SCALING LIMITS OF MARKOV BRANCHING TREES AND GALTON-WATSON TREES CONDITIONED ON THE NUMBER OF VERTICES WITH OUT-DEGREE IN A GIVEN SET

DOUGLAS RIZZOLO

Abstract. We obtain scaling limits for Markov branching trees whose size is specified by the number of nodes whose out-degree lies in a given set. This extends recent results of Haas and Miermont in [5], which considered the case when the size of a tree is either its number of leaves or its number of vertices. We use our result to prove that the scaling limit of finite variance Galton-Watson trees conditioned on the number of nodes whose out-degree lies in a given set is the Brownian continuum random tree. The key to applying our result for Markov branching trees to conditioned Galton-Watson trees is a generalization of the classical Otter-Dwass formula. This is obtained by showing that the number of vertices in a Galton-Watson tree whose out-degree lies in a given set is distributed like the number of vertices in a Galton-Watson tree with a related offspring distribution.

1. Introduction

Recently there has been considerable interest in the literature in studying the asymptotic properties of random trees. One of the first results that obtained rescaled convergence of trees themselves (as opposed to convergence of some statistics of the trees) was a result of Aldous in [2], which proved that critical Galton-Watson trees with finite variance offspring distribution conditioned on their number of vertices converge to the Brownian continuum random tree when properly rescaled. Subsequent work on scaling limits of random trees proceeded largely in two directions, one focusing on conditioned Galton-Watson trees and other trees with nice encodings by continuous functions see [8] for an overview) and another that focused on Markov branching trees. While Galton-Watson trees conditioned on their number of vertices are Markov branching trees, the work on general Markov branching trees typically includes consistency hypotheses that are not satisfied by conditioned Galton-Watson trees (see e.g. [6]). More recently, techniques for handling Markov branching trees that don't necessarily satisfy consistency relations and whose size is their number of leaves (or vertices) were developed in [5]. We will generalize this approach to Markov branching trees whose size is their number of vertices whose out-degree falls in a given set. Our result for Markov branching trees, which requires a bit of setup to state rigorously, is Theorem 5.

As a consequence of Theorem 5 we obtain a new theorem for scaling limits of Galton-Watson trees. In the literature on Galton-Watson trees conditioned on their total number of vertices the techniques employed generally rely heavily on the precise form of the conditioning and cannot easily be modified. By using the framework of Markov branching trees instead,

This material is based upon work supported in part by the National Science Foundation Graduate Research Fellowship under Grant No. DGE 1106400 and in part by NSF DMS-1204840.

we are able to modify the conditioning to condition on the number of vertices with out-degree in a given set. In particular, we prove the following theorem, the notation for which will be fully defined later.

Theorem 1. *Let T be a critical Galton-Watson tree with offspring distribution ξ such that $0 < \sigma^2 = Var(\xi) < \infty$ and fix $A \subseteq \{0, 1, 2, \dots\}$. Suppose that for sufficiently large n the probability that T has exactly n vertices with out-degree in A is positive, and for such n let T_n^A be T conditioned to have exactly n vertices with out-degree in A, considered as a rooted unordered tree with edge lengths 1 and the uniform probability distribution $\mu_{\partial_A T_n^A}$ on its vertices with out-degree in A. Then*

$$\frac{1}{\sqrt{n}} T_n^A \xrightarrow{d} \frac{2}{\sigma\sqrt{\xi(A)}} T_{1/2,\nu_2},$$

where the convergence is with respect to the rooted Gromov-Hausdorff-Prokhorov topology and $T_{1/2,\nu_2}$ is the Brownian continuum random tree.

In the case $A = \mathbb{Z}^+ = \{0, 1, 2, \dots\}$ we recover the classical result about the scaling limit of a Galton-Watson tree conditioned on its number of vertices first obtained in [2]. For other choices of A the result appears to be new. We note, however, that subsequent to the appearance of a draft of this article on the arXiv ([12]), Kortchemski obtained similar results by different techniques in [7] and that local (rather than scaling) limits have been obtained in [1]. The condition that for sufficiently large n the probability that T has exactly n vertices with out-degree in A is positive is purely technical and could be dispensed with at the cost of chasing periodicity considerations through our computations. In addition to generalizing the results of [5], the key to proving this theorem is a generalization of the classical Otter-Dwass formula, which we prove in Section 3.1. The Otter-Dwass formula (see [11]) has been an essential tool in several proofs that the Brownian continuum random tree is the scaling limit of Galton-Watson trees conditioned on their number of vertices, including the original proof in [2] as well as newer proofs in [9] and [5]. While we follow the approach in [5], our generalization of the Otter-Dwass formula should allow for proofs along the lines of [2] and [9] as well. Furthermore, with our results here, it should be straightforward to prove the analogous theorem in the infinite variance case using the approach in [5].

This paper is organized as follows. Section 2 introduces our basic notation, as well as the Markov branching trees and continuum trees that will arise for us as scaling limits. It concludes with our generalization of the scaling limits in [5]. Section 3 is devoted to our study of Galton-Watson trees. We begin by proving our generalization of the Otter-Dwass formula and we then use this to analyze the asymptotics of the partition at the root of a Galton-Watson tree. Bringing this all together, we finish with the proof of Theorem 1.

2. Models of trees

2.1. **Basic notation.** Fix a countably infinite set S; we will consider the vertex sets of all graphs discussed to be subsets of S. A rooted ordered tree is a finite acyclic graph t with a distinguished vertex called the root and such that, if v is a vertex of t, the set of vertices in t that are both adjacent to v and further from the root than v with respect to the graph metric is linearly ordered. Let $\mathcal{T}^S$ be the set of all rooted ordered trees whose vertex set is

a subset of S. For $t, s \in \mathcal{T}^S$, define $t \sim s$ if and only if there is a root and order preserving isomorphism from t to s and let $\mathcal{T} = \mathcal{T}^S/\sim$ be the set of rooted ordered trees considered up to root and order preserving isomorphism. By a similar construction, we let $\mathcal{T}^u$ be the set of rooted unordered trees considered up to root preserving isomorphism. If t is in $\mathcal{T}$ or $\mathcal{T}^u$ and $v \in t$ is a vertex, the out-degree of v is the number of vertices in t that are both adjacent to v and further from the root than v with respect to the graph metric. The out-degree of v will simply be denoted by $\deg(v)$, since we will only ever discuss out-degrees. Fix a set $A \subseteq \mathbb{Z}^+$ and for t in $\mathcal{T}$ or $\mathcal{T}^u$ define $\#_A t$ to be the number of vertices in t whose out-degree is in A. Furthermore, we define $\mathcal{T}_{A,n}$ and $\mathcal{T}^u_{A,n}$ by

$$\mathcal{T}_{A,n} = \{t \in \mathcal{T} : \#_A t = n\} \quad \text{and} \quad \mathcal{T}^u_{A,n} = \{t \in \mathcal{T}^u : \#_A t = n\}.$$

2.2. **Markov branching trees.** In this section we extend the notion of Markov branching trees developed in [5], where Markov branching trees were constructed separately in the cases $A = \{0\}$ and $A = \mathbb{Z}^+$. Here we give a construction for general A. Define $\bar{\mathcal{P}}_{-1} = \emptyset$ and for $n \geq 0$, let $\bar{\mathcal{P}}_n$ be the set of partitions of n; that is

$$\bar{\mathcal{P}}_n = \left\{ (\lambda_1, \dots, \lambda_p) \in \bigcup_{k=0}^{\infty} \{0, 1, 2, \dots\}^k \ : \ \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \text{ and } \sum_{k=1}^{p} \lambda_k = n \right\}.$$

Remark 1. *This definition differs slightly from the standard definition of partitions because we allow blocks of a partition to be* 0.

For $\lambda \in \bar{\mathcal{P}}_n$, let $p(\lambda)$ be the length of λ and $m_j(\lambda)$ the number of blocks in λ equal to j. If $\lambda = (\lambda_1, \dots, \lambda_p) \in \bar{\mathcal{P}}_n$ and $k \geq 0$ we denote by $(\lambda, 0_k)$ the partition

$$(\lambda_1, \dots, \lambda_p, \overbrace{0, \dots, 0}^{k \text{ 0's}})$$

By convention, include $\emptyset = (0_0)$ in $\bar{\mathcal{P}}_0$ and take $p(\emptyset) = 0$. Define $\bar{\mathcal{P}}^A_0 = \bar{\mathcal{P}}_0$ and for $n \geq 1$, define $\bar{\mathcal{P}}^A_n$ by

$$\bar{\mathcal{P}}^A_n = \{\lambda \in \bar{\mathcal{P}}_n : p(\lambda) \notin A\} \cup \{\lambda \in \bar{\mathcal{P}}_{n-1} : p(\lambda) \in A\}.$$

Let (n_k) be an increasing sequence of non-negative integers and let $(q_{n_k})_{k \geq 1}$ be a sequence such that for each k such that:

(i) $n_1 = \mathbf{1}_{\{0 \in A\}}$.
(ii) If $n_k \geq 1$ then q_{n_k} is a probability measure on $\bar{\mathcal{P}}^A_{n_k}$.
(iii) If $n_1 = 0$, q_0 is a probability measure on $\mathcal{T}^u_{A,0}$.

Remark 2. *We emphasize that if $n_1 = 0$ then q_0 is a measure on trees that have no vertices with out-degree in A while for all $n_k \geq 1$, q_{n_k} is a measure on partitions. We do this so that we can present a unified construction of Markov branching trees instead of having to consider the cases $n_1 = 1$ and $n_1 = 0$ separately. Since the assumptions in Theorem 5 imply that the effects of trees that have no vertices with out-degree in A do not influence the scaling limits, we do not need to assume any further structure on q_0.*

For convenience, for $n_k \geq 1$ we will often consider q_{n_k} to be a measure on $\cup_{j \geq 0} \bar{\mathcal{P}}_j$ by extending it to be 0 off of $\bar{\mathcal{P}}_{n_k}$.

Our goal is to construct a sequence of laws $(\mathbf{P}^q_{n_k})_{k=1}^\infty$ such that $\mathbf{P}^q_{n_k}$ is a law on $\mathcal{T}^u_{A,n_k}$ and such that for $n_k \geq 1$ the subtrees above the root are conditionally independent given their sizes and the degree of the root. To facilitate this, we will assume that our sequences $(q_{n_k})_{k\geq 1}$ satisfy the following compatibility condition.

Condition 1. (i) *Either n_1 or n_2 is equal to 1.*

(ii) *For each k such that $n_k \geq 1$, q_{n_k} is concentrated on partitions $\lambda = (\lambda_1, \dots, \lambda_p)$ such that q_{λ_i} is defined for all i.*

(iii) *For all n such that q_n is defined, $\displaystyle\sum_{k=0}^\infty q_n((n, 0_k)) < 1$.*

Remark 3. *Note that these assumptions put nontrivial restrictions on the compatible sequences we consider. For example, if $n_1 = 1$, $1 \notin A$, and $2 \in A$ then q_2 cannot exist. To see this, observe that, since $n_1 = 1$, q_0 is not defined. By (ii), this implies that the only $\lambda \in \bar{\mathcal{P}}^A_2$ that can have $q_2(\lambda) > 0$ is $\lambda = (2)$ because all other elements of $\bar{\mathcal{P}}^A_2$ have a block that is equal to 0. However, by (iii), we must have $q_2((2)) < 1$, which contradicts that q_2 is a probability measure. In terms of the trees we are considering this is to be expected. For example, if $0, 2 \in A$, we must have $n_1 = 1$ since every tree has a leaf. Moreover, if a tree has root degree 2 then it has at least 2 leaves and, as a result, cannot possibly have exactly 2 vertices with out-degree in A.*

If $n_1 = 0$, we let $\mathbf{P}^q_0 = q_0$. For $n_k \geq 1$, define $\alpha_{n_k} = \sum_{j\geq 0} q_{n_k}((n_k, 0_j))$ and let G be geometrically distributed with

$$\mathbb{P}(G = j) = (1 - \alpha_{n_k})\alpha^j_{n_k}.$$

If $n_1 = 1$, $\mathbf{P}^q_1$ is the law of a path with G edges from a root to a leaf (note if $G = 0$, this path is just the single root vertex).

If $k \geq 2$, construct a path with G edges from a from a root to a leaf. For each non-leaf vertex v on this path, independently of everything else, construct a random variable V such that $\mathbb{P}(V = j) = q_{n_k}((n_k, 0_j))/\alpha_{n_k}$ and a vector $(T_1, \dots, T_V)$ of trees that, conditionally given V, are i.i.d with common distribution $\mathbf{P}^q_0$. Then attach the root of each tree T_i to v by an edge. For the leaf, choose a partition Λ according to $q_{n_k}(\cdots | \bar{\mathcal{P}}^A_{n_k} \setminus \cup_{j\geq 0}\{(n_k, 0_j)\})$ and let $(T^1_1, \dots, T^1_{p(\Lambda)})$ be a vector, independent of everything except Λ, such that conditionally given Λ the coordinates are independent and T^1_i has distribution $\mathbf{P}^q_{\Lambda_i}$. Attach the roots of these trees to the leaf by an edge to obtain a tree T. We define $\mathbf{P}^q_{n_k}$ to be the law of T.

To connect with [5], if $(n_k) = (1, 2, 3, \dots)$, the case $A = \{0\}$ corresponds to the $\mathbf{P}^q_n$ defined in [5] and the case $A = \mathbb{Z}^+$ corresponds to the $\mathbf{Q}^q_n$ defined in [5]. Other choices of A interpolate between these two extremes. A sequence $(T_{n_k})_{k\geq 1}$ such that for each k, T_{n_k} has law $\mathbf{P}^q_{n_k}$ for some choice of A and q (independent of n) is called a Markov branching family. For ease of notation, we will generally drop the subscript k and it will be implicit that we are only considering n for which the quantities discussed are defined.

2.3. **Trees as metric measure spaces.** The trees under consideration can naturally be considered as metric spaces with the graph metric. That is, the distance between two vertices is the number of edges on the path connecting them. Let (t, d) be a tree equipped with the

graph metric. For $a > 0$, we define at to be the metric space (t, ad), i.e. the metric is scaled by a. This is equivalent to saying the edges have length a rather than length 1 in the definition of the graph metric. More, generally we can attach a positive length to each edge in t and use these in the definition of the graph metric. Moreover, the trees we are dealing with are rooted so we consider (t, d) as a pointed metric space with the root as the distinguished point. Moreover, we are concerned with the vertices whose out-degree is in A, so we attach a measure $\mu_{\partial_A t}$, which is the uniform probability measure on $\partial_A t = \{v \in t : \deg(v) \in A\}$. If we have a random tree T, this gives rise to a random pointed metric measure space $(T, d, \text{root}, \mu_{\partial_A T})$. To make this last concept rigorous, we need to put a topology on pointed metric measure spaces. This is hard to do in general, but note that the pointed metric measure spaces that come from the trees we are discussing are compact.

Let $\mathcal{M}_w$ be the set of equivalence classes of compact pointed metric measure spaces (equivalence here being up to isometries that preserve the distinguished point and the measure). It is worth pointing out that $\mathcal{M}_w$ actually is a set in the sense of Zermelo-Fraenkel set theory with the axiom of choice, though this takes some work to show (more precisely, there exists a set $\mathcal{M}_w$ of compact pointed metric measure spaces such that every compact pointed metric measure space is equivalent to exactly one element of $\mathcal{M}_w$). We metrize $\mathcal{M}_w$ with the pointed Gromov-Hausdorff-Prokhorov metric (see [5]). Fix $(X, d, \rho, \mu), (X', d', \rho, \mu') \in \mathcal{M}_w$ and define

$$d_{\text{GHP}}(X, X') = \inf_{(M,\delta)} \inf_{\substack{\phi: X \to M \\ \phi': X' \to M}} \left[\delta(\phi(\rho), \phi'(\rho')) \vee \delta_H(\phi(X), \phi'(X')) \vee \delta_P(\phi_* \mu, \phi'_* \mu')\right],$$

where the first infimum is over metric spaces (M, δ), the second infimum if over isometric embeddings ϕ and ϕ' of X and X' into M, δ_H is the Hausdorff distance on compact subsets of M, and δ_P is the Prokhorov distance between the pushforward $\phi_* \mu$ of μ by ϕ and the pushforward $\phi'_* \mu'$ of μ' by ϕ'. Again, the definition of this metric has potential to run into set-theoretic difficulties, but they are not terribly difficult to resolve.

Proposition 1 (Proposition 1 in [5])**.** *The space $(\mathcal{M}_w, d_{GHP})$ is a complete separable metric space.*

We will not need many technical facts about Gromov-Hausdorff-Prokhorov convergence, but we will need the following lemma that shows how the pointed Gromov-Hausdorff-Prokhorov metric can be controlled in some cases by the pointed Hausdorff metric.

Lemma 1. *Let (E, d) be a metric space and let $U, V \subseteq E$ be compact subsets, each with a distinguished point. Let $\mu = n^{-1} \sum_{i=1}^n \delta_{x_i}$ and $\nu = n^{-1} \sum_{i=1}^n \delta_{y_i}$ be probability measures whose supports are contained in U and V respectively. Let $f : \{x_1, \ldots, x_n\} \to \{y_1, \ldots, y_n\}$ be a bijection. Then*

$$d_{GHP}((U, \mu), (V, \nu)) \leq d_H(U, V) + \max\{d(x_i, f(x_i)), 1 \leq i \leq n\},$$

where d_H is the pointed Hausdorff metric on pointed subsets of E.

Proof. Let B be a measurable set and define $B^\epsilon = \{y \in E \ : \ d(y, B) < \epsilon\}$. Suppose that $\epsilon > \max\{d(x_i, f(x_i)), 1 \leq i \leq n\}$. Observe that if $x_i \in B$ then $f(x_i) \in B^\epsilon$. Since f is a bijection, we see that $\sum_{i=1}^n \mathbf{1}_{x_i \in B} \leq \sum_{i=1}^n \mathbf{1}_{y_i \in B^\epsilon}$, so $\mu(B) \leq \nu(B^\epsilon)$. The reverse inequality is shown similarly and the conclusion follows. □

An $\mathbb{R}$-tree is a complete metric space (T, d) with the following properties:

- For $v, w \in T$, there exists a unique isometry $\phi_{v,w} : [0, d(v, w)] \to T$ with $\phi_{v,w}(0) = v$ to $\phi_{v,w}(d(v, w)) = w$.
- For every continuous injective function $c : [0, 1] \to T$ such that $c(0) = v$ and $c(1) = w$, we have $c([0, 1]) = \phi_{v,w}([0, d(v, w)])$.

If (T, d) is a compact $\mathbb{R}$-tree, every choice of root $\rho \in T$ and probability measure μ on T yields an element (T, d, ρ, μ) of $\mathcal{M}_w$. With this choice of root also comes a height function $\mathrm{ht}(v) = d(v, \rho)$. The leaves of T can then be defined as a point $v \in T$ such that v is not in $[[\rho, w[[:= \phi_{\rho,w}([0, \mathrm{ht}(w)))$ for any $w \in T$. The set of leaves is denoted $\mathcal{L}(T)$.

Definition 1. *A continuum tree is an $\mathbb{R}$-tree (T, d, ρ, μ) with a choice of root and probability measure such that μ is non-atomic, $\mu(\mathcal{L}(T)) = 1$, and for every non-leaf vertex w, $\mu\{v \in T : [[\rho, v]] \cap [[\rho, w]] = [[\rho, w]]\} > 0$.*

The last condition says that there is a positive mass of leaves above every non-leaf vertex. We will usually just refer to a continuum tree T, leaving the metric, root, and measure as implicit. A continuum random tree (CRT) is an $(\mathcal{M}_w, d_{GHP})$ valued random variable that is almost surely a continuum tree. The continuum random trees we will be interested in are those associated with self-similar fragmentation processes.

2.4. **Self-similar fragmentations.** In this section, we recall the basic facts about self-similar fragmentations, see [3, Chapter 3] for details. For any set B, let $\mathcal{P}_B$ be the set of countable partitions of B, i.e. countable collections of disjoint sets whose union is B. For $n \in \overline{\mathbb{N}} := \mathbb{N} \cup \{\infty\}$, let $\mathcal{P}_n := \mathcal{P}_{[n]}$, where $[n] = \{1, \dots, n\}$ and $[\infty] = \mathbb{N}$. Suppose that $\pi = (\pi_1, \pi_2, \dots) \in \mathcal{P}_n$ (here and throughout we index the blocks of π in increasing order of their least elements), and $B \subseteq \mathbb{N}$. Define the restriction of π to B, denoted by $\pi_{|B}$ or $\pi \cap B$, to be the partition of $[n] \cap B$ whose elements are the blocks $\pi_i \cap B$, $i \geq 1$. We equip $\mathcal{P}_n$ with the topology induced by the metric $d(\pi, \sigma) := (\inf\{i \ : \pi \cap [i] \neq \sigma \cap [i]\})^{-1}$.

Definition 2 (Definition 3.1 in [3])**.** *Consider two blocks $B \subseteq B' \subseteq \mathbb{N}$. Let π be a partition of B with $\#\pi = n$ non-empty blocks ($n = \infty$ is allowed), and $\pi^{(\cdot)} = \{\pi^{(i)}, i \in [n]\}$ be a sequence in $\mathcal{P}_{B'}$. For every integer i, we consider the partition of the i-th block π_i of π induced by the i-th term $\pi^{(i)}$ of the sequence $\pi^{(\cdot)}$, that is,*

$$\pi^{(i)}_{|\pi_i} = \left(\pi^{(i)}_j \cap \pi_i \; j \in \mathbb{N}\right).$$

As i varies in $[n]$, the collection $\left\{\pi^{(i)}_j \cap \pi_i : i, j \in \mathbb{N}\right\}$ forms a partition of B, which we denote by $\mathrm{Frag}(\pi, \pi^{(\cdot)})$ and call the fragmentation of π by $\pi^{(\cdot)}$.

We will use the Frag function to define the transition kernels of our fragmentation processes. Define

$$\mathcal{S}^{\downarrow} = \left\{(s_1, s_2, \dots) : s_1 \geq s_2 \geq \cdots \geq 0, \sum_{i=1}^{\infty} s_i \leq 1\right\},$$

and

$$\mathcal{S}_1 = \left\{(s_1, s_2, \dots) \in [0, 1]^{\mathbb{N}} \mid \sum_{i=1}^{\infty} s_i \leq 1\right\},$$

and endow both with the topology they inherit as subsets of $[0,1]^{\mathbb{N}}$ with the product topology. Observe that $\mathcal{S}^{\downarrow}$ and $\mathcal{S}_1$ are compact. For a partition $\pi \in \mathcal{P}_\infty$, we define the asymptotic frequency $|\pi_i|$ of the i'th block by $|\pi_i| = \lim_{n\to\infty} n^{-1}|\pi_i \cap [n]|$, provided this limit exists. If all of the blocks of π have asymptotic frequencies, we define $|\pi| \in \mathcal{S}_1$ by $|\pi| = (|\pi_1|, |\pi_2|, \dots)$.

Definition 3 (Definition 3.3 in [3])**.** *Let $(\Pi(t), t \geq 0)$ be an exchangeable, càdlàg $\mathcal{P}_\infty$-valued process such that $\Pi(0) = \mathbf{1}_{\mathbb{N}} := (\mathbb{N}, 0, \dots)$ and such that*

(1) *$\Pi(t)$ almost surely possesses asymptotic frequencies $|\Pi(t)|$ simultaneously for all $t \geq 0$ and*
(2) *if we denote by $B_i(t)$ the block of $\Pi(t)$ which contains i, then the process $t \mapsto |B_i(t)|$ has right-continuous paths.*

We call Π a self-similar fragmentation process with index $\alpha \in \mathbb{R}$ if and only if, for every $t, t' \geq 0$, the conditional distribution of $\Pi(t+t')$ given $\mathcal{F}_t$ is that of the law of $\mathrm{Frag}(\pi, \pi^{(\cdot)})$, where $\pi = \Pi(t)$ and $\pi^{(\cdot)} = \left(\pi^{(i)}, i \in \mathbb{N}\right)$ is a family of independent random partitions such that for $i \in \mathbb{N}$, $\pi^{(i)}$ has the same distribution as $\Pi(t'|\pi_i|^\alpha)$.

Suppose, for the moment, that Π is a self-similar fragmentation process with $\alpha = 0$ (these are also called homogeneous fragmentations). In this case Π is a Feller process as is $\Pi_{|[n]}$ for every n. For $\pi \in \mathcal{P}_n \setminus \{\mathbf{1}_{[n]}\}$, let

$$q_\pi = \lim_{t\to 0^+} \frac{1}{t}\mathbb{P}(\Pi_{|[n]}(t) = \pi).$$

For $\pi \in \mathcal{P}_n$ and $n' \in \{n, n+1, \dots, \infty\}$, define $\mathcal{P}_{n',\pi} = \{\pi' \in \mathcal{P}_{n'} : \pi'_{|[n]} = \pi\}$. One way to construct a homogeneous fragmentation is by specifying these transition rates, which can be done using measures on $\mathcal{S}^{\downarrow}$. The basic building blocks are the so called paintbox partitions that correspond to elements of $\mathcal{S}^{\downarrow}$ as follows. Let $(U_1, U_2, \dots)$ be a sequence of i.i.d uniform $(0,1)$ random variables and for $\mathbf{s} \in \mathcal{S}^{\downarrow}$, let $\rho_{\mathbf{s}}$ be the law of the partition defined by i and j are in the same block if and only if there exists k such that for $\ell \in \{i, j\}$

$$\sum_{m=1}^{k} s_m \leq U_\ell < \sum_{m=1}^{k+1} s_m.$$

Theorem 2 (see Propositions 3.2 and 3.3 in [3])**.** *Given a measure ν on $\mathcal{S}^{\downarrow}$ such that $\nu(\{\mathbf{1}\}) = 0$ and $\int_{\mathcal{S}^{\downarrow}} (1 - s_1)\nu(d\mathbf{s}) < \infty$, define a measure ρ_ν on $\mathcal{P}_\infty$ by*

$$\rho_\nu(\cdot) = \int_{\mathbf{s}\in\mathcal{S}^{\downarrow}} \rho_{\mathbf{s}}(\cdot)\nu(d\mathbf{s}).$$

Then there is a unique homogeneous fragmentation such that $q_\pi = \rho_\nu(\mathcal{P}_{\infty,\pi})$ for all $\pi \in \cup_n(\mathcal{P}_n \setminus \{\mathbf{1}_{[n]}\})$.

If $\Pi^0(t)$ is a homogenous fragmentation corresponding to a measure ν as in the above theorem, we will call ν the splitting measure $\Pi^0(t)$. If $\alpha < 0$, we can construct an α-self-similar fragmentation from Π^0 by a time change. Let $\pi(i,t)$ be the block of Π^0 that contains i at time t and define

$$T_i(t) = \inf\left\{u \geq 0 : \int_0^u |\pi(i,r)|^{-\alpha} dr > t\right\}.$$

For $t \geq 0$, let $\Pi(t)$ be the partition such that i, j are in the same block of $\Pi(t)$ if and only if they are in the same block of $\Pi^0(T_i(t))$. Then $(\Pi(t), t \geq 0)$ is a self-similar fragmentation. Moreover, for we call (α, ν) a fragmentation pair and and we call $(\Pi(t), t \geq 0)$ a fragmentation with characteristics (α, ν).

We will need trees associated to fragmentations with characteristics (α, ν), where $\alpha < 0$ and $\nu(\sum_i s_i < 1) = 0$; henceforth, we let Π be such a self-similar fragmentation. The tree associated a fragmentation processes Π is a continuum random tree that keeps track of when blocks split apart and the sizes of the resulting blocks. For a continuum tree (T, μ) and $t \geq 0$, let $T_1(t), T_2(t), \dots$ be the tree components of $\{v \in T : \mathrm{ht}(v) > t\}$, ranked in decreasing order of μ-mass. We call (T, μ) self-similar with index $\alpha < 0$ if for every $t \geq 0$, conditionally on $(\mu(T_i(t)), i \geq 1)$, $(T_i(t), i \geq 1)$, equipped with the restriction of μ normalized to be a probability measure, has the same law as $(\mu(T_i(t))^{-\alpha} T^{(i)}, i \geq 1)$ where the $T^{(i)}$'s are independent copies of T.

The following summarizes the parts of Theorem 1 and Lemma 5 in [4] that we will need.

Theorem 3. *Let Π be a (α, ν)-self-similar fragmentation with $\alpha < 0$ and ν as above and let $F := |\Pi|^{\downarrow}$ be its ranked sequence of asymptotic frequencies. There exists an α-self-similar CRT $(T_{-\alpha,\nu}, \mu_{-\alpha,\nu})$ such that, writing $F'(t)$ for the decreasing sequence of masses of the connected components of $\{v \in T_{-\alpha,\nu} : ht(v) > t\}$, the process $(F'(t), t \geq 0)$ has the same law as F. Furthermore, $T_{-\alpha,\nu}$ is a.s. compact.*

The choice of where to put negative signs in the notation in the above theorem is to conform with the notation of [5].

Definition 4. *The Brownian CRT is the $-1/2$-self-similar random tree with dislocation measure ν_2 given by*

$$\int_{\mathcal{S}^{\downarrow}} \nu_2(d\mathbf{s}) f(\mathbf{s}) = \int_{1/2}^{1} \sqrt{\frac{2}{\pi s_1^3 (1 - s_1)^3}} ds_1 f(s_1, 1 - s_1, 0, 0, \dots).$$

2.5. **Convergence of Markov branching trees.** We first recall some of the main results of [5]. Let $A \subseteq \mathbb{Z}^+$ and let (q_n) be a compatible sequence of probability measures satisfying the conditions of Section 2.2. Define $\bar{q}_n$ to be the push forward of q_n to $\mathcal{S}^{\downarrow}$ by $\lambda \mapsto \lambda / \sum_i \lambda_i$.

Theorem 4 (Theorems 1 and 2 in [5])**.** *Suppose that $A = \{0\}$ or $A = \mathbb{Z}^+$. Further suppose that there is a fragmentation pair $(-\gamma, \nu)$ with $0 < \gamma < 1$ and a function $\ell : (0, \infty) \to (0, \infty)$, slowly varying at ∞ (or $\gamma = 1$ and $\ell(n) \to 0$) such that, in the sense of weak convergence of finite measures on $\mathcal{S}^{\downarrow}$, we have*

$$n^{\gamma} \ell(n)(1 - s_1) \bar{q}_n(d\mathbf{s}) \to (1 - s_1) \nu(d\mathbf{s}).$$

Let T_n have law $\mathbf{P}_n^q$ and view T_n as a random element of $\mathcal{M}_w$ with the graph distance and the uniform probability measure $\mu_{\partial_A T_n}$ on $\partial_A T_n = \{v \in T_n : deg\ v \in A\}$. Then we have the convergence in distribution

$$\frac{1}{n^{\gamma} \ell(n)} T_n \to T_{\gamma,\nu},$$

with respect to the rooted Gromov-Hausdorff-Prokhorov topology.

The case where $A = \{0\}$ this is a special case of Theorem 1 in [5] and the case $A = \mathbb{Z}^+$ is Theorem 2 in the same paper. The case $A = \mathbb{Z}^+$ is proved by reduction to the $A = \{0\}$ case. We extend this to the case of general A. Fix, for the moment, a tree t. For a vertex $v \in t$, let t_v be the subtree of t above v. We call the elements of $\{t_v : v \in t\}$ the fringe subtrees of t and say that v is the root of t_v. Furthermore, we partially order the fringe subtrees of t by inclusion.

Theorem 5. *The conclusions of Theorem 4 are valid if the only assumption on $A \subseteq \mathbb{Z}^+$ is that $0 \in A$. If $0 \notin A$, the conclusions of Theorem 4 remain valid if we further assume that $\Delta_n = o(n^\gamma \ell(n))$ in probability, where Δ_n is the maximal diameter of all fringe subtrees of t of T_n such that $\#_A t = 0$.*

Remark 4. *The condition that $\Delta_n = o(n^\gamma \ell(n))$ in probability expresses a relationship between the number of subtrees that have no vertices with out-degree in A and the height of these trees. For example, it is always satisfied if there is a uniform upper bound on the trees in the support of $\mathbf{P}_0^q$. In our application to Galton-Watson trees, this condition is easily checked, but it may be of interest to investigate general conditions on $\mathbf{P}_0^q$ and $(q_{n_k})_{k\geq 1}$ that guarantee it is satisfied.*

We will prove Theorem 5 by reduction to Theorem 4. For a tree t, we define t° to be the tree obtained from t as follows:

(1) Attach a mark to each vertex in t whose out-degree is in A.
(2) For each vertex v, delete the tree t_v (which is the subtree of t above v) if it contains no marked vertices.
(3) For each vertex v, if t_v is a maximal among fringe subtrees containing exactly one marked vertex, replace t_v by a marked leaf.
(4) Attach a leaf to each marked vertex that is not a leaf at the completion of Step (3).
(5) Remove all the marks from the tree.

Suppose that $(q_{n_k})_{k\geq 1}$ satisfies the hypotheses above. For notational convenience we extend q_n to $\cup_k \bar{\mathcal{P}}_k^A$ by setting $q_n(\lambda) = 0$ for $\lambda \notin \bar{\mathcal{P}}_n^A$. Define measures

$$q_n^1(\lambda) = \sum_{k=0}^{\infty} q_n((\lambda, 0_k))\mathbf{1}(p(\lambda) + k \notin A) \quad \text{and} \quad q_n^2(\lambda) = \sum_{k=0}^{\infty} q_n((\lambda, 0_k))\mathbf{1}(p(\lambda) + k \in A).$$

We can then define probability measures q_n° on $\bar{\mathcal{P}}_n^{\{0\}}$ by $q_1^\circ(\emptyset) = 1$ and for $n \geq 2$

$$q_n^\circ(\lambda) = \begin{cases} q_n^1(\lambda) & \text{if } \lambda_{p(\lambda)} \geq 2 \\ q_n^1(\lambda) + q_n^2(\lambda') & \text{if } \lambda = (\lambda', 1) \\ 0 & \text{otherwise.} \end{cases}$$

Lemma 2. *If $n \geq 1$ and T_n is distributed like $\mathbf{P}_n^q$ then T_n° is distributed like $\mathbf{P}_n^{q^\circ}$.*

Proof. We proceed by induction on n. For $n = 1$, it is clear from Step (3) in the construction of T_1° that T_1° is a tree with a single vertex; $\mathbf{P}_n^{q^\circ}$ is concentrated on this tree since $q_1^\circ(\emptyset) = 1$. For $n \geq 2$, we condition on the root partition. First suppose that the root partition is not of the form $(n, 0_k)$. In this case, in both T_n° and a tree with law $\mathbf{P}_n^{q^\circ}$, given the root partition the subtrees attached to the root are independent and have the appropriate distributions by the

inductive hypothesis. When the root partition is of the form $(n, 0_k)$ note that in both trees the lowest vertex whose out-degree is not 1 is attached to the root by a path of geometric length and the parameters of the geometric variable are the same. Moreover, in both T_n° and a tree distributed like $\mathbf{P}_n^{q^\circ}$, the fringe subtree above the lowest vertex whose out-degree is not 1 is distributed like the tree conditioned on the root vertex not being of the form $(n, 0_k)$. Thus, by induction, we need only check that the laws of the partitions at the root agree. This, however, is immediate from the construction of q_n°. □

Lemma 3. *If*

$$n^\gamma \ell(n)(1-s_1)\bar{q}_n(d\mathbf{s}) \to (1-s_1)\nu(d\mathbf{s}),$$

then

$$n^\gamma \ell(n)(1-s_1)\bar{q}_n^\circ(d\mathbf{s}) \to (1-s_1)\nu(d\mathbf{s}).$$

Proof. Let $f : \mathcal{S}^\downarrow \to \mathbb{R}$ be Lipschitz continuous (with respect to the uniform norm) with both the uniform norm and Lipschitz constant bounded by K. For $\lambda = (\lambda_1, \dots, \lambda_r, 0_k)$ with $\lambda_r \geq 1$, define $r(\lambda) = r$ and $\iota(\lambda) = (\lambda_1, \dots, \lambda_r, 1, 0_k)$. Observe that for $\lambda \in \bar{\mathcal{P}}_n$,

$$\left| f\left(\frac{\iota(\lambda)}{n+1}\right) - f\left(\frac{\lambda}{n}\right) \right| \leq K \sum_{i=1}^{r(\lambda)} \frac{\lambda_i}{n(n+1)} + \frac{K}{n+1} = \frac{2K}{n+1}.$$

Letting $g(\mathbf{s}) = (1-s_1)f(\mathbf{s})$, we have

$$\begin{aligned} |\bar{q}_n^\circ(g) - \bar{q}_n(g)| &\leq \sum_{\lambda \in \bar{\mathcal{P}}_n^A \cap \bar{\mathcal{P}}_{n-1}} q_n(\lambda) \left| \left(1 - \frac{\lambda_1}{n}\right) f\left(\frac{\iota(\lambda)}{n}\right) - \left(1 - \frac{\lambda_1}{n-1}\right) f\left(\frac{\lambda}{n-1}\right) \right| \\ &\leq \sum_{\lambda \in \bar{\mathcal{P}}_n^A \cap \bar{\mathcal{P}}_{n-1}} q_n(\lambda) \left(\frac{K\lambda_1}{n(n-1)} + \frac{2K}{n} \right) \\ &\leq \frac{3K}{n}. \end{aligned}$$

Multiplying by $n^\gamma \ell(n)$, we see that this upper bound goes to 0 and the result follows. □

Proof of Theorem 5. Let T_n^1 be the tree obtained from T_n by removing all of the maximal fringe subtrees containing no vertices with out-degree in A. By hypothesis, we have

$$d_{\mathrm{GHP}}\left(\frac{1}{n^\gamma \ell(n)} T_n^1, \frac{1}{n^\gamma \ell(n)} T_n \right) \leq \frac{1}{n^\gamma \ell(n)} \Delta_n \to 0$$

in probability. Note that leaves in T_n^1 correspond to the unique vertices with out-degree in A in the fringe subtrees of T_n that are maximal among fringe subtrees with exactly one vertex with out-degree in A. By construction, these vertices are attached to the rest of T_n by independent spines of geometric length with parameter $1 - \sum_{j \geq 0} q_1((1, 0_j))$. Let T_n^2 be the tree obtained from T_n^1 by replacing these spines by a single vertex and moving the mass from the tip of this spine to the single vertex. Since the maximum length of the spines altered in the creation of T_n^2 is $O(\log(n))$ in probability, we have

$$d_{\mathrm{GHP}}\left(\frac{1}{n^\gamma \ell(n)} T_n^1, \frac{1}{n^\gamma \ell(n)} T_n^2 \right) \to 0$$

in probability. Here we have applied Lemma 1 to obtain the Prokhorov part of this convergence with f being the map that takes a vertex at the top of a spine to the single vertex replacing the spine. Finally, observe that T_n° differs from T_n^2 only by the attachment of some additional leaves with a corresponding adjustment of the mass measure. Thus

$$d_{\mathrm{GHP}}\left(\frac{1}{n^\gamma \ell(n)} T_n^2, \frac{1}{n^\gamma \ell(n)} T_n^\circ\right) \to 0$$

in probability. Since $(n^\gamma \ell(n))^{-1} T_n^\circ \to T_{\gamma,\nu}$ by Lemma 3 and Theorem 4, $(n^\gamma \ell(n))^{-1} T_n \to T_{\gamma,\nu}$ as well. □

3. Galton-Watson trees

Let $\xi = (\xi_i)_{i\geq 0}$ be a probability distribution with mean less than or equal to 1, and assume that $\xi_1 < 1$. A Galton-Watson tree with offspring distribution ξ is a random element T of $\mathcal{T}$ with law

$$\mathrm{GW}_\xi(t) = \mathbb{P}(T = t) = \prod_{v\in t} \xi_{\deg(v)}.$$

The fact that ξ has mean less than or equal to 1 implies that the right hand side defines an honest probability distribution of $\mathcal{T}$.

3.1. **Otter-Dwass type formulae.** In this section we develop a transformation of rooted ordered trees that takes Galton-Watson trees to Galton-Watson trees. This transformation is motivated by the observation that the number of leaves in a Galton-Watson tree is distributed like the progeny of a Galton-Watson tree with a related offspring distribution. This simple observation was first made in [10]. Let $\xi = (\xi_i)_{i\geq 0}$ be a probability distribution with mean less than or equal to 1, and assume that $\xi_1 < 1$. Let T be a Galton-Watson tree with offspring distribution ξ and let

$$C(z) = \sum_{i=1}^{\infty} \mathbb{P}(\#_{\{0\}} T = i) z^i$$

be the probability generating function of the number of leaves of T. Furthermore, let

$$\phi(z) = \sum_{i=0}^{\infty} \xi_{i+1} z^i.$$

Decomposing by the root degree, we see that $C(z)$ satisfies the functional equation

$$C(z) = \xi_0 z + C(z)\phi(C(z)).$$

Solving for $C(z)$ yields

$$C(z) = z\left(\frac{\xi_0}{1-\phi(C(z))}\right). \tag{1}$$

Define

$$\theta(z) = \frac{\xi_0}{1-\phi(z)}. \tag{2}$$

Observe that θ has nonnegative coefficients, $[z^0]\theta(z) = \xi_0/(1-\xi_1)$ and $\theta(1) = 1$. Thus the coefficients of θ are a probability distribution, call it $\zeta = (\zeta_i)_{i\geq 0}$.

Proposition 2. *Let T be a Galton-Watson tree with offspring distribution ξ and let T' be a Galton-Watson tree with offspring distribution ζ where ξ and ζ are related as above. Then for all $k \geq 1$, $\mathbb{P}(\#_{\{0\}}T = k) = \mathbb{P}(\#_{\mathbb{Z}^+}T' = k)$. Also, T' is critical (subcritical) if and only if T is critical (subcritical).*

Proof. Let $\tilde{C}(z)$ be the probability generating function of $\#_{\mathbb{Z}^+}T'$. A similar computation as the one above for $C(z)$ shows that $\tilde{C}(z)$ satisfies the functional equation $\tilde{C}(z) = z\theta(\tilde{C}(z))$, which is the same functional equation we showed $C(z)$ solves in Equation 1. The Lagrange inversion formula thus implies that $C(z)$ and $\tilde{C}(z)$ have the same coefficients, i.e. that $\mathbb{P}(\#_{\{0\}}T = k) = \mathbb{P}(\#_{\mathbb{Z}^+}T' = k)$ for all k. From Equation (2), we see that $\theta'(z) = \xi_0\phi'(z)(1-\phi(z))^2$. Let α be the mean of ξ. A short computation shows that that $\phi'(1) = \alpha - 1 + \xi_0$. Since $\phi(1) = 1 - \xi_0$, we see that the mean of ζ is

$$\theta'(1) = \frac{\xi_0(\alpha - 1 + \xi_0)}{\xi_0^2} = \frac{\alpha - 1 + \xi_0}{\xi_0},$$

and the criticality (subcriticality) claim follows. The same approach can also be used to obtain higher moments of ζ. □

Corollary 1. *Let $\mathcal{F}_n$ be an ordered forest of n independent Galton-Watson trees all with offspring distribution ξ. Let ζ be related to ξ as in Proposition 2. Let $(X_i)_{i\geq 1}$ be an i.i.d. sequence of ζ distributed random variables and let $S_k = \sum_{i=1}^k (X_i - 1)$. Let $\#_{\{0\}}\mathcal{F}_n$ denote the number of leaves in $\mathcal{F}_n$. Then for $1 \leq k \leq n$*

$$\mathbb{P}(\#_{\{0\}}\mathcal{F}_k = n) = \frac{k}{n}\mathbb{P}(S_n = -k).$$

Proof. This follows immediately from Proposition 2 and the Otter-Dwass formula (see [11]). □

This relationship between T and T' can also be proved in a more probabilistic fashion. Indeed, by taking such an approach we can get a more general result that includes the results in [10] as a special case. The idea is a lifeline construction. We follow the depth-first walk around the tree and when we encounter a vertex whose degree is in A we label any unlabeled edges on the path from this vertex to the root with the label of this vertex. This labeled path can be considered the lifeline of the vertex. A new tree is constructed by letting the root be first vertex encountered whose degree is in A and attaching vertices whose degree is in A to the earliest vertex whose lifeline touches its own. We will now go through this construction formally.

Suppose that $t \in \mathcal{T}_{A,n}$ for some $n \geq 1$ and label the vertices of t whose out-degrees are in A by the order they appear in the depth first walk of t. We will now color a subset of the edges of t. For each edge $e \in t$, let t_e be the component of $t \setminus e$ that does not contain the root. Color e with the smallest number that is the label of a vertex in t_e, leaving e uncolored if t_e contains no labeled vertices. Note that for any $1 \leq k \leq n$ the subtree spanned by the vertices with labels $\{1, \ldots, k\}$ and the root is colored by $\{1, \ldots, k\}$ and an edge is colored

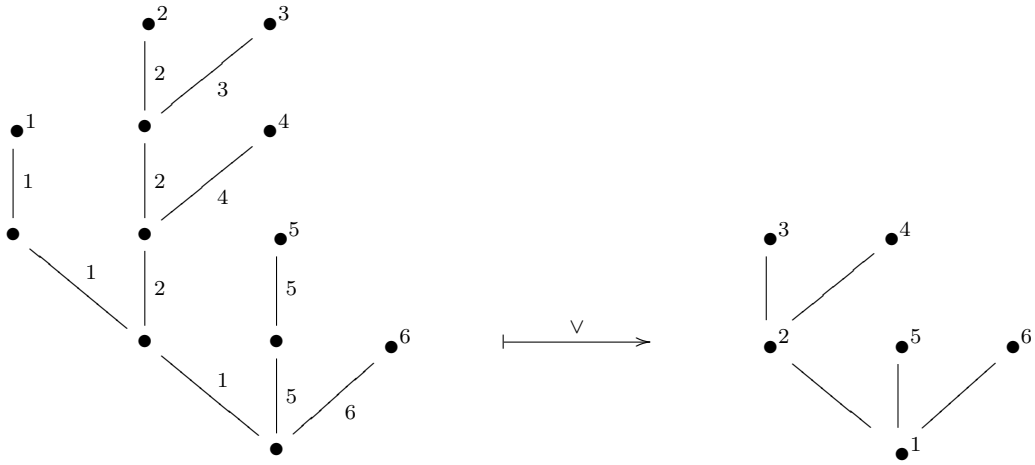

FIGURE 1. A colored tree and its image under $\vee$ when $A = \{0\}$

by an element in $\{1, \ldots, k\}$ if and only if it is in this subtree. Furthermore, the path from the vertex with label k to any edge colored k contains only edges colored k. See Figure 1 for an example of such a coloring when $A = \{0\}$. Call two edges of t adjacent if they share a common vertex.

Lemma 4. *If t is colored as above and $2 \leq j \leq n$, then there is exactly one edge colored j that is adjacent to an edge with a smaller color.*

Proof. The essential property of the labeling is that labels are non-increasing along the path from a labeled vertex to the root. First we show existence. Consider the path from vertex with label j to the root. Let e be the last edge this path that is not contained in the subtree spanned by vertices $\{1, \ldots, j-1\}$. By construction this edge is colored j and is adjacent to an edge colored by an element of $\{1, \ldots, j-1\}$.

To see uniqueness, note that all of labeled edges in the subtree above e have color at least j and all of the edges colored j are in the subtree above e. Thus e is the only edge colored e that is adjacent to an edge with a smaller color. □

With t labeled as above we define a rooted plane tree with n vertices, called the life-line tree and denoted $\check{t}$, as follows. The vertex set of $\check{t}$ is $\{1, 2, \ldots, n\}$, 1 is the root. Furthermore, if $i < j$, i is adjacent to j if i is the smallest number such that there exist adjacent edges e_1, e_2 in t with e_1 colored i and e_2 colored j. Finally, the children of a vertex are ordered by the appearance of the corresponding vertices with out-degree in A in the depth-first search of t. See Figure 1 for an example of this map when $A = \{0\}$.

Lemma 5. *The life-line tree is a connected acyclic graph.*

Proof. Suppose that $\check{t}$ has at least two components. Let j be the smallest vertex not in the same component as 1. By Lemma 4, there exists $1 \leq i < j$ and adjacent edges e_1, e_2 in t labeled i and j respectively. Thus i is adjacent to j, a contradiction.

Suppose that $\check{t}$ contains a cycle. Let j be the largest vertex in this cycle. Then j is adjacent to two smaller vertices, contradicting our definition of $\check{t}$. □

Using this construction we can prove the following theorem.

Theorem 6. *Fix $A \subseteq \mathbb{Z}_+$ and let T' be a critical or subcritical Galton-Watson tree with non-degenerate offspring distribution ξ and let T be T' conditioned to have at least one vertex with out-degree in A. Let $(X_i)_{i\geq 1}$ be i.i.d random variables with distribution ξ. Let $N(X) = \inf\{k : X_k \in A\}$ and $\tau_{-1}(X) = \inf\left\{k : \sum_{i=1}^{k}(X_i - 1) = -1\right\}$. Let $\tilde{X}$ be distributed like*

$$\left(1 + \sum_{i=1}^{N(X)}(X_i - 1)\right) \quad \textit{given} \quad (N(X) \leq \tau_{-1}(X)).$$

Finally, introduce the random variable Y such that, conditionally given $\tilde{X} = k$, Y is binomially distributed with parameters k and $p = \mathbb{P}(N(X) \leq \tau_{-1}(X))$. Then $\check{T}$ is distributed like a Galton-Watson tree whose offspring distribution is the law of Y. Moreover, if T' is critical then so is $\check{T}$ (i.e. $\mathbb{E}Y = 1$) and if additionally $0 < Var(\xi) = \sigma^2 < \infty$ then $Var(Y) = \mathbb{P}(N(X) \leq \tau_{-1}(X))^2\sigma^2/\xi(A)$. We let $\check{\xi}$ be the law of Y.

The claims about the mean and variance of Y in this theorem are very important for our purposes as they are what allow us to derive the constants in Theorem 1. As such, we separate out the computation of these constants into its own lemma.

Lemma 6. *Let be Y as defined in Theorem 6. If ξ has mean 1 then $\mathbb{E}Y = 1$. If additionally $0 < Var(\xi) = \sigma^2 < \infty$ then $Var(Y) = \mathbb{P}(N(X) \leq \tau_{-1}(X))^2\sigma^2/\xi(A)$*

Proof. Let $p = \mathbb{P}(N(X) \leq \tau_{-1}(X))$. By conditioning on the value of $\tilde{X}$, we see that

$$\mathbb{E}Y = \mathbb{E}Y = \sum_{k=1}^{\infty} kp\mathbb{P}(\tilde{X} = k) = p\mathbb{E}\tilde{X}$$

and

$$\mathbb{E}Y^2 = \sum_{k=1}^{\infty}(kp(1-p) + k^2p^2)\mathbb{P}(\tilde{X} = k) = p(1-p)\mathbb{E}\tilde{X} + p^2\mathbb{E}\tilde{X}^2.$$

This reduces the problem to computing the moments of $\tilde{X}$. Assume that ξ has mean 1, so that $\mathbb{E}X_i = 1$. By Wald's first equation we have

$$\begin{aligned} 1 &= \mathbb{E}\left[1 + \sum_{i=1}^{N(X)}(X_i - 1)\right] \\ &= 1 + \mathbb{E}\left[\sum_{i=1}^{N(X)}(X_i - 1),\ N \leq \tau_{-1}(X)\right] + \mathbb{E}\left[\sum_{i=1}^{N(X)}(X_i - 1),\ N(X) > \tau_{-1}(X)\right]. \end{aligned}$$

Using the strong Markov property of $(X_i)_{i\geq 1}$ at the stopping time τ_{-1} and Wald's first equation again we see that

$$\begin{aligned}\mathbb{E}\left[\sum_{i=1}^{N(X)}(X_i-1),\ N(X)>\tau_{-1}(X)\right] &= \mathbb{P}(N(X)>\tau_{-1}(X))\left(-1+\mathbb{E}\left[\sum_{i=1}^{N(X)}(X_i-1)\right]\right).\\ &= -\mathbb{P}(N(X)>\tau_{-1}(X))\end{aligned}$$

Therefore $\mathbb{E}\tilde{X}=\mathbb{P}(N(X)\leq\tau_{-1}(X))^{-1}=p^{-1}$ and consequently $\mathbb{E}Y=p\mathbb{E}\tilde{X}=1$.

Now assume additionally that $0<\mathrm{Var}(\xi)<\infty$. Wald's second equation shows that

$$\begin{aligned}\frac{\sigma^2}{\xi(A)} &= \mathbb{E}\left(\left[\sum_{k=1}^{N(X)}(X_k-1)\right]^2\right)\\ &= \mathbb{E}\left(\left[\sum_{k=1}^{N(X)}(X_k-1)\right]^2,\ N(X)\leq\tau_{-1}(X)\right)+\mathbb{E}\left(\left[\sum_{k=1}^{N(X)}(X_k-1)\right]^2,\ N(X)>\tau_{-1}(X)\right).\end{aligned}$$

Using the same Markov property trick as in the previous computation, we see that

$$\mathbb{E}\left(\left[\sum_{k=1}^{N(X)}(X_k-1)\right]^2,\ N(X)>\tau_{-1}(X)\right)=\mathbb{P}(N(X)>\tau_{-1}(X))\left(1+\frac{\sigma^2}{\xi(A)}\right).$$

Consequently

$$\begin{aligned}\mathbb{E}\tilde{X}^2 &= \frac{1}{\mathbb{P}(N(X)\leq\tau_{-1}(X))}\mathbb{E}\left(\left[1+\sum_{k=1}^{N(X)}(X_k-1)\right]^2,\ N(X)\leq\tau_{-1}(X)\right)\\ &= 1+2p^{-1}\mathbb{E}\left[\sum_{i=1}^{N(X)}(X_i-1),\ N(X)\leq\tau_{-1}(X)\right]+p^{-1}\mathbb{E}\left(\left[\sum_{k=1}^{N(X)}(X_k-1)\right]^2,\ N(X)\leq\tau_{-1}(X)\right)\\ &= 1+2p^{-1}(1-p)+p^{-1}\left[\frac{\sigma^2}{\xi(A)}-(1-p)\left(1+\frac{\sigma^2}{\xi(A)}\right)\right]\\ &= p^{-1}+\frac{\sigma^2}{\xi(A)}.\end{aligned}$$

Thus

$$\mathrm{Var}(Y)=\mathbb{E}Y^2-1=(1-p)+p^2\left(p^{-1}+\frac{\sigma^2}{\xi(A)}\right)-1=p^2\frac{\sigma^2}{\xi(A)},$$

as desired. □

Our proof of Theorem 6 will proceed by induction. The first step is to analyze the degree of the root of $\check{T}$. To do this, we need some notation. For a vertex $v\in t$ define the spine of v to be the vertices in t that are children of v or children one of v's ancestors. The right spine of v, denoted $\mathrm{rspine}(v)$, is the subset of the spine of v consisting of vertices that appear after v in the depth-first walk of t (see Figure 2).

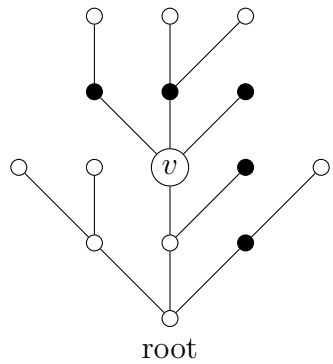

FIGURE 2. The black nodes are the right spine of the vertex v.

Lemma 7. *The distribution of the degree of the root of $\check{T}$ is $\check{\xi}$.*

Proof. Let $(v_1, \ldots, v_n)$ be the vertices of T in depth first order and let v be the first vertex along the depth first path of T such that $\deg(v) \in A$. The degree of the root of $\check{T}$ is the number of vertices in $\mathrm{rspine}(v)$ that are the root of a fringe subtree with at least one vertex whose out-degree is in A. Furthermore, it is clear from the definition of Galton-Watson trees that the fringe subtrees of T rooted at the vertices in $\mathrm{rspine}(v)$ are conditionally independent given the number of vertices in $\mathrm{rspine}(v)$. Hence, conditionally given $\#\mathrm{rspine}(v)$, the degree of the root of $\check{T}$ has distribution $\mathrm{Bin}(\#\mathrm{rspine}(v), p)$, where $p = \mathbb{P}(N(X) \leq \tau_{-1}(X))$ as above. An easy combinatorial argument shows that if k is the index of v in the depth first order of T, then

$$\#\mathrm{rspine}(v) = 1 + \sum_{i=1}^{k} (\deg(v_i) - 1).$$

It is now immediate that the degree of the root of $\check{T}$ has distribution $\check{\xi}$. □

Proof of Theorem 6. It follows from Lemma 7 that the probability that $\check{T}$ has a single vertex equals $\mathbb{P}(Y = 0)$. We will proceed by induction on the number of vertices of t to show that for all t

$$\mathbb{P}(\check{T} = t) = \prod_{v \in t} \mathbb{P}(Y = \deg(v)).$$

Suppose the claim is true for all trees s with fewer than n vertices and let t be a tree with n vertices. Let r be the degree of the root of t and let $(t_1, \ldots, t_r)$ be the subtrees of t attached to the root.

Let us consider the set $S = \{s \in \mathcal{T} : \check{s} = t\}$, so we are trying to find $\mathbb{P}(T \in S)$. Fix $s \in S$ and let v_s be the first vertex along the depth first walk of s whose out-degree is in A. Let $(w_1, \ldots, w_d)$ be the elements of $\mathrm{rspine}(v)$ listed by order of appearance on the depth first walk of s. For $1 \leq i \leq d$, let s_i be the subtree of s rooted at w_i. Since $\check{s} = t$, there is a set $J \subseteq \{1, \ldots, d\}$ of size r such that s_i contains a vertex with out-degree in A if and only if $i \in J$. Let ρ be the unique increasing bijection from $\{1, \ldots, r\}$ to J. Then, again since $\check{s} = t$, it follows from the recursive nature of the construction of $\check{s}$ that $t_i = \check{s}_{\rho(i)}$. It follows

from these considerations that

$$\begin{aligned}\mathbb{P}(T \in S) &= \sum_{d=r}^{\infty} \sum_{s \in S : \#\mathrm{rspine}(v_s)=d} \mathbb{P}(T = s) \\ &= \sum_{d=r}^{\infty} \mathbb{P}(\mathrm{rspine}(v_T) = d) \binom{d}{r} (1-p)^{d-r} \prod_{j=1}^{r} \mathbb{P}(\check{T}' = t_j),\end{aligned}$$

where we have continued to use $p = \mathbb{P}(N(X) \leq \tau_{-1}(X))$. Since t_i has fewer vertices than t, it follows from the induction hypothesis that

$$\mathbb{P}(\check{T}' = t_i) = p\mathbb{P}(\check{T} = t_i) = p \prod_{v \in t_i} \mathbb{P}(Y = \deg(v)).$$

Thus

$$\begin{aligned}\mathbb{P}(\check{T} = t) &= \left[\prod_{j=1}^{r} \prod_{v \in t_j} \mathbb{P}(Y = \deg(v))\right] \sum_{d=r}^{\infty} \mathbb{P}(\mathrm{rspine}(v_T) = d) \binom{d}{r} (1-p)^{d-r} p^r \\ &= \left[\prod_{j=1}^{r} \prod_{v \in t_j} \mathbb{P}(Y = \deg(v))\right] \mathbb{P}(Y = r) \\ &= \prod_{v \in t} \mathbb{P}(Y = \deg(v)),\end{aligned}$$

which shows that $\check{T}$ is a Galton-Watson tree with the appropriate offspring distribution. □

Note that T has n vertices with out-degree in A if and only if $\check{T}$ has n vertices. This allows us to deduce the following corollary.

Corollary 2. *Let $\mathcal{F}_n$ be an ordered forest of n independent Galton-Watson trees with critical or subcritical offspring distribution ξ. Let $(Y_i)_{i \geq 0}$ be i.i.d with common distribution $\check{\xi}$ (see Theorem 6 for the definition of $\check{\xi}$) and let $S_i = (Y_1 - 1) + \cdots + (Y_i - 1)$. Also, let $\mathcal{F}_n^Y$ an ordered forest of n independent Galton-Watson trees with offspring distribution $\check{\xi}$ and let Z_n be independent with distribution Bin$(n, \mathbb{P}(N \leq \tau_{-1}))$. Then*

$$\begin{aligned}\mathbb{P}(\#_A \mathcal{F}_n = k) &= \sum_{j=0}^{n} \frac{j}{k} \binom{n}{j} \mathbb{P}(N \leq \tau_{-1})^j \mathbb{P}(N > \tau_{-1})^{n-j} \mathbb{P}(S_k = -j) \\ &= \mathbb{P}(\#\mathcal{F}_{Z_n}^Y = k)\end{aligned}.$$

We remark that when $n = 1$ a similar formula was obtained in [1] by different methods.

Proof. The sum is obtained by decomposing $\mathbb{P}(\#_A \mathcal{F}_n = k)$ according to which j trees in the forest have a positive number of vertices with out-degree in A. Conditionally given which trees these are, Otter-Dwass formula to obtain the formula for the number of vertices in a forest of j independent Galton-Watson trees distributed like $\check{T}$. One then sums over the possible sets of j trees. □

3.2. **The partition at the root.** Let $\xi = (\xi_i)_{i\geq 0}$ be a probability distribution with mean 1 and variance $0 < \sigma_1^2 < \infty$. Let T be a Galton-Watson tree with offspring distribution ξ (denote the law of T by GW_ξ). Let $A \subseteq \mathbb{Z}^+$ and construct $\check{T}$ as above. By Theorem 6, $\check{T}$ is a Galton-Watson tree with offspring distribution $\check{\xi}$. Again by Theorem 6, $\check{\xi}$ has mean 1 and variance $\sigma^2 = \gamma^2\sigma_1^2/\xi(A)$, where $\gamma := \mathbb{P}(\#_A T \geq 1)$ (see Lemma 6 for computations of the moments of $\check{\xi}$). Assume that for sufficiently large n, $\mathbb{P}(\#_A T = n) > 0$. Let T_n^A be T conditioned to have exactly n vertices with out-degree in A (whenever this conditioning makes sense).

For a t be rooted unordered tree with exactly n vertices with out-degree in A, let $\Pi^A(t)$ be the partition of n or $n-1$ (depending on whether or not the degree of the root of t is in A) defined by the number of vertices with out-degree in A in the subtrees of t attached to the root. We also adopt the convention $\Pi^A(\bullet) = \emptyset$, that is, the partition at the root of the tree with one vertex is the emptyset.

Lemma 8. *(i) Considered as an unordered tree, the law of T_n^A is equal to $\mathbf{P}_n^q$ where q_0 is the law of T_0^A if $0 \notin A$ and for $n \geq 1$ such that T_n^A is defined, and $\lambda = (\lambda_1, \dots, \lambda_p) \in \bar{\mathcal{P}}_n^A$, we have*

$$q_n(\lambda) = \mathbb{P}(\Pi^A(T_n^A) = \lambda) = \frac{p!}{\prod_{j\geq 1} m_j(\lambda)!}\xi(p)\frac{\prod_{i=1}^p \mathbb{P}(\#_A T = \lambda_i)}{\mathbb{P}(\#_A T = n)}.$$

(ii) Let $X_1, X_2, \dots$ be i.i.d. distributed like $\#_A T$ and $\tau_k = X_1 + \cdots + X_k$. We have

$$\mathbb{P}(p(\Pi^A(T_n^A)) = p) = \xi(p)\frac{\mathbb{P}(\tau_p = n - \mathbf{1}(p \in A))}{\mathbb{P}(\tau_1 = n)},$$

and $\mathbb{P}(\Pi^A(T_n^A) \in \cdot \mid \{p(\Pi^A(T_n^A)) = p\})$ is the law of a non-increasing rearrangement of $(X_1, \dots, X_p)$ conditionally on $X_1 + \cdots + X_p = n - \mathbf{1}(p \in A)$.

Proof. (i) Letting $c_\emptyset(T_n^A)$ be the root degree of T_n^A and $a_1, \dots, a_p \in \mathbb{N}$ with sum $n - \mathbf{1}(p \in A)$ we have

$$\mathbb{P}(c_\emptyset(T_n^A) = p, \#_A[(T_n^A)_i] = a_i, 1 \leq i \leq p) = \xi(p)\frac{\prod_{i=1}^p \mathbb{P}(\#_A T = a_i)}{\mathbb{P}(\#_A T = n)}. \tag{3}$$

Part (i) now follows by considering the number of sequences $(a_1, \dots, a_p)$ with the same decreasing rearrangement.

(ii) This follows from Equation (3). □

To simplify notation, let q_n be the law of $\Pi^A(T_n^A)$ and let $\mathbf{1}_p = \mathbf{1}(p \in A)$. Let $(S_r, r \geq 0)$ be a random walk with step distribution $(\check{\xi}_{i+1}, i \geq -1)$. Furthermore, since $\gamma = \mathbb{P}(\#_A T \geq 1)$. By Corollary 2, we have

$$q_n(p(\lambda) = p) = \frac{n\xi(p)}{(n - \mathbf{1}_p)\gamma\mathbb{P}(S_n = -1)}\sum_{j=0}^p \binom{p}{j}\gamma^j(1-\gamma)^{p-j} j\mathbb{P}(S_{n-\mathbf{1}_p} = -j) \tag{4}$$

where $\hat{\xi}(p) = p\xi(p)$ is the size-biased distribution of ξ.

Define $\bar{q}_n$ to be the pushfoward of q_n onto $\mathcal{S}^\downarrow$ by the map $\lambda \mapsto \lambda/\sum_i \lambda_i$.

For a sequence $(x_1, x_2, \dots)$ of non-negative numbers such that $\sum_i x_i < \infty$, let i^* be a random variable with

$$\mathbb{P}(i^* = i) = \frac{x_i}{\sum_{j\geq 1} x_j}.$$

The random variable $x_1^* = x_{i^*}$ is called a size-biased pick from $(x_1, x_2, \dots)$. Given i^*, we remove the i^*'th entry from $(x_1, x_2, \dots)$ and repeat the process. This yields a random re-ordering $(x_1^*, x_2^*, \dots)$ of $(x_1, x_2, \dots)$ called the size-biased order (if ever no positive terms remain, the rest of the size-biased elements are 0). Similarly for a random sequence $(X_1, X_2, \dots)$ we define the size-biased ordering by first conditioning on the value of the sequence. For any non-negative measure μ on $\mathcal{S}^{\downarrow}$, define the size-biased distribution μ^* of μ by

$$\mu^*(f) = \int_{\mathcal{S}^{\downarrow}} \mu(d\mathbf{s})\mathbb{E}[f(\mathbf{s}^*)],$$

where $\mathbf{s}^*$ is the size-biased reordering of $\mathbf{s}$.

Define the measure ν_2 on $\mathcal{S}^{\downarrow}$ by

$$\int_{\mathcal{S}^{\downarrow}} \nu_2(d\mathbf{s})f(\mathbf{s}) = \int_{1/2}^{1} \sqrt{\frac{2}{\pi s_1^3(1-s_1)^3}}\, ds_1 f(s_1, 1-s_1, 0, 0, \dots).$$

Theorem 7. *With the notation above,*

$$\lim_{n\to\infty} n^{1/2}(1-s_1)\bar{q}_n(d\mathbf{s}) = \frac{\sigma_1\sqrt{\xi(A)}}{2}(1-s_1)\nu_2(d\mathbf{s}),$$

where the limit is taken in the sense of weak convergence of finite measures.

Proof. We follow the reductions in Section 5.1 of [5]. By Lemma 16 in [5] (which is a easy variation of Proposition 2.3 in [3]) it is sufficient to show that

$$\lim_{n\to\infty} n^{1/2}((1-s_1)\bar{q}_n(d\mathbf{s}))^* = \frac{\sigma_1\sqrt{\xi(A)}}{2}((1-s_1)\nu_2(d\mathbf{s}))^*.$$

Note that for any finite non-negative measure μ on $\mathcal{S}^{\downarrow}$ and non-negative continuous function $f : \mathcal{S}_1 \to \mathbb{R}$ we have

$$((1-s_1)\mu(d\mathbf{s}))^*(f) = \int_{\mathcal{S}_1} \mu^*(d\mathbf{x})(1-\max\mathbf{x})f(\mathbf{x}).$$

Consequently the theorem follows from the following Proposition. □

Proposition 3. *Let $f : \mathcal{S}_1 \to \mathbb{R}$ be continuous and let $g(\mathbf{x}) = (1-\max\mathbf{x})f(\mathbf{x})$. Then*

$$\sqrt{n}\bar{q}_n^*(g) \to \frac{\sigma_1\sqrt{\xi(A)}}{\sqrt{2\pi}} \int_0^1 \frac{dx}{x^{1/2}(1-x)^{3/2}} g(x, 1-x, 0, \dots).$$

First note that, by linearity, we may assume that $f \geq 0$ and $||f||_\infty \leq 1$. We begin the proof of this Proposition with several Lemmas regarding the concentration of mass of $\bar{q}_n^*$.

Lemma 9. *For every $\epsilon > 0$, $\sqrt{n}q_n(p(\lambda) > \epsilon\sqrt{n}) \to 0$ as $n \to \infty$.*

Lemma 10. *For g as in Proposition 3 we have*

$$\lim_{\eta\downarrow 0}\limsup_{n\to\infty}\sqrt{n}\bar{q}_n^*(|g|\mathbf{1}_{\{x_1>1-\eta\}})=0 \quad \textit{and} \quad \lim_{n\to\infty}\sqrt{n}\bar{q}_n^*(\mathbf{1}_{\{x_1<n^{-7/8}\}})=0.$$

Lemma 11. *For every $\eta>0$ we have*

$$\lim_{n\to\infty}\sqrt{n}\bar{q}_n^*(\mathbf{1}_{\{x_1+x_2<1-\eta\}})=0.$$

Lemma 12. *There exists a function $\beta_\eta=o(\eta)$ as $\eta\downarrow 0$ such that*

$$\begin{aligned}\lim_{\eta\downarrow 0}\liminf_{n\to\infty}\sqrt{n}\bar{q}_n^*(g\mathbf{1}_{\{x_1<1-\eta,x_1+x_2>1-\beta_\eta\}}) &= \lim_{\eta\downarrow 0}\limsup_{n\to\infty}\sqrt{n}\bar{q}_n^*(g\mathbf{1}_{\{x_1<1-\eta,x_1+x_2>1-\beta_\eta\}})\\ &= \frac{\sigma_1\sqrt{\xi(A)}}{\sqrt{2\pi}}\int_0^1\frac{g((x,1-x,0,\dots))}{x^{1/2}(1-x)^{3/2}}dx.\end{aligned}$$

These lemma's are generalizations of the lemmas in Section 5.1 in [5] to our current setting and the proofs are essentially the same. We refer the reader to [5] for the proofs of Lemmas 9, 10, and 11. We include the proof of Lemma 12 because it makes clear how the factor of $\sqrt{\xi(A)}$ appears in the scaling limit.

Proof. Fix $\eta>0$ and suppose that $\eta'\in(0,\eta)$. Using Lemmas 9 and 10 we decompose according to the events $\{p(\lambda)>\epsilon\sqrt{n}\}$ and $\{\mathbf{x}:x_1\le n^{-7/8}\}$ to get

$$\begin{aligned}(5)\quad \sqrt{n}\bar{q}_n^*(g\mathbf{1}_{\{x_1<1-\eta,x_1+x_2>1-\eta'\}}) &= o(1)+\sqrt{n}\sum_{1\le p\le\epsilon n^{1/2}}q_n(p(\lambda)=p)\\ \times\sum_{\substack{n^{1/8}\le m_1\le(1-\eta)(n-\mathbf{1}_p)\\(1-\eta')(n-\mathbf{1}_p)\le m_1+m_2\le n-\mathbf{1}_p}}&\mathbb{E}[g((m_1,m_2,X_3^*,\dots,X_p^*,0,\dots)/(n-\mathbf{1}_p))|\tau_p=n-\mathbf{1}_p,X_1^*=m_1,X_2^*=m_2]\\ &\times\frac{pm_1}{n-\mathbf{1}_p}\frac{\frac{(p-1)m_2}{n-\mathbf{1}_p}}{1-\frac{m_1}{n-\mathbf{1}_p}}\mathbb{P}(X_1=m_1)\mathbb{P}(X_2=m_2)\frac{\mathbb{P}(\tau_{p-2}=n-m_1-m_2-\mathbf{1}_p)}{\mathbb{P}(\tau_p=n-\mathbf{1}_p)}.\end{aligned}$$

Observe that, if $1\ge x_1+x_2\ge 1-\eta'$ and $x_1\le 1-\eta$, then $x_2/(1-x_1)\ge 1-\eta'/\eta$ and $(1-x_1)/x_2\ge 1$.

Using the local limit theorem and formula 4 we observe that n and small ϵ, we have

$$1-\eta\le\frac{q_n(p(\lambda)=p)}{\hat{\xi}(p)}\le 1+\eta,$$

And similarly for sufficiently large n and small ϵ, we have

$$\frac{1-\eta}{\sigma\sqrt{2\pi}}\le(\gamma p)^{-1}n^{3/2}\mathbb{P}(\tau_p=n)\le\frac{1+\eta}{\sigma\sqrt{2\pi}},$$

for all $1\le p\le\epsilon n^{1/2}$. We note in particular that $\tau_1=X_1=_d X_2$. Furthermore, for $n^{1/8}\le m_1\le(1-\eta)n$ and $m_1+m_2\ge(1-\eta')n$ we have $m_2\ge(\eta-\eta')n$ so that m_1 and m_2 go to infinity as n does. Thus, for large n (how large now depends on η') we have

$$\frac{(1-\eta)^2}{2\pi\sigma^2}\le\gamma^{-2}(m_1m_2)^{3/2}\mathbb{P}(X_1=m_1)\mathbb{P}(X_2=m_2)\le\frac{(1+\eta)^2}{2\pi\sigma^2}.$$

Now, recall that f is uniformly continuous on $\mathcal{S}_1$. Furthermore, on the set $\{\mathbf{x} \in \mathcal{S}_1 : x_1 + x_2 > 3/4\}$ we have $\max \mathbf{x} = x_1 \vee x_2$ and $\mathbf{x} \mapsto \max \mathbf{x}$ is thus uniformly continuous on this set. Therefore for $\eta' < (1/4) \wedge \eta^2$ sufficiently small we have

$$|g((m_1, m_2, m_3, \dots)/n) - g((m_1, n - m_1, 0, \dots)/n)| \leq \eta,$$

for every $(m_1, m_2, \dots)$ with sum n sufficiently large such that $m_1 + m_2 \geq (1 - \eta')n$. Take $\beta_\eta := \eta'$.

Given the symmetry of the bounds we have just established it is easy to see that the proofs for the lim sup and lim inf will be nearly identical, one using the upper bounds and the other the lower. We will only write down the proof for the lim inf. For sufficiently large n we have that, up to addition of an $o(1)$ term, $\sqrt{n}\bar{q}_n^*(g\mathbf{1}_{\{x_1<1-\eta, x_1+x_2>1-\eta'\}})$ is bounded below by

$$\begin{aligned}
&\frac{(1-\eta)^3(1-\eta'/\eta)}{(1+\eta)} \sum_{1\leq p\leq \epsilon n^{1/2}} (p-1)\hat{\xi}(p)\frac{1}{n-\mathbf{1}_p} \\
&\qquad \times \sum_{n^{1/8}\leq m_1\leq (1-\eta)(n-\mathbf{1}_p)} (g((m_1, n-m_1-\mathbf{1}_p, 0, \dots)/(n-\mathbf{1}_p)) - \eta) \\
&\qquad\qquad \times \frac{1}{(m_1/(n-\mathbf{1}_p))^{1/2}} \frac{1}{(1-m_1/(n-\mathbf{1}_p))^{3/2}} \frac{1}{\sigma\sqrt{2\pi}} \\
&\qquad\qquad\qquad \times \sum_{(1-\eta')(n-\mathbf{1}_p)-m_1\leq m_2\leq n-m_1-\mathbf{1}_p} \mathbb{P}(\tau_{p-2} = n - m_1 - m_2 - \mathbf{1}_p).
\end{aligned}$$

Observe that this last sum is equal to $\sum_{m=0}^{\eta'(n-\mathbf{1}_p)} \mathbb{P}(\tau_{p-2} = m)$. By the local limit theorem, this can be made arbitrarily close to 1 independent of $1 \leq p \leq n^{1/2}$. Using the convergence of Riemann sums (again care must be taken since the integral we get is improper), we have

$$\begin{aligned}
&\liminf_{n\to\infty} \sqrt{n}\bar{q}_n^*(g\mathbf{1}_{\{x_1<1-\eta, x_1+x_2>1-\eta'\}}) \\
&\geq \frac{(1-\eta)^3(1-\eta'/\eta)}{1+\eta} \sum_{p=1}^{\infty} \gamma(p-1)\hat{\xi}(p) \int_0^{1-\eta} \frac{dx}{\sigma\sqrt{2\pi}x^{1/2}(1-x)^{3/2}} (g(x, 1-x, 0, \dots) - \eta)
\end{aligned}$$

Letting $\eta \downarrow 0$ coupled with observing that $\sum_{p=1}^{\infty}(p-1)\hat{\xi}(p) = \sigma_1^2$ and recalling that $\sigma^2 = \gamma^2\sigma_1^2/\xi(A)$ completes the proof. □

Proof of Proposition 3. Observe that

$$|\bar{q}_n^*(g) - \bar{q}_n^*(g\mathbf{1}_{\{x_1<1-\eta, x_1+x_2>1-\eta'\}})| \leq \bar{q}_n^*(|g|\mathbf{1}_{\{x_1\geq 1-\eta\}}) + \bar{q}_n^*(|g|\mathbf{1}_{\{x_1+x_2\leq 1-\eta'\}}).$$

Fix $\epsilon > 0$ and apply Lemmas 10 and 12 to find η, η' such that

$$\sqrt{n}\bar{q}_n^*(|g|\mathbf{1}_{\{x_1\geq 1-\eta\}}) < \frac{\epsilon}{2}$$

and

$$\left|\sqrt{n}\bar{q}_n^*(g\mathbf{1}_{\{x_1<1-\eta, x_1+x_2>1-\beta_\eta\}}) - \frac{\sigma_1\sqrt{\xi(A)}}{\sqrt{2\pi}} \int_0^1 \frac{g(x, 1-x, 0, 0, \dots)}{x^{1/2}(1-x)^{3/2}} dx\right| \leq \frac{\epsilon}{2},$$

for large enough n. For this choice of η, η' and large n we have

$$\left|\sqrt{n}\bar{q}_n^*(g) - \frac{\sigma_1\sqrt{\xi(A)}}{\sqrt{2\pi}}\int_0^1 \frac{g(x,1-x,0,0,\dots)}{x^{1/2}(1-x)^{3/2}}dx\right| \leq \epsilon + \sqrt{n}\bar{q}_n^*(|g|\mathbf{1}_{\{x_1+x_2\leq 1-\eta'\}}).$$

By Lemma 11 the upper bound goes to ϵ as $n \to \infty$, and the result follows. □

As an immediate corollary of these results, we also identify the unnormalized limit of $\bar{q}_n$.

Corollary 3. $\bar{q}_n \xrightarrow{d} \delta_{(1,0,0,\dots)}$.

Proof. Taking $f \equiv 1$ in Proposition 3 gives $\bar{q}_n(1-s_1) \to 0$. Since L^1 convergence implies convergence in probability, it follows that for all $0 < \eta < 1$ we have $\bar{q}_n(s_1 \geq \eta) \to 1$. □

Note that, as a consequence of Equation (4), we have $q_n(p(\lambda) = p)) \to \hat{\xi}(p)$. Thus, while the degree of the root vertex may be large, only one of the trees attached to the root will have noticeable size.

3.3. **Convergence of Galton-Watson trees.** We are now prepared to prove Theorem 1, which, after all of our work above, is a rather straightforward. The hardest work that remains is to show that $\Delta_n = o(n^\gamma \ell(n))$ in probability, where Δ_n is the maximal diameter of all fringe subtrees of t of T_n such that $\#_A t = 0$ when $0 \notin A$. This is accomplished by the next two lemmas.

Lemma 13. *Let T_n^A be a Galton-Watson tree conditioned to have exactly n vertices with out-degree in A. Then*

$$\frac{\#_{\mathbb{Z}^+}T_n^A}{n} \to \frac{1}{\xi(A)}$$

in probability.

Proof. Let $X = (X_n)_{n\geq 1}$ be an i.i.d sequence with distribution ξ and let

$$\tau_{-1}(X) = \inf\left\{n : \sum_{k=1}^n (X_k - 1) = -1\right\}.$$

Further define $N_k(X)$ to be index of the k'th occurrence in X of an element in A. Using the bijection between trees and their depth first queues, we see that if T is a Galton-Watson tree with offspring distribution ξ then

$$\mathbb{P}(\#_A T = n) = \mathbb{P}\left(N_n(X) \leq \tau_{-1}(X) < N_{n+1}(X)\right)$$

and that $\#_{\mathbb{Z}^+}T_n^A$ is distributed like $\tau_{-1}(X)$ conditioned on $(N_n(X) \leq \tau_{-1}(X) < N_{n+1}(X))$. By Corollary 2 and the local limit theorem, we see that

$$\mathbb{P}(N_n(X) \leq \tau_{-1}(X) < N_{n+1}(X)) = \mathbb{P}(\#_A T = n) = \frac{\mathbb{P}(N_1(X) \leq \tau_1(X))}{n}\mathbb{P}(S_n = -1) \sim cn^{-3/2}$$

for some $c > 0$, where S_n is as defined in Corollary 2. Furthermore, the large deviations concentration of $N_n(X)$ around $1/\xi(A)$ implies that for every $\epsilon > 0$

$$\mathbb{P}\left(\left|\frac{\tau_{-1}(X)}{n} - \frac{1}{\xi(A)}\right| > \epsilon,\ N_n(X) \leq \tau_{-1}(X) < N_{n+1}(X)\right)$$

decays exponentially as n goes to infinity and the claim follows immediately. □

Lemma 14. *Assume that $0 \notin A$ and let T be a Galton-Watson tree with critical, finite variance offspring distribution ξ and let T_n^A be distributed like T conditioned to have exactly n vertices with out-degree in A. Under these conditions, $\Delta_n = O(\log(n))$ in probability, where Δ_n is the maximal diameter of all fringe subtrees of t of T_n such that $\#_A t = 0$.*

Proof. Let $v_1, \ldots, v_d$ be the roots of the maximal fringe subtrees of T_n^A that contain no vertices with out-degree in A, indexed by order of appearance on the depth first walk of T_n^A and let $T_1, \ldots, T_d$ be the fringe subtrees of T_n^A rooted at $v_1, \ldots, v_d$ respectively. It follows from the elementary properties of Galton-Watson trees that, conditionally given d, $T_1, \ldots, T_d$ are i.i.d such that for $t \in \mathcal{T}_{A,0}^u$

$$\mathbb{P}(T_1 = t) = \frac{1}{\mathbb{P}(\#_A T = 0)} \mathbb{P}(T = t) = \prod_{v \in t} \mathbb{P}(\#_A T = 0)^{\deg(v)-1} \xi_{\deg(v)}.$$

The last equality shows that T_1 is distributed like a Galton-Watson tree with offspring distribution μ given by

$$\mu_i = \mathbb{P}(\#_A T = 0)^{i-1} \xi_i \mathbf{1}(i \notin A).$$

Since ξ has mean 1 and $0 < \mathbb{P}(\#_A T = 0) < 1$ we see that μ has mean strictly less than 1. It follows that

$$\mathbb{P}(\text{height}(T_1) > k) \leq \beta^k,$$

where β is the mean of μ. In particular, the distribution of the height of T_1 decays exponentially. By Lemma 13, there is some C such that $\mathbb{P}(d \leq Cn) \to 1$ as $n \to \infty$. Observe that

$$\mathbb{P}\left(\max_{1 \leq i \leq d} \text{height}(T_i) > x, d \leq Cn\right) \leq \mathbb{P}\left(\max_{1 \leq i \leq Cn} Z_i > x\right),$$

where $(Z_i, i \geq 1)$ are i.i.d distributed like $\text{height}(T_1)$. Since the distribution of $\text{height}(T_1)$ decays exponentially, $\max_{1 \leq i \leq Cn} Z_i$ grows logarithmically and the result follows. □

Proof of Theorem 1. Lemma 8 shows that T_n^A has law $\mathbf{P}_n^q$ for a particular choice of $(q_n)_{n \geq 1}$. Theorem 7 and Lemma 14 then show that the hypotheses of Theorem 5 are satisfied. □

4. Acknowledgments

I would like to thank my adviser, Jim Pitman, for suggesting the problem that led to this paper and for helpful conversations and feedback along the way. I would also like to thank the referee for encouragement to remove the condition that $0 \in A$, which appeared in earlier versions of this paper.

References

[1] Romain Abraham and Jean-François Delmas. Local limits of conditioned Galton-Watson trees I: the infinite spine case. *arXiv:1304.4035v3*, 2013.

[2] David Aldous. The continuum random tree. III. *Ann. Probab.*, 21(1):248–289, 1993.

[3] Jean Bertoin. *Random fragmentation and coagulation processes*, volume 102 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2006.

[4] Bénédicte Haas and Grégory Miermont. The genealogy of self-similar fragmentations with negative index as a continuum random tree. *Electron. J. Probab.*, 9:no. 4, 57–97 (electronic), 2004.

[5] Bénédicte Haas and Grégory Miermont. Scaling limits of Markov branching trees with applications to Galton-Watson and random unordered trees. *Ann. Probab.*, 40(6):2589–2666, 2012.
[6] Bénédicte Haas, Grégory Miermont, Jim Pitman, and Matthias Winkel. Continuum tree asymptotics of discrete fragmentations and applications to phylogenetic models. *Ann. Probab.*, 36(5):1790–1837, 2008.
[7] Igor Kortchemski. Invariance principles for Galton–Watson trees conditioned on the number of leaves. *Stochastic Process. Appl.*, 122(9):3126–3172, 2012.
[8] Jean-François Le Gall. Random real trees. *Ann. Fac. Sci. Toulouse Math. (6)*, 15(1):35–62, 2006.
[9] Jean-François Le Gall. Itô's excursion theory and random trees. *Stochastic Process. Appl.*, 120(5):721–749, 2010.
[10] Nariyuki Minami. On the number of vertices with a given degree in a Galton-Watson tree. *Adv. in Appl. Probab.*, 37(1):229–264, 2005.
[11] J. Pitman. *Combinatorial stochastic processes*, volume 1875 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 2006. Lectures from the 32nd Summer School on Probability Theory held in Saint-Flour, July 7–24, 2002.
[12] Douglas Rizzolo. Scaling limits of Markov branching trees and Galton-Watson trees conditioned on the number of vertices with out-degree in a given set. *arXiv:1105.2528v1*, 2011.

Department of Mathematics, University of Washington, WA 98105
E-mail address: drizzolo@math.washington.edu