Probability in the Engineering and Informational Sciences, **21**, 2007, 419–433. Printed in the U.S.A.
DOI: 10.1017/S0269964807000058

ON WEIGHTED PATH LENGTHS AND DISTANCES IN INCREASING TREES

M. KUBA AND A. PANHOLZER
Institut für Diskrete Mathematik und Geometrie
Technische Universität Wien, 1040 Wien, Austria
E-mail: markus.kuba@tuwien.ac.at; alois.panholzer@tuwien.ac.at

We study weighted path lengths (depths) and distances for increasing tree families. For those subclasses of increasing tree families, which can be constructed via an insertion process (e.g., recursive trees, plane-oriented recursive trees, and binary increasing trees), we can determine the limiting distribution that can be characterized as a generalized Dickman's infinitely divisible distribution.

1. INTRODUCTION

The analysis of the length of paths in tree families has received much attention (see e.g., [4,7,12], often due to their importance in the analysis of algorithms. Recently, Aguech, Lasmar, and Mahmoud [1] introduced weighted path lengths in binary search trees. They have shown that different limit laws arise for weighted path lengths depending on the label of the investigated nodes. In particular, they have derived the limit law for the weighted path length to the minimal label, which is Dickman's infinitely divisible distribution.

We extend the studies of weighted path lengths to other important labeled trees, namely to families of increasing trees. We will also consider weighted distances in increasing trees, where, for both parameters, we will encounter a generalized Dickman distribution as the limit law. In order to obtain the limiting distribution results of the parameters considered, we will use suitable distributional decompositions of the (ordinary) path lengths and distances, which will be extended to corresponding decompositions of their weighted counterparts. Previous results concerning ordinary depths (path lengths) and distances for increasing tree families can be found in Dobrow [5], Dobrow and Smythe [6], and Kuba and Panholzer [10]. The distributional decomposition of the weighted parameters and an application of Curtiss' theorem leads to limiting distribution results for the random variables considered.

© 2007 Cambridge University Press 0269-9648/07 $25.00

https://doi.org/10.1017/S0269964807000058 Published online by Cambridge University Press

The arguments used in the present article are related to Hwang and Tsai [8], where the limit law of the algorithm Quickselect was obtained with a similar technique.

Increasing trees are labeled trees for which the nodes of a tree of size n are labeled by distinct integers of the set $\{1, \dots, n\}$ in such a way that each sequence of labels along any branch starting at the root is increasing. We are concerned throughout this article with certain combinatorial models of increasing trees, called simple families of increasing trees. One can consider them as increasingly labeled, simply generated trees (simply generated trees have been introduced by Meir and Moon [16]). A formal description of simple families of increasing trees is given in Subsection 2.1. A thorough study of simple families (= varieties) of increasing trees was conducted in Bergeron, Flajolet, and Salvy [2].

The interest in simple families of increasing trees stems from the fact that several important tree models, such as recursive trees, plane-oriented recursive trees, and binary increasing trees (also called tournament trees), are special instances of simple families of increasing trees. These tree models are of importance in many applications (see the survey paper of Mahmoud and Smythe [15] and also [14]), for example, they are used to describe the spread of epidemics, for pyramid schemes and they are used as a simplified growth model of the World Wide Web: Plane-oriented recursive trees are a special instance of the so-called Albert–Barabási model for scale-free networks (see, e.g., [3]).

The above-mentioned tree families (i.e., recursive trees, plane-oriented recursive trees, and binary increasing trees) can be described not only combinatorially but also in a probabilistic way via a tree evolution process. This means that for every tree T' of size n with vertices $v_1, \dots, v_n$, one can give probabilities $p_{T'}(v_1), \dots, p_{T'}(v_n)$, such that when starting with a *random tree* T' of size n of the tree family considered, choosing a vertex v_i in T' according to the probabilities $p_{T'}(v_i)$, and attaching node $n+1$ to it, one obtains again a *random tree* T of size $n+1$ of the tree family considered. See, for example, [15] for such a description of recursive trees and plane-oriented recursive trees. However, it has been shown in [17] that it is not possible to describe every simple family of increasing trees as defined in Subsection 2.1 via a tree evolution process. Moreover, a full characterization of those simple families of increasing trees having this additional property has been given there. This characterization is restated in Subsection 2.2. Throughout this article we will choose the term *grown simple families of increasing trees* for the subclass of increasing tree families, which can be generated by a tree evolution process. We will restrict ourselves to the study of this subclass of simple families of increasing trees.

In a rooted tree, the depth of node v, also called the level of node v or path length from v to the root, is measured by the number of edges lying on the unique path from the root to node v. For labeled rooted trees, we consider a generalization of the depth. Let $W_{n,j}$ denote the *weighted depth* (or *weighted path length*) of node j in a size $n \geq j$ random grown simple increasing tree, which is the sum of the labels of the nodes encountered on the path from j to the root labeled 1, where $W_{1,1} = 1$ and $W_{n,j} \geq j+1$. For instance, if the nodes labeled $\lambda_1, \dots, \lambda_k$, with $1 = \lambda_1 < \lambda_2 < \cdots < \lambda_k = j$, are visited on the path from the root to j, then the weighted depth equals $\sum_{l=1}^{k} \lambda_k$.

https://doi.org/10.1017/S0269964807000058 Published online by Cambridge University Press

Further, we consider the *weighted distance* $\mathcal{W}_{n,j}$ between node j and node n in a size n random grown, simple increasing tree, which is the sum of the labels of the nodes encountered on the path from n to the node labeled j (hence, $\mathcal{W}_{n,1} = W_{n,n}$).

In the tree of Figure 1, we have $W_{9,1} = 1$, $W_{9,2} = 3$, $W_{9,3} = 4$, $W_{9,4} = 8$, $W_{9,5} = 6$, $W_{9,6} = 12$, $W_{9,7} = 8$, $W_{9,8} = 20$, and $W_{9,9} = 17$, and $\mathcal{W}_{9,4} = 24$ and $\mathcal{W}_{9,8} = 36$.

We will show that $W_{n,j}$ and $\mathcal{W}_{n,j}$, appropriately scaled, lead to generalized Dickman distributions.

Related results for binary search trees are due to Aguech et al. [1]. Further, we refer to the limit law of Quickselect; see Mahmoud, Modarres, and Smythe [13] and Hwang and Tsai [8], which is also Dickman's infinitely divisible distribution.

We denote by $X \overset{(d)}{=} Y$ the equality in distribution of the random variables X and Y and by $X_n \overset{(d)}{\to} X$, the weak convergence (i.e., the convergence in distribution) of the sequence of random variables X_n to a random variable X. For independent random variables X and Y, we write $X \oplus Y$ for the sum of X and Y. For not necessarily mutually independent random variables X and Y, we write $X + Y$. We denote by $H_n := \sum_{k=1}^{n} 1/k$ the nth harmonic number, and by $H_n^{(a)} := \sum_{k=1}^{n} 1/k^a$, the nth harmonic number of order a. We denote the indicator of the event $\mathcal{A}_k$ by $\mathbb{I}(\mathcal{A}_k)$.

In the next section we gather some results about increasing tree families and the generalized Dickman distribution. In Sections 3 and 4 we state the main limiting distribution results of this work concerning the weighted depth and the weighted distance. Section 5 is devoted to the derivation of the distribution laws of the weighted depths and distances. The last section contains the proofs of the limiting distribution results.

2. PRELIMINARIES

2.1 Combinatorial Description of Increasing Tree Families

Formally, a class $\mathcal{T}$ of a simple family of increasing trees can be defined in the following way. We start with a sequence of nonnegative numbers $(\varphi_k)_{k\geq 0}$ with $\varphi_0 > 0$. The sequence $(\varphi_k)_{k\geq 0}$ is called the degree-weight sequence. We assume that there exists a $k \geq 2$ with $\varphi_k > 0$. The degree-weight sequence is used to define the weight $w(T)$ of any ordered tree T by $w(T) := \prod_v \varphi_{d(v)}$, where v ranges over all vertices of T and $d(v)$ is the out-degree of v. Furthermore, $\mathcal{L}(T)$ denotes the set of different increasing

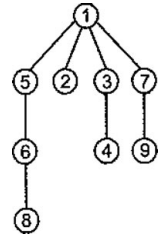


FIGURE 1. A size 9 recursive tree.

https://doi.org/10.1017/S0269964807000058 Published online by Cambridge University Press

labelings of the tree T with distinct integers $\{1, 2, \ldots, |T|\}$, where $|T|$ denotes the size of the tree T and $L(T) := |\mathcal{L}(T)|$ its cardinality. Then the family $\mathcal{T}$ consists of all trees T together with their weights $w(T)$ and the set of increasing labelings $\mathcal{L}(T)$.

For a given degree-weight sequence $(\varphi_k)_{k\geq 0}$ with a degree-weight generating function $\varphi(t) := \sum_{k\geq 0} \varphi_k t^k$, we now define the total weights by $T_n := \sum_{|T|=n} w(T) \cdot L(T)$. It follows then that the exponential generating function $T(z) := \sum_{n\geq 1} T_n(z^n/n!)$ satisfies the autonomous first-order differential equation

$$T'(z) = \varphi(T(z)), \quad T(0) = 0. \tag{1}$$

Remark 1: The formal definition of simple families of increasing trees is similar to the definition of simply generated trees, which are used as an underlying tree model. Note that simply generated trees are unlabeled trees, whereas simple families of increasing trees are increasingly labeled.

2.2. Characterization of Grown Simple Families of Increasing Trees

It turned out that a special subclass of simple families of increasing trees can be constructed via a tree evolution process. We will describe now the characterization of grown simple increasing tree families via the degree-weight generating function $\varphi(t)$ as obtained in [17].

LEMMA 1 ([17]): *A simple family of increasing trees $\mathcal{T}$ can be constructed via a tree evolution process and is thus a grown simple family of increasing trees iff the degree-weight generating function $\varphi(t) = \sum_{k\geq 0} \varphi_k t^k$ is given by one of the following three formulas (with constants $c_1, c_2 \in \mathbb{R}$):*

Case A (recursive trees):

$$\varphi(t) = \varphi_0 e^{c_1 t/\varphi_0} \quad \text{for } \varphi_0 > 0,\ c_1 > 0 \quad (\text{defining } c_2 := 0)$$

Case B (d-ary trees):

$$\varphi(t) = \varphi_0 \left(1 + \frac{c_2 t}{\varphi_0}\right)^d \quad \text{for } \varphi_0 > 0,\ c_2 > 0,\ d := \frac{c_1}{c_2} + 1 \in \{2, 3, 4, \ldots\}$$

Case C (generalized plane-oriented recursive trees):

$$\varphi(t) = \frac{\varphi_0}{\big(1 + (c_2 t/\varphi_0)\big)^{-(c_1/c_2)-1}} \quad \text{for } \varphi_0 > 0,\ 0 < -c_2 < c_1.$$

The tree evolution process that generates random trees of arbitrary size n is for grown simple families of increasing trees given as follows:

- *Step* 1: *The process starts with the root labeled by* 1.

https://doi.org/10.1017/S0269964807000058 Published online by Cambridge University Press

- *Step $i+1$: At step $i+1$, the node with label $i+1$ is attached to any previous node v (with out-degree $d(v)$) of the already grown tree of size i with probabilities $p(v)$ given by*

$$p(v) = \begin{cases} \dfrac{1}{i} & \text{for Case A} \\ \dfrac{d - d(v)}{(d-1)i + 1} & \text{for Case B} \\ \dfrac{d(v) + \alpha}{(\alpha+1)i - 1} \quad \text{with } \alpha := -1 - \dfrac{c_1}{c_2} > 0 & \text{for Case C.} \end{cases}$$

The constants c_1 and c_2 appearing in Lemma 1 come from an equivalent characterization of grown simple families of increasing trees. The total weights T_n of trees of size n of $\mathcal{T}$ satisfy for all $n \in \mathbb{N}$ the equation

$$\frac{T_{n+1}}{T_n} = c_1 n + c_2. \tag{2}$$

For grown simple increasing tree families, the constants c_1 and c_2 describe the growth of the total weights from a tree of size n to a tree of size $n+1$:

$$T_n = \varphi_0 \prod_{k=1}^{n-1} (c_1 k + c_2) = \varphi_0 c_1^{n-1} (n-1)! \binom{n - 1 + \frac{c_2}{c_2}}{n-1}. \tag{3}$$

For recursive trees, we set $c_2 = 0$, and for d-ary increasing trees, $d = (c_1/c_2) + 1$.

Since grown simple increasing trees with degree-weight generating function $\varphi(t)$ are a subclass of simple families of increasing trees, the generating function $T(z)$ of the total weights also satisfies the differential equation (1).

Solving either the differential equation (1) or using (2), one obtains the following explicit formulas for the exponential generating function $T(z)$:

$$T(z) = \begin{cases} \dfrac{\varphi_0}{c_1} \log\left(\dfrac{1}{1 - c_1 z}\right), & \text{Case A} \\ \dfrac{\varphi_0}{c_2} \left(\dfrac{1}{(1 - (d-1)c_2 z)^{1/(d-1)}} - 1\right), & \text{Case B} \\ \dfrac{\varphi_0}{c_2} \left(\dfrac{1}{(1 - c_1 z)^{c_2/c_1}} - 1\right), & \text{Case C.} \end{cases}$$

2.3. The Dickman Distribution and Generalizations

The Dickman function $\rho(u)$, which appears in analytic number theory, is defined as the continuous solution of the differential-difference equation

$$u\rho'(u) + \rho(u-1) = 0 \qquad (u > 1),$$

with initial condition $\rho(u) = 1$ for $0 \le u \le 1$ and with $\rho(u)$ differentiable on $(1, \infty)$. It

https://doi.org/10.1017/S0269964807000058 Published online by Cambridge University Press

is known that the Dickman function is positive and decreasing over the entire interval $(1, \infty)$. Note that

$$\int_0^\infty \rho(v)\, dv = e^\gamma,$$

where γ denotes Euler's constant. For simplicity of reference, we call the distribution with the density function $e^{-\gamma}\rho(v)$ the Dickman distribution.

Penrose and Wade introduced in [18] the *generalized Dickman distribution*. Given $\theta > 0$, a random variable X has a generalized Dickman distribution with shape parameter θ, or

$$X \stackrel{(d)}{=} \mathrm{GD}(\theta), \tag{4}$$

if it satisfies the distributional fixed-point identity

$$X \stackrel{(d)}{=} \mathrm{U}^{1/\theta}(1 + X),$$

where U is uniform on $(0, 1]$ and is independent of the X on the right. Some other known properties of the generalized Dickman distribution stated in [18] are listed as follows.

- If $X \stackrel{(d)}{=} \mathrm{GD}(\theta)$, then the Laplace transform of X is given by

$$\psi(t) = \mathbb{E}(e^{-tX}) = \exp\left(\theta \int_0^t \frac{e^{-s} - 1}{s}\, ds\right), \quad t \in \mathbb{R}.$$

- The $\mathrm{GD}(\theta)$ distribution is infinitely divisible.
- If $X \stackrel{(d)}{=} \mathrm{GD}(\theta)$, then the moments $\mathbb{E}(X^k)$ satisfy $\mathbb{E}(X^0) = 1$, and for integer $k \geq 1$,

$$\mathbb{E}(X^k) = \frac{\theta}{k} \sum_{j=0}^{k-1} \binom{k}{j} \mathbb{E}(X^j).$$

For $\theta = 1$, the $\mathrm{GD}(\theta)$ distribution is just the ordinary Dickman distribution. For more properties of the $\mathrm{GD}(\theta)$ distribution, see Penrose and Wade [18].

Now we state the main results of this article: the limit laws of the weighted depths and distances.

3. RESULTS FOR WEIGHTED DEPTHS

THEOREM 1: *The limiting distribution of the random variable* $(W_{n,j} - j)/j$ *is a generalized Dickman distribution with parameter* $\theta = 1 + (c_2/c_1) > 0$,

$$\lim_{j\to\infty} \frac{W_{n,j} - j}{j} \stackrel{(d)}{=} \mathrm{GD}\left(1 + \frac{c_2}{c_1}\right), \tag{5}$$

https://doi.org/10.1017/S0269964807000058 Published online by Cambridge University Press

or, equivalently, letting $\psi_j(t) := \mathbb{E}(e^{-t(W_{n,j}-j)})$, *we have*

$$\lim_{j\to\infty} \psi_j\left(\frac{t}{j}\right) = \exp\left(\left(1+\frac{c_2}{c_1}\right)\int_0^t \frac{e^{-v}-1}{v}\,dv\right) \tag{6}$$

for constants c_1 *and* c_2 *as in* (2).

Remark 2: Note that the random variable $W_{n,j}$ is independent of n and only depends on j: $W_{n,j} \overset{(d)}{=} W_{j,j}$. Therefore, we consider the limit $j\to\infty$.

For the reader's convenience, we separately present the limit laws for the most important subfamilies of grown simple increasing trees.

COROLLARY 1: *For the three most prominent tree families we obtain the following result*:

- *Recursive trees* ($c_1 = 1,\ c_2 = 0$): *The limit distribution of* $(W_{n,j}-j)/j$ *is Dickman*, GD(1):

$$\lim_{j\to\infty} \mathbb{P}\left\{\frac{W_{n,j}-j}{j} \le x\right\} = e^{-\gamma}\int_0^x \rho(v)\,dv, \qquad x>0. \tag{7}$$

- *Binary increasing trees* ($c_1 = c_2 = 1$): *The limit distribution of* $(W_{n,j}-j)/j$ *is the convolution of two Dickman distributions*, GD(2):

$$\lim_{j\to\infty} \mathbb{P}\left\{\frac{W_{n,j}-j}{j} \le x\right\} = e^{-2\gamma}\int_0^x \rho(v)\rho(x-v)\,dv, \qquad x>0. \tag{8}$$

- *Plane-oriented recursive trees* ($c_1 = 2,\ c_2 = -1$): *The limit distribution of* $(W_{n,j}-j)/j$ *is a generalized Dickman distribution* GD(θ) *with parameter* $\theta = 1/2$.

Next, we will state the main results for weighted distances.

4. RESULTS FOR WEIGHTED DISTANCES

THEOREM 2: *For fixed* j *and* $n\to\infty$ *the limiting distribution of the weighted distance* $\mathcal{W}_{n,j}$ *between node* j *and node* n *in a size* n *grown simple increasing tree is a generalized Dickman distribution with parameter* $\theta = 1 + (c_2/c_1) > 0$:

$$\lim_{n\to\infty} \frac{\mathcal{W}_{n,j}-n}{n} \overset{(d)}{=} \mathrm{GD}\left(1+\frac{c_2}{c_1}\right); \tag{9}$$

or, equivalently, $\mathbb{E}(e^{-t(\mathcal{W}_{n,j}-n)}) = \psi_n(t)$:

$$\lim_{j\to\infty} \psi_n\left(\frac{t}{n}\right) = \exp\left(\left(1+\frac{c_2}{c_1}\right)\int_0^t \frac{e^{-v}-1}{v}\,dv\right) \tag{10}$$

for constants c_1 *and* c_2 *as in* (2).

https://doi.org/10.1017/S0269964807000058 Published online by Cambridge University Press

THEOREM 3: *The limiting distribution of the weighted distance $\mathcal{W}_{n,j}$ between node j and node n in a grown simple families of increasing tree satisfying $c_1 = -2c_2$ depends on the growth of j.*

- *The region $j = o(n)$: The limiting distribution is a generalized Dickman distribution with parameter $\theta = 1/2$. Let $\mathbb{E}(e^{-t(\mathcal{W}_{n,j}-n-j)}) = \psi_n(t)$; then*

$$\lim_{n\to\infty} \psi_n\left(\frac{t}{n}\right) = \exp\left(\frac{1}{2}\int_0^t \frac{e^{-v}-1}{v}\, dv\right). \tag{11}$$

- *The region $j = \mu n$, with $0 < \mu < 1$: The limiting distribution can be characterized via its Laplace transform. Let $\mathbb{E}(e^{-t(\mathcal{W}_{n,j}-n-j)}) = \psi_n(t)$.*

$$\lim_{n\to\infty} \psi_n\left(\frac{t}{n}\right) = \exp\left(\frac{1}{2}\int_0^{\mu t} \frac{e^{-v}-1}{v}\, dv + \int_{\mu t}^t \frac{e^{-v}-1}{v}\, dv\right). \tag{12}$$

- *The region $n - j = o(n)$: The limiting distribution is a Dickman distribution.*

$$\lim_{n\to\infty} \mathbb{P}\left\{\frac{\mathcal{W}_{n,j}-j-n}{n} \le x\right\} = e^{-\gamma}\int_0^x \rho(v)\, dv, \qquad x > 0. \tag{13}$$

Remark 3: Basically, Theorem 2 just tells that for fixed j the random variable $\mathcal{W}_{n,j}$ behaves as the weighted depth $W_{n,1}$.

Unfortunately, we are not able to describe the limiting distribution for the full range of j for the all grown simple families of increasing trees; for example, up to now we are not able to determine the weighted distance between nodes $n-1$ and n for all grown simple families of increasing trees. We can only obtain the limit law for the subfamily admitting $c_1 = -2c_2$ (containing plane-oriented recursive trees), due to the specific decomposition of the underlying ordinary distance for this subfamily.

5. DERIVING THE DISTRIBUTION LAW

We will derive distributional decompositions for the weighted depths and distances by using the underlying decompositions of the ordinary depths and distances.

5.1. Weighted Depth

Dobrow and Smythe [6] have already provided a characterization of the ordinary depth, which can be adapted for describing the weighted depth in grown simple families of increasing trees (see also [10,17]).

LEMMA 2 (Dobrow and Smythe [6]): *The ordinary depth of node j in a size $n \ge j$ random grown simple increasing tree $D_{n,j}$ admits the following distribution law:*

$$D_{n,j} \overset{(d)}{=} D_{j,j} \overset{(d)}{=} \bigoplus_{k=1}^{j-1} B_k,$$

https://doi.org/10.1017/S0269964807000058 Published online by Cambridge University Press

where $B_k \overset{(d)}{=} \mathrm{Be}(p_k)$ *with* $p_k = (1 + (c_2/c_1))/(k + (c_2/c_1))$ *for* $1 \le k \le j-1$. *Note that* $B_k \overset{(d)}{=} \mathrm{Be}(p_k) \overset{(d)}{=} \mathbb{I}(A_k)$, *where* A_k *denotes the event that node k is on the path from the root to node j. Further, the* B_k*'s are independent.*

As an immediate consequence of Lemma 2 we obtain the distribution law of the weighted depth $W_{n,j}$ by simply multiplying all the indicators $\mathbb{I}(A_k)$ with k. This corresponds to counting the labels of all nodes on the unique path from j to the root.

PROPOSITION 1: *The weighted depth of node j in a size* $n \ge j$ *random grown simple increasing tree* $W_{n,j}$ *admits the following distribution law:*

$$W_{n,j} \overset{(d)}{=} W_{j,j} \overset{(d)}{=} j \oplus \bigoplus_{k=1}^{j-1} \mathcal{B}_k, \tag{14}$$

where $\mathcal{B}_k = \mathcal{B}_k(c_1, c_2)$ *are a sequence of independent random variables such that*

$$\mathbb{P}(\mathcal{B}_k = k) = \frac{1 + (c_2/c_1)}{k + (c_2/c_1)} \quad \text{and} \quad \mathbb{P}(\mathcal{B}_k = 0) = 1 - \frac{1 + (c_2/c_1)}{k + (c_2/c_1)}. \tag{15}$$

Note that $\mathcal{B}_k \overset{(d)}{=} k \cdot \mathbb{I}(A_k)$, *where, again,* A_k *denotes the event that node k is on the path from the root to node j. Further, the* $\mathcal{B}_k$*'s are independent.*

We also get as a by-product the expectation and variance of $W_{n,j}$.

COROLLARY 2: *The expectation and the variance of the random variable* $W_{n,j}$ *are given as follows*:

$$\begin{aligned}
\mathbb{E}(W_{n,j}) &= \left(1 + \frac{c_2}{c_1}\right)(j-1) - \frac{c_2}{c_1}\left(1 + \frac{c_2}{c_1}\right)(H_{j-1+(c_2/c_1)} - H_{c_2/c_1}), \\
\mathbb{V}(W_{n,j}) &= \frac{1 + (c_2/c_1)}{2} j(j-1) - \left(1 + \frac{c_2}{c_1}\right)\left(1 + 2\frac{c_2}{c_1}\right)(j-1) \\
&\quad + \left(1 + \frac{c_2}{c_1}\right)\frac{c_2}{c_1}\left(2 + 3\frac{c_2}{c_1}\right)(H_{j-1+(c_2/c_1)} - H_{c_2/c_1}) \\
&\quad - \frac{c_2^2}{c_1^2}\left(1 + \frac{c_2}{c_1}\right)^2 \left(H^{(2)}_{j-1+(c_2/c_1)} - H^{(2)}_{(c_2/c_1)}\right).
\end{aligned} \tag{16}$$

5.2. Weighted Distance

We will use the description for the ordinary distance $\Delta_{n,j}$ between n and j to derive the distribution for the weighted distance $\mathcal{W}_{n,j}$. Dobrow and Smythe provided the distribution law of the (ordinary) distance in [6].

https://doi.org/10.1017/S0269964807000058 Published online by Cambridge University Press

LEMMA 3 (Dobrow and Smythe [6]): *The distance of between node j and node n in a random grown simple increasing tree $\Delta_{n,j}$ admits the following distribution law:*

$$\Delta_{n,j} \stackrel{(d)}{=} \Delta_{j+1,j} \oplus \bigoplus_{k=j+1}^{n-1} B_k, \tag{17}$$

where $B_k \stackrel{(d)}{=} \mathrm{Be}(p_k)$ with $p_k = (1 + (c_2/c_1))/(k + (c_2/c_1))$ for $j + 1 \le k \le n - 1$.

Remark 4: The Bernoulli random variables B_k, appearing in Lemma 3, which characterize the distribution law of $\Delta_{n,j}$ in grown simple families of increasing trees, can be written as indicators $B_k \stackrel{(d)}{=} \mathbb{I}(A_k)$, where A_k denotes the event that node k is on the path from the node n to node j, which was shown in [10].

By using the decomposition of $\Delta_{n,j}$ in Lemma 3, we can simply multiply for $j + 1 \le k \le n - 1$ the indicators $\mathbb{I}(A_k)$ with k in order to obtain the following result.

PROPOSITION 2: *The weighted distance $\mathcal{W}_{n,j}$ between node j and node n in a size $n \ge j$ random grown simple increasing tree admits the following distribution law*:

$$\mathcal{W}_{n,j} \stackrel{(d)}{=} ((\mathcal{W}_{j+1,j}) - (j + 1)) \oplus \bigoplus_{k=j+1}^{n-1} \mathcal{B}_k \oplus n,$$

where the random variables $\mathcal{B}_k \stackrel{(d)}{=} k \cdot \mathbb{I}(A_k)$ and A_k denotes the event that node k is on the path from node n to node j. Further, the $\mathcal{B}_k$'s and $\mathcal{W}_{j+1,j}$ are independent.

Remark 5: Note that by directly translating Lemma 3, the weight of node $j + 1$ is counted one time too many. This leads to the correcting term $j + 1$.

We have not specified the distribution law of $\mathcal{W}_{j+1,j}$ yet. It turned out in [10] that in most cases, there is no simple description of $\Delta_{j+1,j}$, which can be used to determine the distribution of the weighted counterpart $\mathcal{W}_{j+1,j}$. For plane-oriented trees, we state the following result.

PROPOSITION 3: *For grown simple families of increasing trees with $c_1 = -2c_2$, the distribution law of $\mathcal{W}_{j+1,j}$ is given by*

$$\mathcal{W}_{j+1,j} \stackrel{(d)}{=} (2j + 1) \oplus \bigoplus_{k=1}^{j-1} \mathcal{C}_k \stackrel{(d)}{=} (2j + 1) \oplus \bigoplus_{k=1}^{j-1} k \cdot \mathbb{I}(\mathcal{A}_k), \tag{18}$$

where $\mathcal{C}_k$ are a sequence of independent random variables such that

$$\mathbb{P}(\mathcal{C}_k = k) = \frac{2}{2k+1} \quad \textit{and} \quad \mathbb{P}(\mathcal{C}_k = 0) = 1 - \frac{2}{2k+1}. \tag{19}$$

We denote by $\mathcal{A}_k$ the event that node k is on the path from node $j + 1$ to node j. The indicator variables $\mathbb{I}(\mathcal{A}_k)$'s are mutually independent.

https://doi.org/10.1017/S0269964807000058 Published online by Cambridge University Press

Note that a decomposition of $\Delta_{j+1,j}$ of the form

$$\Delta_{j+1,j} \stackrel{(d)}{=} 1 \oplus \sum_{k=1}^{j-1} \mathbb{I}(A_k) \tag{20}$$

is possible for arbitrary grown simple families of increasing trees, but only in the case $c_1 = -2c_2$ are the indicator variables mutually independent; for example, for recursive trees we get

$$\begin{aligned}\mathbb{P}\{A_{j-1}\} &= \mathbb{P}\{A_{j-1}|\Delta_{j,j-1} = 1\}\mathbb{P}\{\Delta_{j,j-1} = 1\} \\ &\quad + \mathbb{P}\{A_{j-1}|\Delta_{j,j-1} > 1\}\mathbb{P}\{\Delta_{j,j-1} > 1\} \\ &= \frac{j-1}{(j-1)j} + \frac{j-2}{(j-1)j} \\ &= \frac{2j-3}{j(j-1)}.\end{aligned}$$

Assuming that the A_k's are mutually independent, we get $\mathbb{P}\{A_k\} = (2k-1)/(k(k+1))$. However, it can easily be seen that $\mathbb{P}\{A_{j-1}\}\mathbb{P}\{A_{j-2}\} \neq \mathbb{P}\{A_{j-1}A_{j-2}\}$, which leads to a contradiction. Hence, for grown simple families of increasing trees with $c_1 \neq -2c_2$, we cannot specify the distribution of the weighted distance $\mathcal{W}_{j+1,j}$. However, it is obvious that for fixed j and $n \to \infty$, the contribution of $\mathcal{W}_{j+1,j}$ can be neglected due to the fact that

$$\mathcal{W}_{j+1,j} \leq 2j + 1 + \sum_{k=1}^{j-1} k = \frac{j(j+3)}{2} + 1.$$

For grown simple families of increasing trees admitting $c_1 = -2c_2$, we multiply the indicator functions $\mathbb{I}(A_k)$ with k, which gives Proposition 3.

6. DERIVING THE LIMITING DISTRIBUTION

6.1. Weighted Depth

First, we present an extension of a result of Hwang and Tsai [8, p. 4] concerning the generalized Dickman distribution.

LEMMA 4: *The limiting distribution of the random variable*

$$X_j = \bigoplus_{k=1}^{j} \mathcal{B}_k(c_1, c_2), \tag{21}$$

where the $\mathcal{B}_k$'s are defined as in Proposition 1, is for $j \to \infty$ asymptotically generalized Dickman GD(θ) *with parameter* $\theta = 1 + (c_2/c_1)$. *For* $c_1 = 1$ *and* $c_2 = 0$, *this just*

https://doi.org/10.1017/S0269964807000058 Published online by Cambridge University Press

tells that the limiting distribution of X_j is asymptotically Dickman:

$$\lim_{j\to\infty} \mathbb{P}\left\{\frac{X_j}{j} < x\right\} = e^{-\gamma}\int_0^x \rho(v)\,dv \quad (x>0).$$

The proof of Lemma 4 is completely analogous to the result of Hwang and Tsai ($c_1 = 1$, $c_2 = 0$) in [8].

We will follow the proof of Lemma 4 in [8] in order to prove Theorem 1. Let $\psi_j(t) := \mathbb{E}(e^{-t(W_{n,j}-j)})$ denote the Laplace transform of the shifted random variable $W_{n,j} - j$. We have

$$\begin{aligned}\phi_j(t) = \mathbb{E}(e^{-t(W_{n,j}-j)}) &= \prod_{1\le k\le j-1} \frac{k-1+e^{-tk}(1+(c_2/c_1))}{k+(c_2/c_1)}\\ &= \prod_{1\le k\le j-1}\left(1+\frac{(e^{-tk}-1)(1+(c_2/c_1))}{k+(c_2/c_1)}\right).\end{aligned}$$

It suffices, by Lévy's continuity theorem and (2.3), to show that

$$\begin{aligned}\lim_{j\to\infty}\phi_j\left(\frac{t}{j}\right) &= \lim_{j\to\infty}\mathbb{E}\left(e^{-t((W_{n,j}-j)/j)}\right)\\ &= \exp\left(\left(1+\frac{c_2}{c_1}\right)\int_0^t \frac{e^{-v}-1}{v}\,dv\right)\end{aligned}$$

for finite and real t. Now,

$$\begin{aligned}\phi_j\left(\frac{t}{j}\right) &= \exp\left(\log\left(\phi_j\left(\frac{t}{j}\right)\right)\right)\\ &= \exp\left(\sum_{k=1}^{j-1}\log\left(1+\frac{(e^{-tk}-1)(1+(c_2/c_1))}{k+(c_2/c_1)}\right)\right)\\ &= \exp\left(\left(1+\frac{c_2}{c_1}\right)\sum_{k=1}^{j-1}\frac{e^{-kt/j}-1}{k+(c_2/c_1)} + R_j(t)\right),\end{aligned}$$

where

$$R_j(t) := \sum_{l\ge 2}\frac{(-1)^{l-1}}{l}\sum_{k=1}^{j-1}\frac{((1+(c_2/c_1)))^l(e^{-kt/j}-1)^l}{(k+(c_2/c_1))^l} = \mathcal{O}\left(\frac{|t^2|}{j}\right).$$

Now, we apply the Euler–MacLaurin summation formula to show

$$\sum_{k=1}^{j-1}\frac{e^{-kt/j}-1}{k+(c_2/c_1)} = \int_0^t \frac{e^{-v}-1}{v}dv + \mathcal{O}\left(\frac{|t^2|}{j}\right).$$

https://doi.org/10.1017/S0269964807000058 Published online by Cambridge University Press

6.2. Weighted Distances

We only sketch the proofs for the weighted distances since they are similar to the proof in Subsection 6.1. Let $\psi_n(t) := \mathbb{E}(e^{-t(\mathcal{W}_{n,j}-j-n)})$ denote the Laplace transform of the shifted random variable $\mathcal{W}_{n,j} - j - n$.

$$\begin{aligned}\phi_n(t) &= \mathbb{E}(e^{-t(\mathcal{W}_{n,j}-j-n)}) \\ &= \mathbb{E}(e^{-t(\mathcal{W}_{j+1,j}-j)}) \prod_{j+1\le k\le n-1} \frac{k-1+e^{-tk}(1+(c_2/c_1))}{k+(c_2/c_1)} \\ &= \mathbb{E}(e^{-t(\mathcal{W}_{j+1,j}-j)}) \prod_{j+1\le k\le n-1} \left(1+\frac{(e^{-tk}-1)(1+(c_2/c_1))}{k+(c_2/c_1)}\right).\end{aligned}$$

Hence,

$$\phi_n\left(\frac{t}{n}\right) = \mathbb{E}\left(e^{-t(\mathcal{W}_{j+1,j}-j)/n}\right) \exp\left(\left(1+\frac{c_2}{c_1}\right) \times \sum_{k=j+1}^{n-1} \frac{e^{-tk}-1}{k+(c_2/c_1)} + R_{n,j}(t)\right),$$

where for fixed j, it can be seen that

$$R_{n,j}(t) := \mathcal{O}\left(\frac{|t^2|}{n}\right). \tag{22}$$

For fixed j, we use, again, the Euler–MacLaurin summation to show that

$$\sum_{k=j+1}^{n-1} \frac{e^{-tk/n}-1}{k+(c_2/c_1)} = \int_0^t \frac{e^{-v}-1}{v}dv + \mathcal{O}\left(\frac{|t^2|}{n}\right).$$

For grown simple families of increasing trees satisfying $c_1 = -2c_2$, we get

$$\phi_n(t) = \mathbb{E}(e^{-t(\mathcal{W}_{n,j}-j-n)}) = \left(\prod_{1\le k\le j-1}\left(1+\frac{e^{-tk}-1}{k+(1/2)}\right)\right) \times \left(\prod_{j+1\le k\le n-1}\left(1+\frac{(e^{-tk}-1)(1/2)}{k-(1/2)}\right)\right)$$

https://doi.org/10.1017/S0269964807000058 Published online by Cambridge University Press

and, further,

$$\phi_n\left(\frac{t}{n}\right) = \exp\left(\sum_{k=1}^{j-1}\frac{e^{-tk/n}-1}{k+(c_2/c_1)} + \frac{1}{2}\sum_{k=j+1}^{n-1}\frac{e^{-tk/n}-1}{k+(c_2/c_1)} + \hat{R}_{n,j}(t) + \tilde{R}_{n,j}(t)\right).$$

An application of the Euler–MacLaurin summation then provides Theorem 3.

7. CONCLUSION

We presented the limit laws of weighted depths and distances in increasing trees. Note that it is possible to set up an even more general (f-weighted) depth and distance by introducing a weight function $f(x)$ and defining

$$W_{j,j}^{[f]} \stackrel{(d)}{=} f(j) \oplus \bigoplus_{k=1}^{j-1} f(k) \cdot \mathbb{I}(A_k),$$

where A_k denotes the event that node k is on the path from j to the root. Furthermore,

$$\mathcal{W}_{n,j}^{[f]} \stackrel{(d)}{=} f(n) \oplus f(j) \oplus \bigoplus_{k=1}^{j-1} f(k) \cdot \mathbb{I}(\mathcal{A}_k) \oplus \bigoplus_{k=j+1}^{n-1} f(k) \cdot \mathbb{I}(\mathcal{A}_k),$$

where, for $1 \le k \le j-1$ and $j+1 \le k \le n-1$, $\mathcal{A}_k$ denotes the event that node k is on the path from j to node n; for example, $f(x) \equiv 1$ leads to the ordinary depth and distance (shifted by 1), where $f(x) = x$ leads to the weighted depth and weighted distance studied in this article. Hence, choosing $f(x) = x^\alpha$ would provide a natural transition between ordinary depths and weighted depths.

Acknowledgments

This work was supported by the Austrian Science Foundation FWF (grant S9608-N13). The authors thank an anonymous referee for many valuable comments.

References

1. Aguech, R., Lasmar, N., & Mahmoud, H. (2007). Extremal weighted path lengths in random binary search trees. *Probability in the Engineering and Informational Sciences* 21(1): 133–141.
2. Bergeron, F., Flajolet, P., & Salvy, B. (1992). *Varieties of increasing trees*. Lecture Notes in Computer Science, Vol. 581. Berlin: Springer-Verlag, pp. 24–48.
3. Bollobas, B. & Riordan, O.M. (2003). Mathematical results on scale-free random graphs. In Stefan Bornhold & Hans Geor (eds.), *Handbook of graphs and networks*. Weinheim: Wiley-VCH, pp. 1–34.
4. Devroye, L. & Neininger, R. (2004). Distances and finger search in random binary search trees. *SIAM Journal on Computing* 33: 647–658.

https://doi.org/10.1017/S0269964807000058 Published online by Cambridge University Press

5. Dobrow, R. (1996). On the distribution of distances in recursive trees. *Journal of Applied Probability* 33: 749–757.
6. Dobrow, R. & Smythe, R. (1996). Poisson approximations for functionals of random trees. *Random Structures and Algorithms* 9: 79–92.
7. Grübel, R. & Stefanoski, N. (2005). Mixed Poisson approximation of node depth distributions in random binary search trees. *Annals of Applied Probability* 15: 279–297.
8. Hwang, H.-K. & Tsai, T.-H. (2002). Quickselect and the Dickman function. *Combinatorics, Probability and Computing* 11(4): 353–371.
9. Kuba, M. (2006). Analysis of node isolation procedures and label-based parameters in tree structures. PhD thesis, Technische Universität Wien.
10. Kuba, M. & Panholzer, A. (2005). On the distribution of distances between specified nodes in increasing trees. http://www.info.tuwien.ac.at/panholzer/distances3.pdf.
11. Loève, M. (1977). *Probability theory I*, 4th ed. New York: Springer-Verlag.
12. Mahmoud, H. (1991). Limiting distributions for path lengths in recursive trees. *Probability in the Engineering and Informational Sciences* 5: 53–59.
13. Mahmoud, H., Modarres, R., & Smythe, R. (1995). Analysis of QUICKSELECT: An algorithm for order statistics. *RAIRO Theoretical Informatics and Applications* 29: 255–276.
14. Mahmoud, H. & Smythe, R. (1991). On the distribution of leaves in rooted subtrees of recursive trees. *Annals of Applied Probability* 1: 406–418.
15. Mahmoud, H. & Smythe, R. (1995). A survey of recursive trees. *Theoretical Probability and Mathematical Statistics* 51: 1–37.
16. Meir, A. & Moon, J.W. (1978). On the altitude of nodes in random trees. *Canadian Journal of Mathematics* 30: 997–1015.
17. Panholzer, A. & Prodinger, H. (2006). The level of nodes in increasing trees revisited. http://www.info.tuwien.ac.at/panholzer/levelincrev.pdf.
18. Penrose, M.D. & Wade, A.R. (2004). Random minimal directed spanning trees and Dickman-type distributions. *Advances in Applied Probability* 36: 691–714.

https://doi.org/10.1017/S0269964807000058 Published online by Cambridge University Press