J. Appl. Prob. **54,** 55–65 (2017)
*doi:*10.1017/jpr.2016.86

LOCAL LIMITS OF GALTON–WATSON TREES CONDITIONED ON THE NUMBER OF PROTECTED NODES

ROMAIN ABRAHAM,* *Université d'Orléans*
AYMEN BOUAZIZ,** *Université de Tunis El Manar*
JEAN-FRANÇOIS DELMAS,*** *Ecole des Ponts*

Abstract

We consider a marking procedure of the vertices of a tree where each vertex is marked independently from the others with a probability that depends only on its out-degree. We prove that a critical Galton–Watson tree conditioned on having a large number of marked vertices converges in distribution to the associated size-biased tree. We then apply this result to give the limit in distribution of a critical Galton–Watson tree conditioned on having a large number of protected nodes.

Keywords: Galton–Watson tree; random tree; local limit; protected node

2010 Mathematics Subject Classification: Primary 60J80; 60B10
Secondary 05C05

1. Introduction

In [6] Kesten proved that a critical or subcritical Galton–Watson (GW) tree conditioned on reaching at least height h converges in distribution (for the local topology on trees) as h goes to ∞ toward the so-called sized-biased tree (that we call here Kesten's tree and whose distribution is described in Section 3.2). Since then, other conditionings have been considered, see [1], [2], [4], and the references therein for recent developments on the subject.

A protected node is a node that is not a leaf and none of its offspring is a leaf. Precise asymptotics for the number of protected nodes in a conditioned GW tree have already been obtained in [3], [5], for instance. Let $A(\boldsymbol{t})$ be the number of protected nodes in the tree $\boldsymbol{t}$. We remark that this functional A is clearly monotone in the sense of [4] (using, for instance, (5.1)); therefore, using Theorem 2.1 of [4], we immediately find that a critical GW tree τ conditioned on $\{A(\tau) > n\}$ converges in distribution toward Kesten's tree as n goes to ∞. Conditioning on $\{A(\tau) = n\}$ needs extra work and is the main objective of this paper. Using the general result of [1], if we have the following limit

$$\lim_{n\to+\infty} \frac{\mathbb{P}(A(\tau)=n+1)}{\mathbb{P}(A(\tau)=n)} = 1, \tag{1.1}$$

then the critical GW tree τ conditioned on $\{A(\tau) = n\}$ converges in distribution also toward Kesten's tree; see Theorem 5.1.

Received 10 September 2015; revision received 22 June 2016.
* Postal address: Laboratoire MAPMO, Université d'Orléans, BP 6759, 45067 Orléans Cedex 2, France.
Email address: romain.abraham@univ-orleans.fr
** Postal address: Institut préparatoire aux études scientifiques et techniques, Université de Tunis El Manar, 2070 La Marsa, Tunis, Tunisie.
*** Postal address: CERMICS, Ecole des Ponts, Université Paris-Est, Champs-sur-Marne, 77455 Marne-la-Vallée Cedex 2, France.

In fact, the limit (1.1) can be seen as a special case of a more general problem: conditionally given the tree, we mark the nodes of the tree independently of the rest of the tree with a probability that depends only on the number of offspring of the nodes. Then we prove that a critical GW tree conditioned on the total number of marked nodes being large converges in distribution toward Kesten's tree; see Theorem 3.1.

The paper is then organised as follows. We first recall briefly the framework of discrete trees, then we consider in Section 3 the problem of a marked GW tree and the proofs of the results are given in Section 4. In particular, in Lemma 4.1 we prove the limit (1.1) when A is the number of marked nodes, and we deduce the convergence of a critical GW tree conditioned on the number of marked nodes toward Kesten's tree in Theorem 3.1. Finally, in Section 5 we explain how the problem of protected nodes can be viewed as a problem on marked nodes. and deduce the convergence in distribution of a critical GW tree conditioned on the number of protected nodes toward Kesten's tree in Theorem 5.1.

2. Technical background on GW trees

2.1. The set of discrete trees

We denote by $\mathbb{N} = \{0, 1, 2, \ldots\}$ the set of nonnegative integers and by $\mathbb{N}^* = \{1, 2, \ldots\}$ the set of positive integers.

If E is a subset of $\mathbb{N}^*$, we call the span of E the greatest common divisor of E. If X is an integer-valued random variable, we call the span of X the span of $\{n > 0;\ \mathbb{P}(X = n) > 0\}$.

We recall Neveu's formalism [7] for ordered rooted trees. Let $\mathcal{U} = \bigcup_{n\geq 0}(\mathbb{N}^*)^n$ be the set of finite sequences of positive integers with the convention $(\mathbb{N}^*)^0 = \{\varnothing\}$. For $u \in \mathcal{U}$, its length or generation $|u| \in \mathbb{N}$ is defined by $u \in (\mathbb{N}^*)^{|u|}$. If u and v are two sequences of $\mathcal{U}$, we denote by uv the concatenation of the two sequences, with the convention that $uv = u$ if $v = \varnothing$ and $uv = v$ if $u = \varnothing$. The set of ancestors of u is the set

$$\mathrm{An}(u) = \{v \in \mathcal{U};\ \text{there exists } w \in \mathcal{U} \text{ such that } u = vw\}.$$

Note that u belongs to $\mathrm{An}(u)$. For two distinct elements u and v of $\mathcal{U}$, we denote by $u < v$ the lexicographic order on $\mathcal{U}$, i.e. $u < v$ if $u \in \mathrm{An}(v)$ and $u \neq v$ or if $u = wiu'$ and $v = wjv'$ for some $i, j \in \mathbb{N}^*$ with $i < j$. We write $u \leq v$ if $u = v$ or $u < v$.

A tree $\boldsymbol{t}$ is a subset of $\mathcal{U}$ that satisfies the following.

- $\varnothing \in \boldsymbol{t}$.
- If $u \in \boldsymbol{t}$ then $\mathrm{An}(u) \subset \boldsymbol{t}$.
- For every $u \in \boldsymbol{t}$, there exists $k_u(\boldsymbol{t}) \in \mathbb{N}$ such that, for every $i \in \mathbb{N}^*$, $ui \in \boldsymbol{t}$ if and only if $1 \leq i \leq k_u(\boldsymbol{t})$.

The vertex $\varnothing$ is called the root of $\boldsymbol{t}$. The integer $k_u(\boldsymbol{t})$ represents the number of offspring of the vertex $u \in \boldsymbol{t}$. The set of children of a vertex $u \in \boldsymbol{t}$ is given by

$$C_u(\boldsymbol{t}) = \{ui;\ 1 \leq i \leq k_u(\boldsymbol{t})\}.$$

By convention, we set $k_u(\boldsymbol{t}) = -1$ if $u \notin \boldsymbol{t}$.

A vertex $u \in \boldsymbol{t}$ is called a leaf if $k_u(\boldsymbol{t}) = 0$. We denote by $\mathcal{L}_0(\boldsymbol{t})$ the set of leaves of $\boldsymbol{t}$. A vertex $u \in \boldsymbol{t}$ is called a *protected node* if $C_u(\boldsymbol{t}) \neq \varnothing$ and $C_u(\boldsymbol{t}) \cap \mathcal{L}_0(\boldsymbol{t}) = \varnothing$, that is, u is not a leaf and none of its children is a leaf. For $u \in \boldsymbol{t}$, we define $F_u(\boldsymbol{t})$, the fringe subtree of $\boldsymbol{t}$ above u, as

$$F_u(\boldsymbol{t}) = \{v \in \boldsymbol{t};\ u \in \mathrm{An}(v)\} = \{uv;\ v \in S_u(\boldsymbol{t})\}$$

with $S_u(\boldsymbol{t}) = \{v \in \mathcal{U};\ uv \in \boldsymbol{t}\}$.

Note that $S_u(\boldsymbol{t})$ is a tree. We denote by $\mathbb{T}$ the set of trees and by $\mathbb{T}_0 = \{\boldsymbol{t} \in \mathbb{T};\ \mathrm{card}(\boldsymbol{t}) < +\infty\}$ the subset of finite trees.

We say that a sequence of trees $(\boldsymbol{t}_n,\ n \in \mathbb{N})$ converges locally to a tree $\boldsymbol{t}$ if and only if $\lim_{n\to\infty} k_u(\boldsymbol{t}_n) = k_u(\boldsymbol{t})$ for all $u \in \mathcal{U}$. Let $(T_n,\ n \in \mathbb{N})$ and T be $\mathbb{T}$-valued random variables. We denote by $\mathrm{dist}(T)$ the distribution of the random variable T and write

$$\lim_{n\to+\infty} \mathrm{dist}(T_n) = \mathrm{dist}(T)$$

for the convergence in distribution of the sequence $(T_n,\ n \in \mathbb{N})$ to T with respect to the local topology.

If $\boldsymbol{t}, \boldsymbol{t}' \in \mathbb{T}$ and $x \in \mathcal{L}_0(\boldsymbol{t})$, we denote by

$$\boldsymbol{t} \circledast_x \boldsymbol{t}' = \{u \in \boldsymbol{t}\} \cup \{xv;\ v \in \boldsymbol{t}'\}$$

the tree obtained by grafting the tree $\boldsymbol{t}'$ on the leaf x of the tree $\boldsymbol{t}$. For every $\boldsymbol{t} \in \mathbb{T}$ and every $x \in \mathcal{L}_0(\boldsymbol{t})$, we shall consider the set of trees obtained by grafting a tree on the leaf x of $\boldsymbol{t}$, i.e.

$$\mathbb{T}(\boldsymbol{t}, x) = \{\boldsymbol{t} \circledast_x \boldsymbol{t}';\ \boldsymbol{t}' \in \mathbb{T}\}.$$

2.2. GW trees

Let $p = (p(n),\ n \in \mathbb{N})$ be a probability distribution on $\mathbb{N}$. We assume that

$$p(0) > 0, \qquad p(0) + p(1) < 1, \quad \text{and} \quad \mu := \sum_{n=0}^{+\infty} np(n) < +\infty. \tag{2.1}$$

A $\mathbb{T}$-valued random variable τ is a GW tree with offspring distribution p if the distribution of $k_\varnothing(\tau)$ is p and it enjoys the branching property: for $n \in \mathbb{N}^*$, conditionally on $\{k_\varnothing(\tau) = n\}$, the subtrees $(S_1(\tau), \dots, S_n(\tau))$ are independent and distributed as the original tree τ.

The GW tree and the offspring distribution are called critical (respectively subcritical, supercritical) if $\mu = 1$ (respectively $\mu < 1$, $\mu > 1$).

3. Conditioning on the number of marked vertices

3.1. Definition of the marking procedure

We begin with a fixed tree $\boldsymbol{t}$. We add marks on the vertices of $\boldsymbol{t}$ in an independent way such that the probability of adding a mark on a node u depends only on the number of children of u. More precisely, we consider a mark function $q\colon \mathbb{N} \to [0, 1]$ and a family of independent Bernoulli random variables $(Z_u(\boldsymbol{t}),\ u \in \boldsymbol{t})$ such that, for all $u \in \boldsymbol{t}$,

$$\mathbb{P}(Z_u(\boldsymbol{t}) = 1) = 1 - \mathbb{P}(Z_u(\boldsymbol{t}) = 0) = q(k_u(\boldsymbol{t})).$$

The vertex u is said to have a mark if $Z_u(\boldsymbol{t}) = 1$. We denote by $\mathcal{M}(\boldsymbol{t}) = \{u \in \boldsymbol{t};\ Z_u(\boldsymbol{t}) = 1\}$ the set of marked vertices and by $M(\boldsymbol{t})$ its cardinal. We call $(\boldsymbol{t}, \mathcal{M}(\boldsymbol{t}))$ a marked tree.

A marked GW tree with offspring distribution p and mark function q is a couple $(\tau, \mathcal{M}(\tau))$, with τ a GW tree with offspring distribution p and conditionally on $\{\tau = \boldsymbol{t}\}$ the set of marked vertices $\mathcal{M}(\tau)$ is distributed as $\mathcal{M}(\boldsymbol{t})$.

Remark 3.1. Note that, for $\mathcal{A} \subseteq \mathbb{N}$, if we set $q(k) = \mathbf{1}_{\{k\in\mathcal{A}\}}$ then the set $\mathcal{M}(\boldsymbol{t})$ is just the set of vertices with out-degree (i.e. number of offspring) in $\mathcal{A}$ considered in [1], [8]. Hence, the above construction can be seen as an extension of this case.

3.2. Kesten's tree

Let p be an offspring distribution satisfying assumption (2.1) with $\mu \le 1$ (i.e. the associated GW process is critical or subcritical). We denote by $p^* = (p^*(n) = np(n)/\mu,\ n \in \mathbb{N})$ the corresponding size-biased distribution.

We define an infinite random tree τ^* (the size-biased tree that we call Kesten's tree in this paper) whose distribution is described as follows.

There exists a unique infinite sequence $(v_k,\ k \in \mathbb{N}^*)$ of positive integers such that, for every $h \in \mathbb{N}$, $v_1 \cdots v_h \in \tau^*$, with the convention that $v_1 \cdots v_h = \varnothing$ if $h = 0$. The joint distribution of $(v_k,\ k \in \mathbb{N}^*)$ and τ^* is determined recursively as follows. For each $h \in \mathbb{N}$, conditionally given $(v_1, \dots, v_h)$ and $\{u \in \tau^*;\ |u| \le h\}$ the tree τ^* up to level h, we have the following.

- The number of children $(k_u(\tau^*),\ u \in \tau^*,\ |u| = h)$ are independent and distributed according to p if $u \ne v_1 \cdots v_h$ and according to p^* if $u = v_1 \dots v_h$.
- Given $\{u \in \tau^*;\ |u| \le h+1\}$ and $(v_1, \dots, v_h)$, the integer v_{h+1} is uniformly distributed on the set of integers $\{1, \dots, k_{v_1\cdots v_h}(\tau^*)\}$.

Remark 3.2. Note that by construction, almost surely τ^* has a unique infinite spine. And following Kesten [6], the random tree τ^* can be viewed as the tree τ conditioned on non-extinction.

For $\boldsymbol{t} \in \mathbb{T}_0$ and $x \in \mathcal{L}_0(\boldsymbol{t})$, we have

$$\mathbb{P}(\tau^* \in \mathbb{T}(t, x)) = \frac{\mathbb{P}(\tau = t)}{\mu^{|x|} p(0)}.$$

3.3. Main theorem

Theorem 3.1. *Let p be a critical offspring distribution that satisfies assumption (2.1). Let $(\tau, \mathcal{M}(\tau))$ be a marked GW tree with offspring distribution p and mark function q such that $p(k)q(k) > 0$ for some $k \in \mathbb{N}$. For every $n \in \mathbb{N}^*$, let τ_n be a tree whose distribution is the conditional distribution of τ given $\{M(\tau) = n\}$. Let τ^* be a Kesten's tree associated with p. Then we have*

$$\lim_{n\to+\infty} \operatorname{dist}(\tau_n) = \operatorname{dist}(\tau^*),$$

where the limit has to be understood along a subsequence for which $\mathbb{P}(M(\tau) = n) > 0$.

Remark 3.3. If, for every $k \in \mathbb{N}$, $0 < q(k) < 1$ then $\mathbb{P}(M(\tau)) = n) > 0$ for every $n \in \mathbb{N}$, hence, the above conditioning is always valid.

4. Proof of Theorem 3.1

Set $\gamma = \mathbb{P}(M(\tau) > 0)$. Since there exists $k \in \mathbb{N}$ with $p(k)q(k) > 0$, we have $\gamma > 0$. A sufficient condition (but not necessary) to have $\mathbb{P}(M(\tau) = n) > 0$ for every large enough n is to assume that $\gamma < 1$ (see Lemma 4.2 and Section 4.4). Taking $q = \mathbf{1}_{\mathcal{A}}$, see Remark 3.1 for $0 \in \mathcal{A} \subset \mathbb{N}$ implies that $\gamma = 1$ and some periodicity may occur.

The following result is the analogue in the random case of Theorem 3.1 in [1] and its proof is in fact a straightforward adaptation of the proof in [1] by using the following.

(i) $M(\boldsymbol{t}) \le \mathrm{card}(\boldsymbol{t})$.

(ii) For every $\boldsymbol{t} \in \mathbb{T}_0$, $x \in \mathcal{L}_0(\boldsymbol{t})$, and $\boldsymbol{t}' \in \mathbb{T}$, it follows that $M(\boldsymbol{t} \circledast_x \boldsymbol{t}')$ is distributed as $\hat{M}(\boldsymbol{t}') + M(\boldsymbol{t}) - \mathbf{1}_{\{Z_x(\boldsymbol{t})=1\}}$, where $\hat{M}(\boldsymbol{t}')$ is distributed as $M(\boldsymbol{t}')$ and is independent of $\mathcal{M}(\boldsymbol{t})$.

Proposition 4.1. *Let* $n_0 \in \mathbb{N} \cup \{\infty\}$. *Assume that* $\mathbb{P}(M(\tau) \in [n, n+n_0)) > 0$ *for large enough* n. *Then, if*

$$\lim_{n\to+\infty} \frac{\mathbb{P}(M(\tau) \in [n+1, n+1+n_0)}{\mathbb{P}(M(\tau) \in [n, n+n_0))} = 1, \tag{4.1}$$

we have

$$\lim_{n\to+\infty} \mathrm{dist}(\tau \mid M(\tau) \in [n, n+n_0)) = \mathrm{dist}(\tau^*).$$

Proof. According to Lemma 2.1 in [1], a sequence $(T_n,\ n \in \mathbb{N})$ of finite random trees converges in distribution (with respect to the local topology) to some Kesten's tree τ^* if and only if, for every finite tree $\boldsymbol{t} \in \mathbb{T}_0$ and every leaf $x \in \mathcal{L}_0(\boldsymbol{t})$,

$$\lim_{n\to+\infty} \mathbb{P}((T_n \in \mathbb{T}(\boldsymbol{t}, x)) = \mathbb{P}(\tau^* \in \mathbb{T}(\boldsymbol{t}, x)) \quad \text{and} \quad \lim_{n\to+\infty} \mathbb{P}(T_n = \boldsymbol{t}) = 0. \tag{4.2}$$

Let $\boldsymbol{t} \in \mathbb{T}_0$ and $x \in \mathcal{L}_0(\boldsymbol{t})$. We set $D(\boldsymbol{t}, x) = M(\boldsymbol{t}) - \mathbf{1}_{\{Z_x(\boldsymbol{t})=1\}}$. Note that $D(\boldsymbol{t}, x) \le \mathrm{card}(\boldsymbol{t}) - 1$. Elementary computations yield, for every $\boldsymbol{t}' \in \mathbb{T}_0$,

$$\mathbb{P}(\tau = \boldsymbol{t} \circledast_x \boldsymbol{t}') = \frac{1}{p(0)}\mathbb{P}(\tau = \boldsymbol{t})\mathbb{P}(\tau = \boldsymbol{t}') \quad \text{and} \quad \mathbb{P}(\tau^* \in \mathbb{T}(\boldsymbol{t}, x)) = \frac{1}{p(0)}\mathbb{P}(\tau = \boldsymbol{t}).$$

As τ is almost surely finite, we have

$$\begin{aligned}
&\mathbb{P}(\tau \in \mathbb{T}(\boldsymbol{t}, x), M(\tau) \in [n, n+n_0)) \\
&\quad = \sum_{\boldsymbol{t}' \in \mathbb{T}_0} \mathbb{P}(\tau = \boldsymbol{t} \circledast_x \boldsymbol{t}', M(\tau) \in [n, n+n_0)) \\
&\quad = \sum_{\boldsymbol{t}' \in \mathbb{T}_0} \mathbb{P}(\tau = \boldsymbol{t} \circledast_x \boldsymbol{t}')\mathbb{P}(M(\boldsymbol{t} \circledast_x \boldsymbol{t}') \in [n, n+n_0)) \\
&\quad = \sum_{\boldsymbol{t}' \in \mathbb{T}_0} \frac{\mathbb{P}(\tau = \boldsymbol{t})\mathbb{P}(\tau = \boldsymbol{t}')}{p(0)}\mathbb{P}(\hat{M}(\boldsymbol{t}') + D(\boldsymbol{t}, x) \in [n, n+n_0)) \\
&\quad = \mathbb{P}(\tau^* \in \mathbb{T}(\boldsymbol{t}, x))\mathbb{P}(\hat{M}(\tau) + D(\boldsymbol{t}, x) \in [n, n+n_0)).
\end{aligned}$$

Note that

$$\begin{aligned}
&\mathbb{P}(\hat{M}(\tau) + D(\boldsymbol{t}, x) \in [n, n+n_0)) \\
&\quad = \sum_{k=0}^{\mathrm{card}(\boldsymbol{t})-1} \mathbb{P}(\hat{M}(\tau) + D(\boldsymbol{t}, x) \in [n, n+n_0) \mid D(\boldsymbol{t}, x) = k)\mathbb{P}(D(\boldsymbol{t}, x) = k) \\
&\quad = \sum_{k=0}^{\mathrm{card}(\boldsymbol{t})-1} \mathbb{P}(M(\tau) \in [n-k, n+n_0-k))\mathbb{P}(D(\boldsymbol{t}, x) = k).
\end{aligned}$$

Then, using assumption (4.1), we obtain

$$\lim_{n\to+\infty} \frac{\mathbb{P}(\hat{M}(\tau) + D(\boldsymbol{t}, x) \in [n, n+n_0))}{\mathbb{P}(M(\tau) \in [n, n+n_0))} = 1,$$

that is,

$$\lim_{n\to+\infty} \mathbb{P}(\tau \in \mathbb{T}(\boldsymbol{t}, x) \mid M(\tau) \in [n, n+n_0)) = \mathbb{P}(\tau^* \in \mathbb{T}(\boldsymbol{t}, x)).$$

This proves the first limit of (4.2).

The second limit is immediate since, for every $n \geq \text{card}(\boldsymbol{t})$,

$$\mathbb{P}(\tau = \boldsymbol{t} \mid M(\tau) \in [n, n+n_0)) = 0. \qquad \square$$

The main ingredient for the proof of Theorem 3.1 is the following lemma.

Lemma 4.1. *Let d be the span of the random variable* $M(\tau) - 1$. *We have*

$$\lim_{n\to+\infty} \frac{\mathbb{P}(M(\tau) \in [n+1, n+1+d))}{\mathbb{P}(M(\tau) \in [n, n+d))} = 1. \tag{4.3}$$

The end of this section is devoted to the proof of Lemma 4.1, see Section 4.4, which follows the ideas of the proof of Theorem 5.1 of [1].

4.1. Transformation of a subset of a tree onto a tree

We recall Rizzolo's map [8] which, from $\boldsymbol{t} \in \mathbb{T}_0$ and a nonempty subset A of $\boldsymbol{t}$, builds a tree $\boldsymbol{t}_A$ such that $\text{card}(A) = \text{card}(\boldsymbol{t}_A)$. We will give a recursive construction of this map $\phi\colon (\boldsymbol{t}, A) \mapsto \boldsymbol{t}_A = \phi(\boldsymbol{t}, A)$. We will check in the next section that this map is such that if τ is a GW tree then τ_A will also be a GW tree for a well chosen subset A of τ. In Figure 1 we show an example of a tree $\boldsymbol{t}$, a set A, and the associated tree $\boldsymbol{t}_A$ which helps to understand the construction.

For a vertex $u \in \boldsymbol{t}$, recall that $C_u(\boldsymbol{t})$ is the set of children of u in $\boldsymbol{t}$. We define, for $u \in \boldsymbol{t}$,

$$R_u(\boldsymbol{t}) = \bigcup_{w \in \text{An}(u)} \{v \in C_w(\boldsymbol{t});\ u < v\}$$

the vertices of $\boldsymbol{t}$ which are larger than u for the lexicographic order and are children of u or of one of its ancestors. For a vertex $u \in \boldsymbol{t}$, we shall consider A_u the set of elements of A in the fringe subtree above u, i.e.

$$A_u = A \cap F_u(\boldsymbol{t}) = A \cap \{uv;\ v \in S_u(\boldsymbol{t})\}. \tag{4.4}$$

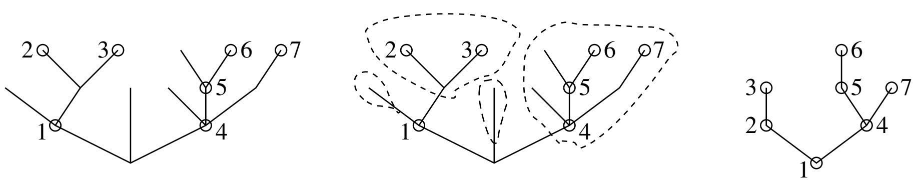

FIGURE 1: *Left:* a tree $\boldsymbol{t}$ and the set A. *Centre:* the fringe subtrees rooted at the vertices in $R_{u_0}(\boldsymbol{t})$. *Right:* the tree $\boldsymbol{t}_A$. The labels have no signification, they only show which node of $\boldsymbol{t}$ corresponds to a node of $\boldsymbol{t}_A$.

Let $\boldsymbol{t} \in \mathbb{T}_0$ and $A \subset \boldsymbol{t}$ such that $A \neq \varnothing$. We shall define $\boldsymbol{t}_A = \phi(\boldsymbol{t}, A)$ recursively. Let u_0 be the smallest (for the lexicographic order) element of A. Consider the fringe subtrees of $\boldsymbol{t}$ that are rooted at the vertices in $R_{u_0}(\boldsymbol{t})$ and contain at least one vertex in A, that is $(F_u(\boldsymbol{t});\ u \in R^A_{u_0}(\boldsymbol{t}))$, with

$$R^A_{u_0}(\boldsymbol{t}) = \{u \in R_{u_0}(\boldsymbol{t});\ A_u \neq \varnothing\} = \{u \in R_{u_0}(\boldsymbol{t});\ \text{there exists } v \in A \text{ such that } u \in \mathrm{An}(v)\}.$$

Define the number of children of the root of tree $\boldsymbol{t}_A$ as the number of those fringe subtrees

$$k_\varnothing(\boldsymbol{t}_A) = \mathrm{card}(R^A_{u_0}(\boldsymbol{t})).$$

If $k_\varnothing(\boldsymbol{t}_A) = 0$ set $\boldsymbol{t}_A = \{\varnothing\}$. Otherwise, let $u_1 < \cdots < u_{k_\varnothing(\boldsymbol{t}_A)}$ be the ordered elements of $R^A_{u_0}(\boldsymbol{t})$ with respect to the lexicographic order on $\mathcal{U}$. And we define $\boldsymbol{t}_A = \phi(\boldsymbol{t}, A)$ recursively by

$$F_i(\boldsymbol{t}_A) = \phi(F_{u_i}(\boldsymbol{t}), A_{u_i}) \quad \text{for } 1 \leq i \leq k_\varnothing(\boldsymbol{t}_A). \tag{4.5}$$

Since $\mathrm{card}(A_{u_i}) < \mathrm{card}(A)$, we deduce that $\boldsymbol{t}_A = \phi(\boldsymbol{t}, A)$ is well defined and is a tree by construction. Furthermore, we clearly have that A and $\boldsymbol{t}_A$ have the same cardinal, i.e.

$$\mathrm{card}(\boldsymbol{t}_A) = \mathrm{card}(A). \tag{4.6}$$

4.2. Distribution of the number of marked nodes

Let $(\tau, \mathcal{M}(\tau))$ be a marked GW tree with critical offspring distribution p satisfying (2.1) and mark function q. Recall that $\gamma = \mathbb{P}(M(\tau) > 0) = \mathbb{P}(\mathcal{M}(\tau) \neq \varnothing)$.

Let $((X_i, Z_i), i \in \mathbb{N}^*)$ be independent and identically distributed random variables such that X_i is distributed according to p and Z_i is conditionally on X_i Bernoulli with parameter $q(X_i)$. We have the following definitions.

- $G = \inf\{k \in \mathbb{N}^*;\ \sum_{i=1}^k (X_i - 1) = -1\}$.
- $N = \inf\{k \in \mathbb{N}^*;\ Z_k = 1\}$.
- $\tilde{X}$ a random variable distributed as $1 + \sum_{i=1}^N (X_i - 1)$ conditionally on $\{N \leq G\}$.
- Y a random variable which is conditionally on $\tilde{X}$ binomial with parameter $(\tilde{X}, \gamma)$.

We say that a probability distribution on $\mathbb{N}$ is aperiodic if the span of its support restricted to $\mathbb{N}^*$ is 1. The following result is immediate as the distribution p of X_1 satisfies (2.1).

Lemma 4.2. *The distribution of Y satisfies (2.1) and if $\gamma < 1$ then it is aperiodic.*

Recall that for a tree $\boldsymbol{t} \in \mathbb{T}_0$, we have

$$\sum_{u \in \boldsymbol{t}} (k_u(\boldsymbol{t}) - 1) = -1 \tag{4.7}$$

and $\sum_{u \in \boldsymbol{t}, u < v} (k_u(\boldsymbol{t}) - 1) > -1$ for any $v \in \boldsymbol{t}$. We deduce that G is distributed according to $\mathrm{card}(\tau)$ and, thus, N is distributed like the index of the first marked vertex along the depth-first walk of τ. Then, we have

$$\gamma = \mathbb{P}(N \leq G). \tag{4.8}$$

We denote by $(\tau^0, \mathcal{M}(\tau^0))$ a random marked tree distributed as $(\tau, \mathcal{M}(\tau))$ conditioned on $\{\mathcal{M}(\tau) \neq \varnothing\}$. By construction, $\mathrm{card}(\tau^0)$ is distributed as G conditioned on $\{N \leq G\}$.

Lemma 4.3. *Under the hypothesis of this section, it holds that $\tau^0_{\mathcal{M}(\tau^0)} = \phi(\tau^0, \mathcal{M}(\tau^0))$ is a critical GW tree with the law of Y as offspring distribution.*

4.3. Proof of Lemma 4.3

In order to simplify notation, we write $\tilde{\tau}$ for $\tau^0_{\mathcal{M}(\tau^0)} = \phi(\tau^0, \mathcal{M}(\tau^0))$ and, for $u \in \tau^0$, we set R_u for $R_u(\tau^0)$.

Lemma 4.4. *The random tree* $\tilde{\tau}$ *is a GW tree with offspring distribution the law of* Y.

Proof. Let u_0 be the smallest (for the lexicographic order) element of $\mathcal{M}(\tau^0)$. The branching property of GW trees implies that, conditionally given u_0 and R_{u_0}, the fringe subtrees of τ^0 rooted at the vertices in R_{u_0}, $(S_u(\tau^0),\ u \in R_{u_0})$ are independent and distributed as τ. Recall notation (4.4) so that the set of marked vertices of the fringe subtree rooted at u is $\mathcal{M}_u(\tau^0) = \mathcal{M}(\tau^0) \cap F_u(\tau^0)$. Define $\tilde{\mathcal{M}}_u(\tau^0) = \{v;\ uv \in \mathcal{M}_u(\tau^0)\}$ the corresponding marked vertices of $S_u(\boldsymbol{t})$. Then, the construction of the marks $\mathcal{M}(\tau)$ implies that the corresponding marked trees $((S_u(\tau^0), \tilde{\mathcal{M}}_u(\tau^0)),\ u \in R_{u_0})$ are independent and distributed as $(\tau, \mathcal{M}(\tau))$. Note that, for $u \in R_{u_0}$, the fringe subtree $F_u(\tau^0)$ contains at least one mark if and only if u belongs to

$$R_{u_0}^{\mathcal{M}(\tau^0)} = \{u \in R_{u_0};\ \text{there exists } v \in \mathcal{M}(\tau^0) \text{ such that } u \in \mathrm{An}(v)\}.$$

Then by considering only the fringe subtrees containing at least one mark, we find that, conditionally on $R_{u_0}^{\mathcal{M}(\tau^0)}$, the subtrees $((S_u(\tau^0), \tilde{\mathcal{M}}_u(\tau^0)),\ u \in R_{u_0}^{\mathcal{M}(\tau^0)})$ are independent and distributed as $(\tau^0, \mathcal{M}(\tau^0))$. We deduce from the recursive construction of the map ϕ, see (4.5), that $\tilde{\tau}$ is a GW tree. Note that the offspring distribution of $\tilde{\tau}$ is given by the distribution of the cardinal of $R_{u_0}^{\mathcal{M}(\tau^0)}$. We now compute the corresponding offspring distribution. We first give an elementary formula for the cardinal of $R_u(\boldsymbol{t})$. Let $\boldsymbol{t} \in \mathbb{T}_0$ and $u \in \boldsymbol{t}$. Consider the tree $\boldsymbol{t}' = R_u(\boldsymbol{t}) \cup \{v \in \boldsymbol{t};\ v \le u\}$. Using (4.7) for $\boldsymbol{t}'$, we obtain

$$-1 = \sum_{v \in \boldsymbol{t}'} (k_v(\boldsymbol{t}') - 1) = \sum_{v \in \boldsymbol{t};\, v \le u} (k_v(\boldsymbol{t}') - 1) + \sum_{v \in R_u(\boldsymbol{t})} (-1).$$

From this we obtain $\mathrm{card}(R_u(\boldsymbol{t})) = 1 + \sum_{v \in \boldsymbol{t};\, v \le u} (k_v(\boldsymbol{t}') - 1)$. We deduce from the definition of $\tilde{X}$ that $\mathrm{card}(R_{u_0})$ is distributed as $\tilde{X}$. We deduce from the first part of the proof that conditionally on $\mathrm{card}(R_{u_0})$, the distribution of $\mathrm{card}(R_{u_0}^{\mathcal{M}(\tau^0)})$ is binomial with parameter $(\mathrm{card}(R_{u_0}(\tau^0)), \gamma)$. It follows that the offspring distribution of $\tilde{\tau}$ is given by the law of Y. □

Lemma 4.5. *The GW tree* $\tilde{\tau}$ *is critical.*

Proof. Since the offspring distribution is the law of Y, we need to check that $\mathbb{E}[Y] = 1$ is $\gamma \mathbb{E}[\tilde{X}] = 1$ since Y is conditionally on $\tilde{X}$ binomial with parameter $(\tilde{X}, \gamma)$.

Recall that N has finite expectation as $\mathbb{P}(Z_1 = 1) > 0$, is not independent of $(X_i)_{i \in \mathbb{N}^*}$, and is a stopping time with respect to the filtration generated by $((X_i,\ Z_i),\ i \in \mathbb{N}^*)$. Using Wald's equality and $\mathbb{E}[X_i] = 1$, we obtain $\mathbb{E}[\sum_{i=1}^N (X_i - 1)] = 0$ and, thus, using the definition of $\tilde{X}$ as well as (4.8),

$$\gamma \mathbb{E}[\tilde{X}] = \gamma + \mathbb{E}\Bigg[\sum_{i=1}^N (X_i - 1)\, \mathbf{1}_{\{N \le G\}}\Bigg] = \gamma - \mathbb{E}\Bigg[\sum_{i=1}^N (X_i - 1)\, \mathbf{1}_{\{N > G\}}\Bigg].$$

We have

$$\mathbb{E}\Big[\sum_{i=1}^{N}(X_i-1)\,\mathbf{1}_{\{N>G\}}\Big]=\mathbb{E}\Big[\sum_{i=1}^{G}(X_i-1)\,\mathbf{1}_{\{N>G\}}\Big]+\mathbb{P}(N>G)\mathbb{E}\Big[\sum_{i=1}^{N}(X_i-1)\Big]$$
$$=-\mathbb{P}(N>G)$$
$$=\gamma-1,$$

where we used the strong Markov property of $((X_i, Z_i), i\in\mathbb{N}^*)$ at the stopping time G for the first equation, the definition of T and Wald's equality for the second, and (4.8) for the third. We deduce that $\mathbb{E}[Y]=\gamma\mathbb{E}[\tilde{X}]=1$, which completes the proof. □

4.4. Proof of (4.3)

According to Lemma 4.3 and (4.6), it follows that $M(\tau^0)$ is distributed as the total size of a critical GW whose offspring distribution satisfies (2.1). The proof of Proposition 4.3 of [1] (see Equation (4.15) in [1]) entails that if τ' is a critical GW tree, then, if d denotes the span of the random variable $\mathrm{card}(\tau')-1$, we have

$$\lim_{n\to\infty}\frac{\mathbb{P}(\mathrm{card}(\tau')\in[n+1,n+1+d))}{\mathbb{P}(\mathrm{card}(\tau')\in[n,n+d))}=1. \qquad \square$$

5. Protected nodes

Recall that a node of a tree $\boldsymbol{t}$ is protected if it is not a leaf and none of its offspring is a leaf. We denote by $A(\boldsymbol{t})$ the number of protected nodes of the tree $\boldsymbol{t}$.

Theorem 5.1. *Let τ be a critical GW tree with offspring distribution p satisfying (2.1) and let τ^* be the associated Kesten's tree. Let τ_n be a random tree distributed as τ conditionally given $\{A(\tau)=n\}$. Then*

$$\lim_{n\to+\infty}\mathrm{dist}(\tau_n)=\mathrm{dist}(\tau^*).$$

Proof. Note that $\mathbb{P}(A(\tau)=n)>0$ for all $n\in\mathbb{N}$. Note that the functional A satisfies the additive property of [1], namely, for every $\boldsymbol{t}\in\mathbb{T}$, every $x\in\mathcal{L}_0(\boldsymbol{t})$, and every $\boldsymbol{t}'\in\mathbb{T}$ that is not reduced to the root, we have

$$A(\boldsymbol{t}\circledast_x\boldsymbol{t}')=A(\boldsymbol{t})+A(\boldsymbol{t}')+D(\boldsymbol{t},x), \tag{5.1}$$

where $D(\boldsymbol{t},x)=1$ if x is the only child of its first ancestor which is a leaf (therefore, this ancestor becomes a protected node in $\boldsymbol{t}\circledast_x\boldsymbol{t}'$) and $D(\boldsymbol{t},x)=0$ otherwise. According to Theorem 3.1 of [1], to complete the proof it is enough to check that

$$\lim_{n\to+\infty}\frac{\mathbb{P}(A(\tau)=n+1)}{\mathbb{P}(A(\tau)=n)}=1. \tag{5.2}$$

For a tree $\boldsymbol{t}\neq\{\varnothing\}$, let $\boldsymbol{t}_{\mathbb{N}^*}=\phi(\boldsymbol{t},\boldsymbol{t}\setminus\mathcal{L}_0(\boldsymbol{t}))$ be the tree obtained from $\boldsymbol{t}$ by removing the leaves. Let τ^0 be a random tree distributed as τ conditioned to $\{k_\varnothing(\tau)>0\}$. Using Theorem 6 and Corollary 2 of [8] with $A=\mathbb{N}^*$ (or Lemma 4.3 with $q(k)=\mathbf{1}_{\{k>0\}}$), it follows that $\tau^0_{\mathbb{N}^*}$ is a critical GW tree with offspring distribution

$$p_{\mathbb{N}^*}(k)=\sum_{n=\max(k,1)}^{+\infty}p(n)\binom{n}{k}(p(0))^{n-k}(1-p(0))^{k-1},\qquad k\in\mathbb{N}.$$

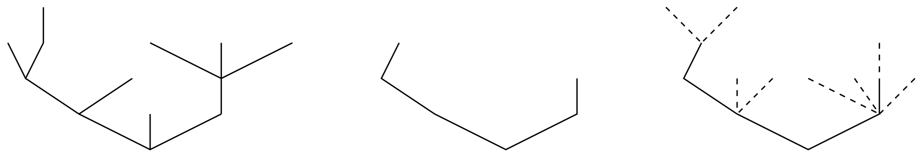

FIGURE 2: The trees τ^0, $\tau^0_{\mathbb{N}^*}$, and $\hat{\tau}$.

Conditionally given $\{\tau^0_{\mathbb{N}^*} = \boldsymbol{t}\}$, we consider independent random variables $(W(u),\ u \in \boldsymbol{t})$ taking values in $\mathbb{N}^*$ whose distributions are given, for all $u \in \boldsymbol{t}$, by $\mathbb{P}(W(u) = 0) = 0$ for $k_u(\boldsymbol{t}) = 0$ and otherwise, for $k_u(\boldsymbol{t}) + n > 0$ (remark that $p_{\mathbb{N}^*}(k_u(\boldsymbol{t})) > 0$), by

$$\mathbb{P}(W(u) = n) = \frac{p(k_u(\boldsymbol{t}) + n)}{p_{\mathbb{N}^*}(k_u(\boldsymbol{t}))}\binom{k_u(\boldsymbol{t}) + n}{n} p(0)^n (1 - p(0))^{k_u(\boldsymbol{t})-1}.$$

In particular, for $k_u(\boldsymbol{t}) > 0$, we have

$$\mathbb{P}(W(u) = 0) = \frac{p(k_u(\boldsymbol{t}))}{p_{\mathbb{N}^*}(k_u(\boldsymbol{t}))}(1 - p(0))^{k_u(\boldsymbol{t})-1}. \tag{5.3}$$

Then, we define a new tree $\hat{\tau}$ by grafting, on every vertex u of $\tau^0_{\mathbb{N}^*}$, $W(u)$ leaves in a uniform manner; see Figure 2.

More precisely, given $\tau^0_{\mathbb{N}^*}$ and $(W(u),\ u \in \tau^0_{\mathbb{N}^*})$, we define a tree $\hat{\tau}$ and a random map $\psi\colon \tau^0_{\mathbb{N}^*} \longmapsto \hat{\tau}$ recursively in the following way. We set $\psi(\varnothing) = \varnothing$. Then, given $k_\varnothing(\tau^0_{\mathbb{N}^*}) = k$, we set $k_\varnothing(\hat{\tau}) = k + W(\varnothing)$. We also consider a family $(i_1, \ldots, i_k)$ of integer-valued random variables such that $(i_1, i_2 - i_1, \ldots, i_k - i_{k-1}, W(u) + k + 1 - i_k)$ is a uniform positive partition of $W(u) + k + 1$. Then, for every $j \le k$ such that $j \notin \{i_1, \ldots, i_k\}$, we set $k_j(\hat{\tau}) = 0$, i.e. these are leaves of $\hat{\tau}$. For every $1 \le j \le k$, we set $\psi(j) = i_j$ and apply to them the same construction as for the root and so on. □

Lemma 5.1. *The new tree $\hat{\tau}$ is distributed as the original tree τ^0.*

Proof. Let $\boldsymbol{t} \in \mathbb{T}_0$. As $\mathbb{P}(\hat{\tau} = \{\varnothing\}) = 0$, we assume that $k_\varnothing(\boldsymbol{t}) > 0$. Let $\boldsymbol{t}_{\mathbb{N}^*}$ be the tree obtained from $\boldsymbol{t}$ by removing the leaves. Using (4.7), we have

$$\begin{aligned}\mathbb{P}(\hat{\tau} = \boldsymbol{t}) &= \prod_{u \in t_{\mathbb{N}^*}} p_{\mathbb{N}^*}(k_u(\boldsymbol{t}_{\mathbb{N}^*}))\mathbb{P}(W(u) = k_u(\boldsymbol{t}) - k_u(\boldsymbol{t}_{\mathbb{N}^*}))\binom{k_u(\boldsymbol{t})}{k_u(\boldsymbol{t}) - k_u(\boldsymbol{t}_{\mathbb{N}^*})}^{-1} \\ &= \frac{\mathbb{P}(\tau = t)}{1 - p(0)} \\ &= \mathbb{P}(\tau^0 = t).\end{aligned}$$

Note that the protected nodes of $\hat{\tau}$ are exactly the nodes of $\tau^0_{\mathbb{N}^*}$ on which we did not add leaves, i.e. for which $W(u) = 0$. If we set $\mathcal{M}(\tau^0_{\mathbb{N}^*}) = \{u \in \tau^0_{\mathbb{N}^*},\ W(u) = 0\}$, we have $M(\tau^0_{\mathbb{N}^*}) = A(\hat{\tau})$.

Using (5.3), we find that the corresponding mark function q is given by

$$q(k) = \frac{p(k)(1 - p(0))^{k-1}}{p_{\mathbb{N}^*}(k)} \mathbf{1}_{\{k \ge 1\}}.$$

As $\hat{\tau}$ is distributed as τ^0, we have

$$\lim_{n\to+\infty} \frac{\mathbb{P}(A(\tau^0) = n + 1)}{\mathbb{P}(A(\tau^0) = n)} = \lim_{n\to+\infty} \frac{\mathbb{P}(A(\hat{\tau}) = n + 1)}{\mathbb{P}(A(\hat{\tau}) = n)} = \lim_{n\to+\infty} \frac{\mathbb{P}(M(\tau^0_{\mathbb{N}^*}) = n + 1)}{\mathbb{P}(M(\tau^0_{\mathbb{N}^*}) = n)}.$$

As $\tau^0_{\mathbb{N}^*}$ is a critical GW tree, from Lemma 4.1 we deduce that

$$\lim_{n\to+\infty} \frac{\mathbb{P}(M(\tau^0_{\mathbb{N}^*}) = n+1)}{\mathbb{P}(M(\tau^0_{\mathbb{N}^*}) = n)} = 1.$$

As $\mathbb{P}(A(\tau) = n) = \mathbb{P}(A(\tau) = n \mid k_\varnothing(\tau) > 0)\mathbb{P}(k_\varnothing(\tau) > 0)$ and $\mathbb{P}(A(\tau) = n \mid k_\varnothing(\tau) > 0) = \mathbb{P}(A(\tau^0) = n)$ for $n \geq 2$, we obtain (5.2) and, hence, complete the proof. □

Acknowledgement

We would like to thank the anonymous referee for his/her useful comments which helped to improve the paper.

References

[1] Abraham, R. and Delmas, J.-F. (2014). Local limits of conditioned Galton–Watson trees: the infinite spine case. *Electron. J. Prob.* **19,** 19 pp.

[2] Abraham, R. and Delmas, J.-F. (2015). An introduction to Galton–Watson trees and their local limits. Lecture given in Hammamet, December 2014. Available at https://arxiv.org/abs/1506.05571.

[3] Devroye, L. and Janson, S. (2014). Protected nodes and fringe subtrees in some random trees. *Electron. Commun. Prob.* **19,** 10 pp.

[4] He, X. (2015). Local convergence of critical random trees and continuous-state branching processes. Preprint Available at https://arxiv.org/abs/1503.00951.

[5] Janson, S. (2016). Asymptotic normality of fringe subtrees and additive functionals in conditioned Galton–Watson trees. *Random Structures Algorithms* **48,** 57–101.

[6] Kesten, H. (1986). Subdiffusive behavior of random walk on a random cluster. *Ann. Inst. H. Poincaré Prob. Statist.* **22,** 425–487.

[7] Neveu, J. (1986). Arbres et processus de Galton–Watson. *Ann. Inst. H. Poincaré Prob. Statist.* **22,** 199–207.

[8] Rizzolo, D. (2015). Scaling limits of Markov branching trees and Galton–Watson trees conditioned on the number of vertices with out-degree in a given set. *Ann. Inst. H. Poincaré Prob. Statist.* **51,** 512–532.