PROTECTED NODES AND FRINGE SUBTREES IN SOME RANDOM TREES

LUC DEVROYE AND SVANTE JANSON

Abstract. We study protected nodes in various classes of random rooted trees by putting them in the general context of fringe subtrees introduced by Aldous (1991). Several types of random trees are considered: simply generated trees (or conditioned Galton–Watson trees), which includes several cases treated separately by other authors, binary search trees and random recursive trees. This gives unified and simple proofs of several earlier results, as well as new results.

1. Introduction

Several recent papers study protected nodes in various classes of random rooted trees, where a node is said to be *protected* if it is not a leaf and, furthermore, none of its children is a leaf. (Equivalently, a node is protected if and only if the distance to any descendant that is a leaf is at least 2; for generalizations, see Section 5.) See Cheon and Shapiro [5] (uniformly random ordered trees, Motzkin trees, full binary trees, binary trees, full ternary trees), Mansour [17] (k-ary trees), Du and Prodinger [10] (digital search trees), Mahmoud and Ward [15] (binary search trees), Mahmoud and Ward [16] (random recursive trees), Bóna [4] (binary search trees).

The purpose of the present paper is to extend and sharpen some of these results by putting them in the general context of *fringe subtrees* introduced by Aldous [1].

If T is any rooted tree, and v is a node in T, let T_v be the subtree rooted at v. By taking v uniformly at random from the nodes of T, we obtain a random rooted tree which we call the *random fringe subtree* of T and denote by T_*.

Note that a node v is protected if and only if the subtree T_v has a protected root. Hence, if $\mathcal{E}_{\mathsf{p}}$ is the set of trees that have a protected root, then v is protected in T if and only if $T_v \in \mathcal{E}_{\mathsf{p}}$. In particular, taking v uniformly at random, for any given tree T,

$$p_{\mathsf{p}}(T) := \mathbb{P}(\text{a uniformly random node } v \text{ is protected}) = \mathbb{P}(T_* \in \mathcal{E}_{\mathsf{p}}). \quad (1.1)$$

and we immediately obtain results for protected nodes from more general results for fringe subtrees, see Section 3.

Date: 2 October, 2013.

2010 *Mathematics Subject Classification.* 60C05; 05C05.

SJ partly supported by the Knut and Alice Wallenberg Foundation.

When T is a random tree, we can think of T_* in two ways, called *annealed* and *quenched* using terminology from statistical physics. In the annealed version we take a random tree T and a uniformly random node v in it, yielding a random fringe subtree T_*.

In the quenched version we do the random choices in two steps. First we choose a random tree T. We then fix T and choose $v \in T$ uniformly at random, yielding a random fringe subtree T_* depending on T. We thus obtain for every choice of T a probability distribution $\mathcal{L}(T_*)$ on the set $\mathfrak{T}$ of all rooted trees; this distribution depends on the random tree T and is thus a random probability distribution. In other words, we consider the conditional distribution $\mathcal{L}(T_* \mid T)$ of T_* given T. We can now study properties of this random probability distribution. Averaging over T, we obtain the distribution of T_* in the annealed version, so results in the quenched version are generally stronger than in the annealed version.

Returning to protected nodes, we see that in the quenched point of view, we consider $n_{\mathsf{p}}(T)$, the number of protected nodes in a tree T, and $p_{\mathsf{p}}(T) = n_{\mathsf{p}}(T)/|T|$, the probability that a randomly chosen node in T is protected, and we regard these functions of T as random variables depending on a random tree T. Thus (1.1) can now be written

$$p_{\mathsf{p}}(T) = \mathbb{P}(T_* \in \mathcal{E}_{\mathsf{p}} \mid T). \tag{1.2}$$

In the annealed version we more simply consider the probability that a random node in a random tree T is protected, which equals the expectation

$$\mathbb{E}\, p_{\mathsf{p}}(T) = \mathbb{P}(T_* \in \mathcal{E}_{\mathsf{p}}). \tag{1.3}$$

The first class of random trees that we consider in this paper are the simply generated random trees; these are defined using a weight sequence $(w_k)_{k=0}^{\infty}$ which we regard as fixed, see Section 2 for the definition and the connection to conditioned Galton–Watson trees. It is well-known that suitable choices of $(w_k)_{k=0}^{\infty}$ yield several important classes of random trees, see e.g. Aldous [2], Devroye [6], Drmota [9], Janson [13] and Section 4.

Let

$$\Phi(t) := \sum_{k=0}^{\infty} w_k t^k \tag{1.4}$$

be the generating function of the weight sequence, and let $\rho \in [0, \infty]$ be its radius of convergence. We define an important parameter $\tau \geqslant 0$ by:

(i) τ is the unique number in $[0, \rho]$ such that

$$\tau \Phi'(\tau) = \Phi(\tau) < \infty, \tag{1.5}$$

if there exists any such τ.

(ii) If (1.5) has no solution, then $\tau := \rho$.

See further [13, Section 7], where several properties and equivalent characterizations are given. (For example, τ is the minimum point in $[0, \rho]$ of $\Phi(t)/t$. Furthermore, $\Phi(\tau) < \infty$ also in case (ii), and $\tau > 0 \iff \rho > 0$.)

We define another weight sequence $(\pi_k)_{k=0}^\infty$ by

$$\pi_k := \frac{w_k\tau^k}{\Phi(\tau)}; \tag{1.6}$$

this weight sequence has the generating function

$$\Phi_\tau(t) := \Phi(\tau t)/\Phi(\tau). \tag{1.7}$$

Note that $\sum_{k=0}^\infty \pi_k = 1$; thus $(\pi_k)_{k=0}^\infty$ is a probability distribution on the non-negative integers $\mathbb{Z}_{\geqslant 0} := \{0, 1, 2, \dots\}$.

Theorem 1.1. *Let $\mathcal{T}_n$ be a simply generated random tree with n nodes. Then, with notations as above, the following holds as $n \to \infty$.*

(i) *(Annealed version.) The probability $p_n = \mathbb{E}\, p_{\mathsf{p}}(\mathcal{T}_n)$ that a random node in a random tree $\mathcal{T}_n$ is protected tends to a limit p_* as $n \to \infty$, with*

$$p_* := \Phi_\tau(1-\pi_0) - \pi_0 = \frac{\Phi\big(\tau - \tau w_0/\Phi(\tau)\big) - w_0}{\Phi(\tau)}. \tag{1.8}$$

(ii) *(Quenched version.) The proportion of nodes in $\mathcal{T}_n$ that are protected, i.e. $p_{\mathsf{p}}(\mathcal{T}_n) = n_{\mathsf{p}}(\mathcal{T}_n)/n$, converges in probability to p_* as $n \to \infty$.*

The main idea of this paper, viz. to study protected nodes by studying fringe subtrees, applies also to other types of random trees. We consider binary search trees in Section 6 and random recusive trees in Section 7.

Protected nodes have been studied also for digital search trees [10] and tries [11]. As far as we know, the fringe subtrees of these random trees have not been studied in general; this will be dealt with elsewhere.

Acknowledgement. This research was mainly done during the 23rd International Meeting on Probabilistic, Combinatorial and Asymptotic Methods for the Analysis of Algorithms (AofA 2012) in Montreal, June 2012. We thank the organizers for providing this opportunity and several participants for helpful comments.

2. Simply generated trees and Galton–Watson trees

All trees in this paper are *rooted* and *ordered* (= *plane*). (For unordered trees, see Example 4.2.) We denote the outdegree of a node $v \in T$ by $d^+(v)$. Note that a tree is uniquely determined by its sequence of outdegrees, taken in e.g. breadth-first order. See further e.g. [9] and [13]. We let $\mathfrak{T}$ denote the set of all ordered rooted trees, and $\mathfrak{T}_n := \{T \in \mathfrak{T} : |T| = n\}$ the set of all ordered rooted trees with with n nodes. By a *random tree* we mean a random element of $\mathfrak{T}$ with some given but arbitrary distribution. (No uniformity is implied unless we say so.)

Given a weight sequence $(w_k)_{k=0}^\infty$, we define the weight of a tree T to be $w(T) := \prod_{v\in T} w_{d^+(v)}$. For $n \geqslant 1$, we define the simply generated random tree $\mathcal{T}_n$ as the random tree obtained by selecting an ordered rooted tree in $\mathfrak{T}_n$ with probability proportional to its weight. (We consider only n such that there is at least one tree in $\mathfrak{T}_n$ with positive weight.)

It is well-known that simply generated random trees are essentially the same as conditioned Galton–Watson trees. Given a probability distribution $(\pi_k)_{k=0}^\infty$ on $\mathbb{Z}_{\geqslant 0}$, let $\mathcal{T}$ be the corresponding *Galton–Watson tree*; this is a random tree where each node has a random number of children, and these numbers all are independent and with the distribution $(\pi_k)_{k=0}^\infty$. Furthermore, let $\mathcal{T}_n$ be $\mathcal{T}$ conditioned on having exactly n nodes; this is called a *conditioned Galton–Watson tree.* (We consider only n such that $\mathbb{P}(|\mathcal{T}| = n) > 0$.) It is easy to see that the conditioned Galton–Watson tree $\mathcal{T}_n$ coincides with the simply generated random tree defined using the weight sequence $(\pi_k)_{k=0}^\infty$. Moreover, if $(w_k)_{k=0}^\infty$ is any weight sequence with radius of convergence $\rho > 0$ (this is satisfied in virtually all applications), let $(\pi_k)_{k=0}^\infty$ be given by (1.6). Then the simply generated random tree defined by $(w_k)_{k=0}^\infty$ coincides with the simply generated random tree defined by $(\pi_k)_{k=0}^\infty$, and thus with the conditioned Galton–Watson tree defined by $(\pi_k)_{k=0}^\infty$, see e.g. [14] and [13]. (There are also other probability distributions yielding the same conditioned Galton–Watson tree, but the choice in (1.6) is the canonical one, see [13].)

It is easy to see that the probability distribution $(\pi_k)_{k=0}^\infty$ has expectation $\tau\Phi'(\tau)/\Phi(\tau)$, which equals 1 in case (i) above (i.e., when (1.5) holds), but is less than 1 in case (ii) (i.e., when (1.5) has no solution). Thus, $(\pi_k)_{k=0}^\infty$ yields a critical Galton–Watson tree $\mathcal{T}$ in case (i), but $\mathcal{T}$ is subcritical in case (ii). In both cases, $\mathcal{T}$ is a.s. finite.

3. Proof of Theorem 1.1

The proof is based on the fact that the random fringe subtrees of a conditioned Galton–Watson tree converge in distribution to the corresponding (unconditional) Galton–Watson tree, as stated in the following theorem. Part (i) was proved by Aldous [1] under some extra conditions, and by Bennies and Kersting [3] under fewer extra conditions; the general case and (ii) are proved in [13, Theorem 7.12].

Theorem 3.1. *Let $\mathcal{T}_n$ be a simply generated random tree with n nodes. Then, with notations as above, the following holds as $n \to \infty$.*

(i) *(Annealed version.) The fringe subtree $\mathcal{T}_{n,*}$ converges in distribution to the Galton–Watson tree $\mathcal{T}$. I.e., for every fixed tree T,*

$$\mathbb{P}(\mathcal{T}_{n,*} = T) \to \mathbb{P}(\mathcal{T} = T). \tag{3.1}$$

(ii) *(Quenched version.) The conditional distributions $\mathcal{L}(\mathcal{T}_{n,*} \mid \mathcal{T}_n)$ converge to the distribution of $\mathcal{T}$ in probability. I.e., for every fixed tree T,*

$$\mathbb{P}(\mathcal{T}_{n,*} = T \mid \mathcal{T}_n) \xrightarrow{\text{p}} \mathbb{P}(\mathcal{T} = T). \tag{3.2}$$

□

Note that the set of (finite) ordered trees is a countable discrete set; this justifies that it is enough to consider point probabilities in (3.1) and (3.2).

Proof of Theorem 1.1. For the annealed version, it follows immediately from (1.1) and (3.1), which can be written $\mathcal{T}_{n,*} \xrightarrow{\text{d}} \mathcal{T}$, that

$$p_n = \mathbb{P}(\mathcal{T}_{n,*} \in \mathcal{E}_{\mathsf{p}}) \to \mathbb{P}(\mathcal{T} \in \mathcal{E}_{\mathsf{p}}). \tag{3.3}$$

For the quenched version, conditioning on $\mathcal{T}_n$, we similarly obtain by (3.2),

$$p_{\mathsf{p}}(\mathcal{T}_n) = \mathbb{P}(\mathcal{T}_{n,*} \in \mathcal{E}_{\mathsf{p}} \mid \mathcal{T}_n) \xrightarrow{\text{p}} \mathbb{P}(\mathcal{T} \in \mathcal{E}_{\mathsf{p}}). \tag{3.4}$$

It remains only to calculate $\mathbb{P}(\mathcal{T} \in \mathcal{E}_{\mathsf{p}})$. This is easy, by conditioning on the root degree, k say. If $k = 0$, then the root is a leaf and not protected, and if $k > 0$, the root is protected if and only if each of its k children has at least one child, which has probability $(1 - \pi_0)^k$. Hence,

$$\mathbb{P}(\mathcal{T} \in \mathcal{E}_{\mathsf{p}}) = \sum_{k=1}^{\infty} \pi_k (1 - \pi_0)^k = \Phi_\tau(1 - \pi_0) - \pi_0. \tag{3.5}$$

Finally, $\pi_0 = w_0/\Phi(\tau)$ by (1.6), and $\Phi_\tau(1 - \pi_0) = \Phi(\tau - \tau\pi_0)/\Phi(\tau)$ by (1.7). □

4. Examples

We give several examples of random trees where Theorem 1.1 applies. We focus on the calculation of p_*, since the other conclusions are the same for all random trees considered here. We omit some steps in the calculations, see e.g. [13, Section 10] for further details.

Example 4.1 (ordered trees)**.** The weight sequence $w_k = 1$ yields uniformly random *ordered trees*. In this case, $\Phi(t) = \sum_{k=0}^{\infty} t^k = 1/(1-t)$ and (1.5) has the solution $\tau = 1/2$, yielding $\pi_k = 2^{-k-1}$ (a geometric Ge(1/2) distribution) and $\Phi_\tau(t) = 1/(2-t)$. Thus $\pi_0 = 1/2$ and, by (1.8),

$$p_* = \Phi_\tau\Big(\frac{1}{2}\Big) - \frac{1}{2} = \frac{1}{2 - \frac{1}{2}} - \frac{1}{2} = \frac{1}{6}. \tag{4.1}$$

We thus recover from the annealed version in Theorem 1.1 the result by Cheon and Shapiro [5] that the average proportion of protected nodes in a random ordered tree converges to 1/6 as the size goes to infinity. Moreover, the quenched version shows that holds also for most individual trees. More precisely, $p_{\mathsf{p}}(\mathcal{T}_n) \xrightarrow{\text{p}} 1/6$, i.e., for any $\varepsilon > 0$, the probability that a uniformly random ordered tree with n nodes has between $(1/6 - \varepsilon)n$ and $(1/6 + \varepsilon)n$ protected nodes tends to 1 as $n \to \infty$.

Example 4.2 (unordered trees)**.** We have assumed that the trees are ordered, but we can treat also unordered labelled trees by giving the children of each node a (uniform) random ordering. As is well known, a uniformly random *unordered labelled rooted tree* (sometimes called *Cayley tree*) then becomes simply generated with weights $w_k = 1/k!$. In this case, $\Phi(t) = \sum_{k=0}^{\infty} t^k/k! = e^t$ and (1.5) has the solution $\tau = 1$, yielding

$\pi_k = e^{-1}/k!$ (a Poisson Po(1) distribution) and $\Phi_\tau(t) = e^{t-1}$. Thus $\pi_0 = e^{-1}$ and, by (1.8),

$$p_* = \Phi_\tau\bigl(1 - e^{-1}\bigr) - e^{-1} = e^{-e^{-1}} - e^{-1} \approx 0.32432. \tag{4.2}$$

Example 4.3 (full d-ary trees)**.** Uniformly random *full d-ary trees* are simply generated random trees with $w_k = 1$ if $k = 0$ or $k = d$, and $w_k = 0$ otherwise. (Here $d \geqslant 2$ is a fixed integer. In this case, the number of nodes n has to be $1 \pmod d$.) We have $\Phi(t) = 1 + t^d$ and $\tau = (d-1)^{-1/d}$, yielding $\pi_0 = (d-1)/d$, $\pi_d = 1/d$, and $\Phi_\tau(t) = (d - 1 + t^d)/d$. Consequently, (1.8) yields

$$p_* = \pi_d(1 - \pi_0)^d = 1/d^{d+1}. \tag{4.3}$$

Thus, Theorem 1.1 shows that the proportion of protected nodes tends to $1/d^{d+1}$.

This was found by Mansour [17] (for the annealed version); note that [17] states the result in terms of the number of internal nodes. Since a full d-ary tree with m internal nodes has $dm + 1$ nodes, the proportion of internal nodes that are protected tends to $1/d^d$.

The special case $d = 2$ yields full binary trees, for which we find $p_* = 1/8$. (The proportion $1/4$ given in [5] seems to be a mistake.)

The special case $d = 3$ yields full ternary trees, for which we find $p_* = 1/81$, in accordance with [5].

Example 4.4 (d-ary trees)**.** Uniformly random *d-ary trees* are simply generated random trees with $w_k = \binom{d}{k}$. (Again, $d \geqslant 2$ is a fixed integer.) In this case, $\Phi(t) = (1 + t)^d$ and $\tau = 1/(d - 1)$, yielding $\pi_k = \binom{d}{k}(\frac{1}{d})^k(\frac{d-1}{d})^{d-k}$ (a binomial $\mathrm{Bi}(d, 1/d)$ distribution) and $\Phi_\tau(t) = ((d-1+t)/d)^d$. Consequently, $\pi_0 = (1 - 1/d)^d$ and

$$p_* = \left(\frac{d - \pi_0}{d}\right)^d - \pi_0^d = \left(1 - \frac{(d-1)^d}{d^{d+1}}\right)^d - \frac{(d-1)^d}{d^d}. \tag{4.4}$$

In particular, for $d = 2$ (binary trees), we obtain $p_* = 33/64$. (The proportion $9/256$ given in [5] seems to be a mistake.)

Example 4.5 (Motzkin trees)**.** A *Motzkin tree* has each outdegree 0,1 or 2. Taking $w_0 = w_1 = w_2 = 1$ and $w_k = 0$ for $k \geqslant 3$ yields a uniformly random Motzkin tree. We have $\Phi(t) = 1 + t + t^2$ and (1.5) has the soultion $\tau = 1$, yielding $\pi_k = 1/3$, $0 \leqslant k \leqslant 2$, and $\Phi_\tau(t) = (1 + t + t^2)/3$. Thus, by (1.8),

$$p_* = \frac{1}{3}\left(\frac{2}{3} + \left(\frac{2}{3}\right)^2\right) = \frac{10}{27}. \tag{4.5}$$

Hence, the proportion of protected nodes in a uniformly random Motzkin tree tends to $10/27$, as shown (in the annealed version) by Cheon and Shapiro [5].

5. ℓ-PROTECTED NODES

More generally, given an integer $\ell \geqslant 1$, we say that a node in a rooted tree is *ℓ-protected* if it has distance at least ℓ to every leaf that is a descendant of it. Thus 2-protected = protected and 1-protected = non-leaf (internal node).

The results above generalize immediately to ℓ-protected nodes for any fixed $\ell \geqslant 1$. Given a tree T, let $p_{\mathsf{p},\ell}(T)$ be the proportion of nodes in T that are ℓ-protected, and let $p_{*,\ell}$ be the probability that the root of the Galton–Watson tree $\mathcal{T}$ is ℓ-protected.

Theorem 5.1. *Let $\mathcal{T}_n$ be a simply generated random tree with n nodes. Then, with notations as above, the following holds as $n \to \infty$.*

(i) *(Annealed version.) The probability $p_{n,\ell} = \mathbb{E}\, p_{\mathsf{p},\ell}(\mathcal{T}_n)$ that a random node in a random tree $\mathcal{T}_n$ is ℓ-protected tends to $p_{*,\ell}$ as $n \to \infty$, with $p_{*,\ell}$ given by the recursion*

$$p_{*,\ell} := \Phi_{\mathcal{T}}(p_{*,\ell-1}) - \pi_0, \qquad \ell \geqslant 1, \tag{5.1}$$

with $p_{*,0} = 1$ and $p_{*,1} = 1 - \pi_0$.

(ii) *(Quenched version.) The proportion of nodes in $\mathcal{T}_n$ that are ℓ-protected, i.e. $p_{\mathsf{p},\ell}(\mathcal{T}_n)$, converges in probability to $p_{*,\ell}$ as $n \to \infty$.*

Proof. The convergence to $p_{*,\ell}$ follows from Theorem 3.1 as in the proof of Theorem 1.1. The recursion (5.1) follows since the root is ℓ-protected if and only if it has outdegree > 0 and each child is $(\ell-1)$-protected. □

Example 5.2. For uniformly random ordered trees, $\Phi_{\mathcal{T}}(t) = 1/(2-t)$, see Example 4.1, and thus the recursion (5.1) is

$$p_{*,\ell} = \frac{1}{2 - p_{*,\ell-1}} - \frac{1}{2} = \frac{p_{*,\ell-1}}{4 - 2p_{*,\ell-1}}, \qquad \ell \geqslant 1, \tag{5.2}$$

yielding $1/p_{*,\ell} = 4/p_{*,\ell-1} - 2$ with the solution $1/p_{*,\ell} = (4^\ell + 2)/3$, i.e.

$$p_{*,\ell} = \frac{3}{4^\ell + 2}, \qquad \ell \geqslant 0. \tag{5.3}$$

In particular, $p_{*,1} = 1/2$, $p_{*,2} = 1/6$, $p_{*,3} = 1/22$, $p_{*,4} = 1/86$.

Hence, for each fixed $\ell \geqslant 1$, the proportion of ℓ-protected nodes in a uniform random ordered tree tends to $3/(4^\ell + 2)$.

Example 5.3. For uniformly random unordered labelled rooted trees we have by Example 4.2 $\pi_0 = e^{-1}$ and $\Phi_{\mathcal{T}}(t) = e^{t-1}$. Thus (5.1) yields

$$p_{*,1} = 1 - e^{-1} \approx 0.63212, \tag{5.4}$$

$$p_{*,2} = e^{-e^{-1}} - e^{-1} \approx 0.32432, \tag{5.5}$$

as in Example 4.2, and

$$p_{*,3} = \exp\left(e^{-e^{-1}} - e^{-1} - 1\right) - e^{-1} \approx 0.14093. \tag{5.6}$$

6. Binary search trees

A random binary search tree with n nodes is a binary tree obtained by inserting, in the standard manner, n independently and identically distributed (i.i.d.) uniform $[0,1]$ random variables $X_1, \dots, X_n$ into an initially empty tree, see e.g. [9]. Let $\mathcal{T}_n$ be a random binary search tree with n nodes. Aldous [1] showed that there exists a random limiting fringe tree $\hat{T}$ in this case too such that (3.1) and (3.2) hold (with $\mathcal{T}$ replaced by $\hat{T}$); in fact, the convergence in (3.2) holds a.s. The limit tree $\hat{T}$ can be described as a binary search tree $\mathcal{T}_N$ with a random size N; this is easily seen by the recursive construction of the binary search tree, letting N be the limiting distribution of the subtree size $|\mathcal{T}_{n,*}|$ of a random node, and a calculation shows that $\mathbb{P}(N = n) = 2/(n+1)(n+2)$, $n \geqslant 1$ [1]. See also Devroye [7] for a simple direct proof.

Moreover, Aldous [1] also shows that $\hat{T}$ may be constructed as follows: Let $\tilde{T}_t$, $t \geqslant 0$, be a random process of a binary tree growing in continuous time, starting with $\tilde{T}_0$ being a single root, and adding left and right children with intensity 1 at all possible places. In other words, given any $\tilde{T}_t$ at a time $t \geqslant 0$, any possible child of an existing node (excluding children already existing) is added after an exponential Exp(1) waiting time; all waiting times being independent. It is well-known and easy to see that at any fixed time $t > 0$, the conditional distribution of $\tilde{T}_t$ given $|\tilde{T}_t| = n$ equals the distribution of $\mathcal{T}_n$. Moreover, if we instead take $\tilde{T}_X$ at a random time $X \sim \mathrm{Exp}(1)$ (independent of everything else), then $\tilde{T}_X \overset{\mathrm{d}}{=} \hat{T}$.

We can now repeat the proof of Theorem 1.1 and obtain the same results as above, with

$$p_* = \mathbb{P}(\text{the root of } \hat{T} \text{ is protected}) = \mathbb{P}(\tilde{T}_X \in \mathcal{E}_{\mathsf{p}}) = \int_0^\infty \mathbb{P}(\tilde{T}_t \in \mathcal{E}_{\mathsf{p}}) e^{-t} \, \mathrm{d}t \tag{6.1}$$

In order to evaluate p_*, we consider first $\tilde{T}_t$ for a given t. The probability, $q_1(t)$ say, that the root of $\tilde{T}_t$ is a leaf is e^{-2t}. Similarly, if the left child appears at time s, then the probability that it still is a leaf at some later time $t > s$ is $e^{-2(t-s)}$. Hence, the probability, $r_1(t)$ say, that there is a left child that is a leaf is

$$r_1(t) := \int_0^t e^{-2(t-s)} e^{-s} \, \mathrm{d}s = \int_0^t e^{s-2t} \, \mathrm{d}s = e^{-t} - e^{-2t}. \tag{6.2}$$

The probability that the root has at least one child that is a leaf is thus, by symmetry and independence, $1 - (1 - r_1(t))^2 = 2r_1(t) - r_1(t)^2$ and the probability that the root in $\tilde{T}_t$ is not protected is

$$\begin{aligned} q_1(t) + 2r_1(t) - r_1(t)^2 &= e^{-2t} + 2e^{-t} - 2e^{-2t} - (e^{-t} - e^{-2t})^2 \\ &= 2e^{-t} - 2e^{-2t} + 2e^{-3t} - e^{-4t}. \end{aligned} \tag{6.3}$$

Hence we obtain from (6.1)

$$p_* = 1 - \int_0^\infty \bigl(2e^{-t} - 2e^{-2t} + 2e^{-3t} - e^{-4t}\bigr)e^{-t}\,\mathrm{d}t = \frac{11}{30}, \tag{6.4}$$

in accordance with Mahmoud and Ward [15] and Bóna [4].

More generally, let $q_\ell(t)$ be the probability that the root of $\tilde{T}_t$ is *not* ℓ-protected, $\ell \geqslant 1$, and let $r_\ell(t)$ be the probability that the root in $\tilde{T}_t$ has a left child that is not ℓ-protected. The same argument as above yields the recursion, for $\ell \geqslant 2$,

$$q_\ell(t) = q_1(t) + 2r_{\ell-1}(t) - r_{\ell-1}(t)^2, \tag{6.5}$$

$$r_{\ell-1}(t) = \int_0^t q_{\ell-1}(t-s)e^{-s}\,\mathrm{d}s = e^{-t}\int_0^t q_{\ell-1}(s)e^{s}\,\mathrm{d}s, \tag{6.6}$$

and then the asymptotic proportion of ℓ-protected nodes is found as

$$p_{*,\ell} = 1 - \int_0^\infty q_\ell(t)e^{-t}\,\mathrm{d}t. \tag{6.7}$$

A Maple calculation yields $p_{*,1} = 2/3$, $p_{*,2} = 11/30$, $p_{*,3} = 1249/8100$, $p_{*,4} = 103365591157608217/2294809143026400000 \approx 0.04504$, in agreement with Bóna [4], who calculates these values by a different method.

Remark 6.1. Bóna [4] considers c_ℓ, the asymptotic probability that a random node is at *level* ℓ, meaning that the distance to the nearest leaf that is a descendant is $\ell - 1$; thus a node is ℓ-protected if it is at a level strictly larger than ℓ, and $c_\ell = p_{*,\ell-1} - p_{*,\ell}$, with $p_{*,0} = 1$.

In the quenched version, asymptotic normality of the number of protected nodes was shown by Mahmoud and Ward [15]. Alternatively, this follows easily by the method of Devroye [7], see [12] for details.

7. Random recursive trees

A uniform random recursive tree (URRT) $\mathcal{T}_n$ of order n is a tree with n nodes labeled $\{1, \dots, n\}$. The root is labelled 1, and for $2 \leqslant i \leqslant n$, the node labelled i chooses a vertex in $\{1, \dots, i-1\}$ uniformly at random as its parent. See e.g. [8], [9], [18]. This case is very similar to the random binary search tree in Section 6: Aldous [1] has shown the existence of a random limiting fringe tree $\hat{T}$, and again $\hat{T}$ can be described as $\mathcal{T}_N$, now with $\mathbb{P}(N = n) = 1/n(n+1)$. Moreover, $\hat{T}$ can be constructed as $\tilde{T}_X$ with $X \sim \mathrm{Exp}(1)$ in this case too, where now $\tilde{T}_t$ is the random tree process where each node gets a new child with i.i.d. exponential waiting times with intensity 1. (The Yule tree process.)

The children of the root arrive in a Poisson process with intensity 1; hence the number of children of the root in $\tilde{T}_t$ has the distribution $\mathrm{Po}(t)$, and the probability that the root is a leaf is $\mathbb{P}(\mathrm{Po}(t) = 0) = e^{-t}$. Moreover, a child that is born at time s is still a leaf at time $t > s$ with probability $e^{-(t-s)}$. Hence children of the root that remain leaves at time t are born with intensity

$e^{-(t-s)}$, $s \in (0,t)$, and since a thinning of a Poisson process is a Poisson process, it follows that the number of children of the root that are leaves at time t has a Poisson distribution with expectation $\int_0^t e^{-(t-s)}\,\mathrm{d}s = 1 - e^{-t}$. Consequently, the probability that the root of $\tilde{T}_t$ has no child that is a leaf is $\exp(-(1-e^{-t}))$. Subtracting the probability that the root has no child at all, we obtain the probability $p_2(t)$ that the root of $\tilde{T}_t$ is protected as

$$p_2(t) = \exp\bigl(e^{-t} - 1\bigr) - e^{-t} \tag{7.1}$$

and thus

$$\begin{aligned} p_* &= \int_0^\infty p_2(t)e^{-t}\,\mathrm{d}t = \int_0^\infty \exp\bigl(e^{-t}-1\bigr)e^{-t}\,\mathrm{d}t - \int_0^\infty e^{-2t}\,\mathrm{d}t \\ &= \int_0^1 \exp(x-1)\,\mathrm{d}x - \frac{1}{2} = \frac{1}{2} - e^{-1}, \end{aligned} \tag{7.2}$$

in accordance with Mahmoud and Ward [16].

We can treat ℓ-protected nodes too in random recursive trees by the same method. If $p_\ell(t)$ is the probability that the root is ℓ-protected in $\tilde{T}_t$, and $q_\ell(t) = 1 - p_\ell(t)$, then the number of children of the root that are not $(\ell-1)$-protected at time t is Poisson distributed with mean $\int_0^t q_{\ell-1}(t-s)\,\mathrm{d}s = \int_0^t q_{\ell-1}(s)\,\mathrm{d}s$, yielding the recursion, for $\ell \geqslant 1$,

$$p_\ell(t) = \exp\left(-\int_0^t q_{\ell-1}(s)\,\mathrm{d}s\right) - \exp(-t) = e^{-t}\left(\exp\left(\int_0^t p_{\ell-1}(s)\,\mathrm{d}s\right) - 1\right), \tag{7.3}$$

with $p_0(t) = 1$ and $p_1(t) = 1 - e^{-t}$. In principle, $p_{*,\ell}$ can be computed as $\int_0^\infty p_\ell(t)e^{-t}\,\mathrm{d}t$, but in this case we do not know any closed form for $\ell > 2$.

References

[1] David Aldous, Asymptotic fringe distributions for general families of random trees. *Ann. Appl. Probab.* **1** (1991), no. 2, 228–266.

[2] David Aldous, The continuum random tree II: an overview. *Stochastic Analysis (Durham, 1990)*, 23–70, London Math. Soc. Lecture Note Ser. 167, Cambridge Univ. Press, Cambridge, 1991.

[3] Jürgen Bennies and Götz Kersting, A random walk approach to Galton–Watson trees. *J. Theoret. Probab.* **13** (2000), no. 3, 777–803.

[4] Miklós Bóna, k-protected vertices in binary search trees. *Adv. Appl. Math.*, to appear. `arXiv:1304.6105.`

[5] Gi-Sang Cheon and Louis W. Shapiro, Protected points in ordered trees. *Applied Mathematics Letters* **21** (2008), no. 5, 516–520.

[6] Luc Devroye, Branching processes and their applications in the analysis of tree structures and tree algorithms. *Probabilistic Methods for Algorithmic Discrete Mathematics*, eds. M. Habib, C. McDiarmid, J. Ramirez and B. Reed, Springer, Berlin, 1998, pp. 249–314.

[7] Luc Devroye, Limit laws for sums of functions of subtrees of random binary search trees. *SIAM J. Comput.* **32** (2002/03), no. 1, 152–171.

[8] Luc Devroye, Omar Fawzi and Nicolas Fraiman, Depth properties of scaled attachment random recursive trees. *Random Structures Algorithms* **41** (2012), no. 1, 66–98.

[9] Michael Drmota, *Random Trees*, Springer, Vienna, 2009.

[10] Rosena R.X. Du and Helmut Prodinger, Notes on protected nodes in digital search trees. *Applied Mathematics Letters* **25** (2012), no. 6, 1025–1028.

[11] Jeffrey Gaither, Yushi Homma, Mark Sellke and Mark Daniel Ward, On the number of 2-protected nodes in tries and suffix trees. *Proceedings, 2012 Conference on Analysis of Algorithms, AofA '12 (Montreal, 2012)*, *DMTCS Proceedings* 2012, 381–398.

[12] Cecilia Holmgren and Svante Janson, Limit laws for functions of fringe trees for binary search trees and recursive trees. In preparation.

[13] Svante Janson, Simply generated trees, conditioned Galton–Watson trees, random allocations and condensation. *Probability Surveys* **9** (2012), 103–252.

[14] Douglas P. Kennedy, The Galton–Watson process conditioned on the total progeny. *J. Appl. Probab.* **12** (1975), 800–806.

[15] Hosam H. Mahmoud and Mark Daniel Ward, Asymptotic distribution of two-protected nodes in random binary search trees. *Applied Mathematics Letters* **25** (2012), no. 12, 2218–2222.

[16] Hosam H. Mahmoud and Mark Daniel Ward, Asymptotic properties of protected nodes in random recursive trees. Preprint, 2013.

[17] Toufik Mansour, Protected points in k-ary trees. *Applied Mathematics Letters* **24** (2011), no. 4, 478–480.

[18] Robert T. Smythe and Hosam M. Mahmoud, A survey of recursive trees. *Theory Probab. Math. Statist.* **51** (1995), 1–27 (1996); translated from *Teor. Ĭmovīr. Mat. Stat.* **51** (1994), 1–29 (Ukrainian).

School of Computer Science, McGill University, 3480 University Street, Montréal, Québec, H3A 2A7, Canada

E-mail address: luc@cs.mcgill.ca

URL: http://luc.devroye.org

Department of Mathematics, Uppsala University, PO Box 480, SE-751 06 Uppsala, Sweden

E-mail address: svante.janson@math.uu.se

URL: http://www2.math.uu.se/∼svante/