HAL
open science

# The total Steiner $k$-distance for $b$-ary recursive trees and linear recursive trees

Götz Olaf Munsonius

## To cite this version:

Götz Olaf Munsonius. The total Steiner $k$-distance for $b$-ary recursive trees and linear recursive trees. 21st International Meeting on Probabilistic, Combinatorial, and Asymptotic Methods in the Analysis of Algorithms (AofA'10), 2010, Vienna, Austria. pp.527-548, 10.46298/dmtcs.2779 . hal-01185577

**HAL Id: hal-01185577**

**https://inria.hal.science/hal-01185577v1**

Submitted on 20 Aug 2015

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

*AofA'10* DMTCS proc. **AM**, 2010, 527–548

# *The total Steiner $k$-distance for $b$-ary recursive trees and linear recursive trees*

Götz Olaf Munsonius

*Department for Mathematics and Computer Science, J.W. Goethe-University Frankfurt, 60054 Frankfurt a. M., Germany*

We prove a limit theorem for the total Steiner $k$-distance of a random $b$-ary recursive tree with weighted edges. The total Steiner $k$-distance is the sum of all Steiner $k$-distances in a tree and it generalises the Wiener index. The limit theorem is obtained by using a limit theorem in the general setting of the contraction method. In order to use the contraction method we prove a recursion formula and determine the asymptotic expansion of the expectation using the so-called Master Theorem by Roura (2001).

In a second step we prove a transformation of the total Steiner $k$-distance of $b$-ary trees with weighted edges to arbitrary recursive trees. This transformation yields the limit theorem for the total Steiner $k$-distance of the linear recursive trees when the parameter of these trees is a non-negative integer.

**Keywords:** internal path length, Wiener index, total Steiner $k$-distance, recursive trees, contraction method

## 1 Introduction

The analysis of algorithms often leads to the consideration of functionals that deal with distances between nodes in trees. For example the internal path length of a binary search tree, which is the sum of all distances between the nodes and the root, equals the number of comparisons that are needed to build up the tree. Distances are also of interest in the area of complex networks. Considering a computer network the distance between two computers is the number of communication links between them. Hence, in this case the distance is a measure for the speed of data transfers.

More generally, for a set of $k$ nodes in a network the size of the smallest subtree containing this set is a quantity which is related to the communication potential between these nodes. The smallest spanning subtree is also called Steiner tree and its size is the so-called Steiner distance of the given set. As pointed out by Dankelmann et al. (1996), to compare different networks relative to efficient communication potential their typical Steiner distances can be used. As stated in Guillemin and Robert (2009), there is also a relation between the Steiner distance and the efficiency of a traceroute algorithm. Thus, it is of interest to understand Steiner distances in networks.

1365–8050 © 2010 Discrete Mathematics and Theoretical Computer Science (DMTCS), Nancy, France

In the present paper, we consider the total Steiner $k$-distance of random trees. This is the sum of Steiner distances for all possible choices of $k$ nodes and it generalises the Wiener index (see Dobrynin et al. (2001) for a survey). The total Steiner $k$-distance of a tree of size $n$ divided by $\binom{n}{k}$ is the average Steiner $k$-distance of this tree. If not all nodes are lying in the same subtree the Steiner tree coincides with the so-called ancestor tree, which is the smallest subtree that contains the given set of nodes and additionally the root. We define the $k$-th internal path length as the sum of the sizes of ancestor trees for all possible choices of $k$ nodes. This is a generalisation of the internal path length.

The underlying tree model in this paper is a $b$-ary recursive tree with weighted edges. A recursive tree is a labelled tree in which the labels along each path starting from the root increase. We consider ordered recursive trees where the number of children is bounded by $b$ for each node. We additionally attach random weights to the edges of this trees. By proper choices of these edge weights the results can be transferred to random tree models with unbounded outdegree, such as the random recursive tree, the random plane oriented recursive tree and the random linear recursive tree due to Pittel (1994).

The $b$-ary recursive tree with weighted edges can be considered as a special case of the tree model in the paper of Broutin and Devroye (2006) in discrete time where the lifetimes of the edges are independent exponentially distributed random variables. The shape of the random tree is also obtained as an increasing tree due to Bergeron et al. (1992) and is a special case of the general model of random trees in Broutin et al. (2008).

We define the random $b$-ary recursive tree with weighted edges in Section 2 and state some basic properties. In Section 3 we introduce the $k$-th internal path length and the total Steiner $k$-distance and show the basic recursion formula in higher dimensions for these functionals in analogy to the two-dimensional recursion formula for the internal path length and the Wiener index in Neininger (2002). In preparation for using the contraction method in $L_2$ setting, the asymptotic expansions of the expectations are determined in Section 4 using the Master Theorem by Roura (2001). Finally the limit theorem in the case of $b$-ary recursive trees is stated and proved in Section 5. In the last Section 6 a transformation of the results for the $b$-ary recursive tree to linear recursive trees is given.

The internal path length has been studied for severals random trees. The expectation of the internal path length for increasing trees which include random recursive trees, plane oriented recursive trees and binary search trees is stated in Bergeron et al. (1992). The distribution of the internal path length of random binary search trees, random recursive trees, random $m$-ary search trees and split trees is investigated in Régnier (1989), Rösler (1991), Mahmoud (1991), Dobrow and Fill (1999), Neininger and Rüschendorf (1999), Neininger and Rüschendorf (2004) and others. Distributional properties of the Wiener index of random trees are studied in Neininger (2002) and Ali Khan and Neininger (2007) for the random binary search tree and the random recursive tree and in Janson (2003) for simply generated trees. In Janson and Chassaing (2004) the Wiener index of random quadrangulations is investigated.

The Steiner distance for a random set of $k$ nodes in various random trees is studied in several papers (see Panholzer (2004a), Panholzer (2004b), Morris et al. (2004), Panholzer and Prodinger (2004a), Panholzer and Prodinger (2004b) and Guillemin and Robert (2009)). There is also considered the size of the ancestor tree of $k$ random nodes in random trees.

The total Steiner $k$-distance is mentioned in Clark et al. (1999) and the expectation for simply

generated trees is determined. In Dankelmann et al. (1996) upper and lower bounds of the average Steiner $k$-distance for deterministic weighted graphs and trees are shown.

We fix some notation. For random variables we write $\stackrel{\mathcal{D}}{=}$ for equality in distribution and $\mathcal{L}(X)$ for the distribution of $X$. Let $\mathcal{M}_{0,2}^d$ be the set of centred probability measures on $\mathbb{R}^d$ with finite second moments. The Wasserstein-metric $l_2$ on $\mathcal{M}_{0,2}^d$ is defined by

$$l_2(\mu,\nu) := \inf\{\|X-Y\|_2 : \mathcal{L}(X)=\mu, \mathcal{L}(Y)=\nu\} \tag{1}$$

where the $L_2$-norm $\|\cdot\|_2$ is given by $\|X\|_2 = (E[\|X\|^2])^{1/2}$. It is well known that convergence with respect to the metric $l_2$ (denoted by $\xrightarrow{l_2}$) is equivalent to weak convergence plus convergence of the second moments (see *e.g.* Bickel and Freedman (1981)).

For $r \in \mathbb{N}$ we denote by $x^{(r)} = x(x-1)\cdots(x-r+1)$ the lower factorial and accordingly $x^{[r]} = x(x+1)\cdots(x+r-1)$ the upper factorial.

## 2 The model of the random $b$-ary recursive tree with weighted edges

We consider the infinite $b$-ary ordered tree $\mathcal{T}_b(\infty)$, with the set of nodes $\mathcal{V}(\infty) := \bigcup_{k\in\mathbb{N}}\{1,\dots,b\}^k \cup \{0\}$. A *b-ary recursive tree* with $n$ nodes is obtained by choosing a subset $V \subset \mathcal{V}(\infty)$ of size $n$ such that for all $u \in V$ the whole path from $u$ to the root belongs to $V$, together with a labelling function $l : V \to \{1,\dots,n\}$ such that the labels on every path to the root are decreasing. We denote by $\mathcal{T}_b(n)$ the set of all $b$-ary recursive trees with $n$ nodes. A *random b-ary recursive tree* $T_n$ with $n$ nodes is a uniformly distributed element of $\mathcal{T}_b(n)$.

Let $Z := (Z_1,\dots,Z_b) \in \mathbb{R}_{\geq 0}^b$ be a random vector and attach to every node $u$ of $\mathcal{T}_\infty$ 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 $Z^{(u)}$ are independent of $T_n$, we refer to $T_n$ with the family $\{Z^{(u)} | u \in \mathcal{V}_\infty\}$ as a *random b-ary recursive tree with edge weights* $Z$.

We will assume for the rest of the paper that the entries of the vector of edge weights are identically distributed, *i.e.* for all $i,j \in \{1,\dots,b\}$ we have

$$Z_i \stackrel{\mathcal{D}}{=} Z_j. \tag{2}$$

The distribution of the random $b$-ary recursive tree is independent under permutations of the subtrees. It follows that the assumption (2) does not mean a constraint since we can always consider the tree with a uniformly distributed permutation of the subtrees.

We introduce some further notation. By $T_n$ we denote the random $b$-ary recursive tree of size $n$ with edge weights $Z$. The subtrees rooted to the children of the root are denoted by $T_{n,1},\dots,T_{n,b}$ from left to right, *i.e.* the subtree $T_{n,i}$ contains all nodes of $T_n$ that lie in the set $\{i\} \cup (\{i\} \times \bigcup_{k\in\mathbb{N}}\{1,\dots,b\}^k)$. We define $I_{n,j} := |T_{n,j}|$ the number of internal nodes in the subtree $T_{n,j}$ and $I_n := (I_{n,1},\dots,I_{n,b})$. The weight of the edge between the root of $T_n$ and the one of $T_{n,i}$ is given by $Z_i$. The notation $u \in T_n$ means that $u$ is a node of $T_n$.

We can describe the random $b$-ary recursive tree by an evolution process. We start with the root as an internal and its $b$ children as external nodes at the next level. In each step we choose a random element uniformly distributed on the set of all external nodes. The chosen node becomes an internal one, and its $b$ children join the set of external nodes. All nodes are labelled in the order of their appearance.

Considering this evolution process, we see that the subtree sizes $I_n = (I_{n,1}, \ldots, I_{n,b})$ of a random recursive $b$-ary tree can be described by a Pólya urn with $b$ colours. Starting with one ball of each colour (corresponding to the external children of the root), each drawn ball is returned to the urn with $b-1$ additional balls of the same colour. Then, the number of drawings of one colour corresponds to the number of internal nodes in the corresponding subtree. We summarise some well known results about this Pólya urn needed later (see *e.g.* Johnson and Kotz (1977)). The explicit formula for the distribution of the subtree size is given by

$$P\left(I_{n+1,1} = k\right) = \frac{1}{b-1} \frac{\Gamma\left(k + \frac{1}{b-1}\right)}{\Gamma(k+1)} \frac{\Gamma(n+1)}{\Gamma\left(n+1+\frac{1}{b-1}\right)}. \tag{3}$$

and the joint factorial moment formula is for $(r_1, \ldots, r_b) \in \mathbb{Z}^b_{\geq 0}$, $r = \sum_{i=1}^b r_i$ and $\alpha = b/(b-1)$

$$E\left[\prod_{j=1}^b I_{n+1,j}^{(r_j)}\right] = \frac{n^{(r)}}{\alpha^{[r]}} \prod_{j=1}^b \left(\frac{1}{b-1}\right)^{[r_j]}. \tag{4}$$

Further, the normalised subtree sizes $I_n/n$ converges almost surely to a Dirichlet $\beta_{(\frac{1}{b-1},\ldots,\frac{1}{b-1})}$ distributed random vector $D$ with parameters $(\frac{1}{b-1}, \ldots, \frac{1}{b-1})$

$$\frac{I_n}{n} \longrightarrow D \quad a.s. \tag{5}$$

and in particular

$$\frac{I_{n,i}}{n} \longrightarrow D_i \quad a.s., \tag{6}$$

for $i = 1, \ldots, b$ where $\mathcal{L}(D_i) = \beta_{(\frac{1}{b-1},1)}$ is the beta distribution (see Athreya (1969)).

Using this convergence result, one can show that for natural numbers $k$, $l \in \mathbb{N}$ with $l \leq k$ there exist constants $\tilde{c}_1(k,l)$ and $\tilde{c}_2(k,l)$ independent of $n$ such that for $n \to \infty$

$$E\left[I_{n,1}^l \binom{n - I_{n,1}}{k-l} \log I_{n,1}\right] = \tilde{c}_1(k,l) n^k \log n + \tilde{c}_2(k,l) n^k + o(n^k) \tag{7}$$

and

$$E\left[I_{n,1}^l \binom{n - I_{n,1}}{k-l}\right] = \tilde{c}_1(k,l) n^k + o(n^k). \tag{8}$$

In particular, it turns out that

$$\tilde{c}_1(k,l) = \frac{1}{b-1} \frac{\Gamma\left(l + \frac{1}{b-1}\right)}{\Gamma\left(k + \frac{b}{b-1}\right)} \tag{9}$$

and with $D_1$ as in (6)

$$\tilde{c}_2(k,l) = \frac{1}{(k-l)!} E\left[D_1^l (1-D_1)^{k-l} \log D_1\right]. \tag{10}$$

A prove is given in Munsonius (2010).

## 3 The $k$-th internal path length and the total Steiner $k$-distance

We consider a labelled rooted tree $T$ with labels $\{1, \ldots, n\}$ and weighted edges. For a subset $M \subset \{1, \ldots, n\}$ of labels, the ancestor tree for $M$ in $T$ is defined by the smallest subtree of $T$ which contains all nodes with labels in $M$ and the root. Let $D_M(T)$ be the sum of the edge weights of the ancestor tree of the nodes with labels in $M$ in the given tree $T$. We define the generalised *$k$-th internal path length* of $T$ by

$$P_k(T) = \sum_{\substack{M \subset \{1,\ldots,n\} \\ |M|=k}} D_M(T). \tag{11}$$

Similarly, we define $S_M(T)$ as the sum of the edge weights of the spanning subtree in $T$ of the nodes with labels in $M$. This is also known as the Steiner distance of $M$. The *total Steiner $k$-distance* is then defined by

$$W_k(T) = \sum_{\substack{M \subset \{1,\ldots,n\} \\ |M|=k}} S_M(T). \tag{12}$$

As special cases, $P_1$ is the internal path length and $W_2$ the Wiener index.

In the case that $T$ is a random $b$-ary recursive tree of size $n$ we write $D_{M,n} := D_M(T)$, $S_{M,n} := S_M(T)$, $P_{k,n} := P_k(T)$ and $W_{k,n} := W_k(T)$. It turns out that we obtain a recursion in dimension $k$ for $P_{k,n}$ and $W_{k,n}$. In Neininger (2002) this recursion is stated in the case $k = 2$.

**Fig. 1:** The ancestor tree (left) and the Steiner tree (right) for the set $M = \{6, 9\}$ with thick edges.

**Lemma 3.1** *Let $T_n$ be a random $b$-ary recursive tree of size $n$ with weighted edges. Further, we denote by $P_{k,n}$ resp. $W_{k,n}$ the $k$-th internal path length resp. total Steiner $k$-distance of $T_n$ and by $P_{k,n,i}$ resp. $W_{k,n,i}$ the corresponding sizes of the subtrees $T_{n,i}$ for $i = 1, \ldots, b$. Then the following recursions hold:*

$$P_{k,n} = \sum_{i=1}^{b} \left\{ P_{k,n,i} + \sum_{l=1}^{k-1} P_{l,n,i} \binom{n - I_{n,i}}{k-l} + Z_i \left[ \binom{n}{k} - \binom{n - I_{n,i}}{k} \right] \right\} \tag{13}$$

*and*

$$W_{k,n} = \sum_{i=1}^{b} \left\{ W_{k,n,i} + \sum_{l=1}^{k-1} P_{l,n,i} \binom{n - I_{n,i}}{k-l} + Z_i \left[ \binom{n}{k} - \binom{n - I_{n,i}}{k} - \binom{I_{n,i}}{k} \right] \right\}. \tag{14}$$

**Proof:** For all subsets $M \subset \{1, \ldots, n\}$ with $|M| = k$ that are subsets of the labels in only one subtree $T_{n,i}$ the summand $D_{M,n}$ is given by

$$D_{M,n} = D_{M,n,i} + Z_i \tag{15}$$

where $D_{M,n,i}$ is the size of the ancestor tree for $M$ in $T_{n,i}$. The sum over all such subsets yields

$$\sum_{\substack{M \subset T_{n,i} \\ |M|=k}} D_{M,n} = P_{k,n,i} + Z_i \binom{I_{n,i}}{k}. \tag{16}$$

If there are $l < k$ elements of $M$ in the set of labels that are in $T_{n,i}$, there are $\binom{n-I_{n,i}}{k-l}$ constellations for the remaining $k-l$ nodes. For each such constellation we get the sum over all subsets of size $l$ in the subtree $T_{n,i}$. This yields the summand

$$\binom{n-I_{n,i}}{k-l} P_{l,n,i} + Z_i \left[ \binom{I_{n,i}}{l} \binom{n-I_{n,i}}{k-l} \right]. \tag{17}$$

Considering the above findings, we obtain

$$P_{k,n} = \sum_{i=1}^{b} \left\{ P_{k,n,i} + \sum_{l=1}^{k-1} P_{l,n,i} \binom{n-I_{n,i}}{k-l} + Z_i \left[ \sum_{l=1}^{k-1} \binom{I_{n,i}}{l} \binom{n-I_{n,i}}{k-l} + \binom{I_{n,i}}{k} \right] \right\}. \tag{18}$$

Since we have

$$\sum_{l=0}^{k} \binom{I_{n,i}}{l} \binom{n-I_{n,i}}{k-l} = \binom{n}{k} \tag{19}$$

the claim for the $k$-th internal path length follows.

If not all elements of $M$ are in one subtree we have $S_{M,n} = D_{M,n}$. Hence, the total Steiner $k$-distance differs from the $k$-th internal path length only in the summands where all nodes are in the same subtree. It is

$$\sum_{\substack{M \subset T_{n,i} \\ |M|=k}} S_{M,n} = W_{k,n,i}. \tag{20}$$

Comparing this to the computations above, we immediately get the claim for the total Steiner $k$-distance. □

From the description by the evolution process, we see that conditioned upon $(I_{n,1}, \ldots, I_{n,b})$ the subtrees are independent, random $b$-ary recursive trees of sizes $I_{n,1}, \ldots, I_{n,b}$ respectively (after changing the labels in an obvious way). Together with the lemma above this yields the following corollary.

**Corollary 3.2** *For $k \in \mathbb{N}$, we set $X_{k,n} = (P_{1,n}, P_{2,n}, \ldots, P_{k-1,n}, W_{k,n})^T$, the vector consisting of the $l$-th internal path lengths ($l = 1, \ldots, k-1$) and the total Steiner $k$-distance of a random $b$-ary recursive tree of size $n$ with edge weights $Z$. Then we have*

$$X_{k,n} \stackrel{\mathcal{D}}{=} \sum_{i=1}^{b} A_i(n) X^{(i)}_{k,I_{n,i}} + b(n) \tag{21}$$

*where* $A_1^{(n)}, \dots, A_b^{(n)}$ *are random* $k \times k$ *matrices given by*

$$(A_i(n))_{l,m} = \begin{cases} \binom{n-I_{n,i}}{l-m}, & \text{for } 1 \le m \le l \le k, \\ 0, & \text{otherwise,} \end{cases} \tag{22}$$

*and* $b(n)$ *is a random* $k$*-dimensional vector with*

$$(b(n))_l = \sum_{i=1}^{b} Z_i \left[ \binom{n}{l} - \binom{n-I_{n,i}}{l} - \mathbb{1}_{\{l=k\}} \binom{I_{n,i}}{k} \right] \tag{23}$$

*for* $l \in \{1, \dots, k\}$*. Moreover,* $(A_1^{(n)}, \dots, A_b^{(n)}, b(n)), (X_{k,j}^{(1)})_{j \ge 0}, \dots, (X_{k,j}^{(b)})_{j \ge 0}$ *are independent and* $X_{k,n}^{(i)} \stackrel{\mathcal{D}}{=} X_{k,n}$ *for all* $n \in \mathbb{N}$*.*

## 4 The expectation of $P_{k,n}$ and $W_{k,n}$

We are going to use the contraction method to show a limit theorem for these functionals. Since we apply the contraction theorem in $L_2$ we have to centre the functionals. Therefore, we have to identify the asymptotic expansion of the expectation.

It follows from Corollary 3.2 that

$$\begin{aligned} E[P_{k,n}] &= bE\left[P_{k,I_{n,1}}\right] + E[t_k(n)] && (24) \\ &= b\sum_{j=0}^{n-1} E\left[P_{k,I_{n,1}} \mid I_{n,1} = j\right] P(I_{n,1} = j) + E[t_k(n)] && (25) \\ &= b\sum_{j=0}^{n-1} E\left[P_{k,j}\right] P(I_{n,1} = j) + E[t_k(n)] && (26) \end{aligned}$$

where

$$E[t_k(n)] := b\sum_{l=1}^{k-1} E\left[P_{l,I_{n,1}} \binom{n-I_{n,1}}{k-l}\right] + bE[Z_1]\left(\binom{n}{k} - E\left[\binom{n-I_{n,1}}{k}\right]\right) \tag{27}$$

as well as

$$E[W_{k,n}] = b\sum_{j=0}^{n-1} E\left[W_{k,j}\right] P(I_{n,1} = j) + \tilde{t}_k(n) \tag{28}$$

where $\tilde{t}_k(n) = E[t_k(n)] - bE[Z_1]E\left[\binom{I_{n,1}}{k}\right]$.

In order to apply the contraction method it suffices to know an asymptotic expansion of the expectation which is fairly good. In Roura (2001), recursions of the given type are considered and the asymptotic expansions of the solutions are identified.

We need two notions from Roura (2001).

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

**Definition 4.2** *We say that*

$$F_n = \begin{cases} b_n, & \text{if } 0 \leq n < N \\ t_n + \sum_{0 \leq k < n} \omega_{n,k} F_k, & \text{if } n \geq N \end{cases} \tag{29}$$

*is a* continuous recursive definition *of $F_n$ iff 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 $z_{n,j} = j/M_n$, $0 \leq j \leq M_n$, with $I_{n,j} = [z_{n,j} n, z_{n,j+1} n)$, $0 \leq j < M_n$, and with*

$$\varepsilon_{n,j} = \left| \sum_{k \in I_{n,j}} \omega_{n,k} - \int_{z_{n,j}}^{z_{n,j+1}} \omega(z)\, dz \right|, \quad 0 \leq j < M_n, \tag{30}$$

$$\sum_{0 \leq j < M_n} \varepsilon_{n,j} = O(n^{-\varrho}) \text{ for some } \varrho > 0. \tag{31}$$

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

**Theorem 4.3 (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 $\mu_n$. Let $\varphi(x) = \int_0^1 \omega(z) z^x\, dz$, and $\mathcal{H} = 1 - \varphi(a)$. If $\mathcal{H} > 0$, then*

$$\lim_{n \to \infty} \frac{F_n}{t_n} = \frac{1}{\mathcal{H}}. \tag{32}$$

Theorem 4.3 also holds in the case $B < 0$ which is seen directly by consideration of the sequence $-F_n$.

To determine the asymptotic expansion via this theorem we have to find the asymptotic expansion of $E[t(n)]$ and we have to show that (26) is a continuous recursive definition of $E[P_{k,n}]$.

**Lemma 4.4** *For $N = 1$, $b_1 = 0$ and*

$$\omega_{n,k} = b \frac{k}{n} P\left(I_{n,1} = k\right) \tag{33}$$

*equation* (29) *is a continuous recursive definition with shape function*

$$\omega(z) = \frac{b}{b-1} z^{\frac{1}{b-1}}. \tag{34}$$

The proof uses Stirlings formula for the Gamma function and can be found in Munsonius (2010). It is omitted here for technical reasons.

We now obtain the asymptotic expansion of the expectation of $P_{k,n}$ by induction on $k$.

**Theorem 4.5** *For $k \geq 1$ let $P_{k,n}$ be the $k$-th internal path length of a random $b$-ary recursive tree of size $n$ with edge weights $Z$. Then there exist constants $c_{p,k}^{(1)}$ and $c_{p,k}^{(2)}$ such that for $n \to \infty$*

$$E[P_{k,n}] = c_{p,k}^{(1)} n^k \log n + c_{p,k}^{(2)} n^k + o(n^k). \tag{35}$$

**Proof:** We use induction on $k$. For $k = 1$ it is $P_{1,n}$ the internal path length, *i.e.* the sum of all weighted depths. In Javanian and Vahidi-Asl (2006) the expectation of the depth of the $n$-th node in a random $b$-ary recursive tree without edge weights is identified. Since the edge weights and the shape of the tree are independent, the expectation of the weighted depth is obtained by Wald's equality. By summing up these values we have

$$E[P_{1,n}] = \frac{b}{b-1} E[Z_1] n \log n + c_{p,1}^{(2)} n + o(n) \tag{36}$$

where

$$c_{p,1}^{(2)} = E[Z_1] \left( \frac{b}{b-1}\gamma - \frac{3b+2}{2b-2} + \frac{b(b-2)}{(b-1)^2} \sum_{j=1}^{\infty} \frac{1}{(j+2)\left(j + \frac{b}{b-1}\right)} \right) \tag{37}$$

with the Euler constant $\gamma \approx 0.577$. The details of this computation can be found in Munsonius (2010).

Now assume the claim is proved for $E[P_{l,n}]$ for all $1 \leq l < k$. We want to use Theorem 4.3. Therefore, we identify the asymptotic expansion of $E[t_k(n)]$ in (27).

By induction hypothesis we get

$$E\left[P_{l,I_{n,1}} \binom{n - I_{n,1}}{k-l}\right] = \sum_{j=0}^{n-1} E\left[P_{l,j}\right] \binom{n-j}{k-l} P\left(I_{n,1} = j\right) \tag{38}$$

$$= \sum_{j=0}^{n-1} \left\{ c_{p,l}^{(1)} j^l \log j + c_{p,l}^{(2)} j^l + o(j^l) \right\} \binom{n-j}{k-l} P\left(I_{n,1} = j\right) \tag{39}$$

$$= E\left[ \left\{ c_{p,l}^{(1)} I_{n,1}^l \log I_{n,1} + c_{p,l}^{(2)} I_{n,1}^l + o\left(I_{n,1}^l\right) \right\} \binom{n - I_{n,1}}{k-l} \right]. \tag{40}$$

With equations (7) and (8) we obtain

$$E[t_k(n)] = \tilde{C}_k^{(1)} n^k \log n + \tilde{C}_k^{(2)} n^k + o(n^k) \tag{41}$$

with

$$\tilde{C}_k^{(1)} := b \sum_{l=1}^{k-1} c_{p,l}^{(1)} \tilde{c}_1(k,l) \tag{42}$$

and

$$\tilde{C}_k^{(2)} := b \sum_{l=1}^{k-1} \left( c_{p,l}^{(1)} \tilde{c}_2(k,l) + c_{p,l}^{(2)} \tilde{c}_1(k,l) \right) + bE[Z_1] \left( \frac{1}{k!} - \tilde{c}_1(k,0) \right). \tag{43}$$

We set

$$c_{p,k}^{(1)} := \frac{k(b-1)+1}{(k-1)(b-1)}\tilde{C}_k^{(1)} = \frac{kb-(k-1)}{(k-1)(b-1)} b \sum_{l=1}^{k-1} c_{p,l}^{(1)} \tilde{c}_1(k,l). \tag{44}$$

Now we deduce a representation of $E[P_{k,n}]$ as a continuous recursive definition in the sense of Definition 4.2. But instead of $E[P_{k,n}]$, we consider

$$G_{k,n} := \frac{1}{n}\left(E[P_{k,n}] - c_{p,k}^{(1)} n^k \log n\right) \tag{45}$$

and obtain by equation (26)

$$\begin{aligned} G_{k,n} &= \frac{E[P_{k,n}] - c_{p,k}^{(1)} n^k \log n}{n} && (46)\\ &= \sum_{j=1}^{n-1} \frac{E\left[P_{k,j}\right] - c_{p,k}^{(1)} j^k \log j}{j} \frac{j}{n} bP\left(I_{n,1}=j\right) && (47)\\ &\quad + \frac{1}{n}\left(E[t_k(n)] + \sum_{j=0}^{n-1} c_{p,k}^{(1)} j^k \log(j) bP\left(I_{n,1}=j\right) - c_{p,k}^{(1)} n^k \log n\right) && (48)\\ &= \sum_{j=0}^{n-1} \omega_{n,j} G_{k,j} + s_k(n), && (49) \end{aligned}$$

with $G_{k,0} := 0$, $\omega_{n,j}$ being given in Lemma 4.4 and

$$s_k(n) := \frac{1}{n}\left(E[t_k(n)] + c_{p,k}^{(1)} bE\left[I_{n,1}^k \log I_{n,1}\right] - c_{p,k}^{(1)} n^k \log n\right). \tag{50}$$

By Lemma 4.4 this is a continuous recursive definition. We determine the main term of $s_k(n)$. Using equations (7) and (41) as well as the definition (44) of $c_{p,k}^{(1)}$ and $\tilde{c}_1(k,k) = 1/(k(b-1)+1)$ (see equation (9)) we obtain

$$\begin{aligned} s_k(n) &= \left(\tilde{C}_k^{(1)} + bc_{p,k}^{(1)}\tilde{c}_1(k,k) - c_{p,k}^{(1)}\right) n^{k-1}\log n + \left(\tilde{C}_k^{(2)} + bc_{p,k}^{(1)}\tilde{c}_2(k,k)\right) n^{k-1} + o(n^{k-1}) && (51)\\ &= \underbrace{\left(\tilde{C}_k^{(2)} + bc_{p,k}^{(1)}\tilde{c}_2(k,k)\right)}_{=:\hat{C}_k} n^{k-1} + o(n^{k-1}). && (52) \end{aligned}$$

We set $\xi_n = 1$, $a = k-1$, $c = 0$ and $B = \hat{C}_k \neq 0$. Using the terminology of Theorem 4.3 we show $\mathcal{H} = 1 - \varphi(k-1) > 0$. It is for $b \geq 2$ and $k > 1$

$$\varphi(k-1) = \int_0^1 \frac{b}{b-1} z^{\frac{1}{b-1}+k-1}\, dz = \frac{b}{kb-(k-1)} < 1. \tag{53}$$

Therefore, $\mathcal{H} = (b-1)(k-1)/(k(b-1)+1) > 0$ and Theorem 4.3 yields $\lim_{n\to\infty} G_{k,n}/s_k(n) = 1/\mathcal{H}$. Hence we get

$$E[P_{k,n}] - c_{p,k}^{(1)} n^k \log n = c_{p,k}^{(2)} n^k + o(n^k), \tag{54}$$

where

$$c_{p,k}^{(2)} = \frac{k(b-1)+1}{(b-1)(k-1)} \left( \tilde{C}_k^{(2)} + b c_{p,k}^{(1)} \tilde{c}_2(k,k) \right) \tag{55}$$

for $k > 1$ and the proof is finished. □

The same method yields the asymptotic expansion of the expectation of $W_{k,n}$.

**Theorem 4.6** *For $k \geq 2$ let $W_{k,n}$ be the total Steiner k-distance of a random b-ary recursive tree of size n with edge weights Z. Then there exist constants $c_{w,k}^{(1)}$ and $c_{w,k}^{(2)}$ such that for $n \to \infty$*

$$E[W_{k,n}] = c_{w,k}^{(1)} n^k \log n + c_{w,k}^{(2)} n^k + o(n^k). \tag{56}$$

**Proof:** We set

$$H_{k,n} := \frac{1}{n} \left( E[W_{k,n}] - c_{p,k}^{(1)} n^k \log n \right), \tag{57}$$

with $c_{p,k}^{(1)}$ as in Theorem 4.5, and obtain by equation (28) the recursion

$$H_{k,n} = \sum_{j=0}^{n-1} \omega_{n,j} H_{k,j} + r_k(n) \tag{58}$$

where $r_k(n)$ is given—analogous to (50)—by

$$r_k(n) = s_k(n) - \frac{1}{n} b E[Z_1] E\left[ \binom{I_{n,1}}{k} \right] \tag{59}$$

with $s_k(n)$ as in (50). With the factorial moments in equation (4) as well as (50) in combination with (52) we see that

$$r_k(n) = \left( \tilde{C}_k^{(2)} + b c_{p,k}^{(1)} \tilde{c}_2(k,k) \right) n^{k-1} - \frac{1}{n} b E[Z_1] E\left[ \binom{I_{n,1}}{k} \right] + o(n^{k-1}) \tag{60}$$

$$= \left( \tilde{C}_k^{(2)} + b c_{p,k}^{(1)} \tilde{c}_2(k,k) - b E[Z_1] \frac{1}{k!(k(b-1)+1)} \right) n^{k-1} + o(n^{k-1}). \tag{61}$$

As in the proof of Theorem 4.5 we have a continuous recursive definition for $H_{k,n}$. Again, the main term of the additive term $r_k(n)$ is of order $n^{k-1}$. Hence, we finally receive by Theorem 4.3 as in the proof of Theorem 4.5 the claim with

$$c_{w,k}^{(1)} = c_{p,k}^{(1)} \tag{62}$$

and with (55)

$$c_{w,k}^{(2)} = \frac{k(b-1)+1}{(k-1)(b-1)} \left( \tilde{C}_k^{(2)} + b c_{p,k}^{(1)} \tilde{c}_2(k,k) - \frac{b}{k(b-1)+1} \frac{1}{k!} E[Z_1] \right) \tag{63}$$

$$= c_{p,k}^{(2)} - \frac{b}{(k-1)(b-1)k!} E[Z_1]. \tag{64}$$

□

**Remark 4.7** *While it does not seem that there is an easy way to compute the constants $c^{(2)}_{w,k}$ and $c^{(2)}_{p,k}$ it is easy to determine the constants $c^{(1)}_{w,k} = c^{(1)}_{p,k}$ inductively. By definition* (9) *of $\tilde{c}_1(k,l)$ we get starting with* (44)

$$c^{(1)}_{p,k} = \frac{k(b-1)+1}{(k-1)(b-1)} \frac{b}{b-1} \frac{1}{\Gamma\left(k+\frac{1}{b-1}\right)\left(k+\frac{1}{b-1}\right)} \sum_{l=1}^{k-1} c^{(1)}_{p,l} \Gamma\left(l+\frac{1}{b-1}\right). \tag{65}$$

*We substitute $\Gamma\left(k+\frac{1}{b-1}\right) c^{(1)}_{p,k} =: B_k$ and get by recursive computations*

$$B_k = \frac{b}{(k-1)(b-1)} \sum_{l=1}^{k-1} B_l \tag{66}$$

$$= \frac{b}{(k-1)(b-1)} \left( \frac{(k-2)(b-1)}{b} \frac{b}{(k-2)(b-1)} \sum_{l=1}^{k-2} B_l + B_{k-1} \right) \tag{67}$$

$$= \frac{b}{(k-1)(b-1)} \left( \frac{(k-2)(b-1)}{b} + 1 \right) B_{k-1} \tag{68}$$

$$= \frac{1}{(k-1)!} \frac{\Gamma\left(k+\frac{1}{b-1}\right)}{\Gamma\left(1+\frac{1}{b-1}\right)} B_1. \tag{69}$$

*By resubstitution and with $c^{(1)}_{p,1} = \frac{b}{b-1} E[Z_1]$ we finally conclude*

$$c^{(1)}_{p,k} = \frac{1}{(k-1)!} \frac{b}{b-1} E[Z_1]. \tag{70}$$

# 5 The limit theorem for $P_{k,n}$ and $W_{k,n}$

We finally show a limit theorem for the $k$-th internal path length and the total Steiner $k$-distance. We can use the multivariate contraction method in $L_2$ setting (Neininger (2001)) because we know the asymptotic expansion of the expectations fairly well. Corollary 3.2 provides the needed recursion formula.

**Theorem 5.1 (Limit theorem for $b$-ary trees with edge weights)** *For $k \in \mathbb{N}$ we set*

$$X_{k,n} = (P_{1,n}, P_{2,n}, \ldots, P_{k-1,n}, W_{k,n})^T \tag{71}$$

*the vector consisting of the $l$-th internal path lengths ($l = 1, \ldots, k-1$) and the total Steiner $k$-distance of a random $b$-ary recursive tree of size $n$ with edge weights $Z$ where $\mathrm{Var}(Z_1) < \infty$. Then we have for*

$$X^*_{k,n} := \begin{bmatrix} \frac{1}{n} & 0 & \cdots & 0 \\ 0 & \frac{1}{n^2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \frac{1}{n^k} \end{bmatrix} (X_{k,n} - E[X_{k,n}]) \tag{72}$$

*and* $n \to \infty$

$$l_2(X^*_{k,n}, X_k) \to 0 \tag{73}$$

*where* $X_k$ *is the unique distributional fixed point of the map* $T : \mathcal{M}^k_{0,2} \to \mathcal{M}^k_{0,2}$ *given for* $\mu \in \mathcal{M}^k_{0,2}$ *by*

$$T(\mu) := \mathcal{L}\left(\sum_{i=1}^{b} A^*_i X^{(i)}_k + b^*\right) \tag{74}$$

*with random* $k \times k$ *matrices* $A^*_i$ *given by*

$$(A^*_i)_{l,m} = \begin{cases} \frac{1}{(l-m)!} D^m_i (1-D_i)^{l-m}, & \text{for } m \le l \le k, \\ 0, & \text{otherwise,} \end{cases} \tag{75}$$

*and a random* $k$*-dimensional vector* $b^*$ *given by*

$$b^*_l = c^{(1)}_{p,l} \sum_{i=1}^{b} D_i \log D_i + \frac{1}{l!} \sum_{i=1}^{b} Z_i \left(1 - (1-D_i)^l - \mathbb{1}_{\{l=k\}} D^k_i\right) \tag{76}$$

$$+ \sum_{i=1}^{b} \sum_{m=1}^{l} \frac{c^{(2)}_{p,m}}{(l-m)!} D^m_i (1-D_i)^{l-m} - c^{(2)}_{p,l} + \mathbb{1}_{\{l=k\}} \left(c^{(2)}_{p,k} - c^{(2)}_{w,k}\right) \left(1 - \sum_{i=1}^{b} D^k_i\right), \tag{77}$$

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

**Proof:** We denote by $\delta(x)$ the diagonal matrix of size $k$ with entries $x, x^2, x^3, \ldots, x^k$, *i.e.* $(\delta(x))_{ij} = x^i \mathbb{1}_{\{i=j\}}$ for $i, j \in \{1, \ldots, k\}$. Then we obtain by Corollary 3.2

$$\delta(1/n)(X_{k,n} - E[X_{k,n}]) \stackrel{\mathcal{D}}{=} \sum_{i=1}^{b} \delta(1/n) A_i(n) \delta(I_{n,i}) X^{(i)*}_{k,I_{n,i}} + b^{(n)} \tag{78}$$

with $X^{(i)*}_{k,I_{n,i}} = \delta(I_{n,i})^{-1} \left(X_{k,I_{n,i}} - E\left[X_{k,I_{n,i}}\right]\right)$ and for $l \le k$ because of $c^{(1)}_{p,k} = c^{(1)}_{w,k}$

$$b^{(n)}_l = \frac{1}{n^l} \left( \sum_{i=1}^{b} Z_i \left[ \binom{n}{l} - \binom{n - I_{n,i}}{l} - \mathbb{1}_{\{l=k\}} \binom{I_{n,i}}{k} \right] - c^{(1)}_{p,l} n^l \log n - c^{(2)}_{p,l} n^l \right. \tag{79}$$

$$+ \sum_{i=1}^{b} \sum_{m=1}^{l} (A_i)_{l,m} \left[ c^{(1)}_{p,m} I^m_{n,i} \log I_{n,i} + c^{(2)}_{p,m} I^m_{n,i} \right] \tag{80}$$

$$\left. + \mathbb{1}_{\{l=k\}} \left(c^{(2)}_{p,k} - c^{(2)}_{w,k}\right) \left(n^k - \sum_{i=1}^{b} I^k_{n,i}\right) \right) + o(1). \tag{81}$$

To be able to apply the contraction theorem (Neininger, 2001, Theorem 4.1) it remains to show that for $n \to \infty$

$$\left(A^{(n)}_1, \ldots, A^{(n)}_b, b^{(n)}\right) \xrightarrow{l_2} \left(A^*_1, \ldots, A^*_b, b^*\right), \tag{82}$$

$$\sum_{i=1}^{b} E\left\|(A_i^*)^T A_i^*\right\|_{\mathrm{op}} < 1 \tag{83}$$

and

$$E\left[\mathbb{1}_{\{I_{n,i}\le l\}\cup\{I_{n,i}=n\}}\left\|(A_i^{(n)})^T A_i^{(n)}\right\|_{\mathrm{op}}\right] \to 0 \tag{84}$$

for all $l \in \mathbb{N}$.

We start by showing (82). We have for $m \le l \le k$

$$\Big(\delta(1/n)A_i(n)\delta(I_{n,i})\Big)_{l,m} = \frac{1}{n^l} I_{n,i}^m \binom{n-I_{n,i}}{l-m} \tag{85}$$

$$= \frac{1}{(l-m)!}\frac{I_{n,i}^m}{n^m}\prod_{\nu=0}^{l-m-1}\left(1-\frac{I_{n,i}+\nu}{n}\right). \tag{86}$$

It is not difficult to show $(\delta(1/n)A_i(n)\delta(I_{n,i}))_{l,m} \to D_i^m(1-D_i)^{l-m}/(l-m)!$ almost surely using the convergence in (6). By deterministic boundedness we get convergence with respect to the $l_2$-metric. Since $(\delta(1/n)A_i(n)\delta(I_{n,i}))_{l,m} = 0$ in the cases where $m > l$ we obtain

$$\delta(1/n)A_i\delta(I_{n,i}) \xrightarrow{l_2} A_i^* \tag{87}$$

with $A_i^*$ as in the theorem.

To show the convergence of $b^{(n)}$ we first consider the term

$$(\tilde{b}_n)_l := \frac{1}{n^l}\sum_{i=1}^{b}\left\{\sum_{m=1}^{l}\binom{n-I_{n,i}}{l-m}c_{p,m}^{(1)} I_{n,i}^m \log I_{n,i}\right\} - c_{p,l}^{(1)}\log n \tag{88}$$

$$= \frac{1}{n^l}\sum_{i=1}^{b}\left\{\sum_{m=1}^{l}\frac{1}{(m-1)!}\frac{b}{b-1}E[Z_1]I_{n,i}^m\binom{n-I_{n,i}}{l-m}\log I_{n,i}\right\} - c_{p,l}^{(1)} n^{-1}\sum_{i=1}^{b} I_{n,i}\log n + o(1). \tag{89}$$

With $x^{(l)} = x^l + o(x^l)$ we have

$$\sum_{m=1}^{l}\frac{1}{(m-1)!}\binom{n-I_{n,i}}{l-m}I_{n,i}^{m-1} = \sum_{m=0}^{l-1}\frac{(n-I_{n,i})^{(l-m-1)}}{m!(l-m-1)!}I_{n,i}^m \tag{90}$$

$$= \frac{1}{(l-1)!}\sum_{m=0}^{l-1}\binom{l-1}{m}I_{n,i}^m(n-I_{n,i})^{l-1-m} + o(n^{l-1}) \tag{91}$$

$$= \frac{1}{(l-1)!}n^{l-1} + o(n^{l-1}). \tag{92}$$

Using this we obtain from (89)

$$(\tilde{b}_n)_l = \sum_{i=1}^{b} \frac{b}{b-1} \frac{1}{(l-1)!} E[Z_1] \frac{I_{n,i}}{n} \log I_{n,i} - c_{p,l}^{(1)} \sum_{i=1}^{b} \frac{I_{n,i}}{n} \log n + o(1) \tag{93}$$

$$= c_{p,l}^{(1)} \sum_{i=1}^{b} \frac{I_{n,i}}{n} \log \frac{I_{n,i}}{n} + o(1). \tag{94}$$

The convergence (6), the assumption $\mathrm{Var}(Z_1) < \infty$ and deterministic boundedness yields

$$b_l^{(n)} \xrightarrow{l_2} c_{p,l}^{(1)} \sum_{i=1}^{b} D_i \log D_i + \frac{1}{l!} \sum_{i=1}^{b} Z_i \left(1 - (1-D_i)^l - \mathbb{1}_{\{l=k\}} D_i^k\right) \tag{95}$$

$$+ \sum_{i=1}^{b} \sum_{m=1}^{l} c_{p,m}^{(2)} (A_i^*)_{l,m} - c_{p,l}^{(2)} + \mathbb{1}_{\{l=k\}} \left(c_{p,k}^{(2)} - c_{w,k}^{(2)}\right) \left(1 - \sum_{i=1}^{b} D_i^k\right) \tag{96}$$

$$=: b_l^* \tag{97}$$

and (82) is proved.

To prove (83) we observe that the eigenvalues of $A_i^*$ are given by $D_i^l$ for $l = 1, \ldots, k$. Since $D_i$ is bounded by 1 and non-negative, it is $\|A_i^*\|_{\mathrm{op}} = \|(A_i^*)^T\|_{\mathrm{op}} = D_i$. We use the inequality $\|AB\|_{\mathrm{op}} \leq \|A\|_{\mathrm{op}} \|B\|_{\mathrm{op}}$. With $\sum_{i=1}^{b} D_i^2 < 1$ almost surely we finally conclude

$$E\left[\sum_{i=1}^{b} \left\|(A_i^*)^T A_i^*\right\|_{\mathrm{op}}\right] \leq E\left[\sum_{i=1}^{b} D_i^2\right] < 1. \tag{98}$$

The condition (84) follows from the deterministic boundedness of $\|A_i^{(n)}\|_{\mathrm{op}}$ and

$$\lim_{n\to\infty} P\left(\{I_{n,i} \leq l\} \cup \{I_{n,i} = n\}\right) = 0 \tag{99}$$

which can be seen directly from equation (3). The contraction theorem (Neininger, 2001, Theorem 4.1) yields the claim. □

**Remark 5.2** *In particular, the Limit Theorem 5.1 implies* $\mathrm{Var}(P_{k,n}) = Cn^{2k} + o(n^{2k})$ *and* $\mathrm{Var}(W_{k,n}) = \tilde{C}n^{2k} + o(n^{2k})$ *with two constants* $C, \tilde{C} > 0$.

## 6 Bijection to linear recursive trees

In this section we transfer the results for the random $b$-ary recursive tree with weighted edges to linear recursive trees. The model of linear recursive trees is introduced by Pittel (1994). In this tree every node $u$ has a weight $w_u$. Starting with the root, the tree grows node by node. In each step the new node is attached to a randomly chosen node of the previous ones. The probability that node $u$ is chosen is proportional to the weight $w_u$ of the node. In the case of linear recursive trees the weight is given by $w_u = 1 + \beta \deg(u)$ where $\deg(u)$ is the number of

children of $u$ and $\beta \in \mathbb{R}_{\geq 0}$ is the parameter of the tree. As a special case the random recursive tree can be considered as a linear recursive tree with parameter $\beta = 0$. This procedure provides recursive trees with unbounded degrees of the nodes.

To transfer the results of the first part to linear recursive trees we choose a special weight vector and define an equivalence relation on the set of $b$-ary recursive trees. We choose for the edges the weight vector $(1, 0, \ldots, 0) \in \mathbb{R}^b$. The weighted depth of a node is the sum of the weights of the edges along the path from this node to the root. For every node $u$ which is the root or the leftmost child of another node we can define the set of nodes $N(u)$ which contains all nodes which lie in the subtree rooted to $u$ and have the same weighted depth as $u$. We say that two $b$-ary recursive trees are equivalent if and only if they differ only in permutations within the sets $N(u)$ for all leftmost children $u$ and the root. Here the exchange of two nodes means the exchange of the whole subtrees rooted to these nodes. Let $\psi$ denote the projection of $\mathcal{T}_b(n)$ onto the set of equivalence classes and $\mathcal{T}(n)$ denote the set of all recursive trees of size $n$. Then we have the following lemma.

**Lemma 6.1** *For any $n \in \mathbb{N}$ there exists a bijection*

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

**Proof:** Given a recursive tree $T \in \mathcal{T}(n+1)$, we define its image under the bijection $\varphi$. The node $u$ with label 2 in $T$ has to be a child of the root. We identify this node with the root $\varphi(u)$ of $\varphi(T)$. The siblings of the node with label 2 in $T$ are identified with the nodes in the set $N(\varphi(u))$ (the little squares in Figure 2). For every node $v$ in $T$, we now recursively identify its child with the smallest label (denoted by $w$) with the leftmost child $(\varphi(v), 1)$ of $\varphi(v)$ in $\varphi(T)$. The siblings of the node $w$ correspond to the nodes in $N(\varphi(v), 1)$. This construction defines a bijection. □

A more detailed prove is stated in Munsonius (2010).

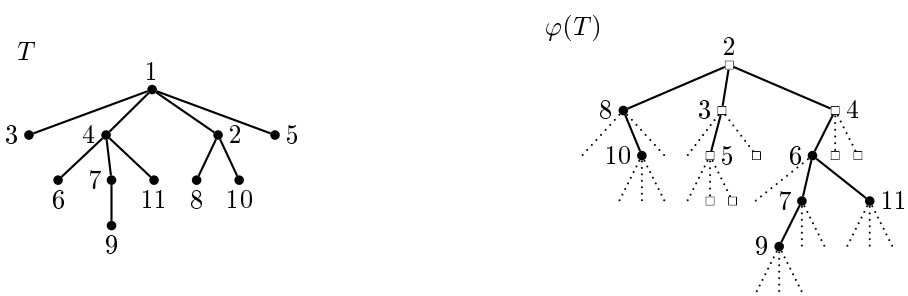


**Fig. 2:** A recursive tree $T$ with eleven nodes and its image $\varphi(T)$ (without edge weights)

From the construction of this bijection we can deduce a relation between the sizes of the subtrees in $T$ and $\varphi(T)$.

**Lemma 6.2** *Let $T$ be a linear recursive tree. For every node $v \in T$ we have*

$$|t_v| = 1 + |\tilde{t}_{(\varphi(v),1)}| \tag{100}$$

*where $t_v$ is the subtree of $T$ rooted to $v$ and $\tilde{t}_{(\varphi(v),1)}$ is the subtree of $\varphi(T)$ rooted to $(\varphi(v), 1)$.*

**Proof:** The subtree of $v \in T$ consists of $v$ and the subtrees rooted to the children of $v$. By the bijection $\varphi$ the first child of $v$ is mapped onto $(\varphi(v), 1)$. The other children of $v$ are mapped to nodes in the subtree $\tilde{t}_{(\varphi(v),1)}$. Hence for every node $u \neq v$ in $t_v$ the image $\varphi(u)$ belongs to $\tilde{t}_{(\varphi(v),1)}$.

On the other hand, if a node $\varphi(u)$ is in $\tilde{t}_{(\varphi(v),1)}$ then $u$ is a child of $v$ or it belongs to the subtree rooted at one of the children of $v$. Hence every node in $t_v$ except $v$ corresponds to a unique node in $\tilde{t}_{(\varphi(v),1)}$ and the claim follows. □

Moreover, we have that $\varphi$ maps the distribution of a random linear recursive tree on the distribution of a random $b$-ary recursive tree.

**Lemma 6.3** *For $b \in \mathbb{Z}_{\geq 0}$ let $T(n+1)$ be a random linear recursive tree of size $n+1$ with parameter $b$ and $T_{b+2}(n)$ be a random $b+2$-ary recursive tree of size $n$. Then we have*

$$\psi(T_{b+2}(n)) \stackrel{\mathcal{D}}{=} \varphi(T(n+1)). \tag{101}$$

This can be shown by induction on the size of the trees and is implicit stated in Broutin and Devroye (2006). The key idea is to equip the linear recursive tree with external nodes which all have the same probability to become the next internal one. When an external node becomes an internal one there appear $b+1$ new external siblings and one external child of the new internal node. These correspond to the $b+2$ external children of a new internal node in the $b+2$-ary recursive tree.

To translate the limit theorem for the total Steiner $k$-distance and the $k$-th internal path length, it remains to investigate the transformation of these functionals under the maps $\varphi$ and $\psi$. We recall the notation. It denotes $P_k(T)$ the $k$-th internal path length and $W_k(T)$ the total Steiner $k$-distance of a tree $T$. To obtain a transformation formula for $P_k$ and $W_k$ we first prove a representation which uses the sizes of subtrees.

**Lemma 6.4** *Given a tree $T$ with weighted edges and a node $v \in T$ we denote by $Z_v$ the weight of the edge which is between node $v$ and its parent. Further, let $t_v$ be the subtree rooted to $v$ and $|t_v|$ the size of $t_v$. Then we have*

$$P_k(T) = \sum_{v \in T} Z_v \left[ \binom{|T|}{k} - \binom{|T| - |t_v|}{k} \right] \tag{102}$$

*and*

$$W_k(T) = \sum_{v \in T} Z_v \left[ \binom{|T|}{k} - \binom{|T| - |t_v|}{k} - \binom{|t_v|}{k} \right]. \tag{103}$$

**Proof:** To prove this lemma we count how often the weight of an edge is added for the $k$-th internal path length. Each edge subdivides the tree into two subtrees, one of them containing the root. For a set $M$ of $k$ nodes, a given edge is not in the ancestor tree of $M$ if and only if all nodes of $M$ are in the subtree containing the root. Therefore, there are exactly $\binom{|T|}{k} - \binom{|T|-|t_v|}{k}$ subsets of $k$ nodes in $T$ for which the edge with weight $Z_v$ is contained in the ancestor tree.

The same arguments provide the formula for the total Steiner $k$-distance. Here, a given edge is not in the Steiner tree if and only if all nodes of $M$ are in only one of the two subtrees. □

The two Lemmata 6.2 and 6.4 yield the transformation formulae for the total Steiner $k$-distance and the $k$-th internal path length under the map $\varphi$.

**Lemma 6.5** *Let $\varphi$ be the bijection of Lemma 6.1, $T$ a recursive tree with $n+1$ nodes and $\varphi(T)$ provided with the edge weight vector $(1,0,\ldots,0)$. Then we have*

$$P_k(T) = P_k(\varphi(T)) + n\binom{n}{k-1} \tag{104}$$

*and*

$$W_k(T) = W_k(\varphi(T)) + W_{k-1}(\varphi(T)) - P_{k-1}(\varphi(T)) + n\binom{n}{k-1}. \tag{105}$$

**Proof:** We first consider the $k$-th internal path length. Since in $T$ all edge weights are equal to 1 we get from Lemma 6.4

$$P_k(T) = \sum_{u\in T}\left[\binom{n+1}{k} - \binom{n+1-|t_u|}{k}\right]. \tag{106}$$

In the tree $\varphi(T)$ all weights are 0 except the ones between one node and its leftmost child. Hence we obtain with Lemma 6.4

$$\begin{aligned} P_k(\varphi(T)) + n\binom{n}{k-1} &= \sum_{v\in\varphi(T)}\left[\binom{n}{k} - \binom{n-|\tilde{t}_{(v,1)}|}{k}\right] + n\binom{n}{k-1} && (107)\\ &= \sum_{u\in T}\left[\binom{n+1}{k} - \binom{n+1-|t_u|}{k}\right] && (108)\\ &= P_k(T). && (109) \end{aligned}$$

With the same arguments we conclude

$$\begin{aligned} &W_k(\varphi(T)) + W_{k-1}(\varphi(T)) - P_{k-1}(\varphi(T)) + n\binom{n}{k-1} && (110)\\ &= \sum_{v\in\varphi(T)}\left[\binom{n}{k} - \binom{n-|\tilde{t}_{(v,1)}|}{k} - \binom{|\tilde{t}_{(v,1)}|}{k}\right] && (111)\\ &\quad + \sum_{v\in\varphi(T)}\left[\binom{n}{k-1} - \binom{n-|\tilde{t}_{(v,1)}|}{k-1} - \binom{|\tilde{t}_{(v,1)}|}{k-1}\right] && (112)\\ &\quad - \sum_{v\in\varphi(T)}\left[\binom{n}{k-1} - \binom{n-|\tilde{t}_{(v,1)}|}{k-1}\right] + n\binom{n}{k-1} && (113)\\ &= \sum_{v\in\varphi(T)}\left[\binom{n+1}{k} - \binom{n-|\tilde{t}_{(v,1)}|}{k} - \binom{|\tilde{t}_{(v,1)}|+1}{k}\right] && (114)\\ &= \sum_{u\in T}\left[\binom{n+1}{k} - \binom{n+1-|t_u|}{k} - \binom{|t_u|}{k}\right] = W_k(T). && (115) \end{aligned}$$

□

We finally obtain the limit theorem for the total Steiner $k$-distance and the $k$-th internal path length for linear recursive trees.

**Theorem 6.6 (Limit theorem for linear recursive trees)** *For $k \in \mathbb{N}$ we denote by*

$$X_{k,n} = (P_{1,n}, P_{2,n}, \dots, P_{k-1,n}, W_{k,n})^T \tag{116}$$

*the vector consisting of the $l$-th internal path lengths ($l = 1, \dots, k-1$) and the total Steiner $k$-distance of a random linear recursive tree of size $n$ with parameter $b-2 \in \mathbb{Z}_{\geq 0}$. Then we have for $n \to \infty$*

$$E[P_{k,n}] = \frac{1}{(k-1)!}\frac{1}{b-1}n^k \log n + \left(c_{p,k}^{(2)} + \frac{1}{(k-1)!}\right) n^k + o(n^k), \tag{117}$$

$$E[W_{k,n}] = \frac{1}{(k-1)!}\frac{1}{b-1}n^k \log n + \left(c_{w,k}^{(2)} + \frac{1}{(k-1)!}\right) n^k + o(n^k) \tag{118}$$

*and*

$$X_{k,n}^* := \begin{bmatrix} \frac{1}{n} & 0 & \cdots & 0 \\ 0 & \frac{1}{n^2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \frac{1}{n^k} \end{bmatrix} (X_{k,n} - E[X_{k,n}]) \xrightarrow{l_2} X_k, \tag{119}$$

*where the distribution of $X_k$ is given in Theorem 5.1 and $c_{p,k}^{(2)}$ and $c_{w,k}^{(2)}$ are the same as in Theorem 4.5 and Theorem 4.6.*

**Proof:** Let $T_b(n-1)$ be a random $b$-ary recursive tree with edge weights $Z^{(0)}$ where

$$P(Z^{(0)} = e_i) = \frac{1}{b} \tag{120}$$

for all $i = 1, \dots, b$, where $e_i$ is the vector consisting of 0 everywhere except for the $i$-th entry which is 1. We denote by $\tilde{T}_b(n-1)$ the tree obtained by ordering the siblings (and the corresponding subtrees) in $T_b(n-1)$ in such a way that the edge with weight 1 is always the leftmost one. The $k$-th internal path length and the total Steiner $k$-distance are invariant under this rearrangement. Further, the sizes of the subtrees which belong to the edges with edge weight 1 depend only on the image under $\psi$. Thus, we see that $P_k$ and $W_k$ are invariant under the map $\psi$. In total we have

$$P_k(\psi(\tilde{T}_b(n-1))) = P_k(T_b(n-1)) \tag{121}$$

and

$$W_k(\psi(\tilde{T}_b(n-1))) = W_k(T_b(n-1)). \tag{122}$$

Let $T_n$ be a random linear recursive tree of size $n$ with weight function $u \mapsto 1 + (b-2)\deg(u)$ and $\varphi$ be the bijection of Lemma 6.1. By Lemma 6.3 it holds $\varphi(T_n) \stackrel{\mathcal{D}}{=} \psi(\tilde{T}_b(n-1))$ and therefore according to (121) and (122)

$$P_k(\varphi(T_n)) \stackrel{\mathcal{D}}{=} P_k(T_b(n-1)) \tag{123}$$

as well as

$$W_k(\varphi(T_n)) \stackrel{\mathcal{D}}{=} W_k(T_b(n-1)). \tag{124}$$

This implies by Lemma 6.5

$$E[P_{k,n}] = E[P_k(T_b(n-1))] + \frac{1}{(k-1)!}n^k + o(n^k) \tag{125}$$

and

$$E[W_{k,n}] = E[W_k(T_b(n-1))] + E[W_{k-1}(T_b(n-1))] - E[P_{k-1}(T_b(n-1))] + \frac{1}{(k-1)!}n^k + o(n^k). \tag{126}$$

With Theorems 4.5 and 4.6 the asymptotic formulae for the expectations follow.

Let $\tilde{X}^*_{k,n-1}$ be the vector corresponding to $X^*_{k,n}$ for the random $b$-ary recursive tree $T_b(n-1)$. Lemma 6.5 yields with equations (123) and (124)

$$X^*_{k,n} \stackrel{\mathcal{D}}{=} \tilde{X}^*_{k,n-1} + \begin{pmatrix} 0 \\ \vdots \\ 0 \\ Y \end{pmatrix} \tag{127}$$

where

$$Y := \frac{W_{k-1}(T_b(n-1)) - E[W_{k-1}(T_b(n-1))]}{n^k} - \frac{P_{k-1}(T_b(n-1)) - E[P_{k-1}(T_b(n-1))]}{n^k}. \tag{128}$$

Remark 5.2 implies $Y \xrightarrow{l_2} 0$. Thus, the claim follows by Theorem 5.1. □

**Remark 6.7** *Random plane oriented recursive trees without the order of the nodes equals in distribution the random linear recursive tree with parameter $\beta = 1$. Since the $k$-th internal path length as well as the total Steiner $k$-distance are invariant under changing of the order of the tree the Limit Theorem 6.6 with $b = 3$ provides in particular the corresponding limit theorem for the plane oriented recursive tree.*

## Acknowledgements

The author thanks Ralph Neininger for posing the problem and leading towards an approach to the Steiner $k$-distance. Furthermore, he thanks an unknown referee for several hints to literature and improvements of the paper.

## References

T. Ali Khan and R. Neininger. Tail bounds for the Wiener index of random trees. In *2007 Conference on Analysis of Algorithms, AofA 07*, Discrete Math. Theor. Comput. Sci. Proc., AH, pages 279–289. Assoc. Discrete Math. Theor. Comput. Sci., Nancy, 2007.

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

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

P. J. Bickel and D. A. Freedman. Some asymptotic theory for the bootstrap. *Ann. Statist.*, 9(6): 1196–1217, 1981.

N. Broutin and L. Devroye. Large deviations for the weighted height of an extended class of trees. *Algorithmica*, 46(3–4):271–297, 2006.

N. Broutin, L. Devroye, E. McLeish, and M. de la Salle. The height of increasing trees. *Random Structures Algorithms*, 32(4):494–518, 2008.

L. H. Clark, A. Meir, and J. W. Moon. On the Steiner distance of trees from certain families. *Australas. J. Combin.*, 20:47–68, 1999.

P. Dankelmann, O. R. Oellermann, and H. C. Swart. The average Steiner distance of a graph. *J. Graph Theory*, 22(1):15–22, 1996.

R. P. Dobrow and J. A. Fill. Total path length for random recursive trees. *Combin. Probab. Comput.*, 8(4):317–333, 1999. Random graphs and combinatorial structures (Oberwolfach, 1997).

A. A. Dobrynin, R. Entringer, and I. Gutman. Wiener index of trees: theory and applications. *Acta Appl. Math.*, 66(3):211–249, 2001.

F. Guillemin and P. Robert. Analysis of Steiner subtrees of random trees for traceroute algorithms. *Random Structures Algorithms*, 35(2):194–215, 2009.

S. Janson. The Wiener index of simply generated random trees. *Random Structures Algorithms*, 22(4):337–358, 2003.

S. Janson and P. Chassaing. The center of mass of the ISE and the Wiener index of trees. *Electron. Comm. Probab.*, 9:178–187 (electronic), 2004.

M. Javanian and M. Q. Vahidi-Asl. Depth of nodes in random recursive $k$-ary trees. *Inform. Process. Lett.*, 98(3):115–118, 2006.

N. L. Johnson and S. Kotz. *Urn Models and Their Application.* John Wiley & Sons, New York-London-Sydney, 1977. An approach to modern discrete probability theory, Wiley Series in Probability and Mathematical Statistics.

H. M. Mahmoud. Limiting distributions for path lengths in recursive trees. *Probab. Engrg. Inform. Sci.*, 5(1):53–59, 1991.

K. Morris, A. Panholzer, and H. Prodinger. On some parameters in heap ordered trees. *Combin. Probab. Comput.*, 13(4–5):677–696, 2004.

G. O. Munsonius. *Limit Theorems for Functionals of Recursive Trees.* PhD thesis, University of Freiburg, Germany, 2010.
`http://www.freidok.uni-freiburg.de/volltexte/7472/`

R. Neininger. On a multivariate contraction method for random recursive structures with applications to Quicksort. *Random Structures Algorithms*, 19(3–4):498–524, 2001. Analysis of algorithms (Krynica Morska, 2000).

R. Neininger. The Wiener index of random trees. *Combin. Probab. Comput.*, 11(6):587–597, 2002.

R. Neininger and L. Rüschendorf. A general limit theorem for recursive algorithms and combinatorial structures. *Ann. Appl. Probab.*, 14(1):378–418, 2004.

R. Neininger and L. Rüschendorf. On the internal path length of $d$-dimensional quad trees. *Random Structures Algorithms*, 15(1):25–41, 1999.

A. Panholzer. The distribution of the size of the ancestor-tree and of the induced spanning subtree for random trees. *Random Structures Algorithms*, 25(2):179–207, 2004a.

A. Panholzer. Distribution of the Steiner distance in generalized $M$-ary search trees. *Combin. Probab. Comput.*, 13(4–5):717–733, 2004b.

A. Panholzer and H. Prodinger. Spanning tree size in random binary search trees. *Ann. Appl. Probab.*, 14(2):718–733, 2004a.

A. Panholzer and H. Prodinger. Analysis of some statistics for increasing tree families. *Discrete Math. Theor. Comput. Sci.*, 6(2):437–460 (electronic), 2004b.

B. Pittel. Note on the heights of random recursive trees and random $m$-ary search trees. *Random Structures Algorithms*, 5(2):337–347, 1994.

M. Régnier. A limiting distribution for quicksort. *RAIRO Inform. Théor. Appl.*, 23(3):335–343, 1989.

U. Rösler. A limit theorem for "Quicksort". *RAIRO Inform. Théor. Appl.*, 25(1):85–100, 1991.

S. Roura. Improved master theorems for divide-and-conquer recurrences. *J. ACM*, 48(2):170–205, 2001.