Applied Probability Trust (22 June 2016)

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 trees; local limits; protected nodes

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

1. Introduction

In [6], Kesten proved that a critical or sub-critical Galton-Watson (GW) tree conditioned on reaching at least height h converges in distribution (for the local topology on trees) as h goes to infinity 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 offsprings 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 $A(\mathbf{t})$ be the number of protected nodes in the tree $\mathbf{t}$. 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 get that a critical GW tree τ conditioned on $\{A(\tau) > n\}$ converges in distribution toward Kesten's tree as n goes to infinity. 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

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

* Postal address: Romain Abraham, Laboratoire MAPMO, CNRS, UMR 7349, Fédération Denis Poisson, FR 2964, Université d'Orléans, B.P. 6759, 45067 Orléans cedex 2, France.
** Postal address: Aymen Bouaziz, Institut préparatoire aux études scientifiques et techniques, 2070 La Marsa, Tunisie.
*** Postal address: Jean-François Delmas, CERMICS, Ecole des Ponts, UPE, Champs-sur-Marne, France.

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

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 offsprings 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 organized 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, we prove the limit (1.1) when A is the number of marked nodes in Lemma 4.1 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. We finally explain in Section 5 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 non-negative 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 $(\mathbb{N}^*)^0 = \{\emptyset\}$. 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 $v = \emptyset$ and $uv = v$ if $u = \emptyset$. The set of ancestors of u is the set

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

Notice 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 $\mathbf{t}$ is a subset of $\mathcal{U}$ that satisfies:

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

The vertex $\emptyset$ is called the root of $\mathbf{t}$. The integer $k_u(\mathbf{t})$ represents the number of offsprings of the vertex $u \in \mathbf{t}$. The set of children of a vertex $u \in \mathbf{t}$ is given by:

$$C_u(\mathbf{t}) = \{ui;\ 1 \leq i \leq k_u(\mathbf{t})\}. \tag{2.1}$$

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

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

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

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

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

We say that a sequence of trees $(\mathbf{t}_n,\ n \in \mathbb{N})$ converges locally to a tree $\mathbf{t}$ if and only if $\lim_{n\to\infty} k_u(\mathbf{t}_n) = k_u(\mathbf{t})$ 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

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

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

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

$$\mathbf{t} \circledast_x \mathbf{t}' = \{u \in \mathbf{t}\} \cup \{xv;\ v \in \mathbf{t}'\} \tag{2.2}$$

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

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

2.2. Galton Watson trees

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

$$p(0) > 0,\ p(0) + p(1) < 1,\ \text{and } \mu := \sum_{n=0}^{+\infty} np(n) < +\infty. \tag{2.3}$$

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

The GW tree and the offspring distribution are called critical (resp. sub-critical, super-critical) if $\mu = 1$ (resp. $\mu < 1,\ \mu > 1$).

3. Conditioning on the number of marked vertices

3.1. Definition of the marking procedure

We begin with a fixed tree $\mathbf{t}$. We add marks on the vertices of $\mathbf{t}$ 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 $q : \mathbb{N} \longrightarrow [0, 1]$ and a family of independent Bernoulli random variables $(Z_u(\mathbf{t}),\ u \in \mathbf{t})$ such that for all $u \in \mathbf{t}$:

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

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

Remark 3.1. Notice that for $\mathcal{A} \subseteq \mathbb{N}$, if we set $q(k) = \mathbf{1}_{\{k\in\mathcal{A}\}}$, then the set $\mathcal{M}(\mathbf{t})$ is just the set of vertices with out-degree (i.e. number of offsprings) 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 (2.3) with $\mu \le 1$ (i.e. the associated GW process is critical or sub-critical). 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 $v_1 \cdots v_h = \emptyset$ 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 $(v_1, \ldots, v_h)$ and $\{u \in \tau^*;\, |u| \le h\}$ the tree τ^* up to level h, we have:

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

Remark 3.2. Notice that by construction, a.s. 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 $\mathbf{t} \in \mathbb{T}_0$ and $x \in \mathcal{L}_0(\mathbf{t})$, 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 (2.3). 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 n large enough 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 $\gamma = 1$ 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 straighforward adaptation of the proof in [1] by using:

(i) $M(\mathbf{t}) \leq \mathrm{Card}(\mathbf{t})$.

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

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

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

we have:

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

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 $\mathbf{t} \in \mathbb{T}_0$ and every leaf $x \in \mathcal{L}_0(\mathbf{t})$,

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

Let $\mathbf{t} \in \mathbb{T}_0$ and $x \in \mathcal{L}_0(\mathbf{t})$. We set $D(\mathbf{t}, x) = M(\mathbf{t}) - \mathbf{1}_{\{Z_x(\mathbf{t})=1\}}$. Notice that $D(\mathbf{t}, x) \leq \mathrm{Card}(\mathbf{t}) - 1$. Elementary computations give for every $\mathbf{t}' \in \mathbb{T}_0$ that:

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

As τ is a.s. finite, we have:

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

Notice that:

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

Then we obtain using Assumption (4.1) that:

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

that is

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

This proves the first limit of (4.2).

The second limit is immediate since, for every $n \geq \mathrm{Card}(\mathbf{t})$,

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

□

The main ingredient for the proof of Theorem 3.1 is then 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 $\mathbf{t} \in \mathbb{T}_0$ and a non-empty subset A of $\mathbf{t}$ builds a tree $\mathbf{t}_A$ such that $\mathrm{Card}(A) = \mathrm{Card}(\mathbf{t}_A)$. We will give a recursive construction of this map ϕ: $(\mathbf{t}, A) \mapsto \mathbf{t}_A = \phi(\mathbf{t}, A)$. 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. Figure 1 below shows an example of a tree $\mathbf{t}$, a set A and the associated tree $\mathbf{t}_A$ which helps to understand the construction.

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

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

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

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

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

$$R_{u_0}^A(\mathbf{t}) = \{u \in R_{u_0}(\mathbf{t});\ A_u \neq \emptyset\} = \{u \in R_{u_0}(\mathbf{t});\ \exists v \in A \text{ such that } u \in \mathrm{An}(v)\}.$$

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

$$k_\emptyset(\mathbf{t}_A) = \mathrm{Card}(R_{u_0}^A(\mathbf{t})).$$

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

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

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

$$\mathrm{Card}(\mathbf{t}_A) = \mathrm{Card}(A). \tag{4.6}$$

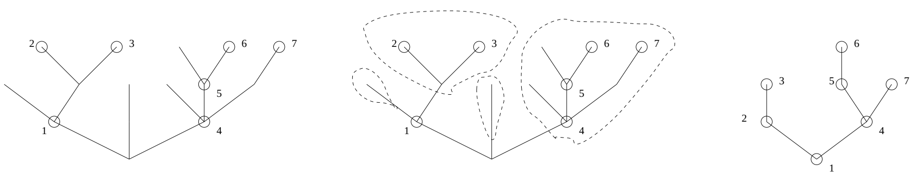

FIGURE 1: Left: A tree $\mathbf{t}$ and the set A. Center: The fringe subtrees rooted at the vertices in $R_{u_0}(\mathbf{t})$. Right: the tree $\mathbf{t}_A$. The labels have no signification, they only show which node of $\mathbf{t}$ corresponds to a node of $\mathbf{t}_A$

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 (2.3) and mark function q. Recall $\gamma = \mathbb{P}(M(\tau) > 0) = \mathbb{P}(\mathcal{M}(\tau) \neq \emptyset)$.

Let $((X_i, Z_i), i \in \mathbb{N}^*)$ be i.i.d. 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 define:

- $G = \inf\{k \in \mathbb{N}^*;\ \sum_{i=1}^k (X_i - 1) = -1\}$.
- $N = \inf\{k \in \mathbb{N}^*;\ Z_k = 1\}$.
- $\tilde{X}$ 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 (2.3).

Lemma 4.2. *The distribution of Y satisfies (2.3) and if $\gamma < 1$ then it is aperiodic.*

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

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

and $\sum_{u\in\mathbf{t}, u<v} (k_u(\mathbf{t})-1) > -1$ for any $v \in \mathbf{t}$. 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 $\{\mathcal{M}(\tau) \neq \emptyset\}$. By construction, $\mathrm{Card}(\tau^0)$ is distributed as G conditioned on $\{N \leq G\}$.

Lemma 4.3. *Under the hypothesis of this section, we have 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 $\mathcal{M}_u(\tau^0) = \mathcal{M}(\tau^0) \bigcap F_u(\tau^0)$. Define $\tilde{\mathcal{M}}_u(\tau^0) = \{v;\, uv \in \mathcal{M}_u(\tau^0)\}$ the corresponding marked vertices of $S_u(\mathbf{t})$. 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 $(\tau, \mathcal{M}(\tau))$. Notice that for $u \in R_{u_0}$, the fringe subtree $F_u(\tau^0)$ contains at least one mark iff u belongs to

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

Then by considering only the fringe subtrees containing at least one mark, we get that, conditionally on $R^{\mathcal{M}(\tau^0)}_{u_0}$, the subtrees $((S_u(\tau^0), \tilde{\mathcal{M}}_u(\tau^0)), u \in R^{\mathcal{M}(\tau^0)}_{u_0})$ 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. Notice that the offspring distribution of $\tilde{\tau}$ is given by the distribution of the cardinal of $R^{\mathcal{M}(\tau^0)}_{u_0}$. We now compute the corresponding offspring distribution. We first give an elementary formula for the cardinal of $R_u(\mathbf{t})$. Let $\mathbf{t} \in \mathbb{T}_0$ and $u \in \mathbf{t}$. Consider the tree $\mathbf{t}' = R_u(\mathbf{t}) \bigcup \{v \in \mathbf{t};\, v \leq u\}$. Using (4.7) for $\mathbf{t}'$, we get:

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

This gives $\mathrm{Card}(R_u(\mathbf{t})) = 1 + \sum_{v\in\mathbf{t};\, v\leq u} (k_v(\mathbf{t}') - 1)$. 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 $\text{Card}(R_{u_0})$, the distribution of $\text{Card}(R_{u_0}^{\mathcal{M}(\tau^0)})$ is binomial with parameter $(\text{Card}(R_{u_0}(\tau^0)), \gamma)$. This gives 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 $\mathbb{E}[Y] = 1$ that is $\gamma\mathbb{E}[\tilde{X}] = 1$ since Y is conditionally on $\tilde{X}$ binomial with parameter $(\tilde{X}, \gamma)$.

Recall 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 get $\mathbb{E}\left[\sum_{i=1}^{N}(X_i - 1)\right] = 0$ and thus using the definition of $\tilde{X}$ as well as (4.8):

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

We have:

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

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 ends the proof. □

4.4. Proof of (4.3)

According to Lemma 4.3 and (4.6), we have that $M(\tau^0)$ is distributed as the total size of a critical GW whose offspring distribution satisfies (2.3). 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 $\text{Card}(\tau') - 1$, we have

$$\lim_{n\to\infty} \frac{\mathbb{P}(\text{Card}(\tau') \in [n + 1, n + 1 + d))}{\mathbb{P}(\text{Card}(\tau') \in [n, n + d))} = 1.$$

5. Protected nodes

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

Theorem 5.1. *Let τ be a critical GW tree with offspring distribution p satisfying (2.3) 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:*

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

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

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

where $D(\mathbf{t}, x) = 1$ if x is the only child of its first ancestor which is a leaf (therefore this ancestor becomes a protected node in $\mathbf{t} \circledast_x \mathbf{t}'$) and $D(\mathbf{t}, x) = 0$ otherwise. According to Theorem 3.1 of [1], to end 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 $\mathbf{t} \neq \{\emptyset\}$, let $\mathbf{t}_{\mathbb{N}^*} = \phi(\mathbf{t}, \mathbf{t} \setminus \mathcal{L}_0(\mathbf{t}))$ be the tree obtained from $\mathbf{t}$ by removing the leaves. Let τ^0 be a random tree distributed as τ conditioned to $\{k_\emptyset(\tau) > 0\}$. Using Theorem 6 and Corollary 2 of [8] with $A = \mathbb{N}^*$ (or Lemma 4.3 with $q(k) = \mathbf{1}_{\{k>0\}}$), we have that $\tau^0_{\mathbb{N}^*}$ is a critical GW tree with offspring distribution:

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

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

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

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

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

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.

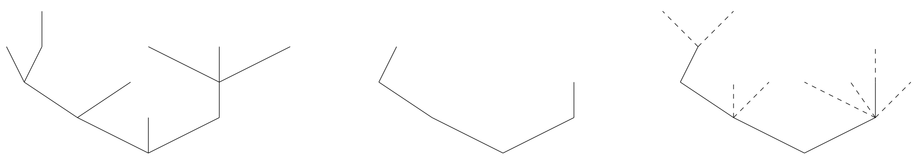

FIGURE 2: The trees τ^0, $\tau^0_{\mathbb{N}^*}$ and $\hat{\tau}$

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 $\psi : \tau^0_{\mathbb{N}^*} \longmapsto \hat{\tau}$ recursively in the following way. We set $\psi(\emptyset) = \emptyset$. Then, given $k_\emptyset(\tau^0_{\mathbb{N}^*}) = k$, we set $k_\emptyset(\hat{\tau}) = k + W(\emptyset)$. We also consider a family $(i_1, \dots, i_k)$ of integer-valued random variables such that $(i_1, i_2 - i_1, \dots, i_k - i_{k-1}, W(u) + k + 1 - i_k)$ is a uniform positive partition of $W(u) + k + 1$. Then, for every $j \leq k$ such that $j \notin \{i_1, \dots, i_k\}$, we set $k_j(\hat{\tau}) = 0$ i.e. these are leaves of $\hat{\tau}$. For every $1 \leq j \leq k$, we set $\psi(j) = i_j$ and we 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 $\mathbf{t} \in \mathbb{T}_0$. As $\mathbb{P}(\hat{\tau} = \{\emptyset\}) = 0$, we assume that $k_\emptyset(\mathbf{t}) > 0$. Let $\mathbf{t}_{\mathbb{N}^*}$ be the tree obtained from $\mathbf{t}$ by removing the leaves. Using (4.7), we have:

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

□

Notice 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 get that the corresponding mark function q is given by:

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

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, we deduce from Lemma 4.1 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 $\mathbb{P}(A(\tau) = n) = \mathbb{P}(A(\tau) = n | k_\emptyset(\tau) > 0)\mathbb{P}(k_\emptyset(\tau) > 0)$ and $\mathbb{P}(A(\tau) = n | k_\emptyset(\tau) > 0) = \mathbb{P}(A(\tau^0) = n)$ for $n \geq 2$, we obtain (5.2) and hence end the proof. □

Acknowledgement

We would like to thank the referee for his 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. *Elec. J. of Probab.* **19,** 1–19.

[2] Abraham, R. and Delmas, J.-F. (2015). An introduction to Galton-Watson trees and their local limits. arXiv:1506.05571.

[3] Devroye, L. and Janson, S. (2014). Protected nodes and fringe subtrees in some random trees. *Elec. Comm. in Probab.* **19,** 1–10.

[4] He, X. (2015). Local convergence of critical random trees and continuum condensation tree. arXiv:1503.00951.

[5] Janson, S. (2015). Asymptotic normality of fringe subtrees and additive functionals in conditioned Galton-Watson trees. *Random Struct. and Alg.* **To appear,**.

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

[7] Neveu, J. (1986). Arbres et processus de Galton-Watson. *Ann. de l'Inst. Henri Poincaré* **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. de l'Inst. Henri Poincaré* **51,** 512–532.