

ARTICLE

# On bucket increasing trees, clustered increasing trees and increasing diamonds

Markus Kuba<sup>1\*</sup> and Alois Panholzer<sup>2</sup>

<sup>1</sup>Department Applied Mathematics and Physics, University of Applied Sciences - Technikum Wien, Höchstädtplatz 5, 1200 Wien, Austria and <sup>2</sup>Institut für Diskrete Mathematik und Geometrie, Technische Universität Wien, Wiedner Hauptstr. 8-10/104, 1040 Wien, Austria

\*Corresponding author. Email: [kuba@technikum-wien.at](mailto:kuba@technikum-wien.at)

(Received 20 March 2020; revised 5 January 2021; accepted 15 November 2020; first published online 13 October 2021)

## Abstract

In this work we analyse bucket increasing tree families. We introduce two simple stochastic growth processes, generating random bucket increasing trees of size  $n$ , complementing the earlier result of Mahmoud and Smythe (1995, *Theoret. Comput. Sci.* **144** 221–249.) for bucket recursive trees. On the combinatorial side, we define multilabelled generalisations of the tree families  $d$ -ary increasing trees and generalised plane-oriented recursive trees. Additionally, we introduce a clustering process for ordinary increasing trees and relate it to bucket increasing trees. We discuss in detail the bucket size two and present a bijection between such bucket increasing tree families and certain families of graphs called increasing diamonds, providing an explanation for phenomena observed by Bodini et al. (2016, *Lect. Notes Comput. Sci.* **9644** 207–219.). Concerning structural properties of bucket increasing trees, we analyse the tree parameter  $K_n$ . It counts the initial bucket size of the node containing label  $n$  in a tree of size  $n$  and is closely related to the distribution of node types. Additionally, we analyse the parameters descendants of label  $j$  and degree of the bucket containing label  $j$ , providing distributional decompositions, complementing and extending earlier results (Kuba and Panholzer (2010), *Theoret. Comput. Sci.* **411**(34–36) 3255–3273.).

**Keywords:** Bucket increasing trees; clustered trees; stochastic growth processes; descendants; node-degrees; limiting distributions

**2020 MSC Codes:** Primary: 05C05, Secondary: 60F05

## 1. Introduction

*Increasing trees* or *increasingly labelled trees* are rooted labelled trees. The nodes of a tree  $T$  of size  $|T| = n$  are labelled with distinct integers from a label set  $\mathcal{M}$  of size  $|\mathcal{M}| = n$ . Here, the size  $|T|$  of a tree denotes the number of vertices of  $T$  (and thus coincides with the number of labels). One chooses as label set the first  $n$  positive integers, i.e.,  $\mathcal{M} = [n] := \{1, 2, \dots, n\}$ , in such a way that the label of any node in the tree is smaller than the labels of its children. As a consequence, the labels of each path from the root to an arbitrary node in the tree are forming an increasing sequence, which explains the name of such a labelling. Various increasing tree models turned out to be appropriate in order to describe the growth behaviour of quantities in a number of applications and occurred in the probabilistic literature. E.g., they are used to describe the spread of epidemics, to model pyramid schemes, and as a simplified growth model of the world wide web. See Mahmoud and Smythe [19] for a survey collecting results about recursive trees, a subfamily of increasing trees, prior 1995. For recent results about increasing trees we refer to the books of Drmota [5], Flajolet and Sedgewick [6] and references therein.

In above definition of increasing trees each node in the tree gets exactly one label. Here in this work we discuss an extension of increasing trees to *bucket increasing trees*. These are *multilabelled increasing tree families*. All the nodes  $v$  in the tree  $T$  are *buckets* with a maximal capacity of  $b$  labels. Here and throughout this work the integer  $b \in \mathbb{N}$  denotes the maximal capacity or bucket size. The integer  $c = c(v)$  denotes the current capacity or load of a node  $v \in T$ , with  $1 \leq c \leq b$ . We always call a node  $v$  with capacity  $c(v) = b$  *saturated* and otherwise *unsaturated*. We impose the additional restriction that all internal nodes (i.e., non-leaves) in the tree must be saturated. In contrast, the leaves might be either saturated or unsaturated. The size of a tree  $|T|$  is here and throughout this work measured by the sum of all node capacities  $c(v)$ :  $|T| = \sum_{v \in T} c(v)$ ; equivalently, it is given by the number of labels in the tree  $T$ . We note in passing that tree families where all leaves have to be saturated have been considered and discussed recently in [14, 15] in connection with so-called hook-length formulas. Moreover, closely related combinatorial objects named increasing diamonds have been studied in [4], as discussed later.

A specific family of bucket increasing trees called *bucket recursive trees* has been introduced by Mahmoud and Smythe [18]. The probabilistic description of random bucket recursive trees is given by a *stochastic growth rule*: a tree grows by progressive attraction of increasing integer labels. Such growth rules are part of a general (preferential) attachment rule with fertility and ageing, see Berger et al. [1, 2]. A general combinatorial model for bucket recursive trees has been introduced in [13]. Note that also different models of bucket recursive trees appeared, see for example [9], where the capacities of internal nodes might also vary. Technically, this work and related articles are published in a journal, though they do not seem to be written to professional standards.<sup>1</sup>

The aims of this paper are the following: first, we discuss the general combinatorial framework for bucket increasing trees and the corresponding random tree model. We introduce and analyse two new families of bucket increasing trees. For the two new families of  $(b, d)$ -ary increasing trees and  $(b, \alpha)$ -plane-oriented recursive trees we present both, a stochastic growth rule and a combinatorial description. Then, we prove that both descriptions are equivalent. As a byproduct of our study we complement previous results [13, 18] for bucket recursive trees. Furthermore we present a clustering map for ordinary increasing trees, which maps them to bucket increasing trees. In the special case of bucket size  $b = 2$  we state a bijection between bucket increasing trees and so-called increasing diamonds studied in [4], providing an explanation for phenomena observed in [4]; see Corollary 19 for the precise statement.

We obtain the combinatorial description using a generalisation of a class of weighted tree families, so called simple families of increasing trees, to a class of bucket trees, which we call families of bucket increasing trees. All considered tree families will then be special instances of a bucket increasing tree family. As in the previous analysis of bucket recursive trees the gain of the combinatorial description provided here is that the natural combinatorial decomposition of a bucket increasing tree into a root bucket and its subtrees will lead to a recursive description of several important tree parameters in random bucket recursive trees. This combinatorial decomposition can be translated into certain differential equations for suitably defined generating functions. On the other hand the stochastic growth rules will allow us to present decompositions of random variables of interest.

The combination of both methods – a combinatorial approach leading to exact expressions and the stochastic growth rule leading to decompositions – turns out to be a particularly useful tool for a variety of parameters. First we study the random variables (r.v. for short)  $K_n$  counting the size of the bucket containing label  $n$ . Explicit results for the probability mass function of  $K_n$  are obtained by the combinatorial approach. The explicit expression for  $\mathbb{P}\{K_n = m\}$  readily leads to a

<sup>1</sup>We do not see convincing evidence of a suitable level of peer review: the suggested combinatorial model turns out to be not consistent with the proposed stochastic growth rule; moreover, unfortunately the combinatorial treatment of the enumeration problem as well as of the parameters considered also does not match the given description.

discrete limit law for  $K_n$  as  $n$  tends to infinity. Then, we relate  $K_n$  to the random variables  $N_{n,k}$ , counting the number of buckets of capacity  $k$ ,  $1 \leq k \leq b$ .

Next, we turn to the analysis of the r.v.  $Y_{n,j}$ , counting the number of descendants of label  $j$ . The results for  $K_n$  and its limit law are used to obtain the exact distribution of  $Y_{n,j}$ , as well as several limit laws depending on the growth of  $j = j(n)$  as  $n$  tends to infinity. Moreover, we study the random variable  $X_{n,j}$  counting the out-degree of the bucket containing label  $j$  in a size  $n$  bucket tree. The analysis of  $X_{n,j}$  is based on a stopping time closely related to  $Y_{n,j}$ . We also provide decompositions of the random variables  $X_{n,j}$  and  $Y_{n,j}$  in terms of  $K_n$  for fixed  $n$ .

**1.1. Notation.**

We denote with  $X \stackrel{(d)}{=} Y$  the equality in distribution of two random variables  $X$  and  $Y$ . We write  $X_n \xrightarrow{(d)} X$  for the weak convergence (i.e., convergence in distribution) of a sequence of random variables  $X_n$  to a r.v.  $X$ . Let  $H_n := \sum_{k=1}^n \frac{1}{k}$  denote the harmonic numbers and  $H_{n+\alpha} - H_\alpha := \sum_{k=1}^n \frac{1}{k+\alpha}$  the continuation of the harmonic numbers for a complex  $\alpha \in \mathbb{C} \setminus \{-1, -2, -3, \dots\}$ . Here and throughout this work we use the notation  $x^{\underline{s}} := x(x-1) \dots (x-(s-1))$  for the falling factorials, and  $x^{\overline{s}} := x(x+1) \dots (x+s-1)$  for the rising factorials,  $s \in \mathbb{N}_0$ .<sup>2</sup> Throughout this work we will use the abbreviation (b,d)-ary ITs for (b,d)-ary bucket increasing trees and (b,  $\alpha$ )-PORTs for bucket (b,  $\alpha$ )-plane-oriented increasing trees.

**2. Description of bucket increasing trees**

**2.1. Tree evolution processes**

In [18] a stochastic growth rule for bucket recursive trees has been given, generalising the rule for ordinary random recursive trees (which are the special instance  $b = 1$ ): a bucket tree grows by progressive attraction of increasing integer labels: when inserting element  $n + 1$  into an existing bucket recursive tree containing  $n$  elements (i.e., containing the labels  $\{1, 2, \dots, n\}$ ) all  $n$  existing elements in the tree compete to attract the element  $n + 1$ , where all existing elements have equal chance to recruit the new element. If the element winning this competition is contained in a node with less than  $b$  elements (an unsaturated bucket or node), element  $n + 1$  is added to this node, otherwise if the winning element is contained in a node with already  $b$  elements (a saturated bucket or node), element  $n + 1$  is attached to this node as a new bucket containing only the element  $n + 1$ . Starting with a single bucket as root node containing only element 1 leads after  $n - 1$  insertion steps, where the labels  $2, 3, \dots, n$  are successively inserted according to this growth rule, to a so called random bucket recursive tree with  $n$  elements and maximal bucket size  $b$ . Of course, the above growth rule for inserting the element  $n + 1$  could also be formulated by saying that, for an existing bucket recursive tree  $T$  with  $n$  elements, the probability that a certain node  $v \in T$  attracts the new element  $n + 1$  is proportional to the number of elements contained in  $v$ , let us say  $k$  with  $1 \leq k \leq b$ , and is thus given by  $\frac{k}{n}$ . Summarising this procedure we obtain the following definition.

**Definition 1** (Bucket recursive trees). We start with a single bucket as root node containing only label 1. Given a tree  $T$  of size  $n \geq 1$ . Let  $p(v) = \mathbb{P}(n + 1 <_t v \mid c(v))$  denote the probability that node  $v \in T$  attracts label  $n + 1$  conditioned on its capacity  $c(v)$ . The family of random *bucket recursive trees* is generated according to the probabilities

$$p(v) = \frac{c(v)}{n},$$

with capacity  $1 \leq c(v) \leq b$ , thus independent of the out-degree  $\deg^+(v) \geq 0$  of node  $v$ .

<sup>2</sup>The notation  $x^{\underline{s}}$  and  $x^{\overline{s}}$  was introduced and popularised by Knuth; alternative notations for the falling factorials include the Pochhammer symbol  $(x)_s$ , which is unfortunately sometimes also used for the rising factorials.

Note that here and throughout this work the capacities  $c(v) = c_n(v)$  and the out-degree  $\text{deg}^+(v) = \text{deg}_n^+(v)$  of a node  $v$  in a tree  $T$  are always dependent on the size  $|T| = n$ .

In the following we present two new stochastic growth rules. They generate families of random bucket increasing trees with bucket size  $b \geq 1$ . Concerning the created trees, they are by definition *unordered*. For both rules there is an additional dependence on the out-degree  $\text{deg}^+(v)$  of the nodes.

**Definition 2** (( $b,d$ )-ary increasing). We start, case  $n = 1$ , with a single bucket as root node containing only label 1. Given a tree  $T$  of size  $n \geq 1$ . Let  $p(v)$  denote the probability that node  $v \in T$  attracts label  $n + 1$  in a bucket increasing tree of size  $n \in \mathbb{N}$ .

The family of random ( $b,d$ )-ary increasing trees, with  $d \in \mathbb{Q}$  such that  $(d - 1)b \in \mathbb{N}$ , is generated according to the probabilities

$$p(v) = \frac{(d - 1)c(v) + 1 - \text{deg}^+(v)}{(d - 1)n + 1},$$

with  $1 \leq c(v) \leq b$  and  $\text{deg}^+(v) \geq 0$ .

**Definition 3** (( $b,\alpha$ )-plane oriented recursive trees). We start, case  $n = 1$ , with a single bucket as root node containing only label 1. Given a tree  $T$  of size  $n \geq 1$ . The family of random ( $b,\alpha$ )-plane oriented recursive trees, with  $\alpha > 0$ , is generated according to the probabilities  $p(v)$  that node  $v \in T$  attracts label  $n + 1$ :

$$p(v) = \frac{\text{deg}^+(v) + (\alpha + 1)c(v) - 1}{(\alpha + 1)n - 1},$$

with  $1 \leq c(v) \leq b$  and  $\text{deg}^+(v) \geq 0$ .

**Remark 4.** Panholzer and Prodinger [21] characterised increasing tree families, which can be constructed by a simple stochastic growth rule. They obtained a unifying description based on two parameter, leading to three different families of trees. The two growth processes for ( $b,d$ )-ary increasing trees and ( $b,\alpha$ )-PORTs together with the process for random bucket recursive trees stated before, generalise the processes of [21], namely, they correspond to the special case of bucket size  $b = 1$ . For general  $b \in \mathbb{N}$ , we can show [10] that the three processes introduced before are the only ones, which can be constructed by a simple stochastic growth rule.

**Remark 5** (Linear bucket increasing trees). Plane-oriented recursive trees (PORTs, i.e., (1,1)-PORTs) and their generalisation also appeared in the literature under different names: Pittel [22] calls such tree families linear increasing trees, also unifying recursive trees and  $d$ -ary increasing trees into a single family. Closely related tree families are so-called scale-free trees, which can also be generated by a preferential attachment rule.

In the spirit of Pittel, we can define *linear bucket increasing trees* as follows. Given a tree  $T$  of size  $n \geq 1$ , a label  $n + 1$  is attracted by a node  $v \in T$  that is chosen with probability proportional to

$$\alpha \cdot (c(v) - 1) + \beta \cdot \text{deg}^+(v) + m,$$

with  $1 \leq c(v) \leq b$  and  $\text{deg}^+(v) \geq 0$ . Here  $\alpha, \beta, m \in \mathbb{R}$  denote real parameters such that the sum is non-negative. For  $b = 1$  we have  $c(v) = 1$  and we reobtain Pittel’s linear increasing trees.

**2.2 Combinatorial description of bucket increasing tree families**

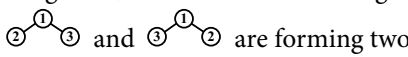
Our basic objects are rooted ordered trees  $T \in \mathcal{B}$ . Here, the order of the subtrees of a node is of relevance. The nodes of the trees are buckets with an integer capacity  $c$ , with  $1 \leq c \leq b$  for a given maximal integer bucket size  $b \geq 1$ . We assume that all internal nodes (i.e., non-leaves) in the tree must be saturated, while the leaves might be either saturated or unsaturated. Here  $\mathcal{B}$  denotes the family of all bucket ordered trees with maximal bucket size  $b$ . A tree  $T$  defined in this way is

called a bucket ordered tree with maximal bucket size  $b$ . As already mentioned we define for bucket ordered trees the size  $|T|$  of a tree  $T$  via  $|T| = \sum_v c(v)$ , where  $c(v)$  ranges over all vertices of  $T$ . An increasing labelling  $\ell(T)$  of a bucket ordered tree  $T$  is then a labelling of  $T$ , where the labels  $\{1, 2, \dots, |T|\}$  are distributed amongst the nodes of  $T$ , such that the following conditions are satisfied: (i) every node  $v$  contains exactly  $c(v)$  labels, (ii) the labels within a node are arranged in increasing order, (iii) each sequence of labels along any path starting at the root is increasing. A bucket ordered increasing tree  $\tilde{T}$  is given by a pair  $\tilde{T} = (T, \ell(T))$ .

Then a class  $\mathcal{T}$  of a family of bucket increasing trees with maximal bucket size  $b$  can be defined in the following way. A sequence of non-negative numbers  $(\varphi_k)_{k \geq 0}$  with  $\varphi_0 > 0$  and a sequence of non-negative numbers  $\psi_1, \psi_2, \dots, \psi_{b-1}$  is used to define the weight  $w(T)$  of any bucket ordered tree  $T$  by  $w(T) := \prod_v w(v)$ , where  $v$  ranges over all vertices of  $T$ . The weight  $w(v)$  of a node  $v$  is given as follows, where  $\text{deg}^+(v)$  denotes the out-degree (i.e., the number of children) of node  $v$ :

$$w(v) = \begin{cases} \varphi_{\text{deg}^+(v)}, & \text{if } c(v) = b, \\ \psi_{c(v)}, & \text{if } c(v) < b. \end{cases}$$

Thus, for saturated nodes the weight depends on the out-degree  $\text{deg}^+(v)$  and is described by the sequence  $\varphi_k$ , whereas for unsaturated nodes the weight depends on the capacity  $c(v)$  and is described by the sequence  $\psi_k$ .

Furthermore,  $\mathcal{L}(T)$  denotes the set of different increasing labellings  $\ell(T)$  of the tree  $T$  with distinct integers  $\{1, 2, \dots, |T|\}$ , where  $L(T) := |\mathcal{L}(T)|$  denotes its cardinality. The family  $\mathcal{T}$  consists of all trees  $\tilde{T} = (T, \ell(T))$ , with their weights  $w(T)$  and the set of increasing labellings  $\mathcal{L}(T)$  and we define  $w(\tilde{T}) := w(T)$ . Concerning bucket ordered increasing trees, note that the left-to-right order of the subtrees of the nodes is relevant. E.g., the trees  are forming two different trees.

For a given degree-weight sequence  $(\varphi_k)_{k \geq 0}$  with a degree-weight generating function  $\varphi(t) := \sum_{k \geq 0} \varphi_k t^k$  and a bucket-weight sequence  $\psi_1, \dots, \psi_{b-1}$ , we define now the total weights  $T_n$  by

$$T_n := \sum_{T \in \mathcal{B}: |T|=n} w(T) \cdot L(T) = \sum_{\tilde{T}=(T,\ell(T)) \in \mathcal{T}: |T|=n} w(\tilde{T}).$$

It is advantageous for such enumeration problems to describe a family of increasing trees  $\mathcal{T}$  by the following formal recursive equation:

$$\begin{aligned} \mathcal{T} &= \psi_1 \cdot \boxed{1} \dot{\cup} \psi_2 \cdot \boxed{1|2} \dot{\cup} \dots \dot{\cup} \psi_{b-1} \cdot \boxed{1|2|\dots|b-1} \dot{\cup} \\ &\quad \varphi_0 \cdot \boxed{1|2|\dots|b} \dot{\cup} \varphi_1 \cdot \boxed{1|2|\dots|b} \times \mathcal{T} \dot{\cup} \varphi_2 \cdot \boxed{1|2|\dots|b} \times \mathcal{T} * \mathcal{T} \dot{\cup} \varphi_3 \cdot \boxed{1|2|\dots|b} \times \mathcal{T} * \mathcal{T} * \mathcal{T} \dot{\cup} \dots \\ &= \psi_1 \cdot \boxed{1} \dot{\cup} \psi_2 \cdot \boxed{1|2} \dot{\cup} \dots \dot{\cup} \psi_{b-1} \cdot \boxed{1|2|\dots|b-1} \dot{\cup} \boxed{1|2|\dots|b} \times \varphi(\mathcal{T}), \end{aligned} \tag{1}$$

where  $\boxed{1|2|\dots|k}$  denotes a bucket of capacity  $k$  labelled by  $1, 2, \dots, k$ ,  $\times$  the cartesian product,  $*$  the partition product for labelled objects, and  $\varphi(\mathcal{T})$  the substituted structure. On the other hand, we can use standard notation [6]. Let  $\mathcal{Z}$  denote the atomic class (i.e., a single (uni)labelled node),  $\mathcal{A}^\square * \mathcal{B}$  the boxed product (i.e., the smallest label is constrained to lie in the  $\mathcal{A}$  component) of the combinatorial classes  $\mathcal{A}$  and  $\mathcal{B}$ . Then,

$$\begin{aligned} \mathcal{T} &= \psi_1 \cdot \mathcal{Z}^\square + \psi_2 \cdot (\mathcal{Z}^\square)^2 + \dots + \psi_{b-1} \cdot (\mathcal{Z}^\square)^{b-1} + (\mathcal{Z}^\square)^b * \varphi(\mathcal{T}) \\ &= \sum_{k=1}^{b-1} \psi_k \cdot (\mathcal{Z}^\square)^k + (\mathcal{Z}^\square)^b * \varphi(\mathcal{T}). \end{aligned}$$

Here the meaning of  $(\mathcal{Z}^\square)^k * \mathcal{B}$  is  $\mathcal{Z}^\square * (\mathcal{Z}^\square * (\dots * (\mathcal{Z}^\square * \mathcal{B})))$ , with  $k$  occurrences of  $\mathcal{Z}^\square$ .

Using above formal description, one can show that the exponential generating function  $T(z) := \sum_{n \geq 1} T_n \frac{z^n}{n!}$  of the total weights  $T_n$  is characterised by the following result of [13].

**Proposition 6.** *The exponential generating function  $T(z)$  of bucket increasing trees with degree-weight generating function  $\varphi(t)$  satisfies an ordinary differential equation of order  $b$ :*

$$\frac{d^b}{dz^b} T(z) = \varphi(T(z)), \tag{2}$$

with initial conditions

$$T(0) = 0, \quad T^{(k)}(0) = \psi_k, \quad \text{for } 1 \leq k \leq b - 1.$$

**Example 7** (Bucket size one – ordinary increasing trees). In the case of bucket size  $b = 1$  we obtain ordinary increasing trees. We have a simple description using the *boxed product*:

$$\mathcal{T} = \mathcal{Z}^\square * \varphi(\mathcal{T}). \tag{3}$$

The exponential generating function  $T(z)$  of bucket increasing trees with degree-weight generating function  $\varphi(t)$  is implicitly defined by

$$\int_0^{T(z)} \frac{dt}{\varphi(t)} = z.$$

Prominent varieties include recursive trees, binary increasing and plane-oriented recursive trees; for an overview see, e.g., [5, 6].

**Example 8** (Bucket size two – bilabelled trees). Trees with bucket size  $b = 2$ , degree-weight generating function  $\varphi(t)$  and weight  $\psi = \psi_1 \geq 0$  correspond to so-called bilabelled trees. The family  $\widehat{\mathcal{T}}$  can be described by the following symbolic equation:

$$\mathcal{T} = \mathcal{Z}^\square * \left( \mathcal{Z}^\square * \varphi(\mathcal{T}) \right), \tag{4}$$

The differential equation  $T''(z) = \varphi(T(z))$  can be readily translated to a first-order equation. Namely, multiplication with  $T'(z)$  and integration leads to the first-order differential equation

$$T'(z) = \sqrt{\psi^2 + 2 \cdot \Phi(T(z))}, \quad T(0) = 0, \tag{5}$$

with  $\Phi(x) = \int_0^x \varphi(t) dt$ . Hence,  $T = T(z)$  is implicitly given via

$$\int_0^T \frac{dx}{\sqrt{\psi^2 + 2 \cdot \Phi(x)}} = z.$$

The special choice  $\psi = 0$  leads to so-called strict bilabelled increasing tree families. Here, due to the choice  $\psi = 0$ , all nodes have to be saturated. Such tree families naturally give rise to hook-length formulas; see [15] for many examples.

**Example 9** (Strict  $b$  labelled tree families). Families of strictly  $b$ -labelled trees have been recently studied due to their connection to hook-length formulas [14, 15]. Here  $\psi_1 = \dots = \psi_{b-1} = 0$ , thus only saturated nodes are allowed.

### 3. Combinatorial models of the tree evolution processes

#### 3.1 Random tree models

For each class  $\mathcal{T}$  of bucket increasing trees associated to a given degree-weight sequence  $(\varphi_k)_{k \geq 0}$  and bucket-weight sequence  $\psi_1, \dots, \psi_{b-1}$  we define in a natural way probability models for

random (ordered) bucket increasing trees  $\mathcal{T}_n$  of size  $n$ . We assume that each increasingly labelled bucket ordered increasing tree  $\tilde{T} = (T, \ell(T)) \in \mathcal{T}_n$  of size  $n$  is chosen with a probability proportional to its weight  $w(\tilde{T})$ .

**Definition 10** (Random bucket ordered increasing trees). A probability measure on ordered bucket increasing trees  $\tilde{T} \in \mathcal{T}_n$  of size  $n$  is defined by

$$\mathbb{P}^{[\mathcal{T}]}(\{\tilde{T}\}) = \frac{w(\tilde{T})}{T_n} = \frac{\left(\prod_{v \in \tilde{T}: c(v) < b} \psi_{c(v)}\right) \cdot \left(\prod_{v \in \tilde{T}: c(v) = b} \varphi_{\text{deg}^+(v)}\right)}{T_n}.$$

We speak then about *random ordered bucket increasing trees* of size  $n$  of the family  $\mathcal{T}$  under the *random tree model*.

From a combinatorial point of view it is often convenient to work with ordered trees. However, the trees generated by the tree evolution processes are by definition unordered. Thus, we turn our attention to unordered trees. Our basic objects are unordered bucket trees  $T^{[U]}$  together with an increasing labelling  $\ell(T^{[U]})$ . In order to obtain a measure on unordered bucket increasing trees  $\mathcal{T}^{[U]}$  we proceed in a standard way embedding unordered trees into ordered trees. Each unordered bucket increasing tree  $\tilde{T}^{[U]} = (T^{[U]}, \ell(T^{[U]}))$  corresponds to

$$\prod_{v \in T^{[U]}} \text{deg}^+(v)!$$

different ordered bucket increasing trees. We obtain a canonical ordered representative  $\tilde{T} \in \mathcal{T}$  of the unordered tree  $\tilde{T}^{[U]} = (T^{[U]}, \ell(T^{[U]}))$  by ordering the subtrees of each node  $v \in \tilde{T}^{[U]}$  in an increasing left-to-right way according to the smallest labels contained in the respective subtrees taking into account the labelling  $\ell(T^{[U]})$ . Let  $f: \mathcal{T}^{[U]} \rightarrow \mathcal{T}$  denote this injective ordering map.

**Definition 11** (Random unordered bucket increasing trees). A probability measure  $\mathbb{P}^{[\mathcal{T}]}$  on unordered bucket increasing trees  $\tilde{T}^{[U]} \in \mathcal{T}^{[U]}_n$  of size  $n$  is defined by

$$\mathbb{P}^{[\mathcal{T}]}(\{\tilde{T}^{[U]}\}) = \frac{(w \circ f)(\tilde{T}^{[U]}) \cdot \prod_{v \in T^{[U]}} \text{deg}^+(v)!}{T_n}.$$

We speak then about *random unordered bucket increasing trees* of size  $n$  of the family  $\mathcal{T}^{[U]}$  under the *random tree model*.

Possibly different models of randomness are introduced when generating an unordered tree of size  $n$  according to one of the three tree evolution processes described before. Given an integer  $j \geq 2$  and any tree  $\tilde{T}^{[U]} = (T^{[U]}, \ell(T^{[U]}))$ , we denote by  $\text{attr}(j)$  the node in  $\tilde{T}^{[U]}$  that attracted label  $j$ , and by  $\tilde{T}_{<j}^{[U]}$  the tree obtained from  $\tilde{T}^{[U]}$  when restricting to labels less than  $j$ . Then we define  $\pi^{(j)}$  as the map  $\pi^{(j)}: \tilde{T}^{[U]} \rightarrow [0, 1]$  that gives the probability that label  $j$  will be attracted by node  $\text{attr}(j)$  in  $\tilde{T}_{<j}^{[U]}$ :

$$\pi^{(j)}(\tilde{T}^{[U]}) = \begin{cases} \mathbb{P}\{j <_t \text{attr}(j) \mid \tilde{T}_{<j}^{[U]}\}, & |\tilde{T}^{[U]}| \geq j, \\ 0, & |\tilde{T}^{[U]}| < j, \end{cases}$$

with  $\mathbb{P}\{j <_t \text{attr}(j) \mid \tilde{T}_{<j}^{[U]}\}$  being determined by the tree evolution processes.



The point-wise product  $\mathbb{P}^{[e]} = \prod_{j=2}^n \pi^{(j)}$  is then a probability measure on unordered bucket increasing trees  $\mathcal{T}_n^{[U]}$  of size  $n$ . In the following we denote with a superscript the source of randomness on  $\mathcal{T}_n$ :  $\mathbb{P}^{[T]}$  for the random tree model and  $\mathbb{P}^{[e]}$  for the tree evolution process. It holds

$$\mathbb{P}^{[e]} \{ \tilde{T}^{[U]} \} = \prod_{j=2}^n \pi^{(j)} \left( \tilde{T}^{[U]} \right).$$

Let  $\text{deg}_{<j}^+(v)$  be the out-degree of node  $v$  when restricting to labels less than  $j$ . Then, a probability measure on ordered trees is obtained via

$$\mathbb{P}^{[e]} \{ T \} = \prod_{j=2}^n \frac{\pi^{(j)} \left( \tilde{T}^{[U]} \right)}{\text{deg}_{<j}^+(\text{attr}(j))}.$$

### 3.2 Combinatorial models

It was already proven in [13] that bucket recursive trees generated according to the growth process as stated in Subsection 2.1 can be considered as a certain bucket increasing tree family. In the following we extend this result, where, for the sake of completeness, we also collect the findings of [13] for bucket recursive trees.

**Theorem 12** (Combinatorial models for families generated by a stochastic growth rule). *The tree evolution processes that generate families of random unordered bucket recursive trees, (b,d)-ary increasing trees and (b, α)-plane oriented recursive trees, with bucket size  $b \geq 1$ , can be realised combinatorially by suitably chosen sequences of degree-weights  $(\varphi_k)_{k \geq 0}$  and bucket weights  $\psi_1, \dots, \psi_{b-1}$ . Given an arbitrary ordered bucket increasing tree  $T \in \mathcal{T}$  of size  $|T| = n$ , then it holds that under the random tree model the probability that a new element  $n + 1$  is attracted by a node  $v \in T$  with capacity  $c(v) = k$  is given by  $p(v) = \mathbb{P}\{n + 1 <_t v \mid c(v) = k\}$  as defined in Definitions 1-3 for the corresponding tree evolution process.*

Consequently, both models of randomness for (un)ordered trees coincide:  $\mathbb{P}^{[e]} = \mathbb{P}^{[T]}$ .

1. Bucket recursive trees: a combinatorial model can be obtained from the sequences

$$\varphi_k = \frac{(b - 1)!b^k}{k!}, \quad \text{for } k \geq 0, \quad \psi_k = (k - 1)!, \quad \text{for } 1 \leq k \leq b - 1,$$

such that  $\varphi(t) = \sum_{k \geq 0} \varphi_k t^k = (b - 1)! \cdot \exp(bt)$ . For this model of bucket recursive trees the exponential generating function  $T(z)$  and the total weights  $T_n$  are given by

$$T(z) = \log \left( \frac{1}{1 - z} \right), \quad T_n = (n - 1)!$$

2. (b,d)-ary increasing trees: a combinatorial model can be obtained from the sequences

$$\varphi_k = (b - 1)!(d - 1)^{b-1} \binom{b - 1 + \frac{1}{d-1}}{b - 1} \binom{b(d - 1) + 1}{k}, \quad \text{for } k \geq 0,$$

$$\psi_k = (k - 1)!(d - 1)^{k-1} \binom{k - 1 + \frac{1}{d-1}}{k - 1}, \quad \text{for } 1 \leq k \leq b - 1,$$

such that  $\varphi(t) = (b - 1)!(d - 1)^{b-1} \binom{b-1+\frac{1}{d-1}}{b-1} (1 + t)^{b(d-1)+1}$ . For this model of (b,d)-ary increasing trees the exponential generating function  $T(z)$  and the total weights  $T_n = [z^n]T(z)$  are given by



$$T(z) = \frac{1}{(1 - (d - 1)z)^{\frac{1}{d-1}}} - 1, \quad T_n = (n - 1)!(d - 1)^{n-1} \binom{n - 1 + \frac{1}{d-1}}{n - 1}.$$

3.  $(b, \alpha)$ -plane oriented recursive trees: a combinatorial model can be obtained from the sequences

$$\begin{aligned} \varphi_k &= (b - 1)!(\alpha + 1)^{b-1} \binom{b - 1 - \frac{1}{\alpha+1}}{b - 1} \binom{(\alpha + 1)b - 2 + k}{k}, \quad \text{for } k \geq 0, \\ \psi_k &= (k - 1)!(\alpha + 1)^{k-1} \binom{k - 1 - \frac{1}{\alpha+1}}{k - 1}, \quad \text{for } 1 \leq k \leq b - 1, \end{aligned}$$

such that  $\varphi(t) = \frac{(b-1)!(\alpha+1)^{b-1} t^{b-1-\frac{1}{\alpha+1}}}{(1-t)^{(\alpha+1)b-1}}$ . For this model of  $(b, \alpha)$ -plane oriented recursive trees the exponential generating function  $T(z)$  and the total weights  $T_n = [z^n]T(z)$  are given by

$$T(z) = 1 - (1 - (\alpha + 1)z)^{\frac{1}{\alpha+1}}, \quad T_n = (n - 1)!(\alpha + 1)^{n-1} \binom{n - 1 - \frac{1}{\alpha+1}}{n - 1}.$$

**Proof.** To prove that these choices of sequences  $(\varphi_k)_{k \in \mathbb{N}}$  and  $(\psi_k)_{1 \leq k \leq b-1}$  are actually models for bucket increasing trees generated according to the stochastic growth rules defined in Subsection 2.1, we have to show that the combinatorial families  $\mathcal{T}$  of bucket increasing trees have the same stochastic growth rules as the counterparts created probabilistically. Given an arbitrary bucket increasing tree  $T \in \mathcal{T}$  of size  $|T| = n$ , then the probability that a new element  $n + 1$  is attracted by a node  $v \in T$  with capacity  $c(v) = k$  has to coincide with the corresponding probability stated in Definitions 1–3.

We use now the notation  $T \rightarrow T'$  to denote that  $T'$  is obtained from  $T$  with  $|T| = n$  by incorporating element  $n + 1$ , i.e., either by attaching element  $n + 1$  to a saturated node  $v \in T$  at one of the  $\deg^+(v) + 1$  possible positions (recall that bucket increasing trees are per definition ordered trees and thus the order of the subtrees is of relevance) by creating a new bucket of capacity 1 containing element  $n + 1$  or by adding element  $n + 1$  to an unsaturated node  $v \in T$  by increasing the capacity of  $v$  by 1. If we want to express that node  $v \in T$  has attracted the element  $n + 1$  leading from  $T$  to  $T'$  we use the notation  $T \xrightarrow{v} T'$ . If there exists a stochastic growth rule for a bucket increasing tree family  $\mathcal{T}$ , then it must hold that for a given tree  $T \in \mathcal{T}$  of size  $|T| = n$  and a given node  $v \in T$  the probability  $p_T(v)$ , which gives the probability that element  $n + 1$  is attracted by node  $v \in T$ , is given as follows:

$$p_T(v) = \frac{\sum_{T' \in \mathcal{T}: T \xrightarrow{v} T'} w(T')}{\sum_{\tilde{T} \in \mathcal{T}: T \rightarrow \tilde{T}} w(\tilde{T})} = \frac{\sum_{T' \in \mathcal{T}: T \xrightarrow{v} T'} \frac{w(T')}{w(T)}}{\sum_{\tilde{T} \in \mathcal{T}: T \rightarrow \tilde{T}} \frac{w(\tilde{T})}{w(T)}}. \tag{6}$$

The remaining task is to simplify the expression above into the form stated in Definitions 1–3. For a certain tree  $\tilde{T}$  with  $T \xrightarrow{u} \tilde{T}$  and  $u \in T$  the quotient of the weight of the trees  $\tilde{T}$  and  $T$  is due to the definition of bucket increasing trees given as follows, where we define for simplicity  $\psi_b := \varphi_0$ :

$$\frac{w(\tilde{T})}{w(T)} = \begin{cases} \psi_1 \frac{\varphi_{k+1}}{\varphi_k}, & \text{for } c(u) = b \text{ and } \deg^+(u) = k, \\ \frac{\psi_{k+1}}{\psi_k}, & \text{for } c(u) = k < b. \end{cases}$$

For a given tree  $T \in \mathcal{T}$  we define by  $m_k := |\{u \in T: c(u) = k < b\}|$  the number of unsaturated nodes of  $T$  with capacity  $k < b$  and by  $n_k := |\{u \in T: c(u) = b \text{ and } \deg^+(u) = k\}|$  the number of saturated nodes of  $T$  with out-degree  $k \geq 0$ . It holds then

$$n = \sum_{u \in T} c(u) = \sum_{k=1}^{b-1} km_k + b \sum_{k \geq 0} n_k \tag{7}$$

and (where we use that there are  $k + 1$  possibilities of attaching a new node to a saturated node  $u \in T$  with out-degree  $\text{deg}^+(u) = k$ ):

$$\sum_{\tilde{T} \in \mathcal{T}: T \rightarrow \tilde{T}} \frac{w(\tilde{T})}{w(T)} = \sum_{k=1}^{b-1} m_k \frac{\psi_{k+1}}{\psi_k} + \sum_{k \geq 0} n_k (k + 1) \psi_1 \frac{\varphi_{k+1}}{\varphi_k}.$$

Moreover, we also have the relation

$$1 = \sum_{k=1}^{b-1} m_k - \sum_{k \geq 0} (k - 1)n_k, \tag{8}$$

which follows as the difference between the node-sum and edge-sum equation for the tree  $T$ :

$$\# \text{ nodes} = \sum_{k=1}^{b-1} m_k + \sum_{k \geq 0} n_k, \quad \# \text{ edges} = \# \text{ nodes} - 1 = \sum_{k \geq 0} kn_k.$$

First we turn our attention to the family of  $(b,d)$ -ary increasing trees and the weights as given in Theorem 12. We have

$$\sum_{T' \in \mathcal{T}: T \xrightarrow{v} T'} \frac{w(T')}{w(T)} = \begin{cases} (k + 1)\psi_1 \frac{\varphi_{k+1}}{\varphi_k} = b(d - 1) + 1 - k, & \text{for } c(v) = b \text{ and } \text{deg}^+(v) = k, \\ \frac{\psi_{k+1}}{\psi_k} = k(d - 1) + 1, & \text{for } c(v) = k < b, \end{cases}$$

and consequently

$$\begin{aligned} \sum_{\tilde{T} \in \mathcal{T}: T \rightarrow \tilde{T}} \frac{w(\tilde{T})}{w(T)} &= \sum_{k=1}^{b-1} (k(d - 1) + 1)m_k + \sum_{k \geq 0} n_k (b(d - 1) + 1 - k) \\ &= (d - 1) \left( \sum_{k=1}^{b-1} km_k + b \sum_{k \geq 0} n_k \right) + \sum_{k=1}^{b-1} m_k - \sum_{k \geq 0} (k - 1)n_k = (d - 1)n + 1, \end{aligned}$$

due to equations (7) and (8). Thus, with this choice of weight sequences  $(\varphi_k)_k$  and  $(\psi_k)_k$ , the probability  $p_T(v)$  that in a bucket increasing tree  $T$  of size  $|T| = n$  the node  $v$  with capacity  $c(v) = k$  attracts element  $n + 1$  coincides with the corresponding probability of the stochastic growth rule for  $(b,d)$ -ary increasing trees given in Definition 2.

We obtain then from equation (2) that the exponential generating function  $T(z) := \sum_{n \geq 1} T_n \frac{z^n}{n!}$  of the total-weight  $T_n$  of bucket increasing trees of size  $n$  satisfies the differential equation

$$\frac{d^b}{dz^b} T(z) = (b - 1)!(d - 1)^{b-1} \binom{b - 1 + \frac{1}{d-1}}{b - 1} (1 + T(z))^{b(d-1)+1}, \tag{9}$$

with initial conditions  $T(0) = 0$  and  $\left. \frac{d^k}{dz^k} T(z) \right|_{z=0} = (k - 1)!(d - 1)^{k-1} \binom{k-1 + \frac{1}{d-1}}{k-1}$ , for  $1 \leq k \leq b - 1$ .

The solution of this equation is given by

$$T(z) = \frac{1}{(1 - (d - 1)z)^{\frac{1}{d-1}}} - 1 = \sum_{n \geq 1} (n - 1)!(d - 1)^{n-1} \binom{n - 1 + \frac{1}{d-1}}{n - 1} \frac{z^n}{n!}, \tag{10}$$

as can be checked easily. Hence the total weight of all size  $n$   $(b,d)$ -ary increasing trees is given by  $T_n = (n - 1)!(d - 1)^{n-1} \binom{n-1 + \frac{1}{d-1}}{n-1}$  as stated in Theorem 12.

**Table 1.** Summary: bucket increasing tree families and their growth processes

Tree family	Growth process : $p(v)$
Bucket recursive trees	$\frac{c(v)}{n}$
$(b, d)$ -ary increasing trees	$\frac{(d-1)c(v) + 1 - \text{deg}^+(v)}{(d-1)n + 1}$
$(b, \alpha)$ -PORT	$\frac{\text{deg}^+(v) + (\alpha + 1)c(v) - 1}{(\alpha + 1)n - 1}$

For  $(b, \alpha)$ -plane oriented recursive trees and weights as given in Theorem 12 we obtain

$$\sum_{T' \in \mathcal{T}: T \xrightarrow{v} T'} \frac{w(T')}{w(T)} = \begin{cases} (k+1)\psi_1 \frac{\varphi_{k+1}}{\varphi_k} = (\alpha+1)b - 1 + k, & \text{for } c(v) = b \text{ and } \text{deg}^+(v) = k, \\ \frac{\psi_{k+1}}{\psi_k} = k(\alpha+1) - 1, & \text{for } c(v) = k < b, \end{cases}$$

and consequently

$$\begin{aligned} \sum_{\tilde{T} \in \mathcal{T}: T \rightarrow \tilde{T}} \frac{w(\tilde{T})}{w(T)} &= \sum_{k=1}^{b-1} (k(\alpha+1) - 1)m_k + \sum_{k \geq 0} n_k((\alpha+1)b - 1 + k) \\ &= (\alpha+1) \left( \sum_{k=1}^{b-1} km_k + b \sum_{k \geq 0} n_k \right) - \left( \sum_{k=1}^{b-1} m_k - \sum_{k \geq 0} (k-1)n_k \right) = (\alpha+1)n - 1, \end{aligned}$$

due to equations (7) and (8). Again, with this choice of weight sequences  $(\varphi_k)_k$  and  $(\psi_k)_k$ , it follows that the probability  $p_T(v)$  that in a bucket increasing tree  $T$  of size  $|T| = n$  the node  $v$  with capacity  $c(v) = k$  attracts element  $n + 1$  coincides with the corresponding probability in the stochastic growth rule for  $(b, \alpha)$ -plane oriented recursive trees given in Definition 3.

We obtain then from equation (2) that the exponential generating function  $T(z) := \sum_{n \geq 1} T_n \frac{z^n}{n!}$  of the total-weight  $T_n$  of  $(b, \alpha)$ -plane oriented recursive trees of size  $n$  satisfies the differential equation

$$\frac{d^b}{dz^b} T(z) = (b-1)!(\alpha+1)^{b-1} \binom{b-1-\frac{1}{\alpha+1}}{b-1} \frac{1}{(1-T(z))^{b(\alpha+1)-1}}, \tag{11}$$

with initial conditions  $T(0) = 0$  and  $\left. \frac{d^k}{dz^k} T(z) \right|_{z=0} = (k-1)!(\alpha+1)^{k-1} \binom{k-1-\frac{1}{\alpha+1}}{k-1}$ , for  $1 \leq k \leq b-1$ . Again it can be checked easily that the solution of this equation is given by

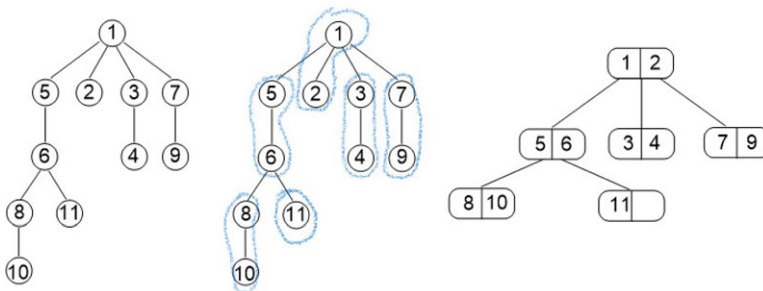
$$T(z) = (1 - (\alpha+1)z)^{\frac{1}{\alpha+1}} - 1 = \sum_{n \geq 1} (n-1)!(\alpha+1)^{n-1} \binom{n-1-\frac{1}{\alpha+1}}{n-1} \frac{z^n}{n!}. \tag{12}$$

Hence the total weight of all size  $n$   $(b, \alpha)$ -plane oriented recursive trees is given by  $T_n = (n-1)!(\alpha+1)^{n-1} \binom{n-1-\frac{1}{\alpha+1}}{n-1}$ , which finishes the proof of Theorem 12  $\square$

In Tables 1–2 we summarise the combinatorial properties as well as the growth processes of the bucket increasing tree families considered in this work.

**Table 2.** Summary: bucket increasing tree families and their combinatorial properties

Tree family	Degree-weight GF $\varphi(t)$	initial weights $\psi_k, 1 \leq k \leq b - 1$
Bucket recursive trees	$(b - 1)! \exp(b \cdot t)$	$(k - 1)!$
$(b, d)$ -ary increasing trees	$(b - 1)!(d - 1)^{b-1} \binom{b-1+\frac{1}{d-1}}{b-1} (1+t)^{b(d-1)+1}$	$(k - 1)!(d - 1)^{k-1} \binom{k-1+\frac{1}{d-1}}{k-1}$
$(b, \alpha)$ -PORT	$\frac{(b - 1)!(\alpha + 1)^{b-1} \binom{b-1-\frac{1}{\alpha+1}}{b-1}}{(1 - t)^{(\alpha+1)b-1}}$	$(k - 1)!(\alpha + 1)^{k-1} \binom{k-1-\frac{1}{\alpha+1}}{k-1}$



**Figure 1.** A plane-oriented recursive tree  $T$  of size eleven, the clustered tree with  $b = 2$  and the corresponding bilabelled increasing ordered tree  $\tilde{T} = \mathcal{C}(T)$ .

**4. Clustering process and weight sequences**

Up to this point, no indication has been given how to find the weight sequences in Theorem 12. Similarly, we did not give any motivation behind the definition of the growth processes. Here, as Algorithm 1 we will present a clustering map for ordinary increasing trees, case  $b = 1$ . It leads directly to the stated weight sequences and serves as a motivation behind the definition of the growth processes and weight sequences.

Let  $\mathcal{T}^{[b]}$  denote the family of ordered bucket increasing trees with maximal bucket size  $b$ , such that  $\mathcal{T}^{[1]}$  denotes the family of ordinary ordered increasing trees. Algorithm 1, which defines a map  $\mathcal{C}: \mathcal{T}^{[1]} \rightarrow \mathcal{T}^{[b]}$ , is given as follows; furthermore it is illustrated in Figure 1.

**Proposition 13.** *The map  $\mathcal{C}: \mathcal{T}^{[1]} \rightarrow \mathcal{T}^{[b]}$  defined in Algorithm 1 is not injective, but surjective.*

**Remark 14.** Note that for arbitrary increasing tree families the map  $\mathcal{C}$  does neither take into account the weight sequence  $(\phi_k)_{k \geq 0}$  of  $\mathcal{T}^{[1]}$ , nor it determines  $(\varphi_k)_{k \geq 0}$  and  $(\psi_k)_{1 \leq k \leq b-1}$  of  $\mathcal{T}^{[b]}$ . We will see later, see Examples 16 and 17, that the map can be modified to obtain bijections between ordinary increasing trees and specific families of bucket increasing trees.

**Proof.** Apparently, all trees  $T \in \mathcal{T}^{[1]}$  of size  $k, 1 \leq k \leq b$ , are mapped to a tree  $\mathcal{C}(T)$  consisting only of a single bucket  $v$  of capacity  $c(v) = k$ , thus the map is not injective, except for degenerate choices of the weight sequence. However, the map  $\mathcal{C}$  is surjective: every size  $n$  bucket increasing tree can be created by clustering an ordinary increasing tree. Given a bucket increasing tree  $\tilde{T} \in \mathcal{T}^{[b]}$ , we can replace all buckets by corresponding increasing chains, holding the corresponding labels. This gives an ordinary ordered increasing tree  $T \in \mathcal{T}^{[1]}$  with  $\mathcal{C}(T) = \tilde{T}$ . □

Given a family of ordinary increasing trees  $\mathcal{T}^{[1]}$  with weights  $(\phi_k)_{k \geq 0}$ , we are interest in the resulting clustered tree families and its properties. Since we can freely choose the weight sequences

**Algorithm 1** ClusteringIncreasingTrees (T,b)

**Input:** Ordered Increasing tree  $T \in \mathcal{T}^{[1]}$ , integer  $b \geq 2$   
**Result:** Ordered bucket increasing tree  $\tilde{T} = \mathcal{C}(T) \in \mathcal{T}^{[b]}$

- 1:  $V \leftarrow V(T), \tilde{T} \leftarrow \emptyset$
- 2: **do**
- 3:   Choose  $v_{\min} \in V$ , the node with minimal label
- 4:   **if**  $|\text{subtree}(v_{\min})| \geq b$  **then**  $s \leftarrow b$
- 5:   **else**  $s \leftarrow |\text{subtree}(v_{\min})|$
- 6:   **end if**
- 7:   Merge  $s$  smallest labelled nodes  $v_{\min} = v_1, v_2, \dots, v_s \in \text{subtree}(v_{\min})$  into new bucket  $v$ .
- 8:    $V \leftarrow V \setminus \{v_1, \dots, v_s\}$
- 9:   Redirect all the edges starting at any of these  $s$  nodes to the new bucket.
- 10:  $\tilde{T} \leftarrow \tilde{T} \cup \{v\}$
- 11: **While**  $V \neq \emptyset$

$(\varphi_k)_{k \geq 0}$  and  $(\psi_k)_{1 \leq k \leq b-1}$  of the associated bucket increasing families, we choose them in a weight preserving way, depending on the given weights  $(\phi_k)_{k \geq 0}$ .

**Definition 15** (Weight preserving bucket trees). Given a family of ordinary increasing trees  $\mathcal{T}^{[1]}$  with weights  $(\phi_k)_{k \geq 0}$ . We call a family  $\mathcal{T}^{[b]}$  of bucket increasing trees with weights  $(\varphi_k)_{k \geq 0}$  and  $(\psi_k)_{1 \leq k \leq b-1}$  *weight preserving*, if it holds, for all  $\tilde{T} \in \mathcal{T}_n$ ,

$$w(\tilde{T}) = \sum_{T \in \mathcal{T}^{[1]}: \mathcal{C}(T) = \tilde{T}} w(T). \tag{13}$$

Given  $(\phi_k)_{k \geq 0}$ , the total weights  $T_n$  of the bucket increasing trees, with weight preserving  $(\varphi_k)_{k \geq 0}$  and  $(\psi_k)_{1 \leq k \leq b-1}$ , equals the weight  $T_n$  of their ordinary increasing trees counterpart:

$$\sum_{\tilde{T} \in \mathcal{T}^{[b]}, |\tilde{T}| = n} w(\tilde{T}) = \sum_{\tilde{T} \in \mathcal{T}^{[b]}, |\tilde{T}| = n} \sum_{T \in \mathcal{T}^{[1]}: \mathcal{C}(T) = \tilde{T}} w(T) = \sum_{T \in \mathcal{T}^{[1]}, |T| = n} w(T) = T_n.$$

Moreover, we directly obtain from (13)

$$\psi_k = T_k, \quad 1 \leq k \leq b - 1, \quad \varphi_0 = T_b,$$

where  $T_k$  denotes the total weight of size  $k$  trees of the ordinary increasing tree family.

Next, we consider star-shaped trees  $\tilde{T}$  of size  $b + k$  and root degree  $k$ . We get

$$w(\tilde{T}) = \varphi_k \cdot \psi_1^k.$$

On the other hand, we can describe all trees  $T \in \mathcal{T}^{[1]}$  satisfying  $\mathcal{C}(T) = \tilde{T}$ . Namely, all such trees can be created by taking an arbitrary tree of size  $b$  and attaching  $k$  new nodes labelled  $k + 1, \dots, k + b$  to any of the existing  $b$  nodes. We denote by  $v_1, \dots, v_b$  the nodes labelled  $1, \dots, b$  in  $T$ , by  $\text{deg}_{\leq b}^+(v_m)$  the out-degree of node  $v_m$  when restricting to labels  $\leq b$ , and by  $j_m$  the number of nodes with labels  $> b$  attached to  $v_m$ . This gives

$$\sum_{T \in \mathcal{T}^{[1]}: \mathcal{C}(T) = \tilde{T}} w(T) = \phi_0^k \sum_{T \in \mathcal{T}^{[1]}: |T| = b} w(T) \sum_{\substack{j_s \geq 0 \\ \sum_{s=1}^b j_s = k}} \prod_{m=1}^b \frac{\phi_{\text{deg}_{\leq b}^+(v_m) + j_m}}{\phi_{\text{deg}_{\leq b}^+(v_m)}} \binom{\text{deg}_{\leq b}^+(v_m) + j_m}{j_m}.$$

We use  $\psi_1 = \phi_0 = T_1$  and obtain for  $\varphi_k$  the expression

$$\varphi_k = \sum_{T \in \mathcal{T}^{[1]}: |T|=b} w(T) \sum_{\substack{\sum_{s=1}^b j_s = k, \\ j_s \geq 0}} \prod_{m=1}^b \frac{\phi_{\deg_{\leq b}^+(v_m)+j_m}}{\phi_{\deg_{\leq b}^+(v_m)}} \binom{\deg_{\leq b}^+(v_m)+j_m}{j_m}. \tag{14}$$

As mentioned earlier, it has been shown in [21] that only three families of increasing trees can be constructed using a stochastic growth process. These are ordinary recursive trees, generalised plane-oriented recursive trees and  $d$ -ary increasing trees, and they are determined by the degree-weight sequences  $\phi_k = \frac{1}{k!}$ ,  $\phi_k = \binom{k+\alpha-1}{k}$  and  $\phi_k = \binom{d}{k}$ ,  $k \geq 0$ , respectively. The three weight sequences  $(\phi_k)_{k \geq 0}$  together with (14) directly lead to the result of Theorem 12. Below we present the calculations for  $(b, \alpha)$ -plane oriented recursive trees derived from generalised plane-oriented recursive trees with  $\phi_k = \binom{k+\alpha-1}{k}$ . Using

$$\frac{\phi_{\deg_{\leq b}^+(v_m)+j_m}}{\phi_{\deg_{\leq b}^+(v_m)}} \binom{\deg_{\leq b}^+(v_m)+j_m}{j_m} = \binom{\deg_{\leq b}^+(v_m)+j_m+\alpha-1}{j_m}$$

and

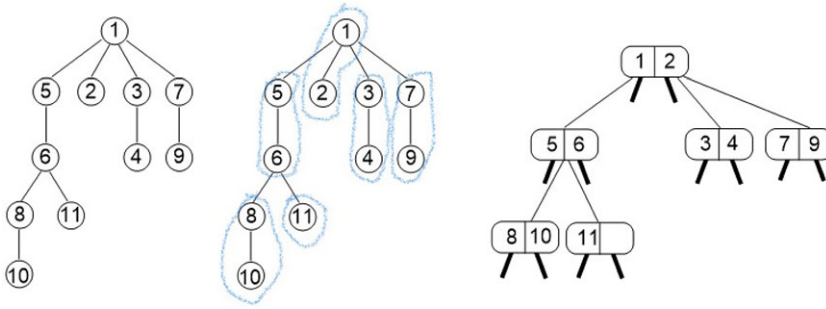
$$\begin{aligned} \sum_{\substack{\sum_{s=1}^b j_s = k, \\ j_s \geq 0}} \prod_{m=1}^b \binom{\deg_{\leq b}^+(v_m)+j_m+\alpha-1}{j_m} &= \sum_{\substack{\sum_{s=1}^b j_s = k, \\ j_s \geq 0}} \prod_{m=1}^b [t^{j_m}] \frac{1}{(1-t)^{\deg_{\leq b}^+(v_m)+\alpha}} \\ &= [t^k] \frac{1}{(1-t)^{b-1+b\alpha}} = \binom{b+b\alpha+k-2}{k}, \end{aligned}$$

we get indeed the result stated in Theorem 12:

$$\varphi_k = \sum_{T \in \mathcal{T}_1: |T|=b} w(T) \cdot \binom{b+b\alpha+k-2}{k} = T_b \cdot \binom{b+b\alpha+k-2}{k}. \tag{15}$$

**Example 16** (Bijection between PORTs and three-bundled bilabelled bucket ordered increasing trees). Following the terminology of [8], we call an ordered (bucket) increasing tree  $d$ -bundled, if every node has  $d$  positions, with a (possibly empty) sequence of  $d$ -bundled trees (with disjoint sets of labels) attached to each position. Equivalently, one may think of each node of the ordered tree as having  $d - 1$  separation walls, which can be regarded as a special type of edges or half-edges, that separate the subtrees of each node into  $d$  bundles. It is known that  $d$ -bundled increasing trees correspond to ordinary ordered increasing trees with degree-weight generating function  $\varphi(t) = 1/(1 - t)^d$ ; analogous the family  $\mathcal{B}_d$  of  $d$ -bundled bucket increasing trees corresponds to the family of bucket ordered increasing trees with  $\varphi(t) = 1/(1 - t)^d$ .

Given an ordinary plane-oriented recursive tree with degree-weight generating function  $\phi(t) = 1/(1 - t)$ , we modify the clustering map  $\mathcal{C}$  defined in Algorithm 1 to a map  $\mathcal{C}^*$ , such that the corresponding bilabelled bucket ordered increasing tree is three-bundled: when redirecting edges to the buckets, we keep track of their previous ancestor. Given a resulting bucket  $v = (\ell_{\min}, \ell_{\max})$ , edges originally incident to the node with smaller label  $\ell_{\min}$  that are lying to the left of the node with larger label  $\ell_{\max}$  are grouped into the first bundle, whereas edges incident to  $\ell_{\min}$  lying to the right of  $\ell_{\max}$  are grouped into the third bundle. All edges incident to the node with larger label  $\ell_{\max}$  are put into the second bundle, the centre.



**Figure 2.** The plane-oriented recursive tree  $T$  of size eleven from Figure 1, the corresponding clustered tree with  $b = 2$ , and the corresponding bilabelled 3-bundled increasing ordered tree  $\tilde{T} = \mathcal{C}^*(T)$ .

Then, the map  $\mathcal{C}_* : \mathcal{T}^{[1]} \rightarrow \mathcal{B}_3$ , with bucket size  $b = 2$  of bucket three-bundled ordered increasing trees, is a bijection. This bijection is illustrated in Figure 2.

**Example 17** (Bijection between recursive trees and two-bundled bilabelled bucket recursive trees). We introduce two-bundled bucket recursive trees  $\mathcal{BR}_2$  with bucket size two. Each saturated node has an half-edge that separates its subtrees into two bundles. The tree family corresponds to the degree-weight generating function  $\varphi(t) = \exp(2t)$  and  $\psi_1 = 1$ . Then, we modify the clustering map  $\mathcal{C}$  again to a map  $\hat{\mathcal{C}}$ , such that the corresponding bilabelled bucket recursive tree is two-bundled: when redirecting edges to the buckets, we keep track of their previous ancestor. Given a resulting bucket  $v = (\ell_{\min}, \ell_{\max})$ , all edges originally incident to the smaller label  $\ell_{\min}$  are grouped into the first bundle, whereas all edges incident to the larger label  $\ell_{\max}$  are put into the second bundle.

Then, the map  $\hat{\mathcal{C}} : \mathcal{R} \rightarrow \mathcal{BR}_2$  from ordinary recursive trees to two-bundled bilabelled bucket recursive trees is a bijection.

Similar bijections for bucket size  $b = 2$  can be obtained also for binary increasing trees and variants.

### 5. Increasing diamonds and bucket increasing trees

#### 5.1 Symbolic description and bijections

We consider trees with bucket size  $b = 2$ , degree-weight generating function  $\varphi(t)$  and weight  $\psi = \psi_1 = 1$ . As stated in Example 8 they allow the symbolic description

$$\mathcal{T} = \mathcal{Z}^\square + \mathcal{Z}^\square * (\mathcal{Z}^\square * \varphi(\mathcal{T})). \tag{16}$$

The exponential generating function  $T(z)$  satisfies the first-order differential equation

$$T'(z) = \sqrt{1 + 2 \cdot \Phi(T(z))}, \quad T(0) = 0,$$

with  $\Phi(x) = \int_0^x \varphi(t) dt$ . Moreover,  $T = T(z)$  is implicitly given via

$$\int_0^T \frac{dx}{\sqrt{1 + 2 \cdot \Phi(x)}} = z.$$

On the other hand, increasing diamonds as proposed in [4] are defined by the symbolic equation

$$\mathcal{F} = \mathcal{Z}^\square + \mathcal{Z}^\square * \varphi(\mathcal{F}) * \mathcal{Z}^\blacksquare, \tag{17}$$



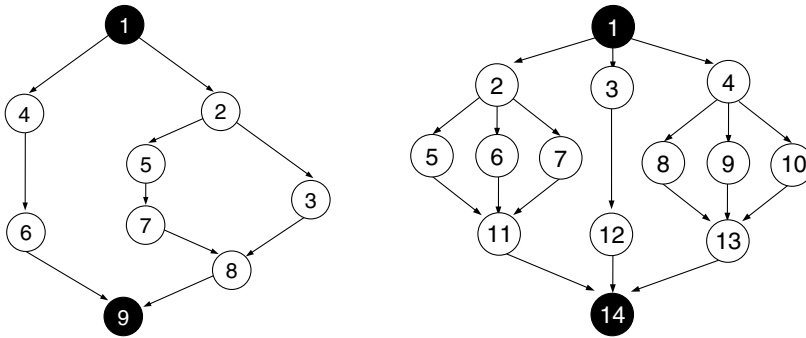


Figure 3. Two different increasing diamonds of size nine and fourteen, respectively.

where the latter boxed product constrains the largest label. They constitute a combinatorial family of labelled, directed acyclic graphs (DAGs) with a source and a sink such that the labels along any path are increasing. The functional operation  $\varphi$  occurring in (17) is here specifying possible degrees respectively their weights, see [4] for more details. Figure 3 illustrates two examples of increasing diamonds.

The symbolic description (17) of increasing diamonds leads to exactly the same exponential generating function as for bucket increasing trees with bucket size  $b = 2$  described formally via (16). Thus, it is natural to ask for a bijection between these combinatorial objects of a given size, where the size of an increasing diamond is given by the number of its nodes.

Before stating such a bijection we note that it is convenient to think of an increasing diamond  $F \in \mathcal{F}$  as having three types of nodes stemming from the recursive combinatorial construction given in (17), which partitions the set of nodes into small nodes, inner nodes and large nodes, i.e.,

$$V(F) = V_S(F) \dot{\cup} V_I(F) \dot{\cup} V_L(F).$$

Namely, if  $F$  has size one its only vertex is assigned to an inner node. Otherwise, the smallest node and the largest node of  $F$  are assigned to the respective node type; furthermore by removing these two nodes of  $F$  we obtain a (possibly empty) sequence of increasing diamonds  $F_1, \dots, F_r$ , and to each of these structures we apply this assignment recursively.

**Theorem 18.** *The family of ordered increasing diamonds  $\mathcal{F}_n$  of size  $n$ , with degree-weight generating function  $\varphi(t)$ , are in a natural bijection  $\mathcal{M}$  with ordered bucket increasing trees  $\mathcal{T}_n$  of size  $n$  with the same degree-weight generating function,  $\mathcal{F}_n \cong \mathcal{T}_n$ .*

The bijection  $\mathcal{M}$  together with the characterisation given in Theorem 12 provides a characterisation of families  $\mathcal{F}$  of increasing diamonds with a total weight  $T_n$ , resembling the weights of ordinary increasing trees and thus leading to simple counting formulas. In particular, this explains the strikingly simple formula observed in [4], which we collect in the next statement.

**Corollary 19.** *The number  $T_n$  of increasing diamonds of size  $n$  with degree-weight generating function  $\varphi(t) = 1/(1 - t)^3$  equals the number of plane-oriented recursive trees of size  $n$ . It is given by the simple formula*

$$T_n = (2n - 3)!! = \frac{(2n - 2)!!}{2^{n-1}(n - 1)!}, \quad n \geq 1.$$

*Proof of Corollary 19.* By bijection  $\mathcal{M}$  in Theorem 18 increasing diamonds  $\mathcal{F}$  with degree-weight generating function  $\varphi(t) = 1/(1 - t)^3$  are mapped to three-bundled bilabelled bucket ordered increasing trees  $\mathcal{B}_3$  with the same degree-weight generating function. Furthermore, the refined clustering map  $\mathcal{C}_*$  maps the family  $\mathcal{B}_3$  bijectively to a family of ordinary increasing

---

**Algorithm 2:** DiamondToBucket (F)

---

**Input:** Increasing Diamond  $F \in \mathcal{F}$

**Result:** Bucket size two increasing tree  $\tilde{T} \in \mathcal{T}$  of same size and with same label set

- 1:  $\hat{T} \leftarrow \text{DiamondToIncdesc}(F)$
  - 2:  $\tilde{T} \leftarrow \text{IncdescToBucket}(\hat{T})$
  - 3: Return  $\tilde{T}$
- 

trees  $\mathcal{T}^{[1]}$ , called plane-oriented recursive trees, with degree-weight generating function  $\phi(t) = 1/(1 - t)$ . Symbolically, we have

$$(C_*^{-1} \circ \mathcal{M})(\mathcal{F}) = \mathcal{T}^{[1]}.$$

The number of plane-oriented recursive trees of size  $n$  is well known [3, 5, 6, 23, 24] and obtained from the the classical generating function of the total numbers  $T(z) = 1 - \sqrt{1 - 2z}$ , satisfying

$$T'(z) = \phi(T(z)) = \frac{1}{1 - T(z)}, \quad T(0) = 0.$$

Extraction of coefficients gives

$$T_n = n![z^n]T(z) = (2n - 3)!!$$

**Proof.** Our bijection  $\mathcal{M}: \mathcal{F}_n \rightarrow \mathcal{T}_n$  uses two steps: first we construct recursively intermediate objects that we call increasing-decreasing bilabelled trees, where each bucket holds the smallest as well as the largest label in its subtree. Thus, when considering a tree  $\hat{T} \in \mathcal{ID}$  of this family and restricting to the smaller or larger labels in each bucket, it is an increasing or decreasing tree, respectively. Then we give a recursive procedure that transforms, by using cyclic permutations of the labels, such trees to bucket increasing trees of the same shape. These recursive procedures are given as Algorithms 3–4; combining them leads to Algorithm 2 and thus the map  $\mathcal{M}$ . We further observe that the weight sequence  $(\varphi_k)_{k \geq 0}$  is not involved in the bijection, as the weights are directly preserved: if a certain substructure of an increasing diamonds  $F$  decomposes into the node with smallest label, the node with largest label and  $r \geq 0$  increasing subdiamonds, then the corresponding subtree in the bucket increasing tree  $\tilde{T} = \mathcal{M}(F)$  decomposes into the root bucket and exactly  $r$  subtrees. We also note that the reverse map  $\mathcal{M}^{-1}$  is readily obtained by inverting the recursive procedures in a natural way.  $\square$

The bijection  $\mathcal{M}$  is illustrated in Figures 4–5 for the increasing diamonds given in Figure 3.

The map also gives a correspondence between quantities in increasing diamonds and bucket increasing trees. Let  $I_n$  denote the random variable counting the number of inner nodes in a random increasing diamond of size  $n$ . Moreover, let  $N_n$  denote the number of nodes with capacity one in a random bucket increasing tree of size  $n$ . Then we get the following corollary.

**Corollary 20.** *Under the bijection  $\mathcal{M}$  given in Theorem 18 the number of inner nodes in an increasing diamond  $F$  coincides with the number of nodes of capacity one in the corresponding bucket increasing tree  $\mathcal{M}(F)$ . Consequently, the r.v.  $I_n$  and  $N_n$  are equal in distribution,  $I_n \stackrel{(d)}{=} N_n$ .*

We note that a refined enumeration of increasing diamonds according to the number of inner nodes is possible in a rather direct way. Let

$$f(z, u) = \sum_{n \geq 1} \sum_{j=1}^n \frac{T_{nj}}{n!} z^n u^j = \sum_{n \geq 1} \frac{T_n \cdot \mathbb{E}(u^{I_n})}{n!} z^n$$

denote the refined generating function with  $f(z, 1) = T(z)$ .

**Algorithm 3:** DiamondToIncdesc ( $F$ )

**Input:** Increasing diamond  $F \in \mathcal{F}$   
**Result:** Increasing-decreasing bilabelled tree  $\hat{T} \in \mathcal{ID}$  of same size and with same label set

- 1:  $\hat{F} \leftarrow F$
- 2:  $n \leftarrow \text{size of } \hat{F}$
- 3: **if**  $n = 1$  **then**
- 4:     form a bucket  $\hat{b} = (v_l | )$  with  $v_l$  the single node of  $\hat{F}$
- 5:      $\hat{T} \leftarrow$  tree consisting of single bucket  $\hat{b}$
- 6:     Return  $\hat{T}$
- 7: **else**
- 8:     form a bucket  $\hat{b} = (v_S | v_L)$  with  $v_S$  and  $v_L$  the nodes with smallest and largest label of  $\hat{F}$
- 9:      $\hat{F}_1, \dots, \hat{F}_r \leftarrow$  sequence of increasing diamonds obtained by removing  $v_S$  and  $v_L$  from  $\hat{F}$
- 10:    **for**  $j = 1$  to  $r$  **do**
- 11:        $\hat{T}_j \leftarrow \text{DiamondToIncdesc}(\hat{F}_j)$
- 12:    **end for**
- 13:     $\hat{T} \leftarrow$  tree with root bucket  $\hat{b}$  and subtrees  $\hat{T}_1, \dots, \hat{T}_r$  attached to it
- 14:    Return  $\hat{T}$
- 15: **end if**

**Algorithm 4:** IncdescToBucket( $\hat{T}$ )

**Input:** Increasing-decreasing bilabelled tree  $\hat{T} \in \mathcal{ID}$   
**Result:** Bucket size 2 increasing tree  $\tilde{T} \in \mathcal{T}$  of same size, same shape and with same label set

- 1:  $\tilde{T} \leftarrow \hat{T}$
- 2:  $n \leftarrow \text{size of } \tilde{T}$
- 3: **if**  $n = 1$  **then**
- 4:     Return  $\tilde{T}$
- 5: **else**
- 6:     let  $\ell_1 < \ell_2 < \dots < \ell_n$  be the labels of  $\tilde{T}$  in increasing order
- 7:     define  $\pi \leftarrow (\ell_1)(\ell_2 \ell_3 \dots \ell_n)$  permutation in cycle notation
- 8:     permute labels of  $\tilde{T}$  according to permutation  $\pi$
- 9:      $\tilde{b} \leftarrow$  root bucket of  $\tilde{T}$
- 10:     $\hat{T}_1, \hat{T}_2, \dots, \hat{T}_r \leftarrow$  subtrees of root of  $\tilde{T}$
- 11:    **for**  $j = 1$  to  $r$  **do**
- 12:        $\tilde{T}_j \leftarrow \text{IncdescToBucket}(\hat{T}_j)$
- 13:    **end for**
- 14:     $\tilde{T} \leftarrow$  tree with root bucket  $\tilde{b}$  and subtrees  $\tilde{T}_1, \dots, \tilde{T}_r$  attached to it
- 15:    Return  $\tilde{T}$
- 16: **end if**

Then from the symbolic equation (17) and the resulting differential equation for  $f(z,u)$  one gets that  $f = f(z, u)$  is characterised implicitly via

$$\int_0^f \frac{dx}{\sqrt{u^2 + 2 \cdot \Phi(x)}} = z.$$

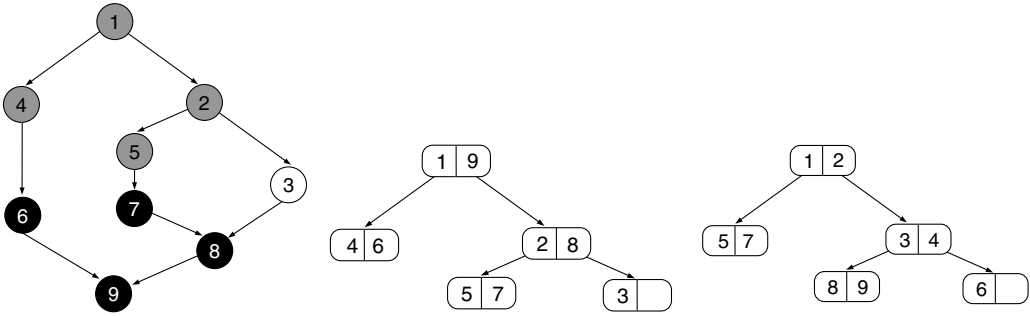


Figure 4. The increasing diamond of size nine from Figure 2 with three different label types is mapped to an increasing-decreasing bilabelled tree and then to a bucket increasing tree.

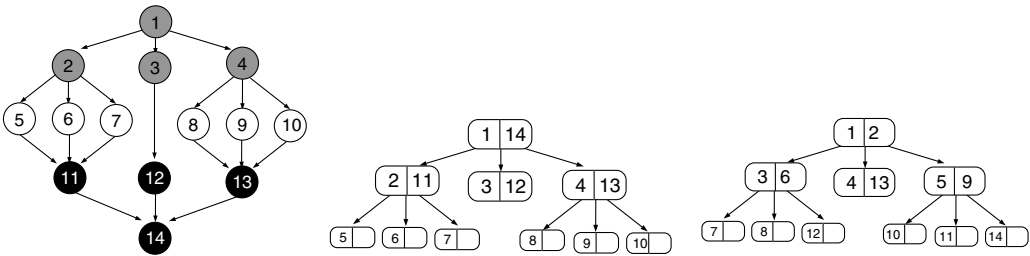


Figure 5. The increasing diamond from Figure 2 of size 14 with three different label types is mapped to an increasing-decreasing bilabelled tree and then to a bucket increasing tree.

Many interesting concrete generating functions and examples can be obtained by specialising  $\varphi(t)$ ; the authors are currently investigating into this matter.

**6. Initial bucket size of a specified element**

We consider now the random variable  $K_n$ , which denotes the size of the bucket containing element  $n$  in a random bucket increasing tree (with maximal bucket size  $b$ ) of size  $n$ . Note that by definition  $1 \leq K_n \leq b$ . As it will turn out later, the precise analysis of  $K_n$  is required in order to obtain distributional decompositions for two further r.v. of interest, the number of descendants  $Y_{n,j}$  of label  $j$  as well as the out-degree  $X_{n,j}$  node  $j$ . We derive the probability mass function of  $K_n$  using a generating function approach. Then, we relate the parameter  $K_n$  to the distribution of node types in bucket increasing trees.

**6.1. The generating functions approach**

In order to study  $K_n$  for bucket increasing trees we introduce the bivariate generating function

$$N(z, v) := \sum_{n \geq 0} \sum_{m \geq 0} \mathbb{P}\{K_{n+1} = m\} T_{n+1} \frac{z^n}{n!} v^m. \tag{18}$$

To establish a functional equation for  $N(z, v)$  from the formal recursive equation (1) it is now convenient to think of specifically bicoloured bucket increasing trees, where the colouring is as follows: exactly one element, namely the element with largest label, is coloured *red*, and all elements having a label smaller than the red element are coloured *black*. Let us first assume that the

red element of  $T$  is not contained in the root node. Then the red element is located in one of the  $r$  subtrees of the root of  $T$ ; let us assume that it is in the  $j$ -th subtree. Let us now consider these  $r$  subtrees. After order preserving relabellings, each subtree  $S_1, \dots, S_r$  is an bucket increasing tree by itself, where one of the  $r$  subtrees contains the red element. Note the obvious fact that the size of the bucket of the red element is the same in  $T$  and in the respective subtree  $S_j$ .

We introduce now generating functions, with exponential variable  $z$ , where  $z$  marks the black elements,  $f(z) = \sum_{n \geq 0} f_n \frac{z^n}{n!}$  for sequences  $f_n$  and  $f(z, \nu) = \sum_{n, m \geq 0} f_{n, m} \frac{z^n}{n!} \nu^m$  for sequences  $f_{n, m}$ , where  $\nu$  counts the initial bucket size of the red element. With this setting, the total weight of all suitably bicoloured bucket increasing trees, where the initial bucket size of the red element is exactly  $m$ , is given by  $\mathbb{P}\{K_{n+1} = m\}T_{n+1}$ , and thus its generating function is given by

$$\sum_{n \geq 0} \sum_{m \geq 0} \mathbb{P}\{K_{n+1} = m\}T_{n+1} \frac{z^n}{n!} \nu^m = N(z, \nu),$$

whereas the total weight of suitably monocoloured ordinary bucket increasing trees is  $T_n$  and its generating function is given by

$$\sum_{n \geq 1} T_n \frac{z^n}{n!} = T(z).$$

The  $r - 1$  monocoloured trees and the bicoloured bucket tree lead then to the expression  $T(z)^{r-1} \cdot N(z, \nu)$ . Since the red element can be in the first, second,  $\dots$ ,  $r$ -th subtree, we additionally get a factor  $r$ . Furthermore, the event that the root has out-degree  $r$  leads to a factor  $\varphi_r$ . Summing over all  $r \geq 1$  leads thus to  $\sum_{r \geq 1} r \varphi_r T(z)^{r-1} N(z, \nu) = \varphi'(T(z))N(z, \nu)$ . Since the elements labelled by  $1, 2, \dots, b$  contained in the root node are all coloured black (which again means that  $b$  elements in a labelled object are fixed), equation (1) leads thus to the following differential equation of order  $b$  for  $N(z, \nu)$ :

$$\frac{\partial^b}{\partial z^b} N(z, \nu) = \varphi'(T(z)) \cdot N(z, \nu). \tag{19}$$

The cases, where the red element is contained in the root of the tree do not appear explicitly in the differential equation itself, but will be described by the initial conditions. Since  $\mathbb{P}\{K_n = n\} = 1$ , for  $1 \leq n \leq b$  (if element  $n$  is contained in the root node then all elements with a label  $\leq n$  are also contained in the root node), we obtain the following initial conditions, for  $0 \leq \ell \leq b - 1$ :

$$\left. \frac{\partial^\ell}{\partial z^\ell} N(z, \nu) \right|_{z=0} = \sum_{m \geq 0} \mathbb{P}\{K_{\ell+1} = m\}T_{\ell+1} \nu^m = T_{\ell+1} \nu^{\ell+1},$$

with  $T_n = n![z^n]T(z)$  and  $T(z)$  as characterised in Proposition 12 for the particular tree families.

Now we can specify the sequences according to the tree family of interest. Note that for bucket recursive trees the initial bucket size of node  $n$  was already (implicitly) characterised in [13], hence we will skip this case. For (b,d)-ary increasing trees we obtain the following Cauchy-Euler type differential equation together with the initial conditions for the bivariate generating function  $N(z, \nu)$ :

$$\frac{\partial^b}{\partial z^b} N(z, \nu) = \frac{T_{b+1}}{(1 - (d - 1)z)^b} N(z, \nu), \quad \left. \frac{\partial^\ell}{\partial z^\ell} N(z, \nu) \right|_{z=0} = T_{\ell+1} \nu^{\ell+1}, \quad \text{for } 0 \leq \ell \leq b - 1. \tag{20}$$

For  $(b, \alpha)$ -plane oriented recursive trees we obtain a very similar Cauchy-Euler type differential equation together with the initial conditions for the bivariate generating function  $N(z, v)$ :

$$\frac{\partial^b}{\partial z^b} N(z, v) = \frac{T_{b+1}}{(1 - (\alpha + 1)z)^b} N(z, v), \quad \frac{\partial^\ell}{\partial z^\ell} N(z, v) \Big|_{z=0} = T_{\ell+1} v^{\ell+1}, \quad \text{for } 0 \leq \ell \leq b - 1. \tag{21}$$

**6.2. The distribution of the initial bucket size**

In order to obtain the exact distribution of the r.v.  $K_n$  we will give the exact solution of the homogeneous differential equations (20), (21), which are of Cauchy-Euler-type. Plugging in the Ansatz  $N(z, v) = \frac{1}{(1 - (d-1)z)^\lambda}$  for  $(b, d)$ -ary increasing trees, and  $N(z, v) = \frac{1}{(1 - (\alpha+1)z)^\lambda}$  for  $(b, \alpha)$ -plane oriented recursive trees, with unspecified  $\lambda$ , into equations (20), (21) leads to the indicial equation

$$\lambda^{\bar{b}} = \begin{cases} \frac{T_{b+1}}{(d-1)^b}, \\ \frac{T_{b+1}}{(\alpha+1)^b}, \end{cases} \quad \text{or equivalently} \quad \binom{\lambda + b - 1}{b} = \begin{cases} \binom{b + \frac{1}{d-1}}{b}, & (b, d)\text{-ary ITs,} \\ \binom{b - \frac{1}{\alpha+1}}{b}, & (b, \alpha)\text{-PORTs.} \end{cases} \tag{22}$$

Similar equations have been studied in [18] for bucket recursive trees. It is convenient to give a unified analysis of the indicial equations for all three models. Let

$$\kappa = \begin{cases} 0, & \text{bucket recursive trees,} \\ \frac{1}{d-1}, & (b, d)\text{-ary ITs,} \\ -\frac{1}{\alpha+1}, & (b, \alpha)\text{-PORTs.} \end{cases} \tag{23}$$

Then, the indicial equations can be written in a unified way:

$$\lambda^{\bar{b}} = (b + \kappa)^{\bar{b}}. \tag{24}$$

Equations of this or similar kind have been treated in [13, 17, 18]. It follows from these considerations that, for  $\kappa$  given in (23), all solutions  $\lambda_1, \lambda_2, \dots, \lambda_b$  of (24) are simple, and when arranging them in descending order of real parts it further holds  $1 + \frac{1}{d-1} = \lambda_1 > \Re(\lambda_2) \geq \Re(\lambda_3) \geq \dots \geq \Re(\lambda_b)$  for  $(b, d)$ -ary increasing trees and  $1 - \frac{1}{\alpha+1} = \lambda_1 > \Re(\lambda_2) \geq \Re(\lambda_3) \geq \dots \geq \Re(\lambda_b)$  for  $(b, \alpha)$ -plane oriented recursive trees. Thus the general solutions of (20), (21) are given by

$$N(z, v) = \begin{cases} \sum_{i=1}^b \frac{\beta_i(v)}{(1 - (d-1)z)^{\lambda_i}}, & (b, d)\text{-ary ITs,} \\ \sum_{i=1}^b \frac{\beta_i(v)}{(1 - (\alpha+1)z)^{\lambda_i}}, & (b, \alpha)\text{-PORTs,} \end{cases} \tag{25}$$

with certain functions  $\beta_i(u, v)$ , which are specified by the initial conditions as given in (20), (21). When these initial conditions are plugged into (25) this leads to the following system of linear equations for the unknown functions  $\beta_i(v)$ ,  $1 \leq i \leq b$ :

$$\sum_{i=1}^b \lambda_i \bar{\nu} \beta_i(\nu) = \begin{cases} \nu^{\ell+1} \ell! \binom{\ell + \frac{1}{d-1}}{\ell}, & \text{(b,d)-ary ITs,} \\ \nu^{\ell+1} \ell! \binom{\ell - \frac{1}{\alpha+1}}{\ell}, & \text{(b, \alpha)-PORTs.} \end{cases}$$

Using the abbreviations

$$s_\ell := s_\ell(\nu) := \begin{cases} \nu^{\ell+1} \binom{\ell + \frac{1}{d-1}}{\ell}, & \text{(b,d)-ary ITs,} \\ \nu^{\ell+1} \binom{\ell - \frac{1}{\alpha+1}}{\ell}, & \text{(b, \alpha)-PORTs,} \end{cases} \tag{26}$$

we obtain the following system of linear equations for the unknown  $\beta_i = \beta_i(\nu)$ ,  $1 \leq i \leq b$ :

$$\sum_{i=1}^b \binom{\lambda_i + \ell - 1}{\ell} \beta_i = s_\ell, \quad \text{for } 0 \leq \ell \leq b - 1. \tag{27}$$

To obtain the solution of (27) we can use results and respective computations given in [13, equations (22) and (30)], slightly adapted to  $\kappa$  occurring in the indicial equation (24) of the tree families considered.

**Lemma 21** ([13]). *For given numbers  $\lambda_i$ , with  $1 \leq i \leq b$ , specified as the solutions of equation (24), and numbers  $s_\ell$ ,  $0 \leq \ell \leq b - 1$ , the system of linear equations with unknowns  $\beta_i$ ,*

$$\sum_{i=1}^b \binom{\lambda_i + \ell - 1}{\ell - 1} \beta_i = s_\ell, \quad 0 \leq \ell \leq b - 1,$$

has the exact solution

$$\beta_i = \begin{cases} \sum_{r=0}^{b-1} s_r \frac{\binom{\lambda_i+b-1}{b-r-1}}{\binom{b}{r}(b-r)\binom{b+\frac{1}{d-1}}{b} (H_{\lambda_i+b-1} - H_{\lambda_i-1})}, & 1 \leq i \leq b, \quad \text{(b,d)-ary ITs,} \\ \sum_{r=0}^{b-1} s_r \frac{\binom{\lambda_i+b-1}{b-r-1}}{\binom{b}{r}(b-r)\binom{b-\frac{1}{\alpha+1}}{b} (H_{\lambda_i+b-1} - H_{\lambda_i-1})}, & 1 \leq i \leq b, \quad \text{(b, \alpha)-PORTs.} \end{cases}$$

An application of Lemma 21 immediately gives the values  $\beta_i(\nu)$  with respect to the initial conditions (26). Thus extracting coefficients yields

$$[z^{n-1} \nu^m] N(z, \nu) = \mathbb{P}\{K_n = m\} \frac{T_n}{(n-1)!} = \begin{cases} \sum_{i=1}^b \binom{\lambda_i + n - 2}{n-1} (d-1)^{n-1} [\nu^m] \beta_i(\nu), & \text{(b,d)-ary ITs,} \\ \sum_{i=1}^b \binom{\lambda_i + n - 2}{n-1} (\alpha+1)^{n-1} [\nu^m] \beta_i(\nu), & \text{(b, \alpha)-PORTs,} \end{cases}$$



and by specifying  $T_n$  as given in Theorem 12 and  $s_\ell(v)$  as given in (26) we obtain the exact distribution of  $K_n$  stated in the following theorem.

**Theorem 22.** *The probability mass function of the random variable  $K_n$  counting the initial bucket size of node  $n$  in a random bucket tree of size  $n$  is given by the following closed formula.*

$$\mathbb{P}\{K_n = m\} = \begin{cases} \sum_{i=1}^b \frac{(\lambda_i+n-2) \binom{\lambda_i+b-1}{b-m} \binom{m-1+\frac{1}{d-1}}{m-1}}{\binom{n-1+\frac{1}{d-1}}{n-1} \binom{b}{m-1} (b-m+1) \binom{b+\frac{1}{d-1}}{b} (H_{\lambda_i+b-1} - H_{\lambda_i-1})}, & \text{for } (b,d)\text{-ary ITs,} \\ \sum_{i=1}^b \frac{(\lambda_i+n-2) \binom{\lambda_i+b-1}{b-m} \binom{m-1-\frac{1}{\alpha+1}}{m-1}}{\binom{n-1-\frac{1}{\alpha+1}}{n-1} \binom{b}{m-1} (b-m+1) \binom{b-\frac{1}{\alpha+1}}{b} (H_{\lambda_i+b-1} - H_{\lambda_i-1})}, & \text{for } (b, \alpha)\text{-PORTs,} \end{cases}$$

for  $1 \leq m \leq b$ .

For  $n$  tending to infinity the random variable  $K_n$  converges in distribution to a limit  $K$ , whose discrete distribution is given as follows.

$$\mathbb{P}\{K = m\} = \begin{cases} \frac{\binom{b+\frac{1}{d-1}}{b-m} \binom{m-1+\frac{1}{d-1}}{m-1}}{\binom{b}{m-1} (b-m+1) \binom{b+\frac{1}{d-1}}{b} (H_{b+\frac{1}{d-1}} - H_{\frac{1}{d-1}})}, & (b, d)\text{-ary ITs,} \\ \frac{\binom{b-\frac{1}{\alpha+1}}{b-m} \binom{m-1-\frac{1}{\alpha+1}}{m-1}}{\binom{b}{m-1} (b-m+1) \binom{b-\frac{1}{\alpha+1}}{b} (H_{b-\frac{1}{\alpha+1}} - H_{-\frac{1}{\alpha+1}})}, & (b, \alpha)\text{-PORTs,} \end{cases}$$

for  $1 \leq m \leq b$ .

**Remark 23.** The corresponding results for bucket recursive trees were already derived in [13] (although not stated explicitly):

$$\mathbb{P}\{K_n = m\} = \sum_{i=1}^b \frac{\binom{\lambda_i+b-1}{b-m} \binom{\lambda_i+n-2}{n-1}}{\binom{b}{m-1} (b-m+1) (H_{\lambda_i+b-1} - H_{\lambda_i-1})},$$

for  $1 \leq m \leq b$ . Furthermore, for bucket recursive trees the random variable  $K_n$  converges, for  $n \rightarrow \infty$ , in distribution to a Zipf-distributed limit  $K$ :

$$\mathbb{P}\{K = m\} = \frac{1}{mH_b}, \quad 1 \leq m \leq b.$$

**Proof.** It remains to show the stated limiting distribution results. We apply Stirling’s formula for the Gamma-function

$$\Gamma(z) = \left(\frac{z}{e}\right)^z \frac{\sqrt{2\pi}}{\sqrt{z}} \left(1 + \mathcal{O}\left(\frac{1}{z}\right)\right),$$

and get

$$\binom{n-1+\frac{1}{d-1}}{n-1} = \frac{\Gamma\left(n+\frac{1}{d-1}\right)}{\Gamma\left(1+\frac{1}{d-1}\right)\Gamma(n)} = \frac{n^{\frac{1}{d-1}}}{\Gamma\left(1+\frac{1}{d-1}\right)} \left(1 + \mathcal{O}\left(n^{-1}\right)\right),$$

as well as

$$\binom{n-1-\frac{1}{\alpha+1}}{n-1} = \frac{\Gamma\left(n-\frac{1}{\alpha+1}\right)}{\Gamma\left(1-\frac{1}{\alpha+1}\right)\Gamma(n)} = \frac{n^{-\frac{1}{\alpha+1}}}{\Gamma\left(1-\frac{1}{\alpha+1}\right)} (1 + \mathcal{O}(n^{-1})).$$

Concerning the roots of the indicial equations recall from remarks stated past equation (24) that

$$\lambda_1 = \begin{cases} 1 + \frac{1}{d-1}, & \text{(b,d)-ary ITs,} \\ 1 - \frac{1}{\alpha+1}, & \text{(b, } \alpha\text{)-PORTs.} \end{cases}$$

For (b,d)-ary increasing trees we obtain the following asymptotic expansions:

$$\binom{n-2+\lambda_i}{n-1} = \frac{n^{\lambda_i-1}}{\Gamma(\lambda_i)} (1 + \mathcal{O}(n^{-1})) = \begin{cases} \frac{n^{\frac{1}{d-1}}}{\Gamma\left(1+\frac{1}{d-1}\right)} (1 + \mathcal{O}(n^{-1})), & i = 1, \\ \mathcal{O}\left(n^{\lambda_i-1}\right), & 2 \leq i \leq b. \end{cases}$$

For (b,  $\alpha$ )-plane oriented recursive trees we have the corresponding results

$$\binom{n-2+\lambda_i}{n-1} = \begin{cases} \frac{n^{-\frac{1}{\alpha+1}}}{\Gamma\left(1-\frac{1}{\alpha+1}\right)} (1 + \mathcal{O}(n^{-1})), & i = 1, \\ \mathcal{O}\left(n^{\lambda_i-1}\right), & 2 \leq i \leq b. \end{cases}$$

Thus, for  $n \rightarrow \infty$ , the dominant contribution in the asymptotic expansions of  $\mathbb{P}\{K_n = m\}$  and the finite sum stems from the index  $i = 1$  and  $\lambda_1$ , leading to the stated result.  $\square$

**6.3 Relation to node types in bucket increasing trees**

Let  $N_{n,j}$  denote the random variable counting the number of nodes with capacity  $c(v) = j, 1 \leq j \leq b$ . Furthermore, let  $\mathbf{N}_n$  denote the random vector  $(N_{n,1}, \dots, N_{n,b})$ . Mahmoud and Smythe [18] considered bucket recursive trees. They proved a multivariate central limit theorem for  $\mathbf{N}_n$  for trees with bucket size  $b \leq 26$ . For trees with  $b > 26$  a phase change in the limiting distribution of  $\mathbf{N}_n$  was detected and the central limit theorem does not hold anymore. In the following we are going to analyse the limiting distribution of the random vector  $\mathbf{N}_n$  for (b,d)-ary increasing trees and (b,  $\alpha$ )-plane oriented recursive trees. We also summarise the main results for bucket recursive trees. There is a close connection between the random variables  $K_n$  and  $\mathbf{N}_n$ . The distribution of the initial bucket size  $K_{n+1}$  depends on the different node types present at time  $n, n \geq 1$ . Let  $v_k = v_{n,m,k}$  denote the buckets of capacity  $m$  contributing to  $N_{n,m}$ , with  $1 \leq m \leq b$  and  $1 \leq k \leq N_{n,j}$ . Then, by definition of the growth processes

$$\mathbb{P}\{K_{n+1} = 1 \mid \mathbf{N}_n\} = \sum_{k=1}^{N_{n,b}} \mathbb{P}\{n+1 <_t v_{n,b,k}\}$$

and, for  $2 \leq m \leq b$ ,

$$\mathbb{P}\{K_{n+1} = m \mid \mathbf{N}_n\} = \sum_{k=1}^{N_{n,m-1}} \mathbb{P}\{n+1 <_t v_{n,m-1,k}\}.$$

Of course, for each  $k$ , the probabilities  $\mathbb{P}\{n + 1 <_t v_{n,b,k}\}$  coincide and are given according to Definitions 1–3:

$$\mathbb{P}\{n + 1 <_t v_{n,m,k}\} = \begin{cases} \frac{m}{n}, & \text{bucket recursive trees,} \\ \frac{(d - 1)m + 1 - \text{deg}^+(v)}{(d - 1)n + 1}, & \text{(b,d)-ary ITs,} \\ \frac{\text{deg}^+(v) + (\alpha + 1)m - 1}{(\alpha + 1)n - 1}, & \text{(b, } \alpha\text{)-PORTs.} \end{cases}$$

Note that for  $1 \leq m \leq b - 1$  the individual nodes  $v_k = v_{n,m,k}$  are unsaturated, such that  $\text{deg}^+(v_{n,m,k}) = 0$ . Consequently, summation leads to the following result.

**Proposition 24.** *The random variable  $K_n$  counting the initial size of the bucket containing label  $n$  is related to the number of nodes  $N_n = (N_{n,1}, \dots, N_{n,b})$  with respective capacities as follows. For  $2 \leq m \leq b$  it holds*

$$\mathbb{P}\{K_{n+1} = m\} = \mathbb{E}(N_{n,m-1}) \cdot \begin{cases} \frac{m - 1}{n}, & \text{bucket recursive trees,} \\ \frac{(d - 1)(m - 1) + 1}{(d - 1)n + 1}, & \text{(b,d)-ary ITs,} \\ \frac{(\alpha + 1)(m - 1) - 1}{(\alpha + 1)n - 1}, & \text{(b, } \alpha\text{)-PORTs.} \end{cases}$$

and

$$\mathbb{P}\{K_{n+1} = 1\} = \begin{cases} \mathbb{E}(N_{n,b}) \cdot \frac{b}{n}, & \text{bucket recursive trees,} \\ \mathbb{E}(N_{n,b}) \cdot \frac{(d - 1)b + 1}{(d - 1)n + 1} + \frac{1 - \sum_{j=1}^b \mathbb{E}(N_{n,j})}{(d - 1)b + 1}, & \text{(b,d)-ary ITs,} \\ \mathbb{E}(N_{n,b}) \cdot \frac{(\alpha + 1)b - 1}{(\alpha + 1)n - 1} + \frac{-1 + \sum_{j=1}^b \mathbb{E}(N_{n,j})}{(\alpha + 1)n - 1}, & \text{(b, } \alpha\text{)-PORTs.} \end{cases}$$

In the following we consider generalised Pólya-Eggenberger urn models (see, e.g., [7]) with  $b$  different types of balls. For bucket recursive trees Mahmoud and Smythe [18] studied the following urn model.

**Urn 25** (Bucket recursive trees) Consider a balanced urn with balls of  $b$  colours and let  $Q_{n,m}$  denote the number of balls of type  $m$ ,  $1 \leq m \leq b$ , in the urn after  $n$  draws.  $(Q_{n,1}, \dots, Q_{n,b})$  denotes the corresponding random vector at time  $n$ . At each time step, draw one ball at random from the urn, observe its colour, and add balls according to the ball replacement matrix

$$M = \begin{pmatrix} -1 & 2 & 0 & \dots & 0 \\ 0 & -2 & 3 & \dots & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & \dots & 0 & -(b - 1) & b \\ 1 & 0 & 0 & \dots & 0 \end{pmatrix}$$

and initial composition a single ball of type one. Then, the random variables  $Q_{n,m}$  are related to the node types via

$$N_{n,m} = \frac{Q_{n,m}}{m}, \quad 1 \leq m \leq b.$$

The characteristic polynomial of  $M$  is given

$$\chi_M(\lambda) = \det(M - \lambda I) = (-1)^b (\lambda^{\bar{b}} - b!). \tag{28}$$

For  $b \leq 26$  and  $b > 26$ , respectively, there occurs a phase change: the limit law changes from normal to non-normal, which is due to the structure of the characteristic polynomial and the general results revealed in [7]. We assume that the eigenvalues  $\lambda_1, \dots, \lambda_b$  are indexed according to their real parts

$$1 = \Re \lambda_1 \geq \Re \lambda_2 \geq \dots \geq \Re \lambda_b.$$

If the real part of  $\lambda_2$  exceeds  $\frac{1}{2}$  then the limit law changes. In the following we denote the random vector of ball types as  $\mathbf{Q}_n = (Q_{n,1}, \dots, Q_{n,b})$ .

**Theorem 26** (Mahmoud and Smythe [18]). *For  $b \leq 26$  the limit law of  $\frac{1}{\sqrt{n}}(\mathbf{Q}_n - \mathbb{E}(\mathbf{Q}_n))$  is asymptotically normal. For  $b > 26$  there is no normal limit law (under the same normalisation).*

Here we introduce two new urn models.

**Urn 27** ( $(b, \alpha)$ -plane oriented recursive trees) Consider the urn with ball replacement matrix

$$M = \begin{pmatrix} -\alpha & 2\alpha + 1 & 0 & \dots & 0 \\ 0 & -(2\alpha + 1) & 3\alpha + 2 & \dots & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & \dots & 0 & -((b-1)\alpha + b - 2) & b\alpha + b - 1 \\ \alpha & 0 & 0 & \dots & 1 \end{pmatrix}$$

and initial composition  $\alpha$  balls of type one. Then, the random variables  $Q_{n,m}$  are related to the node types by

$$N_{n,m} = \frac{Q_{n,m}}{m\alpha + m - 1}, \quad 1 \leq m \leq b.$$

The characteristic polynomial of  $M$  is given

$$\chi_M(\lambda) = \det(M - \lambda I) = (-1)^b (\alpha + 1)^b \left( \left( \frac{\lambda - 1}{\alpha + 1} \right)^{\bar{b}} - \left( \frac{\alpha}{\alpha + 1} \right)^{\bar{b}} \right). \tag{29}$$

**Urn 28** ( $(b,d)$ -ary increasing trees) Consider the urn with ball replacement matrix

$$M = \begin{pmatrix} -d & 2d - 1 & 0 & \dots & 0 \\ 0 & -(2d - 1) & 3d - 2 & \dots & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & \dots & 0 & -((b-1)d - (b-2)) & bd - (b-1) \\ d & 0 & 0 & \dots & -1 \end{pmatrix}$$

and initial composition  $d$  balls of type one. Then, the random variables  $Q_{n,m}$  are related to the node types by

$$N_{n,m} = \frac{Q_{n,m}}{md - (m - 1)}, \quad 1 \leq m \leq b.$$

The characteristic polynomial of  $M$  is given

$$\chi_M(\lambda) = \det(M - \lambda I) = (-1)^b (d - 1)^b \left( \left( \frac{\lambda + 1}{d - 1} \right)^{\bar{b}} - \left( \frac{d}{d - 1} \right)^{\bar{b}} \right). \tag{30}$$

As indicated by the connection between  $K_n$  and  $\mathbf{N}_n$ , there is also a close connection between the characteristic polynomials (28), (29), (30) and the indicial equation (24):

$$\lambda^{\bar{b}} = (b + \kappa)^{\bar{b}}.$$

In particular, there is an affine transformation between these two polynomials. We note in passing that the general theorems of Janson [7] and Müller [20] allow to describe a phase change in the limit laws for  $\mathbf{N}_n$  similar to the results of Mahmoud and Smythe [18].

### 7. Applications

In the following we present a few applications of the results obtained in Section 6. The stochastic growth rule discussed in Subsection 2.1 and the analysis of the initial bucket size  $K_n$  can be used to analyse several parameters. We consider in the following the random variable  $Y_{n,j}$ , which counts the number of descendants of element  $j$ , i.e., the total number of elements with a label greater or equal  $j$  contained in the subtree rooted with the bucket containing element  $j$ , in a random bucket increasing tree (with maximal bucket size  $b$ ) of size  $n$ . For this random variable we provide the exact distribution, as well as a decomposition of the random variable of interest in terms of the initial bucket size  $K_n$ . Then we apply our results to the root degree as well as the out-degree  $X_{n,j}$  of the node containing element  $j$ . There we also present a decomposition of the random variable of interest, complementing earlier results.

#### 7.1 Descendants in bucket increasing trees

In order to avoid degeneracy we assume that  $j \geq b + 1$  (for  $1 \leq j \leq b$  it holds  $Y_{n,j} = n + 1 - j$ ). Explicit results for the probability mass functions and the moments of both r.v.  $Y_{n,j}$  and  $X_{n,j}$  can be obtained in principle by purely combinatorial means and a generating functions approach similar to the parameter initial bucket size treated above. However, here we take a different point of view utilising the previous results for the initial bucket size to provide concise decompositions of the random variables based on results already known in the literature. Such decompositions readily lead to limit laws and seem to be more difficult to obtain by using a purely analytic combinatorial approach.

In order to obtain to analyse  $Y_{n,j}$ , we introduce a refinement of this r.v.: let  $Y_{n,\ell,j} = Y_{n,j} \mid K_j = \ell$  denote  $Y_{n,j}$  conditioned on the event  $K_j = \ell$ . According to the stochastic growth processes defined in Subsection 2.1 we obtain the recurrence relation

$$\mathbb{P}\{Y_{n+1,\ell,j} = m\} = \frac{c_1(m + \ell - 2) + c_2}{c_1 n + c_2} \mathbb{P}\{Y_{n,\ell,j} = m - 1\} + \frac{c_1(n + 1 - m - \ell)}{c_1 n + c_2} \mathbb{P}\{Y_{n,\ell,j} = m\},$$

for  $m \geq 1$  and the initial value  $Y_{j,\ell,j} = 1$ . The parameters  $c_1, c_2$  occurring in this description of the law of  $Y_{n,\ell,j}$  are determined by the fraction  $\frac{c_2}{c_1} = \kappa$  as given in (23). Alternatively,  $Y_{n,\ell,j}$  can be described as follows.

**Proposition 29.** *The r.v.  $Y_{n,\ell,j}$  can be described by a sum of dependent random variables  $A_{i,\ell,j}, i \geq j$ , all taking values in  $\{0, 1\}$ , and initial value  $A_{j,\ell,j} = 1$ , where  $A_{i,\ell,j}$  denotes the indicator variable of the event that label  $i$  is a descendant of label  $j$  conditioned on the event  $K_j = \ell$ , and we get*

$$Y_{n,\ell,j} = \sum_{i=j}^n A_{i,\ell,j}, \quad \mathbb{P}\{A_{i+1,\ell,j} = 1 \mid Y_{i,\ell,j}\} = \frac{c_1(\ell - 1 + Y_{i,\ell,j}) + c_2}{c_1 i + c_2}. \tag{31}$$

**Proof.** It suffices to show that the probabilities  $\mathbb{P}\{A_{i+1,\ell,j} = 1 \mid Y_{i,\ell,j}\}$  are indeed given as stated above. Given  $K_j$  assume that node  $v$  containing label  $j$  has  $m \geq 1$  descendants at time  $i \geq j$ . If  $m < b + 1 - K_j$  then node  $v$  has out-degree zero and capacity  $c(v) = K_j + m - 1$ . It attracts a new label  $i + 1$  with probability  $p(v) = p_{i+1}(v)$  determined directly by the stochastic growth rules:

$$p(v) = \begin{cases} \frac{K_j + m - 1}{i}, & \text{bucket recursive trees,} \\ \frac{(d - 1)(K_j + m - 1) + 1}{(d - 1)i + 1}, & \text{(b,d)-ary ITs,} \\ \frac{(\alpha + 1)(K_j + m - 1) - 1}{(\alpha + 1)i - 1}, & \text{(b, } \alpha\text{)-PORTs.} \end{cases}$$

Otherwise, assume that the number of descendants is given by  $m \geq b + 1 - K_j$ . Thus, node  $v$  is saturated and the remaining  $m - (b + 1 - K_j)$  labels are distributed amongst the  $r$  non-root nodes  $u_1, \dots, u_r$  in the subtree rooted at  $v$ . The probability  $\mathbb{P}\{i + 1 <_d j\}$  that label  $i + 1$  is a descendant of label  $j$  is given by

$$\mathbb{P}\{i + 1 <_d j\} = \mathbb{P}\{i + 1 <_c v\} + \sum_{k=1}^r \mathbb{P}\{i + 1 <_c u_k\},$$

where  $\mathbb{P}\{i + 1 <_c x\}$  denotes the probability that label  $i + 1$  is contained in the node  $x$ . We note that

$$\sum_{k=1}^r c(u_k) = m - (b + 1 - K_j), \quad \text{deg}^+(v) + \sum_{k=1}^r \text{deg}^+(u_k) = r.$$

For bucket recursive trees we obtain

$$\mathbb{P}\{i + 1 <_d j\} = \frac{b}{i} + \sum_{k=1}^r \frac{c(u_k)}{i} = \frac{b + m - (b + 1 - K_j)}{i} = \frac{K_j + m - 1}{i}.$$

For (b,d)-ary increasing trees we have

$$\begin{aligned} \mathbb{P}\{i + 1 <_d j\} &= \frac{(d - 1)c(v) + 1 - \text{deg}^+(v)}{(d - 1)i + 1} + \sum_{k=1}^r \frac{(d - 1)c(u_k) + 1 - \text{deg}^+(u_k)}{(d - 1)i + 1} \\ &= \frac{(d - 1)b + r + 1 + (d - 1)(m - (b + 1 - K_j)) - r}{(d - 1)i + 1} = \frac{(d - 1)(K_j + m - 1) + 1}{(d - 1)i + 1}. \end{aligned}$$

Finally, for  $(b, \alpha)$ -plane oriented recursive trees we get

$$\begin{aligned} \mathbb{P}\{i + 1 <_d j\} &= \frac{\text{deg}^+(v) + (\alpha + 1)c(v) - 1}{(\alpha + 1)i - 1} + \sum_{k=1}^r \frac{\text{deg}^+(u_k) + (\alpha + 1)c(u_k) - 1}{(\alpha + 1)i - 1} \\ &= \frac{r + (\alpha + 1)(b + m - (b + 1 - K_j)) - r - 1}{(\alpha + 1)i - 1} = \frac{(\alpha + 1)(K_j + m - 1) - 1}{(\alpha + 1)i - 1}. \end{aligned}$$

Summarising we obtain the probabilities

$$\mathbb{P}\{i + 1 <_d j \mid Y_{i,\ell,j} = m\} = \begin{cases} \frac{\ell - 1 + m}{i}, & \text{bucket recursive trees,} \\ \frac{(d - 1)(\ell - 1 + m) + 1}{(d - 1)i + 1}, & \text{(b,d)-ary ITs,} \\ \frac{(\alpha + 1)(\ell - 1 + m) - 1}{(\alpha + 1)i - 1}, & \text{(b, \alpha)-PORTs,} \end{cases}$$

thus leading to the decomposition stated in (31). □

We remark in passing that, alternatively,  $Y_{n,\ell,j}$  can also be described via a generalised Pólya urn model; see [16].

It is a key observation that the distribution of  $Y_{n,\ell,j}$  is identical to the distribution of a r.v.  $D_{n,\ell,j}$ , counting so-called generalised descendants  $D_{n,\ell,j}$  in ordinary families of increasing trees, which has been introduced and studied in [12]. Namely, the r.v.  $D_{n,\ell,j}$  also admits a description as a sum of dependent r.v. equivalent to (31), see [12, equations (7)–(8)], thus  $Y_{n,\ell,j} \stackrel{(d)}{=} D_{n,\ell,j}$ . Unconditioning immediately leads to the following result.

**Proposition 30** (Decomposition of the number of descendants). *The random variable  $Y_{n,j}$  counting the number of descendants of label  $j$  in a bucket increasing tree of size  $n$  for the families bucket recursive trees,  $(b,d)$ -ary increasing trees, and  $(b, \alpha)$ -plane oriented recursive trees is related to the random variable  $D_{n,\ell,j}$ , counting generalised descendants with parameter  $\ell$  for the corresponding ordinary increasing tree families as studied in [12], where the parameter  $\ell$  is given by the random variable  $K_j$  measuring the initial bucket size of label  $j$ :*

$$Y_{n,j} \stackrel{(d)}{=} D_{n,K_j,j}.$$

The probability mass function of  $D_{n,\ell,j}$  as well as limit laws have been obtained in [12] using lattice path counting arguments. Using them this easily leads to the following limiting distribution result of  $Y_{n,j}$ , also slightly refining the results of [13] for bucket recursive trees. We thus omit the details.

**Corollary 31** (Limit laws for the number of descendants). *For  $n \rightarrow \infty$  and  $j = j(n)$ , the random variable  $Y_{n,j}$  has the following limit laws, depending on the random initial bucket size  $K_j$  or its limit  $K$ .*

1. *The region for  $j$  fixed. The normalised random variable  $Y_{n,j}$  is asymptotically Beta-distributed,  $Y_{n,j}/n \xrightarrow{(d)} \beta(K_j + \frac{c_2}{c_1}, j - K_j)$ .*
2. *The region for small  $j$ :  $j \rightarrow \infty$  such that  $j = o(n)$ . The normalised random variable  $jY_{n,j}/n$  is asymptotically Gamma-distributed,  $jY_{n,j}/n \xrightarrow{(d)} \gamma(K + \frac{c_2}{c_1}, 1)$ .*



3. The central region for  $j: j \rightarrow \infty$  such that  $j \sim \rho n$ , with  $0 < \rho < 1$ . The shifted random variable  $Y_{n,j} - 1$  is asymptotically negative binomial-distributed,  $Y_{n,j} - 1 \xrightarrow{(d)} \text{NegBin}(K + \frac{c_2}{c_1}, \rho)$ .
4. The region for large  $j: j \rightarrow \infty$  such that  $k := n - j = o(n)$ . The random variable  $Y_{n,j}$  has asymptotically all its mass concentrated at 1,  $\mathbb{P}\{Y_{n,j} = 1\} = 1 + \mathcal{O}(\frac{k}{n})$ .

**Remark 32.** We mention the possibility to improve the distributional convergence of  $Y_{n,j}$  in several ways. First, one can prove moment convergence. Second, local limit theorems can be deduced using the explicit expressions for the probability mass functions of  $K_j$  and  $D_{n,\ell,j}$ . Third, for fixed  $j$  one can measure the difference between  $Y_{n,j}/n$  and the limiting beta random variable in terms of a so-called martingale tail sum using discrete martingales.

**7.2 Node degrees in bucket increasing trees**

Let  $X_{n,j}$  denote the random variable counting the out-degree of the bucket containing label  $j$  in a size  $n$  random bucket increasing tree. By definition, we can decompose  $X_{n,j}$  into a sequence of dependent indicator variables

$$X_{n,j} = \sum_{\ell=j+1}^n \mathbb{I}\{\ell <_a j\},$$

where  $\{\ell <_a j\}$  denotes the event that label  $\ell$  is attached as a new node to the node containing label  $j$ . Similar to the case  $b = 1$ , see for example [11], the random variable  $X_{n,j}$  does not obey a uniform behaviour, but it is similar to the r.v. number of descendants considered before. Basically, we will observe that for  $b > 1$  the random variable  $X_{n,j}$  behaves similar to the case  $b = 1$  once the bucket containing the label  $j$  is fully saturated. Compared to the two “stages” of  $Y_{n,j}$ , insertion of label  $j$  into a node of size  $K_j$  and then attraction of new labels, we have three stages: first, label  $j$  is inserted into a node  $v$  of size  $K_j$ ; second, node  $v$  attracts new labels until it is fully saturated; then, until time  $n$  the node  $v$  attracts new labels and its out-degree increases. In order to state the precise decompositions of  $X_{n,j}$  we utilise our previous results for  $K_j$  and  $Y_{n,j}$ . We also collect known results about additional random variables.

We introduce first the stopping time  $\tau_{n,j} := \min_{j \leq \ell \leq n} \{Y_{\ell,j} = b + 1 - K_j\}$  until the node containing label  $j$  is saturated. Let  $\nabla$  denote the backward difference operator, i.e.,  $\nabla x_k := x_k - x_{k-1}$ . The random variable  $\tau_{n,j}$  can be expressed in terms of indicator variables as follows:

$$\tau_{n,j} = \sum_{k=j}^n \mathbb{I}\{Y_{k,j} = b + 1 - K_j, \nabla Y_{k,j} = 1\} \cdot k + \mathbb{I}\{Y_{n,j} < b + 1 - K_j\} \cdot n,$$

where  $Y_{j-1,j} = 0$ . Let  $\mathbb{P}\{m <_t v \mid c(v) = b - 1\}$  denote the conditional probability that node  $v$ , containing label  $j$ , attracts label  $m$  conditioned on the event  $c(v) = b - 1$ , as determined by the stochastic growth processes of the three tree families considered:

$$\mathbb{P}\{m <_t v \mid c(v) = b - 1\} = \begin{cases} \frac{b - 1}{m - 1}, & \text{bucket recursive trees,} \\ \frac{(d - 1)(b - 1) + 1}{(d - 1)(m - 1) + 1}, & \text{(b,d)-ary ITs,} \\ \frac{(\alpha + 1)(b - 1) - 1}{(\alpha + 1)(m - 1) - 1}, & \text{(b, \alpha)-PORTs.} \end{cases}$$

The probability mass function  $\mathbb{P}\{\tau_{n,j} = m\}$  is readily obtained in terms of the probability mass functions of  $K_j, Y_{n,j}$  and thus  $Y_{n,\ell,j}$ , where the constants  $c_1, c_2$  are determined by (23) via  $c_2/c_1 = \kappa$ .

**Lemma 33** (Distribution of the saturation time). *For  $1 \leq j \leq b$  the random variable  $\tau_{n,j}$  is deterministic:  $\tau_{n,j} = b$ .*

For  $j \geq b + 1$  the probability mass function  $\mathbb{P}\{\tau_{n,j} = m\}$  is given as follows:  
 For  $m = j$  we have

$$\mathbb{P}\{\tau_{n,j} = j\} = \mathbb{P}\{K_j = b\},$$

for  $j + 1 \leq m < n$  we get

$$\mathbb{P}\{\tau_{n,j} = m\} = \sum_{\ell=1}^{b-1} \mathbb{P}\{K_j = \ell\} \cdot \mathbb{P}\{Y_{m-1,\ell,j} = b - \ell\} \cdot \mathbb{P}\{m <_{\ell} v \mid c(v) = b - 1\},$$

whereas for  $m = n$  we obtain

$$\mathbb{P}\{\tau_{n,j} = n\} = \sum_{\ell=1}^{b-1} \mathbb{P}\{K_j = \ell\} \cdot (\mathbb{P}\{Y_{n,\ell,j} < b + 1 - \ell\} + \mathbb{P}\{Y_{n-1,\ell,j} = b - \ell\} \cdot \mathbb{P}\{n <_{\ell} v \mid c(v) = b - 1\}).$$

**Remark 34.** In the case  $j = b + 1$  the expressions above simplify due to the fact that  $K_{b+1} = 1$ , which is evident from the stochastic growth rules.

**Proof.** For  $1 \leq j \leq b$  there is only one bucket present. It is saturated after the insertion of label  $b$ . Assume that  $j \geq b + 1$ . The probability  $\mathbb{P}\{\tau_{n,j} = j\}$  is given by

$$\mathbb{P}\{Y_{j,j} = b + 1 - K_j, \nabla Y_{j,j} = 1\} = \mathbb{P}\{1 = b + 1 - K_j, Y_{j,j} - Y_{j-1,j} = 1\} = \mathbb{P}\{K_j = b\}.$$

For  $j < m < n$  we obtain the probability of the event  $\{Y_{k,j} = b + 1 - K_j, \nabla Y_{k,j} = 1\}$  by conditioning on the initial bucket size  $K_j$ . Finally, for  $j = n$  we also take into account the probability of the event  $\{Y_{n,j} < b + 1 - K_j\}$ . □

Let  $\text{Be}(p)$  denote a Bernoulli-distributed random variable:

$$\mathbb{P}\{\text{Be}(p) = 1\} = p, \quad \mathbb{P}\{\text{Be}(p) = 0\} = 1 - p.$$

Furthermore, let  $W_N(w_0, b_0)$  denote the number of white balls at time  $N$  in a triangular urn model with balls of two colours and initial values  $w_0 > 0$  and  $b_0 > 0$ , whose balanced replacement matrix given by  $\begin{pmatrix} 1 & \alpha \\ 0 & 1 + \alpha \end{pmatrix}$ .

We obtain the following result.

**Theorem 35.** *The random variable  $X_{n,j}$  counting the out-degree of label  $j$  in a random bucket increasing tree of size  $n$ , with  $1 \leq j \leq n$ , satisfies the following.*

- For bucket recursive trees,  $X_{n,j}$  is distributed as a random sum of mutually independent indicator variables:

$$X_{n,j} \stackrel{(d)}{=} \sum_{\ell=\tau_{n,j}+1}^n \mathbb{I}(\ell <_c j), \quad \text{with } \mathbb{I}(\ell <_c j) = \text{Be}\left(\frac{b}{\ell}\right).$$

- For  $(b, \alpha)$ -plane oriented recursive trees,  $X_{n,j}$  is distributed as the number of white balls in the balanced triangular urn model described above:

$$X_{n,j} \stackrel{(d)}{=} W_{n-\tau_{n,j}}(w_0, b_0), \quad w_0 = b(\alpha + 1) - 1, \quad b_0 = (\alpha + 1)(\tau_{n,j} - b).$$

In both cases, the stopping time  $\tau_{n,j} := \min_{j \leq \ell \leq n} \{Y_{\ell,j} = b + 1 - K_j\}$  depends on  $K_j$  as well as on  $Y_{n,j}$ .

**Remark 36.** A similar result holds for (b,d)-ary increasing trees; however, the random variable  $X_{n,j}$  is by definition bounded, thus we leave the result to the interested reader. We further note that the probability mass functions of  $X_{n,j}$  can alternatively be obtained by a generating functions approach, but without revealing the structural decompositions (compare with the corresponding results of [13] for bucket recursive trees).

## 8. Conclusion

In this work we introduced two new families of bucket increasing trees, which can be generated by a stochastic growth process.

We introduced a clustering process  $\mathcal{C}$  for ordinary increasing trees that generate bucket increasing trees. Moreover, by modifying the map  $\mathcal{C}$  we obtain bijections between certain ordinary increasing tree families and families of bilabelled bucket increasing trees. Additionally, we obtained a bijection between increasing diamonds and bilabelled bucket increasing trees.

We analysed structural properties of bucket increasing trees, in particular, the tree parameter  $K_n$ , counting the initial bucket size of the node containing label  $n$  in a random tree of size  $n$ . Using the combinatorial description as well as the tree evolution process, a study of further quantities in bucket increasing tree families is possible, e.g., we want to mention node distances. Moreover, there exist relations of bucket increasing trees to further combinatorial structures as, e.g., certain models of series-parallel networks, see [17].

## Acknowledgements

The authors warmly thank Henning Sulzbach for his feedback and the anonymous referees for their insightful comments, improving the presentation of this work.

## References

- [1] Berger, N., Borgs, C., Chayes, J. T., D'Souza, R. M. and Kleinberg, R. D. (2004) Competition-induced preferential attachment. In *Proceedings of the 31st International Colloquium on Automata, Languages and Programming (ICALP)*, Vol. 3142 of *Lecture Notes in Computer Science*, pp. 208–221.
- [2] Berger, N., Borgs, C., Chayes, J. T., D'Souza, R. M. and Kleinberg, R. D. (2005) Degree distribution of competition-induced preferential attachment graphs. *Comb. Prob. Comput.* **14** 697–721.
- [3] Bergeron, F., Flajolet, P. and Salvy, B. (1992) Varieties of increasing trees. *Lect. Notes Comput. Sci.* **581** 24–48.
- [4] Bodini, O., Dien, M., Fontaine, X., Genitrini, A. and Hwang, H.-K. (2016) Increasing diamonds. *Lect. Notes Comput. Sci.* **9644** 207–219.
- [5] Drmota, M. (2009) *Random Trees*. Springer.
- [6] Flajolet, P. and Sedgewick, R. (2009) *Analytic Combinatorics*. Cambridge University Press.
- [7] Janson, S. (2004) Functional limit theorems for multitype branching processes and generalized Pólya urns. *Stochast. Process. Appl.* **110** 177–245.
- [8] Janson, S., Kuba, M. and Panholzer, A. (2011) Generalized Stirling permutations, families of increasing trees and URN models. *J. Comb. Theory Ser. A* **118** 94–114.
- [9] Kazemi, R. (2014) Depth in bucket recursive trees with variable capacities of buckets. *Acta Mathematica Sinica, English Series* **30** 305–310.
- [10] Kuba, M. and Panholzer, A. (in preparation) Characterization of growth processes generating bucket increasing trees.
- [11] Kuba, M. and Panholzer, A. (2007) On the degree distribution of the nodes in increasing trees. *J. Comb. Theory Ser. A* **114** 597–618.
- [12] Kuba, M. and Panholzer, A. (2009) A combinatorial approach for analyzing the number of descendants in increasing trees and related parameters. *Quaestiones Mathematicae* **32** 91–114.
- [13] Kuba, M. and Panholzer, A. (2010) A combinatorial approach to the analysis of bucket recursive trees and variants. *Theoret. Comput. Sci.* **411**(34–36) 3255–3273.
- [14] Kuba, M. and Panholzer, A. (2012) Bilabelled increasing trees and hook-length formulas. *Eur. J. Comb.* **33** 248–258.
- [15] Kuba, M. and Panholzer, A. (2016) Combinatorial families of multilabelled increasing trees and hook-length formulas. *Discrete Math.* **339** 227–254.
- [16] Kuba, M. and Panholzer, A. (2016) Moment sequences and mixed Poisson distributions. *Prob. Surv.* **13** 89–155.

- [17] Kuba, M. and Panholzer, A. (2019) Combinatorial analysis of growth models for series-parallel networks. *Comb. Prob. Comput.* **28** 574–599.
- [18] Mahmoud, H. and Smythe, R. (1995) Probabilistic analysis of bucket recursive trees. *Theoret. Comput. Sci.* **144** 221–249.
- [19] Mahmoud, H. and Smythe, R. (1995) A survey of recursive trees. *Theoret. Prob. Math. Stat.* **51** 1–37.
- [20] Müller, N. S. (2019) Central limit theorem analogues for multicolour urn models. submitted.
- [21] Panholzer, A. and Prodinger, H. (2007) The level of nodes in increasing trees revisited. *Random Struct. Alg.* **31** 203–226.
- [22] Pittel, B. (1994) Note on the heights of random recursive trees and random m-ary search trees. *Random Struct. Alg.* **5** 337–347.
- [23] Prodinger, H. (1996) Depth and path length of heap ordered trees. *Int. J. Found. Comput. Sci.* **7** 293–299.
- [24] Prodinger, H. (1996) Descendants in heap ordered trees - or - a triumph of computer algebra. *Electron. J. Comb.* **3**(#R29).