 when we studied the tightness of the tilt. Then replacing  $\mathbf{p}$  in the proof by  $\mathbf{q}$ , the quantity of interest  $\sup_{t \in [0, 1]} |W_m(t)| = \sigma(\mathbf{q}) \mathcal{R}_1(m)$  where  $\mathcal{R}_1(m)$  is as defined in (4.17) and  $\sigma(\mathbf{q}) := \sqrt{\sum_i q_i^2}$ . Now (4.17) and (4.18) imply the existence of a constant  $C$  (independent of  $m$ ) such that for all  $m$  and  $x \geq e$ ,

$$\mathbb{P} \left( \sup_{t \in [0, 1]} |W_m(t)| > x \sigma(\mathbf{q}) \right) \leq \exp(-Cx \log(\log x)).$$

Since  $\sigma(\mathbf{q}) \leq \sqrt{q_{\max}} \rightarrow 0$  as  $m \rightarrow \infty$ , this completes the proof.  $\blacksquare$

We now have all the ingredients for the proof of Proposition 4.20. We prove parts (a), (b) and (c) one by one.

**Proof of Proposition 4.20 (a):** Recall from Definition 4.16 the tree  $r_{IJ}^{(m)}$  that contains all the edge lengths and hub information in  $\mathcal{R}_{IJ}^{(m)}$  but ignores root-to-leaf measures and lead values  $\mathfrak{G}_{(m)}(\cdot)$ . By [27, Corollary 15] or [13, Proposition 3], for fixed  $J \geq 1$ , we have

$$\left(r_{IJ}^{(m)} : I' \geq 0\right) \xrightarrow{w} \left(r_{IJ}^{(\infty)} : I' \geq 0\right) \quad (4.44)$$

with respect to the product topology on  $\prod_{I' \geq 0} \mathbf{T}_{I'J}$ . Using Lemma 4.21, Lemma 4.22 and Skorohod embedding, we assume that we are working on a probability space that supports a sequence of unordered  $\mathbf{p}$ -trees  $\{\mathcal{T}_m^{\mathbf{p}, \text{uo}} : m \geq 1\}$ , sampled vertices  $\{V_j^{(m)} : 1 \leq j \leq J, m \geq 1\}$  using the associated sequence of probability mass functions  $\{\mathbf{p}(m) : m \geq 1\}$ , an ICRT  $\mathcal{T}_{(\infty)}^{\theta}$ , and sampled vertices  $\{V_j^{(\infty)} : 1 \leq j \leq J\}$  using the mass measure such that the following hold:

(A) Convergence in (4.44) happens almost surely:

$$\left(r_{IJ}^{(m)} : I' \geq 0\right) \xrightarrow{\text{a.s.}} \left(r_{IJ}^{(\infty)} : I' \geq 0\right) \text{ as } m \rightarrow \infty \quad (4.45)$$

coordinatewise, where the underlying tree corresponding to  $r_{IJ}^{(m)}$  is spanned by the root of  $\mathcal{T}_m^{\mathbf{p}, \text{uo}}$  and  $V_j^{(m)}$ ,  $1 \leq j \leq J$ .

(B) Writing  $s_m(i) := \sum_{v \in \mathcal{D}_m(i)} p_v$  for the sum of weights of children of  $i$  in  $\mathcal{T}_m^{\mathbf{p}, \text{uo}}$ , we have

$$\left(\frac{s_m(i)}{\sigma(\mathbf{p})} : i \geq 1\right) \xrightarrow{\text{a.s.}} (\theta_i : i \geq 1) \quad (4.46)$$

coordinatewise. (We can assume that this holds because of Lemma 4.21.)

(C) For fixed hub  $i \geq 1$  and  $m \geq 1$ , write

$$q_{m,i}(v) := \frac{p_v}{s_m(i)}, \quad v \in \mathcal{D}_m(i), \quad q_{m,i}^{\max} := \max_{v \in \mathcal{D}_m(i)} q_{m,i}(v). \quad (4.47)$$

Then we assume (using Lemma 4.22 and (4.43)) that for all  $i \geq 1$

$$q_{m,i}^{\max} \xrightarrow{\text{a.s.}} 0 \text{ and } d_m(i) \xrightarrow{\text{a.s.}} \infty.$$

Now, for each  $z \in [m]$  and  $i \geq 1$ , if  $i \in [\rho, z]$  (where  $\rho = \rho_m$  is the root of  $\mathcal{T}_m^{\mathbf{p}, \text{uo}}$ ), write  $c(i; z) \in \mathcal{D}_m(i)$  for the child of  $i$  that is the ancestor of  $z$ . Next, construct a collection  $\{U_{m,i}(v) : m \geq 1, i \geq 1, v \in [m]\}$  of uniform  $[0, 1]$  random variables on the same space such that

- (a)  $\{\mathcal{T}_m^{\mathbf{p}, \text{uo}}, U_{m,i}(v) : i \geq 1, v \in [m]\}$  are jointly independent for each  $m \geq 1$ ; and
- (b) for each  $i \leq R$  and  $j \leq J$  for which  $i \in [\rho, V_j^{(\infty)}]$ ,  $U_{m,i}(c(i; V_j^{(m)}))$  is a constant sequence (in  $m$ ) eventually.

As described below Theorem 4.15, we can use these uniform random variables to generate the sequence of ordered  $\mathbf{p}$ -trees  $\{\mathcal{T}_m^{\mathbf{p}}\}$  from  $\{\mathcal{T}_m^{\mathbf{p}, \text{uo}}\}$  as follows: Let  $\mathbf{U}_{m,i} := \{U_{m,i}(v) : v \in \mathcal{D}_m(i)\}$ . Think of these as “ages” of the children and arrange the children from left to right in decreasing order of their ages.

Once this ordering has been defined, we can construct the function  $\mathfrak{G}_{(m)}(\cdot)$  as in (4.22). In this case we can write this function explicitly in terms of the associated uniform random variables as follows. Define

$$\begin{aligned} \mathbb{O}_{(m),i}(z) &:= \frac{1}{\sigma(\mathbf{p})} \mathbb{1}\{i \in [\rho, z]\} \sum_{v \in \mathcal{D}_m(i)} p_v \mathbb{1}\{U_{m,i}(v) < U_{m,i}(c(i; z))\} \\ &= \mathbb{1}\{i \in [\rho, z]\} \frac{s_m(i)}{\sigma(\mathbf{p})} \sum_{v \in \mathcal{D}_m(i)} q_{m,i}(v) \mathbb{1}\{U_{m,i}(v) < U_{m,i}(c(i; z))\}. \end{aligned}$$

Then

$$(\sigma(\mathbf{p}))^{-1} \mathfrak{G}_{(m)}^R(z) = \sum_{i \leq R} \mathcal{O}_{(m),i}(z). \quad (4.48)$$

Similarly, the root-to-leaf measure  $Q_v^{(m),R}$  (recall (4.40)) can also be expressed in terms of this function.

Now using (4.45), for every fixed hub  $i \leq R$ ,  $j \leq J$ , and a.e. sample point  $\omega$ , one of the following two holds:

- (a)  $i \notin [\rho, V_j^{(\infty)}]$ , in which case there exists  $m = m(\omega)$  such that  $i \notin [\rho, V_j^{(m)}]$  for all  $m > m(\omega)$ .
- (b)  $i \in [\rho, V_j^{(\infty)}]$ , in which case there exists  $m = m(\omega)$  such that  $i \in [\rho, V_j^{(m)}]$  for all  $m > m(\omega)$ .

When the latter happens, using Lemma 4.23 together with (4.46) and (4.47), we get

$$\left| \mathcal{O}_{(m),i}(V_j^{(m)}) - \theta_i U_{m,i}(c(i; V_j^{(m)})) \right| \xrightarrow{\mathbb{P}} 0 \quad \text{as } m \rightarrow \infty.$$

By construction,  $U_{m,i}(c(i; V_j^{(m)}))$  is eventually constant in  $m$  on the event  $\{i \in [\rho, V_j^{(\infty)}]\}$ . This immediately implies convergence of the (scaled) truncated leaf values  $\mathfrak{G}_{(m)}^R(V_j^{(m)})/\sigma(\mathbf{p})$  (see (4.48)) for  $1 \leq j \leq J$ , and similarly the truncated root to leaf measures  $Q_{V_j^{(m)}}^{(m),R}$  jointly with the convergence in (4.45) and thus yields the convergence  $\mathcal{R}_{IJ}^{(m),R} \xrightarrow{d} \mathcal{R}_{IJ}^{(\infty),R}$ .  $\blacksquare$

**Proof of Proposition 4.20 (b):** Recall from Section 2.2 that  $\mathcal{R}_{IJ}^{(\infty)}$  is obtained by applying the stick-breaking construction to  $[0, \eta_J]$ , and leaf  $j+$  in  $\mathcal{R}_{IJ}^{(\infty)}$  corresponds to the vertex coming from  $\eta_j$ . It is easy to see from the definition of  $\mathfrak{G}_{(\infty)}^R$  and  $Q_v^{(\infty),R}$  that it suffices to prove

$$0 \leq \mathcal{E}_R^{(1)} := \sum_{j=1}^J (\mathfrak{G}_{(\infty)}(\eta_j) - \mathfrak{G}_{(\infty)}^R(\eta_j)) \xrightarrow{\mathbb{P}} 0, \quad \text{as } R \rightarrow \infty.$$

For every hub  $i \geq 1$  and leaf  $\eta_j$ , write  $\{i \rightarrow \eta_j\}$  if  $\eta_j$  is a descendant of  $i$  (namely  $i \in [\rho, \eta_j]$ ). Then note that

$$\mathcal{E}_R^{(1)} = \sum_{j=1}^J \sum_{i=R+1}^{\infty} \sum_{k=1}^{\infty} \theta_i U_k^{(i)} \mathbb{1}\{\eta_j \in \mathcal{T}_k^{(i)}\} \leq \sum_{j=1}^J \sum_{i=R+1}^{\infty} \theta_i \mathbb{1}\{i \rightarrow \eta_j\} := \mathcal{E}_R^{(2)}.$$

Thus, it is enough to show that given  $\varepsilon > 0$ , we can find  $R = R(\varepsilon) < \infty$  such that  $\mathbb{P}(\mathcal{E}_R^{(2)} > \varepsilon) < \varepsilon$ . To this end, first choose  $K_\varepsilon$  large enough so that  $\mathbb{P}(\eta_J > K_\varepsilon) < \varepsilon/2$ , and then choose  $R_\varepsilon$  large enough so that

$$\frac{JK_\varepsilon}{\varepsilon} \sum_{i=R_\varepsilon+1}^{\infty} \theta_i^2 < \varepsilon/2.$$

Then note that

$$\begin{aligned} \mathbb{P}(\mathcal{E}_{R_\varepsilon}^{(2)} > \varepsilon) &\leq \mathbb{P}(\eta_J > K_\varepsilon) + \mathbb{P}\left(J \sum_{i=R_\varepsilon+1}^{\infty} \theta_i \mathbb{1}\{i \text{th hub appears before time } K_\varepsilon\} > \varepsilon\right) \\ &\leq \frac{\varepsilon}{2} + \frac{J}{\varepsilon} \sum_{i=R_\varepsilon+1}^{\infty} \theta_i (1 - \exp(-\theta_i K_\varepsilon)) \\ &< \varepsilon \quad \text{by choice of } R_\varepsilon, \end{aligned}$$

where the first term in the second inequality follows from the choice of  $K_\varepsilon$ , while the second term comes from the stick-breaking construction of  $\mathcal{T}_{(\infty)}^\theta$  using the countable collection of Poisson point processes. This completes the proof.  $\blacksquare$

**Proof of Proposition 4.20(c):** Recall that the tree  $\mathcal{R}_{IJ}^{(m)}$  (and  $\mathcal{R}_{IJ}^{(m),R}$ ) can be thought of as being made up of  $2J + 1$  coordinates:

- (a) One coordinate for the shape and edge length information along with the labels smaller than  $I$  namely  $r_{IJ}^{(m)}$  (see Definition 4.16). Note that this is the same for both  $\mathcal{R}_{IJ}^{(m)}$  and  $\mathcal{R}_{IJ}^{(m),R}$ .
- (b)  $J$  coordinates for the leaf values  $\mathfrak{G}_{(m)}(V_j)/\sigma(\mathbf{p})$  (resp.  $\mathfrak{G}_{(m)}^R(V_j)/\sigma(\mathbf{p})$ ).
- (c)  $J$  coordinates for the measured metric spaces  $\mathcal{M}_j^{(m)} := ([\rho, V_j^{(m)}], Q_{V_j}^{(m)})$  (resp.  $\mathcal{M}_j^{(m),R} := ([\rho, V_j^{(m)}], Q_{V_j}^{(m),R})$ ).

Since  $\mathbf{T}_{IJ}^*$  assumes the product topology on these coordinates, it is enough to show the required estimate in Proposition 4.20 (c) with functions of the form

$$f(\mathbf{t}, (a_i)_{1 \leq j \leq J}, (M_j)_{1 \leq j \leq J}) := F(\mathbf{t}) \prod_{1 \leq j \leq J} g_j(a_i) \prod_{1 \leq j \leq J} h_j(M_j).$$

Here  $\mathbf{t} \in \mathbf{T}_{IJ}$ ,  $a_j \in \mathbb{R}$  are associated leaf values and  $M_j$  are the paths from the root to leaf  $j$  with an associated probability measure and  $f$ ,  $g_j$  and  $h_j$  are bounded *uniformly continuous* functions on the spaces  $\mathbf{T}_{IJ}$ ,  $\mathbb{R}$  and  $\mathcal{S}$  (measured compact metric spaces) respectively. To simplify notation, we will simply write this as  $f(\mathbf{t})$ .

Now we can go from  $\mathcal{R}_{IJ}^{(m)}$  to  $\mathcal{R}_{IJ}^{(m),R}$  by flipping one coordinate at a time. Thus writing

$$f_1^{(-i)}(\mathbf{t}) := F(\mathbf{t}) \prod_{\substack{1 \leq j \leq J \\ j \neq i}} g_j(a_i) \prod_{1 \leq j \leq J} h_j(M_j), \quad f_2^{(-i)}(\mathbf{t}) := F(\mathbf{t}) \prod_{1 \leq j \leq J} g_j(a_i) \prod_{\substack{1 \leq j \leq J \\ j \neq i}} h_j(M_j),$$

we get

$$\begin{aligned} |\mathbb{E}(f(\mathcal{R}_{IJ}^{(m),R})) - \mathbb{E}(f(\mathcal{R}_{IJ}^{(m)}))| &\leq \sum_{j=1}^J \|f_1^{(-j)}\|_\infty \mathbb{E} \left( \left| g_j \left( \frac{\mathfrak{G}_{(m)}(V_j)}{\sigma(\mathbf{p})} \right) - g_j \left( \frac{\mathfrak{G}_{(m)}^R(V_j)}{\sigma(\mathbf{p})} \right) \right| \right) \\ &\quad + \sum_{j=1}^J \|f_2^{(-j)}\|_\infty \mathbb{E} (|h_j(\mathcal{M}_j^{(m)}) - h_j(\mathcal{M}_j^{(m),R})|) \end{aligned} \quad (4.49)$$

Since  $V_j$ 's have been sampled in an i.i.d. fashion from  $\mathbf{p}$ , it is enough to show that for any two bounded uniformly continuous functions  $h, g$  on  $\mathbb{R}$  and  $\mathcal{S}$  respectively,

$$\limsup_{R \rightarrow \infty} \limsup_{m \rightarrow \infty} \mathbb{E} \left( \left| g \left( \frac{\mathfrak{G}_{(m)}(V_1^{(m)})}{\sigma(\mathbf{p})} \right) - g \left( \frac{\mathfrak{G}_{(m)}^R(V_1^{(m)})}{\sigma(\mathbf{p})} \right) \right| \right) = 0, \quad (4.50)$$

and

$$\limsup_{R \rightarrow \infty} \limsup_{m \rightarrow \infty} \mathbb{E} (|h(\mathcal{M}_1^{(m)}) - h(\mathcal{M}_1^{(m),R})|) = 0. \quad (4.51)$$

Now consider the measured metric spaces  $\mathcal{M}_1^{(m)}$  and  $\mathcal{M}_1^{(m),R}$ . As remarked above, they share the same metric space, namely the path  $[\rho, V_1^{(m)}]$ . The only difference is in the associated probability measures. Consider the natural correspondence  $C = \{(x, x) : x \in [\rho, V_1^{(m)}]\}$  between  $\mathcal{M}_1^{(m)}$  and  $\mathcal{M}_1^{(m),R}$ . Further, define a probability measure  $\pi$  on  $[\rho, V_1^{(m)}] \times [\rho, V_1^{(m)}]$  as

$$\pi(i, i) := \begin{cases} \sum_u p_u \mathbb{1}\{u \in \mathcal{R}^C(i, [\rho, V_1^{(m)}])\} / \mathfrak{G}_{(m)}(V_1^{(m)}), & \text{if } i \in [\rho, V_1^{(m)}] \text{ and } i \leq R, \\ [\mathfrak{G}_{(m)}(V_1^{(m)}) - \mathfrak{G}_{(m)}^R(V_1^{(m)})] / \mathfrak{G}_{(m)}(V_1^{(m)}), & \text{if } i = \rho. \end{cases} \quad (4.52)$$

Writing  $\pi_1$  and  $\pi_2$  for the marginals of  $\pi$ , we have, using the above choice of correspondence  $C$  and of the measure  $\pi$ ,

$$d_{\text{GHP}}^{\text{pt}}(\mathcal{M}_1^{(m)}, \mathcal{M}_1^{(m),R}) \leq (\|\pi_1 - Q_{V_1}^{(m)}\| + \|\pi_2 - Q_{V_1}^{(m),R}\|) \leq \frac{2[\mathfrak{G}_{(m)}(V_1) - \mathfrak{G}_{(m)}^R(V_1)]}{\mathfrak{G}_{(m)}(V_1)}. \quad (4.53)$$



Now suppose we show (4.50). Using part (a) and part (b) of Proposition 4.20, we get  $(\sigma(\mathbf{p}))^{-1} \mathfrak{G}_{(m)}(V_1^{(m)}) \xrightarrow{d} \mathfrak{G}_{(\infty)}(V_1^{(\infty)}) > 0$ . Now using the bound in (4.53) and uniform continuity of  $h$ , we see that (4.51) is true. Hence it is enough to prove (4.50).

Recall from Section 4.2.1 the construction of  $V_1^{(m)}$  and the tree simultaneously via the birthday construction, where  $V_1^{(m)}$  is obtained as the value before the first repeat time, namely  $Y_{R_1-1}$ . Fix  $\varepsilon > 0$ . By [27, Theorem 4], under Assumptions 4.4 we may choose  $K_\varepsilon$  large so that the first repeat time satisfies  $\mathbb{P}(R_1 > K_\varepsilon/\sigma(\mathbf{p})) < \varepsilon$  for all  $m \geq 1$ . Next, by uniform continuity of  $g$ , choose  $\delta \in (0, 1)$  such that  $|g(x) - g(y)| < \varepsilon$  if  $|x - y| < \delta$ . Finally choose  $R$  large so that for all  $m$ ,

$$\frac{K_\varepsilon^2}{\delta \wedge \varepsilon} \sum_{i=R+1}^m \frac{p_i^2}{\sigma^2(\mathbf{p})} < \varepsilon.$$

First, by choice of  $K_\varepsilon$  and boundedness of  $g$ ,

$$\left| \mathbb{E} \left[ g \left( \frac{\mathfrak{G}_{(m)}(V_1^{(m)})}{\sigma(\mathbf{p})} \right) \right] - \mathbb{E} \left[ g \left( \frac{\mathfrak{G}_{(m)}(V_1^{(m)})}{\sigma(\mathbf{p})} \right) \mathbb{1} \left\{ R_1 \leq \frac{K_\varepsilon}{\sigma(\mathbf{p})} \right\} \right] \right| \leq \|g\|_\infty \varepsilon, \quad (4.54)$$

and a similar inequality holds true if we replace the functional  $\mathfrak{G}_{(m)}$  by  $\mathfrak{G}_{(m)}^R$ . Next, writing

$$\mathcal{E}_{(m)}^{(1)}(R) := \left| \mathbb{E} \left[ \left\{ g \left( \frac{\mathfrak{G}_{(m)}(V_1^{(m)})}{\sigma(\mathbf{p})} \right) - g \left( \frac{\mathfrak{G}_{(m)}^R(V_1^{(m)})}{\sigma(\mathbf{p})} \right) \right\} \mathbb{1} \left\{ R_1 \leq \frac{K_\varepsilon}{\sigma(\mathbf{p})} \right\} \right] \right|,$$

we have

$$\mathcal{E}_{(m)}^{(1)}(R) \leq \varepsilon + 2\|g\|_\infty \mathbb{P} \left( R_1 \leq \frac{K_\varepsilon}{\sigma(\mathbf{p})}, \frac{(\mathfrak{G}_{(m)}(V_1^{(m)}) - \mathfrak{G}_{(m)}^R(V_1^{(m)}))}{\sigma(\mathbf{p})} \geq \delta \right) \quad (4.55)$$

by our choice of  $\delta$ . The difference  $\mathfrak{G}_{(m)}(V_1^{(m)}) - \mathfrak{G}_{(m)}^R(V_1^{(m)})$  is a tricky object for which we will need a tractable upper bound. Recall that we have used  $\mathcal{T}_1^{\mathcal{B}}$  for the birthday tree in (4.9) constructed by time  $R_1$ . For every vertex  $i \in \mathcal{T}_1^{\mathcal{B}}$ , let  $\mathcal{J}(i)$  be the first child of  $i$  in the birthday construction (the first **new, i.e., previously un-sampled** vertex sampled immediately after a prior sampling of  $i$ ). This will be an empty set if  $i$  is a leaf in the eventual full tree  $\mathcal{T}_m^{\mathbf{p}}$ . Recall that  $\{i \rightsquigarrow j\}$  was used to denote the event that  $j$  is a child of  $i$  in  $\mathcal{T}_m^{\mathbf{p}}$ . Then note that

$$\mathfrak{G}_{(m)}(V_1^{(m)}) - \mathfrak{G}_{(m)}^R(V_1^{(m)}) \leq \sum_{i \geq R+1} \mathbb{1} \{ i \in \mathcal{T}_1^{\mathcal{B}} \} \sum_{j \in [m]} p_j \mathbb{1} \{ i \rightsquigarrow j, j \neq \mathcal{J}(i) \}.$$

Thus,

$$\begin{aligned} & \mathbb{P} \left( R_1 \leq \frac{K_\varepsilon}{\sigma(\mathbf{p})}, \frac{(\mathfrak{G}_{(m)}(V_1^{(m)}) - \mathfrak{G}_{(m)}^R(V_1^{(m)}))}{\sigma(\mathbf{p})} \geq \delta \right) \\ & \leq \frac{1}{\delta} \sum_{i=R+1}^m \sum_{j \in [m]} \frac{p_j}{\sigma(\mathbf{p})} \mathbb{P} \left( i \text{ appears before } \frac{K_\varepsilon}{\sigma(\mathbf{p})}, i \rightsquigarrow j, j \neq \mathcal{J}(i) \right) =: \mathcal{E}_{(m)}^{(2)}(R). \end{aligned} \quad (4.56)$$

For  $i \neq j \in [m]$ , define the event  $E_{ij} := \left\{ i \text{ appears before } \frac{K_\varepsilon}{\sigma(\mathbf{p})}, i \rightsquigarrow j, j \neq \mathcal{J}(i) \right\}$ . Then for  $E_{ij}$  to happen, the following needs to happen in the birthday construction: (a) There is an  $0 \leq r_1 \leq K_\varepsilon/\sigma(\mathbf{p})$  such that till time  $r_1$ , neither  $i$  or  $j$  have been sampled. (b) At time  $r_1 + 1$  vertex  $i$  is sampled. (c) There is an  $r_2 \geq 0$  such that in the times  $[r_1 + 1, r_1 + 1 + r_2]$  samples,  $j$  does not appear. (d) Then at time  $r_1 + r_2 + 2$ , vertex  $i$  is sampled again. (e) In the next time step  $r_1 + r_2 + 3$

vertex  $j$  is sampled. Therefore,

$$\mathbb{P}(E_{ij}) \leq \sum_{r_1=0}^{K_\varepsilon/\sigma(\mathbf{p})} \sum_{r_2=0}^{\infty} (1-p_i-p_j)^{r_1} p_i (1-p_j)^{r_2} p_i p_j \leq p_i^2 \frac{K_\varepsilon}{\sigma(\mathbf{p})}.$$

Using this in (4.56), we get

$$\mathcal{E}_{(m)}^{(2)}(R) \leq \frac{K_\varepsilon}{\delta} \sum_{R+1}^m \frac{p_i^2}{[\sigma(\mathbf{p})]^2} \leq \varepsilon, \text{ by our choice of } R. \quad (4.57)$$

Combining (4.54), (4.55), (4.56) and (4.57) now gives the following lemma which completes the proof of (4.50) and thus the proof of part (c) of the proposition.  $\blacksquare$

**Lemma 4.24.** *Given  $\varepsilon > 0$  choose  $K_\varepsilon, \delta$  and  $R$  as above. Then, for all  $m \geq 1$ ,*

$$\left| \mathbb{E} \left[ g \left( \frac{\mathfrak{G}_{(m)}(V_1^{(m)})}{\sigma(\mathbf{p})} \right) \right] - \mathbb{E} \left[ g \left( \frac{\mathfrak{G}_{(m)}^R(V_1^{(m)})}{\sigma(\mathbf{p})} \right) \right] \right| \leq \varepsilon (4 \|g\|_\infty + 1).$$

**Proof of Theorem 4.18:** We now prove continuity of the function  $g_\phi^{(k)}$  on the space  $\mathbf{T}_{I,(k+\ell)}^*$ . In fact, we will give a quantitative estimate. Since we are assuming the discrete topology on the coordinate corresponding to the shape, without loss of generality we will work with two trees  $\mathbf{t}, \bar{\mathbf{t}} \in \mathbf{T}_{I,(k+\ell)}^*$  having the same shape. We need to distinguish the labels for the root and the leaves in the two trees; so write  $0+$  (respectively  $\bar{0}+$ ) for the root of  $\mathbf{t}$  (respectively  $\bar{\mathbf{t}}$ ) and write  $\{j+ : 1 \leq j \leq k+\ell\}$  (respectively  $\{\bar{j}+ : 1 \leq j \leq k+\ell\}$ ) for the collection of leaves in  $\mathbf{t}$  (respectively  $\bar{\mathbf{t}}$ ). Finally, let  $\nu_j$  be the corresponding probability measure on the path  $\mathcal{M}_j := [0+, j+]$  for  $1 \leq j \leq k$ , and analogously let  $\bar{\nu}_j$  be the probability measure on  $\bar{\mathcal{M}}_j := [\bar{0}+, \bar{j}+]$ . View these paths as pointed measured metric spaces pointed at the roots  $0+$  and  $\bar{0}+$  respectively. Now let  $\varepsilon_j := d_{\text{GHP}}^{\text{pt}}(\mathcal{M}_j, \bar{\mathcal{M}}_j)$ , where  $d_{\text{GHP}}^{\text{pt}}$  is the pointed Gromov-Hausdorff-Prokhorov metric defined in Section 2.1.

Write  $L = \binom{\ell}{2}$ . Let  $\phi : \mathbb{R}_+^L \rightarrow \mathbb{R}$  be a bounded continuous function. For  $K > 0$ , let  $\square(K) = [0, K]^L$ , and for  $\delta > 0$ , define

$$\text{osc}_\phi(\delta, K) := \sup_{\substack{\mathbf{x}, \mathbf{y} \in \square(K) \\ \|\mathbf{x} - \mathbf{y}\|_\infty < \delta}} |\phi(\mathbf{x}) - \phi(\mathbf{y})|.$$

Finally, define

$$\varepsilon := 4 \sum_{j=1}^k \varepsilon_j + (k+1) \sum_e |l_e(\mathbf{t}) - l_e(\bar{\mathbf{t}})|, \quad (4.58)$$

where  $l_e(\cdot)$  denotes the length of the edge  $e$  and we have used the fact that both trees have the same shape. Write  $\text{ht}(\mathbf{t})$  for the height of tree  $\mathbf{t}$  (not graph distance, rather in terms of maximal distance from the root when incorporating edge lengths). The following proposition completes the proof of Theorem 4.18:

**Proposition 4.25.** *For two trees  $\mathbf{t}, \bar{\mathbf{t}} \in \mathbf{T}_{I,(k+\ell)}^*$  having the same shape, and with  $\varepsilon$  as in (4.58),*

$$|g_\phi^{(k)}(\mathbf{t}) - g_\phi^{(k)}(\bar{\mathbf{t}})| \leq 2\varepsilon \|\phi\|_\infty + \text{osc}_\phi\left(\varepsilon, 2\text{ht}(\mathbf{t}) + 2\text{ht}(\bar{\mathbf{t}})\right).$$

**Proof:** For each  $j \leq k$ , choose a correspondence  $C_j$  and a measure  $\pi_j$  on the product space  $[0+, j+] \times [\bar{0}+, \bar{j}+]$  such that the following conditions are met: **(a)**  $(0+, \bar{0}+) \in C_j$ ; **(b)** the distortion satisfies  $\text{dis}(C_j) < 3\varepsilon_j$ ; **(c)** the measure of the complement satisfies  $\pi_j(C_j^c) < 2\varepsilon_j$ ; **(d)** and finally

$$\|v_j - p_*\pi_j\| + \|\bar{v}_j - \bar{p}_*\pi_j\| < 2\varepsilon_j, \quad (4.59)$$

where  $p_*\pi_j$  and  $\bar{p}_*\pi_j$  are the marginals of  $\pi_j$ . Now sample  $(X_j^*, \bar{X}_j^*) \sim \pi_j$  from  $[0+, j+] \times [\bar{0}+, \bar{j}+]$  independently for  $1 \leq j \leq k$ . By (4.59), we can couple  $(X_j^*, \bar{X}_j^*)$  with two random variables  $X_j, \bar{X}_j$  (again independently for  $1 \leq j \leq k$ ) such that  $X_j \sim v_j$  and  $\bar{X}_j \sim \bar{v}_j$ , and further

$$\mathbb{P}(X_j \neq X_j^*) + \mathbb{P}(\bar{X}_j \neq \bar{X}_j^*) < 2\varepsilon_j. \quad (4.60)$$

Using conditions (b) and (c), we get

$$\mathbb{P}\left(\left|d_{\mathbf{t}}(0+, X_j^*) - d_{\bar{\mathbf{t}}}(\bar{0}+, \bar{X}_j^*)\right| > 3\varepsilon_j\right) \leq 2\varepsilon_j, \quad (4.61)$$

where  $d_{\mathbf{t}}$  is the distance metric on tree  $\mathbf{t}$  which incorporates the edge lengths. Now write  $E$  for the following “good event”:

$$E := \bigcap_{j=1}^k \left\{ X_j = X_j^*, \bar{X}_j = \bar{X}_j^*, \left|d_{\mathbf{t}}(0+, X_j^*) - d_{\bar{\mathbf{t}}}(\bar{0}+, \bar{X}_j^*)\right| \leq 3\varepsilon_j \right\}.$$

It follows from (4.60) and (4.61) that

$$\mathbb{P}(E^c) \leq 4 \sum_{j=1}^k \varepsilon_j. \quad (4.62)$$

Now we are going to create “shortcuts” by gluing the leaves to the corresponding sampled points. Let  $\mathbf{S}$  (resp.  $\bar{\mathbf{S}}$ ) be the (random) metric space obtained by identifying each of the leaves  $j+$  (resp.  $\bar{j}+$ ) with  $X_j$  (resp.  $\bar{X}_j$ ) in  $\mathbf{t}$  (resp.  $\bar{\mathbf{t}}$ ) for  $1 \leq j \leq k$  and write  $d_{\mathbf{S}}$  (resp.  $d_{\bar{\mathbf{S}}}$ ) for the induced metric. Then by definition,

$$g_{\phi}^{(k)}(\mathbf{t}) = \mathbb{E} \left[ \phi \left( d_{\mathbf{S}}((k+i_1)+, (k+i_2)+) : 1 \leq i_1 < i_2 \leq \ell \right) \right],$$

and an analogous expression holds for  $g_{\phi}^{(k)}(\bar{\mathbf{t}})$ . This gives

$$\begin{aligned} |g_{\phi}^{(k)}(\mathbf{t}) - g_{\phi}^{(k)}(\bar{\mathbf{t}})| &\leq \mathbb{E} \left( \left| \phi \left( d_{\mathbf{S}}((k+i_1)+, (k+i_2)+) : 1 \leq i_1 < i_2 \leq \ell \right) \right. \right. \\ &\quad \left. \left. - \phi \left( d_{\bar{\mathbf{S}}}(\overline{(k+i_1)+}, \overline{(k+i_2)+}) : 1 \leq i_1 < i_2 \leq \ell \right) \right| \right). \end{aligned} \quad (4.63)$$

Consider the map from  $\mathbf{t}$  to  $\bar{\mathbf{t}}$  which takes every vertex to the corresponding vertex and points on each edge are mapped by linear interpolation (using the edge lengths) to points on the corresponding edge. Consider  $a \in [0, j+]$  and let  $\bar{a} \in [\bar{0}, \bar{j}+]$  be the corresponding point in  $\bar{\mathbf{t}}$  for some  $j \leq k$ . Then note that

$$\begin{aligned} \left| d_{\mathbf{t}}(a, X_j) - d_{\bar{\mathbf{t}}}(\bar{a}, \bar{X}_j) \right| &\leq \left| d_{\mathbf{t}}(0+, X_j) - d_{\bar{\mathbf{t}}}(\bar{0}+, \bar{X}_j) \right| + \left| d_{\mathbf{t}}(0+, a) - d_{\bar{\mathbf{t}}}(\bar{0}+, \bar{a}) \right| \\ &\leq 3\varepsilon_j + \sum_e |l_e(\mathbf{t}) - l_e(\bar{\mathbf{t}})| \end{aligned} \quad (4.64)$$

on the set  $E$ .

Now consider a shortest path in  $\mathbf{S}$  connecting  $(k+i_1)+$  and  $(k+i_2)+$ . We can go from  $\overline{(k+i_1)+}$  to  $\overline{(k+i_2)+}$  by taking the same route in  $\bar{\mathbf{S}}$ , i.e., by traversing the same edges and taking the same

shortcuts in the same order. We make the following observations: (i) The difference between distance traversed while crossing the edge  $e$  is  $|l_e(\mathbf{t}) - l_e(\bar{\mathbf{t}})|$ . (ii) By (4.64), on the set  $E$ , taking a “shortcut” contributes at most  $(3\varepsilon_j + \sum_e |l_e(\mathbf{t}) - l_e(\bar{\mathbf{t}})|)$  to the difference between distance traversed. Since we have to take at most  $k$  shortcuts, we immediately get

$$d_{\bar{\mathbf{S}}}(\overline{(k+i_1)+}, \overline{(k+i_2)+}) \leq d_{\mathbf{S}}((k+i_1)+, (k+i_2)+) + 3 \sum_{j=1}^k \varepsilon_j + (k+1) \sum_e |l_e(\mathbf{t}) - l_e(\bar{\mathbf{t}})|$$

on the set  $E$ . By symmetry, a similar inequality holds if we interchange the roles of  $\mathbf{S}$  and  $\bar{\mathbf{S}}$ . This observation combined with (4.62) and (4.63) yields the result.  $\blacksquare$

## 5. PROOFS: CONVERGENCE IN GROMOV-WEAK TOPOLOGY

Recall from Proposition 4.1 that conditional on the partition of the vertices  $\{\mathcal{V}^{(i)} : i \geq 1\}$  into the connected components, the actual structure of the components of  $\mathcal{G}(\mathbf{x}, t)$  can be generated independently as the connected graph  $\tilde{\mathcal{G}}_{|\mathcal{V}^{(i)}|}(a_n^{(i)}, \mathbf{p}_n^{(i)})$  where  $a_n^{(i)}, \mathbf{p}_n^{(i)}$  are as in Proposition 4.1 and given  $m, \mathbf{p}, a$ ,  $\tilde{\mathcal{G}}_m(a, \mathbf{p})$  is the connected random graph model studied in the previous section. For Theorem 1.7, the time scale  $t = t_n$  of interest in the expression of  $a_n^{(i)}$  is

$$t_n := \lambda + \frac{1}{\sigma_2(\mathbf{x}^{(m)})}, \quad (5.1)$$

for fixed  $\lambda \in \mathbb{R}$ . Let  $\mathcal{N}(\mathbb{R}_+)$  denote the space of counting measures on  $\mathbb{R}_+$  equipped with the vague topology. Define  $\mathbf{Y}_n^{(i)} := (p_v / \sigma(\mathbf{p}), v \in \mathcal{V}^{(i)})$  and view  $(a_n^{(i)} \sigma(\mathbf{p}_n^{(i)}), \mathbf{Y}_n^{(i)})$  as a random element of  $\mathbb{S} := \mathbb{R}_+ \times \mathcal{N}(\mathbb{R}_+)$  (equipped with the product topology). Finally, define

$$\mathcal{P}_n := ((a_n^{(i)} \sigma(\mathbf{p}_n^{(i)}), \mathbf{Y}_n^{(i)}) : i \geq 1)$$

viewed as an element of  $\mathbb{S}^\infty$ , again equipped with the product topology induced by a single coordinate  $\mathbb{S}$ . Now given an infinite vector  $\mathbf{c} \in l_0$  recall the process  $\tilde{V}_\lambda^{\mathbf{c}}(\cdot)$  as in (1.19), the corresponding excursions  $\mathcal{Z}(\lambda)$  as in (1.20) and the corresponding excursion lengths in (1.21). Finally recall the definitions of  $\tilde{\gamma}^{(i)}, \boldsymbol{\theta}^{(i)}$  from (2.13). Writing these out explicitly, define

$$\mathcal{P}_\infty := ((\tilde{\gamma}^{(i)}, \boldsymbol{\theta}^{(i)}) : i \geq 1) = \left( Z_i(\lambda) \sqrt{\sum_{v \in \mathcal{Z}_i(\lambda)} c_v^2}, \left( \frac{c_j}{\sqrt{\sum_{v \in \mathcal{Z}_i(\lambda)} c_v^2}} : j \in \mathcal{Z}_i(\lambda) \right) : i \geq 1 \right). \quad (5.2)$$

**Proposition 5.1.** *The following hold under Assumption 1.5:*

- (i) For every  $i \geq 1$ ,  $\sigma(\mathbf{p}_n^{(i)}) \xrightarrow{\mathbf{P}} 0$  as  $n \rightarrow \infty$ .
- (ii)  $\mathcal{P}_n \xrightarrow{d} \mathcal{P}_\infty$  on  $\mathbb{S}^\infty$  as  $n \rightarrow \infty$ . Further for every fixed  $i \geq 1$ , almost surely,

$$\sum_{v \in \mathcal{Z}_i(\lambda)} c_v = \infty. \quad (5.3)$$

**Proof of Theorem 1.7:** We prove the theorem assuming Proposition 5.1. By an application of Skorohod embedding we may assume that we are working on a probability space where the convergence in Proposition 5.1 happens almost surely. In particular, in this space, Assumption 4.4 is satisfied almost surely for  $\mathbf{p}_n^{(i)}$  for any fixed  $i \geq 1$ . Now an application of Theorem 4.5 completes the proof.  $\blacksquare$

**5.1. Verification of weight assumptions in maximal components.** Here we give the proof of Proposition 5.1. To ease notation, we will throughout assume  $\lambda = 0$ . The general case follows in an identical fashion, but this assumption simplifies notation. We will write  $V^c$  instead of  $V_0^c$  for the process in (1.18) with  $\lambda = 0$  and simply write  $\mathcal{C}_i$  for  $\mathcal{C}_i([\sigma_2(\mathbf{x}^{(n)})]^{-1})$ .

We start by describing an exploration scheme (developed in [9]) which simultaneously constructs the graph  $\mathcal{G}_n(\mathbf{x}, t)$  and a “breadth first” walk. This was carefully analyzed in [10] to prove Theorem 1.6.

For every ordered pair  $(u, v)$ , let  $\eta_{u,v}$  be an exponential random variable with rate  $tx_v$  (independent across ordered pairs). Note that there is a simple relation between the connection probabilities of  $\mathcal{G}_n(\mathbf{x}, t)$  given by (1.12) and the above random variables given by:

$$q_{uv} := \mathbb{P}(\eta_{uv} < x_u). \quad (5.4)$$

At each stage  $i \geq 1$ , we have a collection of active vertices  $\mathcal{A}(i)$ , a collection of explored vertices  $\mathcal{O}(i)$  and a collection of unexplored vertices  $\mathcal{U}(i) = [n] \setminus \mathcal{A}(i) \cup \mathcal{O}(i)$ .

Initialize with  $\mathcal{O}(1) = \emptyset$  and  $\mathcal{A}(1) = \{v(1)\}$ , where the first vertex  $v(1)$  is chosen by size-biased sampling, namely with probability proportional to vertex weights  $\mathbf{x}$ . When possible we will suppress dependence on  $n$  to ease notation. Now let  $\mathcal{D}(v(1)) := \{v : \eta_{v(1),v} \leq x_{v(1)}\}$  denote the collection of “children” of  $v(1)$  and note that by (5.4) this generates the right connection probabilities in  $\mathcal{G}_n(\mathbf{x}, t)$ . Think of the associated  $\eta_{v(1),v}$  values (for vertices connected to  $v(1)$ ) as “birth-times” of these connections in the interval  $[0, x_{v(1)}]$  and label the corresponding vertices as  $v(2), v(3), \dots, v(|\mathcal{D}(v(1))| + 1)$ . Update the process as  $\mathcal{O}(2) := \{v(1)\}$ ,  $\mathcal{A}(2) := \mathcal{D}(v(1))$  and  $\mathcal{U}(2) = \mathcal{U}(1) \setminus \mathcal{D}(v(1))$ .

Associate with this construction a breadth-first walk as follows:

$$Z_n(0) := 0, \quad Z_n(u) := -u + \sum_v x_v \mathbb{1}\{\eta_{v(1),v} \leq u\}, \quad 0 \leq u \leq x_{v(1)}. \quad (5.5)$$

Recursively for  $i \geq 2$  let  $T_{i-1} := \sum_{j=1}^{i-1} x_{v(j)}$ . At this “time” we will explore the unexplored neighbors of  $v(i)$ . By this time, there are  $|\mathcal{U}(i)| := i - 1 + |\mathcal{A}(i)|$  vertices that have either been explored or are active. Let  $\mathcal{D}(v(i)) := \{v \in \mathcal{U}(i) : \eta_{v(i),v} \leq x_{v(i)}\}$  and again label these as  $v(i + |\mathcal{A}(i)|), v(i + |\mathcal{A}(i)| + 1), \dots, v(i + |\mathcal{A}(i)| + |\mathcal{D}(v(i))| - 1)$  in increasing order of their  $\eta_{v(i),v}$  values. Update  $\mathcal{O}(i+1) = \mathcal{O}(i) \cup \{v(i)\}$ ,  $\mathcal{A}(i+1) = \mathcal{A}(i) \cup \mathcal{D}(v(i)) \setminus \{v(i)\}$  and  $\mathcal{U}(i+1) = \mathcal{U}(i) \setminus \mathcal{D}(v(i))$ . Again update the walk as

$$Z(T_{i-1} + u) = Z(T_{i-1}) - u + \sum_{v \in \mathcal{D}(v(i))} x_v \mathbb{1}\{\eta_{v(i),v} \leq u\}, \quad 0 \leq u \leq x_{v(i)}. \quad (5.6)$$

After finishing a component (which happens when  $\mathcal{A}(i) = \emptyset$  for some  $i \geq 2$ ), choose the next vertex to explore in a size-biased manner from the unexplored set  $\mathcal{U}(i)$ . If  $\mathcal{U}(i) = \emptyset$ , then we have finished constructing the partition of the graph into the connected components.

Now note the following important properties of this exploration:

- (a) The ordering  $(v(1), v(2), \dots, v(n))$  is a size-biased reordering of the vertex set  $[n]$ .
- (b) If we start a new component at some stage  $i$  with vertex  $v(i)$ , and finish exploring the component at stage  $j \geq i$ , then the walk satisfies

$$Z(T_j) = Z(T_{i-1}) - x_{v(i)}, \quad Z(u) \geq Z(T_j) \text{ on } T_{i-1} < u < T_j.$$

Thus, the size of the component of  $v(i)$ ,  $\sum_{l=i}^j x_{v(l)}$  is essentially the length of the excursion of the walk beyond past minima.

As a starting point in proving Theorem 1.6, Aldous and Limic [10] show the following result. Their result is more general (incorporating the presence of a “Brownian component”) but we state their result as applied to our setting.

**Proposition 5.2** ([10, Proposition 9]). *Consider the process  $\{\bar{Z}_n(s) : s \geq 0\}$  defined by setting  $\bar{Z}_n(s) := Z(s)/\sigma_2$ . Then under Assumption 1.5  $\bar{Z}_n \xrightarrow{d} V^c$  as  $n \rightarrow \infty$ .*

Using this result, Aldous and Limic [10] show that the corresponding maximal excursions beyond past minima of  $\bar{Z}_n$  also converge to the maximal excursions beyond past minima of  $V_\lambda^c$ , namely the excursion lengths of the reflected process  $\bar{V}_\lambda^c$  (see (1.19)) from zero. A consequence of the proof of Theorem 1.6 in [10] using Proposition 5.2 is the following result:

**Lemma 5.3.** *Fix  $K$  and let  $\mathcal{E}_n(K)$  be the time required for the above construction to explore the maximal  $K$  components  $\{\mathcal{C}_i : 1 \leq i \leq K\}$ . Then  $\{\mathcal{E}_n(K) : K \geq 1\}$  is tight.*

In other words, for any fixed  $K \geq 1$ , the maximal length excursions of  $\bar{V}^c$  are found in finite time. Thus, even though the total weight of vertices  $\sigma_1 \rightarrow \infty$ , when exploring the graph in size-biased fashion, under Assumption 1.5 one needs only a finite amount of “time” to find the maximal components. Here time is measured in terms of the weight of vertices already explored. Now define

$$S_{n,2}(u) = \sum_{i: T_i \leq u} \left( \frac{x_{v(i)}}{\sigma_2} \right)^2, \quad R_n^\varepsilon(u) := \sum_{i: T_i \leq u} \frac{x_{v(i)}^2}{\sigma_2^2} \mathbb{1}_{\{x_{v(i)} < \sigma_2 \varepsilon\}}. \quad (5.7)$$

Thus,  $S_{n,2}(t)$  is the normalized sum of *squares* of vertex weights of vertices explored by time  $t$  and  $R_n^\varepsilon$  is the normalized sum of these squares where we only retain explored vertices with weight at most  $\varepsilon \sigma_2$ . Using the same set of exponential random variables  $\{\xi_j : j \geq 1\}$  that arose in the definition of the process  $V^c$  in (1.18) define a new process

$$S_{\infty,2}(u) := \sum_{j=1}^{\infty} c_j^2 \mathbb{1}_{\{\xi_j \leq u\}}. \quad (5.8)$$

The same proof techniques as in [10] now implies the following. Since the ideas basically follow from [10] we only sketch the proof.

**Lemma 5.4.** *Assumption 1.5 implies the joint convergence of the processes  $(\bar{Z}_n(\cdot), S_{n,2}(\cdot)) \xrightarrow{d} (V^c(\cdot), S_{\infty,2}(\cdot))$  as  $n \rightarrow \infty$ .*

**Proof:** Fix  $K \geq 1$ , and for each  $i \geq 1$ , let  $\xi_i^{(n)}$  denote the time when vertex  $i$  is added to the collection of active vertices. Now consider the  $K + 1$  dimensional stochastic process

$$\mathbf{Y}_n^K(s) := \left( \bar{Z}_n(s), \frac{x_1}{\sigma_2} \mathbb{1}_{\{\xi_1^{(n)} \leq s\}}, \dots, \frac{x_K}{\sigma_2} \mathbb{1}_{\{\xi_K^{(n)} \leq s\}} \right), \quad s \geq 0.$$

Write

$$\mathbf{Y}_\infty^K(s) := (V^c(s), c_1 \mathbb{1}_{\{\xi_1 \leq s\}}, \dots, c_K \mathbb{1}_{\{\xi_K \leq s\}}).$$

In the proof of Proposition 5.2, Aldous and Limic showed that  $\mathbf{Y}_n^K \xrightarrow{d} \mathbf{Y}_\infty^K$  for every fixed  $K \geq 1$ . Thus to complete the proof, it is enough to show, for every fixed  $A > 0$  and  $\eta > 0$ ,  $\limsup_{\varepsilon \rightarrow 0} \limsup_{n \rightarrow \infty} \mathbb{P}(R_n^\varepsilon(A) > \eta) = 0$ . Now as described on [10, Page 17], we can couple  $(\xi_1^{(n)}, \xi_2^{(n)}, \dots, \xi_n^{(n)})$  with a sequence of independent exponential random variables

$(\tilde{\xi}_1^{(n)}, \tilde{\xi}_2^{(n)}, \dots, \tilde{\xi}_n^{(n)})$  with  $\tilde{\xi}_j^{(n)}$  having rate  $x_j/\sigma_2$  such that  $\tilde{\xi}_j^{(n)} \leq \xi_j^{(n)}$ . Now write

$$\tilde{R}_n^\varepsilon(t) := \sum_{j: x_j < \varepsilon \sigma_2} \frac{x_j^2}{\sigma_2^2} \mathbb{1}_{\{\tilde{\xi}_j^{(n)} \leq t\}}. \quad (5.9)$$

Then it is enough to show

$$\limsup_{\varepsilon \rightarrow 0} \limsup_{n \rightarrow \infty} \mathbb{E}(\tilde{R}_n^\varepsilon(A)) = 0, \quad (5.10)$$

which is trivial since

$$\mathbb{E}(\tilde{R}_n^\varepsilon(A)) \leq A \sum_{j: x_j \leq \varepsilon \sigma_2} \left( \frac{x_j}{\sigma_2} \right)^3 \rightarrow A \sum_{j: c_j < \varepsilon} c_j^3. \quad (5.11)$$

We have used both (1.15) and (1.16) in the last convergence assertion. Thus, first letting  $n \rightarrow \infty$  and then  $\varepsilon \rightarrow 0$  completes the proof.  $\blacksquare$

We can now complete the proof of Proposition 5.1. First, note that to prove (5.3), it is enough to show that for any two rationals  $r < s$ ,  $\sum_j c_j \mathbb{1}_{\{r \leq \xi_j \leq s\}} = \infty$  almost surely where  $\xi_j$  are the associated exponential rate  $c_j$  random variables. This, however, is trivially true as  $\sum_j c_j^2 = \infty$ .

To prove the other assertions, define, for  $i \geq 1$ , the point processes  $\Xi_n^{(i)} := \{x_u/\sigma_2 : u \in \mathcal{C}_i\}$ , namely the rescaled vertex weights in the  $i$ th maximal component. Analogously define  $\Xi_\infty^{(i)} = \{c_v : v \in \mathcal{Z}_i\}$ , namely the collection of jumps in the  $i$ th largest excursion of  $\tilde{V}^c$ . Let

$$s_n^{(i)} = \sum_{v \in \mathcal{C}_i} \frac{x_v^2}{\sigma_2^2}, \quad \text{and} \quad s_\infty^{(i)} := \sum_{v \in \mathcal{Z}_i} c_v^2, \quad (5.12)$$

for the normalized sum of squares of vertex weights in a component. Define

$$\tilde{\mathcal{P}}_n := ((\text{mass}(\mathcal{C}_i), s_n^{(i)}, \Xi_n^{(i)}), i \geq 1), \quad \tilde{\mathcal{P}}_\infty := ((Z_i, s_\infty^{(i)}, \Xi_\infty^{(i)}), i \geq 1).$$

We will view these as random elements of  $\tilde{\mathbb{S}}^\infty$  where  $\tilde{\mathbb{S}} := \mathbb{R}^2 \times \mathcal{N}(\mathbb{R})$ . Lemma 5.3 and Lemma 5.4 now imply the following:

**Lemma 5.5.** *As  $n \rightarrow \infty$   $\tilde{\mathcal{P}}_n \xrightarrow{d} \tilde{\mathcal{P}}_\infty$  on  $\tilde{\mathbb{S}}^\infty$ .*

Expressing the functionals that arise in Proposition 5.1 in terms of vertex weights in maximal components completes the proof. Indeed,

$$\sigma(\mathbf{p}_n^{(i)}) = \frac{\sqrt{\sum_{v \in \mathcal{C}_i} x_v^2}}{\sum_{v \in \mathcal{C}_i} x_v} = \frac{\sigma_2 \sqrt{s_n^{(i)}}}{\text{mass}(\mathcal{C}_i)} \rightarrow 0,$$

as  $n \rightarrow \infty$ . The proof of  $\mathcal{P}_n \xrightarrow{d} \mathcal{P}_\infty$  is similar.  $\blacksquare$

**5.2. Gromov-weak convergence in Theorem 1.1.** That convergence in (1.9) holds with respect to Gromov-weak topology is an easy consequence of Theorem 1.7. Indeed, setting

$$x_i = n^{-\frac{\tau-2}{\tau-1}} w_i \quad \text{and} \quad t_n = \frac{1}{\ell_n} \left( 1 + \lambda n^{-\frac{\tau-3}{\tau-1}} \right) n^{\frac{2(\tau-2)}{\tau-1}}, \quad (5.13)$$

we can write  $\text{NR}_n(\mathbf{w}(\lambda))$  as the model  $\mathcal{G}(\mathbf{x}, t_n)$  where  $\mathbf{x} = \mathbf{x}^{(n)} := (x_i : i \in [n])$ . A direct computation will show that  $\mathbf{x}^{(n)}$  satisfies Assumption 1.5 with the entrance boundary  $\mathbf{c}^{\text{nr}}$  defined in (2.16). Note also that

$$t_n - \frac{1}{\sigma_2(\mathbf{x}^{(n)})} = \frac{n^{\frac{2(\tau-2)}{\tau-1}}}{\sum_{i \in [n]} w_i^2} \left( \frac{1}{\ell_n} \sum_{i \in [n]} w_i^2 - 1 \right) + \frac{\lambda}{\ell_n / n}.$$

Under the assumptions of Theorem 1.1,  $\ell_n/n \rightarrow \mathbb{E} W$  and  $\sum_i w_i^2/n \rightarrow \mathbb{E} W^2 = \mathbb{E} W$ . Further, by [17, Lemma 2.2],

$$\frac{1}{\ell_n} \sum_{i \in [n]} w_i^2 = 1 + \zeta n^{-\frac{\tau-3}{\tau-1}} + o(n^{-\frac{\tau-3}{\tau-1}}),$$

where  $\zeta$  is as defined in (2.17). Combining these observations, we see that

$$t_n - (\sigma_2(\mathbf{x}^{(n)}))^{-1} \rightarrow t_\lambda^{\text{nr}} \text{ as } n \rightarrow \infty,$$

where  $t_\lambda^{\text{nr}}$  is as in (2.17). Since  $n^{\frac{\tau-3}{\tau-1}} \sigma_2(\mathbf{x}^{(n)}) \rightarrow \mathbb{E} W$ , we conclude that  $\mathbf{M}_\infty^{\text{nr}}(\lambda)$  defined in (2.18) is the Gromov-weak limit of  $n^{-\frac{\tau-3}{\tau-1}} \mathbf{M}_n^{\text{nr}}(\lambda)$ , where  $\mathbf{M}_n^{\text{nr}}(\lambda)$  is as in (1.8).

**Remark 7.** Theorem 1.7 is stated for a fixed  $\lambda \in \mathbb{R}$ , but in the argument just given, we have to work with a sequence, namely  $t_n - (\sigma_2(\mathbf{x}^{(n)}))^{-1}$  converging to  $t_\lambda^{\text{nr}}$ . This, however, does not make any difference. Indeed, the proof of [10, Proposition 9] can be imitated to prove the same result in the setup where we have a sequence converging to  $t$  instead of a fixed  $t$ , and no new idea is involved here. (In [10, Lemma 27], Aldous and Limic prove a similar result for the multiplicative coalescent. They do not, however, explicitly state the convergence of the associated process under the same assumption.)

## 6. PROOFS: CONVERGENCE IN GROMOV-HAUSDORFF-PROKHOROV TOPOLOGY

In this section, we improve Gromov-weak convergence in Theorem 1.1 to Gromov-Hausdorff-Prokhorov convergence. To do so, we will rely on [14, Theorem 6.1] which gives a criterion for convergence in Gromov-Hausdorff-weak topology. We do not give the definition of Gromov-Hausdorff-weak topology and instead refer the reader to [14, Definition 5.1]. Convergence in Gromov-Hausdorff-weak topology implies convergence in Gromov-Hausdorff-Prokhorov topology when we are working with metric measure spaces having full support (i.e. support of the measure is the entire metric space). This is true in our situation. Indeed, it is a trivial fact that  $\mathcal{C}_i(\lambda)$  has full support. Further, the mass measure on an inhomogeneous continuum random tree has full support which implies that the same is true for  $M_i^{\text{nr}}(\lambda)$ .

Applying [14, Theorem 6.1] to our situation, we see that it is enough to prove the following lemma.

**Lemma 6.1** (Global lower mass-bound). *Let  $\mathcal{C}_i(\lambda)$  be the  $i$ th largest component of  $\text{NR}_n(\mathbf{w}(\lambda))$ . Then the following assertion is true: For each  $i \geq 1$ ,  $v \in [n]$  and  $\delta > 0$ , let  $B(v, \delta)$  denote the intrinsic ball (in  $\text{NR}_n(\mathbf{w}(\lambda))$ ) of radius  $\delta n^{(\tau-3)/(\tau-1)}$  around  $v$  and set*

$$m_i^{(n)}(\delta) = \inf \left\{ n^{-\frac{\tau-2}{\tau-1}} \sum_{j \in B(v, \delta)} w_j \mid v \in \mathcal{C}_i(\lambda) \right\}.$$

*Then the sequence  $\left\{ \left( m_i^{(n)}(\delta) \right)^{-1} \right\}_{n \geq 1}$  is tight.*

This ensures compactness of the spaces  $M_i^{\text{nr}}(\lambda)$  which, in turn, implies compactness of the spaces  $M_i^{\mathbf{c}}(\lambda)$  when  $\mathbf{c} = (c_1, c_2, \dots)$  is of the form (1.22), thus proving the first assertion in Theorem 1.8.

Before moving on to the proof of Lemma 6.1, we state a result which essentially says that instead of looking at the largest components, we can work with the components of high-weight vertices. This observation will be used to prove the global lower-mass bound.



**Proposition 6.2.** *For every  $\varepsilon > 0$  and  $k \geq 1$ , there exists  $K = K(\varepsilon, k, \lambda) > 0$  such that*

$$\mathbb{P}([K] \cap \mathcal{C}_i(\lambda) = \emptyset \text{ for some } 1 \leq i \leq k) \leq \varepsilon.$$

The above proposition follows trivially from [17, Theorem 1.6 (a)] and [17, Theorem 1.1].

**6.1. Bound on size of  $\varepsilon n^{\frac{\tau-3}{\tau-1}}$ -nets for the largest components.** For convenience, we set

$$\eta = (\tau - 3)/(\tau - 1) \text{ and } \rho = (\tau - 2)/(\tau - 1). \quad (6.1)$$

The purpose of this section is to prove a strong result (Proposition 6.3 stated below) that gives control over the number of intrinsic balls of radius  $\varepsilon n^\eta$  needed to cover the largest components. This acts as a crucial ingredient in the proof of Lemma 6.1 as well as the proof of the bound on the upper box-counting dimension.

**Proposition 6.3** (Small diameter after removing high-weight vertices). *For every  $\varepsilon, \delta > 0$ , and  $N = N(\varepsilon) := \varepsilon^{-\delta-1/\eta}$ ,*

$$\mathbb{P}(\text{diam}(\text{NR}_n(\mathbf{w}(\lambda)) \setminus [N]) > \varepsilon n^\eta) \leq c_\delta \exp(-C/\varepsilon^{\delta\eta}), \quad (6.2)$$

*for all  $n$  sufficiently large, a positive constant  $c_\delta$  depending on  $\delta$  and a universal constant  $C > 0$ . Here  $\text{NR}_n(\mathbf{w}(\lambda)) \setminus [N]$  denotes the graph obtained by removing all vertices with labels in  $[N]$  and the edges incident to them from the graph  $\text{NR}_n(\mathbf{w}(\lambda))$*

We continue to prove Proposition 6.3. Write

$$E_n = \{\text{diam}(\text{NR}_n(\mathbf{w}(\lambda)) \setminus [N]) \leq \varepsilon n^\eta\}. \quad (6.3)$$

The proof consists of four steps. In the first step, we reduce the proof to the study of the height of mixed-Poisson branching processes. In the second step, we ensure that we can take  $\lambda = 0$ , while in the third step, we study the survival probability of such critical infinite-variance branching processes. In the fourth and final step, we prove the claim.

**Comparison to mixed Poisson branching processes.** Let  $\mathcal{C}_{\text{res}}(i)$  be the cluster of  $i$  in the (restricted) random graph on vertex set  $[n] \setminus [i-1]$  with edge probabilities  $q_{k\ell}(\mathbf{w}(\lambda))$  for  $k, \ell \in [n] \setminus [i-1]$ , where  $q_{k\ell}(\mathbf{w}(\lambda))$  is as in (1.1).

Note that the event  $E_n^c$  implies the existence of  $i > N$  such that the following happens: **(a)** The diameter of the component of  $i$  in  $\text{NR}_n(\mathbf{w}(\lambda)) \setminus [N]$  is bigger than  $\varepsilon n^\eta$ . **(b)** No  $j \in \{N+1, \dots, i-1\}$  belongs to the component of  $i$  in  $\text{NR}_n(\mathbf{w}(\lambda)) \setminus [N]$ . In particular,  $\text{diam}(\mathcal{C}_{\text{res}}(i)) \geq \varepsilon n^\eta$  for this  $i$ . Thus,

$$\mathbb{P}(E_n^c) \leq \sum_{i>N} \mathbb{P}(\text{diam}(\mathcal{C}_{\text{res}}(i)) > \varepsilon n^\eta). \quad (6.4)$$

Now the random graph  $\text{NR}_n(\mathbf{w}(\lambda))$  restricted to  $[n] \setminus [i-1]$  is the Norros-Reittu random graph  $\text{NR}_n(\mathbf{w}^{(i)}(\lambda))$ , where  $\mathbf{w}^{(i)}(\lambda) = (w_j^{(i)}(\lambda) : j \in [n] \setminus [i-1])$ ,  $w_j^{(i)}(\lambda) = w_j(\lambda)\ell_n^{(i)}/\ell_n$ , and  $\ell_n^{(i)} = \sum_{k=i}^n w_k$ . Indeed, this follows from the simple observation:

$$\frac{w_k^{(i)}(\lambda)w_\ell^{(i)}(\lambda)}{\sum_{r=i}^n w_r^{(i)}(\lambda)} = \left(1 + \frac{\lambda}{n^\eta}\right) \frac{w_k w_\ell}{\ell_n}.$$

Write  $W_n^{(i)}(\lambda)$  for a random variable whose distribution is given by  $(n-i+1)^{-1} \sum_{j=i}^n \delta_{w_j^{(i)}(\lambda)}$ , and for any non-negative random variable  $X$  with  $\mathbb{E} X > 0$ , let  $X^\circ$  be the random variable having the size-biased distribution given by

$$\mathbb{P}(X^\circ \leq x) = \mathbb{E}(X \mathbb{1}_{\{X \leq x\}}) / \mathbb{E} X.$$

We will use the following comparison to a mixed Poisson branching process:

**Lemma 6.4** (Domination by a mixed-Poisson branching process). *Fix  $i \in [n]$  and consider  $\text{NR}_n(\mathbf{w}^{(i)}(\lambda))$ . Then, there exists a coupling of  $\mathcal{C}_{\text{res}}(i)$  and a branching process where the root has a  $\text{Poi}(w_i^{(i)}(\lambda))$  offspring distribution while every other vertex has a  $\text{Poi}((W_n^{(i)}(\lambda))^\circ)$  offspring distribution such that in the breadth-first exploration of  $\mathcal{C}_{\text{res}}(i)$  starting from  $i$ , each vertex  $v \in \mathcal{C}_{\text{res}}(i)$  has at most the number of children as in the branching process.*

**Proof:** See [57, Proposition 3.1]. ■

It immediately follows from Lemma 6.4 that

$$\mathbb{P}(\text{diam}(\mathcal{C}_{\text{res}}(i)) > \varepsilon n^\eta) \leq \mathbb{P}(\text{ht}(\bar{T}_n^{(i)}(\lambda)) > \varepsilon n^\eta/2), \quad (6.5)$$

where  $\bar{T}_n^{(i)}(\lambda)$  is a mixed Poisson branching process tree whose root has a  $\text{Poi}(w_i^{(i)}(\lambda))$  offspring distribution and every other vertex has a  $\text{Poi}((W_n^{(i)}(\lambda))^\circ)$  offspring distribution. As before,  $\text{ht}(\mathbf{t})$  denotes the height of the tree  $\mathbf{t}$ .

When  $\text{ht}(\bar{T}_n^{(i)}(\lambda)) > \varepsilon n^\eta/2$ , at least one of the subtrees of the root needs to have height at least  $\varepsilon n^\eta/2$ . Combining this observation with (6.4) and (6.5), we get

$$\mathbb{P}(E_n^c) \leq \sum_{i>N} \mathbb{E}[\text{Poi}(w_i^{(i)}(\lambda))] \mathbb{P}(\text{ht}(T_n^{(i)}(\lambda)) \geq \varepsilon n^\eta/2) \leq \sum_{i>N} w_i^{(i)}(\lambda) \mathbb{P}(\text{ht}(T_n^{(i)}(\lambda)) \geq \varepsilon n^\eta/2), \quad (6.6)$$

where  $T_n^{(i)}(\lambda)$  is a branching process tree where every vertex has a  $\text{Poi}((W_n^{(i)}(\lambda))^\circ)$  offspring distribution.

We make the convention of writing  $T_n^{(i)}$ ,  $W_n^{(i)}$  etc. instead of  $T_n^{(i)}(0)$ ,  $W_n^{(i)}(0)$  etc. With this notation, it is easy to see that  $W_n^{(i)}(\lambda) \stackrel{d}{=} (1 + \lambda n^{-\eta}) W_n^{(i)}$  and hence  $(W_n^{(i)}(\lambda))^\circ \stackrel{d}{=} (1 + \lambda n^{-\eta})(W_n^{(i)})^\circ$ .

**The survival probability of mixed Poisson branching processes.** We would like to compare our mixed-Poisson branching process with an offspring distribution that is independent of  $n$ . For this, we rely on the following two lemmas:

**Lemma 6.5** (Mixed-Poisson branching processes of different parameters). *Let  $T_n^{(i)}$  and  $T_n^{(i)}(\lambda)$  be as above. Assume further that  $\lambda \geq 0$ . Then, for each  $k \geq 1$ ,*

$$\mathbb{P}(\text{ht}(T_n^{(i)}(\lambda)) \geq k) \leq (1 + \lambda n^{-\eta})^k \cdot \mathbb{P}(\text{ht}(T_n^{(i)}) \geq k).$$

**Proof:** We follow [44, Proof of Lemma 3.4(1)]. Writing  $\delta = 1 + \lambda n^{-\eta}$ , we note that we can obtain  $T_n^{(i)}$  as a subtree of  $T_n^{(i)}(\lambda)$  by killing every child independently with probability  $1 - \delta^{-1}$ . Write  $\mathcal{A}$  for the event in which  $\text{ht}(T_n^{(i)}(\lambda)) \geq k$  and no vertex in the leftmost path of length  $k$  starting from the root in  $T_n^{(i)}(\lambda)$  is killed. Then

$$\mathbb{P}(\mathcal{A}) = \delta^{-k} \mathbb{P}(\text{ht}(T_n^{(i)}(\lambda)) \geq k).$$

Indeed, the probability of the leftmost path surviving is precisely  $1/\delta^k$ . To finish the proof, note that  $\mathcal{A}$  implies  $\text{ht}(T_n^{(i)}) \geq k$ , so that

$$\mathbb{P}(\text{ht}(T_n^{(i)}) \geq k) \geq \mathbb{P}(\mathcal{A}) = \delta^{-k} \mathbb{P}(\text{ht}(T_n^{(i)}(\lambda)) \geq k),$$

which is the desired inequality. ■

**Lemma 6.6** (Stochastic bound by  $n$ -independent variable). *The random variable  $(W_n^{(i)})^\circ$  is stochastically upper bounded by  $W^\circ$  where  $W \sim F$ , i.e.,  $(W_n^{(i)})^\circ \stackrel{\text{st}}{\leq} W^\circ$ .*

**Proof:** We note that

$$\begin{aligned}\mathbb{P}(W_n^{(i)} \leq x) &= \frac{1}{n-i+1} \sum_{j \in [n] \setminus [i-1]} \mathbb{1}\{w_j^{(i)} \leq x\} = \frac{1}{n-i+1} \sum_{j \in [n] \setminus [i-1]} \mathbb{1}\{w_j \leq x \ell_n / \ell_n^{(i)}\} \\ &\geq \frac{1}{n-i+1} \sum_{j \in [n] \setminus [i-1]} \mathbb{1}\{w_j \leq x\} \geq \frac{1}{n} \sum_{j \in [n]} \mathbb{1}\{w_j \leq x\}.\end{aligned}$$

The last inequality holds because we are removing the largest weight vertices, and can be checked easily by induction on  $i$ . We conclude that  $W_n^{(i)} \stackrel{\text{st}}{\leq} W_n$  which is distributed as the empirical weight distribution of  $(w_i)_{i \in [n]}$ . Further, since  $w_i = [1 - F]^{-1}(i/n)$ ,  $W_n \stackrel{\text{st}}{\leq} W$  (see, e.g., [42, (6.1.17)]). Finally, size-biasing preserves stochastic domination (see, e.g., [42, Proposition 2.13]), which implies that  $(W_n^{(i)})^\circ \stackrel{\text{st}}{\leq} W_n^\circ \stackrel{\text{st}}{\leq} W^\circ$ . This yields the desired result.  $\blacksquare$

We continue to study the survival probability of mixed-Poisson branching processes with infinite variance offspring distribution.

**Lemma 6.7** (Survival probability of infinite-variance MPBP). *Let  $T$  denote a mixed-Poisson branching process tree with offspring distribution  $\text{Poi}(W^\circ)$ . Then, there exists a constant  $c_{6.7}$  such that for all  $m \geq 1$ ,*

$$\mathbb{P}(\text{ht}(T) \geq m) \leq c_{6.7} m^{-1/(\tau-3)}.$$

**Proof:** This is a well-known result. We sketch the proof briefly for completeness. Recall the following facts about  $W^\circ$ : **(a)**  $\mathbb{E}[W^\circ] = v = 1$  and **(b)** for  $x \rightarrow \infty$ ,  $\mathbb{P}(W^\circ > x) = cx^{-(\tau-2)}(1 + o(1))$ . By the *Otter-Dwass formula*, which describes the distribution of the total progeny of a branching process (see [36] for the special case when the branching process starts with a single individual, [58] for the more general case, and [43] for a simple proof based on induction), we have

$$\mathbb{P}(|T| = k) = \frac{1}{k} \mathbb{P}\left(\sum_{i=1}^k X_i = k-1\right),$$

where  $X_i$  are i.i.d. random variables distributed as  $W^\circ$ . By [41, Proposition 2.7], in our situation,  $\mathbb{P}(\sum_{i=1}^k X_i = k-1) \leq ck^{-1/(\tau-2)}$ , so that

$$\mathbb{P}(|T| = k) \leq ck^{-(\tau-1)/(\tau-2)} \text{ and } \mathbb{P}(|T| \geq k) \leq ck^{-1/(\tau-2)}. \quad (6.7)$$

Take  $k = m^{(\tau-2)/(\tau-3)}$  in the second inequality in (6.7) to get

$$\mathbb{P}(\text{ht}(T) \geq m) \leq cm^{-1/(\tau-3)} + \mathbb{P}(\text{ht}(T) \geq m, |T| \leq m^{(\tau-2)/(\tau-3)}),$$

where  $|T|$  denotes the total number of vertices in  $T$ . We condition on the size  $|T|$  and write

$$\begin{aligned}\mathbb{P}(\text{ht}(T) \geq m, |T| \leq m^{(\tau-2)/(\tau-3)}) &= \sum_{k=1}^{m^{(\tau-2)/(\tau-3)}} \mathbb{P}(\text{ht}(T) \geq m \mid |T| = k) \mathbb{P}(|T| = k) \\ &\leq c \sum_{k=1}^{m^{(\tau-2)/(\tau-3)}} \mathbb{P}(\text{ht}(T) \geq m \mid |T| = k) k^{-\frac{\tau-1}{\tau-2}}.\end{aligned} \quad (6.8)$$

By [50, Theorem 4], there exists a  $\kappa > 1$  such that, uniformly for  $u \geq 1$ ,

$$\mathbb{P}(\text{ht}(T) \geq uk^{(\tau-3)/(\tau-2)} \mid |T| = k) \leq e^{-au^\kappa}.$$

Combining this with (6.8), we get

$$\mathbb{P}(\text{ht}(T) \geq m, |T| \leq m^{(\tau-2)/(\tau-3)}) \leq \sum_{k=1}^{m^{\frac{\tau-2}{\tau-3}}} \exp\left(-a\left(mk^{-\frac{\tau-3}{\tau-2}}\right)^K\right) k^{-\frac{\tau-1}{\tau-2}} = \Theta\left(m^{-\frac{1}{\tau-3}}\right).$$

■

**Proof of Proposition 6.3:** Clearly

$$\mathbb{P}(\text{ht}(T_n^{(i)}) \geq m) \leq \mathbb{E}\left[\left(W_n^{(i)}\right)^\circ\right] \mathbb{P}(\text{ht}(T_n^{(i)}) \geq m-1) =: v_n^{(i)} \mathbb{P}(\text{ht}(T_n^{(i)}) \geq m-1),$$

where

$$v_n^{(i)} = \mathbb{E}\left[\left(W_n^{(i)}\right)^\circ\right] = \frac{\sum_{j \geq i} (w_j^{(i)})^2}{\sum_{j \geq i} w_j^{(i)}} = \left(\frac{\ell_n^{(i)}}{\ell_n}\right) \frac{\sum_{j \geq i} w_j^2}{\sum_{j \geq i} w_j} = \frac{\sum_{j \geq i} w_j^2}{\ell_n}.$$

Iterating this  $\varepsilon n^\eta/4$  times, we get

$$\begin{aligned} \mathbb{P}(\text{ht}(T_n^{(i)}) \geq \varepsilon n^\eta/2) &\leq (v_n^{(i)})^{\varepsilon n^\eta/4} \mathbb{P}(\text{ht}(T_n^{(i)}) \geq \varepsilon n^\eta/4) \\ &\leq (v_n^{(i)})^{\varepsilon n^\eta/4} \mathbb{P}(\text{ht}(T) \geq \varepsilon n^\eta/4) \leq (v_n^{(i)})^{\varepsilon n^\eta/4} \times c_{6.7} \left(\frac{4}{\varepsilon}\right)^{\frac{1}{\tau-3}} \frac{1}{n^{1/(\tau-1)}}, \end{aligned} \quad (6.9)$$

where the second inequality is a consequence of Lemma 6.6 and the last step follows from Lemma 6.7.

Substituting the estimate (6.9) into (6.6) leads to

$$\mathbb{P}(E_n^c) \leq c \varepsilon^{-\frac{1}{\tau-3}} n^{-\frac{1}{\tau-1}} \left(1 + \frac{\max\{\lambda, 0\}}{n^\eta}\right)^{1+\varepsilon n^\eta/2} \sum_{i > N} w_i (v_n^{(i)})^{\varepsilon n^\eta/4} \quad (6.10)$$

for some constant  $c$ . Here we have used Lemma 6.5 and the simple fact:  $w_i^{(i)} \leq w_i$ .

Next, note that it is an easy consequence of (1.3) that there exist constants  $c', c'' > 0$  such that for all  $i \in [n]$ ,

$$w_i \leq c' \left(\frac{n}{i}\right)^{1/(\tau-1)} \text{ and } \sum_{j=1}^i w_j^2 \geq c'' \sum_{j=1}^i \left(\frac{n}{j}\right)^{2/(\tau-1)}. \quad (6.11)$$

Further, [17, Lemma 2.2] implies that  $v_n^{(1)} < 1$  for large  $n$ . Hence, for every  $i \geq 2$ ,

$$v_n^{(i)} = v_n^{(1)} - \frac{1}{\ell_n} \sum_{j=1}^{i-1} w_j^2 \leq 1 - C n^{-\eta} i^\eta \leq \exp(-C n^{-\eta} i^\eta)$$

for some  $C > 0$ . Here, we have used the second inequality in (6.11). Combining this estimate with (6.10) and the first inequality in (6.11), we end up with

$$\mathbb{P}(E_n^c) \leq C' \varepsilon^{-1/(\tau-3)} \sum_{i > N} i^{-1/(\tau-1)} \exp(-C \varepsilon i^\eta/4)$$

for some  $C' > 0$ . Taking  $N = \varepsilon^{-\delta-1/\eta}$ , we get

$$\begin{aligned} \mathbb{P}(E_n^c) &\leq C' \varepsilon^{-1/(\tau-3)} N^{-1/(\tau-1)} \sum_{i>N} \exp(-C\varepsilon i^\eta/4) \\ &\leq C' \varepsilon^{\delta/(\tau-1)} \sum_{k=0}^{\infty} \sum_{i=N2^k}^{N2^{k+1}-1} \exp(-C\varepsilon i^\eta/4) \\ &\leq C' \varepsilon^{\delta/(\tau-1)} N \sum_{k=0}^{\infty} 2^k \exp(-C\varepsilon (N2^k)^\eta/4). \end{aligned} \quad (6.12)$$

Note that  $\varepsilon N^\eta = \varepsilon^{-\delta\eta}$ . A little more work after plugging this into (6.12) will lead to (6.2).  $\blacksquare$

**6.2. Proof of global lower-mass bound.** In this section, we complete the proof of Lemma 6.1. We start with some preliminaries:

**Lemma 6.8** (Weight of size-biased reordering). *Let  $\pi_v(1) = v$  and  $(\pi_v(i) : i \in [n] \setminus \{1\})$  be a size-biased reordering on  $[n] \setminus \{v\}$  where size of vertex  $v'$  is proportional to  $w_{v'}$  for  $v' \in [n] \setminus \{v\}$ . Then, for every  $k = o(1)$ , there exists a  $J > 0$  such that*

$$\mathbb{P}\left(\exists v : \sum_{j=1}^k w_{\pi_v(j)} \leq k/2\right) \leq ne^{-Jk}.$$

**Proof:** See [17, Proof of Lemma 5.1].  $\blacksquare$

Recall the definitions of  $\eta$  and  $\rho$  from (6.1). Recall that for  $v \in [n]$ ,  $B(v, \delta)$  denotes the intrinsic ball (in  $\text{NR}_n(\mathbf{w}(\lambda))$ ) around  $v$  or radius  $\delta n^\eta$ . We will use the following bound on the weight of balls:

**Lemma 6.9** (Weights of balls around high-weight vertices cannot be too small). *For every  $\varepsilon > 0$  and  $i \geq 1$ , there exist  $n_{i,\varepsilon}$  large and  $\delta_{i,\varepsilon} > 0$  such that for all  $n \geq n_{i,\varepsilon}$  and  $\delta \in (0, \delta_{i,\varepsilon}]$ , we have*

$$\mathbb{P}\left(\sum_{j \in B(i, \delta)} w_j \leq \left(\frac{c_F}{2i}\right)^{1/(\tau-1)} \frac{\delta n^\rho}{2}\right) \leq n \exp\left(-\frac{J\delta n^\rho}{i^{1/(\tau-1)}}\right) + \frac{\varepsilon}{2i}. \quad (6.13)$$

**Proof:** We will rely on a cluster exploration used in [17] which we describe next. We denote by  $(Z_l(i))_{l \geq 0}$  the exploration process of the cluster  $\mathcal{C}(i)$ , starting from  $i$ , in the breadth-first search, where  $Z_0(i) = 1$  and where  $Z_1(i)$  denotes the number of potential neighbors of the initial vertex  $i$ . The variable  $Z_l(i)$  has the interpretation of the number of potential neighbors of the first  $l$  explored potential vertices in the cluster whose neighbors have not yet been explored. As a result, we explore by taking one vertex of the ‘stack’ of size  $Z_l(i)$ , drawing its mark and checking whether it is a real vertex, followed by drawing its number of potential neighbors. Thus, we set  $Z_0(i) = 1$ ,  $Z_1(i) = \text{Poi}(w_i)$ , and note that, for  $l \geq 2$ ,  $Z_l(i)$  satisfies the recursion relation

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

where  $X_l$  denotes the number of potential neighbors of the  $l$ th potential vertex that is explored, where  $X_1 = X_1(i) = \text{Poi}(w_i)$ . More precisely, when we explore the  $l$ th potential vertex, we start by drawing its mark  $M_l$  in an i.i.d. way with distribution

$$\mathbb{P}(M = m) = w_m / \ell_n, \quad 1 \leq m \leq n.$$

When we have already explored a vertex with the same mark as the one drawn, we turn the status of the vertex to be explored to inactive, the potential vertex does not become a real vertex, and we proceed with the next potential vertex. When, instead, it receives a mark that we have not

yet seen, then the potential vertex becomes a real vertex, its mark  $M_l \in [n]$  indicating to which vertex in  $[n]$  the  $l$ th explored vertex corresponds, so that  $M_l \in \mathcal{C}(i)$ . We then draw  $X_l = \text{Poi}(w_{M_l})$ , and  $X_l$  denotes the number of potential vertices incident to the real vertex  $M_l$ . Again, upon exploration, these potential vertices might become real vertices, and this occurs precisely when their mark corresponds to a vertex in  $[n]$  that has not appeared in the cluster exploration so far. We call the above procedure of drawing a mark for a potential vertex to investigate whether it corresponds to a real vertex a *vertex check*. Let

$$\mathcal{Z}_t^{(n)}(i) = n^{-1/(\tau-1)} Z_{\lfloor tn^\rho \rfloor}(i) \text{ for } t > 0.$$

Then, by imitating the techniques used in the proof of Theorem 2.4 of [17], we get

$$(\mathcal{Z}_t^{(n)}(i))_{t>0} \xrightarrow{d} (\mathcal{S}_t(i))_{t>0}.$$

([17, Theorem 2.4] states the result for  $i = 1$ . However the exact same proof goes through for any  $i \geq 2$ .) The limiting process  $(\mathcal{S}_t(i))_{t>0}$  is defined as follows: Let

$$a = c_F^{1/(\tau-1)} / \mathbb{E}[W] \text{ and } b = b(i) = (c_F/i)^{1/(\tau-1)}. \quad (6.14)$$

We let  $(\mathcal{J}_i(t))_{i \geq 1}$  denote independent increasing indicator processes defined by

$$\mathcal{J}_i(s) = \mathbb{1} \{ \text{Exp}(ai^{-1/(\tau-1)}) \in [0, s] \}, \quad s \geq 0,$$

so that

$$\mathbb{P}(\mathcal{J}_i(s) = 0 \forall s \in [0, t]) = \exp(-at/i^{1/(\tau-1)}).$$

Then we define

$$\mathcal{S}_t(i) = b - abt + ct + \sum_{j \neq i}^{\infty} \frac{b}{j^{1/(\tau-1)}} \left[ \mathcal{J}_j(t) - \frac{at}{j^{1/(\tau-1)}} \right] \quad (6.15)$$

for all  $t \geq 0$ , where  $c = \zeta - ab$  and  $\zeta$  is as in (2.17). We call  $(\mathcal{S}_t)_{t \geq 0}$  a *thinned Lévy process*.

Let  $\mathcal{H}_n^{(i)}(u)$  denote the hitting time of  $u$  of the process  $(\mathcal{Z}_t^{(n)}(i))_{t>0}$ . Then, by [17, Corollary 3.4],  $\mathcal{H}_n^{(i)}(u) \xrightarrow{d} \mathcal{H}_{\mathcal{S}_t(i)}(u)$ , the hitting time of  $u$  of the process  $(\mathcal{S}_t(i))_{t>0}$ . This implies the existence of a  $B_{\varepsilon, i}$  (independent of  $n$ ) and an integer  $n_{i, \varepsilon}$  such that

$$\mathbb{P}(\mathcal{H}_n^{(i)}((c_F/2i)^{1/(\tau-1)}) \leq B_{\varepsilon, i}) \leq \varepsilon 2^{-i} \text{ for } n \geq n_{i, \varepsilon}, \quad (6.16)$$

since the limiting process  $(\mathcal{S}_t(i))_{t>0}$  starts from  $(c_F/i)^{1/(\tau-1)}$  and takes a positive amount of time to reach  $(c_F/2i)^{1/(\tau-1)}$ .

Let  $|B(i, r)|$  denote the number of vertices in  $B(i, r)$ . Let  $\delta_{\varepsilon, i}$  be so small that

$$(c_F/2i)^{1/(\tau-1)} \delta_{\varepsilon, i} < B_{\varepsilon, i}. \quad (6.17)$$

Then we claim that for all  $\delta \in (0, \delta_{\varepsilon, i}]$ ,

$$\mathbb{P}(|B(i, \delta)| \leq (c_F/2i)^{1/(\tau-1)} \delta n^\rho) \leq \mathbb{P}(\mathcal{H}_n^{(i)}((c_F/2i)^{1/(\tau-1)}) \leq B_{\varepsilon, i}). \quad (6.18)$$

That (6.18) holds can be seen as follows. For  $|B(i, \delta)| \leq (c_F/2i)^{1/(\tau-1)} \delta n^\rho$  to occur, there has to exist some  $j \in [1, \delta n^\eta]$  such that the number of vertices at distance  $j$  from  $i$  is smaller than  $(c_F/2i)^{1/(\tau-1)} \delta n^\rho / (\delta n^\eta)$ , i.e.,

$$\min_{1 \leq j \leq \delta n^\eta} |\partial B(i, j n^{-\eta})| \leq (c_F/2i)^{1/(\tau-1)} n^{1/(\tau-1)}. \quad (6.19)$$

Now the number of vertices at distance  $j$  from  $i$  is precisely the number of vertices in generation  $j$  of the breadth-first exploration process, and hence this number (scaled by  $n^\rho$ ) appears in the

function  $\mathcal{X}_t^{(n)}(i)$ . Thus, (6.19) implies that  $(\mathcal{X}_t^{(n)}(i))_{t>0}$  has to hit  $(c_F/2i)^{1/(\tau-1)}$  before we have finished exploring up to generation  $\delta n^\eta$ , i.e., we must have that

$$\mathcal{H}_n^{(i)}((c_F/2i)^{1/(\tau-1)}) \leq \frac{|B(i, \delta)|}{n^\rho} \leq \left(\frac{c_F}{2i}\right)^{1/(\tau-1)} \delta < B_{\varepsilon, i},$$

where the last inequality holds by (6.17) and because  $\delta \in (0, \delta_{\varepsilon, i}]$ .

Combining (6.16) and (6.18), we conclude that for all  $\delta \in (0, \delta_{\varepsilon, i}]$  and  $n \geq n_{i, \varepsilon}$ ,

$$\mathbb{P}\left(|B(i, \delta)| \leq (c_F/2i)^{1/(\tau-1)} \delta n^\rho\right) \leq \varepsilon 2^{-i}. \quad (6.20)$$

This explains the second term in (6.13).

To see what happens when  $|B(i, \delta)| \geq (c_F/2i)^{1/(\tau-1)} \delta n^\rho$ , recall that the vertices appear in a size-biased fashion in our exploration process. Hence

$$\begin{aligned} \mathbb{P}\left(\sum_{j \in B(i, \delta)} w_j \leq \left(\frac{c_F}{2i}\right)^{1/(\tau-1)} \frac{\delta n^\rho}{2}, |B(i, \delta)| \geq \left(\frac{c_F}{2i}\right)^{1/(\tau-1)} \delta n^\rho\right) \\ \leq \mathbb{P}\left(\sum_{j=1}^{\delta n^\rho (c_F/(2i))^{1/(\tau-1)}} w_{\pi_i(j)} \leq \left(\frac{c_F}{2i}\right)^{1/(\tau-1)} \frac{\delta n^\rho}{2}\right) \\ \leq n \exp\left(-\frac{J \delta n^\rho}{i^{1/(\tau-1)}}\right), \end{aligned} \quad (6.21)$$

by Lemma 6.8. Combining (6.20) and (6.21) proves the claim.  $\blacksquare$

**Lemma 6.10.** *For  $v \in [n]$ , let  $\mathcal{C}(v)$  denote the component of  $v$  in  $\text{NR}_n(\mathbf{w}(\lambda))$ . Then for every fixed  $i \geq 1$  and  $\varepsilon_1, \varepsilon_2 > 0$ , there exist  $\xi = \xi_{\varepsilon_1, \varepsilon_2}^{(i)} > 0$  and an integer  $\bar{n}_{\varepsilon_1, \varepsilon_2}^{(i)}$  such that*

$$\mathbb{P}\left(\min_{v \in \mathcal{C}(i)} \left(\sum_{j \in B(v, \varepsilon_1)} w_j\right) \leq \xi n^\rho\right) \leq \varepsilon_2 \text{ for } n \geq \bar{n}.$$

**Proof:** Recall Proposition 6.3, and choose  $N_{\varepsilon_1, \varepsilon_2}$  and  $n_{\varepsilon_1, \varepsilon_2}$  large so that

$$\mathbb{P}(\text{diam}(\text{NR}_n(\mathbf{w}(\lambda)) \setminus [N]) \leq \varepsilon_1 n^\eta/2) \geq 1 - \varepsilon_2 \quad (6.22)$$

for all  $n \geq n_{\varepsilon_1, \varepsilon_2}$ . Let

$$F_1 = \{\text{diam}(\text{NR}_n(\mathbf{w}(\lambda)) \setminus [N]) \leq \varepsilon_1 n^\eta/2\} \text{ and } F_2 = \{\text{diam}(\mathcal{C}(i)) > \varepsilon_1 n^\eta/2\}.$$

Clearly, on the set  $F_1 \cap F_2$ ,

$$\min_{v \in \mathcal{C}(i)} \left\{ \sum_{j \in B(v, \varepsilon_1)} w_j \right\} \geq \min_{k \in [N_{\varepsilon_1, \varepsilon_2}]} \left\{ \sum_{j \in B(k, \varepsilon_1/2)} w_j \right\}. \quad (6.23)$$

Recall the definition of  $\delta_{\varepsilon, i}$  in (6.17), and let

$$\Delta_{\varepsilon_1, \varepsilon_2} = \varepsilon_1 \wedge (\delta_{\varepsilon_2, 1} \wedge \cdots \wedge \delta_{\varepsilon_2, N_{\varepsilon_1, \varepsilon_2}})/2.$$

Then (6.23) implies

$$\min_{v \in \mathcal{C}(i)} \sum_{j \in B(v, \varepsilon_1)} w_j \geq \min_{k \in [N_{\varepsilon_1, \varepsilon_2}]} \sum_{j \in B(k, \Delta_{\varepsilon_1, \varepsilon_2})} w_j$$

on the set  $F_1 \cap F_2$ . Hence, for all  $n \geq n_{\varepsilon_1, \varepsilon_2}$ ,

$$\begin{aligned}
\mathbb{P} \left( F_1 \cap F_2 \cap \left\{ \min_{v \in \mathcal{C}(i)} \left\{ \sum_{j \in B(v, \varepsilon_1)} w_j \right\} \leq \left( \frac{c_F}{2N_{\varepsilon_1, \varepsilon_2}} \right)^{1/(\tau-1)} \frac{\Delta_{\varepsilon_1, \varepsilon_2} n^\rho}{2} \right\} \right) \\
\leq \sum_{k=1}^{N_{\varepsilon_1, \varepsilon_2}} \mathbb{P} \left( \sum_{j \in B(k, \Delta_{\varepsilon_1, \varepsilon_2})} w_j \leq \left( \frac{c_F}{2N_{\varepsilon_1, \varepsilon_2}} \right)^{1/(\tau-1)} \frac{\Delta_{\varepsilon_1, \varepsilon_2}}{2} n^\rho \right) \\
\leq \sum_{k=1}^{N_{\varepsilon_1, \varepsilon_2}} \left( n \exp \left( -\frac{J \Delta_{\varepsilon_1, \varepsilon_2} n^\rho}{N_{\varepsilon_1, \varepsilon_2}^{1/(\tau-1)}} \right) + \frac{\varepsilon_2}{2^k} \right) \\
\leq n^2 \exp \left( -\frac{J \Delta_{\varepsilon_1, \varepsilon_2} n^\rho}{N_{\varepsilon_1, \varepsilon_2}^{1/(\tau-1)}} \right) + \varepsilon_2,
\end{aligned} \tag{6.24}$$

where the second inequality is a consequence of Lemma 6.9.

Next, on the set  $F_1 \cap F_2^c$ ,

$$\sum_{j \in B(v, \varepsilon_1)} w_j = \sum_{j \in \mathcal{C}(i)} w_j$$

for any  $v \in \mathcal{C}(i)$ . Further, by [17, Theorem 1.4],  $n^{-\rho} \sum_{j \in \mathcal{C}(i)} w_j$  converges in distribution to a positive random variable. Hence, there exists  $\xi_{\varepsilon_2}^{(i)} > 0$  such that

$$\begin{aligned}
\limsup_{n \rightarrow \infty} \mathbb{P} \left( F_1 \cap F_2^c \cap \left\{ \min_{v \in \mathcal{C}(i)} \left( \sum_{j \in B(v, \varepsilon_1)} w_j \right) \leq \xi_{\varepsilon_2}^{(i)} n^\rho \right\} \right) \\
\leq \limsup_{n \rightarrow \infty} \mathbb{P} \left( \sum_{j \in \mathcal{C}(i)} w_j \leq \xi_{\varepsilon_2}^{(i)} n^\rho \right) \leq \varepsilon_2.
\end{aligned} \tag{6.25}$$

The result follows upon combining (6.22), (6.24) and (6.25). ■

We are now ready for

**Proof of Lemma 6.1:** Using Proposition 6.2, for any  $i \geq 1$  and  $\varepsilon > 0$ , we can choose  $K$  such that

$$\mathbb{P}(\mathcal{C}_i(\lambda) \cap [K] = \emptyset) \leq \varepsilon/2. \tag{6.26}$$

By Lemma 6.10, we can choose  $\xi > 0$  and an integer  $\bar{n}$  such that

$$\mathbb{P} \left( \min_{v \in \mathcal{C}(k)} \left( \sum_{j \in B(v, \delta)} w_j \right) \leq \xi n^\rho \right) \leq \varepsilon/(2K) \tag{6.27}$$

for all  $n \geq \bar{n}$  and  $k \in [K]$ . Combining (6.26) and (6.27), we see that

$$\mathbb{P} \left( \left( m_i^{(n)}(\delta) \right)^{-1} > 1/\xi \right) \leq \varepsilon \text{ for } n \geq \bar{n}$$

which yields the desired tightness. ■

## 7. PROOFS: FRACTAL DIMENSION

In this section, we prove the assertion about the upper box-counting dimension. We first prove a similar result for the component of  $j$ ,  $\mathcal{C}(j)$ . Consider  $\mathcal{C}(1)$ , and as usual, view  $\mathcal{C}(1)$  as a metric measure space via the graph distance and by assigning mass  $p_v := w_v / (\sum_{\ell \in \mathcal{C}(1)} w_\ell)$  to vertex  $v \in \mathcal{C}(1)$ . Set  $\mathbf{p} := (p_v : v \in \mathcal{C}(1))$ . Now note that conditional on the set  $\{v \in \mathcal{C}(1)\}$ ,  $\mathcal{C}(1)$  is the graph  $\tilde{\mathcal{G}}_m(\mathbf{p}, a)$  where  $a = (1 + \lambda n^{-\eta}) (\sum_{j \in \mathcal{C}(1)} w_j)^2 / \ell_n$ . Using [17, Proposition 3.7] and [17, Lemma 3.1], it is easy to verify that the conditions in Assumption 4.4 hold with this choice



of  $a$  and  $\mathbf{p}$ . Thus, by Theorem 4.5,  $n^{-\eta}\mathcal{C}(1)$  converges in Gromov-weak topology to a limiting space which we denote by  $\mathcal{M}(1)$ . Further, the sequence  $\{n^{-\eta}\mathcal{C}(1)\}_{n \geq 1}$  satisfies the global lower mass-bound property by Lemma 6.10. Hence

$$n^{-\eta}\mathcal{C}(1) \xrightarrow{w} \mathcal{M}(1) \quad (7.1)$$

with respect to Gromov-Hausdorff-Prokhorov topology. By similar arguments, we can show that  $n^{-\eta}\mathcal{C}(j) \xrightarrow{w} \mathcal{M}(j)$  with respect to Gromov-Hausdorff-Prokhorov topology for any  $j \geq 1$  and an appropriate (random) compact metric measure space  $\mathcal{M}(j)$ . The key ingredient in the proof is the following lemma.

**Lemma 7.1.** *Writing  $\pi = (\tau - 2)/(\tau - 3)$ , we have*

$$\mathbb{P}\left(\overline{\dim}(\mathcal{M}(j)) > \pi\right) = 0$$

for every  $j \geq 1$ .

**Proof:** For simplicity, we work with  $j = 1$ . The proof is similar for any  $j \geq 2$ . Recall that  $\mathcal{N}(\mathcal{M}, \delta)$  denotes the minimum number of open balls of radius  $\delta$  needed to cover the compact space  $\mathcal{M}$ . Write

$$\mathfrak{N}_{(\infty)}(\varepsilon) := \mathcal{N}(\mathcal{M}(1), \varepsilon) \text{ and } \mathfrak{N}_{(n)}(\varepsilon) := \mathcal{N}(\mathcal{C}(1), \varepsilon n^\eta).$$

Since the convergence in (7.1) holds with respect to Gromov-Hausdorff topology,

$$\mathbb{P}\left(\mathfrak{N}_{(\infty)}(2\varepsilon) > x\right) \leq \limsup_n \mathbb{P}\left(\mathfrak{N}_{(n)}(\varepsilon) > x\right) \quad (7.2)$$

for every  $x, \varepsilon > 0$ .

Fix an arbitrary  $\delta > 0$  and for any  $\varepsilon > 0$ , define

$$x_\varepsilon := \varepsilon^{-\delta-\pi}, \quad u_\varepsilon := |\log \varepsilon|, \quad \delta' := \frac{\delta}{2} \left( \frac{\tau-1}{\tau-2} \right), \text{ and } N(\varepsilon) = \varepsilon^{-\delta'-1/\eta}. \quad (7.3)$$

Let  $E_n$  be the event defined in (6.3). Clearly, on the event  $E_n \cap \{\mathfrak{N}_{(n)}(\varepsilon) > x_\varepsilon\}$ , any  $v \in \mathcal{C}(1)$  is within distance  $\varepsilon n^\eta$  from a point in  $\mathcal{C}(1) \cap [N(\varepsilon)]$ . Hence

$$\mathbb{P}\left(\mathfrak{N}_{(n)}(\varepsilon) > x_\varepsilon\right) \leq \mathbb{P}(E_n^c) + \mathbb{P}(|\mathcal{C}(1) \cap [N(\varepsilon)]| \geq x_\varepsilon), \quad (7.4)$$

and by Proposition 6.3,

$$\limsup_n \mathbb{P}(E_n^c) \leq c_{\delta'} \exp\left(-C/\varepsilon^{\delta'\eta}\right). \quad (7.5)$$

It remains to prove a bound on  $\mathbb{P}(|\mathcal{C}(1) \cap [N(\varepsilon)]| \geq x_\varepsilon)$ . To this end, note that by [17, Proposition 3.7],

$$|\mathcal{C}(1) \cap [N(\varepsilon)]| \xrightarrow{w} \sum_{q=1}^{N(\varepsilon)} \mathcal{I}_q\left(\mathcal{H}_{\mathcal{S}(1)}(0)\right), \quad (7.6)$$

where  $\mathcal{I}_q(\cdot)$  and  $\mathcal{H}_{\mathcal{S}(1)}(\cdot)$  are as defined around (6.15). Further, [45, Theorem 1.4] implies the existence of positive constants  $A_1$  and  $A_2$  such that

$$\mathbb{P}\left(\mathcal{H}_{\mathcal{S}(1)}(0) > u_\varepsilon\right) \leq A_1 \exp(-A_2 u_\varepsilon^{\tau-1}). \quad (7.7)$$

Combining (7.4), (7.5), (7.6) and (7.7), we conclude that

$$\limsup_n \mathbb{P}\left(\mathfrak{N}_{(n)}(\varepsilon) > x_\varepsilon\right) \leq c_{\delta'} \exp\left(-C\varepsilon^{-\delta'\eta}\right) + A_1 \exp(-A_2 u_\varepsilon^{\tau-1}) + \mathbb{P}\left(\sum_{q=1}^{N(\varepsilon)} \mathcal{I}_q(u_\varepsilon) \geq x_\varepsilon\right). \quad (7.8)$$

Now  $\mathcal{J}_q(u_\varepsilon)$  are i.i.d. Bernoulli random variables with

$$\mathbb{P}(\mathcal{J}_q(u_\varepsilon) = 1) = 1 - \exp(-au_\varepsilon/q^{1/(\tau-1)}) =: p_q$$

where  $a$  is as in (6.14). Choose  $\nu > 0$  small so that  $e^\nu - 1 \leq 2\nu$ . Clearly

$$\mathbb{E} \exp(\nu \mathcal{J}_q(u_\varepsilon)) = 1 + p_q(e^\nu - 1) \leq \exp(p_q(e^\nu - 1)) \leq \exp(2\nu p_q).$$

Hence, there exists a constant  $A_3 > 0$  such that

$$\begin{aligned} \mathbb{P}\left(\sum_{q=1}^{N(\varepsilon)} \mathcal{J}_q(u_\varepsilon) \geq x_\varepsilon\right) &\leq \exp\left(-\nu x_\varepsilon + 2\nu \sum_{q=1}^{N(\varepsilon)} p_q\right) \\ &\leq \exp\left(-\nu x_\varepsilon + 2\nu A_3 u_\varepsilon N(\varepsilon)^\rho\right) \\ &= \exp\left(-\nu x_\varepsilon + 2\nu A_3 u_\varepsilon \varepsilon^{-\frac{\delta}{2}-\pi}\right). \end{aligned} \tag{7.9}$$

Combining (7.2), (7.8) and (7.9), we see that  $\sum_{k=1}^\infty \mathbb{P}(\mathfrak{N}_{(\infty)}(2/k) > k^{\delta+\pi}) < \infty$ . Since  $\delta > 0$  was arbitrary, we conclude that

$$\limsup_k \frac{\log(\mathfrak{N}_{(\infty)}(2/k))}{\log(k/2)} \leq \pi \text{ a.s.}$$

By sandwiching  $\varepsilon$  between  $2/(k-1)$  and  $2/k$ , we get the desired bound on  $\overline{\dim}(\mathcal{M}(1))$ .  $\blacksquare$

**Proof of (1.10) and (1.23):** We only give the proof of (1.10). This will imply (1.23) because of (2.18). Fix  $i \geq 1$  and let

$$K_n := \min \{j \in [n] \mid j \in \mathcal{C}_i(\lambda)\}.$$

By Proposition 6.2,  $K_n$  is tight. By passing to a subsequence if necessary, we can assume that we are working on a space where

$$(n^{-\eta} \mathbf{M}_n^{\text{nr}}(\lambda), K_n) \rightarrow (\mathbf{M}_\infty^{\text{nr}}(\lambda), K_\infty) \quad \text{a.s.}$$

for some (integer valued) random variable  $K_\infty$ . Then

$$\mathbb{P}\left(\overline{\dim}(M_i^{\text{nr}}(\lambda)) > \pi\right) = \sum_{j=1}^\infty \mathbb{P}\left(\overline{\dim}(M_i^{\text{nr}}(\lambda)) > \pi, K_\infty = j\right).$$

By Lemma 7.1,  $\mathbb{P}\left(\overline{\dim}(M_i^{\text{nr}}(\lambda)) > \pi, K_\infty = j\right) = 0$  for every  $j \geq 1$ . This completes the proof.  $\blacksquare$

## 8. OPEN PROBLEMS

In Theorem 1.7, we have considered a general entrance boundary  $\mathbf{c} \in l_0$ . To study specific properties of the limit objects, we focused mainly on the special case  $\mathbf{c} = \mathbf{c}(\alpha, \tau)$  as in (1.22) and in this case, we have shown compactness as well as established an upper bound on the box counting dimension in Theorem 1.8. An important problem in this context is to establish necessary and sufficient conditions on  $\mathbf{c}$  that ensure compactness of the limiting spaces.

Another motivation for pursuing this problem comes from the following simple Corollary of Theorem 1.8: For any  $i \geq 1$ , consider the sequence  $\boldsymbol{\theta}^{(i)}$  as in (2.13). Then  $\mathcal{T}_{(\infty)}^{\boldsymbol{\theta}^{(i)}}$  is almost surely compact. Similarly, compactness of  $\mathcal{M}(1)$  (as defined in (7.1)) implies compactness of the associated ICRT  $\mathcal{T}_{(\infty)}^{\bar{\boldsymbol{\theta}}}$  where  $\bar{\boldsymbol{\theta}} = (\bar{\theta}_i : i \geq 1)$  is given by the following prescription: Let  $q_k$  be such

that

$$\sum_{q=1}^{q_k} \mathcal{I}_q(\mathcal{H}_{\mathcal{S}(1)}(0)) = k$$

where  $\mathcal{I}_q(\cdot)$  and  $\mathcal{H}_{\mathcal{S}(1)}(\cdot)$  are as defined around (6.15). Define

$$\bar{\theta}_i = \frac{q_i^{-1/(\tau-1)}}{\left(\sum_{k=1}^{\infty} q_k^{-2/(\tau-1)}\right)^{1/2}} \text{ for } i \geq 1.$$

These can be thought of as “annealed results,” since  $\theta^{(i)}$  and  $\bar{\theta}$  are random. No result is known in this direction without a prior distribution on  $\theta$ , i.e., sufficient conditions on non-random  $\theta \in \Theta$  that ensure compactness of the tree  $\mathcal{T}_{(\infty)}^{\theta}$  are not known. In [11, Section 7], Aldous, Miermont and Pitman conjecture that boundedness of  $\mathcal{T}_{(\infty)}^{\theta}$  for  $\theta \in \Theta$  is equivalent to  $\int_1^{\infty} (\psi_{\theta}(u))^{-1} du$ , where  $\psi_{\theta}$ , in our situation, is given by

$$\psi_{\theta}(u) = \sum_{i=1}^{\infty} (\exp(-u\theta_i) - 1 + u\theta_i). \quad (8.1)$$

This conjecture, however, is open to date. Our proof technique demonstrates a method of proving such annealed results via approximation by random graphs. Thus, classification of those  $\mathbf{c} \in l_0$  for which the spaces  $M_i^{\mathbf{c}}(\lambda)$  are compact will lead to a broad class of prior distributions on  $\theta$  for which  $\mathcal{T}_{(\infty)}^{\theta}$  is compact.

**Problem 8.1.** *Find necessary and sufficient conditions on  $\mathbf{c}$  that ensure compactness of the spaces  $M_i^{\mathbf{c}}(\lambda)$  for  $i \geq 1$ .*

Another related problem is to find the fractal dimensions of the limiting spaces. As a corollary to Theorem 1.8, we get

$$\overline{\dim}(\mathcal{T}_{(\infty)}^{\theta^{(i)}}) \leq (\tau - 2)/(\tau - 3) \text{ a.s.} \quad (8.2)$$

where  $\theta^{(i)}$  is as in (2.13) corresponding to  $\mathbf{c}$  of the form (1.22). Lemma 7.1 shows that the assertion in (8.2) remains true if we replace  $\theta^{(i)}$  by  $\bar{\theta}$ . Now, it is not hard to prove that

$$\inf_j \bar{\theta}_j j^{1/(\tau-2)} > 0 \text{ a.s. and } \sup_j \bar{\theta}_j j^{1/(\tau-2)} < \infty \text{ a.s.}$$

It then follows that

$$\tau - 2 = \sup \left\{ a \geq 0 : \lim_{u \rightarrow \infty} u^{-a} \psi_{\bar{\theta}}(u) = \infty \right\} = \inf \left\{ a \geq 0 : \lim_{u \rightarrow \infty} u^{-a} \psi_{\bar{\theta}}(u) = 0 \right\} \text{ a.s.,}$$

which in turn implies that the Minkowski dimension, Hausdorff dimension and packing dimension of a  $\psi_{\bar{\theta}}$  Lévy tree is  $(\tau - 2)/(\tau - 3)$  a.s. (see [34, 40]). Using the analogy between ICRTs and Lévy trees as in [11, Section 7], it is natural to expect that the same is true for  $\mathcal{T}_{(\infty)}^{\bar{\theta}}$  and hence for  $\mathcal{M}(1)$ . This is the heuristic behind Conjecture 1.2.

**Problem 8.2.** *Prove Conjecture 1.2.*

## ACKNOWLEDGEMENTS

The authors are indebted to Grégory Miermont for many enlightening discussions about inhomogeneous continuum random trees. SS thanks ENS Lyon for hospitality and accommodation during visits. The authors thank Igor Kortchemski for drawing their attention to his recent preprint [50]. SB has been partially supported by NSF-DMS grants 1105581 and 1310002 and SES grant 1357622. RvdH and SS have been supported in part by the Netherlands Organisation for Scientific Research (NWO) through the Gravitation Networks grant 024.002.003. In addition, RvdH has been supported by VICI grant 639.033.806.

## REFERENCES

- [1] R. Abraham, J.-F. Delmas, and P. Hoscheit, *A note on the Gromov-Hausdorff-Prokhorov distance between (locally) compact metric measure spaces*, Electron. J. Probab. **18** (2013), no. 14, 1–21.
- [2] D. Achlioptas, R. M D’Souza, and J. Spencer, *Explosive percolation in random networks*, Science **323** (2009), no. 5920, 1453–1455.
- [3] L. Addario-Berry, N. Broutin, and C. Goldschmidt, *Critical random graphs: limiting constructions and distributional properties*, Electron. J. Probab. **15** (2010), no. 25, 741–775. MR2650781 (2011d:60025)
- [4] L. Addario-Berry, N. Broutin, and C. Goldschmidt, *The continuum limit of critical random graphs*, Probab. Theory Related Fields **152** (2012), no. 3-4, 367–406. MR2892951
- [5] L. Addario-Berry, N. Broutin, C. Goldschmidt, and G. Miermont, *The scaling limit of the minimum spanning tree of the complete graph*, arXiv preprint arXiv:1301.1664 (2013).
- [6] R. Albert and A.-L. Barabási, *Statistical mechanics of complex networks*, Reviews of modern physics **74** (2002), no. 1, 47.
- [7] D. Aldous, *The continuum random tree. I*, Ann. Probab. **19** ((1991)), 1–28.
- [8] D. Aldous, *The continuum random tree III*, Ann. Probab. **21** ((1993)), 248–289.
- [9] D. Aldous, *Brownian excursions, critical random graphs and the multiplicative coalescent*, Ann. Probab. **25** (1997), no. 2, 812–854. MR1434128 (98d:60019)
- [10] D. Aldous and V. Limic, *The entrance boundary of the multiplicative coalescent*, Electron. J. Probab. **3** (1998), No. 3, 59 pp. (electronic). MR1491528 (99d:60086)
- [11] D. Aldous, G. Miermont, and J. Pitman, *The exploration process of inhomogeneous continuum random trees, and an extension of Jeulin’s local time identity*, Probab. Theory Related Fields **129** (2004), no. 2, 182–218. MR2063375 (2005f:60023)
- [12] D. Aldous and J. Pitman, *A family of random trees with random edge lengths*, Random Structures Algorithms **15** (1999), no. 2, 176–195. MR1704343 (2000k:05242)
- [13] D. Aldous and J. Pitman, *Inhomogeneous continuum random trees and the entrance boundary of the additive coalescent*, Probab. Theory Related Fields **118** (2000), no. 4, 455–482. MR1808372 (2002a:60012)
- [14] S. Athreya, W. Löhner, and A. Winter, *The gap between Gromov-vague and Gromov-Hausdorff-vague topology*, arXiv preprint arXiv:1407.6309 (2014).
- [15] J. Bertoin, *Random fragmentation and coagulation processes*, Cambridge Studies in Advanced Mathematics, vol. 102, Cambridge University Press, Cambridge, 2006. MR2253162 (2007k:60004)
- [16] S. Bhamidi, R. van der Hofstad, and J. S. H. van Leeuwen, *Scaling limits for critical inhomogeneous random graphs with finite third moments*, Electronic Journal of Probability **15** (2010), no. 54, 1682.
- [17] S. Bhamidi, R. van der Hofstad, and J. S. H. van Leeuwen, *Novel scaling limits for critical inhomogeneous random graphs*, Ann. Probab. **40** (2012), no. 6, 2299–2361. MR3050505
- [18] 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).
- [19] S. Bhamidi, S. Dhara, R. v. d. Hofstad, and S. Sen, *Continuum scaling limits of the configuration model*, In preparation (2016).
- [20] S. Bhamidi, S. Sen, and X. Wang, *Continuum limit of critical inhomogeneous random graphs*, arXiv preprint arXiv:1404.4118 (2014).
- [21] B. Bollobás, *Random graphs*, Cambridge University Press, 2001.
- [22] B. Bollobás, S. Janson, and O. Riordan, *The phase transition in inhomogeneous random graphs*, Random Structures Algorithms **31** (2007), no. 1, 3–122. MR2337396 (2008e:05124)

- [23] L. A. Braunstein, S. V. Buldyrev, R. Cohen, S. Havlin, and H. E. Stanley, *Optimal paths in disordered complex networks*, Physical review letters **91** ((2003)), no. 16, 168701.
- [24] L. A. Braunstein, Z. Wu, Y. Chen, S. V. Buldyrev, T. Kalisky, S. Sreenivasan, R. Cohen, E. Lopez, S. Havlin, and H. E. Stanley, *Optimal path and minimal spanning trees in random weighted networks*, International Journal of Bifurcation and Chaos **17** (2007), no. 07, 2215–2255.
- [25] T. Britton, M. Deijfen, and A. Martin-Löf, *Generating simple random graphs with prescribed degree distribution*, J. Stat. Phys. **124** (2006), no. 6, 1377–1397. MR2266448 (2007g:05168)
- [26] D. Burago, Y. Burago, and S. Ivanov, *A course in metric geometry*, Graduate Studies in Mathematics, vol. 33, American Mathematical Society, Providence, RI, 2001. MR1835418 (2002e:53053)
- [27] M. Camarri and J. Pitman, *Limit distributions and random trees derived from the birthday problem with unequal probabilities*, Electron. J. Probab. **5** (2000), no. 2, 18 pp. (electronic). MR1741774 (2001c:60080)
- [28] Y. Chen, E. López, S. Havlin, and H. E. Stanley, *Universal behavior of optimal paths in weighted networks with general disorder*, Physical review letters **96** (2006), no. 6, 068702.
- [29] F. Chung and L. Lu, *The average distances in random graphs with given expected degrees*, Proc. Natl. Acad. Sci. USA **99** (2002), no. 25, 15879–15882 (electronic). MR1944974 (2003k:05124)
- [30] F. Chung and L. Lu, *Connected components in random graphs with given expected degree sequences*, Ann. Comb. **6** (2002), no. 2, 125–145. MR1955514 (2003k:05123)
- [31] F. Chung and L. Lu, *The average distance in a random graph with given expected degrees*, Internet Math. **1** (2003), no. 1, 91–113. MR2076728 (2005e:05122)
- [32] F. Chung and L. Lu, *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 American Mathematical Society, Providence, RI, 2006. MR2248695 (2007i:05169)
- [33] S. N. Dorogovtsev and J. F. F. Mendes, *Evolution of networks*, Advances in physics **51** (2002), no. 4, 1079–1187.
- [34] T. Duquesne and J.-F. Le Gall, *Probabilistic and fractal aspects of lévy trees*, Probability Theory and Related Fields **131** (2005), no. 4, 553–603.
- [35] R. Durrett, *Random graph dynamics*, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press, Cambridge, 2007. MR2271734 (2008c:05167)
- [36] M. Dwass, *The total progeny in a branching process and a related random walk*, J. Appl. Prob. **6** (1969), 682–686.
- [37] S. N. Evans, *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. MR2351587 (2009d:60014)
- [38] A. Greven, P. Pfaffelhuber, and A. Winter, *Convergence in distribution of random metric measure spaces ( $\Lambda$ -coalescent measure trees)*, Probab. Theory Related Fields **145** (2009), no. 1-2, 285–322. MR2520129 (2011c:60008)
- [39] M. Gromov, *Metric structures for Riemannian and non-Riemannian spaces*, English, Modern Birkhäuser Classics, Birkhäuser Boston, Inc., Boston, MA, 2007. Based on the 1981 French original, With appendices by M. Katz, P. Pansu and S. Semmes, Translated from the French by Sean Michael Bates. MR2307192 (2007k:53049)
- [40] B. Haas and G. Miermont, *The genealogy of self-similar fragmentations with negative index as a continuum random tree*, Electron. J. Probab. **9** (2004), no. 4, 57–97 (electronic). MR2041829 (2004m:60086)
- [41] R. v. d. Hofstad, *Critical behavior in inhomogeneous random graphs*, Random Structures and Algorithms **42** (2013), no. 4, 480–508.
- [42] R. v. d. Hofstad, *Random graphs and complex networks. Vol. I*, (2014). In preparation, see <http://www.win.tue.nl/~rhofstad/NotesRGCN.pdf>.
- [43] R. v. d. Hofstad and M. Keane, *An elementary proof of the hitting time theorem*, Amer. Math. Monthly **115** (2008), no. 8, 753–756. MR2456097
- [44] R. v. d. Hofstad and A. Nachmias, *Hypercube percolation*, Preprint (2012). To appear in *Journ. Europ. Math. Soc.*
- [45] Rvd. Hofstad, S. Kliem, and JSHv. Leeuwaarden, *Cluster tails for critical power-law inhomogeneous random graphs*, arXiv preprint arXiv:1404.1727 (2014).
- [46] S. Janson, *Asymptotic equivalence and contiguity of some random graphs*, Random Structures Algorithms **36** (2010), no. 1, 26–45. MR2591045 (2011h:60021)
- [47] S. Janson, T. Luczak, and A. Rucinski, *Random graphs*, Vol. 45, John Wiley & Sons, 2011.
- [48] A. Joseph et al., *The component sizes of a critical random graph with given degree sequence*, The Annals of Applied Probability **24** (2014), no. 6, 2560–2594.

- [49] T. Klein and E. Rio, *Concentration around the mean for maxima of empirical processes*, Ann. Probab. **33** (2005), no. 3, 1060–1077. MR2135312 (2006c:60022)
- [50] I. Kortchemski, *Sub-exponential tail bounds for conditioned stable Bienaymé-Galton-Watson trees*, arXiv preprint arXiv:1504.04358 (2015).
- [51] J.-F. Le Gall, *Random trees and applications*, Probab. Surv. **2** (2005), 245–311. MR2203728 (2007h:60078)
- [52] T. Łuczak, *Component behavior near the critical point of the random graph process*, Random Structures & Algorithms **1** (1990), no. 3, 287–310.
- [53] T. Łuczak, B. Pittel, and J. C. Wierman, *The structure of a random graph at the point of the phase transition*, Transactions of the American Mathematical Society **341** (1994), no. 2, 721–748.
- [54] P. Massart, *The tight constant in the Dvoretzky-Kiefer-Wolfowitz inequality*, Ann. Probab. **18** (1990), no. 3, 1269–1283. MR1062069 (91i:60052)
- [55] M. Newman, *Networks: an introduction*, Oxford University Press, 2010.
- [56] M. E. Newman, *The structure and function of complex networks*, SIAM review **45** (2003), no. 2, 167–256.
- [57] I. Norros and H. Reittu, *On a conditionally Poissonian graph process*, Adv. in Appl. Probab. **38** ((2006)), no. 1, 59–75. MR2213964
- [58] R. Otter, *The multiplicative process*, Ann. Math. Statist. **20** ((1949)), 206–224. MR0030716 (11,41a)
- [59] J. Pitman, *Random mappings, forests, and subsets associated with Abel-Cayley-Hurwitz multinomial expansions*, Sémin. Lothar. Combin. **46** (2001/02), Art. B46h, 45 pp. (electronic). MR1877634 (2002m:60017)
- [60] O. Riordan and L. Warnke, *Explosive percolation is continuous*, Science **333** (2011), no. 6040, 322–324.
- [61] O. Schramm, *Scaling limits of loop-erased random walks and uniform spanning trees*, Israel Journal of Mathematics **118** (2000), no. 1, 221–288.
- [62] R. van der Hofstad, *Critical behavior in inhomogeneous random graphs*, Random Structures Algorithms **42** (2013), no. 4, 480–508. MR3068034
- [63] Z. Wu, L. A. Braunstein, S. Havlin, and H. E. Stanley, *Transport in weighted networks: partition into superhighways and roads*, Physical review letters **96** (2006), no. 14, 148702.