Deterministic edge-weights in increasing tree families.

Markus Kuba and Stephan Wagner*

Institut für Diskrete Mathematik und Geometrie
Technische Universität Wien
Wiedner Hauptstr. 8-10/104, 1040 Wien, Austria
kuba@dmg.tuwien.ac.at

and

Department of Mathematical Sciences
Stellenbosch University
Private Bag X1, Matieland 7602, South Africa
swagner@sun.ac.za

April 2, 2008

Abstract

In this work we study edge weights for two specific families of increasing trees, which include binary increasing trees and plane oriented recursive trees as special instances, where plane-oriented recursive trees serve as a combinatorial model of scale-free random trees given by the $m = 1$ case of the Barabási-Albert model. An edge $e = (k, l)$, connecting the nodes labeled k and l, respectively, in an increasing tree, is associated with the weight $w_e = |k - l|$. We are interested in the distribution of the number of edges with a fixed edge weight j in a random generalized plane oriented recursive tree or random d-ary increasing tree. We provide exact formulas for expectation and variance and prove a normal limit law for this quantity. A combinatorial approach is also presented and applied to a related parameter, the maximum edge weight.

Keywords: Increasing trees, Scale-free trees, plane oriented recursive trees, deterministic edge-weights, j-independence, limiting distribution

2000 Mathematics Subject Classification 05C05, 60F05.

1 Introduction

1.1 Increasing trees

Increasing trees are rooted labeled trees where the nodes of a tree of size n are labeled by distinct integers from the set $\{1, \dots, n\}$ in such a way that the sequence of labels along any branch starting at the root is increasing. In this paper, we will consider two specific combinatorial models of increasing trees, namely the family of so-called generalized plane oriented recursive trees (often abbreviated as "gports") and d-ary increasing trees.

The interest in these two tree families stems from the fact that several important tree models, such as plane-oriented recursive trees and binary increasing trees (also called tournament trees), are special instances of these families. These tree models are of importance in many applications. Plane-oriented recursive trees are a special instance of the well known Barabási-Albert model [2] for scale-free networks (see also [7]), which is used as a simplified growth model of the world wide web [1]. Binary increasing trees ($d = 2$) are of special importance in computer science, since they are isomorphic to binary search trees, which in turn serve as an analytic model for the famous Quicksort algorithm [11].

Generalized plane oriented recursive trees and d-ary increasing trees can also be described via a tree evolution process, as pointed out in [14]. For every tree T' of size n with vertices $v_1, \dots, v_n$ one can give probabilities $p_{T'}(v_1), \dots, p_{T'}(v_n)$, such that when starting with a *random tree* T' of size n of the tree

*This work was supported by the Austrian Science Foundation FWF, grants S9608 and S9611.

family considered, choosing a vertex v_i in T' at random according to the probabilities $p_{T'}(v_i)$ and attaching node $n+1$ to it, one obtains again a *random tree* T of size $n+1$ of the tree family considered.

1.2 Deterministic edge weights

Let T denote an increasing tree of size n, where T is either a generalized plane oriented recursive tree or a d-ary increasing tree. We consider edge-weighted increasing trees, where every edge $e \in E = E(T)$ of the tree will be weighted deterministically as follows. If the edge $e = (k, l)$ is adjacent to the nodes (labeled) k and l, then we define the weight w_e of the edge e as $w_e := |k - l|$. The notion of edge weights provides a natural new cost measure for constructing increasing trees (i.e. scale free networks): the smaller the sum of the edge weights, the cheaper the construction. Let us denote by E_n the set of edges of a random increasing tree of size n. The aim of this paper is to study the random variable $S_{n,j} := \sum_{e \in E_n} \mathbb{I}_{\{w_e = j\}}$, counting the number of edge weights of size j in a size n random increasing tree. Here, $\mathbb{I}_{\{w_e=j\}}$ stands for the indicator variable of the event that e has weight j. An alternative representation of $S_{n,j}$ is obtained by the growth process generating random increasing trees of size n:

$$S_{n,j} = \sum_{k=j+1}^{n} \mathbb{I}_{\{k <_c k-j\}}, \tag{1}$$

where $k <_c k - j$ denotes the event that node k is a child of (attached to) node $k - j$. In the following we will use both combinatorial and probabilistic methods to analyze the distribution of $S_{n,j}$. In order to obtain exact results for expectation and variance we will proceed similarly to [7]. Using a decomposition for $S_{n,j}$ and a theorem concerning weakly dependent random variables, we will be able to show that for arbitrary but fixed $j \in \mathbb{N}$, the random variable $S_{n,j}$ is asymptotically normal distributed.

A combinatorial approach via generating functions allows one to compute all probabilities $\mathbb{P}(S_{n,j} = m)$, but it seems to be impossible to derive the normal law from it—as we will see, the recursions lead to large systems of differential equations with no nice explicit solutions. However, the same combinatorial approach turns out to be somewhat more useful for the analysis of the probability $\mathbb{P}(M_n \leq k)$, where the random variable $M_n = \max_{e \in E_n} w_e$ is the maximal edge weight in a random increasing tree of size n. The analysis of the random variables $S_{n,j}$ and M_n is much easier for recursive trees. An extensive study for recursive trees was conducted in [13], where it was also shown that for recursive trees the edge weights w_e have an intimate relationship with entries in inversion tables of permutations.

1.3 Notation

We use the abbreviations $x^{\underline{l}} := x(x-1)\cdots(x-l+1)$ and $x^{\overline{l}} := x(x+1)\cdots(x+l-1)$ for the falling and rising factorials, respectively. Furthermore, we denote by $X \stackrel{\mathcal{L}}{=} Y$ the equality in distribution of random variables X and Y, by $X \oplus Y$ the sum of two independent random variables and by $X + Y$ the sum of not necessarily independent random variables. Moreover, we write $X_1 \oplus \cdots \oplus X_l$ for the sum of mutually independent random variables. We also denote by $X_n \xrightarrow{\mathcal{L}} X$ the weak convergence, i. e. the convergence in distribution, of the sequence of random variables X_n to a random variable X.

Throughout this work we often use the abbreviation “gports”, standing for generalized plane oriented recursive trees. Note that we use the following notations interchangeably: $\alpha = -\frac{c_1}{c_2} - 1 = -d$.

1.4 Plan of the Paper

The paper is organized as follows: in the next section, we describe the construction of the tree families we investigate. Then, we study the distribution of edge weights first in the simple case $j = 1$, then in general. By means of an approach that is due to Bollobás and Riordan [7], we find an explicit formula for the probability that a certain set of edges is contained in a random tree. This allows us to determine exact and asymptotic formulas for the mean and variance. Finally, we prove a central limit theorem for the number of edges with a specific weight. After a short section on edge weight tables, we consider a combinatorial approach that is applied to the study of the quantities “number of edges with a given weight” and “maximum edge weight”. However, it turns out that the probabilistic approach usually yields much stronger results.

2 Preliminaries

2.1 A combinatorial description of increasing trees

In the following we give a general combinatorial definition of increasing trees (including the families of gports and d-ary increasing trees). Formally, a *simple family of increasing trees* $\mathcal{T}$ can be defined in the following way. We start with a sequence of non-negative numbers $(\varphi_k)_{k\geq 0}$, where $\varphi_0 > 0$. The sequence $(\varphi_k)_{k\geq 0}$ is called the degree-weight sequence. We assume that there exists a $k \geq 2$ with $\varphi_k > 0$. The degree-weight sequence is used to define the weight $w(T)$ of any ordered tree T by $w(T) := \prod_v \varphi_{d(v)}$, where v ranges over all vertices of T and $d(v)$ is the out-degree of v. Furthermore, $\mathcal{L}(T)$ denotes the set of different increasing labellings of the tree T with distinct integers $\{1, 2, \dots, |T|\}$, where $|T|$ denotes the size of the tree T, and $L(T) := \big|\mathcal{L}(T)\big|$ its cardinality. Then the family $\mathcal{T}$ consists of all trees T together with their weights $w(T)$ and the set of increasing labellings $\mathcal{L}(T)$. For a given degree-weight sequence $(\varphi_k)_{k\geq 0}$ with a degree-weight generating function $\varphi(t) := \sum_{k\geq 0} \varphi_k t^k$, we now define the total weights by $T_n := \sum_{|T|=n} w(T) \cdot L(T)$. It follows that the exponential generating function $T(z) := \sum_{n\geq 1} T_n \frac{z^n}{n!}$ satisfies the autonomous first order differential equation

$$T'(z) = \varphi\big(T(z)\big), \quad T(0) = 0. \tag{2}$$

This can be deduced from the fact that a simple family of increasing trees $\mathcal{T}$ is described by the formal recursive equation

$$\mathcal{T} = \textcircled{1} \times \Big(\varphi_0 \cdot \{\epsilon\} \mathbin{\dot\cup} \varphi_1 \cdot \mathcal{T} \mathbin{\dot\cup} \varphi_2 \cdot \mathcal{T} * \mathcal{T} \mathbin{\dot\cup} \varphi_3 \cdot \mathcal{T} * \mathcal{T} * \mathcal{T} \mathbin{\dot\cup} \cdots\Big) = \textcircled{1} \times \varphi(\mathcal{T}), \tag{3}$$

where $\textcircled{1}$ denotes the node labeled 1, $\times$ the Cartesian product, $\dot\cup$ the disjoint union, $*$ the partition product for labeled objects, and $\varphi(\mathcal{T})$ the substituted structure (see for instance [15] or [6]). In short, this formal recursive equation corresponds to the fact that we may describe a tree as a root node with several subtrees of the same family attached to it. Next we are going to specify the degree-weight generating function for the tree families that are investigated in this paper. Generalized plane-oriented recursive trees and d-ary increasing trees are characterized by the degree-weight generating functions

$$\varphi(t) = \begin{cases} \dfrac{\varphi_0}{(1 + \frac{c_2 t}{\varphi_0})^{\alpha}}, & \text{for generalized plane-oriented recursive trees} \\ \varphi_0 \Big(1 + \frac{c_2 t}{\varphi_0}\Big)^d & \text{for } d\text{-ary increasing trees ,} \end{cases}$$

where $\alpha := -\frac{c_1}{c_2} - 1$ with $\varphi_0 > 0$ and $0 < -c_2 < c_1$; and $d := \frac{c_1}{c_2} + 1 \in \{2, 3, 4, \dots\}$ with $\varphi_0 > 0$ and $c_2 > 0$. Identifying d and $-\alpha$, we see that the definitions are very similar. By solving the differential equation (2) with respect to the degree-weight generating functions $\varphi(t)$, and extraction of coefficients one obtains a formula for the total weight T_n of generalized plane-oriented recursive trees, and d-ary increasing trees,

$$T_n = \varphi_0 c_1^{n-1} (n-1)! \binom{n - 1 + \frac{c_2}{c_1}}{n - 1}, \tag{4}$$

with φ_0, c_1 and c_2 as specified for the particular tree family.

2.2 Description via a tree evolution processes

As mentioned before, generalized plane oriented recursive trees and d-ary increasing trees can be generated by an evolution process. This description is a consequence of the considerations made in [14]. The process generates random trees of arbitrary size n. The process starts with the root labeled by 1. At step $i + 1$ the node with label $i + 1$ is attached to any previous node v (with out-degree $d^+(v)$) of the already grown tree of size i with probability $p(v)$ equal to

$$p(v) = \begin{cases} \dfrac{\deg^+(v) + \alpha}{(\alpha + 1)i - 1}, & \text{for generalized plane-oriented recursive trees,} \\ \dfrac{d - \deg^+(v)}{(d - 1)i + 1}, & \text{for } d\text{-ary increasing trees,} \end{cases}$$

with $d \in \mathbb{N} \setminus \{1\}$ and $\alpha > 0$ as before. $d - \deg^+(v)$ and $\deg^+(v) + \alpha$ are interpreted as the number of places where a new node can be attached to v, even if the latter is not necessarily an integer. Hence, the process associated to generalized plane oriented recursive trees generalizes the preferential attachment rule of the Barabási-Albert model $m = 1$.

2.3 Examples

Example 1. Plane-oriented recursive trees are the family of plane increasing trees such that all node degrees are allowed and assigned equal weights. The degree-weight generating function is thus $\varphi(t) = \frac{1}{1-t}$. Equation (2) leads to

$$T(z) = 1 - \sqrt{1-2z}, \quad \text{and} \quad T_n = \tfrac{(n-1)!}{2^{n-1}}\tbinom{2n-2}{n-1} = 1 \cdot 3 \cdot 5 \cdots (2n-3) = (2n-3)!!, \quad n \geq 1.$$

Moreover, the probability of attaching to a node v in a tree of size i is given by $p(v) = \frac{\deg^+(v)+1}{2i-1}$, which corresponds to the case $m = 1$ of the Barabási-Albert model.

Example 2. Binary increasing trees have the degree-weight generating function $\varphi(t) = (1+t)^2$. Thus it follows that

$$T(z) = \frac{z}{1-z}, \quad \text{and} \quad T_n = n!, \ \text{ for } n \geq 1.$$

Moreover, the the probability of attaching to a node v in a tree of size i is given by $p(v) = \frac{2-\deg^+(v)}{i+1}$.

Bearing the special cases of ordinary plane oriented trees and d-ary trees in mind, we will use the expression "number of increasing trees" (with a certain number of nodes and within a given family of increasing trees), even though "total weight" would be more appropriate (note that the total weight is not even necessarily an integer if α is not).

3 The distribution of edge weights: case $j = 1$

First of all, we will discuss the case $j = 1$, which turns out to be somewhat simpler compared to the general case. We obtain explicit results for the probability distribution and a normal limit law as n tends to infinity. The key tool for studying $S_{n,1}$ is the following Lemma, which provides the independence of the indicator variables.

Lemma 1. *The random variable $S_{n,1}$, counting the number of edge weights of size 1 in a size n random gport or d-ary increasing tree, satisfies the decomposition*

$$S_{n,j} = \bigoplus_{k=2}^{n} \mathbb{I}_{\{k <_c k-1\}}, \tag{5}$$

with the indicators being mutually independent.

Proof. We simply condition on the event that node $n-1$ is adjacent to node n.

$$\begin{aligned}\mathbb{P}\{S_{n,1} = m\} &= \mathbb{P}\{S_{n,1} = m | n <_c n-1\}\mathbb{P}\{n <_c n-1\} + \mathbb{P}\{S_{n,1} = m | n \not<_c n-1\}\mathbb{P}\{n \not<_c n-1\} \\ &= \mathbb{P}\{S_{n-1,1} = m-1\}\mathbb{P}\{n <_c n-1\} + \mathbb{P}\{S_{n-1,1} = m\}\mathbb{P}\{n \not<_c n-1\}.\end{aligned}$$

Hence, we obtain the stated result by iterating this argument. □

The following lemma gives an explicit formula for the probabilities $\mathbb{P}\{k <_c i\} = \mathbb{E}(\mathbb{I}_{\{k <_c i\}})$.

Lemma 2 (Dobrow and Smythe [9])**.** *The probability that the node k is attached to node i, with $1 \leq i < k$, in a size n random grown simple increasing tree is given by*

$$\mathbb{P}\{k <_c i\} = \mathbb{P}\{i+1 <_c i\} \prod_{l=i+1}^{k-1} \mathbb{P}\{l+1 \not<_c l\} = \frac{1+\frac{c_2}{c_1}}{i+\frac{c_2}{c_1}} \prod_{l=i+1}^{k-1}\left(1 - \frac{1+\frac{c_2}{c_1}}{l+\frac{c_2}{c_1}}\right) = \frac{\binom{i-1+\frac{c_2}{c_1}}{i-1}}{\binom{k-1+\frac{c_2}{c_1}}{k-2}}.$$

Now we are ready to state our result concerning $S_{n,1}$, which also appeared in a different context in Dobrow and Smythe [9], and Panholzer and Prodinger [14], concerning the depth of node n in a size n random increasing tree.

Theorem 1. *The random variable $S_{n,1}$ satisfies the following distributional decomposition.*

$$S_{n,1} \stackrel{(d)}{=} B_1 \oplus \cdots \oplus B_{n-1}, \tag{6}$$

where $B_k \stackrel{(d)}{=} \mathrm{Be}(\frac{1+\frac{c_2}{c_1}}{k+\frac{c_2}{c_1}})$ is Bernoulli distributed for $1 \le k \le n-1$. The probability distribution of $S_{n,1}$ is given by

$$\mathbb{P}\{S_{n,1} = m\} = \frac{(1+\frac{c_2}{c_1})^m}{(n-1+\frac{c_2}{c_1})^{\underline{n-1}}} \begin{bmatrix} n-1 \\ m \end{bmatrix}, \tag{7}$$

where $\begin{bmatrix} n \\ m \end{bmatrix}$ denotes the signless Stirling numbers of the first kind. The expectation and the variance of $S_{n,1}$ are given by the following exact and asymptotic expressions.

$$\begin{aligned}
\mathbb{E}(S_{n,1}) &= \Big(1+\frac{c_2}{c_1}\Big)\Big(\Psi\Big(n+\frac{c_2}{c_1}\Big) - \Psi\Big(1+\frac{c_2}{c_1}\Big)\Big) \sim \Big(1+\frac{c_2}{c_1}\Big)\log(n) + \mathcal{O}(1),\\
\mathbb{V}(S_{n,1}) &= \Big(1+\frac{c_2}{c_1}\Big)\Big(\Psi\Big(n+\frac{c_2}{c_1}\Big) - \Psi\Big(1+\frac{c_2}{c_1}\Big)\Big) + \Big(1+\frac{c_2}{c_1}\Big)^2\Big(\Psi_1\Big(n+\frac{c_2}{c_1}\Big) - \Psi_1\Big(1+\frac{c_2}{c_1}\Big)\Big)\\
&\sim \Big(1+\frac{c_2}{c_1}\Big)\log n + \mathcal{O}(1),
\end{aligned}$$

where $\Psi(z) = \frac{d}{dz}\log(\Gamma(z)) = \frac{\Gamma'(z)}{\Gamma(z)}$ denotes the Digamma function and $\Psi_1(z) = \frac{d^2}{dz^2}\log(\Gamma(z))$ the Trigamma function. Furthermore the centered and normalized random variable $S_{n,1}^*$ is asymptotically normal distributed:

$$S_{n,1}^* = \frac{S_n - \mathbb{E}(S_{n,1})}{\sqrt{\mathbb{V}(S_{n,1})}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1), \tag{8}$$

where $\mathcal{N}(0,1)$ denotes the standard normal distribution.

Proof (sketch). The expectation and the variance follow immediately from the distributional decomposition and the asymptotic expansion of the Ψ-function. The probabilities can easily be obtained from the probability generating function as follows.

$$\begin{aligned}
\mathbb{P}\{S_{n,1} = m\} &= [v^m]p_n(v) = [v^m]\prod_{k=1}^{n-1}\left(\frac{k-1+v(1+\frac{c_2}{c_1})}{k+\frac{c_2}{c_1}}\right) = [v^m]\frac{\binom{v(1+\frac{c_2}{c_1})+n-2}{n-1}}{\binom{n-1+\frac{c_2}{c_1}}{n-1}}\\
&= \frac{[z^n v^m]}{\binom{n-1+\frac{c_2}{c_1}}{n-1}}\sum_{k\ge 1}\binom{v(1+\frac{c_2}{c_1})+k-2}{k-1}z^k = \frac{[z^{n-1}v^m]}{\binom{n-1+\frac{c_2}{c_1}}{n-1}}\frac{1}{(1-z)^{v(1+\frac{c_2}{c_1})}}\\
&= \frac{(1+\frac{c_2}{c_1})^m[z^{n-1}v^m]}{\binom{n-1+\frac{c_2}{c_1}}{n-1}}\frac{1}{(1-z)^v} = \frac{(1+\frac{c_2}{c_1})^m}{(n-1+\frac{c_2}{c_1})^{\underline{n-1}}}\begin{bmatrix} n-1 \\ m \end{bmatrix},
\end{aligned} \tag{9}$$

where we have used $[z^n]1/(1-z)^{\alpha+1} = \binom{\alpha+n}{n}$ and the expansion of the generating function for the signless Stirling numbers of the first kind, which is given by

$$1/(1-z)^v = \sum_{n\ge 0}\sum_{m\ge 0}\begin{bmatrix} n \\ m \end{bmatrix}\frac{z^n}{n!}v^m.$$

For the normal limit law either apply Hwang's Quasi Power theorem [10], as done in [14], or use Poisson approximation techniques [9]. □

Remark 1. Note that by Lemma 5 the random variable $S_{n,1}$ satisfies the same distribution as the random variable D_n counting the depth of node n in a size n random increasing tree. The depth has been studied before independently by Dobrow and Smythe [9] and Panholzer and Prodinger [14].

4 The distribution of edge weights: case $j > 1$

In the case $j > 1$ the indicator variables $\mathbb{I}_{\{k <_c k-j\}}$ are by definition of the growth process not mutually independent any more. For example, for plane oriented recursive trees we have

$$\mathbb{P}\{n <_c n-2 | n-1 <_c n-2\} = \frac{2}{2n-3} \neq \mathbb{P}\{n <_c n-2 | n-1 \not<_c n-2\} = \frac{1}{2n-3}. \tag{10}$$

Therefore we turn our attention to the derivation of an exact formula for the variance of $S_{n,j}$ by other means. For the exact variance we need to calculate probabilities of the form $\mathbb{P}\{k_1 <_c i_1, k_2 <_c i_2\}$, assuming that $i_1 < k_1$, $i_2 < k_2$. In order to derive these probabilities we use an approach (due to Bollobás and Riordan

[7]) that involves the calculation of a more general quantity. We determine the probability that a subgraph S is present in a random tree of size n, i.e. we calculate probabilities $\mathbb{P}\{k_1 <_c i_1, \dots k_l <_c i_l\}$. Note that in [7] the probabilities $\mathbb{P}\{k_1 <_c i_1, \dots k_l <_c i_l\}$ were derived for the special case of plane oriented recursive trees (corresponding to $\alpha = 1$ in the growth process), we also refer to [8] for an application of such a result.

4.1 An exact formula for the expectation and variance for $j > 1$

We fix a graph S with nodes $V(S)$ and edges $E(S)$. $E(S)$ represents the collection of events $\{k_1 <_c i_1\}, \{k_2 <_c i_2\}, \dots$ S is a possible subgraph of B_n for large n, where B_n denotes a tree of size n. Furthermore orient each edge $e = (i, j) \in E(S)$ with $i < j$ from j to i. We write $V^+(S)$ for the set of vertices of S from which edges leave and $V^-(S)$ for those vertices at which edges arrive. Note that usually $V^+(S) \cap V^-(S) \neq \emptyset$. For $i \in V^-(S)$ let $g_S^{[in]}(i)$ denote the in-degree of i in S. We obtain the following result.

Theorem 2. *The probability p_S that a given graph S is a subgraph of B_n is given by the following formulas.*

$$p_S = \prod_{i \in V^-(S)} \alpha^{\overline{g_S^{[in]}(i)}} \prod_{i \in V^+(S)} \frac{1}{(\alpha+1)(i-1)-1} \prod_{k \notin V^+(S)} \Big(1 + \frac{C_S(k)}{(\alpha+1)(k-1)-1}\Big)$$

for generalized plane oriented recursive trees and

$$p_S = \prod_{i \in V^-(S)} d^{\underline{g_S^{[in]}(i)}} \prod_{i \in V^+(S)} \frac{1}{(d-1)(i-1)+1} \prod_{k \notin V^+(S)} \Big(1 - \frac{C_S(k)}{(d-1)(k-1)+1}\Big),$$

for d-ary increasing trees. Here $C_S(m)$ denotes the number of edges $e = (i, l) \in E(S)$ with $i < m$ and $l \geq m$.

Before we turn to the proof of Theorem 2, we state two of its applications. For $j > 1$ we do not have a decomposition of $S_{n,j}$ with mutually independent indicator variables for generalized plane oriented recursive trees and d-ary increasing trees. Nevertheless, the next result shows that for $j \in \mathbb{N}$ there is only a local dependency structure of the indicator variables.

Corollary 1. *The indicator variables $\mathbb{I}_{A_k}$ of the events $A_k = \{k <_c k - j\}$, for $j + 1 \leq k \leq n$, are weakly dependent, or j-independent, which means whenever I and L are two subsets of the positive integers $\{j + 1, \dots, n\}$ with $\min\{|i - l| : i \in I, l \in L\} > j - 1$, then the subsystems $(\mathbb{I}_{A_i}, i \in I)$ and $(\mathbb{I}_{A_l}, l \in L)$ are independent.*

Remark 2. Note that Theorem 2 together with Corollary 1 extend the results of Lemma 2 and parts of Theorem 1. Moreover, using the convention $d = -\alpha$, the two formulas stated in Theorem 2 are basically equivalent.

Corollary 2. *The probability $\mathbb{P}\{k_1 <_c i_1, k_2 <_c i_2\}$, with $1 \leq i_1 < k_1$, $1 \leq i_2 < k_2$ and $k_2 > k_1$, is given by the following closed formulas, using $\alpha = -1 - \frac{c_1}{c_2} = -d$,*

$$\begin{aligned}
&\textit{Case } i_2 \geq k_1: \quad \mathbb{P}\{k_1 <_c i_1, k_2 <_c i_2\} = \mathbb{P}\{k_1 <_c i_1\}\mathbb{P}\{k_2 <_c i_2\} = \frac{\binom{i_1-1+\frac{c_2}{c_1}}{i_1-1}\binom{i_2-1+\frac{c_2}{c_1}}{i_2-1}}{\binom{k_1-1+\frac{c_2}{c_1}}{k_1-2}\binom{k_2-1+\frac{c_2}{c_1}}{k_2-2}},\\
&\textit{Case } i_1 < i_2 < k_1: \quad \mathbb{P}\{k_1 <_c i_1, k_2 <_c i_2\} = \frac{(1+\frac{c_2}{c_1})}{(k_1-1)} \frac{\binom{i_1-1+\frac{c_2}{c_1}}{i_1-1}\binom{k_1-2-\frac{c_2}{c_1}}{k_1-2}}{\binom{k_2-1+\frac{c_2}{c_1}}{k_2-2}\binom{i_2-1-\frac{c_2}{c_1}}{i_2-1}}\\
&\textit{Case } i = i_1 = i_2: \quad \mathbb{P}\{k_1 <_c i, k_2 <_c i\} = \frac{1}{k_1-1} \frac{\binom{i-1+\frac{c_2}{c_1}}{i-1}\binom{k_1-2-\frac{c_2}{c_1}}{k_1-2}}{\binom{k_2-1+\frac{c_2}{c_1}}{k_2-2}\binom{i-1-\frac{c_2}{c_1}}{i-1}}\\
&\textit{Case } i_1 > i_2: \quad \mathbb{P}\{k_1 <_c i_1, k_2 <_c i_2\} = \frac{(1+\frac{c_2}{c_1})}{(k_1-1)} \frac{\binom{i_2-1+\frac{c_2}{c_1}}{i_2-1}\binom{k_1-2-\frac{c_2}{c_1}}{k_1-2}}{\binom{k_2-1+\frac{c_2}{c_1}}{k_2-2}\binom{i_1-1-\frac{c_2}{c_1}}{i_1-1}}.
\end{aligned} \tag{11}$$

Our proof of Theorem 2 closely follows the arguments of [7] for the special case of plane oriented recursive trees. Let S_n denote the restriction of S up to time n: S_n consists of edges $E = (i, l)$, with $i, l \le n$. First we need some notations. We denote by $X_{n,i}$ the outdegree and by

$$\beta_{n,i} = \begin{cases} X_{n,i} + \alpha, \\ d - X_{n,i}, \end{cases} \tag{12}$$

the "actual node degree" responsible for the connectivity of node i in a random size n increasing tree. We will refer to $\beta_{n,i}$ simply as node degree. Further for $n \ge i$ let $r_{n,i}$ denote the number of edges $(k, i) \in E(S)$ with $k > n$. Thus $r_{n,i}$ is just the number of edges coming to node i after time n. We consider the random variable

$$Y_n = \prod_{i,l \in E(S_n)} \mathbb{I}_{\{(i,l) \in E(B_n)\}} \times \begin{cases} \prod\limits_{i \in V(S), i \le n} \beta_{n,i}^{\overline{r_{n,i}}}, & \text{for gports} \\ \prod\limits_{i \in V(S), i \le n} \beta_{n,i}^{\underline{r_{n,i}}}, & \text{for } d\text{-ary increasing trees,} \end{cases} \tag{13}$$

with $\lambda_n = \mathbb{E}(Y_n)$. For large n, one has $r_{n,i} \equiv 0$, $Y_n = \prod_{i,l \in E(S_n)} \mathbb{I}_{\{(i,l) \in E(B_n)\}}$ and $\lambda_n = \mathbb{P}\{S \subset B_n\}$. Theorem 2 will follow directly from our next lemma by means of induction.

Remark 3. For $\alpha = 1$ we have ordinary plane oriented recursive trees, already treated in [7]. Note that for recursive trees things are much easier since $\mathbb{P}\{k_1 <_c i_1, \dots, k_l <_c i_l\} = \prod_{j=1}^{l} (k_j - 1)^{-1}$. Furthermore, recall that for d-ary increasing trees there are at most d edges coming into each node by definition.

Lemma 3. *For $n \ge 0$ the numbers λ_n satisfy the following recurrences.*

- *There is an edge $e = (k, n+1) \in E(S)$ with $k \le n$:*

$$\lambda_{n+1} = \lambda_n \times \begin{cases} \dfrac{\alpha^{\overline{r_{n+1,n+1}}}}{(\alpha+1)n - 1} & \textit{for gports,} \\ \dfrac{d^{\underline{r_{n+1,n+1}}}}{(d-1)n + 1} & \textit{for } d\textit{-ary increasing trees,} \end{cases} \tag{14}$$

- *There is no edge $e = (k, n+1) \in E(S)$ with $k \le n$:*

$$\lambda_{n+1} = \lambda_n \times \begin{cases} \alpha^{\overline{r_{n+1,n+1}}} \Big(1 + \dfrac{C_S(n+1)}{(\alpha+1)n - 1}\Big), & \textit{for gports} \\ d^{\underline{r_{n+1,n+1}}} \Big(1 - \dfrac{C_S(n+1)}{(d-1)n + 1}\Big), & \textit{for } d\textit{-ary increasing trees,} \end{cases} \tag{15}$$

  where $C_S(n+1) = \sum_{k \in V(S), k \le n} r_{n,k}$ denotes the number of edges $e = (k, l) \in E(S)$ with $k \le n$ and $l \ge n+1$.

Proof. We will focus on generalized plane oriented trees and only state some of the analogous formulas for d-ary increasing trees. The outdegree of node $n+1$ in a size $n+1$ tree is always 0, and so we can decompose Y_{n+1} as follows:

$$Y_{n+1} = \beta_{n+1,n+1}^{\overline{r_{n+1,n+1}}} Z_{n+1} = \alpha^{\overline{r_{n+1,n+1}}} Z_{n+1}, \tag{16}$$

with

$$Z_{n+1} = \prod_{i,l \in E(S_{n+1})} \mathbb{I}_{\{(i,l) \in E(B_{n+1})\}} \prod_{i \in V(S), i \le n} \beta_{n+1,i}^{\overline{r_{n+1,i}}}. \tag{17}$$

First, we consider the case that S does not contain an edge $e = (k, n+1)$ with $1 \le k \le n$. Then $S_n = S_{n+1}$ and also $r_{n+1,i} = r_{n,i}$ for each $i \le n$. Hence

$$Z_{n+1} = \prod_{i,l \in E(S_n)} \mathbb{I}_{\{(i,l) \in E(B_n)\}} \prod_{i \in V(S), i \le n} \beta_{n+1,i}^{\overline{r_{n,i}}}, \tag{18}$$

which is exactly the formula for Y_n except for the node degrees $\beta_{n,i}$. Now if node $n+1$ does not attach to any of the vertices of S we have the equality $Z_{n+1} = Y_n$. We consider the random attachment of node

$n+1$. If node $n+1$ attaches to a node $i \in S$ then $\beta_{n+1,i} = \beta_{n,i}+1$ for gports (or $\beta_{n+1,i} = \beta_{n,i}-1$ for d-ary increasing trees) and

$$\beta_{n+1,i}^{\overline{r_{n,i}}} = (\beta_{n,i}+1)^{\overline{r_{n,i}}} = \frac{\beta_{n,i}+r_{n,i}}{\beta_{n,i}} \beta_{n,i}^{\overline{r_{n,i}}} \tag{19}$$

resp.

$$\beta_{n+1,i}^{\underline{r_{n,i}}} = (\beta_{n,i}-1)^{\underline{r_{n,i}}} = \frac{\beta_{n,i}-r_{n,i}}{\beta_{n,i}} \beta_{n,i}^{\underline{r_{n,i}}}, \tag{20}$$

and all other degrees stay the same, so that we get $Z_{n+1} - Y_n = Y_n r_{n,i}/\beta_{n,i}$ resp. $Z_{n+1} - Y_n = -Y_n r_{n,i}/\beta_{n,i}$. In this setting the probability $p_{n+1,i}$ of the event $\{n+1 <_c i\}$, $i \in S$, $i \le n$ is given by

$$p_{n+1,i} = \begin{cases} \frac{X_{n,i}+\alpha}{(\alpha+1)n-1} = \frac{\beta_{n,i}}{(\alpha+1)n-1} & \text{for gports,} \\ \frac{d-X_{n,i}}{(d-1)n-1} = \frac{\beta_{n,i}}{(d-1)n+1} & \text{for } d\text{-ary increasing trees .} \end{cases} \tag{21}$$

Thus the expected difference is given by

$$\mathbb{E}(Z_{n+1} - Y_n | B_n) = \begin{cases} \sum\limits_{i \in V(S), i \le n} \dfrac{p_{n+1,i} Y_n r_{n,i}}{\beta_{n,i}} = \dfrac{Y_n C_S(n+1)}{(\alpha+1)n-1}, & \text{for gports,} \\ -\sum\limits_{i \in V(S), i \le n} \dfrac{p_{n+1,i} Y_n r_{n,i}}{\beta_{n,i}} = -\dfrac{Y_n C_S(n+1)}{(d-1)n-1}, & \text{for } d\text{-ary increasing trees .} \end{cases} \tag{22}$$

Therefore $\mathbb{E}(Z_{n+1}) = \lambda_n(1 + \frac{C_S(n+1)}{(\alpha+1)n-1})$ resp. $\mathbb{E}(Z_{n+1}) = \lambda_n(1 - \frac{C_S(n+1)}{(d-1)n-1})$. Now suppose that there is an edge $e = (n+1, k) \in E(S)$ with $k \le n$. In this case $Y_{n+1} = 0$ unless node $n+1$ is attached to k, which happens with probability $p_{n+1,k}$. Under the assumption $\{n+1 <_c k\}$ we have

$$\prod_{i,l \in E(S_{n+1})} \mathbb{I}_{\{(i,l) \in E(B_{n+1})\}} = \prod_{i,l \in E(S_n)} \mathbb{I}_{\{(i,l) \in E(B_n)\}}, \tag{23}$$

and the node degrees change as follows. We have $\beta_{n+1,i} = \beta_{n,i}$ for $1 \le i \le n, i \ne k$ and $\beta_{n+1,k} = \beta_{n,k}+1$ (or $\beta_{n+1,k} = \beta_{n,k}-1$ for d-ary increasing trees). Furthermore $r_{n+1,i} = r_{n,i}$ for $1 \le i \le n, i \ne k$ and $r_{n+1,k} = r_{n,k} - 1$. Hence,

$$\prod_{i \in V(S), i \le n} \beta_{n+1,i}^{\overline{r_{n+1,i}}} = (\beta_{n,k}+1)^{\overline{r_{n,k}-1}} \prod_{i \in V(S), i \le n, i \ne k} \beta_{n,i}^{\overline{r_{n,i}}} = \frac{1}{\beta_{n,k}} \prod_{i \in V(S), i \le n} \beta_{n,i}^{\overline{r_{n,i}}}, \tag{24}$$

which finally leads to $\mathbb{E}(Z_{n+1}) = \mathbb{E}(Y_n)\beta_{n,k}^{-1} p_{n+1,k} = \frac{\mathbb{E}(Y_n)}{(\alpha+1)n-1}$. □

Theorem 3. *The expectation and the variance of $S_{n,j}$ are given by the following closed formulas.*

$$\mathbb{E}(S_{n,j}) = \sum_{k=j+1}^{n} \mathbb{E}(\mathbb{I}_{\{k <_c k-j\}}) = \sum_{k=j+1}^{n} \frac{\binom{k-j-1+\frac{c_2}{c_1}}{k-j-1}}{\binom{k-1+\frac{c_2}{c_1}}{k-2}},$$
$$\mathbb{V}(S_{n,j}) = \sum_{k=j+1}^{n} \sum_{l=j+1, l \ne k}^{n} \mathbb{P}\{k <_c k-j, l <_c l-j\} + \mathbb{E}(S_{n,j}) - \mathbb{E}(S_{n,j})^2,$$

with $\mathbb{P}\{k <_c k-j, l <_c l-j\}$ as given in Corollary 2. Moreover, the expectation and the variance satisfy the asymptotic expansion

$$\mathbb{E}(S_{n,j}) = (1 + \frac{c_2}{c_1}) \log n + \mathcal{O}(1), \quad \mathbb{V}(S_{n,j}) = (1 + \frac{c_2}{c_1}) \log n + \mathcal{O}(1).$$

Proof. The asymptotic results for the expectation are readily obtained from the exact formula. The exact result for the variance is a consequence of the relation

$$\mathbb{E}(S_{n,j}^2) = \mathbb{E}\Big(\Big(\sum_{k=j+1}^{n} \mathbb{I}_{\{k <_c k-j\}}\Big)^2\Big) = \mathbb{E}(S_{n,j}) + \sum_{k=j+1}^{n} \sum_{l=j+1, l \ne k}^{n} \mathbb{P}\{k <_c k-j, l <_c l-j\}.$$

Moreover, we have

$$\begin{aligned}
&\sum_{k=j+1}^{n}\sum_{l=j+1,l\neq k}^{n}\mathbb{P}\{k<_c k-j, l<_c l-j\}-\mathbb{E}(S_{n,j})^2\\
&=2\sum_{k=j+1}^{n}\sum_{l=k+1}^{k+j-1}\Big(\mathbb{P}\{k<_c k-j, l<_c l-j\}-\mathbb{P}\{k<_c k-j\}\mathbb{P}\{l<_c l-j\}\Big),
\end{aligned}$$

according to Corollary 2, where the factor 2 is due to symmetry between k and l. Furthermore, we have the upper bound

$$\Big(\mathbb{P}\{k<_c k-j, l<_c l-j\}-\mathbb{P}\{k<_c k-j\}\mathbb{P}\{l<_c l-j\}\Big)\leq\frac{2\kappa}{(k-2)^2},$$

in the range $j+1\leq k\leq n$ and $k+1\leq l\leq k+j-1$, with $\kappa=\max\{j+1,d\}$. Hence,

$$\sum_{k=j+1}^{n}\sum_{l=j+1,l\neq k}^{n}\mathbb{P}\{k<_c k-j, l<_c l-j\}-\mathbb{E}(S_{n,j})^2\leq 2j\kappa\sum_{k=j+1}^{n}\frac{1}{(k-2)^2}\leq j\kappa\frac{\pi^2}{3},$$

which proves the stated result. □

4.2 Central limit theorem

By Corollary 1 we already know that the indicator variables are j-independent. We will use a simplified version of a result of Barbour et al. [3], see also [4].

Theorem 4 ([3], [4])**.** *Suppose that $(Y_n)_{n\in\mathbb{N}}$ is a sequence of random variables with $\mathbb{E}(Y_n)=0$ and bounded third moment $\mathbb{E}(|Y_n|^3)<\infty$, that are j-independent. Set $Z_n=Y_1+\cdots+Y_n$ and $\sigma_n^2:=\mathbb{V}(Z_n)$. If $\lim_{n\to\infty}\frac{1}{\sigma_n^3}\sum_{l=1}^{n}\mathbb{E}(|Y_l|^3)=0$, then Z_n satisfies a central limit theorem, $\frac{Z_n}{\sqrt{\mathbb{V}(Z_n)}}\to\mathcal{N}(0,1)$.*

Now we are ready to state the central limit theorem for $S_{n,j}$.

Theorem 5. *For arbitrary but fixed $j\in\mathbb{N}$ and n tending to infinity, the suitably shifted and normalized random variable $S_{n,j}=\sum_{e\in E_n}\mathbb{I}_{\{w(e)=j\}}$ is asymptotically normal distributed,*

$$\frac{S_{n,j}-\mathbb{E}(S_{n,j})}{\sqrt{\mathbb{V}(S_{n,j})}}\xrightarrow{d}\mathcal{N}(0,1).$$

Proof. We want to apply Theorem 4 to the centered random variable $\tilde{S}_{n,j}=S_{n,j}-\mathbb{E}(S_{n,j})=\sum_{l=j+1}^{n}Y_l$, with

$$Y_l=\mathbb{I}_{\{l<_c l-j\}}-\mathbb{P}\{l<_c l-j\}.$$

By construction, $\mathbb{E}(Y_l)=0$, and by Corollary 1 the centered indicator variables Y_l are j-independent. Let Ω denote the sample space of all trees of size n and $\mathbb{P}$ the probability measure on Ω. We have $\Omega=\Omega_1\cup\Omega_1^c$, where $\Omega_1=\Omega_1(k)=\{\omega\in\Omega:\mathbb{I}_{\{k<_c k-j\}}(\omega)=1\}$, and furthermore

$$\begin{aligned}
\mathbb{E}(|Y_k|^3)&=\int_\Omega|\mathbb{I}_{\{k<_c k-j\}}-\mathbb{P}\{k<_c k-j\}|^3(\omega)d\mathbb{P}\\
&=\int_{\Omega_1}(1-\mathbb{P}\{k<_c k-j\})^3d\mathbb{P}+\int_{\Omega_1^c}\mathbb{P}\{k<_c k-j\}^3d\mathbb{P}\\
&=\mathbb{P}\{k<_c k-j\}(1-\mathbb{P}\{k<_c k-j\})^3+(1-\mathbb{P}\{k<_c k-j\})\mathbb{P}\{k<_c k-j\}^3.
\end{aligned}$$

Therefore we get the estimates

$$\begin{aligned}
\sum_{k=j+1}^{n}\mathbb{E}(|Y_l|^3)&=\sum_{k=j+1}^{n}(\mathbb{P}\{k<_c k-j\}(1-\mathbb{P}\{k<_c k-j\})^3+(1-\mathbb{P}\{k<_c k-j\})\mathbb{P}\{k<_c k-j\}^3)\\
&\leq\sum_{k=j+1}^{n}(\mathbb{P}\{k<_c k-j\}+\mathbb{P}\{k<_c k-j\}^3)=\Big(1+\frac{c_2}{c_1}\Big)\log n+\mathcal{O}(1).
\end{aligned}$$

Since $\sigma_n^2=\mathbb{V}(\tilde{S}_{n,j})=\mathbb{V}(S_{n,j})=\big(1+\frac{c_2}{c_1}\big)\log n+\mathcal{O}(1)$, the conditions of Theorem 4 are satisfied, which implies the asymptotic normality of $\tilde{S}_{n,j}/\sqrt{\mathbb{V}(S_{n,j})}$. □

Remark 4. Note that with a bit more effort, one can also obtain the speed of convergence with respect to the metric d_1 [4]: one has $d_1\Big(\mathcal{L}(\frac{\tilde{S}_{n,j}}{\sqrt{\mathbb{V}(S_{n,j})}}),\mathcal{N}(0,1)\Big) \le \frac{C_j}{\sqrt{\log n}}$, where C_j is a constant depending on j. For two probability measures P and Q their d_1-distance is defined as $d_1(P,Q) := \sup_{||h||=1} |\mathbb{E}(h(X)) - E(h(Y))|$, where X and Y are random variables with distribution P and Q, respectively. We refer the reader to [3] and [4].

5 Representation of increasing trees via edge weight tables

It was shown in [13] that the family of recursive trees can be represented by a so-called edge-weight table, corresponding to the inversion table of permutations. Moreover, it was asked for a corresponding notion for other tree families. Here we will introduce edge-weight tables for plane oriented recursive trees and d-ary increasing trees. Such sequences may be important regarding the automatic generating of all trees of a given family.

Let $\mathcal{C}_n$ denote the family of sequences $\sigma = a_1b_1a_2b_2\dots a_nb_n$ of length $2n$, $n\ge 1$, such that $1\le a_i\le i$ and $1\le b_i\le |\{j|j+1-a_j = i+1-a_i, 1\le j\le i-1\}|+1$. Moreover, let $\mathcal{D}_n=\mathcal{D}_n(d)$ denote the family of sequences $a_1b_1a_2b_2\dots a_nb_n$ of length $2n$, $n\ge 1$, such that $1\le a_i\le i$, $1\le b_i\le d$, and if $b_i=b_j$, then $i+1-a_i\neq j+1-a_j$ for $1\le i<j\le n$.

Proposition 1. *The family of plane oriented increasing trees of size $n+1$ is in bijection with the family $\mathcal{C}_n$. Furthermore, the family of d-ary increasing trees of size $n+1$ is in bijection with the family $\mathcal{D}_n$.*

Proof. We use a recursive construction. For a given size $n+1$ plane oriented recursive tree, we note the edge-weight induced by the node labeled $n+1$ and its position, going from left to right, which gives a_nb_n. Now we remove node $n+1$ and proceed recursively. Conversely, for a given edge-weight table $\sigma\in\mathcal{C}_n$ we recursively construct the size $n+1$ tree by attaching node i to the node labeled $i+1-a_i$ at position b_i, $1\le i\le n$.

For d-ary increasing trees we proceed analogously, denoting the edge-weight induced by the node labeled i and its position, $i=n+1,\dots,2$. □

6 A combinatorial approach: recurrences and a system of differential equations

At first sight the most natural approach for the analysis of $S_{n,j}$ seems to be the usage of the combinatorial description of increasing trees according to (3), which was often useful for similar problems. Unfortunately, this approach is not easily applicable since the subtrees are relabeled in the description, whereas we cannot simply drop the labeling of the subtrees without further considerations.

In order to analyze $S_{n,j}$ combinatorially for $j>1$, we have to proceed in a different way. The main idea is to partition the T_n different size-n increasing trees into classes according to the out-degrees of the nodes $n-j+1, n-j+2,\dots,n-1$, which are relevant for $S_{n,j}$. Then one can set up suitable recurrences for the arising tree classes, always keeping track of the behavior of all relevant outdegrees.

First we need some notation. Let $\mathbf{g}_{j-1}=(g_1,g_2,\dots,g_{j-1})$ denote a vector of size $j-1$, which will encode the outdegrees of the nodes $n-1,n-2,\dots,n-j+1$, and $G_k=\sum_{i=1}^k g_i$ the sum of the first k entries, $1\le k\le j-1$. Furthermore, W_{j-1} denotes the set of vectors $\mathbf{g}_{j-1}$ satisfying $g_k\ge 0$ and

$$0\le G_k=\sum_{i=1}^k g_i\le k$$

for $1\le k\le j-1$, which is the natural restriction for increasing trees. We denote by $T_n^{[\mathbf{g}_{j-1}]} = T_n^{[g_1,g_2,\dots,g_{j-1}]}$ the number of increasing trees of size n, where the distribution of the outdegrees of the nodes $(n-1,n-2,\dots,n-j+1)$ is given by the vector $\mathbf{g}_{j-1}\in W_{j-1}$, i.e. node $n-k$ has outdegree g_k for $1\le k\le j-1$. The total number is obtained from the $T_n^{[\mathbf{g}_{j-1}]}$ by summation over all possible degree sequences of $n-1,\dots,n+j-1$,

$$T_n=\sum_{\mathbf{g}_{j-1}\in W_{j-1}} T_n^{[\mathbf{g}_{j-1}]}.$$

Our first result is an explicit formula for $T_n^{[\mathbf{g}_{j-1}]}$, the number of all size-n trees with degrees prescribed by a sequence $\mathbf{g}_{j-1}$. In the following, we will state all results first for gports, then for d-ary trees. The proofs are only given for gports, the situation for d-ary trees being completely analogous.

Theorem 6. *The number $T_n^{[\mathbf{g}_{j-1}]}$ of all size-n ($n \geq j$) increasing trees, where the outdegrees of the nodes $n-1, \dots, n-j+1$ are prescribed by $\mathbf{g}_{j-1} \in W_{j-1}$, is given as follows: for generalized plane oriented recursive trees,*

$$T_n^{[\mathbf{g}_{j-1}]} = T_{n-j+1} \cdot \big((\alpha+1)(n-j)\big)^{\overline{j-1-G_{j-1}}} \cdot \prod_{i=1}^{j-1} \binom{i - G_{i-1}}{g_i} \alpha^{\overline{g_i}}.$$

For d-ary increasing trees,

$$T_n^{[\mathbf{g}_{j-1}]} = T_{n-j+1} \cdot \big((d-1)(n-j)\big)^{\underline{j-1-G_{j-1}}} \cdot \prod_{i=1}^{j-1} \binom{i - G_{i-1}}{g_i} d^{\underline{g_i}}.$$

Proof. Given the tree induced by the first $n-j+1$ nodes (there are T_{n-j+1} possibilities for this tree), we can choose the g_i children of node $n-i$ ($i = 1, \dots, j-1$) out of a set of $i - G_{i-1}$ nodes with larger number, which can be attached in $\alpha^{\overline{g_i}}$ different ways. Finally, we have to attach the remaining $j-1-G_{j-1}$ nodes from the set $\{n, n-1, \dots, n-j+2\}$ to nodes with smaller labels, which gives rise to the second factor in our formula. ☐

Since we will use the (refined) quantities $T_n^{[\mathbf{g}_{j-1}]}$ to describe a system of differential equations, we are interested in the cardinality of the system depending on j.

Proposition 2. *The cardinality of W_{j-1} is given by the j-th Catalan number $C_j = \binom{2j}{j}/(j+1)$.*

Proof (sketch). Observe that we can interpret the elements $\mathbf{g}_{j-1} = (g_1, g_2, \dots, g_{j-1})$ of W_{j-1} as lattice paths with steps $(1,0), (1,1), \dots, (1, j-1)$, starting at $(0,0)$, which never exceed the diagonal $y = x$. ☐

Example 3. As an example, let us consider plane oriented recursive trees ($\alpha = 1$) with prescribed outdegrees for the nodes $n-1$ and $n-2$. We have $|W_2| = C_3 = 5$ and by Theorem 6

$$T_n^{[0,0]} = 2(n-3)(2(n-3)+1)(2n-7)!!, \quad T_n^{[0,1]} = 4(n-3)(2n-7)!!,$$
$$T_n^{[0,2]} = 2(2n-7)!!, \quad T_n^{[1,0]} = 2(n-3)(2n-7)!!, \quad T_n^{[1,1]} = (2n-7)!!.$$

Now let $T_{n,j,m}^{[\mathbf{g}_{j-1}]}$ denote the number of size-n increasing trees with m edge weights of size j and outdegrees specified by $\mathbf{g}_{j-1} \in W_{j-1}$ as before. We have the relation

$$T_n \mathbb{P}\{S_{n,j} = m\} = \sum_{\mathbf{g}_{j-1} \in W_{j-1}} T_{n,m,j}^{[\mathbf{g}_{j-1}]}.$$

Furthermore let $W_{j-1}(\mathbf{g}_{j-1}) \subset W_{j-1}$ denote the set of vectors $\mathbf{l}_{j-1} = (l_1, \dots, l_{j-1}) \in W_{j-1}$ such that $\mathbf{l}_{j-1}$ has the form $\mathbf{l}_{j-1} = (g_2, g_3, \dots, g_{j-1}, i) - \mathbf{e}_k$, with $1 \leq k \leq j-2$ and $0 \leq i \leq j - G_{j-1}$, where $\mathbf{e}_k$ denotes a unit vector. In other words, $W_{j-1}(\mathbf{g}_{j-1)}$ consists of all vectors $\mathbf{l}_{j-1}$ in W_{j-1} with $l_h = g_{h+1}$ for $h \in \{1, \dots, j-2\} \setminus \{k\}$, and $l_k = g_{k+1} - 1$, where $1 \leq k \leq j-2$. We obtain the following recurrences for $T_{n,j,m}^{[\mathbf{g}_{j-1}]}$ by distinguishing two cases for g_1.

Proposition 3. *For $n \geq j+1$ and $m \geq 0$ the quantities $T_{n,j,m}^{[\mathbf{g}_{j-1}]}$, with $\mathbf{g}_{j-1} \in W_{j-1}$, satisfy the following system of recurrence relations. For $g_1 = 1$,*

$$T_{n,j,m}^{[\mathbf{g}_{j-1}]} = \sum_{i=0}^{j-G_{j-1}} A \cdot T_{n-1,j,m}^{[g_2,\dots,g_{j-1},i]}, \quad A := \begin{cases} \alpha, \\ d. \end{cases}$$

For $g_1 = 0$,

$$T_{n,j,m}^{[\mathbf{g}_{j-1}]} = \sum_{\mathbf{l}_{j-1} \in W_j(\mathbf{g}_j)} B \cdot T_{n-1,j,m}^{[\mathbf{l}_{j-1}]} + \sum_{i=0}^{j-1-G_{j-1}} C \cdot T_{n-1,j,m}^{[g_2,\dots,g_{j-1},i]} + \sum_{i=0}^{j-1-G_{j-1}} D \cdot T_{n-1,j,m-1}^{[g_2,\dots,g_{j-1},i]},$$

with $B = B(\mathbf{l}_{j-1})$, $C = C(i)$ and $D = D(i)$ given by

$$B := \begin{cases} \sum_{h=1}^{j-2}(g_{h+1} - l_h)(l_h + \alpha), \\ \sum_{h=1}^{j-2}(g_{h+1} - l_h)(d - l_h), \end{cases} \quad C := \begin{cases} (\alpha+1)(n-1) - 1 - i - G_{j-1} - j\alpha, \\ (d-1)(n-1) - 1 + i + G_{j-1} - jd, \end{cases} \quad D := \begin{cases} i + \alpha, \\ d - i, \end{cases}$$

and initial values $T^{[\mathbf{g}_{j-1}]}_{j,j,0} = T^{[\mathbf{g}_{j-1}]}_{j}$ given by Theorem 6.

Proof. In the case $g_1 = 1$, the newly inserted node labeled n must be attached to node $n-1$. Hence we have to consider trees with $n-1$ nodes and m edge weights of size j, where the outdegrees of nodes $n-2, \dots, n-j+1$ are given by $g_2, \dots, g_{j-1}$ and the outdegree i of node $n-j$ is between zero and

$$j - 1 - \sum_{k=2}^{j-1} g_k = j - 1 - (G_{j-1} - g_1) = j - G_{j-1}.$$

The other case $g_1 = 0$, where node n is not attached to node $n-1$, splits into three possible cases: node n is attached to one of the nodes $n-2, \dots, n-j+1$, or node n is attached to node $n-j$, increasing the number of edge weights of size j by one, or node n is not attached to any of the nodes $n-2, \dots, n-j$. First we consider the case that node n is attached to $n-k$, with $2 \le k \le j-1$. Then there are $g_{k+1} - 1 + \alpha = l_k + \alpha$ possible positions to attach node n to any $n-k$. Note that under the assumption $\mathbf{l}_{j-1} \in W_j(\mathbf{g}_j)$ with $l_k = g_{k+1} - 1$, we have $\sum_{h=1}^{j-2}(g_{h+1} - l_h)(l_h + \alpha) = l_k + \alpha$, as required.

Next we look at the case that node n is attached to node $n-j$. Assuming that node $n-j$ has outdegree i, we have $i + \alpha$ different places to attach node n to node $n-j$.

Finally we consider the case that node n is not attached to any of the nodes node $n-k$, $2 \le k \le j$. Hence, assuming again that node $n-j$ has outdegree i, $0 \le i \le j - 1 - G_{j-1}$, we have

$$(\alpha+1)(n-1) - 1 - (i+\alpha) - \sum_{k=2}^{j-1}(g_k + \alpha) - \alpha = (\alpha+1)(n-1) - 1 - i - G_{j-1} - j\alpha$$

different places to attach node n to the tree of size $n-1$, which finishes the proof of our formula. □

Note that for $m = 0$ one has to skip the terms including $T^{[g_2,\dots,g_{j-1},i]}_{n-1,j,-1}$. Now we introduce the bivariate generating functions

$$F^{[\mathbf{g}_{j-1}]}(z, v) = \sum_{n \ge j+1} \sum_{m \ge 0} T^{[\mathbf{g}_{j-1}]}_{n,m,j} \frac{z^{n-j}}{(n-j)!} v^m, \tag{25}$$

for $\mathbf{g}_{j-1} \in W_{j-1}$. By multiplying our recurrence relations by $v^m z^{n-j-1}/(n-j-1)!$ and summing over $n \ge j+1, m \ge 0$ the recurrences above can by translated into a system of linear differential equations.

6.1 The case $j = 2$

Let us now consider the case $j = 2$ for gports as an illustration. The initial values are given by $T^{[1]}_{2,0,2} = T_2$ and $T^{[i]}_{2,m,2} = 0$ for all other i and m. For the sake of simplicity we will drop the dependence on $j = 2$. By Proposition 3 we get the recurrences

$$\begin{aligned} T^{[0]}_{n,m} &= ((\alpha+1)(n-3)+1)T^{[0]}_{n-1,m} + \alpha T^{[0]}_{n-1,m-1} + (\alpha+1)(n-3)T^{[1]}_{n-1,m} + (\alpha+1)T^{[1]}_{n-1,m-1}, \\ T^{[1]}_{n,m} &= \alpha T^{[0]}_{n-1,m} + \alpha T^{[1]}_{n-1,m}. \end{aligned} \tag{26}$$

Following (25), we set up the generating functions

$$F^{[0]}(z, v) = \sum_{n \ge 3} \sum_{m \ge 0} T^{[0]}_{n,m} \frac{z^{n-2}}{(n-2)!} v^m, \quad F^{[1]}(z, v) = \sum_{n \ge 3} \sum_{m \ge 0} T^{[1]}_{n,m} \frac{z^{n-2}}{(n-2)!} v^m.$$

Multiplication by $v^m z^{n-3}/(n-3)!$ and summation over $n \ge 3$ and $m \ge 0$ leads to the following system of linear differential equations.

$$\begin{aligned} \frac{\partial}{\partial z} F^{[0]}(z, v) &= (\alpha v + 1)F^{[0]}(z, v) + (\alpha+1)vF^{[1]}(z, v) + (\alpha+1)vT_2 \\ &\quad + (\alpha+1)z\frac{\partial}{\partial z}F^{[0]}(z, v) + (\alpha+1)z\frac{\partial}{\partial z}F^{[1]}(z, v), \\ \frac{\partial}{\partial z} F^{[1]}(z, v) &= \alpha F^{[0]}(z, v) + \alpha F^{[1]}(z, v) + \alpha T_2. \end{aligned} \tag{27}$$

Unfortunately, this system of differential equations is not explicitly solvable. However, one can easily determine the first few coefficients from it; in the case of ordinary plane oriented recursive trees ($\alpha = 1$), one obtains

$$F^{[0]}(z,v) = 2vz + (1+4v+v^2)z^2 + \frac{7+26v+11v^2+v^3}{3}z^3 + \frac{58+222v+119v^2+20v^3+v^4}{12} + \dots$$

and

$$F^{[1]}(z,v) = z + \frac{1+2v}{2}z^2 + \frac{3+10v+2v^2}{6}z^3 + \frac{17+62v+24v^2+2v^3}{24}z^4 + \dots$$

and altogether

$$\begin{aligned} F^{[0]}(z,v) + F^{[1]}(z,v) &= (1+2v)z + \frac{3+10v+2v^2}{2}z^2 \\ &\quad + \frac{17+62v+24v^2+2v^3}{6}z^3 + \frac{133+506v+262v^2+42v^3+2v^4}{24}z^4 + \dots \end{aligned}$$

6.2 Maximal edge weight

To show the usefulness of our approach, we consider a related problem: let $p_{n,m} = \mathbb{P}\{M_n \le m\}$ denote the probability that the maximal edge weight $M_n = \max_{e\in E_n} w_e$ in a size n random increasing tree is less or equal m. In order to study this probability, we use two different approaches. For large m (i.e. $m = n-k$ with fixed k), one can apply the principle of inclusion and exclusion to get an expression for the probabilities $p_{n,n-k}$ as follows:

Theorem 7. *The probability that the maximal edge weight M_n is less or equal $n-k$, with $2 \le k \le n-1$, is given by*

$$\mathbb{P}\{M_n \le n-k\} = 1 + \sum_{l=1}^{k-1}(-1)^l \sum_{\substack{n+2-k\le i_1<\dots<i_l\le n \\ 1\le j_{i_h}\le i_h-(n+1-k)}} \mathbb{P}\{i_1 <_c j_1, \dots, i_l <_c j_l\},$$

with $\mathbb{P}\{i_1 <_c j_1, \dots, i_l <_c j_l\}$ as given by Theorem 2.

Example 4. By application of Theorem 7 we obtain e. g. for $k=2$

$$\mathbb{P}\{M_n \le n-2\} = 1 - \frac{1}{\binom{n-1+\frac{c_2}{c_1}}{n-2}}.$$

For small m we have to proceed differently. Let $p_{n,m}^{[\mathbf{g}_{m-1}]}$ denote the probability that the maximal edge weight M_n is less or equal m and that the outdegrees of nodes $n-1, \dots, n-m+1$ are given by $g_1, \dots, g_{m-1}$.

Proposition 4. *For $n \ge 2$, the probabilities $p_{n,m}^{[\mathbf{g}_{m-1}]}$, with $\mathbf{g}_{m-1} \in W_{m-1}$, satisfy the following system of recurrence relations. For $g_1 = 1$ we have*

$$p_{n,m}^{[\mathbf{g}_{m-1}]} = \sum_{i=0}^{m-G_{m-1}} A \cdot p_{n-1,m}^{[g_2,\dots,g_{m-1},i]}, \quad A := \begin{cases} \frac{\alpha}{(\alpha+1)(n-1)-1}, \\ \frac{d}{(d-1)(n-1)+1}. \end{cases} \tag{28}$$

For $g_1 = 0$ we have

$$\begin{aligned} p_{n,m}^{[\mathbf{g}_{m-1}]} &= \sum_{\mathbf{l}_{m-1}\in W_m(\mathbf{g}_m)} B \cdot p_{n-1,m}^{[\mathbf{l}_{m-1}]} + \sum_{i=0}^{m-1-G_{m-1}} C \cdot p_{n-1,m}^{[g_2,\dots,g_{m-1},i]}, \\ B = B(\mathbf{l}_{m-1}) &:= \begin{cases} \sum_{h=1}^{m-2}(g_{h+1}-l_h)\frac{l_h+\alpha}{(\alpha+1)(n-1)-1}, \\ \sum_{h=1}^{m-2}(g_{h+1}-l_h)\frac{d-l_h}{(d-1)(n-1)+1}, \end{cases} \quad C = C(i) := \begin{cases} \frac{i+\alpha}{(\alpha+1)(n-1)-1}, \\ \frac{d-i}{(d-1)(n-1)+1}. \end{cases} \end{aligned} \tag{29}$$

with initial values $p_{2,m}^{[1]} = 1$ and $p_{2,m}^{[0]} = 0$ for all m.

By using our earlier results concerning $S_{n,1}$ we immediately obtain

$$p_{n,1} = \mathbb{P}\{M_n \leq 1\} = \mathbb{P}\{S_{n,1} = n-1\} = \begin{cases} \frac{(\frac{\alpha}{\alpha+1})^{n-1}}{(n-1)!\binom{n-1-\frac{1}{\alpha+1}}{n-1}}, \\ \frac{(\frac{d}{d-1})^{n-1}}{(n-1)!\binom{n-1+\frac{1}{d-1}}{n-1}}. \end{cases}$$

Finally, we compute $p_{n,2}$ by means of Proposition 4. Unlike the differential equations obtained in the previous section, the differential equations for the generating functions

$$P_m^{[\mathbf{g}_{m-1}]}(z) := \sum_{n \geq 2} p_{n,m}^{[\mathbf{g}_{m-1}]} z^{n-1-1/(\alpha+1)}$$

(in the case of d-ary increasing tree, the exponent has to be modified to $n-1+\frac{1}{d-1}$; this somewhat artificial choice results in simpler differential equations) will be linear with constant coefficients and therefore explicitly solvable. We illustrate this in the case $m=2$, where we get the recurrence relations

$$\begin{aligned} ((\alpha+1)(n-1)-1)p_{n,2}^{[0]} &= \alpha p_{n-1,2}^{[0]} + (\alpha+1)p_{n-1,2}^{[1]}, \\ ((\alpha+1)(n-1)-1)p_{n,2}^{[1]} &= \alpha p_{n-1,2}^{[0]} + \alpha p_{n-1,2}^{[1]} \end{aligned}$$

for gports and

$$\begin{aligned} ((d-1)(n-1)+1)p_{n,2}^{[0]} &= d p_{n-1,2}^{[0]} + (d-1)p_{n-1,2}^{[1]}, \\ ((d-1)(n-1)+1)p_{n,2}^{[1]} &= d p_{n-1,2}^{[0]} + d p_{n-1,2}^{[1]} \end{aligned}$$

for d-ary trees. Let us consider the latter case in more detail: introduction of the generating functions

$$P_2^{[i]}(z) := \sum_{n \geq 2} p_{n,2}^{[i]} z^{n-1+1/(d-1)}$$

yields the differential equations

$$\begin{aligned} (d-1)\frac{d}{dz}P_2^{[0]}(z) &= dP_2^{[0]}(z) + (d-1)P_2^{[1]}(z), \\ (d-1)\frac{d}{dz}P_2^{[1]}(z) &= dP_2^{[0]}(z) + dP_2^{[1]}(z) + dz^{1/(d-1)}. \end{aligned}$$

A particularly nice special case is $d=2$, where one gets

$$\begin{aligned} P_2^{[0]}(z) &= 2 + z + \frac{3\sqrt{2}-4}{4}\, e^{(2+\sqrt{2})z} - \frac{3\sqrt{2}+4}{4}\, e^{(2-\sqrt{2})z}, \\ P_2^{[1]}(z) &= -3 - 2z + \frac{3-2\sqrt{2}}{2}\, e^{(2+\sqrt{2})z} - \frac{3+2\sqrt{2}}{2}\, e^{(2-\sqrt{2})z}. \end{aligned}$$

Putting these together, one obtains the simple explicit formula

$$p_{n,2} = \frac{1}{2n!}\left((2+\sqrt{2})^{n-1} + (2-\sqrt{2})^{n-1}\right).$$

Generally, the asymptotics of the probability depend on the largest eigenvalue of a matrix of dimension $|W_{m-1}|$; by Proposition 2, this is equal to $\frac{1}{m+1}\binom{2m}{m}$. In the case $m=2$, the largest eigenvalue is

$$\frac{\alpha+\sqrt{\alpha^2+\alpha}}{\alpha+1} \quad \text{resp.} \quad \frac{d+\sqrt{d^2-d}}{d-1},$$

but it seems that there is no nice explicit formula for the general case.

References

[1] R. Albert, H. Jeong, A. -L. Barabási, The diameter of the world wide web. *Nature* 401, 130-131, 1999.

[2] A. -L. Barabási and R. Albert, Emergence of scaling in random networks. *Science* 286, 509-512, 1999.

[3] A. D. Barbour and G. K. Eagleson, Multiple comparisons and sums of dissociated random variables, *Adv. Appl. Probab.* 17, 147–162, 1985.

[4] A. D. Barbour, M. Karoński and A. Ruciński, A Central Limit Theorem for Decomposable Random Variables with Applications to Random Graphs, *Journal of Combinatorial Theory B*, 47, 125–145, 1989.

[5] F. Bergeron, P. Flajolet and B. Salvy, Varieties of Increasing Trees, *Lecture Notes in Computer Science* 581, 24–48, 1992.

[6] F. Bergeron, G. Labelle, and P. Leroux, *Combinatorial species and tree-like structures*, volume 67 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 1998.

[7] B. Bollobás and O. Riordan, Mathematical results on scale-free random graphs, *Handbook of Graphs and Networks*, eds. S. Bornholdt & H.G. Schuster, Wiley-VCH, Weinheim (2003), 1-34.

[8] B. Bollobás and O. Riordan, Shortest paths and load scaling in scale-free trees, *Physical Review E* 69, 2004.

[9] R. Dobrow and R. Smythe, Poisson approximations for functionals of random trees. *Random Structures & Algorithms* 9, 79-92, 1996.

[10] H.-K. Hwang, On Convergence Rates in the Central Limit Theorems for Combinatorial Structures, *European Journal of Combinatorics* 19, 329–343, 1998.

[11] C. A. R. Hoare, Quicksort. *Computer Journal*, 5:1015, 1962.

[12] S. Janson, Asymptotic degree distribution in random recursive trees, *Random Structures Algorithms* 26, 69–83, 2005.

[13] M. Kuba and A. Panholzer, On edge-weighted recursive trees and inversions in random permutations, *Discrete Mathematics*, 308, 529–540, 2008.

[14] A. Panholzer and H. Prodinger, The level of nodes in increasing trees revisited, *Random Structures & Algorithms*, 31, 203–226, 2007.

[15] J. Vitter and P. Flajolet, *Average Case Analysis of Algorithms and Data Structures*, in *Handbook of Theoretical Computer Science*, 431–524, Elsevier, Amsterdam, 1990.