AofA'10 DMTCS proc. **AM**, 2010, 129–142

The height of scaled attachment random recursive trees

Luc Devroye[1] Omar Fawzi[1] Nicolas Fraiman[2†]

[1] *School of Computer Science, McGill University, Montreal, Canada H3A 2K6.*
[2] *Department of Mathematics and Statistics, McGill University, Montreal, Canada H3A 2K6.*

We study depth properties of a general class of random recursive trees where each node n attaches to the random node $\lfloor nX_n \rfloor$ and $X_0, \ldots, X_n$ is a sequence of i.i.d. random variables taking values in $[0, 1)$. We call such trees scaled attachment random recursive trees (SARRT). We prove that the height H_n of a SARRT is asymptotically given by $H_n \sim \alpha_{\max} \log n$ where $\alpha_{\max}$ is a constant depending only on the distribution of X_0 whenever X_0 has a bounded density. This gives a new elementary proof for the height of uniform random recursive trees $H_n \sim e \log n$ that does not use branching random walks.

Keywords: Random recursive tree, Random web model, Power of choice, Renewal process, Second moment method

1 Introduction

A *uniform random recursive tree* (URRT) T_n of order n is a tree with $n + 1$ nodes labeled $\{0, 1, \ldots, n\}$ constructed as follows. The root is labelled 0, and for $1 \leq i \leq n$, the node labelled i is inserted and chooses a vertex in $\{0, \ldots, i - 1\}$ uniformly at random as its parent. The asymptotic properties of T_n have been extensively studied starting from Na and Rapoport (1970), Moon (1974) and Gastwirth (1977). For a survey, see Smythe and Mahmoud (1995) or Drmota (2009). In particular, Devroye (1988) and Mahmoud (1991) showed that the depth of node n is $D_n = (1 + o(1)) \log n$ and Devroye (1987) and Pittel (1994) showed that the height $H_n = \max_{1 \leq i \leq n} D_i = (e + o(1)) \log n$ with probability going to 1.

A natural generalization of this model introduced by Devroye and Lu (1995) is to let a vertex choose $k > 1$ parents uniformly. This construction defines a random directed acyclic graph (k-DAG), which was used to model circuits (Tsukiji and Xhafa, 1996; Arya et al., 1999).

The uniformity condition was relaxed by Szymański (1987) by allowing the probability of a node to be chosen as a parent to depend on its degree. When the probability of linking to a node is proportional to its degree, this gives a random plane-oriented recursive tree whose typical depth was studied by Mahmoud (1992) and height by Pittel (1994). When $k > 1$ parents are chosen for each node, the popular preferential attachment model of Barabasi and Albert (1999) is obtained.

Motivated by recent work on distances in random k-DAGs (Devroye and Janson, 2009) and the power of choice in the construction of random trees (D'Souza et al., 2007; Mahmoud, 2009), we introduce a

† Research supported by an NSERC Discovery Grant program.

generalization of uniform random recursive trees. In a *scaled attachment random recursive tree* (SARRT), a node i chooses its parent to be the node labelled $\lfloor iX_i \rfloor$ where $X_0, X_1, \ldots, X_n$ is a sequence of independent random variables distributed as $X \in [0,1)$. Note that the choice of the parent here only depends on the labels of previous nodes and not on their properties relative to the tree (like the degree, for example). In particular, if X is uniform on $[0,1)$ we get a URRT. The distribution $\mathcal{L}(X)$ of X is called the attachment distribution.

We study properties of the depth of nodes in a SARRT with a general attachment distribution. We determine the first-order asymptotics for the depth D_n of the node with label n and the height $H_n = \max_{1 \le i \le n} D_i$. Our result gives a new elementary proof for the height of a URRT that is not based on branching random walks that were used in previous proofs by Devroye (1987) and Pittel (1994).

Furthermore, setting $X = \max(U_1, \ldots, U_k)$ where $U_1, \ldots, U_k$ are independent uniforms, the depth D_i of node i in a SARRT with attachment X is the distance given by following the oldest parent from node i to the root in a random k-DAG (Devroye and Janson, 2009; Mahmoud, 2009). This problem can be seen as a "power of choice" question: how much can one optimize properties of the tree when each node is given k choices of parents? A new node is given k choices of parents, and it selects the best one according to some criterion. In the setting of this paper, we study selection criteria that only depend on the labels or arrival times of the potential parents. Our results describe the influence of such selection criteria on the depth of the last inserted node and the height of the tree.

More precisely, in Section 2, we prove concentration and a central limit theorem for D_n:

$$\frac{D_n}{\log n} \stackrel{\mathcal{P}}{\to} \frac{1}{\mu} \qquad \text{and} \qquad \frac{D_n - \mu^{-1}\log n}{\sigma\sqrt{\mu^{-3}\log n}} \stackrel{\mathcal{L}}{\to} \mathcal{N}(0,1)$$

where μ and σ^2 are simply the expected value and the variance of $-\log X$, $\mathcal{N}(0,1)$ denotes the standard Gaussian distribution and the symbols $\stackrel{\mathcal{P}}{\to}$ and $\stackrel{\mathcal{L}}{\to}$ refer to convergence in probability and convergence in distribution. This generalizes a result of Mahmoud (2009). In Sections 3 and 4, we prove the main theorem (Theorem 2) of this paper: If $\mathcal{L}(X)$ has a bounded density on $(0,1)$, then there exist a constant $\alpha_{\max}$ such that

$$\frac{H_n}{\log n} \stackrel{\mathcal{P}}{\to} \alpha_{\max}.$$

This constant is defined as the solution of an equation involving the rate function associated to $\log X$. The proof of this result uses a second moment method. We bound the joint probabilities of nodes x and y having a large depth by considering a restricted class of long path for which we can bound the probability of merging of two branches.

2 The depth of a node

We look at the sequence of labels from node n to the root as a renewal process. The law of large numbers and central limit theorem for renewal processes are all we use. X is a random variable on $[0,1)$, we denote

$$\mu = \mathbf{E}\{-\log X\} > 0 \qquad \text{and} \qquad \sigma^2 = \mathbf{Var}\{-\log X\}.$$

Consider independent random variables $X_n \stackrel{\mathcal{L}}{=} X$ such that the label of the parent of n is given by $\lfloor nX_n \rfloor$. The root of the tree is labelled 0 and $L(n,j)$ is the (random) label of the j-th grandparent of n on its path to the root. Note that $L(n, j+1) = \lfloor L(n,j) X_{L(n,j)} \rfloor$ and that $L(n, 0) = n$. A tree is thus described by the sequence $X_0, X_1, \ldots, X_n$ of i.i.d. random variables. Then we have that

$$D_n = \min\{j \geq 0 : L(n,j) = 0\} = \min\{j \geq 0 : \lfloor \lfloor \lfloor nX_n \rfloor X_{L(n,1)} \rfloor \ldots X_{L(n,j-1)} \rfloor < 1\}.$$

To analyze the behaviour of D_n, note that

$$nX_n X_{L(n,1)} \ldots X_{L(n,j-1)} - j \leq \lfloor \lfloor \lfloor nX_n \rfloor X_{L(n,1)} \rfloor \ldots X_{L(n,j-1)} \rfloor \leq nX_n X_{L(n,1)} \ldots X_{L(n,j-1)}.$$

Theorem 1

(A) *If* $\mu = \infty$*, then* $\dfrac{D_n}{\log n} \xrightarrow{\mathcal{P}} 0.$

(B) *If* $\mu < \infty$*, then* $\dfrac{D_n}{\log n} \xrightarrow{\mathcal{P}} \dfrac{1}{\mu}.$

(C) *If* $\mu < \infty$ *and* $0 < \sigma^2 < \infty$*, then* $\dfrac{D_n - \log n/\mu}{\sigma\sqrt{\log n/\mu^3}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1).$

Remark 1 *Mahmoud (2009) proved a similar result using generating functions for the particular cases* $X \stackrel{\mathcal{L}}{=} \max(U_1, \ldots, U_k)$ *and* $X \stackrel{\mathcal{L}}{=} \min(U_1, \ldots, U_k)$*, where* $U_1, \ldots, U_k$ *are i.i.d. uniform on* $(0,1)$*.*

Proof: We consider an auxiliary renewal process $R_t = \max\left\{j : \sum_{i=1}^{j} Z_i \leq t\right\}$ with interarrival times distributed as $Z_i \stackrel{\mathcal{L}}{=} -\log X$ for all i. When $\mu < \infty$, the strong law of large numbers for renewal processes gives that $R_t/t \to 1/\mu$ almost surely (See Ross 1996, Proposition 3.3.1).

When $\mu = \infty$, we have $R_t/t \stackrel{\text{a.s.}}{\to} 0$ as $t \to \infty$ by the following truncation argument: let $\widetilde{Z}_i = \min(Z_i, a)$ where a is chosen so that $\mathbf{E}\{\widetilde{Z}_i\} \geq b$. Consider the renewal process $\widetilde{R}_t$ with interarrival times $\widetilde{Z}_i$. By the fact that $R_t \leq \widetilde{R}_t$ and the law of large numbers for $\widetilde{R}_t$ we have, for sufficiently large t, $R_t/t \leq \widetilde{R}_t/t < 2/b$ almost surely. Since b is arbitrary, we have $R_t/t \to 0$ with probability 1.

We upper bound the depth of the node n by

$$\begin{aligned} D_n &\leq \min\left\{j : nX_n X_{L(n,1)} \ldots X_{L(n,j-1)} < 1\right\} \\ &= \min\left\{j : \textstyle\sum_{i=0}^{j-1} -\log X_{L(n,i)} > \log n\right\} \stackrel{\text{def}}{=} \widehat{D}_n. \end{aligned}$$

For $n \geq 1$, $\widehat{D}_n \stackrel{\mathcal{L}}{=} R_{\log n} + 1$. So, we have for any $\varepsilon > 0$ that

$$\mathbf{P}\left\{\frac{D_n}{\log n} > \frac{1}{\mu} + \varepsilon\right\} \leq \mathbf{P}\left\{\frac{\widehat{D}_n}{\log n} > \frac{1}{\mu} + \varepsilon\right\} = \mathbf{P}\left\{\frac{R_{\log n} + 1}{\log n} > \frac{1}{\mu} + \varepsilon\right\} = o(1). \tag{1}$$

Since $D_n > 0$, equation (1) proves part (A) of the theorem (by writing $1/\mu = 0$ when $\mu = \infty$).

Similarly, a lower bound is given by

$$\begin{aligned} D_n &\geq \min\left\{j : nX_n \ldots X_{L(n,j-1)} - j < 1\right\} \\ &\geq \min\left\{j : \textstyle\sum_{i=0}^{j-1} -\log X_{L(n,i)} > \log n - \log j\right\}. \end{aligned}$$

Let $j(n) = \lfloor \log^2 n \rfloor$ and define the event

$$E_n = \left[\sum_{i=0}^{j(n)-1} - \log X_{L(n,i)} > \log n \right].$$

Using the upper bound (1), we have that $\mathbf{P}\{E_n\} \to 1$. Also, we have $\log j \leq 2 \log\log n$ and if we define $f(n) = \log n - 2\log\log n$, then when E_n holds

$$D_n \geq \min\left\{ j : \sum_{i=0}^{j-1} - \log X_{L(n,i)} > f(n) \right\} \stackrel{\text{def}}{=} \overline{D}_n.$$

We have $\overline{D}_n \stackrel{\mathcal{L}}{=} R_{f(n)} + 1$ for $n \geq 2$, and thus,

$$\mathbf{P}\left\{ \frac{\overline{D}_n}{\log n} < \frac{1}{\mu} - \varepsilon \right\} = \mathbf{P}\left\{ \frac{R_{f(n)} + 1}{f(n)} \cdot \frac{f(n)}{\log n} < \frac{1}{\mu} - \varepsilon \right\} = o(1), \tag{2}$$

by the law of large numbers for renewal processes and the fact that

$$\lim_{n\to\infty} \frac{f(n)}{\log n} = 1.$$

Combining (1) and (2) with the fact that $\mathbf{P}\left\{D_n \geq \overline{D}_n\right\} \geq \mathbf{P}\{E_n\}$ we obtain part (B) of the theorem.

Part (C) is proven by similar arguments using the central limit theorem for renewal processes (see Ross 1996, Theorem 3.3.5). □

3 The height of the tree: upper bound

The height of a SARRT is defined by $H_n = \max_{1\leq i\leq n} D_i$. We first deal with its upper tail. For a random variable Y, define its cumulant generating function of Λ_Y and its convex (Fenchel–Legendre) dual Λ_Y^* as follows. For $\lambda \geq 0$ and $z \geq \mathbf{E}\{Y\}$,

$$\Lambda_Y(\lambda) = \log \mathbf{E}\left\{e^{\lambda Y}\right\} \quad \text{and} \quad \Lambda_Y^*(z) = \sup_{\lambda\geq 0}\left\{\lambda z - \Lambda_Y(\lambda)\right\}. \tag{3}$$

Since we only use these functions for $Y = \log X$ (when $X = 0$, $\log X$ is defined to be $-\infty$), the subscript is omitted for this case. We write

$$\Lambda(\lambda) = \log \mathbf{E}\left\{e^{\lambda \log X}\right\} = \log \mathbf{E}\left\{X^\lambda\right\} \quad \text{and} \quad \Lambda^*(z) = \sup_{\lambda\geq 0}\left\{\lambda z - \Lambda(\lambda)\right\}. \tag{4}$$

for the cumulant generating function of $\log X$ and its dual. The quantity

$$\Psi(c) = c\Lambda^*\left(-1/c\right) \tag{5}$$

plays an important role. The objective of this section and the next one is to show that $\frac{H_n}{\log n} \stackrel{\mathcal{P}}{\to} \alpha_{\max}$ where

$$\alpha_{\max} = \inf\left\{ c : \; c > \frac{1}{\mu} \text{ and } \Psi(c) > 1 \right\}. \tag{6}$$

Based on the techniques from Chernoff (1952) and Hoeffding (1963), we can show that the expected number of nodes with depth at least $c \log n$ for $c > \alpha_{\max}$ goes to 0 as $n \to \infty$. This implies the following lemma.

Lemma 1 *For any $c > \alpha_{\max}$, we have $\mathbf{P}\{H_n \geq c \log n\} \to 0$.*

4 The height of the tree: lower bound

4.1 The probability of having a large depth

We begin with precise bounds on the probability of the events $[X_1 \cdots X_t \geq b]$.

Proposition 1 (Cramér, 1938) *Let $Y_1, \ldots, Y_t$ be a sequence of i.i.d. real random variables having a well-defined expected value $\mathbf{E}\{Y_1\} \in \mathbb{R} \cup \{\pm\infty\}$ and let $a \in \mathbb{R}$ be a constant. Then*

$$\mathbf{P}\{Y_1 + \cdots + Y_t \geq ta\} = \exp\Big(-t\Lambda_Y^*(a) + o(t)\Big) \qquad \text{if } a \geq \mathbf{E}\{Y_1\} \text{ and } \mathbf{E}\{Y_1\} \neq +\infty,$$

where Λ_Y^* is as defined in equation (3).

See also Dembo and Zeitouni (1998, chap. 2, page 27). As a result, for products of i.i.d. random variables, by taking the logarithm we get for Λ^* as defined in equation (4) and a positive integer t:

$$\mathbf{P}\left\{X_1 \ldots X_t \geq e^{ta}\right\} = \exp\Big(-t\Lambda^*(a) + o(t)\Big) \qquad \text{if } a \geq -\mu \in [-\infty, 0). \tag{7}$$

The next lemma is based on a rotation argument introduced by Dwass (1969). A proof of a similar lemma can be found in Devroye and Reed (1995).

Lemma 2 *Let t be a positive integer, let $\beta > 0$, and let $X_1, \ldots, X_t$ be i.i.d. random variables. Then*

$$\mathbf{P}\left\{X_1 \geq \beta, X_1X_2 \geq \beta^2, \ldots, X_1 \cdots X_t \geq \beta^t\right\} \geq \frac{1}{t}\mathbf{P}\left\{X_1 \cdots X_t \geq \beta^t\right\}.$$

For convenience of notation, the nodes of the tree are labelled from 0 to $3n$, and we shall study the height H_{3n}. For a node $x \in \{2n+1, \ldots, 3n\}$, $t \in \mathbb{N}$ and $0 < \beta < 1$, define the event

$$A_{x,t}(\beta) = \left[L(x,1) \geq n\beta, L(x,2) \geq n\beta^2, \ldots, L(x,t) \geq n\beta^t\right]. \tag{8}$$

We set $A_{x,0}(\beta) = \left[L(x,0) > n\beta^0\right] = [x > n]$ so that $\mathbf{P}\{A_{x,0}(\beta)\} = 1$. Note that when β is clear from the context, we just write $A_{x,t}$ for $A_{x,t}(\beta)$.

Lemma 3 *Assume X has a bounded density. Let $c \in (1/\mu, \alpha_{\max})$, $\beta = e^{-1/c}$ and $\delta > 0$ such that $\Psi(c) + \delta < 1$ and $\Psi(c) - \delta > 0$ (see equation* (5) *for a definition of Ψ). Then there exists $t_0 = t_0(c, \delta, \mathcal{L}(X))$ such that for all integers $t \geq t_0$, $n \geq t\beta^{-t}$ and $2n+1 \leq x \leq 3n$,*

$$\frac{\beta^t}{t} \leq \frac{\beta^{(\Psi(c)+\delta)t}}{t} \leq \mathbf{P}\{A_{x,t}(\beta)\} \leq \beta^{(\Psi(c)-\delta)t}.$$

Proof: Using standard properties of the function Λ^*, we know that $0 < \Psi(c) < 1$ for $c \in (1/\mu, \alpha_{\max})$. So we can choose $\delta > 0$ with $\Psi(c) + \delta < 1$ and $\Psi(c) - \delta > 0$. We start with the upper bound. Using Chernoff's large deviation bound,

$$\begin{aligned}\mathbf{P}\left\{L(x,t) \geq n\beta^t\right\} &\leq \mathbf{P}\left\{3nX_{L(x,0)}\ldots X_{L(x,t-1)} \geq n\beta^t\right\} \\ &\leq \inf_{\lambda \geq 0} \exp\Big(\lambda(-t\log\beta + \log 3) + \Lambda(\lambda)t\Big) \\ &= \exp\left(-t\Lambda^*\left(-\frac{1}{c} - \frac{\log 3}{t}\right)\right).\end{aligned}$$

By definition of Ψ, we have $\Lambda^*(-1/c) = \Psi(c)/c$. Thus for t large enough, by continuity of Λ^*, $\Lambda^*(-1/c - (\log 3)/t) > (\Psi(c) - \delta)/c$. Thus,

$$\mathbf{P}\left\{L(x,t) \geq n\beta^t\right\} \leq \exp\left(-t(\Psi(c) - \delta)/c\right) = \beta^{(\Psi(c)-\delta)t}.$$

To prove a lower bound on the probability of $A_{x,t}$, we use that for all $s \in \{1, \ldots, t\}$

$$\begin{aligned}[L(x,s) \geq n\beta^s] &\supseteq \left[2nX_{L(x,0)}\cdots X_{L(x,s-1)} - s \geq n\beta^s\right] \\ &\supseteq \left[X_{L(x,0)}\cdots X_{L(x,s-1)} \geq \frac{\beta^s}{2} + \frac{s}{2n}\right] \\ &\supseteq \left[X_{L(x,0)}\cdots X_{L(x,s-1)} \geq \beta^s\right].\end{aligned}$$

The last inclusion holds because we assumed $n \geq t\beta^{-t} \geq s\beta^{-s}$ for all $s \leq t$. Thus, we write

$$\begin{aligned}\mathbf{P}\{A_{x,t}\} &= \mathbf{P}\left\{L(x,1) \geq n\beta, L(x,2) \geq n\beta^2, \ldots, L(x,t) \geq n\beta^t\right\} \\ &\geq \mathbf{P}\left\{X_{L(x,0)} \geq \beta, X_{L(x,0)}X_{L(x,1)} \geq \beta^2, \ldots, X_{L(x,0)}\cdots X_{L(x,t-1)} \geq \beta^t\right\}.\end{aligned}$$

We now use Lemma 2 to get

$$\mathbf{P}\{A_{x,t}\} \geq \frac{1}{t}\mathbf{P}\left\{X_{L(x,0)}\cdots X_{L(x,t-1)} \geq \beta^t\right\}.$$

Using Cramér's theorem (equation (7)),

$$\begin{aligned}\mathbf{P}\left\{X_{L(x,0)}\cdots X_{L(x,t-1)} \geq \beta^t\right\} &= \mathbf{P}\left\{X_{L(x,0)}\cdots X_{L(x,t-1)} \geq e^{-t/c}\right\} \\ &= \exp\Big(-t\Lambda^*(-1/c) + o(t)\Big).\end{aligned}$$

But $\Lambda^*(-1/c) = \Psi(c)/c < (\Psi(c) + \delta)/c$. So for t large enough,

$$\mathbf{P}\left\{X_{L(x,0)}\cdots X_{L(x,t-1)} \geq \beta^t\right\} \geq \exp\left(-(\Psi(c) + \delta)t/c\right) = \beta^{(\Psi(c)+\delta)t}.$$

As a result

$$\mathbf{P}\{A_{x,t}\} \geq \frac{\beta^{(\Psi(c)+\delta)t}}{t} \geq \frac{\beta^t}{t}.$$

□

4.2 The second moment method

The main theorem (Theorem 2) is proven using a second moment method on the number of nodes that have a large depth. To prepare for this, we compute a bound on the joint probability $\mathbf{P}\{A_{x,t} \cap A_{y,t}\}$.

Lemma 4 *Let X have a density bounded by κ on $(0,1)$. Let $x \neq y$ be elements of $\{2n+1,\dots,3n\}$, let t be a positive integer and let $\beta \in (0,1)$. Then*

$$\mathbf{P}\{A_{x,t} \cap A_{y,t}\} \leq \sum_{s=0}^{t-1} \mathbf{P}\{A_{x,t}\} \mathbf{P}\{A_{y,s}\} \frac{(t+1)\kappa}{n\beta^s} + \mathbf{P}\{A_{x,t}\} \mathbf{P}\{A_{y,t}\}.$$

Proof: Define $P_t(x) = \{L(x,0), L(x,1), \dots, L(x,t)\}$ to be the (random) path of length t starting from x towards the root. We introduce the collision time T when the path starting at y meets the path of x. It is convenient to let T be the step before the collision happens: $T = \min\{s \geq 0 : L(y,s+1) \in P_t(x)\}$ if $P_t(x) \cap P_t(y) \neq \emptyset$, and $T = \infty$ otherwise. Then

$$\mathbf{P}\{A_{x,t} \cap A_{y,t}\} = \sum_{s=0}^{t-1} \mathbf{P}\{T = s, A_{x,t} \cap A_{y,t}\} + \mathbf{P}\{T = \infty, A_{x,t} \cap A_{y,t}\}.$$

In order to evaluate this expression, we fix the path $P_t(x)$ from x to its t-th ancestor. Consider the set of possible paths $\mathcal{F} = \{Q \subseteq \{0,\dots,3n\} : x = \max Q,\ |Q| \leq t\}$. For all $s \in \{0,\dots,t-1\}$

$$\begin{aligned}
&\mathbf{P}\{T = s, A_{x,t} \cap A_{y,t}\} \\
&\quad = \sum_{Q \in \mathcal{F}} \mathbf{P}\{T = s, A_{x,t} \cap A_{y,t}, P_t(x) = Q\} \\
&\quad \leq \sum_{Q \in \mathcal{F}} \mathbb{1}_{A_{x,t}}(Q)\, \mathbf{P}\{P_s(y) \cap Q = \emptyset, L(y,s+1) \in Q, A_{y,s}, P_t(x) = Q\} \\
&\quad = \sum_{Q \in \mathcal{F}} \mathbb{1}_{A_{x,t}}(Q) \sum_{u:\ \substack{u \geq n\beta^s \\ u \notin Q}} \mathbf{P}\{P_s(y) \cap Q = \emptyset, L(y,s) = u, \lfloor uX_u \rfloor \in Q, A_{y,s}, P_t(x) = Q\}.
\end{aligned}$$

where $\mathbb{1}_{A_{x,t}}(Q)$ is the indicator of the event $A_{x,t}$ when $P_t(x) = Q$. In order to simplify this expression, we use the independence claim below.

Claim: *For any $Q \subseteq \{0,\dots,3n\}$ and $u \notin Q$, the events $[P_s(y) \cap Q = \emptyset, L(y,s) = u, A_{y,s}]$, $[\lfloor uX_u \rfloor \in Q]$ and $[P_t(x) = Q]$ are mutually independent.*

A proof of this claim can be found in the full version Devroye et al. (2010). It follows that

$$
\begin{aligned}
&\mathbf{P}\left\{T = s, A_{x,t} \cap A_{y,t}\right\} \\
&\le \sum_{Q \in \mathcal{F}} \mathbb{1}_{A_{x,t}}(Q) \sum_{u\colon \substack{u \ge n\beta^s \\ u \notin Q}} \mathbf{P}\left\{P_s(y) \cap Q = \emptyset, L(y,s) = u, A_{y,s}\right\} \mathbf{P}\left\{P_t(x) = Q\right\} \mathbf{P}\left\{\lfloor uX_u \rfloor \in Q\right\} \\
&\le \sum_{Q \in \mathcal{F}} \mathbb{1}_{A_{x,t}}(Q) \mathbf{P}\left\{A_{y,s}\right\} \mathbf{P}\left\{P_t(x) = Q\right\} \sup_{u\colon \substack{u \ge n\beta^s \\ u \notin Q}} \mathbf{P}\left\{\lfloor uX_u \rfloor \in Q\right\} \\
&\le \left(\sum_{Q \in \mathcal{F}} \mathbb{1}_{A_{x,t}}(Q) \mathbf{P}\left\{P_t(x) = Q\right\} \right) \mathbf{P}\left\{A_{y,s}\right\} (t+1) \sup_{\substack{u\colon u \ge n\beta^s \\ w\colon w \ge n\beta^t}} \mathbf{P}\left\{\lfloor uX_u \rfloor = w\right\} \\
&= \mathbf{P}\left\{A_{x,t}\right\} \mathbf{P}\left\{A_{y,s}\right\} (t+1) \sup_{\substack{u\colon u \ge n\beta^s \\ w\colon w \ge n\beta^t}} \mathbf{P}\left\{\lfloor uX_u \rfloor = w\right\}.
\end{aligned}
$$

The last inequality holds because either the event $A_{x,t}$ holds, in which case all nodes in $P_t(x)$ have a label at least $n\beta^t$, or $A_{x,t}$ does not hold and the left hand side is 0. In order to bound the collision probability $\mathbf{P}\left\{\lfloor uX_u \rfloor = w\right\}$, we use the fact that X has a density bounded by κ:

$$
\mathbf{P}\left\{\lfloor uX_u \rfloor = w\right\} \le \mathbf{P}\left\{X_u \in \left[\frac{w}{u}, \frac{w+1}{u}\right)\right\} \le \frac{\kappa}{u}.
$$

Thus,

$$
\mathbf{P}\left\{T = s, A_{x,t} \cap A_{y,t}\right\} \le \mathbf{P}\left\{A_{x,t}\right\} \mathbf{P}\left\{A_{y,s}\right\} \frac{(t+1)\kappa}{n\beta^s}.
$$

Repeating the above argument for $T = \infty$, we get

$$
\begin{aligned}
\mathbf{P}\left\{T = \infty, A_{x,t} \cap A_{y,t}\right\} &\le \sum_{Q \in \mathcal{F}} \mathbb{1}_{A_{x,t}}(Q)\, \mathbf{P}\left\{P_t(y) \cap Q = \emptyset, A_{y,t}, P_t(x) = Q\right\} \\
&\le \left(\sum_{Q \in \mathcal{F}} \mathbb{1}_{A_{x,t}}(Q) \mathbf{P}\left\{P_t(x) = Q\right\} \right) \mathbf{P}\left\{A_{y,t}\right\} \\
&= \mathbf{P}\left\{A_{x,t}\right\} \mathbf{P}\left\{A_{y,t}\right\}.
\end{aligned}
$$

□

Theorem 2 *The height H_n of a* SARRT *with attachment X having a bounded density verifies*

$$
\frac{H_n}{\log n} \xrightarrow{\mathcal{P}} \alpha_{\max} \quad \textit{as} \quad n \to \infty
$$

where $\alpha_{\max}$ is defined in equation (6).

Before proving the theorem, we give a simple application to the power of choice in random recursive trees. Suppose each new node n is given k choices of parents and chooses the one with the largest label. Let R_n^+ be the depth of node n in this tree and $H_n^+ = \max_{1\le i\le n} R_n^+$ be the height of the tree. The question is how does R_n^+ and H_n^+ change as a function of k. Note that R_n^+ can also be seen as the distance in a random k-DAG from n to the root given by following the parent with largest label (Devroye and Janson, 2009).

By applying Theorems 1 and 2 to $X = \max(U_1, \ldots, U_k)$ where $U_1, \ldots, U_k$ are independent uniform $(0, 1)$ random variables, we get

$$\frac{R_n^+}{\log n} \xrightarrow{\mathcal{P}} k \qquad \text{and} \qquad \frac{H_n^+}{\log n} \xrightarrow{\mathcal{P}} \rho_{\max}^+$$

where $\rho_{\max}^+$ is the solution larger than k of the equation $-c + k - c\log\frac{k}{c} = 1$. For example, setting $k = 1$, we recover the asymptotic height $e\log n + o(\log n)$ of a URRT. Table 1 shows approximate values of $\rho_{\max}^+$ for $k = 1, \ldots, 5$.

Similarly, node n could choose the node with smallest label as a parent. Let D_n^- and H_n^- be the depth of node n and the height for the tree constructed in this way. We can apply Theorems 1 and 2 to $X = \min(U_1, \ldots, U_k)$:

$$\frac{R_n^-}{\log n} \xrightarrow{\mathcal{P}} \frac{1}{h_k} \qquad \text{and} \qquad \frac{H_n^-}{\log n} \xrightarrow{\mathcal{P}} \rho_{\max}^-$$

where $h_k = \sum_{i=1}^k \frac{1}{i}$ and $\rho_{\max}^-$ is the solution larger than $\frac{1}{h_k}$ of the equation

$$-\lambda_k^*(-1/c) + \sum_{i=1}^{k} \log\left(1 + \frac{\lambda_k^*(-1/c)}{i}\right) = 1$$

where $\lambda_k^*(z)$ is the solution of $z + \sum_{i=1}^k \frac{1}{1+\lambda_k^*(z)/i} = 0$. Approximate values are shown in Table 1. Note that the value of $\rho_{\max}^+$ for $k = 2$ is the same as the constant in the height of the random binary search tree (Devroye, 1986).

Tab. 1: Approximate numerical values for some constants

k	1	2	3	4	5
$\rho_{\max}^+$	e	4.3111	5.7640	7.1451	8.4805
$\rho_{\max}^-$	e	1.6738	1.3025	1.1060	0.9818

Proof: Fix $\delta \in (0, 1/2)$, $\varepsilon \in (0, 1)$ with $3\delta < \varepsilon$ and $c \in (1/\mu, \alpha_{\max})$. Define $\beta = e^{-1/c}$ and $t = \lfloor (1-\varepsilon)c\log n \rfloor$. Our objective is to show that $\lim_{n\to\infty} \mathbf{P}\{H_{3n} \ge t\} = 1$.

For this we consider the union of the events $A_{x,t}$ defined in equation (8). The fact that $A_{x,t}$ holds implies that $L(x, t) \ge n\beta^t \ge n/n^{1-\varepsilon} = n^\varepsilon \ge 1$, i.e., the depth of node x is at least t. A lower bound on the probability is given by the following second moment inequality (Chung and Erdős, 1952):

$$\mathbf{P}\left\{\bigcup_{x=2n+1}^{3n} A_{x,t}\right\} \ge \frac{\left(\sum_{x=2n+1}^{3n} \mathbf{P}\{A_{x,t}\}\right)^2}{\sum_{x=2n+1}^{3n} \mathbf{P}\{A_{x,t}\} + \sum_{x\ne y} \mathbf{P}\{A_{x,t} \cap A_{y,t}\}}. \tag{9}$$

Let $t_0(c, \delta, \mathcal{L}(X))$ be defined as in Lemma 3. When n is large enough, the conditions $t \geq t_0$ and $n \geq t\beta^{-t}$ are met. So Lemma 3 gives

$$\mathbf{P}\{A_{x,t}\} \geq \frac{\beta^t}{t} \geq \frac{1}{tn^{1-\varepsilon}}. \tag{10}$$

Now, fixing $x \neq y$, we have by Lemma 4:

$$\mathbf{P}\{A_{x,t} \cap A_{y,t}\} \leq \sum_{s=0}^{t-1} \mathbf{P}\{A_{x,t}\} \mathbf{P}\{A_{y,s}\} \frac{(t+1)\kappa}{n\beta^s} + \mathbf{P}\{A_{x,t}\} \mathbf{P}\{A_{y,t}\}.$$

For $s \geq t_0$, we apply Lemma 3 to find an upper bound on $\mathbf{P}\{A_{x,s}\}$:

$$\begin{aligned}\mathbf{P}\{A_{x,t} \cap A_{y,t}\} &\leq \mathbf{P}\{A_{x,t}\} \left(\sum_{s=0}^{t_0-1} \frac{(t+1)\kappa}{n\beta^s} + \sum_{s=t_0}^{t-1} \beta^{(\Psi(c)-\delta)s} \frac{(t+1)\kappa}{n\beta^s} + \mathbf{P}\{A_{y,t}\} \right) \\ &\leq \mathbf{P}\{A_{x,t}\} \left(O\left(\frac{t}{n}\right) + \frac{(t+1)\kappa}{n} \cdot \frac{\beta^{(\Psi(c)-\delta-1)t} - 1}{\beta^{(\Psi(c)-\delta-1)} - 1} + \mathbf{P}\{A_{y,t}\} \right). \end{aligned} \tag{11}$$

We now show that the dominating term is $\mathbf{P}\{A_{x,t}\}\mathbf{P}\{A_{y,t}\}$. Using equation (10),

$$\frac{t/n}{\mathbf{P}\{A_{y,t}\}} \leq \frac{t^2 n^{1-\varepsilon}}{n} = O\left(n^{-\varepsilon/2}\right) \tag{12}$$

as $t = O(\log n)$. Moreover, using the more precise lower bound on $\mathbf{P}\{A_{y,t}\}$ given in Lemma 3,

$$\frac{t\beta^{(\Psi(c)-\delta-1)t}}{n\mathbf{P}\{A_{y,t}\}} \leq \frac{t^2 \beta^{(\Psi(c)-\delta-1)t} \beta^{-(\Psi(c)+\delta)t}}{n} = \frac{t^2 (\beta^{-t})^{2\delta} \beta^{-t}}{n}.$$

By definition of t, $\beta^{-t} \leq n^{1-\varepsilon}$, and thus

$$\frac{t\beta^{(\Psi(c)-\delta-1)t}}{n\mathbf{P}\{A_{y,t}\}} \leq t^2 n^{2\delta-\varepsilon} \leq t^2 n^{-\varepsilon/3} = O(n^{-\varepsilon/4}). \tag{13}$$

Plugging inequalities (12) and (13) into (11), we get

$$\mathbf{P}\{A_{x,t} \cap A_{y,t}\} \leq \mathbf{P}\{A_{x,t}\} \mathbf{P}\{A_{y,t}\} \left(1 + O\left(n^{-\varepsilon/4}\right)\right).$$

Going back to equation (9), we can now bound

$$\sum_{x \neq y} \mathbf{P}\{A_{x,t} \cap A_{y,t}\} \leq \left(\sum_{x=2n+1}^{3n} \mathbf{P}\{A_{x,t}\} \right)^2 \left(1 + O\left(n^{-\varepsilon/4}\right)\right)$$

and

$$\sum_{x=2n+1}^{3n} \mathbf{P}\{A_{x,t}\} \geq n \frac{1}{tn^{1-\varepsilon}} = \frac{n^\varepsilon}{t}.$$

Thus,

$$\mathbf{P}\left\{\bigcup_{x=2n+1}^{3n} A_{x,t}\right\} \geq \frac{1}{\left(\sum_{x=2n+1}^{3n}\mathbf{P}\left\{A_{x,t}\right\}\right)^{-1} + 1 + O\left(n^{-\varepsilon/4}\right)} \geq 1 - O\left(n^{-\varepsilon/4}\right) - O\left(tn^{-\varepsilon}\right).$$

This shows that

$$\mathbf{P}\left\{H_{3n} \geq t\right\} = \mathbf{P}\left\{H_{3n} \geq \lfloor (1-\varepsilon)c\log n\rfloor\right\} \geq 1 - O\left(n^{-\varepsilon/4}\right).$$

Combining this with the upper bound given in Lemma 1, we get the desired result. □

5 Conclusions

To compute the height of the tree, our proof uses the existence of a bounded a density for $\mathcal{L}(X)$ in order to bound the collision probability. The existence of a density is only used to find a lower bound on the height. The upper bound given here (Lemma 1) works for any distribution. It is natural to ask whether this upper bound is tight for a larger family of distributions. In the full version Devroye et al. (2010), we give a complete proof of Theorem 2 for X having any density and possibly an atom at 0. We also study the minimum depth $M_n = \min_{n/2 \leq i \leq n} D_i$ of a SARRT.

One can construct a random k-DAG or SARRD in the same way. Node n chooses k parents $\lfloor nX^{(1)}\rfloor$, $\lfloor nX^{(2)}\rfloor, \ldots, \lfloor nX^{(k)}\rfloor$ where $X^{(1)}, \ldots, X^{(k)}$ are independent copies of a random variable $X \in [0,1)$. The "greedy" distance measures can be computed simply by considering the SARRT with attachment random variable $X_{\min} = \min(X^{(1)}, \ldots, X^{(k)})$ and $X_{\max} = \max(X^{(1)}, \ldots, X^{(k)})$. One could study the shortest and longest path distances in a SARRD, which has been done for the uniform case in Arya et al. (1999), Devroye and Janson (2009), D'Souza et al. (2007), and Tsukiji and Xhafa (1996).

References

S. Arya, M. Golin, and K. Mehlhorn. On the expected depth of random circuits. *Combinatorics, Probability and Computing*, 8:209–228, 1999.

A. Barabasi and R. Albert. Emergence of scaling in random networks. *Science*, 286:509–512, 1999.

H. Chernoff. A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *The Annals of Mathematical Statistics*, 23:493–507, 1952.

K. Chung and P. Erdős. On the application of the Borel-Cantelli lemma. *Transactions of the American Mathematical Society*, pages 179–186, 1952.

H. Cramér. Sur un nouveau théorème-limite de la théorie des probabilités. *Actualités Scientifiques et Industrielles*, 736:5–23, 1938.

A. Dembo and O. Zeitouni. *Large Deviations Techniques and Applications*. Springer Verlag, 1998.

L. Devroye. A note on the height of binary search trees. *Journal of the ACM*, 33(3):489–498, 1986.

L. Devroye. Branching processes in the analysis of the heights of trees. *Acta Informatica*, 24(3):277–298, 1987.

L. Devroye. Applications of the theory of records in the study of random trees. *Acta Informatica*, 26: 123–130, 1988.

L. Devroye and S. Janson. Long and short paths in uniform random recursive dags. *Arkiv för Matematik (to appear)*, 2009. URL `http://arxiv.org/abs/0906.0152v1`.

L. Devroye and J. Lu. The strong convergence of maximal degrees in uniform random recursive trees and dags. *Random Structures and Algorithms*, 7(1):1–14, 1995.

L. Devroye and B. Reed. On the variance of the height of random binary search trees. *SIAM Journal on Computing*, 24:1157–1162, 1995.

L. Devroye, O. Fawzi, and N. Fraiman. Depth properties of scaled attachment Random recursive trees. Manuscript submitted for publication, 2010.

M. Drmota. *Random trees: an interplay between combinatorics and probability*. Springer, 2009.

R. M. D'Souza, P. L. Krapivsky, and C. Moore. The power of choice of choice in growing trees. *The European Physical Journal B - Condensed Matter and Complex Systems*, 59(4):535–543, 2007.

M. Dwass. The total progeny in a branching process and a related random walk. *Journal of Applied Probability*, 6(3):682–686, 1969.

J. L. Gastwirth. A probability model of a pyramid scheme. *The American Statistician*, 31(2):79–82, 1977.

W. Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58:13–30, 1963.

H. Mahmoud. The power of choice in the construction of recursive trees. *Methodology and Computing in Applied Probability (to appear)*, 2009.

H. Mahmoud. Limiting distributions for path lengths in recursive trees. *Probability in the Engineering and Informational Sciences*, 5(01):53–59, 1991.

H. Mahmoud. Distances in random plane-oriented recursive trees. *Journal of Computational and Applied Mathematics*, 41(1-2):237–245, 1992.

J. Moon. The distance between nodes in recursive trees. *London Mathematical Society Lecture Notes*, 13: 125–132, 1974.

H. Na and A. Rapoport. Distribution of nodes of a tree by degree. *Mathematical Biosciences*, 6:313–329, 1970.

B. Pittel. Note on the heights of random recursive trees and random m-ary search trees. *Random Structures and Algorithms*, 5:337–348, 1994.

S. Ross. *Stochastic processes*. Wiley New York, 1996.

R. Smythe and H. Mahmoud. A survey of recursive trees. *Theory of Probability and Mathematical Statistics*, 51:1–27, 1995.

J. Szymański. On a nonuniform random recursive tree. In *Random Graphs' 85: Based on Lectures Presented at the 2nd International Seminar on Random Graphs and Probabilistic Methods in Combinatorics*, volume 33, pages 297–306. North-Holland, 1987.

T. Tsukiji and F. Xhafa. On the depth of randomly generated circuits. In *Proceedings of Fourth European Symposium on Algorithms*, pages 208–220, 1996.