J. Appl. Prob. **48,** 1060–1080 (2011)
Printed in England

LIMIT THEOREMS FOR DEPTHS AND DISTANCES IN WEIGHTED RANDOM *b*-ARY RECURSIVE TREES

GÖTZ OLAF MUNSONIUS * AND
LUDGER RÜSCHENDORF,** *University of Freiburg*

Abstract

Limit theorems are established for some functionals of the distances between two nodes in weighted random b-ary recursive trees. We consider the depth of the nth node and of a random node, the distance between two random nodes, the internal path length, and the Wiener index. As an application, these limit results imply, by an imbedding argument, corresponding limit theorems for further classes of random trees: plane-oriented recursive trees and random linear recursive trees.

Keywords: Random tree; Wiener index; path length; contraction method; plane-oriented recursive tree

2010 Mathematics Subject Classification: Primary 05C05; 60C05; 60F05

1. Introduction

In this paper we establish limit theorems for several basic functionals of the distances between nodes in weighted random b-ary recursive trees. We consider the depth of the nth node, the depth of a randomly chosen node, the distance between two randomly chosen nodes, the internal path length, i.e. the sum of all depths of nodes, and the Wiener index, i.e. the sum of all distances of pairs of nodes in the tree. All these functionals are well motivated and of importance for the structure of the tree and for the closely connected analysis of related algorithms (see, for example, Devroye and Neininger (2004), Mahmoud (1992), Mahmoud and Neininger (2003), and Su *et al.* (2006)). They have been studied in a wide variety of tree models.

In Szymański (1987) a procedure is introduced to also obtain nonuniform distributions on the set of recursive trees, i.e. trees which evolve through a step-by-step insertion of the nodes. This procedure operates by defining a weight function for each node in terms of its degree and attaching a new node randomly to a former node with a probability proportional to its weight. In Quintas and Szymański (1992) a weight function was used which yields trees with bounded degrees, so-called recursive f-trees. A slight modification of this tree model coincides with the b-ary increasing tree introduced in Bergeron *et al.* (1992).

The weighted random b-ary recursive tree is a combination of the b-ary increasing tree and the continuous-time model of b-ary trees introduced in Broutin and Devroye (2006). In the tree model of Broutin and Devroye (2006), a copy of a nonnegative vector $((Z_1, E_1), \dots, (Z_b, E_b))$ is attached independently to any node in an infinite b-ary tree. The components Z_i are random weights of the edges to the b children of a node, the entries E_i describe the lifetimes of

Received 11 May 2010; revision received 23 May 2011.
* Current address: Institute of Mathematics, University of Frankfurt, 60054 Frankfurt am Main, Germany.
Email address: munsonius@math.uni-frankfurt.de
** Postal address: Department of Mathematical Stochastics, University of Freiburg, Eckerstraße 1, 79104 Freiburg, Germany. Email address: ruschen@stochastik.uni-freiburg.de

the children. At time t the tree T_t is given by the set of all those nodes for which the sum of the lifetimes along the path to the root is smaller than t. By a proper choice of the lifetimes, this tree model without edge weights is close to being a random split tree and, thus, includes important families of trees, such as random m-ary search trees, quad-trees, and many others. Despite the bounded branching factor of these trees, it is possible to transfer properties of these weighted random b-ary trees to trees with unbounded branching factor, e.g. to random recursive trees, plane-oriented recursive trees, and random linear recursive trees, as introduced in Pittel (1994). If all lifetimes are independent and exponentially distributed and we consider the tree at the random moment where it has n nodes, due to the lack-of-memory property of the exponential distribution, the shape of the tree (i.e. the tree without the edge weights) coincides with the b-ary increasing tree, in which every external node has the same probability of becoming the next new internal node.

In Section 2 we introduce the weighted random b-ary trees together with some basic properties. In Section 3 we derive limit theorems for the depths of the nth node as well as for a randomly chosen node in the tree and for the distance between two randomly chosen nodes. In Section 4 we establish a limit theorem for the internal path length and the Wiener index based on a suitable two-dimensional recursion for their joint distribution by applying the contraction method. The main problem for the application of the contraction method to this problem is to derive a second-order expansion for the mean of the Wiener index. Finally, in Section 5 we obtain as a consequence of the limit theorems for weighted random b-ary trees corresponding limit results for plane-oriented recursive trees and linear recursive trees.

There are several related results in the literature for the depths and distances of random recursive trees (see Smythe and Mahmoud (1995) for a survey of early results for recursive trees). Limit theorems for the depth of the nth node are given in Devroye (1999) for random split trees and in Mahmoud (1992) for plane-oriented recursive trees. For the depths of a random node as well as for the distance between two random nodes, limit theorems are given in Panholzer and Prodinger (2004a), (2004b), Morris *et al.* (2004), Panholzer (2004a), (2004b), and Kuba and Panholzer (2010) for several random trees.

The internal path length of a tree has been studied for a large class of trees, including in particular random recursive trees, random m-ary search trees, and split trees (see Dobrow and Fill (1999), Rösler (1991), Neininger and Rüschendorf (1999), (2004), and others). The Wiener index has been investigated in Neininger (2002) for binary search trees and random recursive trees and in Janson (2003) for simply generated trees.

For several details and extensions of results in this paper, we refer the reader to the dissertation of Munsonius (2010) on which this paper is based.

We fix some notation for the rest of this paper. We use the notation $f \sim g$ for two functions f and g if $f(x)/g(x) \to 1$ for $x \to \infty$. For a real number x, the largest integer smaller than or equal to x is denoted by $\lfloor x \rfloor$. For random variables or distributions, we write '$\stackrel{\mathrm{D}}{=}$' for equality in distribution and $\mathcal{L}(X)$ for the distribution of X. By $N(0,1)$ we denote the standard normal distribution with expectation 0 and variance 1. The Wasserstein metric ℓ_2 is defined on the set of distributions on $\mathbb{R}^d$ by

$$\ell_2(\mu,\nu) := \inf\{\|X-Y\|_2 \colon \mathcal{L}(X)=\mu,\ \mathcal{L}(Y)=\nu\},$$

where the L_2-norm $\|\cdot\|_2$ is given by $\|X\|_2 = (\mathrm{E}[\|X\|^2])^{1/2}$. We denote convergence in distribution, convergence in probability, and convergence with respect to the L_2-metric by '$\stackrel{\mathrm{D}}{\to}$', '$\stackrel{\mathrm{P}}{\to}$', and '$\stackrel{L_2}{\longrightarrow}$', respectively. Let $\mathcal{M}_{0,2}^2$ be the set of centered probability measures on $\mathbb{R}^2$ with finite second moments.

https://doi.org/10.1239/jap/1324046019 Published online by Cambridge University Press

2. Random weighted b-ary recursive trees

The random b-ary recursive tree is a rooted, ordered, labeled tree where the outdegree is bounded by b and the labels along each path beginning at the root increase. We define this tree model by the following recursive procedure. We consider the infinite complete b-ary rooted, ordered tree and start with the root as the first internal node and its b children as external nodes. Given the random b-ary recursive tree with n internal nodes, the $(n+1)$th internal node is added in the following way. We choose a random node uniformly distributed on the set of all current external nodes, change it to an internal node and add the b children of this new node to the set of external nodes. Finally, the nodes are labeled in the order of their appearance.

Remark 2.1. Considering the above insertion rule, the parent u of the nth internal node is chosen with probability proportional to $b - \deg(u)$, where $\deg(u)$ is the number of internal children of node u in the tree with $n-1$ nodes and each of the $\deg(u)+1$ possible positions for the new node are equally likely. In Panholzer and Prodinger (2007) and Kuba and Panholzer (2010) it was shown that this tree is the same as the b-ary increasing tree, which belongs to the simple families of increasing trees introduced in Bergeron *et al.* (1992). In Drmota (2009, Section 1.3.3) this tree is also called the b-ary recursive tree.

It is well known that, for $b = 2$, the b-ary recursive tree is isomorphic to the random binary search tree.

The random b-ary recursive tree can also be defined as uniformly distributed on the set of ordered, labeled, rooted b-ary trees where the labels increase along each path beginning at the root. Note that in this class we have to distinguish trees where the nodes are in different positions, i.e. a tree where a node is at the leftmost position is not identical to the tree where this node is at the second position from the left also in the case that there are no other siblings of this node. The equivalence of the distributions is already mentioned in Stanley (1997) for the binary case (i.e. $b = 2$). For the general case, this can be seen by induction on the size of the tree (see Munsonius (2010)).

Now, we introduce edge weights. Let $Z := (Z_1, \dots, Z_b) \in \mathbb{R}^b_{\geq 0}$ be a random vector with nonnegative entries and attach to every node u of the complete infinite b-ary tree an independent copy $Z^{(u)}$ of Z. We consider the entries of $Z^{(u)}$ as weights of the edges from u to its b children. If all the $Z^{(u)}$ are independent of T_n, we refer to T_n supplied with the family $\{Z^{(u)}\}$ as a *random b-ary recursive tree with edge weights Z*.

While the entries of the vector Z may depend on each other, we assume throughout this paper that they are identically distributed, i.e. for all $i, j \in \{1, \dots, b\}$, we have

$$Z_i \stackrel{\mathrm{D}}{=} Z_j.$$

This assumption is not restrictive for the intended limit theorems as can be seen by a permutation argument (see Munsonius (2010, pp. 14–15)). Furthermore, we assume that $\mu := \mathrm{E}[Z_1]$ and $0 \leq \sigma^2 := \mathrm{var}(Z_1) < \infty$.

Given a random b-ary recursive tree with weighted edges, we denote by $T_{n,1}, \dots, T_{n,b}$ the subtrees rooted at the children of the root from left to right. Let $I_{n,j} := |T_{n,j}|$ be the number of internal nodes in the subtree $T_{n,j}$, and let $I_n := (I_{n,1}, \dots, I_{n,b})$ be the vector of the subtree sizes. For the edge weight of the edge between the root of T_n and the root of $T_{n,i}$, we write Z_i instead of $Z_i^{(0)}$. From the definition we see that, conditioned upon their sizes, the subtrees are again independent, b-ary recursive trees. This property of T_n is fundamental when using the contraction method.

https://doi.org/10.1239/jap/1324046019 Published online by Cambridge University Press

The subtree sizes $I_n = (I_{n,1}, \dots, I_{n,b})$ of a random b-ary recursive tree can be described by a Pólya urn with b colors, starting with one ball of each color, where each drawn ball is returned to the urn with $b-1$ additional balls of the same color. Then, the number of drawings of one color corresponds to the number of internal nodes in the corresponding subtree. We summarize some well-known results needed later (see, e.g. Johnson and Kotz (1977, Sections 4.5.1 and 6.3.3)). The explicit formula for the distribution of the subtree size is given by

$$\mathrm{P}(I_{n+1,1} = k) = \frac{1}{b-1} \frac{\Gamma(k + 1/(b-1))}{\Gamma(k+1)} \frac{\Gamma(n+1)}{\Gamma(n+1+1/(b-1))}. \tag{2.1}$$

The first and second moments are

$$\mathrm{E}[I_{n,1}] = \frac{1}{b} n, \qquad \mathrm{E}[I_{n,1}^2] = \frac{1}{2b-1} n^2 + \frac{b-1}{b(2b-1)} n, \quad \text{and} \quad \mathrm{E}[I_{n,1} I_{n,2}] = \frac{n(n-1)}{b(2b-1)}. \tag{2.2}$$

For the normalized subtree sizes, we have $I_n/n \to (D_1, \dots, D_b) =: D$ almost surely, where D is a Dirichlet $\beta_{(1/(b-1),\dots,1/(b-1))}$ distributed random vector, with parameters $(1/(b-1), \dots, 1/(b-1))$ (see, e.g. Athreya (1969)).

Furthermore, we have the asymptotic expansions

$$\mathrm{E}[I_{n,1} \log I_{n,1}] = \frac{1}{b} n \log n - \frac{b-1}{b^2} n + o(n) \tag{2.3}$$

and

$$\mathrm{E}[I_{n,1}^2 \log I_{n,1}] = \frac{1}{2b-1} n^2 \log n - \frac{b-1}{(2b-1)^2} n^2 + o(n^2) \tag{2.4}$$

as $n \to \infty$. For details and proofs of (2.3) and (2.4), see Munsonius (2010).

3. Limit theorems for depths and distances

In this section we consider the depth of one (random) node and the distance between two random nodes in a b-ary recursive tree with edge weights. The (weighted) depth of a node is given by the sum of the edge weights along the unique path from the root to that node. The (weighted) distance between two nodes is defined in the same way as the sum of the edge weights along the unique path between these nodes.

With the aid of a central limit theorem given in Javanian and Vahidi-Asl (2006) (see also Kuba and Panholzer (2010)) for the unweighted depth of the nth node, in Theorem 3.1 below we obtain the central limit theorem for the weighted depth of the nth node. We then use this result to derive the central limit theorem of a randomly chosen node D_U in Corollary 3.1 below. The result of Javanian and Vahidi-Asl (2006) corresponds to the case of all edge weights being 1, i.e. $\mu = 1$ and $\sigma^2 = 0$.

Theorem 3.1. (Central limit theorem for D_n.) *Let D_n be the weighted depth of the node with label n in a random b-ary recursive tree with edge weights Z and $0 \le \sigma^2 = \mathrm{var}(Z_1) < \infty$. Then we have, for $n \to \infty$,*

$$\mathrm{E}[D_n] \sim \mu \frac{b}{b-1} \log n \quad \textit{and} \quad \mathrm{var}(D_n) \sim (\mu^2 + \sigma^2) \frac{b}{b-1} \log n. \tag{3.1}$$

https://doi.org/10.1239/jap/1324046019 Published online by Cambridge University Press

Furthermore, for $n \to \infty$, it holds that

$$\frac{D_n - b\mu \log n/(b-1)}{\sqrt{(\sigma^2+\mu^2) b \log n/(b-1)}} \xrightarrow{\mathrm{D}} N(0,1).$$

Proof. Let $\tilde{D}_n$ be the depth of the node with label n in a random b-ary recursive tree with constant edge weights $(1, \dots, 1)$. The weighted depth D_n is the sum of independent, identically distributed random variables, as the path to the root never contains two nodes at the same level. So, for independent copies $\tilde{Z}_k$ of Z_1, we have

$$D_n \stackrel{\mathrm{D}}{=} \sum_{k=0}^{\tilde{D}_n - 1} \tilde{Z}_k.$$

Since $\tilde{D}_n$ is independent of the summands, Wald's equation yields $\mathrm{E}[D_n] = \mu\, \mathrm{E}[\tilde{D}_n]$, and by direct calculation we obtain $\mathrm{var}(D_n) = \mu^2 \mathrm{var}(\tilde{D}_n) + \sigma^2 \mathrm{E}[\tilde{D}_n]$. Thus, the claims for the expectation and variance in (3.1) follow from the results of Javanian and Vahidi-Asl (2006) for $\tilde{D}_n$.

Now, let $x_n = b \log n/(b-1)$, $f(x,y) = \sqrt{x^2/(x^2+y^2)}$, and $Z_i^* := (\tilde{Z}_i - \mu)/\sigma$. Then we obtain the representation

$$\begin{aligned}
\frac{D_n - b\mu \log n/(b-1)}{\sqrt{(\sigma^2+\mu^2) b \log n/(b-1)}} &\stackrel{\mathrm{D}}{=} \frac{\sum_{k=0}^{\tilde{D}_n-1} \tilde{Z}_k - b\mu \log n/(b-1)}{\sqrt{(\sigma^2+\mu^2) b \log n/(b-1)}} \\
&= f(\sigma,\mu) \sqrt{\frac{\lfloor x_n \rfloor}{x_n}} \frac{1}{\sqrt{\lfloor x_n \rfloor}} \sum_{k=0}^{\lfloor x_n \rfloor - 1} \tilde{Z}_k^* + f(\mu,\sigma) \frac{\tilde{D}_n - x_n}{\sqrt{x_n}} \\
&\quad + f(\sigma,\mu) \frac{1}{\sqrt{x_n}} \left(\sum_{k=0}^{\tilde{D}_n - 1} \tilde{Z}_k^* - \sum_{k=0}^{\lfloor x_n \rfloor - 1} \tilde{Z}_k^* \right). \qquad (3.2)
\end{aligned}$$

In the proof of the central limit theorem of Doeblin–Anscombe in Chow and Teicher (1997, Section 9.4), it was shown that, for $n \to \infty$, the last term of (3.2) converges to 0 in probability. Since the first two terms on the right-hand side of (3.2) are independent and both converge in distribution to normal distributions with variances $f(\mu,\sigma)^2$ and $f(\sigma,\mu)^2$, respectively, we obtain, for independent standard normal distributed random variables N and N',

$$\frac{D_n - b\mu \log n/(b-1)}{\sqrt{(\sigma^2+\mu^2) b \log n/(b-1)}} \xrightarrow{\mathrm{D}} \sqrt{\frac{\sigma^2}{\sigma^2+\mu^2}} N + \sqrt{\frac{\mu^2}{\sigma^2+\mu^2}} N' \stackrel{\mathrm{D}}{=} N(0,1).$$

Now we can transfer this result to the depth of a uniformly distributed node. For the unweighted case, this result was proved in Panholzer and Prodinger (2004a) using generating functions.

Corollary 3.1. (Central limit theorem for D_U.) *Let U_n be uniformly distributed on $\{1, \dots, n\}$, and let D_{U_n} be the weighted depth of the node with label U_n in a random b-ary recursive tree with edge weights Z and $0 \le \sigma^2 < \infty$. Then we have, for $n \to \infty$,*

$$\frac{D_{U_n} - b\mu \log n/(b-1)}{\sqrt{(\sigma^2+\mu^2) b \log n/(b-1)}} \xrightarrow{\mathrm{D}} N(0,1).$$

https://doi.org/10.1239/jap/1324046019 Published online by Cambridge University Press

Proof. Let $\varepsilon \in (0, \frac{1}{2})$ and $I_\varepsilon := [\varepsilon n, n]$. For $k \in I_\varepsilon$, we have $|\log(k/n)| \le -\log \varepsilon$ and

$$1 = \lim_{n\to\infty} \frac{\log \varepsilon + \log n}{\log n} \le \lim_{n\to\infty} \frac{\log k}{\log n} \le 1.$$

Together with Theorem 3.1 this yields, for $n \to \infty$,

$$\begin{aligned}\frac{D_k - b\mu \log n/(b-1)}{\sqrt{(\sigma^2+\mu^2) b \log n/(b-1)}} &= \underbrace{\sqrt{\frac{\log k}{\log n}}}_{\to 1} \frac{D_k - b\mu \log k/(b-1)}{\sqrt{(\sigma^2+\mu^2) b \log k/(b-1)}} \\ &\quad + \underbrace{\frac{b\mu \log(k/n)/(b-1)}{\sqrt{(\sigma^2+\mu^2) b \log n/(b-1)}}}_{\to 0} \\ &\xrightarrow{\mathrm{D}} N(0,1). \end{aligned} \tag{3.3}$$

Since $\mathrm{P}(U_n \in I_\varepsilon) \ge 1 - \varepsilon$, the convergence in (3.3) yields

$$\liminf_{\varepsilon\to 0} \lim_{n\to\infty} \mathrm{P}\left(\frac{D_{U_n} - b\mu \log n/(b-1)}{\sqrt{(\sigma^2+\mu^2) b \log n/(b-1)}} \le x,\ U_n \in I_\varepsilon \right) \to \mathrm{P}(N \le x)$$

for a standard normal distributed random variable N. The claim follows with

$$\limsup_{\varepsilon\to 0} \limsup_{n\to\infty} \mathrm{P}\left(\frac{D_{U_n} - b\mu \log n/(b-1)}{\sqrt{(\sigma^2+\mu^2) b \log n/(b-1)}} \le x,\ U_n \notin I_\varepsilon \right) \le \limsup_{\varepsilon\to 0} \mathrm{P}(U_n \notin I_\varepsilon) = 0.$$

We now turn to the distance between two random nodes. In the unweighted case, the central limit theorem is proved by using generating functions given in Panholzer and Prodinger (2004a). We give a short self-contained proof of this result that is based on a simple stochastic argument which traces the problem back to the depth of random nodes.

The distance is given by the sum of the edge weights along the unique path between these nodes. This path can be found by starting at each node and going up in the tree until the two paths coincide. The node at which the two paths meet is called the last common ancestor (LCA) of the nodes.

The key idea is to express the distance as the sum of the respective depths of the two nodes minus two times the depth of the LCA. We first show that the depth of the LCA is bounded in probability. A similar idea was used in the recent thesis of Ryvkina (2008) in the case of random split trees.

Lemma 3.1. (Depth of the LCA.) *Let $\tilde{U}_n$ and $\tilde{V}_n$ be two independent random variables uniformly distributed on $\{1, \dots, n\}$. Denote by $R(n)$ the (unweighted) depth of the LCA of the nodes U_n and V_n with labels $\tilde{U}_n$ and $\tilde{V}_n$, respectively, in a random b-ary recursive tree of size n. For any real sequence f_n with $f_n \to \infty$, we have, as $n \to \infty$,*

$$\frac{R(n)}{f_n} \xrightarrow{\mathrm{P}} 0.$$

Proof. Let $\mathrm{E}[I_{n,1}] = \alpha_1 n$ and $\mathrm{E}[I_{n,1}^2] = \alpha_2 n^2 + \alpha_3 n$ with $\alpha_i \in \mathbb{R}$. First, we note that, for $m \ge 0$,

$$\mathrm{P}(R(n) \ge m) = (b\alpha_2)^m + r(m,n), \tag{3.4}$$

where $r(m,n) \le m(\max\{\alpha_1, \alpha_2, \alpha_3\} b)^m / n$.

https://doi.org/10.1239/jap/1324046019 Published online by Cambridge University Press

This can be seen in the following way. If we have $R(n) \geq m+1$, both nodes have to lie in the same subtree and the depth of the LCA related to this subtree has to be greater than m. Conditioned upon the sizes of the subtrees, the depth of the LCA related to the subtree with size k_i is distributed as $R(k_i)$. We obtain

$$\begin{aligned}\mathrm{P}(R(n) \geq m+1) &= \sum_{k \in \mathbb{N}_0^b} \sum_{i=1}^{b} \mathrm{P}(R(n) \geq m+1,\ U_n, V_n \in T_{n,i} \mid I_n = k)\, \mathrm{P}(I_n = k) \\ &= \sum_{k \in \mathbb{N}_0^b} \sum_{i=1}^{b} \left(\frac{k_i}{n}\right)^2 \mathrm{P}(R(k_i) \geq m)\, \mathrm{P}(I_n = k).\end{aligned}$$

Equation (3.4) can now be proved by induction on m. In our case we have $\max\{\alpha_1, \alpha_2, \alpha_3\} = 1/b$. This yields, for every $\varepsilon > 0$ and any sequence f_n with $f_n \to \infty$ and $f_n = o(n)$,

$$\mathrm{P}(R(n) \geq \varepsilon f_n) \leq (b\alpha_2)^{\varepsilon f_n} + \frac{\varepsilon f_n}{n} \to 0$$

since $0 < \alpha_2 < 1/b$. Then surely $\mathrm{P}(R(n) \geq \varepsilon f_n) \to 0$ also holds for any sequence $f_n \to \infty$.

Lemma 3.2. *Let $\tilde{U}_n$ and $\tilde{V}_n$ be two independent random variables uniformly distributed on $\{1, \ldots, n\}$, and let $\tilde{\Delta}_{U_n,V_n}$ be the (unweighted) distance between the nodes U_n and V_n with labels $\tilde{U}_n$ and $\tilde{V}_n$, respectively, in a random b-ary recursive tree of size n. Then we have, for $n \to \infty$,*

$$\frac{\tilde{\Delta}_{U_n,V_n} - 2b \log n/(b-1)}{\sqrt{2b \log n/(b-1)}} \xrightarrow{\mathrm{D}} N(0,1).$$

Proof. The unweighted distance between U_n and V_n is given by

$$\tilde{\Delta}_{U_n,V_n} = \tilde{D}'_{U_n} + \tilde{D}'_{V_n},$$

where $\tilde{D}'_{U_n} = \tilde{D}_{U_n} - R(n)$ is the unweighted distance between U_n and the LCA of U_n and V_n, and $\tilde{D}'_{V_n}$ is defined similarly. Since $\tilde{D}'_{U_n}$ and $\tilde{D}'_{V_n}$ are independent by the construction of the tree, the claim follows by application of Lemma 3.1 and Corollary 3.1.

Equipped with these preliminaries, we now obtain the central limit theorem for the distance between two uniformly distributed nodes in random weighted b-ary recursive trees.

Theorem 3.2. (Central limit theorem for the distance.) *Let $\tilde{U}_n$ and $\tilde{V}_n$ be two independent random variables uniformly distributed on $\{1, \ldots, n\}$, and let Δ_{U_n,V_n} be the distance between the nodes U_n and V_n with labels $\tilde{U}_n$ and $\tilde{V}_n$, respectively, in a random b-ary recursive tree of size n with edge weights Z where $\mathrm{var}(Z_1) = \sigma^2 \in [0, \infty)$. Then we have, for $n \to \infty$,*

$$\frac{\Delta_{U_n,V_n} - 2b\mu \log n/(b-1)}{\sqrt{2(\sigma^2 + \mu^2) b \log n/(b-1)}} \xrightarrow{\mathrm{D}} N(0,1),$$

where $\mu = \mathrm{E}[Z_1]$.

Proof. We prove the claim analogously to the proof of Theorem 3.1. We make use of the fact that the weighted distance is given by the sum of the edge weights along the path between U_n and V_n. This path consists of $\tilde{\Delta}_{U_n,V_n}$ edges, as given in Lemma 3.2. Except for the two edges

https://doi.org/10.1239/jap/1324046019 Published online by Cambridge University Press

which belong to the LCA of U_n and V_n, the edge weights in the sum building the weighted distance are independent. Hence, we represent Δ_{U_n,V_n} as

$$\Delta_{U_n,V_n} \stackrel{\mathrm{D}}{=} \sum_{i=1}^{\tilde{\Delta}_{U_n,V_n}-2} \tilde{Z}_i + \hat{Z}_1 + \hat{Z}_2,$$

where $\tilde{Z}_1, \dots, \tilde{Z}_n, (\hat{Z}_1, \hat{Z}_2)$ are independent and $\tilde{Z}_i \stackrel{\mathrm{D}}{=} Z_1$.

Using the same arguments as in the proof of Theorem 3.1 as well as Lemma 3.2, we conclude the proof. The additional term $\hat{Z}_1 + \hat{Z}_2$ vanishes due to the scaling.

4. The internal path length and the Wiener index

The internal path length of a tree is the sum of the depths of all nodes. The Wiener index of a tree is the sum of all distances between pairs of nodes. We denote by P_n the internal path length and by W_n the Wiener index of a random b-ary recursive tree of size n with weighted edges.

The vector consisting of the Wiener index and the internal path length satisfies a recursion formula in dimension two. We will use this recursion to establish a limit theorem via the contraction method. Since we apply the contraction theorem in L_2, we have to center this vector. Therefore, we have to derive an asymptotic expansion of the expectation of the internal path length and of the Wiener index. The expectation of the internal path length is given in Bergeron *et al.* (1992) for the unweighted tree. It can also be obtained by summing up the exact expectations for the unweighted depths given in Javanian and Vahidi-Asl (2006).

Unlike for the internal path length, it seems that there is no simple way currently available to determine the expectation of W_n directly. In Roura (2001), the asymptotic expansion of a certain class of recursively defined sequences was proved. We show that the expectation of the Wiener index belongs to this class and we accordingly obtain the needed asymptotic of the expectation.

Lemma 4.1. *Let (W_n, P_n) be the vector containing the Wiener index W_n and the internal path length P_n of a random b-ary recursive tree of size n with edge weights Z. Then we have the recursion formula*

$$\begin{pmatrix} W_n \\ P_n \end{pmatrix} \stackrel{\mathrm{D}}{=} \sum_{i=1}^{b} \begin{bmatrix} 1 & n - I_{n,i} \\ 0 & 1 \end{bmatrix} \begin{pmatrix} W_{I_{n,i}}^{(i)} \\ P_{I_{n,i}}^{(i)} \end{pmatrix} + b(n) \tag{4.1}$$

with

$$b(n) = \begin{pmatrix} \displaystyle\sum_{i=1}^{b} Z_i I_{n,i} + \frac{1}{2} \sum_{i \neq j} (Z_i + Z_j) I_{n,i} I_{n,j} \\ \displaystyle\sum_{i=1}^{b} Z_i I_{n,i} \end{pmatrix}, \tag{4.2}$$

where $(Z_1, \dots, Z_b), (W_n, P_n), (W_n^{(1)}, P_n^{(1)}), \dots, (W_n^{(b)}, P_n^{(b)})$ are independent and

$$(W_n^{(i)}, P_n^{(i)}) \stackrel{\mathrm{D}}{=} (W_n, P_n) \quad \textit{for } i \in \{1, \dots, b\}.$$

Proof. Let T_n be a random b-ary recursive tree with weighted edges. By $P_{n,i}$ we denote the internal path length of $T_{n,i}$. For $u \in T_{n,i}$, let $D_u^{(i)}$ be the depth of node u in $T_{n,i}$. Thus, $D_u^{(i)}$ is the

https://doi.org/10.1239/jap/1324046019 Published online by Cambridge University Press

sum of the weights of the edges along the path from node u to node i. Obviously, $D_u^{(i)} + Z_i = D_u$. So we obtain

$$P_n = \sum_{i=1}^{b} (P_{n,i} + Z_i I_{n,i}). \tag{4.3}$$

The Wiener index is given by

$$W_n := \sum_{\substack{\{u,v\} \subset T_n \\ u \neq v}} \Delta_{u,v},$$

where $\Delta_{u,v}$ is the weighted distance between u and v. We distinguish two cases—either the nodes u and v lie in the same or in different subtrees of the root—and we rewrite the Wiener index as

$$W_n = \sum_{i=1}^{b} \sum_{\{u,v\} \in T_{n,i}} \Delta_{u,v} + \frac{1}{2} \sum_{i \neq j} \sum_{\substack{u \in T_{n,i} \\ v \in T_{n,j}}} \Delta_{u,v} + \sum_{u \neq 0} \Delta_{0,u}.$$

For $u \in T_{n,i}$ and $v \in T_{n,j}$ with $i \neq j$, we have $\Delta_{u,v} = D_u^{(i)} + D_v^{(j)} + Z_i + Z_j$. Summing this equation over $u \in T_{n,i}$ and $v \in T_{n,j}$ we obtain

$$\sum_{\substack{u \in T_{n,i} \\ v \in T_{n,j}}} \Delta_{u,v} = I_{n,j} P_{n,i} + I_{n,i} P_{n,j} + (Z_i + Z_j) I_{n,i} I_{n,j}.$$

With $\sum_{i \neq j} I_{n,j} = n - 1 - I_{n,i}$ and (4.3), we obtain

$$W_n = \sum_{i=1}^{b} (W_{n,i} + (n - I_{n,i}) P_{n,i}) + \sum_{i=1}^{b} Z_i I_{n,i} + \frac{1}{2} \sum_{i \neq j} (Z_i + Z_j) I_{n,i} I_{n,j}.$$

The claim follows since the subtrees are (conditioned upon their sizes) independent random b-ary recursive trees.

In order to apply the contraction theorem to the vector (W_n, P_n), we have to identify the expectation. In Bergeron *et al.* (1992) the first- and second-order terms of the expectation of the internal path length of b-ary recursive trees without edge weights are determined. Since the edge weights and the shape of the tree are independent, using Wald's equation, we obtain the following lemma.

Lemma 4.2. *Let P_n be the internal path length of a random b-ary recursive tree with edge weights Z. Then there exists a constant $c_p \in \mathbb{R}$ such that, for $n \to \infty$,*

$$\mathrm{E}[P_n] = \frac{b}{b-1} \mu n \log n + c_p n + o(n). \tag{4.4}$$

Remark 4.1. Lemma 4.2 can also be proved by a direct calculation using an exact formula for the expectation of the unweighted depth given in Javanian and Vahidi-Asl (2006). We then obtain the constant c_p in (4.4) in terms of an infinite series. As remarked by a reviewer, this series can be expressed in a closed form by using the psi function (also called the digamma function), which is the logarithmic derivative of the gamma function, i.e. $\psi(u) = \Gamma'(u)/\Gamma(u)$, and we obtain

$$c_p = \frac{\mu}{b-1} \left(-1 - b\psi\left(\frac{2b-1}{b-1} \right) \right).$$

https://doi.org/10.1239/jap/1324046019 Published online by Cambridge University Press

It remains to determine an expansion of the expectation of the Wiener index. From (4.1) we obtain

$$\mathrm{E}[W_n] = \sum_{i=1}^{b} \mathrm{E}[W_{I_{n,i}}^{(i)}] + \mathrm{E}[t(n)]$$

with

$$t(n) := \sum_{i=1}^{b} (n - I_{n,i}) P_{I_{n,i}}^{(i)} + b_1(n), \tag{4.5}$$

where $b_1(n)$ denotes the first entry of the vector $b(n)$ in (4.2). Since all subtrees and their sizes are identically distributed, the above equation can be simplified:

$$\mathrm{E}[W_n] = b \sum_{k=0}^{n-1} \mathrm{E}[W_k] \mathrm{P}(I_{n,1} = k) + \mathrm{E}[t(n)]. \tag{4.6}$$

There is no obvious way to solve this recurrence. However, to apply the contraction method, a second-order asymptotic expansion of the expectation is sufficient. In Roura (2001), certain recursions, as in (4.6), were considered and some sufficient conditions for the asymptotic expansion of the solution were identified. We need two notions from Roura (2001).

Definition 4.1. Let $\omega(z) \geq 0$ be a function on $[0, 1]$ such that $1 \leq \int_0^1 \omega(z)\,\mathrm{d}z < \infty$. Furthermore, assume that there is some $\mu < 0$ such that $\int_0^1 \omega(z) z^\mu \,\mathrm{d}z$ converges. Then we say that $\omega(z)$ is a shape function.

Definition 4.2. We say that

$$F_n = \begin{cases} b_n & \text{if } 0 \leq n < N, \\ t_n + \displaystyle\sum_{0 \leq k < n} \omega_{n,k} F_k & \text{if } n \geq N \end{cases} \tag{4.7}$$

is a 'continuous recursive definition' of F_n if and only if there exists some shape function $\omega(z)$, some constant $0 < q \leq 1$, and some function $M_n = \Theta(n^q)$ with integer values such that, with $z_{n,j} = j/M_n$, $0 \leq j \leq M_n$, $I_{n,j} = [z_{n,j} n, z_{n,j+1} n)$, $0 \leq j < M_n$, and

$$\varepsilon_{n,j} = \left| \sum_{k \in I_{n,j}} \omega_{n,k} - \int_{z_{n,j}}^{z_{n,j+1}} \omega(z)\,\mathrm{d}z \right|, \qquad 0 \leq j < M_n,$$

$$\sum_{0 \leq j < M_n} \varepsilon_{n,j} = O(n^{-\varrho}) \quad \text{for some } \varrho > 0.$$

One of the main conclusions of Roura (2001) is the following theorem.

Theorem 4.1. (Roura (2001, Theorem 3.3(1)).) *Let F_n be a function defined by a continuous recursive definition, and let $Bn^a \log^c n \cdot \xi_n$ be the main term of t_n, where $B > 0$, a and c are arbitrary constants, and $\xi_n = \mu_n$ or $\xi_n = 1/\mu_n$ for some sublogarithmical function μ_n. Let $\varphi(x) = \int_0^1 \omega(z) z^x \,\mathrm{d}z$ and $\mathcal{H} = 1 - \varphi(a)$. If $\mathcal{H} > 0$ then*

$$F_n \sim \frac{t_n}{\mathcal{H}}.$$

https://doi.org/10.1239/jap/1324046019 Published online by Cambridge University Press

To determine the asymptotic expansion of $\mathrm{E}[W_n]$ via this theorem, we have to find the asymptotic expansion of $\mathrm{E}[t(n)]$ and show that (4.6) is a continuous recursive definition of $\mathrm{E}[W_n]$.

Lemma 4.3. *For $N = 1$, $b_1 = 0$, and $\omega_{n,k} = bk/n\,\mathrm{P}(I_{n,1} = k)$, (4.7) is a continuous recursive definition with the shape function $\omega(z) = b/(b-1)z^{1/(b-1)}$.*

Proof. We set $M_n = n$, $b_0 = 0$, and $F_0 := 0$. It is clear that the given function $\omega(z)$ is a shape function.

For the proof, it is sufficient to show that

$$\sum_{k=0}^{n-1} \left| \omega_{n,k} - \int_{k/n}^{(k+1)/n} \omega(z)\,\mathrm{d}z \right| = O(n^{-1/(2(b-1))}). \tag{4.8}$$

For $k = 0$, we have $\int_0^{1/n} \omega(z)\,\mathrm{d}z = n^{-b/(b-1)}$. Thus, it suffices to consider the terms with $k \geq 1$. Since ω is increasing, we have

$$\frac{b}{b-1}\frac{1}{n}\left(\frac{k}{n}\right)^{1/(b-1)} \leq \int_{k/n}^{(k+1)/n} \omega(z)\,\mathrm{d}z \leq \frac{b}{b-1}\frac{1}{n}\left(\frac{k+1}{n}\right)^{1/(b-1)}.$$

This implies that

$$\begin{aligned}\left| \int_{k/n}^{(k+1)/n} \omega(z)\,\mathrm{d}z - \frac{b}{b-1}\frac{1}{n}\left(\frac{k}{n}\right)^{1/(b-1)} \right| &\leq \frac{b}{b-1}\frac{1}{n^{1+1/(b-1)}}((k+1)^{1/(b-1)} - k^{1/(b-1)}) \\ &\leq \frac{b}{b-1} n^{-(1+1/(b-1))},\end{aligned}$$

where we have used the fact that $(k+1)^{1/(b-1)} - k^{1/(b-1)} \leq 1$. With the triangle inequality we obtain

$$\left| \omega_{n,k} - \int_{k/n}^{(k+1)/n} \omega(z)\,\mathrm{d}z \right| \leq \left| \omega_{n,k} - \frac{b}{b-1}\frac{1}{n}\left(\frac{k}{n}\right)^{1/(b-1)} \right| + \frac{b}{b-1} n^{-(1+1/(b-1))}.$$

Using (2.1) and Stirling's formula for the gamma function, we obtain, by analytical computations,

$$\left| \omega_{n,k} - \frac{b}{b-1}\frac{1}{n}\left(\frac{k}{n}\right)^{1/(b-1)} \right| = O(n^{-1-1/(2(b-1))}) \tag{4.9}$$

for all $k \in \{1, \ldots, n-1\}$. Summing (4.9) over $k = 1, \ldots, n-1$ yields (4.8), completing the proof.

In order to use Theorem 4.1 to obtain the asymptotic behavior of $\mathrm{E}[W_n]$, it remains to identify the first-order term of $\mathrm{E}[t(n)]$ in (4.5). Using the fact that all subtrees are identically distributed, we obtain, from (4.5),

$$\mathrm{E}[t(n)] = b\,\mathrm{E}[(n - I_{n,1})P^{(1)}_{I_{n,1}}] + b\mu\,\mathrm{E}[I_{n,1}] + b(b-1)\mu\,\mathrm{E}[I_{n,1}I_{n,2}].$$

Since $I_{n,1} < n$, we have $b\mu\,\mathrm{E}[I_{n,1}] = o(n^2)$. By Lemma 4.2, there exists a function ε_1 with $\varepsilon_1(n) = o(n)$ for $n \to \infty$ such that

$$\mathrm{E}[P_n] = \frac{b}{b-1}\mu n \log n + c_p n + \varepsilon_1(n) \quad \text{for all } n \in \mathbb{N}.$$

https://doi.org/10.1239/jap/1324046019 Published online by Cambridge University Press

This yields

$$\begin{aligned}\mathrm{E}[(n-I_{n,1})P^{(1)}_{I_{n,1}}] &= \sum_{k=0}^{n-1}\mathrm{E}[(n-I_{n,1})P^{(1)}_{I_{n,1}} \mid I_{n,1}=k]\mathrm{P}(I_{n,1}=k)\\ &= n\left(\frac{b}{b-1}\mu\,\mathrm{E}[I_{n,1}\log I_{n,1}] + c_p\,\mathrm{E}[I_{n,1}] + \mathrm{E}[\varepsilon_1(I_{n,1})]\right)\\ &\quad - \left(\frac{b}{b-1}\mu\,\mathrm{E}[I^2_{n,1}\log I_{n,1}] + c_p\,\mathrm{E}[I^2_{n,1}]\right) + o(n^2).\end{aligned}$$

Since, almost surely, $I_{n,1}\to\infty$, we have $\mathrm{E}[\varepsilon_1(I_{n,1})]=o(n)$, and using (2.2), (2.3), and (2.4), we finally obtain

$$\mathrm{E}[t(n)] = \frac{b}{2b-1}\mu n^2\log n + \left(\frac{b-1}{2b-1}c_p - \frac{b^2-b}{(2b-1)^2}\mu\right)n^2 + o(n^2). \tag{4.10}$$

Combining these results, we obtain an asymptotic expansion of $\mathrm{E}[W_n]$ of second order.

Theorem 4.2. (Expectation of the Wiener index.) *Let W_n be the Wiener index of a random b-ary recursive tree of size n with edge weights Z. Then there exists a constant $c_w\in\mathbb{R}$ such that, for $n\to\infty$,*

$$\mathrm{E}[W_n] = \frac{b}{b-1}\mu n^2\log n + c_w n^2 + o(n^2).$$

Proof. It suffices to show that, for $n\to\infty$,

$$G_n := \frac{1}{n}\left(\mathrm{E}[W_n] - \frac{b}{b-1}\mu n^2\log n\right) \sim c_w n.$$

From (4.6), we obtain the recursion

$$G_n = \sum_{k=0}^{n-1}\omega_{n,k}G_k + s_n,$$

with $\omega_{n,k}$ as in Lemma 4.3 and

$$\begin{aligned}ns_n &= \mathrm{E}[t(n)] - \frac{b}{b-1}\mu n^2\log n + \sum_{k=0}^{n-1} b\,\mathrm{P}(I_{n,1}=k)\frac{b}{b-1}\mu k^2\log k\\ &= \mathrm{E}[t(n)] - \frac{b}{b-1}\mu n^2\log n + \frac{b^2}{b-1}\mu\,\mathrm{E}[I^2_{n,1}\log I_{n,1}].\end{aligned}$$

Using (2.3), (2.4), and (4.10), we also obtain

$$ns_n = \underbrace{\left(\frac{b-1}{2b-1}c_p - \frac{b}{2b-1}\mu\right)}_{=:\hat{c}}n^2 + o(n^2).$$

In short, we write $s_n=\hat{c}n+o(n)$.

Lemma 4.3 shows that G_n is defined by a continuous recursive definition. The main term of $s(n)$ is given by $\hat{c}n$. We set $\xi_n=1, a=1, c=0$, and $B=\hat{c}$. Then, $B<0$. To use Theorem 4.1,

https://doi.org/10.1239/jap/1324046019 Published online by Cambridge University Press

we need $B > 0$. Multiplying the recursion by -1 shows that Theorem 4.1 also works in the case $B < 0$. In the terminology of Roura (2001), we will show that $\mathcal{H} = 1 - \varphi(1) > 0$. Note that

$$\varphi(1) = \int_0^1 \frac{b}{b-1} z^{1/(b-1)+1} \, \mathrm{d}z = \frac{b}{2b-1} < 1.$$

Therefore, $\mathcal{H} = (b-1)/(2b-1) > 0$ and Theorem 4.1 yields $G_n \sim s(n)/\mathcal{H}$. Thus, we finally obtain the expansion

$$\mathrm{E}[W_n] - \frac{b}{b-1}\mu n^2 \log n = c_w n^2 + o(n^2)$$

with $c_w := c_p - b/(b-1)\mu$.

Upon determining the asymptotic expansion of the expectation, we now use the recursion formula for the vector consisting of the internal path length and the Wiener index to show a limit theorem via the contraction method.

Theorem 4.3. (Limit theorem for (W_n, P_n).) *Let (W_n, P_n) denote the vector of the Wiener index and internal path length of a random b-ary recursive tree of size n with random edge weights Z, where $\sigma^2 = \mathrm{var}(Z_1) < \infty$. Then we have*

$$\ell_2\left(\left(\frac{W_n - \mathrm{E}[W_n]}{n^2}, \frac{P_n - \mathrm{E}[P_n]}{n}\right), (W, P)\right) \to 0,$$

where (W, P) is the unique distributional fixed point of the map $T : \mathcal{M}_{0,2}^2 \to \mathcal{M}_{0,2}^2$ given for $\nu \in \mathcal{M}_{0,2}^2$ by

$$T(\nu) := \mathcal{L}\left(\sum_{i=1}^{b} \begin{bmatrix} D_i^2 & D_i(1-D_i) \\ 0 & D_i \end{bmatrix} \begin{pmatrix} X_1^{(i)} \\ X_2^{(i)} \end{pmatrix} + \begin{pmatrix} b_1^* \\ b_2^* \end{pmatrix}\right)$$

with

$$\begin{pmatrix} b_1^* \\ b_2^* \end{pmatrix} = \frac{b}{b-1}\mu \sum_{i=1}^{b} D_i \log D_i \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \begin{pmatrix} \displaystyle\sum_{i \neq j} \left(\frac{1}{2}(Z_i + Z_j) + \frac{b}{b-1}\mu\right) D_i D_j \\ \displaystyle\sum_{i=1}^{b} Z_i D_i \end{pmatrix},$$

where $D := (D_1, \ldots, D_b)$ has the Dirichlet distribution with parameter $(1/(b-1), \ldots, 1/(b-1))$, $\mathcal{L}(X^{(i)}) = \nu$ for $X^{(i)} := (X_1^{(i)}, X_2^{(i)})$, and $X^{(1)}, \ldots, X^{(b)}$, D, and Z are independent.

Proof. We define $w_n := \mathrm{E}[W_n]$ and $p_n := \mathrm{E}[P_n]$, and, for the standardized vector X_n, we obtain from (4.1) the recursion

$$X_n := \begin{pmatrix} \dfrac{W_n - w_n}{n^2} \\ \dfrac{P_n - p_n}{n} \end{pmatrix} = \sum_{i=1}^{b} A_i^{(n)} X_{I_{n,i}}^{(i)} + b^{(n)}$$

with

$$A_i^{(n)} := \begin{bmatrix} \dfrac{1}{n^2} & 0 \\ 0 & \dfrac{1}{n} \end{bmatrix} \begin{bmatrix} 1 & n - I_{n,i} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} I_{n,i}^2 & 0 \\ 0 & I_{n,i} \end{bmatrix} = \begin{bmatrix} \dfrac{I_{n,i}^2}{n^2} & \dfrac{I_{n,i}(n - I_{n,i})}{n^2} \\ 0 & \dfrac{I_{n,i}}{n} \end{bmatrix}$$

https://doi.org/10.1239/jap/1324046019 Published online by Cambridge University Press

and $b^{(n)} = (b_1^{(n)}, b_2^{(n)})^\top$, where

$$b_1^{(n)} = \frac{1}{n^2}\Bigg\{ \sum_{i=1}^{b} Z_i I_{n,i} + \frac{1}{2}\sum_{i\neq j}(Z_i + Z_j) I_{n,i} I_{n,j} - \frac{b}{b-1}\mu n^2 \log n - c_w n^2 + o(n^2) + \sum_{i=1}^{b} w_{I_{n,i}} + n\sum_{i=1}^{b} p_{I_{n,i}} - \sum_{i=1}^{b} I_{n,i} p_{I_{n,i}} \Bigg\}$$

and

$$b_2^{(n)} := \sum_{i=1}^{b} Z_i \frac{I_{n,i}}{n} - \frac{b}{b-1}\mu \log n - c_p + o(1) + \frac{1}{n}\sum_{i=1}^{b} p_{I_{n,i}}.$$

Using $\sum_{i=1}^{b} I_{n,i} = n-1$, it follows that

$$n\sum_{i=1}^{b} p_{I_{n,i}} - \frac{b}{b-1}\mu n^2 \log n = n\frac{b}{b-1}\mu \sum_{i=1}^{b} I_{n,i} \log \frac{I_{n,i}}{n} + c_p n(n-1) + o(n^2)$$

and

$$\sum_{i=1}^{b} w_{I_{n,i}} - \sum_{i=1}^{b} I_{n,i} p_{I_{n,i}} = (c_w - c_p)\sum_{i=1}^{b} I_{n,i}^2 + o(n^2).$$

The equation

$$1 - \sum_{i=1}^{b}\left(\frac{I_{n,i}}{n}\right)^2 = \sum_{i\neq j} \frac{I_{n,i} I_{n,j}}{n^2} + o(1)$$

yields, with $Z_i I_{n,i} = o(n^2)$ and $c_p - c_w = b/(b-1)\mu$,

$$b_1^{(n)} = \frac{b}{b-1}\mu \sum_{i=1}^{b} \frac{I_{n,i}}{n} \log \frac{I_{n,i}}{n} + \sum_{i\neq j}\left(\frac{1}{2}(Z_i + Z_j) + \frac{b}{b-1}\mu\right)\frac{I_{n,i}}{n}\frac{I_{n,j}}{n} + o(1). \tag{4.11}$$

By similar arguments we have

$$b_2^{(n)} = \frac{b}{b-1}\mu \sum_{i=1}^{b} \frac{I_{n,i}}{n} \log \frac{I_{n,i}}{n} + \sum_{i=1}^{b} Z_i \frac{I_{n,i}}{n} + o(1). \tag{4.12}$$

In order to use the contraction method as in Neininger (2001, Theorem 4.1), it suffices to show that, for $n \to \infty$,

$$(A_1^{(n)}, \dots, A_b^{(n)}, b^{(n)}) \xrightarrow{L_2} (A_1^*, \dots, A_b^*, b^*), \tag{4.13}$$

$$\mathrm{E}[\mathbf{1}_{\{I_{n,i}\le l\}\cup\{I_{n,i}=n\}} \|(A_i^{(n)})^\top A_i^{(n)}\|_{\mathrm{op}}] \to 0 \tag{4.14}$$

for all $l \in \mathbb{N}$, and

$$\sum_{i=1}^{b} \mathrm{E}\,\|(A_i^*)^\top A_i^*\|_{\mathrm{op}} < 1, \tag{4.15}$$

where $\|\cdot\|_{\mathrm{op}}$ is the operator norm.

https://doi.org/10.1239/jap/1324046019 Published online by Cambridge University Press

Let $D := (D_1, \dots, D_b)$ be the almost-sure limit of I_n/n, which is Dirichlet distributed with parameter $(1/(b-1), \dots, 1/(b-1))$. By (4.11) and (4.12), we have, almost surely, $b^{(n)} \to b^*$ as $n \to \infty$ with

$$b^* = \frac{b}{b-1}\mu \sum_{i=1}^{b} D_i \log D_i \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \begin{pmatrix} \sum_{i \neq j} \left(\frac{1}{2}(Z_i + Z_j) + \frac{b}{b-1}\mu \right) D_i D_j \\ \sum_{i=1}^{b} Z_i D_i \end{pmatrix}.$$

By the boundedness of the function $x \mapsto x \log x$ on $[0, 1]$ and as $I_{n,i}/n \in [0, 1]$, there exists a constant C such that

$$|b_1^{(n)}| \leq C + \frac{1}{2} \sum_{i \neq j} |Z_i + Z_j|.$$

By the assumption that $\mathrm{E}[Z_1^2] < \infty$, we obtain the uniform integrability of ${b_1^{(n)}}^2$ and, consequently, the convergence of $b_1^{(n)}$ with respect to the L_2-norm. Similar arguments yield the L_2-convergence of $b_2^{(n)}$ and the convergence of $A_i^{(n)}$ with respect to the L_2-norm to

$$A_i^* = \begin{bmatrix} D_i^2 & D_i(1 - D_i) \\ 0 & D_i \end{bmatrix}.$$

This shows condition (4.13).

Condition (4.14) follows from the deterministic boundedness of $\|A_i^{(n)}\|_{\mathrm{op}}$ and from the fact that

$$\lim_{n \to \infty} \mathrm{P}(\{I_{n,i} \leq l\} \cup \{I_{n,i} = n\}) = 0,$$

which results from (2.1) or the almost-sure convergence of I_n/n to a continuous distribution.

It remains to show (4.15). Solving the characteristic equation for the matrix $(A_i^*)^\top A_i^*$ we find that its eigenvalue $\lambda(D_i)$ being larger in absolute value is given by

$$\lambda(D_i) = D_i^2 \Big(1 - D_i + D_i^2 + (1 - D_i)\sqrt{D_i^2 + 1} \Big).$$

Elementary calculations show that $x > x^2(1 - x + x^2 + (1 - x)\sqrt{x^2 + 1})$ for all $x \in (0, 1)$. Thus, we obtain

$$\mathrm{E}[\lambda(D_i)] < \frac{1}{b-1} \int_0^1 x^{1/(b-1)} \, \mathrm{d}x = \frac{1}{b},$$

which finally yields

$$\mathrm{E}\left[\sum_{i=1}^{b} \|(A_i^*)^\top A_i^*\|_{\mathrm{op}} \right] = \mathrm{E}\left[\sum_{i=1}^{b} \lambda(D_i) \right] < 1.$$

Remark 4.2. Since convergence with respect to the ℓ_2-metric implies convergence of the second moments, Theorem 4.3 shows that, for the variance of P_n,

$$\mathrm{var}(P_n) \sim \mathrm{var}(P) n^2,$$

where $\mathrm{var}(P) < \infty$ can be calculated via the fixed-point equation in Theorem 4.3.

https://doi.org/10.1239/jap/1324046019 Published online by Cambridge University Press

5. Application to linear recursive trees and plane-oriented recursive trees

The results on random weighted b-ary trees also imply limit theorems for further classes of recursive trees with not necessarily bounded outdegree of the nodes, such as random recursive trees or plane-oriented recursive trees (PORTs).

Pittel (1994) introduced the so-called *linear recursive tree* in which, for every new node, the parent u is chosen from the already existent nodes with a probability proportional to $1+\beta\deg(u)$, where $\beta \geq 0$ is the parameter of the tree and $\deg(u)$ denotes the number of internal children of node u. For $\beta = 0$, we obtain the random recursive tree. The PORT—going back to Szymański (1987)—without the consideration of the orientation, corresponds to the $\beta = 1$ case.

For our purpose, we consider the random linear recursive tree with parameter $\beta \in \mathbb{N}_0$ and give a construction in this case. Starting with one internal node and one external child, in each step a uniformly distributed external node is chosen and replaced by an internal one. Furthermore, in each step, $\beta + 1$ external siblings and one external child of the new node are added to the tree. By this construction, the number of external children of a node u is given by $1 + \beta\deg(u)$, which corresponds to the weight defined above. Since the new node is chosen with uniform distribution on the set of external nodes, the probability that an internal node becomes the parent of the new node is proportional to $1 + \beta\deg(u)$. Hence, this construction yields the linear recursive tree with parameter β.

Let T denote the linear recursive tree with parameter β, and consider simultaneously the b-ary recursive tree with $b = \beta + 2$ and edge weights $z := (1, 0, \dots, 0)$ denoted by T'. The tree T with two internal nodes corresponds to the tree T' with one internal node. In both of these trees we have the same number of external nodes. We identify the internal node labeled 2 in T with the root of T' and the external siblings of the first one with the external children of the root in T' where the edges have weight 0. The external child of node 2 in T corresponds in T' to the child of the root where the edge has weight 1.

Now, in both tree models and in each insertion step an external node is chosen, changed to an internal node, and b external nodes are added. We identify these new nodes in the same way as above, i.e. the new external siblings of the new internal node T correspond to the children of the corresponding new internal node in T' where the edge weights are 0 and the child of the new node T is identified with the child of the new node in T' where the edge weight is 1. Then, the depth of a node in the linear recursive tree is equal to the weighted depth of the corresponding node in the b-ary recursive tree plus 1.

This relationship between both tree models was used in Broutin and Devroye (2006) when investigating the height of linear recursive trees. In Figure 1 a linear recursive tree T and its correspondent b-ary recursive tree T' for the $b = 3$ case is shown. The nodes in T' indicated by small squares correspond to the nodes in T which are children of the root.

A fundamental difference between both tree models is that the b-ary recursive tree is an ordered tree, and the linear recursive tree is not. To obtain a transformation between both models, we can define an equivalence relation on the set of ordered b-ary trees which identifies trees that correspond to the same linear recursive tree. If $\psi(\mathcal{T}_b(n))$ denotes the set of equivalence classes of b-ary recursive trees with n nodes and $\mathcal{T}_{n+1}$ denotes the set of all unordered recursive trees with $n + 1$ nodes, we can show the following lemma.

Lemma 5.1. *For any $n \in \mathbb{N}$, there exists a bijection*

$$\varphi\colon \mathcal{T}_{n+1} \to \psi(\mathcal{T}_b(n)).$$

Moreover, let $T_{\mathrm{lin}(b-1)}(n+1)$ be a random linear recursive tree of size $n+1$ and let $T_{b+1}(n)$

https://doi.org/10.1239/jap/1324046019 Published online by Cambridge University Press

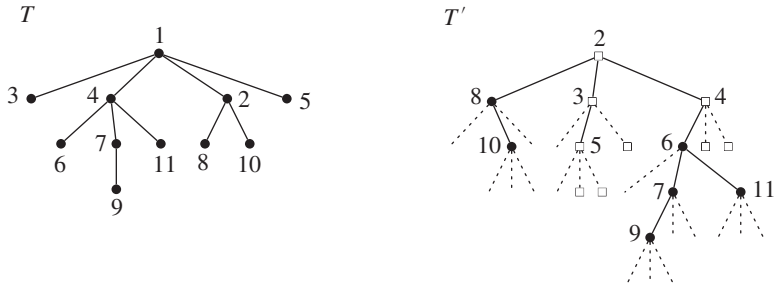

FIGURE 1: A linear recursive tree T with 11 nodes and its correspondent ternary tree T' (without the edge weights).

be a random $(b+1)$*-ary recursive tree of size* n*. Then we have*

$$\psi(T_{b+1}(n)) \stackrel{\mathrm{D}}{=} \varphi(T_{\mathrm{lin}(b-1)}(n+1)).$$

This lemma can be proved rigorously by induction on n (for the details, see Munsonius (2010)).

To transfer the limit results for functionals of random b-ary recursive trees to random linear recursive trees, we have to investigate the behavior of the functionals under the bijection φ. For a node $u \in T$, we denote the subtree of T rooted at u by T_u. Let

$$\Gamma := \{(u, v) \in T \times T \mid v \in T_u\}$$

be the set of all pairs of nodes such that the second node lies in the subtree which is rooted to the first node.

Lemma 5.2. *Let* φ *denote the bijection of Lemma 5.1, let* T *be a recursive tree, and let* $\varphi(T)$ *be weighted with the edge weight vector* $z = (1, 0, \ldots, 0)$*.*

(a) *Let* $D(u)$ *denote the depth of node* u *in* T*, and let* $D_b(w)$ *denote the weighted depth of node* w *in* $\varphi(T)$*. Then we have* $D(u) = D_b(\varphi(u)) + 1$ *for all nodes* u *which are different from the root.*

(b) *Let* $\Delta(u, v)$ *denote the distance between nodes* u *and* v*, where the label of node* u *is less than the label of* v *in the recursive tree* T*, and let* $\Delta_b(\varphi(u), \varphi(v))$ *be the distance between the corresponding nodes in* $\varphi(T)$*. Then*

$$\Delta(u, v) = \Delta_b(\varphi(u), \varphi(v)) + 2\,\mathbf{1}_{(u,v)\notin\Gamma}\,.$$

The proof is by induction on n (see Munsonius (2010)). From these relationships between the depth and the distances in recursive trees and their images under φ, we can now deduce formulae for the internal path length and the Wiener index in linear recursive trees. For a tree T, let $P(T)$ denote its internal path length and let $W(T)$ be its Wiener index.

Corollary 5.1. *Let* φ *be the bijection of Lemma 5.1, let* T *be a recursive tree with* n *nodes, and let* $\varphi(T)$ *be the weighted* b*-ary tree with edge weight vector* $z = (1, 0, \ldots, 0)$*. Then we have*

$$P(T) = P(\varphi(T)) + n - 1 \tag{5.1}$$

https://doi.org/10.1239/jap/1324046019 Published online by Cambridge University Press

and

$$W(T) = W(\varphi(T)) + (n-1)^2 - P(\varphi(T)). \tag{5.2}$$

Proof. Lemma 5.2(a) immediately yields (5.1).

To see (5.2), we argue as follows. By Lemma 5.2(b) we get all distances between nodes other than the root. So we have, with (5.1),

$$\begin{aligned} W(T) &= \sum_{1<u<v} (\Delta_b(\varphi(u), \varphi(v)) + 2\,\mathbf{1}_{(u,v)\notin\Gamma}) + \sum_{1<v} \Delta(1, v) \\ &= W(\varphi(T)) + 2 \sum_{1<u<v} \mathbf{1}_{(u,v)\notin\Gamma} + P(T) \\ &= W(\varphi(T)) + 2\left(\binom{n-1}{2} - |\Gamma|\right) + P(\varphi(T)) + n - 1. \end{aligned}$$

So we have to determine $|\Gamma|$. For $v \in \{2, \dots, n\}$, there are exactly $D(v)$ nodes along the path from v to the root, including the root. This means that there are $D(v) - 1$ tuples in Γ where the second entry is v. Summing over all $v \in \{2, \dots, n\}$ yields

$$|\Gamma| = P(T) - (n-1) = P(\varphi(T)).$$

So, we finally obtain

$$\begin{aligned} W(T) &= W(\varphi(T)) + 2\left(\frac{(n-1)(n-2)}{2} - P(\varphi(T))\right) + P(\varphi(T)) + n - 1 \\ &= W(\varphi(T)) + (n-1)^2 - P(\varphi(T)). \end{aligned}$$

For the functions considered, it does not matter whether we weigh the edges of a random b-ary recursive tree in a definite order or not. In order to apply the results of the last sections, we need the edge weights to be identically distributed. So we take as edge weights the vector $(Z_1, \dots, Z_b)$ with

$$\mathrm{P}((Z_1, \dots, Z_b) = e_i) = \frac{1}{b}$$

for $e_1 = (1, 0, \dots, 0)$, $e_2 = (0, 1, 0, \dots, 0)$, etc, and, thus, $\mu = 1/b$ and $\sigma^2 = (b-1)/b^2$.

As a result, we obtain the corresponding limit theorems for random linear recursive trees combining Lemmas 5.1 and 5.2 with Theorem 3.1.

Theorem 5.1. (Depth of node n.) *Let D_n denote the depth of the node with label n in a random linear recursive tree of size n with weight function $u \mapsto 1 + (b-2)\deg(u)$ for $b \geq 2$. For $n \to \infty$, we have* $\mathrm{E}[D_n] = \log n/(b-1) + o(\log n)$, $\mathrm{var}(D_n) = \log n/(b-1) + o(\log n)$, *and*

$$\frac{D_n - \log n/(b-1)}{\sqrt{\log n/(b-1)}} \xrightarrow{\mathrm{D}} N(0, 1).$$

Similarly, using Corollary 3.1 and Theorem 3.2, we obtain a limit theorem for the depth and distance of random nodes.

Theorem 5.2. (Depth and distance of random nodes.) *Let D_U denote the depth, and let $\Delta_{U,V}$ be the distance of uniformly distributed nodes U and V in a random linear recursive tree of size n with weight function $u \mapsto 1 + (b-2)\deg(u)$ for $b \geq 2$.*

https://doi.org/10.1239/jap/1324046019 Published online by Cambridge University Press

For $n \to \infty$, we have $\mathrm{E}[D_U] = 1/(b-1)\log n + o(\log n)$ *and*

$$\frac{D_U - \mathrm{E}[D_U]}{\sqrt{\log n/(b-1)}} \xrightarrow{\mathrm{D}} N(0,1).$$

Furthermore, $\mathrm{E}[\Delta_{U,V}] = 2/(b-1)\log n + o(\log n)$ *and*

$$\frac{\Delta_{U,V} - \mathrm{E}[\Delta_{U,V}]}{\sqrt{2\log n/(b-1)}} \xrightarrow{\mathrm{D}} N(0,1).$$

Finally, the following limit theorem for the internal path length and the Wiener index is a consequence of the imbedding procedure and Theorem 4.3 for random b-ary recursive trees.

Theorem 5.3. (Limit theorems for P_n and W_n.) *Let W_n denote the Wiener index, and let P_n be the internal path length of a random linear recursive tree of size n with weight function $u \mapsto 1 + (b-2)\deg(u)$ for $b \geq 2$. Then we have, for $n \to \infty$,*

$$\mathrm{E}[P_n] = \frac{1}{b-1} n \log n + (c_p + 1)n + o(n)$$

and

$$\mathrm{E}[W_n] = \frac{1}{b-1} n^2 \log n + \left(c_p + \frac{b-2}{b-1}\right) n^2 + o(n^2),$$

where c_p is given in Remark 4.1. Furthermore, we have

$$\left(\frac{W_n - \mathrm{E}[W_n]}{n^2}, \frac{P_n - \mathrm{E}[P_n]}{n}\right) \xrightarrow{\mathrm{D}} (W, P),$$

where $\mathcal{L}(W, P)$ is given in Theorem 4.3.

Proof. The expectation of P_n and W_n follows directly from Lemma 4.2, Theorem 4.2 (with $c_w = c_p - 1/(b-1)$), Corollary 5.1, and Lemma 5.1.

Let φ be the bijection of Lemma 5.1. Let T_n be a random linear recursive tree of size n with weight function $u \mapsto 1 + (b-2)\deg(u)$. Then, Remark 4.2 yields

$$\frac{P(\varphi(T_n)) - \mathrm{E}[P(\varphi(T_n))]}{n^2} \xrightarrow{\mathrm{P}} 0.$$

The combination of Corollary 5.1 and Theorem 4.3 now implies the claim.

Random PORTs without the order of the nodes are equal in distribution to random linear recursive trees with parameter $\beta = 1$. Since the considered functionals are invariant under changing the order of the tree, the limit theorems above provide in particular the corresponding limit theorems for the PORT. Limit theorems for the depth of a (random) node and the distance between two random nodes, as well as the expectation of the internal path length and of the Wiener index, are given in Morris *et al.* (2004). The limit theorem for the depth of node n in the PORT is proved in Mahmoud (1992). We obtain the results for PORTs as a corollary of Theorem 5.3

Corollary 5.2. (PORTs.) *The depth of the nth node and of a random node, the distance between two random nodes, the internal path length, and the Wiener index of a PORT satisfy the same limit theorem as the random linear recursive trees in the $b = 3$ case.*

https://doi.org/10.1239/jap/1324046019 Published online by Cambridge University Press

Acknowledgement

A series of useful hints, suggestions, and remarks concerning the results in this paper due to Ralph Neininger are gratefully acknowledged.

References

Athreya, K. B. (1969). On a characteristic property of Polya's urn. *Studia Sci. Math. Hung.* **4,** 31–35.

Bergeron, F., Flajolet, P. and Salvy, B. (1992). Varieties of increasing trees. In *CAAP '92* (Rennes, 1992; Lecture Notes Comput. Sci. **581**), Springer, Berlin, pp. 24–48.

Broutin, N. and Devroye, L. (2006). Large deviations for the weighted height of an extended class of trees. *Algorithmica* **46,** 271–297.

Chow, Y. S. and Teicher, H. (1997). *Probability Theory*, 3rd edn. Springer, New York.

Devroye, L. (1999). Universal limit laws for depths in random trees. *SIAM J. Comput.* **28,** 409–432.

Devroye, L. and Neininger, R. (2004). Distances and finger search in random binary search trees. *SIAM J. Comput.* **33,** 647–658.

Dobrow, R. P. and Fill, J. A. (1999). Total path length for random recursive trees. *Combinatorics Prob. Comput.* **8,** 317–333.

Drmota, M. (2009). *Random Trees*. SpringerWienNewYork, Vienna.

Janson, S. (2003). The Wiener index of simply generated random trees. *Random Structures Algorithms* **22,** 337–358.

Javanian, M. and Vahidi-Asl, M. Q. (2006). Depth of nodes in random recursive k-ary trees. *Inform. Process. Lett.* **98,** 115–118.

Johnson, N. L. and Kotz, S. (1977). *Urn Models and Their Application*. John Wiley, New York.

Kuba, M. and Panholzer, A. (2010). On the distribution of distances between specified nodes in increasing trees. *Discrete Appl. Math.* **158,** 489–506.

Mahmoud, H. M. (1992). Distances in random plane-oriented recursive trees. *J. Comput. Appl. Math.* **41,** 237–245.

Mahmoud, H. M. and Neininger, R. (2003). Distribution of distances in random binary search trees. *Ann. Appl. Prob.* **13,** 253–276.

Morris, K., Panholzer, A. and Prodinger, H. (2004). On some parameters in heap ordered trees. *Combinatorics Prob. Comput.* **13,** 677–696.

Munsonius, G. O. (2010). Limit theorems for functionals of recursive trees. Doctoral Thesis, University of Freiburg. Available at http://www.freidok.uni-freiburg.de/volltexte/7472.

Neininger, R. (2001). On a multivariate contraction method for random recursive structures with applications to Quicksort. *Random Structures Algorithms* **19,** 498–524.

Neininger, R. (2002). The Wiener index of random trees. *Combinatorics Prob. Comput.* **11,** 587–597.

Neininger, R. and Rüschendorf, L. (1999). On the internal path length of d-dimensional quad trees. *Random Structures Algorithms* **15,** 25–41.

Neininger, R. and Rüschendorf, L. (2004). A general limit theorem for recursive algorithms and combinatorial structures. *Ann. Appl. Prob.* **14,** 378–418.

Panholzer, A. (2004a). The distribution of the size of the ancestor-tree and of the induced spanning subtree for random trees. *Random Structures Algorithms* **25,** 179–207.

Panholzer, A. (2004b). Distribution of the Steiner distance in generalized M-ary search trees. *Combinatorics Prob. Comput.* **13,** 717–733.

Panholzer, A. and Prodinger, H. (2004a). Analysis of some statistics for increasing tree families. *Discrete Math. Theoret. Comput. Sci.* **6,** 437–460.

Panholzer, A. and Prodinger, H. (2004b). Spanning tree size in random binary search trees. *Ann. Appl. Prob.* **14,** 718–733.

Panholzer, A. and Prodinger, H. (2007). Level of nodes in increasing trees revisited. *Random Structures Algorithms* **31,** 203–226.

Pittel, B. (1994). Note on the heights of random recursive trees and random m-ary search trees. *Random Structures Algorithms* **5,** 337–347.

Quintas, L. V. and Szymański, J. (1992). Nonuniform random recursive trees with bounded degree. In *Sets, Graphs and Numbers* (Budapest, 1991; Colloq. Math. Soc. János Bolyai **60**), North-Holland, Amsterdam, pp. 611–620.

Rösler, U. (1991). A limit theorem for "Quicksort". *RAIRO Inf. Théor. Appl.* **25,** 85–100.

Roura, S. (2001). Improved master theorems for divide-and-conquer recurrences. *J. Assoc. Comput. Mach.* **48,** 170–205.

Ryvkina, J. (2008). Ein universeller zentraler Grenzwertsatz für den Abstand zweier Kugeln in zufälligen Splitbäumen. Diploma Thesis, University of Frankfurt.

SMYTHE, R. T. AND MAHMOUD, H. M. (1995). A survey of recursive trees. *Theory Prob. Math. Statist.* **51,** 1–27.
STANLEY, R. P. (1997). *Enumerative Combinatorics* (Camb. Stud. Adv. Math. **49**), Vol. 1. Cambridge University Press.
SU, C., LIU, J. AND FENG, Q. (2006). A note on the distance in random recursive trees. *Statist. Prob. Lett.* **76,** 1748–1755.
SZYMAŃSKI, J. (1987). On a nonuniform random recursive tree. In *Random Graphs '85* (Poznań, 1985; North-Holland Math. Stud. **144**), North-Holland, Amsterdam, pp. 297–306.

https://doi.org/10.1239/jap/1324046019 Published online by Cambridge University Press