



**HAL**  
open science

## Counting embeddings of rooted trees into families of rooted trees

Bernhard Gittenberger, Zbigniew Gołębiewski, Isabella Larcher, Malgorzata Sulkowska

► **To cite this version:**

Bernhard Gittenberger, Zbigniew Gołębiewski, Isabella Larcher, Malgorzata Sulkowska. Counting embeddings of rooted trees into families of rooted trees. *The Electronic Journal of Combinatorics*, 2022, 29 (3), pp.1-34. 10.37236/10760 . hal-03462575

**HAL Id: hal-03462575**

**<https://hal.science/hal-03462575>**

Submitted on 3 Dec 2021

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

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

# COUNTING EMBEDDINGS OF ROOTED TREES INTO FAMILIES OF ROOTED TREES

BERNHARD GITTENBERGER, ZBIGNIEW GOŁĘBIEWSKI,  
ISABELLA LARCHER, AND MAŁGORZATA SULKOWSKA

ABSTRACT. The number of embeddings of a partially ordered set  $S$  in a partially ordered set  $T$  is the number of subposets of  $T$  isomorphic to  $S$ . If both,  $S$  and  $T$ , have only one unique maximal element, we define good embeddings as those in which the maximal elements of  $S$  and  $T$  overlap. We investigate the number of good and all embeddings of a rooted poset  $S$  in the family of all binary trees on  $n$  elements considering two cases: plane (when the order of descendants matters) and non-plane. Furthermore, we study the number of embeddings of a rooted poset  $S$  in the family of all planted plane trees of size  $n$ . We derive the asymptotic behaviour of good and all embeddings in all cases and we prove that the ratio of good embeddings to all is of the order  $\Theta(1/\sqrt{n})$  in all cases, where we provide the exact constants. Furthermore, we show that this ratio is asymptotically non-decreasing with  $S$ . Finally, we comment on the case when  $S$  is disconnected.

## 1. INTRODUCTION

This paper studies the number of embeddings of a given rooted tree in the family of (plane and non-plane) binary trees, as well as planted plane trees. Here, the notion of embedding is wider than just a copy. We assume the investigated structures to be partially ordered sets (in short: posets) and by saying that there exists an embedding of  $S$  into  $T$  we understand (in the non-plane case) that a poset  $S$  is a subposet of  $T$ . We distinguish between good embeddings in which the roots of  $S$  and  $T$  overlap and bad embeddings in which they do not. The number of good and bad embeddings of a rooted structure in a complete binary tree was first investigated by Morayne [35]. His research was motivated by optimal stopping problems. The ratio of the number of good embeddings to the number of all embeddings and its monotonicity properties were used in estimates of conditional probabilities needed to obtain an optimal policy for the best choice problem considered on a complete (balanced) binary tree. This and similar results first served just as tools but soon became interesting questions about the structural features of posets on their own and resulted in a series of self-standing papers [27, 28, 20]. Counting chains and antichains in trees took a special place in this pool [31, 32, 29].

In this paper we present a follow-up and generalization of the results obtained by Kubicki *et al.* [27, 28] and Georgiou [20]. We give the asymptotic behaviour of the number of good and all embeddings of a rooted tree  $S$  in the family of plane (when the order of descendants matters) and non-plane binary trees, as well as planted plane trees, on  $n$  vertices. We prove that the ratio of the number of good embeddings to the number of all embeddings is of the order  $\Theta(1/\sqrt{n})$  in all cases and provide the exact constants. Furthermore, we show that this ratio is asymptotically non-decreasing in  $S$ . We comment also on the case where  $S$  is disconnected, *i.e.* a forest. In order to obtain those results we use tools of analytic combinatorics that have not been used before in the aforementioned papers.

The results of our paper may also be put into the framework of counting patterns in large structures. This is a vast field where many different types of structures have been considered. We only mention subgraph avoidance (and characterizing whole graph classes like series-parallel or planar graphs in that way) or subgraph counts in random graphs (see [25, 1]), pattern avoidance

---

*Date:* September 29, 2021.

This research has been supported by the ÖAD, grant PL04-2018, and Wrocław University of Science and Technology grant 0401/0052/18.

in permutations (see [4]) or in trees (see [9]), or pattern avoidance in lattice paths and words, where many particular patterns have been treated separately (see [8] or the introduction of [2] for a survey) and eventually put under a unifying umbrella in [2].

The closest to the present work is pattern counting in trees. One of the earliest investigations of this kind was [38], where the enumeration of given stars as subgraphs in trees (equivalently nodes of fixed degree) was treated. Later generalizations are found in [10, 36] (multivariate setting), in [33] (distinct patterns) or [22] (large patterns of that type). A method to deal with general contiguous patterns in trees by means of generating functions was developed in [5], which was partially generalized to planar maps recently [12, 6, 11]. Pattern avoidance in trees was the topic of [39], where also the concept of Wilf equivalence was dealt with, which was adopted from pattern avoidance in permutations.

Except for permutations, where most of the patterns that have been studied so far are non-contiguous, the considered patterns in other domains are typically contiguous. To our knowledge, the first work considering non-contiguous patterns in trees is [7]. In the present paper, the tree which is embedded becomes in general a collection of (partially) non-adjacent nodes in the tree where it is embedded. It can therefore be seen as a non-contiguous pattern occurring in that tree. Thus, our paper deals with certain enumeration problems for non-contiguous patterns in trees.

The paper is organized as follows. Section 2 introduces basic definitions and notation. Section 3 provides possible applications of our results in optimal stopping problems. In Section 4 we obtain generating functions for the number of good and the number of all embeddings of a rooted tree in the family of all plane binary trees with a given number of vertices. Section 5 is devoted to the asymptotics of this number when the size of the underlying tree is tending to infinity. Moreover, we investigate the asymptotics as well as the asymptotic monotonicity of their ratio. In Section 6 we briefly discuss the case when the embedded structure is disconnected, *i.e.* it is a forest. Section 7 deals with the non-plane binary case and in Section 8 the problem is extended to planted plane trees. A discussion of the obtained results as well as an outlook into some related future problems is given in Section 9.

## 2. DEFINITIONS AND NOTATION

By  $\mathcal{B}_n$  we denote the family of (unlabelled) plane binary trees with  $n$  nodes. A binary tree is a tree in which each node has either 0 or 2 descendants and by plane we understand that the order of subtrees of a given node matters, *i.e.* we distinguish between the different embeddings of a tree in the plane. It is commonly known that for odd  $n$  the cardinality  $|\mathcal{B}_n|$  of  $\mathcal{B}_n$  satisfies  $|\mathcal{B}_n| = C_{\frac{n-1}{2}}$ , where  $C_k$  is the  $k$ -th Catalan number given by  $C_k = \frac{1}{k+1} \binom{2k}{k}$ . Note that all binary trees have odd sizes and thus, for even  $n$  the cardinality  $|\mathcal{B}_n|$  is zero. All plane binary trees of size 5 are shown in Figure 1. We assume also that all edges are directed towards the descendants. Therefore, the in-degree of the root, as well as the out-degree of each leaf, is always 0. A vertex is said to be  $d$ -ary if its out-degree equals  $d$ . Subsequently, the root of a tree  $T$  will be denoted by  $\mathbf{1}_T$ .

By  $\mathcal{V}_n$  we denote the family of non-plane binary trees with  $n$  nodes. By non-plane we understand that the subtrees of a given node are treated as a set of subtrees, *i.e.* there is no ordering. E.g., there is only one non-plane binary tree of size 5, see Figure 1. Again for even  $n$  the cardinality  $|\mathcal{V}_n|$  is zero. For odd  $n$  the values  $|\mathcal{V}_n|$  are known as Wedderburn-Etherington numbers and do not have a closed form ( $|\mathcal{V}_1| = 1, |\mathcal{V}_3| = 1, |\mathcal{V}_5| = 1, |\mathcal{V}_7| = 2, |\mathcal{V}_9| = 3, \dots$ ).

Planted plane trees (also known as Catalan trees) are rooted plane trees where each internal node can have arbitrarily many descendants. We denote the family of planted plane trees of size  $n$  by  $\mathcal{T}_n$ . For all  $n$  the cardinality of  $\mathcal{T}_n$  satisfies  $|\mathcal{T}_n| = C_{n-1}$ .

This paper concentrates on investigating the number of embeddings of any rooted tree (or a forest of rooted trees - a disconnected graph whose components are rooted trees) in all trees from either family  $\mathcal{B}_n$ ,  $\mathcal{V}_n$  or  $\mathcal{T}_n$ . An embedding of a rooted tree  $S$  into another rooted tree  $T$  can be seen as a kind of generalized pattern occurrence of  $S$  in  $T$ , defined as follows, where we distinguish between the plane and the non-plane case.

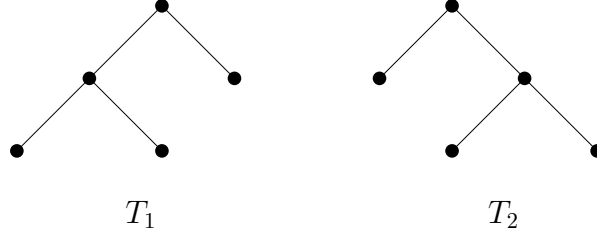


FIGURE 1. The family  $\mathcal{B}_5 = \{T_1, T_2\}$  of plane binary trees is of size  $|\mathcal{B}_5| = \mathcal{C}_2 = 2$ , while the family  $\mathcal{V}_5 = \{T_1\}$  of non-plane binary trees has the size  $|\mathcal{V}_5| = 1$ .

**Definition 1** (non-plane embedding). Let  $S$  and  $T$  be two non-plane rooted trees. When interpreting  $T$  as the cover graph of a partially ordered set (poset), rooted at the root of  $T$ , *i.e.* at the single maximal element of the poset, then an embedding of  $S$  into  $T$  can be defined as any subposet of  $T$  isomorphic to  $S$ .

*Remark 1.* Note that there exists a non-plane embedding of a binary tree  $S$  into a binary tree  $T$  if and only if  $S$  is a minor of  $T$ .

*Remark 2.* Instead of starting from a tree as combinatorial structure and then interpreting it as a poset, we may also start from posets and then define a *tree poset* as a poset  $P$  which has exactly one maximal element and such that any Hasse diagram of  $P$  looks like a (combinatorial) tree. This is equivalent to the definition of a tree poset given in [17].<sup>1</sup> Likewise, an embedding of a tree poset  $S$  into another tree poset  $T$ , as defined in [17], matches exactly the definition of a non-plane embedding given above.

**Definition 2** (plane embedding). Let  $S$  and  $T$  be two plane rooted trees. If we interpret  $T$  to be a Hasse diagram of a poset, then an embedding of  $S$  into  $T$  can be defined as any subposet of  $T$  isomorphic to  $S$  in which the left-to-right order of the children of each node of  $S$  is inherited from  $T$  (thus, a plane version of a subposet).

*Remark 3.* So, in the plane case  $S$  and  $T$  can be interpreted as Hasse diagrams of posets, and whenever  $S$  can be embedded in  $T$  it follows that  $S$  is a subposet of  $T$ . However, note that the respective posets can possibly be represented as different Hasse diagrams in such a way that no embedding of the corresponding trees is possible.

We say that an embedding of  $S$  into  $T$  is *good* if it contains the root of  $T$ . Otherwise we call it a *bad embedding*. If there exists at least one embedding of  $S$  into  $T$ , we write  $S \subseteq T$ . All embeddings of a cherry, (*i.e.* a tree composed only of a root and its two children) in a given binary tree of size 5 are given in Figure 2. Four of them are good and the last one is bad.

Subsequently the size of the tree  $S$  will always be denoted by  $m$ , while the size of  $T$  is consistently denoted by  $n$ . Thus, for the asymptotic analysis of the number of embeddings of a tree  $S$  into a class of trees of size  $n$ , the quantity  $m$  is considered to be a constant, while  $n$  tends to infinity.

For  $S$ , the structure that we embed, we define its *degree distribution sequence* as  $d_S = (d_0, d_1, \dots, d_{m-1})$ , where  $d_i$  is the number of vertices in  $S$  with out-degree equal to  $i$ . Note that  $d_0$  is simply the number of leaves, which will be, interchangeably, denoted by  $l$  (*i.e.*  $l = d_0$ ). Similarly,  $d_1$  is the number of unary nodes, which will be, interchangeably, denoted by  $u$  (*i.e.*  $u = d_1$ ). The number of all embeddings of a given tree  $S$  in  $T$  will be denoted by  $a_T(S)$  and the number of its good embeddings in  $T$  by  $g_T(S)$ . The number of all embeddings of  $S$  in a family  $\mathcal{F} = \{F_1, \dots, F_N\}$  will be denoted by  $a_{\mathcal{F}}(S)$  and understood as the cumulative number of embeddings of  $S$  into all elements of  $\mathcal{F}$ , *i.e.*  $a_{\mathcal{F}}(S) = \sum_{i=1}^N a_{F_i}(S)$ . Analogously, we define the number of good embeddings of  $S$  in  $\mathcal{F}$ :  $g_{\mathcal{F}}(S) = \sum_{i=1}^N g_{F_i}(S)$ . For  $S$  being a cherry and  $\mathcal{B}_5 = \{T_1, T_2\}$ , we obtain  $a_{T_1}(S) = a_{T_2}(S) = 5$ ,  $g_{T_1}(S) = g_{T_2}(S) = 4$ , thus  $a_{\mathcal{B}_5}(S) = 10$  and  $g_{\mathcal{B}_5}(S) = 8$  (compare Figure 2). Notations for  $\mathcal{V}_n$  and  $\mathcal{T}_n$  are analogous.

<sup>1</sup>For the sake of better distinction from a combinatorial tree, we use the term “tree poset” for what is simply called “tree” in [17].

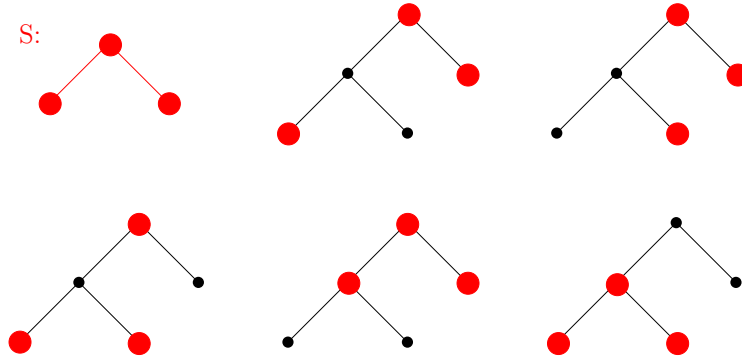


FIGURE 2. All five embeddings of a cherry  $S$  in a given plane binary tree of size 5. Or all four embeddings of a cherry  $S$  in a given non-plane binary tree of size 5, since in the non-plane case the two rightmost pictures in the upper row represent the same embedding (they can easily be mapped onto each other via a simple automorphism that changes the order of the two leftmost leaves).

Throughout this paper we use the standard notation  $f(n) \sim g(n)$  if  $\lim_{n \rightarrow \infty} \frac{f(n)}{g(n)} = 1$ . Also  $[z^n]f(z)$  denotes the coefficient of  $z^n$  in the formal power series  $f(z) = \sum_{n \geq 0} f_n z^n$ , *i.e.*  $[z^n] \left( \sum_{n \geq 0} f_n z^n \right) = f_n$ .

### 3. APPLICATIONS IN OPTIMAL STOPPING PROBLEMS

The most prominent problem in the area of optimal stopping is the so-called “secretary problem” (consult [34, 13, 15, 40]), where one assumes a linear order on the applicants for a secretary position concerning their qualifications. The applicants are interviewed in a random order and the decision whether to hire an applicant has to be made immediately after the interview - a rejected applicant cannot be hired at a later point. Thus, if we interview all the candidates, we have to hire the last applicant. The goal is to find the optimal stopping strategy to hire the best applicant. Thus, we want to stop at the time maximizing the probability that the present applicant is the best one overall, *i.e.* the maximum element in the linear order. It has been proved (see for example [34, 21]) that for a large number of applicants it is optimal to wait until approximately 37% (more precisely  $\frac{100}{e}$ %) of the applicants have been interviewed and then to select the next relatively best one. This optimal algorithm returns the best applicant with asymptotic probability of  $1/e$ . The secretary problem has been extended and generalized in many different directions. One of these is the extension to partially ordered sets, possibly with more than one maximal element, see [41, 23]. Optimal strategies for particular posets were investigated among others in [35, 26]. Versions for unknown poset, when the selector knows in advance only its cardinality, were presented in [37, 16, 18]. Another interesting generalization was to replace the underlying poset structure by a directed graph. This version was first considered on directed paths by Kubicki and Morayne in [30] and later extended to other families of graphs and different versions of the game (consult [42, 24, 3]).

In the remainder of this section we give examples of stopping problems in which either the value  $a_{\mathcal{V}_n}(S)$  or the ratio  $g_{\mathcal{V}_n}(S)/a_{\mathcal{V}_n}(S)$  (both investigated in this paper) plays a crucial role in estimating the conditional probabilities needed to obtain the optimal policy. One can consider analogous examples for the families  $\mathcal{B}_n$  or  $\mathcal{T}_n$  as well.

Let us think about elements of  $\mathcal{V}_n$  as of Hasse diagrams of posets. Consider the following process. Elements (*i.e.* nodes) of some  $T$  from  $\mathcal{V}_n$  appear one by one in a random order (all permutations of elements of  $T$  are equiprobable). At time  $t$ , *i.e.* when  $t$  elements have already appeared, the selector can see a poset induced on those elements. He knows that the underlying structure is drawn uniformly at random from  $\mathcal{V}_n$ .

**Example** (Best choice problem for the family of binary trees). The selector’s task is to stop the process maximizing the probability that the element that has just appeared is the root of the

underlying structure. He wins only if the chosen element is indeed  $\mathbb{1}_T$ . Note that it neither pays off to stop the process when the induced structure is disconnected nor when the currently observed element is not the maximal one in the induced poset. The selector wonders whether to stop only if the emerged element at time  $t$  is the unique maximal element in the induced structure. In order to take a decision whether to stop at time  $t$ , he needs to know the probability of winning if he stops now. Let  $W_t$  denote the event of winning when stopping at time  $t$ ,  $S_t$  the event that at time  $t$  he observes a certain structure  $S$  with degree distribution sequence  $d_S$  and  $R_i$  denote the event that  $T_i$  has been drawn as the underlying structure, where we use the notation  $\mathcal{V}_n = \{T_1, \dots, T_N\}$  with  $N = |\mathcal{V}_n|$ . Then the probability of winning if he stops at time  $t$  is given by

$$\mathbb{P}[W_t|S_t] = \sum_{i=1}^N \mathbb{P}[W_t|S_t \cap R_i] \mathbb{P}[R_i|S_t] = \sum_{i=1}^N \frac{g_{T_i}(S)}{a_{T_i}(S)} \frac{\mathbb{P}[S_t|R_i] \mathbb{P}[R_i]}{\mathbb{P}[S_t]}.$$

Since  $\mathbb{P}[R_i] = 1/N$ ,  $\mathbb{P}[S_t|R_i] = a_{T_i}(S)/\binom{n}{t}$  and

$$\mathbb{P}[S_t] = \sum_{i=1}^N \mathbb{P}[S_t|R_i] \mathbb{P}[R_i] = \sum_{i=1}^N \frac{a_{T_i}(S)}{\binom{n}{t}} \frac{1}{N} = \frac{a_{\mathcal{V}_n}(S)}{N \binom{n}{t}}$$

we get

$$\mathbb{P}[W_t|S_t] = \sum_{i=1}^N \frac{g_{T_i}(S)}{a_{T_i}(S)} \frac{a_{T_i}(S)}{\binom{n}{t}} \frac{1}{N} \frac{N \binom{n}{t}}{a_{\mathcal{V}_n}(S)} = \frac{g_{\mathcal{V}_n}(S)}{a_{\mathcal{V}_n}(S)}.$$

**Example** (Identifying complete balanced binary trees). The selector has to identify whether the underlying structure is a complete balanced binary tree or not. The payoff of the game, if he stops the process at time  $t$ , is  $n - t$  if he guesses correctly and 0 otherwise. He has to maximize the expected payoff. At moment  $t$  he observes a structure  $S$ , which is not necessarily connected. Again, in order to make a decision whether to stop, he needs to know what is the probability that the currently observed structure is a subposet of a complete balanced binary tree. For a rooted tree  $S$  this probability is given by

$$\frac{a_{T_b}(S)}{a_{\mathcal{V}_n}(S)},$$

where  $T_b \in \mathcal{V}_n$  denotes the complete balanced binary tree of size  $n$ .

#### 4. GENERATING FUNCTIONS FOR THE NUMBER OF EMBEDDINGS IN $\mathcal{B}_n$

In this section we derive generating functions for the sequences  $a_{\mathcal{B}_n}(S)$  and  $g_{\mathcal{B}_n}(S)$ , where  $S$  is a given rooted plane tree of size  $m$ . In order to do so, we use the symbolic method (consult [14]).

**Theorem 1.** *Consider a rooted tree  $S$  with degree distribution sequence  $d_S = (l, u, d_2, \dots, d_{m-1})$ . The generating function  $A_S(z)$  of the sequence  $a_{\mathcal{B}_n}(S)$ , which counts the number of all embeddings of  $S$  into all trees of the family  $\mathcal{B}_n$  is given by*

$$A_S(z) = \left( \frac{1}{1 - 2zB(z)} \right)^{m+l-1} z^{l+u-1} B(z)^{l+u} 2^u \prod_{i=3}^{m-1} (C_{i-1})^{d_i},$$

where  $B(z)$  is the generating function of the family of plane binary trees, i.e.

$$B(z) = \frac{1 - \sqrt{1 - 4z^2}}{2z} = C_0z + C_1z^3 + C_2z^5 + C_3z^7 + \dots$$

*Remark 4.* Note that  $A_S(z)$  depends only on the degree distribution sequence  $d_S$ , not the particular shape of  $S$ . Thus, as long as  $d_{S_1}$  and  $d_{S_2}$  are the same,  $A_{S_1}(z)$  and  $A_{S_2}(z)$  coincide even if  $S_1$  and  $S_2$  are not isomorphic. However, we use the subscript  $S$  to provide a transparent notation. Moreover, note that  $A_S(z)$  does also depend on the tree class  $\mathcal{B}_n$  in which we embed the tree  $S$ . In order to avoid a large number of indices we will omit to indicate this dependence and just emphasize at this point that the generating functions  $A_S(z)$  may differ according to the underlying tree classes.

*Proof.* First let us recall that the class  $\mathcal{B}$  of plane binary trees can be specified by

$$\mathcal{B} = \{\bullet\} + \{\bullet\} \times \mathcal{B} \times \mathcal{B},$$

since every node is either a leaf or a binary node with two attached binary trees. By means of the symbolic method (see [14]) we can directly translate this specification into a functional equation that defines the generating function  $B(z)$  of binary trees where  $z$  marks the number of nodes, which gives

$$B(z) = z + zB(z)^2.$$

Solving this equation for  $B(z)$  yields the explicit formula that is given in the theorem. Now, we start the proof of the expression of  $A_S(z)$  with the case where  $S$  is a Motzkin tree, *i.e.* a tree where each internal node has either one or two children. The specification of the class  $\mathcal{M}$  of Motzkin trees is

$$\mathcal{M} = \{\bullet\} + \{\bullet\} \times \mathcal{M} + \{\bullet\} \times \mathcal{M} \times \mathcal{M} \quad (1)$$

and thereby we must distinguish between the three cases whether  $S$  is a single node, or the root of  $S$  is a unary node, or a binary node, and hence falling into the respective subclass of  $\mathcal{M}$  among the subclasses that we find as summands on the right-hand side of (1). The generating function  $A_S(z)$  for the number of embeddings of  $S$  into the family  $\mathcal{B}_n$  can then be recursively defined by

$$A_S(z) = \begin{cases} zB'(z) & \text{if } S = \bullet, \\ \frac{2zB(z)}{1-2zB(z)}A_{\tilde{S}}(z) & \text{if } S = (\bullet, \tilde{S}), \\ \frac{z}{(1-2zB(z))^2}A_{S_L}(z)A_{S_R}(z) & \text{if } S = (\bullet, S_L, S_R), \end{cases} \quad (2)$$

where the three cases correspond to the cases described above. The first case, which yields a factor  $zB'(z)$ , corresponds to marking a node in the underlying tree  $T$  (*i.e.* pointing at a node), because obviously a single vertex can be embedded in every node. We can also interpret it as counting the number of pairs  $(T, E)$  where  $E$  is an embedding of  $S$  into  $T$ .

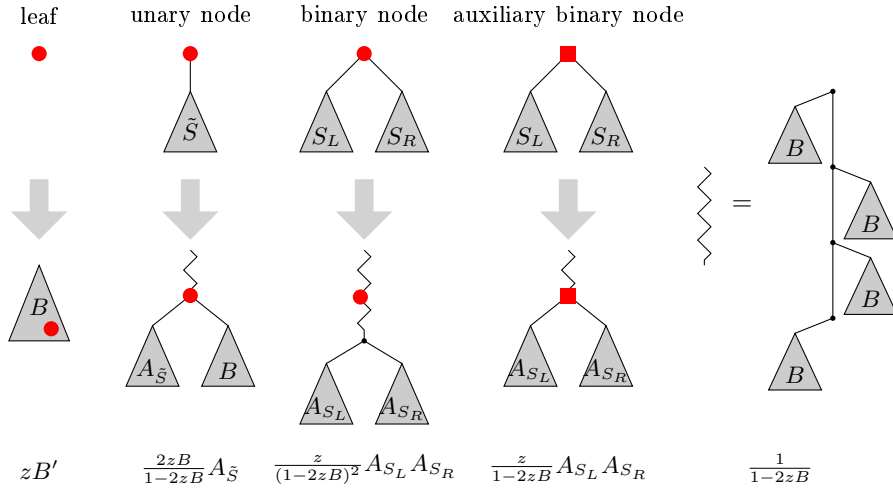


FIGURE 3. Sketch of the recursive construction of the generating function  $A_S(z)$ : When  $S$  is a Motzkin tree consisting of more than one vertex (plane binary case), then the first three cases above can appear. Here  $B$  each time refers to an abstract object representing any tree from family  $\mathcal{B}_n$ . If  $S$  contains vertices with three or more children, then auxiliary vertices occur. They are embedded according to the fourth picture above and depicted as red squares.

Now we show how an embedding of  $S$  into  $T$  can be constructed in a recursive way - see Figure 3 for a visualization of the used approach. We start with the case that the root of  $S$  is a unary node. This root has to be embedded at some point in the tree  $T$ . The part of  $T$  that is above the embedded root of  $S$  can be expressed as a path of left-or-right trees, which contributes

a factor  $\frac{1}{1-2zB(z)}$ . The embedded root of  $S$  itself yields a factor  $z$ , since the generating function of an object of size one is given by  $z$ . To the embedded root we have to attach an additional tree  $T$  in order to create a binary structure, yielding a factor  $B(z)$ , as well as the remaining tree that contains the embedding of  $\tilde{S}$ . The factor 2 that appears in the coefficient in the second case of (2) indicates that we work with plane trees - the substructure  $\tilde{S}$  can be embedded either in the left or in the right subtree of the unary vertex.

The third case of (2), where  $S$  starts with a binary node, is similar to the previous case. Thus, the factor  $\frac{1}{(1-2zB(z))^2}$  corresponds to two consecutive paths of left-or-right trees, which are separated by the embedded root which itself gives the additional factor  $z$ . At some point the lower path splits into two subtrees containing the embeddings of the subtrees  $S_L$  and  $S_R$ .

By simple iteration one can see that in case of embedding a Motzkin tree  $S$ , the generating function  $A_S(z)$  reads as

$$A_S(z) = \left( \frac{z}{(1-2zB(z))^2} \right)^{l-1} \left( \frac{2zB(z)}{1-2zB(z)} \right)^u (zB'(z))^l, \tag{3}$$

where  $l$  denotes the number of leaves and  $u$  the number of unary nodes in  $S$ . The exponent  $l-1$  in (3) arises from the fact that a Motzkin tree with  $l$  leaves has  $l-1$  binary nodes, and for each of these nodes we get the respective factor.

Finally, we consider the general case where  $S$  is an arbitrary plane tree without any restrictions on the degree distribution sequence. Then we proceed as follows. Every  $d$ -ary node with  $d \geq 3$  together with its  $d$  children is replaced by a binary tree having  $d$  leaves, which are then replaced by the successors of the original  $d$ -ary node. There are exactly  $C_{d-1}$  possible ways to construct such a binary tree. Unary and binary nodes stay unaltered. Applying this for all nodes results in constructing a Motzkin tree, called  $S'$ , and the number of Motzkin trees that can be constructed in that way is  $\prod_{i=3}^{m-1} C_{i-1}^{d_i}$ . These Motzkin trees are then embedded with the approach described above.

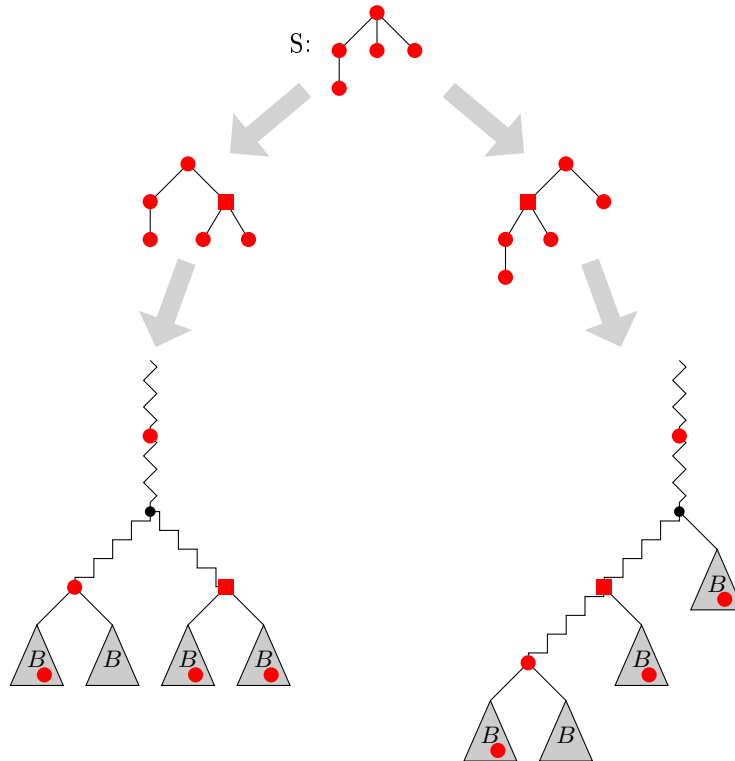


FIGURE 4. Sketch of the principle of embedding an arbitrary plane tree (plane binary case).



But note that replacing a  $d$ -ary node  $v$  with  $d \geq 3$  by a binary tree (with more than 1 internal nodes) introduced further vertices into  $S$ . In particular,  $v$  becomes a binary node, and as there were  $d$  successors before, the binary tree arising from  $v$  must have  $d - 1$  internal nodes, thus giving rise to  $d - 2$  auxiliary vertices. Since  $S$  has  $m$  vertices,  $S'$  has therefore  $m + \sum_{i=3}^{m-1} (i-2)d_i = 2l + u - 1$  vertices.

Furthermore note that the auxiliary vertices have to be embedded into  $T$ , but they do not belong to  $S$ . This causes a special treatment (see Figure 4 for an illustration). Consider two subtrees  $S_1$  and  $S_2$  of  $S'$  whose last common ancestor (in  $S'$ ) is an auxiliary vertex, say  $w$ , and let  $v$  be the predecessor of  $w$  in  $S'$ . Of course, any vertex on the path from  $v$  to  $w$  may serve as auxiliary vertex instead of  $w$ . But as  $w$  does not belong to  $S$ , its actual position is unimportant. Hence, for the sake of not overcounting, the auxiliary vertices are always placed at the last possible position, see Figures 3 and 4. This eventually yields a factor  $z/(1 - 2zB(z))$  for each binary auxiliary node. As there are  $2l + u - m - 1$  auxiliary nodes and  $m - l - u$  other binary nodes, this gives altogether

$$\begin{aligned} A_S(z) &= \left( \frac{z}{1 - 2zB(z)} \right)^{2l+u-m-1} \left( \frac{z}{(1 - 2zB(z))^2} \right)^{m-l-u} \left( \frac{2zB(z)}{1 - 2zB(z)} \right)^u (zB'(z))^l \prod_{i=3}^{m-1} (C_{i-1})^{d_i} \\ &= \frac{2^u z^{l+u-1} B(z)^u}{(1 - 2zB(z))^{m-1}} (zB'(z))^l \prod_{i=3}^{m-1} (C_{i-1})^{d_i}. \end{aligned}$$

Using the identity  $zB'(z) = \frac{B(z)}{1 - 2zB(z)}$ , which holds for plane binary trees, yields the desired result.  $\square$

**Corollary 1.** *Let  $S$  be a rooted tree. The generating function of the sequence  $g_{\mathcal{B}_n}(S)$ , which counts the number of good embeddings of  $S$  into all trees of the family  $\mathcal{B}_n$  is given by*

$$G_S(z) = (1 - 2zB(z))A_S(z).$$

*Proof.* The corollary follows immediately, as the only difference in the case of good embeddings is that the root of  $S$  is always embedded in the root of the underlying tree. Thus, we have to omit the path of left-or-right trees in the beginning. This corresponds to a multiplication by the factor  $(1 - 2zB(z))$ .  $\square$

## 5. ASYMPTOTICS OF THE NUMBER OF EMBEDDINGS IN $\mathcal{B}_n$

In this section we investigate the asymptotics of  $a_{\mathcal{B}_n}(S)$  and  $g_{\mathcal{B}_n}(S)$ , as well as monotonicity of their ratio when  $S$  is a rooted tree. As a tool we use singularity analysis which provides a relation between the behaviour of a generating function near its dominant singularities (*i.e.* its singularities on the circle of convergence) and the asymptotics of its coefficients. The following lemma will be helpful later on.

**Lemma 1** (Compare Theorems VI.4 and VI.5 in [14]). *Define*

$$\Delta_0 = \{z \in \mathbb{C} \mid |z| < \rho + \epsilon, z \neq \rho, |\arg(z - \rho)| > \nu\}$$

for some  $\rho > 0, \epsilon > 0, 0 < \nu < \frac{\pi}{2}$ . Let  $r \geq 0, \rho_j = \rho e^{i\phi_j}$ , for  $j = 0, 1, \dots, r$  with  $\phi_0 = 0$  and  $\phi_1, \dots, \phi_r \in (0, 2\pi)$ . Consider  $T(z) = \sum_{n \geq 0} T_n z^n$  to be an analytic function in  $\Delta := \bigcap_{j=0}^r e^{i\phi_j} \Delta_0$  and satisfying for each  $j = 0, \dots, r$

$$T(z) \sim K_j \left(1 - \frac{z}{\rho_j}\right)^{-\alpha_j}, \quad \text{as } z \rightarrow \rho_j \text{ in } \Delta,$$

where  $\alpha_j \notin \{0, -1, -2, \dots\}$  and the  $K_j$  are constants. Then

$$[z^n]T(z) \sim \sum_{j=0}^r K_j \frac{n^{\alpha_j-1}}{\Gamma(\alpha_j)} \rho_j^{-n}, \quad \text{as } n \rightarrow \infty.$$

*Remark 5.* Note that the assumptions of Lemma 1 imply that  $\{\rho_0, \rho_1, \dots, \rho_r\}$  is exactly the set of all singularities of the power series  $\sum_{n \geq 0} T_n z^n$  on its circle of convergence.

**Theorem 2.** Consider a rooted tree  $S$  with degree distribution sequence  $d_S = (l, u, d_2, \dots, d_{m-1})$ . Let  $C = \prod_{i=3}^{m-1} (C_{i-1})^{d_i}$ . The asymptotics of the number of all embeddings of  $S$  into  $\mathcal{B}_n$  is given by

$$a_{\mathcal{B}_n}(S) \sim \frac{C \cdot 2^{\frac{5-m-3l}{2}}}{\Gamma(\frac{m+l-1}{2})} \cdot 2^n \cdot n^{\frac{m+l-3}{2}}$$

for  $n$  being odd and  $a_{\mathcal{B}_n}(S) = 0$  for  $n$  being even. The asymptotics of the number of good embeddings of  $S$  into  $\mathcal{B}_n$  is given by

$$g_{\mathcal{B}_n}(S) \sim \begin{cases} \frac{C \cdot 2^{\frac{6-m-3l}{2}}}{\Gamma(\frac{m+l-2}{2})} \cdot 2^n \cdot n^{\frac{m+l-4}{2}} & \text{if } m+l-2 > 0 \\ \frac{\sqrt{2} \cdot 2^n}{\sqrt{\pi n^3}} & \text{if } m+l-2 = 0 \end{cases}$$

for  $n$  being odd and  $g_{\mathcal{B}_n}(S) = 0$  for  $n$  being even.

*Proof.* Recall that  $a_{\mathcal{B}_n}(S) = [z^n]A_S(z)$ . The function  $A_S(z)$  has two dominant singularities at  $\rho_0 = 1/2$  and  $\rho_1 = -1/2$ . Expanding  $A_S(z)$  into its Puiseux series at  $z \rightarrow \rho_0 = 1/2$  gives

$$A_S(z) = C \cdot 2^{\frac{3-m-3l}{2}} \cdot \left(1 - \frac{z}{\rho_0}\right)^{-\frac{m+l-1}{2}} \left(1 + O\left(\left(1 - \frac{z}{\rho_0}\right)^{1/2}\right)\right).$$

Note that  $m+l-1 \geq 1$ , since always  $l \geq 1$  and  $m \geq 1$ . Expanding  $A_S(z)$  into a Puiseux series at  $z \rightarrow \rho_1 = -1/2$  gives

$$A_S(z) = -C \cdot 2^{\frac{3-m-3l}{2}} \cdot \left(1 - \frac{z}{\rho_1}\right)^{-\frac{m+l-1}{2}} \left(1 + O\left(\left(1 - \frac{z}{\rho_1}\right)^{1/2}\right)\right).$$

By Lemma 1 we get

$$\begin{aligned} [z^n]A_S(z) &\sim \frac{C \cdot 2^{\frac{3-m-3l}{2}}}{\Gamma(\frac{m+l-1}{2})} \cdot (\rho_0)^{-n} \cdot n^{\frac{m+l-3}{2}} - \frac{C \cdot 2^{\frac{3-m-3l}{2}}}{\Gamma(\frac{m+l-1}{2})} \cdot (\rho_1)^{-n} \cdot n^{\frac{m+l-3}{2}} \\ &= \begin{cases} \frac{C \cdot 2^{\frac{5-m-3l}{2}}}{\Gamma(\frac{m+l-1}{2})} \cdot 2^n \cdot n^{\frac{m+l-3}{2}} & \text{if } n \text{ is odd,} \\ 0 & \text{if } n \text{ is even.} \end{cases} \end{aligned}$$

The asymptotic analysis for the number of good embeddings is analogous. Again,  $g_{\mathcal{B}_n}(S) = [z^n]G_S(z)$  and  $G_S(z)$  has two dominant singularities at  $1/2$  and  $-1/2$ . For  $m+l-2 > 0$  we obtain

$$[z^n]G_S(z) \sim \begin{cases} \frac{C \cdot 2^{\frac{6-m-3l}{2}}}{\Gamma(\frac{m+l-2}{2})} \cdot 2^n \cdot n^{\frac{m+l-4}{2}} & \text{if } n \text{ is odd,} \\ 0 & \text{if } n \text{ is even.} \end{cases}$$

The case  $m+l-2 = 0$  needs to be treated separately. Note that then  $m = 1$  and  $l = 1$ , thus the structure  $S$  that we embed is a single vertex. Therefore the number of good embeddings is just the cardinality of  $\mathcal{B}_n$ , i.e.  $g_{\mathcal{B}_n}(S) = C_{\frac{n-1}{2}} \sim \frac{\sqrt{2} \cdot 2^n}{\sqrt{\pi n^3}}$ . (Note also that for  $S$  being a single vertex  $a_{\mathcal{B}_n}(S) = nC_{\frac{n-1}{2}} \sim \frac{\sqrt{2} \cdot 2^n}{\sqrt{\pi n}}$ .)  $\square$

**Corollary 2.** Consider a rooted tree  $S$  with degree distribution sequence  $d_S = (l, u, d_2, \dots, d_{m-1})$ . Let  $k = \frac{m+l-2}{2}$  and let  $n$  be odd. The asymptotic ratio of the number of good embeddings of  $S$  into  $\mathcal{B}_n$  to the number of all embeddings into  $\mathcal{B}_n$  is given by

$$\frac{g_{\mathcal{B}_n}(S)}{a_{\mathcal{B}_n}(S)} \sim \begin{cases} \frac{\Gamma(k+1/2) \sqrt{2}}{\Gamma(k) \sqrt{n}} & \text{if } k > 0, \\ 1/n & \text{if } k = 0. \end{cases}$$

*Proof.* The corollary follows immediately from Theorem 2.  $\square$

Kubicki *et al.* [27] proved that if  $T$  is a complete balanced binary tree of arbitrary size and  $S_1, S_2$  are rooted trees in which each node has at most 2 descendants (i.e.  $S_1$  and  $S_2$  are Motzkin trees) and  $S_1 \subseteq S_2$ , then  $\frac{g_T(S_1)}{a_T(S_1)} \leq \frac{g_T(S_2)}{a_T(S_2)}$ . They also conjectured that the ratio  $\frac{g_T(S)}{a_T(S)}$  is weakly increasing with  $S$  for  $S$  being any rooted tree. One year later in [28] they also stated an asymptotic

result for the ratio  $\frac{g_T(S)}{a_T(S)}$  when  $S$  is an arbitrary rooted tree and  $T$  a complete binary tree of size  $n$ . They showed that  $\lim_{n \rightarrow \infty} \frac{g_T(S)}{a_T(S)} = 2^{l-1} - 1$  where  $l$  is the number of leaves in  $S$ . Thereby they proved that for any rooted tree  $S$  the asymptotic ratio  $\frac{g_T(S)}{a_T(S)}$  is non-decreasing with  $S$  (the function  $2^{l-1} - 1$  increases with  $l$  and if  $S_1 \subseteq S_2$  then the number of leaves of  $S_2$  equals at least the number of leaves of  $S_1$ ).

The conjecture from [27] was disproved by Georgiou [20] who chose specific ternary trees as embedded structures to construct a counterexample. He also generalized the underlying structure to a complete  $k$ -ary tree and considered strict-order preserving maps instead of embeddings. In this setting he proved that a correlation inequality (corresponding to  $\frac{g_{T_n}(S_1)}{a_{T_n}(S_1)} \leq \frac{g_{T_n}(S_2)}{a_{T_n}(S_2)}$ ) already holds for  $S_1, S_2$  being arbitrary rooted trees such that  $S_1 \subseteq S_2$ .

Referring to the asymptotic result from [28], we show below and in the subsequent sections that in our case the asymptotic ratios  $\frac{\sqrt{n} g_{\mathcal{B}_n}(S)}{a_{\mathcal{B}_n}(S)}$ ,  $\frac{\sqrt{n} g_{T_n}(S)}{a_{T_n}(S)}$  and  $\frac{\sqrt{n} g_{\mathcal{V}_n}(S)}{a_{\mathcal{V}_n}(S)}$  are all weakly increasing with  $S$  for  $S$  being an arbitrary rooted tree. Using this asymptotic result we show later that also the ratios  $\frac{g_{\mathcal{B}_n}(S)}{a_{\mathcal{B}_n}(S)}$ ,  $\frac{g_{T_n}(S)}{a_{T_n}(S)}$  and  $\frac{g_{\mathcal{V}_n}(S)}{a_{\mathcal{V}_n}(S)}$  (unlike in the case from [27]) are eventually weakly increasing with  $S$  for sufficiently large  $n$ . In order to do so, we use Gautschi's inequality given in the following lemma.

**Lemma 2** (Gautschi's inequality, [19]). *Let  $x$  be a positive real number and let  $s \in (0, 1)$ . Then*

$$x^{1-s} < \frac{\Gamma(x+1)}{\Gamma(x+s)} < (x+1)^{1-s}.$$

**Theorem 3.** *Let  $S_1, S_2$  be rooted trees such that  $S_1 \subseteq S_2$ . Then*

$$\lim_{n \rightarrow \infty} \sqrt{n} \frac{g_{\mathcal{B}_n}(S_1)}{a_{\mathcal{B}_n}(S_1)} \leq \lim_{n \rightarrow \infty} \sqrt{n} \frac{g_{\mathcal{B}_n}(S_2)}{a_{\mathcal{B}_n}(S_2)}.$$

*Proof.* Let  $d_{S_1} = (l_1, u_1, \dots)$ ,  $d_{S_2} = (l_2, u_2, \dots)$ ,  $k_1 = \frac{m_1 + l_1 - 2}{2}$ ,  $k_2 = \frac{m_2 + l_2 - 2}{2}$  (where  $m_i$  denotes the size of  $S_i$ ) and  $k_1 > 0$  (the case when  $k_1 = 0$  is trivial). By Corollary 2 we have

$$\lim_{n \rightarrow \infty} \sqrt{n} \frac{g_{\mathcal{B}_n}(S_1)}{a_{\mathcal{B}_n}(S_1)} = \frac{\sqrt{2} \cdot \Gamma(k_1 + 1/2)}{\Gamma(k_1)} \quad \text{and} \quad \lim_{n \rightarrow \infty} \sqrt{n} \frac{g_{\mathcal{B}_n}(S_2)}{a_{\mathcal{B}_n}(S_2)} = \frac{\sqrt{2} \cdot \Gamma(k_2 + 1/2)}{\Gamma(k_2)}.$$

Note that the values  $k_1, k_1 + 1/2, k_2$  and  $k_2 + 1/2$  all belong to the set  $\{\frac{1}{2}, 1, \frac{3}{2}, 2, \frac{5}{2}, \dots\}$ . First, we are going to show that the function  $f(k) = \frac{\Gamma(k+1/2)}{\Gamma(k)}$  is increasing in  $k$  for  $k \in \{\frac{1}{2}, 1, \frac{3}{2}, 2, \frac{5}{2}, \dots\}$ . Indeed, applying twice Gautschi's inequality (Lemma 2) we get for  $k > 1/2$

$$\frac{f(k+1/2)}{f(k)} = \frac{\Gamma(k+1)}{\Gamma(k+1/2)} \frac{\Gamma(k)}{\Gamma(k+1/2)} > k^{1/2} (k+1/2)^{1/2}.$$

Thus, for  $k > \frac{\sqrt{17}-1}{4} \approx 0.78$ , we obtain  $\frac{f(k+1/2)}{f(k)} > 1$ . For  $k = 1/2$  we also have  $\frac{f(k+1/2)}{f(k)} = \frac{\pi}{2} > 1$ .

Now, it suffices to show that whenever  $S_1 \subseteq S_2$ , then  $k_1 \leq k_2$  (equivalently  $m_1 + l_1 \leq m_2 + l_2$ ). Of course,  $m_1 \leq m_2$ . Next, observe that if  $S_1 \subseteq S_2$ , then also  $l_1 \leq l_2$ . Indeed, the number of leaves in a tree is the cardinality of its largest antichain. If  $S_1$  has  $l_1$  leaves and  $S_1 \subseteq S_2$ , then  $S_2$  needs to contain an antichain of cardinality  $l_1$  as a subposet, which means that its number of leaves has to satisfy  $l_2 \geq l_1$ . Together we get  $m_1 + l_1 \leq m_2 + l_2$ .  $\square$

**Theorem 4.** *Let  $S_1, S_2$  be rooted trees such that  $S_1 \subseteq S_2$ . Then for sufficiently large  $n$*

$$\frac{g_{\mathcal{B}_n}(S_1)}{a_{\mathcal{B}_n}(S_1)} \leq \frac{g_{\mathcal{B}_n}(S_2)}{a_{\mathcal{B}_n}(S_2)}.$$

*Proof.* Let  $d_{S_1} = (l_1, u_1, \dots)$ ,  $d_{S_2} = (l_2, u_2, \dots)$ ,  $k_1 = \frac{m_1 + l_1 - 2}{2}$ ,  $k_2 = \frac{m_2 + l_2 - 2}{2}$ . Aiming for a contradiction, assume that  $S_1 \subseteq S_2$  and that there is an increasing sequence  $n_0 < n_1 < n_2 < \dots$  such that  $\frac{g_{\mathcal{B}_n}(S_1)}{a_{\mathcal{B}_n}(S_1)} > \frac{g_{\mathcal{B}_n}(S_2)}{a_{\mathcal{B}_n}(S_2)}$  for all  $n \in \{n_i \mid i \in \mathbb{N}\}$ . Then by Theorem 3

$$\lim_{n \rightarrow \infty} \sqrt{n} \frac{g_{\mathcal{B}_n}(S_1)}{a_{\mathcal{B}_n}(S_1)} = \lim_{n \rightarrow \infty} \sqrt{n} \frac{g_{\mathcal{B}_n}(S_2)}{a_{\mathcal{B}_n}(S_2)} = \frac{\sqrt{2} \cdot \Gamma(k_1 + 1/2)}{\Gamma(k_1)} = \frac{\sqrt{2} \cdot \Gamma(k_2 + 1/2)}{\Gamma(k_2)}.$$

Recall that the function  $f(k) = \frac{\Gamma(k+1/2)}{\Gamma(k)}$  is increasing in  $k$  for  $k \in \{\frac{1}{2}, 1, \frac{3}{2}, 2, \frac{5}{2}, \dots\}$  thus the above equality implies  $k_1 = k_2$ , or equivalently  $m_1 + l_1 = m_2 + l_2$ . By  $S_1 \subseteq S_2$  we have  $l_1 \leq l_2$  and  $m_1 \leq m_2$  (see the proof of Theorem 3), therefore we get  $l_1 = l_2$  and  $m_1 = m_2$ . Thus  $S_1$  and  $S_2$  are isomorphic and  $\frac{g_{\mathcal{B}_n}(S_1)}{a_{\mathcal{B}_n}(S_1)} = \frac{g_{\mathcal{B}_n}(S_2)}{a_{\mathcal{B}_n}(S_2)}$  which is a contradiction.  $\square$

## 6. EMBEDDING DISCONNECTED STRUCTURES IN $\mathcal{B}_n$

In this section we briefly discuss the case of embedding disconnected structures in  $\mathcal{B}_n$ . Note that in this case all the embeddings must be bad (the underlying structure  $T$  has only one maximal element  $\mathbf{1}_T$ ; as long as the induced structure is disconnected, we can be sure that it does not contain the root  $\mathbf{1}_T$ ).

Assume that  $S$  is a forest, *i.e.* a set of rooted trees  $S_1, S_2, \dots, S_r$  ( $r \geq 2$ ) with the degree distribution sequence  $d_S = (l, u, d_2, \dots, d_{m-1})$ . The underlying structure  $T$  is connected, thus  $S_1, S_2, \dots, S_r$  always have a common parent in  $T$ . Let  $\sigma = (\sigma_1, \sigma_2, \dots, \sigma_r)$  be a permutation of the set  $\{1, 2, \dots, r\}$ . Define  $S^{(\sigma)}$  to be a structure constructed as shown in Figure 5 - we add an additional vertex  $\mathbf{1}_{S^{(\sigma)}}$  to  $S$ , which is a common parent of  $S_1, S_2, \dots, S_r$  appearing in the order given by  $\sigma$ . Now, instead of counting the number of embeddings of  $S$  into  $T$  we can simply count the numbers of good embeddings of  $S^{(\sigma)}$  in  $T$  for all permutations  $\sigma$  generating non-isomorphic structures  $S^{(\sigma)}$  and sum them up. Thus,

$$a_{\mathcal{B}_n}(S) = \sum_{\sigma \in \Sigma} g_{\mathcal{B}_n}(S^{(\sigma)}),$$

where  $\Sigma$  is a set of permutations of  $\{1, 2, \dots, r\}$  such that whenever  $\sigma, \tau \in \Sigma$  and  $\sigma \neq \tau$  then  $S^{(\sigma)}$  and  $S^{(\tau)}$  are not isomorphic. Moreover, whenever  $\tau$  is a permutation of  $\{1, 2, \dots, r\}$  and  $\tau \notin \Sigma$  then there exists  $\sigma \in \Sigma$  such that  $S^{(\sigma)}$  and  $S^{(\tau)}$  are isomorphic.

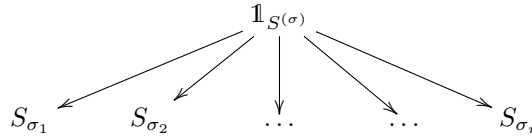


FIGURE 5. The structure of  $S^{(\sigma)}$ ,  $\sigma = (\sigma_1, \sigma_2, \dots, \sigma_r)$ .

Note that the asymptotics of  $g_{\mathcal{B}_n}(S^{(\sigma)})$  is the same for all  $\sigma \in \Sigma$  since the degree distribution sequence of  $S^{(\sigma)}$  is the same for all  $\sigma \in \Sigma$ . It is given by  $d_{S^{(\sigma)}} = (\tilde{d}_0, \tilde{d}_1, \dots, \tilde{d}_{m-1}) = (l, u, \dots, d_{r-1}, d_r + 1, d_{r+1}, \dots, d_{m-1})$ . Therefore, by Theorem 2

$$a_{\mathcal{B}_n}(S) \sim \begin{cases} \frac{m!}{k_1! k_2! \dots k_\ell!} \frac{\tilde{C} \cdot 2^{\frac{6-m-3l}{2}}}{\Gamma(\frac{m+l-2}{2})} \cdot 2^n \cdot n^{\frac{m+l-4}{2}} & \text{if } n \text{ is odd,} \\ 0 & \text{if } n \text{ is even} \end{cases}$$

where  $\ell$  is the number of equivalence classes of the set  $\{S_1, S_2, \dots, S_r\}$  with respect to the equivalence relation of being isomorphic and  $k_1, k_2, \dots, k_\ell$  are the cardinalities of those classes. Here  $\tilde{C} = \prod_{i=3}^{m-1} (C_{i-1})^{\tilde{d}_i}$ . (Note that here we do not consider the case  $m+l-2=0$  from Theorem 2, because by  $r \geq 2$  we always have  $m+l-2 > 0$ .)

## 7. NON-PLANE CASE - EMBEDDINGS IN $\mathcal{V}_n$

In this section we explain how to take advantage of the results obtained for the plane case in order to infer about the asymptotics of good and all embeddings of a rooted tree  $S$  in the family of non-plane binary trees  $\mathcal{V}_n$ .

**Theorem 5.** Consider a rooted tree  $S$  with degree distribution sequence  $d_S = (l, u, d_2, \dots, d_{m-1})$ . The generating function  $A_S(z)$  of the sequence  $a_{\mathcal{V}_n}(S)$ , counting the number of all embeddings of  $S$  into the family  $\mathcal{V}_n$ , is given by

$$A_S(z) = \left( \frac{1}{1 - zV(z)} \right)^{m+l-1} z^{l+u-1} V(z)^{l+u} C_S (1 + o(1)) \quad \text{as } z \rightarrow \pm\rho \quad (4)$$

where  $C_S$  is a constant dependent on the structure of  $S$  and  $V(z)$  is the generating function of the family of non-plane binary trees, satisfying

$$V(z) = z + \frac{z}{2}(V(z)^2 + V(z^2)), \quad (5)$$

which has its dominant singularities at  $z = \pm\rho \approx \pm 0.6346$ .

*Remark 6.* Note that the value for  $\rho$  does not coincide with the one given in [14, Chapter VII], since there the size of a tree corresponds to the number of internal nodes, while we count the total number of nodes. This is the reason why in our model the coefficients  $V_n$  are zero for even  $n$ , which yields a periodicity in the generating function that results in the presence of two dominant singularities. However, the generating function  $N(z)$  of non-plane binary trees where  $z$  solely marks the number of internal vertices can easily be connected with our generating function  $V(z)$  via  $V(z) = zN(z^2)$ . Thus, with the result from [14] that

$$N(z) \sim \frac{1}{\sigma} - a\sqrt{1 - \frac{z}{\sigma}}, \quad \text{as } z \rightarrow \sigma$$

with  $\sigma \approx 0.4027$  and  $a \approx 2.8062$ , we immediately know that there are two dominant singularities of  $V(z) = zN(z^2)$  at  $z = \pm\sqrt{\sigma}$  and we get

$$V(z) = zN(z^2) \sim \pm\sqrt{\sigma} \left( \frac{1}{\sigma} - a\sqrt{2}\sqrt{1 \mp \frac{z}{\sqrt{\sigma}}} \right), \quad \text{as } z \rightarrow \pm\sqrt{\sigma}.$$

Finally, by setting  $\rho = \sqrt{\sigma} \approx 0.6346$  and  $b = a\sqrt{2\sigma} \approx 2.5184$  we have

$$V(z) \sim \pm \left( \frac{1}{\rho} - b\sqrt{1 \mp \frac{z}{\rho}} \right), \quad \text{as } z \rightarrow \pm\rho.$$

*Proof.* Throughout this proof we write  $S_1 \cong S_2$  whenever the structures  $S_1$  and  $S_2$  are isomorphic. This time we introduce a bivariate generating function, where  $z$  still marks the total number of vertices of a tree, while  $u$  is associated with *classes of vertices*. Two vertices  $v, w$  are meant to belong to the same class whenever there exists an isomorphism  $f: T \rightarrow T$  such that  $f(v) = w$ . From [33] we have

$$V(z, u) = zu + \frac{zu}{2}(V(z, u)^2 - V(z^2, u^2) + 2V(z^2, u)). \quad (6)$$

By  $V_u(z, u)$  we denote the derivative of  $V(z, u)$  with respect to  $u$ , i.e.  $V_u(z, u) = \frac{\partial V(z, u)}{\partial u}$ . We proceed as in the plane case and start with recursively defining the generating function  $A_S(z)$  for the number of embeddings of  $S$  into the family  $\mathcal{V}_n$ , when  $S$  is a Motzkin tree:

$$A_S(z) = \begin{cases} V_u(z, 1) & \text{if } S = \bullet \\ \frac{zV(z)}{1-zV(z)} A_{\tilde{S}}(z) & \text{if } S = (\bullet, \tilde{S}) \\ \frac{z}{(1-zV(z))^2} A_{S_L}(z) A_{S_R}(z) & \text{if } S = (\bullet, S_L, S_R) \text{ and } S_L \not\cong S_R \\ \frac{z}{(1-zV(z))^2} \frac{1}{2} (A_{S_L}(z)^2 + A_{S_L}(z^2)) & \text{if } S = (\bullet, S_L, S_R) \text{ and } S_L \cong S_R \end{cases}$$

The idea of setting up this recursive definition for  $A_S(z)$  is similar to the plane case with the following differences. In the first case, corresponding to embedding a single node, we can mark an arbitrary vertex class, instead of an arbitrary vertex, since there might be some non-trivial isomorphisms that would lead to multiple countings of the same embedding. Furthermore, the paths of left-or-right trees from the previous section, yielding a factor  $\frac{1}{1-2zB(z)}$ , are now replaced by paths of trees where we do not distinguish between the left-or-right order, since we are in the non-plane setting. Thus, these paths give a factor  $\frac{1}{1-zV(z)}$ . Finally, in the case when the Motzkin

tree starts with a binary root, we have to distinguish between the cases whether the two attached trees are isomorphic or not. The non-isomorphic case works analogously to its plane version, while in the isomorphic case we have to eliminate potential double-countings by using the same idea as for Equation (5). We do not have to solve the recursion for  $A_S(z)$  explicitly, since we are solely interested in the asymptotic behaviour of its coefficients and it is easy to see that asymptotically the contribution of the term  $A_{S_L}(z^2)$  is negligible. Since  $\rho < 1$  the function  $A_S(z^2)$  is analytic at  $z = \rho$ . Thus,  $[z^n]A_S(z^2) < (\rho + \varepsilon)^{-n}$ , which is exponentially smaller than  $C\rho^{-n}n^\beta = [z^n]A_S(z)$ .

Thus, by iterating we obtain

$$A_S(z) \sim \left( \frac{z}{(1 - zV(z))^2} \right)^{l-1} V_u(z, 1)^l \left( \frac{zV(z)}{1 - zV(z)} \right)^u \left( \frac{1}{2} \right)^s, \quad \text{as } z \rightarrow \rho,$$

where  $l$  denotes the number of leaves,  $u$  the number of unary nodes and  $s$  the number of symmetry nodes in  $S$  (a symmetry node is a parent of two isomorphic subtrees). An analogous expansion holds for  $z \rightarrow -\rho$  (only the  $1 - zV(z)$  in the denominators must be replaced by  $1 + zV(z)$ ).

In the general case where  $S$  is an arbitrary non-plane tree, *i.e.* a Pólya tree, we proceed as in the previous section and consider the embeddings of all non-plane unary-binary trees obtained by replacing  $d$ -ary nodes with  $d \geq 3$  together with their children by binary trees with  $d$  leaves. Thus, again taking into account that there are  $m - l - u$  binary nodes that were already there before the replacement (as binary or  $d$ -ary nodes with  $d \geq 3$ ) and  $2l + u - m - 1$  auxiliary binary nodes that were introduced by the replacement, we get

$$A_S(z) \sim \left( \frac{z}{1 - zV(z)} \right)^{2l+u-m-1} \left( \frac{z}{(1 - zV(z))^2} \right)^{m-l-u} V_u(z, 1)^l \left( \frac{zV(z)}{1 - zV(z)} \right)^u C_S, \quad \text{as } z \rightarrow \rho, \quad (7)$$

and the analogous expansion for  $z \rightarrow -\rho$ . The constant  $C_S$  arises from the isomorphisms and reads as

$$C_S = \sum_{\substack{t \in \mathcal{M}_S \\ s \text{ symmetry node of } t}} \left( \frac{1}{2} \right)^s, \quad (8)$$

where  $\mathcal{M}_S$  denotes the set of all non-plane unary-binary trees obtained from  $S$  by replacing the  $d$ -ary nodes with non-plane binary trees with  $d$  leaves for  $d \geq 3$ . Differentiating Equation (6) with respect to  $u$  and plugging  $u = 1$  yields

$$V_u(z, 1) = \frac{V(z)}{1 - zV(z)}.$$

Finally, substituting this expression for  $V_u(z, 1)$  in Equation (7) yields the desired result. Note that the asymptotic equivalence (7), or (4) respectively, is also true for the case when  $S$  is a single node, *i.e.*  $l = 1$  and  $u = s = 0$ .  $\square$

**Theorem 6.** *Consider a rooted tree  $S$  with degree distribution sequence  $d_S = (l, u, d_2, \dots, d_{m-1})$ . The asymptotics of the number of all embeddings of  $S$  into  $\mathcal{V}_n$  is given by*

$$a_{\mathcal{V}_n}(S) \sim \frac{2C_S b^{-m-l+1} \rho^{-m-l}}{\Gamma(\frac{m+l-1}{2})} \cdot \rho^{-n} \cdot n^{\frac{m+l-3}{2}}$$

for  $n$  being odd and  $a_{\mathcal{V}_n}(S) = 0$  for  $n$  being even. The asymptotics of the number of good embeddings of  $S$  into  $\mathcal{V}_n$  is given by

$$g_{\mathcal{V}_n}(S) \sim \begin{cases} \frac{2C_S b^{-m-l+2} \rho^{-m-l+1}}{\Gamma(\frac{m+l-2}{2})} \cdot \rho^{-n} \cdot n^{\frac{m+l-4}{2}} & \text{if } m+l-2 > 0 \\ \frac{b}{\sqrt{\pi}} \cdot \rho^{-n} \cdot n^{-3/2} & \text{if } m+l-2 = 0, \end{cases}$$

for  $n$  being odd and  $g_{\mathcal{V}_n}(S) = 0$  for  $n$  being even. Here  $b \approx 2.5184$ ,  $\rho \approx 0.6346$  and the constant  $C_S$ , given in (8), depends on the structure of  $S$ .

*Proof.* First, note that  $V(\rho) \sim \frac{1}{\rho}$ , which was already outlined in Remark 6. Therefore, the dominant part of the asymptotics of the coefficients of  $A_S(z)$  comes from the factors  $\frac{1}{1-zV(z)}$ , which give

$$\frac{1}{1-zV(z)} \sim \frac{1}{\rho b \sqrt{1-\frac{z}{\rho}}} \quad \text{for } z \rightarrow \rho.$$

The result for  $a_{\mathcal{V}_n}(S)$  follows immediately by use of Lemma 1. As in the plane case, the generating function  $G_S(z)$  for the good embeddings just differs from  $A_S(z)$  by a factor  $(1-zV(z))$  and thus, the asymptotic behaviour of its coefficients can be determined analogously. Recall that  $m+l-2=0$  represents the case where  $S$  is a single vertex. The number of good embeddings is therefore just the cardinality of  $\mathcal{V}_n$  (see Remark 6).  $\square$

Now we can formulate a corollary analogous to Corollary 2 from the plane case.

**Corollary 3.** *Consider a rooted tree  $S$  with degree distribution sequence  $d_S = (l, u, d_2, \dots, d_{m-1})$ . Let  $k = \frac{m+l-2}{2}$  and let  $n$  be odd. The asymptotic ratio of the number of good embeddings of  $S$  into  $\mathcal{V}_n$  to the number of all embeddings into  $\mathcal{V}_n$  is given by*

$$\frac{g_{\mathcal{V}_n}(S)}{a_{\mathcal{V}_n}(S)} \sim \begin{cases} \frac{\Gamma(k+1/2)}{\Gamma(k)} \frac{b\rho}{\sqrt{n}} & \text{if } k > 0, \\ 1/n & \text{if } k = 0. \end{cases}$$

**Theorem 7.** *Let  $S_1, S_2$  be rooted trees such that  $S_1 \subseteq S_2$ . Then*

$$\lim_{n \rightarrow \infty} \sqrt{n} \frac{g_{\mathcal{V}_n}(S_1)}{a_{\mathcal{V}_n}(S_1)} \leq \lim_{n \rightarrow \infty} \sqrt{n} \frac{g_{\mathcal{V}_n}(S_2)}{a_{\mathcal{V}_n}(S_2)}.$$

*Proof.* By Corollary 3 we get that for any  $S$  with  $d_S = (l, u, d_2, \dots, d_{m-1})$

$$\lim_{n \rightarrow \infty} \sqrt{n} \frac{g_{\mathcal{V}_n}(S)}{a_{\mathcal{V}_n}(S)} = \frac{\Gamma(k+1/2)}{\Gamma(k)} b\rho$$

where  $k = \frac{m+l-2}{2} > 0$ . The rest of the proof is then analogous to the proof of Theorem 3.  $\square$

**Corollary 4.** *Let  $S_1, S_2$  be rooted trees such that  $S_1 \subseteq S_2$ . Then for sufficiently large  $n$*

$$\frac{g_{\mathcal{V}_n}(S_1)}{a_{\mathcal{V}_n}(S_1)} \leq \frac{g_{\mathcal{V}_n}(S_2)}{a_{\mathcal{V}_n}(S_2)}.$$

(Compare Theorem 4 and its proof.)

Now, let us comment on embedding disconnected structures in a non-plane case. Let  $S$  be a forest, *i.e.* a set of rooted trees  $S_1, S_2, \dots, S_r$ ,  $r \geq 2$ . Again, instead of counting all embeddings of  $S$  into  $\mathcal{V}_n$ , we can count the good embeddings of  $\tilde{S}$  in  $\mathcal{V}_n$ , where  $\tilde{S}$  is a forest  $S$  with an additional common parent that clips together all  $S_i$ 's. Note that in the non-plane case the order of  $S_i$ 's does not matter, thus we simply have

$$a_{\mathcal{V}_n}(S) = g_{\mathcal{V}_n}(\tilde{S}).$$

## 8. PLANTED PLANE CASE - EMBEDDINGS IN $\mathcal{T}_n$

In this section we extend the results from plane binary trees to planted plane trees, *i.e.* to rooted trees where each internal node can have arbitrarily many child-nodes and the order of the subtrees is important. The structures that we embed are as well planted plane trees, and therefore every such a tree  $S$  is of the form  $S = (\bullet, S_1, \dots, S_k)$ , where the  $S_i$ 's denote the subtrees that are attached to the root. The following lemma contains the construction of the generating function  $A_S(z)$  of all embeddings of the tree  $S$  in the family  $\mathcal{T}_n$  of planted plane trees of size  $n$ .

**Lemma 3.** *The generating function  $A_S(z)$  of all embeddings of  $S = (\bullet, S_1, \dots, S_k)$  into the family  $\mathcal{T}_n$  of planted plane trees of size  $n$  can be recursively specified as*

$$A_S(z) = \begin{cases} zT'(z) = \frac{T(z)(1-T(z))}{1-2T(z)} & \text{if } k = 0 \\ \frac{T(z)}{1-2T(z)} A_{S_1}(z) & \text{if } k = 1 \\ \frac{T(z)^2}{(1-2T(z))^2(1-T(z))} A_{S_1}(z) A_{S_2}(z) & \text{if } k = 2 \\ \frac{T(z)}{(1-2T(z))^2} \left( \frac{1-2T(z)}{1-T(z)} A_{S_1}(z) A_{S_{2,k}}(z) \right. \\ \quad \left. + \frac{T(z)(1-2T(z))}{(1-T(z))^2} A_{S_{1,k-1}}(z) A_{S_k}(z) \right. \\ \quad \left. + \left( \frac{1-2T(z)}{1-T(z)} \right)^2 (A_{S_{1,2}}(z) A_{S_{3,k}}(z) + \dots + A_{S_{1,k-2}}(z) A_{S_{k-1,k}}(z)) \right) & \text{if } k > 2 \end{cases} \quad (9)$$

where  $T(z)$  denotes the generating function of the family of planted plane trees, *i.e.*

$$T(z) = \frac{1 - \sqrt{1-4z}}{2} = C_0 z + C_1 z^2 + C_2 z^3 + \dots,$$

and  $S_{i,j}$  denotes the tree  $S_{i,j} = (\bullet, S_i, \dots, S_j)$  that consists of a root to which the  $j-i+1$  subtrees  $S_i, \dots, S_j$  are attached (in that order).

*Proof.* The case  $k = 0$  is equivalent to the binary cases, and corresponds to marking an arbitrary node in the tree  $T$ . Differentiating both sides of the specification  $T(z) = \frac{z}{1-T(z)}$  of planted plane trees with respect to  $z$  and solving for  $T'(z)$  yields the equality

$$zT'(z) = \frac{z}{1-2T(z)} = \frac{T(z)(1-T(z))}{1-2T(z)}.$$

Now, let us continue with the proof of the recurrence for the case  $k > 2$ . In order to do so let us observe Figure 6 that visualizes how an embedding of a tree  $S$  in a tree  $T$  can be constructed. We start with a path of left-or-right plane trees, followed by the embedded root node. Attached to the root node there is another such path, ending with the so-called “splitting node”. To the left and the right of this second path there can of course be several planted plane trees attached to the embedded root node, which themselves do not contain any embedded vertices. The two paths that are separated by the embedded root node contribute a factor  $\left( \frac{1}{1 - \frac{z}{(1-T(z))^2}} \right)^2$ ,

which can be simplified to  $\left( \frac{1-T(z)}{1-2T(z)} \right)^2$  by means of the functional equation  $T(z) = \frac{z}{1-T(z)}$ . The root node together with the two sequences of planted plane trees that can be attached to the left or to the right of the path give a factor  $\frac{z}{(1-T(z))^2} = \frac{T(z)}{1-T(z)}$ .

The splitting node can as well have a sequence of plane trees attached, that do not contain any embedded nodes, yielding a factor  $\frac{1}{1-T(z)}$ , but at some point there has to appear the first plane tree that contains some embedded nodes (pictured in blue in Figure 6). All subtrees attached to the splitting node that are to the right of this blue one are comprised in one plane tree (pictured in green in Figure 6). Now we have to distinguish between the cases where a different number of the subtrees  $S_1, \dots, S_k$  are embedded in the left (*i.e.* the blue) subtree, while the remaining ones are embedded in the right (*i.e.* the green) tree. These case distinctions give rise to the recursion (9) for the generating function. The first two summands of the last case in (9), *i.e.* the case  $k > 2$ , represent the cases where one of the  $S_i$ 's is embedded in a separate subtree:



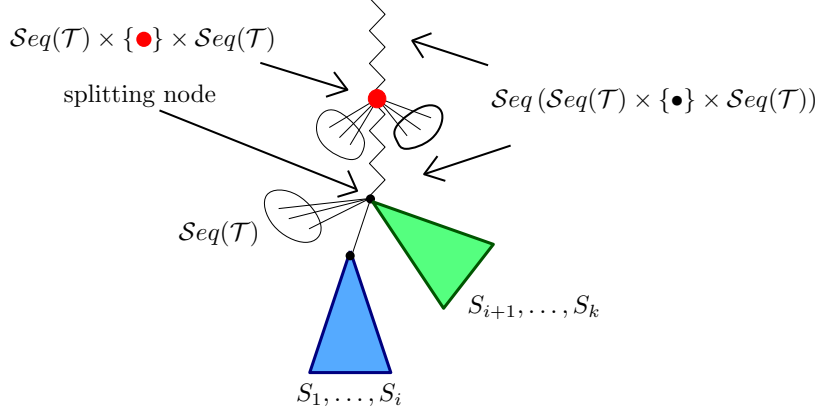


FIGURE 6. Sketch of the principle of embedding a plane tree  $S = (\bullet, S_1, \dots, S_k)$  into the family of planted plane trees. Here the class  $\text{Seq}(\mathcal{T})$  indicates that this part of the structure belongs to this class, i.e., it stands for a sequence of planted plane trees. Likewise,  $\text{Seq}(\text{Seq}(\mathcal{T}) \times \{\bullet\} \times \text{Seq}(\mathcal{T}))$  represents a path of left-or-right planted plane trees.

- Solely  $S_1$  is embedded in the left tree. In this case we count all embeddings of  $S_1$  in the left subtree, giving a factor  $A_{S_1}(z)$ , while in the right subtree we count exclusively the good embeddings of  $S_{2,k} = (\bullet, S_2, \dots, S_k)$ , since the splitting node has to be the embedded root of  $S_{2,k}$  in order to prevent multiple embeddings of the root. We already know that the generating function of good embeddings is obtained from the generating function of all embeddings by multiplication with  $\frac{1-2T(z)}{1-T(z)}$  (corresponding to 1 divided by the generating function of the starting path) and thus we get the factor  $\frac{1-2T(z)}{1-T(z)} A_{S_{2,k}}(z)$ .
- Solely  $S_k$  is embedded in the right tree. Here we count the good embeddings of  $S_{1,k-1}$  in the left tree, as this is general necessary for all cases where we consider more than just one of the  $S_i$ 's to be embedded in the same subtree. However, in this case we have to count only the bad embeddings of  $S_k$  in the right tree, since no node of  $S$  can be embedded into the splitting node, except the root of  $S$ , but then the embedding of  $S_k$  is still a bad embedding into the green tree. Altogether this yields the factor  $\frac{1-2T(z)}{1-T(z)} \frac{T(z)}{1-T(z)} A_{S_{1,k-1}}(z) A_{S_k}(z)$ .

In all other cases where we embed at least two of the subtrees  $S_1, \dots, S_k$  in both the left and the right (*i.e.* the blue and the green) subtree, we consider good embeddings for both subtrees (the blue and the green one, which are then becoming two trees where we embed  $S_{1,i}$  and  $S_{i+1,k}$ , respectively) yielding a factor  $\left(\frac{1-2T(z)}{1-T(z)}\right)^2$ . Together with the factors from the two paths, the embedded root and the sequence of plane trees we get the desired coefficients.

The cases  $k = 1$  and  $k = 2$  can be treated in the exact same way as we just did for  $k > 2$ . However, note that in the case  $k = 1$  the green (*i.e.* the right) tree and two of the sequences of planted plane trees are merged such that we end up with one path, the embedded root of  $S$  together with its two sequences of planted plane trees and finally the attached blue tree that contains the embedding of the only subtree  $S_1$ . This yields the factor

$$\frac{1-T(z)}{1-2T(z)} \frac{T(z)}{1-T(z)} A_{S_1}(z).$$

In the case  $k = 2$  we have the pre-factor  $\frac{T(z)}{(1-2T(z))^2}$  that covers the two paths, the embedded root node with its attached sequences of plane trees and the sequence of plane trees that is attached to the splitting node. Now there is just one splitting option:  $S_1$  has to be embedded in the left tree, where we consider all embeddings, and  $S_2$  has to be embedded in the right tree, where we solely count the bad embeddings of  $S_2$ , since the splitting node must not be an embedded node. It is

easy to verify that this case gives the factor

$$\frac{T(z)^2}{(1-2T(z))^2(1-T(z))} A_{S_1}(z) A_{S_2}(z). \quad \square$$

*Remark 7.* Note that for the cases  $k = 0, 1, 2$  the generating function  $A_S(z)$  of all embeddings of  $S = (\bullet, S_1, \dots, S_k)$  into the family  $\mathcal{T}_n$  of planted plane trees of size  $n$  given in (9) is of the form  $f_k(T) \cdot A_{S_1}(z) \dots A_{S_k}(z)$ , where  $f_k(T)$  is a function that depends only on  $T(z)$  and on the size  $k$  of  $S$ , but not on the specific shape of  $S$ . We want to emphasize that, by digging into the structure of  $S$  and by recursive application of the formulas given in (9), it follows that  $A_S(z)$  is in fact of the form

$$A_S(z) = f(T) \cdot A_{S_1}(z) \cdots A_{S_k}(z),$$

for arbitrary  $S = (\bullet, S_1, \dots, S_k)$ .

Now, we are in the position to obtain the asymptotic number of all and good embeddings of a given plane tree  $S$  in the family of planted plane trees.

**Theorem 8.** *Consider a rooted tree  $S$  of size  $m$  with degree distribution sequence  $d_S = (l, d_1, d_2, \dots, d_{m-1})$ . Let  $C = \prod_{i=1}^{m-1} (C_{i-1})^{d_i}$ . The asymptotics of the number of all embeddings of  $S$  into  $\mathcal{T}_n$  is given by*

$$a_{\mathcal{T}_n}(S) \sim \frac{C \cdot (\frac{1}{2})^{m+l}}{\Gamma(\frac{m+l-1}{2})} \cdot 4^n \cdot n^{\frac{m+l-3}{2}}.$$

The asymptotics of the number of good embeddings of  $S$  into  $\mathcal{T}_n$  is given by

$$g_{\mathcal{T}_n}(S) \sim \frac{2C \cdot (\frac{1}{2})^{m+l}}{\Gamma(\frac{m+l-2}{2})} \cdot 4^n \cdot n^{\frac{m+l-4}{2}}.$$

*Proof.* Triggered by the observation in Remark 7, let us set

$$f_1(z) = \frac{1}{2(1-2T(z))}, \quad \text{and} \quad f_k(z) = \frac{A_S(z)}{\prod_{i=1}^k A_{S_i}(z)} \quad \text{for } k > 1. \quad (10)$$

Then (9) immediately gives  $f_2(z) = T(z)^2 / ((1-2T(z))^2(1-T(z)))$ .

Next, consider the last equation of (9) (the case  $k \geq 3$ ) and observe that all generating functions on the right-hand side which are associated with a composite structure are of the form  $A_{S_{i,j}}(z)$ , where the root of  $S_{i,j}$  has degree at least two. Thus, dividing the equation by  $\prod_{i=1}^k A_{S_i}(z)$  (and cancelling out all single  $A_{S_j}(z)$ ) yields only quotients which can be readily turned into  $f_\ell(z)$  with suitable choices of  $\ell$ , because the case  $\ell = 1$  does not appear here. A straight-forward simplification then gives

$$f_k(z) = \frac{T(z)}{(1-T(z))^2} \sum_{j=1}^{k-1} f_j(z) f_{k-j}(z) \quad \text{for } k \geq 3. \quad (11)$$

Both sides of this equation tend to infinity, as  $z \rightarrow 1/4$ , and we need their singular behaviour for our analysis of  $A_S(z)$ . Hence, we set  $g_k(z) = (1-2T(z))^k f_k(z)$  for  $k \geq 1$ . Plugging this into (11) we observe that  $g_k(z)$  satisfies the same recurrence as  $f_k(z)$ , but with the initial values  $g_1(z) = 1/2$  and  $g_2(z) = T(z)^2 / (1-T(z))$ . As  $T(1/4) = 1/2$ , the functions  $g_k(z)$  are regular at  $z = 1/4$ . By evaluating the recurrence at  $z = 1/4$  and setting  $h_k := 2g_k(1/4)$ , we get a recurrence for  $h_k$ , which is in fact already valid for  $k \geq 2$ :

$$h_1 = 1 \quad \text{and} \quad h_k = \sum_{j=1}^{k-1} h_j h_{k-j} \quad \text{for } k \geq 2.$$

This is exactly the recurrence for the Catalan numbers, and thus,  $h_k = C_{k-1}$ .

Hence, for  $z \rightarrow 1/4$  and  $k \geq 2$  we have

$$f_k(z) \sim \frac{1}{2} C_{k-1} (1-4z)^{-k/2},$$

which implies that as  $z \rightarrow 1/4$  we have

$$A_S(z) \sim \frac{C^{k-1}}{2} (1-4z)^{-k/2} A_{S_1}(z) \dots A_{S_k}(z) = \left( \prod_{i=1}^{m-1} \left( \frac{C_{i-1}}{2} (1-4z)^{-i/2} \right)^{d_i} \right) (A_{\bullet}(z))^l,$$

where  $S = (\bullet, S_1, \dots, S_k)$ ,  $d_i$  denotes the number of nodes with out-degree  $i$ ,  $l$  denotes the number of leaves, *i.e.*  $l = d_0$ , and the equation follows from recursively going into the subtrees  $S_1, \dots, S_k$  and using (10) until one encounters a leaf of  $S$ . Then each leaf yields a factor  $A_{\bullet}(z)$ . Using the equality  $A_{\bullet}(z) = zT'(z) \sim \frac{1}{2}(1-4z)^{-1/2}$ , which follows from (9) and the singular expansion of  $T(z)$ , we get for  $z \rightarrow \frac{1}{4}$

$$A_S(z) \sim \left( \prod_{i=1}^{m-1} \left( \frac{C_{i-1}}{2} \right)^{d_i} \right) (1-4z)^{-(l+\sum_{i=1}^{m-1} id_i)/2} \left( \frac{1}{4} \right)^l. \quad (12)$$

Note that  $\sum_{i=1}^{m-1} id_i = m-1$ , since every vertex with out-degree  $i$  is counted exactly  $i$  times and thus, we simply obtain the total number of nodes with in-degree greater than zero (*i.e.* all nodes except for the root). We also have  $\sum_{i=1}^{m-1} d_i = m-l$  thus (recall that  $C = \prod_{i=1}^{m-1} (C_{i-1})^{d_i}$ )

$$\left( \frac{1}{4} \right)^l \cdot \prod_{i=1}^{m-1} \left( \frac{C_{i-1}}{2} \right)^{d_i} = C \left( \frac{1}{2} \right)^{m+l}.$$

Finally, Lemma 1 gives

$$a_{\mathcal{T}_n}(S) \sim \frac{C \cdot \left(\frac{1}{2}\right)^{m+l}}{\Gamma\left(\frac{m+l-1}{2}\right)} \cdot 4^n \cdot n^{\frac{m+l-3}{2}}.$$

The generating function of the number of good embeddings can be derived from the generating function  $A_S(z)$  by multiplication by the factor  $\frac{1-2T(z)}{1-T(z)}$ . This factor is responsible for getting rid of the path of trees which could appear above embedded root of  $S$  when we were considering all embeddings. Thus we have  $G_S(z) = \frac{1-2T(z)}{1-T(z)} A_S(z)$ . Noticing that

$$\frac{1-2T(z)}{1-T(z)} = 2 \frac{\sqrt{1-4z}}{1+\sqrt{1-4z}},$$

using (12) and applying Lemma 1 yields the desired result.  $\square$

**Corollary 5.** *Consider a rooted tree  $S$  of size  $m$  with  $l$  leaves. The asymptotic ratio of the number of good embeddings of  $S$  into  $\mathcal{T}_n$  to the number of all embeddings in  $\mathcal{T}_n$  is given by*

$$\frac{g_{\mathcal{T}_n}(S)}{a_{\mathcal{T}_n}(S)} \sim \begin{cases} \frac{2\Gamma\left(\frac{m+l-1}{2}\right)}{\Gamma\left(\frac{m+l-2}{2}\right)\sqrt{n}} & \text{if } m > 1, \\ 1/n & \text{if } m = 1. \end{cases}$$

**Theorem 9.** *Let  $S_1, S_2$  be rooted trees such that  $S_1 \subseteq S_2$ . Then*

$$\lim_{n \rightarrow \infty} \sqrt{n} \frac{g_{\mathcal{T}_n}(S_1)}{a_{\mathcal{T}_n}(S_1)} \leq \lim_{n \rightarrow \infty} \sqrt{n} \frac{g_{\mathcal{T}_n}(S_2)}{a_{\mathcal{T}_n}(S_2)}.$$

*Proof.* By Corollary 5 we get that for any  $S$  with  $d_S = (l, u, d_2, \dots, d_{m-1})$

$$\lim_{n \rightarrow \infty} \sqrt{n} \frac{g_{\mathcal{T}_n}(S)}{a_{\mathcal{T}_n}(S)} = \frac{2\Gamma(k+1/2)}{\Gamma(k)}$$

where  $k = \frac{m+l-2}{2} > 0$ . The rest of the proof is then analogous to the proof of Theorem 3.  $\square$

**Corollary 6.** *Let  $S_1, S_2$  be rooted trees such that  $S_1 \subseteq S_2$ . Then for sufficiently large  $n$*

$$\frac{g_{\mathcal{T}_n}(S_1)}{a_{\mathcal{T}_n}(S_1)} \leq \frac{g_{\mathcal{T}_n}(S_2)}{a_{\mathcal{T}_n}(S_2)}.$$

(Compare Theorem 4 and its proof.)

## 9. DISCUSSION

We proved that the ratio of the number of good embeddings to the number of all embeddings of a given tree  $S = (\bullet, S_1, \dots, S_k)$  into the families of trees  $\mathcal{B}_n, \mathcal{V}_n, \mathcal{T}_n$  is asymptotically of the same order for all the three considered families of trees, namely plane binary trees, non-plane binary trees and planted plane trees. Thereby we extended the results of Kubicki *et al.* [27, 28] and Georgiou [20]. We expect that this result will also hold for the family of Pólya trees, which are the closest counterpart to posets that admit a (rooted) treelike shape, *i.e.* they have a single maximal element. In principle, the approach that we used within this paper works for embeddings into the family of Pólya trees as well. However, one would have to consider all possible partitions of  $S_1, \dots, S_k$ , as any collection of isomorphic subtrees within  $S_1, \dots, S_k$  admits non-trivial isomorphisms between the  $S_i$ 's, which can get rather involved and is therefore omitted in this work.

## REFERENCES

- [1] Noga Alon and Joel H. Spencer. *The Probabilistic Method*. Wiley-Interscience Series in Discrete Mathematics and Optimization. John Wiley & Sons, Inc., Hoboken, NJ, third edition, 2008.
- [2] Andrei Asinowski, Axel Bacher, Cyril Banderier, and Bernhard Gittenberger. Analytic combinatorics of lattice paths with forbidden patterns, the vectorial kernel method, and generating functions for pushdown automata. *Algorithmica*, 82(3):386–428, 2020.
- [3] Fabrício Siqueira Benevides and Małgorzata Sulkowska. Percolation and best-choice problem for powers of paths. *J. Applied Probability*, 54(2):343–362, 2017.
- [4] Miklós Bóna. *Combinatorics of Permutations*. Discrete Mathematics and its Applications (Boca Raton). CRC Press, Boca Raton, FL, second edition, 2012.
- [5] Frédéric Chyzak, Michael Drmota, Thomas Klausner, and Gerard Kok. The distribution of patterns in random trees. *Combin. Probab. Comput.*, 17(1):21–59, 2008.
- [6] Gwendal Collet, Michael Drmota, and Lukas D. Klausner. Limit laws of planar maps with prescribed vertex degrees. *Combin. Probab. Comput.*, 28(4):519–541, 2019.
- [7] Michael Dairyko, Lara Pudwell, Samantha Tyner, and Casey Wynn. Non-contiguous pattern avoidance in binary trees. *Electron. J. Combin.*, 19(3), 2012.
- [8] Emeric Deutsch. Dyck path enumeration. *Discrete Math.*, 204(1-3):167–202, 1999.
- [9] Michael Drmota. *Random Trees*. Springer, Vienna-New York, 2009.
- [10] Michael Drmota and Bernhard Gittenberger. The distribution of nodes of given degree in random trees. *J. Graph Theory*, 31(3):227–253, 1999.
- [11] Michael Drmota and Benedikt Stuffer. Pattern occurrences in random planar maps. *Statist. Probab. Lett.*, 158, 2020.
- [12] Michael Drmota and Guan-Ru Yu. The number of double triangles in random planar maps. In *29th International Conference on Probabilistic, Combinatorial and Asymptotic Methods for the Analysis of Algorithms*, volume 110 of *LIPICs. Leibniz Int. Proc. Inform. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern*, 2018.
- [13] Thomas S. Ferguson. Who solved the secretary problem? *Statist. Sci.*, 4(3):282–289, 1989.
- [14] Philippe Flajolet and Robert Sedgewick. *Analytic Combinatorics*. Cambridge University Press, 2009.
- [15] Peter R. Freeman. The secretary problem and its extensions: A review. *Int. Stat. Rev.*, 51:189–206, 1983.
- [16] Ragnar Freij and Johan Wästlund. Partially ordered secretaries. *Electron. Commun. Probab.*, 15:504–507, 2010.
- [17] Anton Freund. From Kruskal's theorem to Friedman's gap condition. *Math. Struct. Comp. Sci.*, 30(8):952–975, 2020.
- [18] Bryn Garrod and Robert Morris. The secretary problem on an unknown poset. *Random Struct. Algor.*, 43(4):429–451, 2013.
- [19] Walter Gautschi. Some elementary inequalities relating to the gamma and incomplete gamma function. *J. Math. Phys. Camb.*, 38:77–81, 1959.
- [20] Nicholas Georgiou. Embeddings and other mappings of rooted trees into complete trees. *Order*, 22(3):257–288, 2005.
- [21] John P. Gilbert and Frederick Mosteller. Recognizing the maximum of a sequence. In *Selected Papers of Frederick Mosteller*, pages 355–398. Springer, 2006.
- [22] Bernhard Gittenberger. Nodes of large degree in random trees and forests. *Random Struct. Algor.*, 28(3):374–385, 2006.
- [23] Alexander V. Gnedin. Multicriteria extensions of the best choice problem: Sequential selection without linear order. *Contemp. Math.*, 125:153–172, 1992.
- [24] Andrzej Grzesik, Michał Morayne, and Małgorzata Sulkowska. From directed path to linear order—the best choice problem for powers of directed paths. *SIAM J. Discrete Math.*, 29(1):500–513, 2015.
- [25] Svante Janson, Tomasz Łuczak, and Andrzej Ruciński. *Random Graphs*. Wiley-Interscience Series in Discrete Mathematics and Optimization. Wiley-Interscience, New York, 2000.

- [26] Wojciech Kaźmierczak. The best choice problem for a union of two linear orders with common maximum. *Discrete Appl. Math.*, 161(18):3090–3096, 2013.
- [27] Grzegorz Kubicki, Jenő Lehel, and Michał Morayne. A ratio inequality for binary trees and the best secretary. *Combin. Probab. Comput.*, 11(2):149–161, 2002.
- [28] Grzegorz Kubicki, Jenő Lehel, and Michał Morayne. An asymptotic ratio in the complete binary tree. *Order*, 20(2):91–97, 2003.
- [29] Grzegorz Kubicki, Jenő Lehel, and Michał Morayne. Counting chains and antichains in the complete binary tree. *Ars Combinatoria*, 79, 2006.
- [30] Grzegorz Kubicki and Michał Morayne. Graph-theoretic generalization of the secretary problem: The directed path case. *SIAM J. Discrete Math.*, 19(3):622–632, 2005.
- [31] Małgorzata Kuchta, Michał Morayne, and Jarosław Niemiec. Counting embeddings of a chain into a tree. *Discrete Math.*, 297(1-3):49–59, 2005.
- [32] Małgorzata Kuchta, Michał Morayne, and Jarosław Niemiec. Counting embeddings of a chain into a binary tree. *Ars Combinatoria*, 91, 2009.
- [33] Xueliang Li, Yiyang Li, and Yongtang Shi. The asymptotic number of non-isomorphic rooted trees obtained by rooting a tree. *J. Math. Anal. Appl.*, 434(1):1–11, 2016.
- [34] Dennis V. Lindley. Dynamic programming and decision theory. *Appl. Stat. - J. Roy. St. C*, 10(1):39–51, 1961.
- [35] Michał Morayne. Partial-order analogue of the secretary problem - the binary tree case. *Discrete Math.*, 184(1-3):165–181, 1998.
- [36] Konstantinos Panagiotou and Makrand Sinha. Vertices of degree  $k$  in random unlabeled trees. *J. Graph Theory*, 69(2):114–130, 2012.
- [37] John Preater. The best-choice problem for partially ordered objects. *Oper. Res. Lett.*, 25(4):187–190, 1999.
- [38] Robert W. Robinson and Allen J. Schwenk. The distribution of degrees in a large random tree. *Discrete Math.*, 12(4):359–372, 1975.
- [39] Eric S. Rowland. Pattern avoidance in binary trees. *J. Combin. Theory Ser. A*, 117(6):741–758, 2010.
- [40] Stephen M. Samuels. Secretary problems. In *Handbook of sequential analysis*, volume 118 of *Statist. Textbooks Monogr.*, pages 381–405. Dekker, New York, 1991.
- [41] Wolfgang Stadje. Efficient stopping of a random series of partially ordered points. *Multiple Criteria Decision Making Theory and Application. Lect. Notes Econ. Math.*, 177:430–447, 1980.
- [42] Małgorzata Sulkowska. The best choice problem for upward directed graphs. *Discrete Optim.*, 9(3):200–204, 2012.

DEPARTMENT OF DISCRETE MATHEMATICS AND GEOMETRY, TECHNISCHE UNIVERSITÄT WIEN, WIEDNER HAUPTSTRASSE 8-10/104, A-1040 WIEN, AUSTRIA.

*Email address:* gittenberger@dmg.tuwien.ac.at

DEPARTMENT OF FUNDAMENTALS OF COMPUTER SCIENCE, WROCLAW UNIVERSITY OF SCIENCE AND TECHNOLOGY, UL. WYBRZEŻE WYSPIAŃSKIEGO 27, 50-370, WROCLAW, POLAND.

*Email address:* zbigniew.golebiewski@pwr.edu.pl

DEPARTMENT OF DISCRETE MATHEMATICS AND GEOMETRY, TECHNISCHE UNIVERSITÄT WIEN, WIEDNER HAUPTSTRASSE 8-10/104, A-1040 WIEN, AUSTRIA.

*Email address:* isabella.larcher@tuwien.ac.at

DEPARTMENT OF FUNDAMENTALS OF COMPUTER SCIENCE, WROCLAW UNIVERSITY OF SCIENCE AND TECHNOLOGY, UL. WYBRZEŻE WYSPIAŃSKIEGO 27, 50-370, WROCLAW, POLAND.

UNIVERSITÉ CÔTE D'AZUR, CNRS, INRIA, I3S, FRANCE.

*Email address:* malgorzata.sulkowska@pwr.edu.pl