# ON THE SHAPE OF RANDOM PÓLYA STRUCTURES

BERNHARD GITTENBERGER, EMMA YU JIN AND MICHAEL WALLNER

Abstract. Panagiotou and Stufler recently proved an important fact on their way to establish the scaling limits of random Pólya trees: a uniform random Pólya tree of size $n$ consists of a conditioned critical Galton-Watson tree $C_n$ and many small forests, where with probability tending to one, as $n$ tends to infinity, any forest $F_n(v)$, that is attached to a node $v$ in $C_n$, is maximally of size $|F_n(v)| = O(\log n)$. Their proof used the framework of a Boltzmann sampler and deviation inequalities.

In this paper, first, we employ a unified framework in analytic combinatorics to prove this fact with additional improvements for $|F_n(v)|$, namely $|F_n(v)| = \Theta(\log n)$. Second, we give a combinatorial interpretation of the rational weights of these forests and the defining substitution process in terms of automorphisms associated to a given Pólya tree. Third, we derive the limit probability that for a random node $v$ the attached forest $F_n(v)$ is of a given size. Moreover, structural properties of those forests like the number of their components are studied. Finally, we extend all results to other Pólya structures.

## 1. Introduction

In this section we first recall the asymptotic estimation of the number of Pólya trees with $n$ nodes from the literature [11, 12, 14]. Then, we briefly discuss simply generated trees and give finally an outline of the paper.

1.1. **Pólya trees.** A *Pólya tree* is a rooted unlabelled non-plane tree. The *size* of a tree is given by the number of its nodes. We denote by $t_n$ the number of Pólya trees of size $n$ and by $T(z) = \sum_{n\geq 1} t_n z^n$ the corresponding ordinary generating function. By Pólya's enumeration theory [14], the generating function $T(z)$ satisfies

$$T(z) = z \exp\left(\sum_{i=1}^{\infty} \frac{T(z^i)}{i}\right). \tag{1.1}$$

The first few terms of $T(z)$ are then

$$T(z) = z + z^2 + 2z^3 + 4z^4 + 9z^5 + 20z^6 + 48z^7 + 115z^8 + 286z^9 + 719z^{10} + \cdots, \tag{1.2}$$

(see OEIS A000081, [15]). By differentiating both sides of (1.1) with respect to $z$, one can derive a recurrence relation of $t_n$ (see [11, Chapter 29] and [12]), which is

$$t_n = \frac{1}{n-1} \sum_{i=1}^{n-1} t_{n-i} \sum_{m|i} m t_m, \quad \text{for } n > 1, \text{ and } t_1 = 1.$$

Corresponding author email: gittenberger@dmg.tuwien.ac.at. This work was supported by the Austrian Research Fund (FWF), grant SFB F50-03.

A preliminary version of this paper was published in the Proceedings of ANALCO 2017.

Pólya [14] showed that the radius of convergence $\rho$ of $T(z)$ satisfies $0 < \rho < 1$ and that $\rho$ is the unique singularity on the circle of convergence. Subsequently, Otter [12] proved that $T(\rho) = 1$ as well as the singular expansion

$$T(z) = 1 - b\,(\rho - z)^{1/2} + c(\rho - z) + \mathcal{O}\left((\rho - z)^{3/2}\right), \tag{1.3}$$

where $\rho \approx 0.3383219$, $b \approx 2.68112$ and $c = b^2/3 \approx 2.39614$. Moreover, he derived

$$t_n = \frac{b\sqrt{\rho}}{2\sqrt{\pi}} \frac{\rho^{-n}}{\sqrt{n^3}} \left(1 + \mathcal{O}\left(\frac{1}{n}\right)\right).$$

1.2. **Relation between Pólya and simply generated trees.** We will see that $T(z)$ is connected with the *exponential generating function* of Cayley trees. "With a minor abuse of notation" (*cf.* [9, Ex. 10.2]), Cayley trees belong to the class of *simply generated trees*. Simply generated trees have been introduced by Meir and Moon [10] to describe a weighted version of rooted trees. They are defined by the functional equation

$$y(z) = z\Phi(y(z)), \qquad \text{with} \quad \Phi(z) = \sum_{j\geq 0} \phi_j\, z^j, \quad \phi_j \geq 0. \tag{1.4}$$

The power series $y(x) = \sum_{n\geq 1} y_n x^n$ has nonnegative coefficients and is the generating function of *weighted* simply generated trees. One usually assumes that $\phi_0 > 0$ and $\phi_j > 0$ for some $j \geq 2$ to exclude the trivial cases. In particular, in the above-mentioned sense, *Cayley trees* can be seen as simply generated trees which are characterized by $\Phi(z) = \exp(z)$. It is well known that the number of rooted Cayley trees of size $n$ is given by $n^{n-1}$. Let

$$C(z) = \sum_{n\geq 0} n^{n-1} \frac{z^n}{n!}, \tag{1.5}$$

be the associated exponential generating function. Then, by construction it satisfies $C(z) = z\exp(C(z))$. In contrast, Pólya trees are not simply generated (see [5] for a simple proof of this fact). Note that though $T(z)$ and $C(z)$ are closely related, Pólya trees are not related to Cayley trees in a strict sense, but rather to a certain class of weighted *unlabeled non-plane* trees, which will be called $C$-trees in the sequel and have the *ordinary* generating function $C(z)$.

Informally speaking, a Pólya tree is constructed from a $C$-tree where to each node a forest is attached. The family where the forests are taken from will be called $D$-forests. So, a Pólya tree is (as above "with a minor abuse of notation") a simply generated tree with small decorations on each of its vertices. This follows from a limit theorem by Panagiotou and Stufler [13] where trees are seen as random metric spaces which converge to a limit space, the so-called scaling limit, when suitably rescaled. Their proof uses advanced probabilistic methods. A goal of this paper is to understand this limit from an analytic combinatorics [6] point of view and to offer a somewhat more elementary approach to this limit theorem. We will not reprove the complete limit theorem, but we set up a description in terms of generating functions which exhibits combinatorially the above mentioned relation between Pólya trees and simply generated trees. Moreover, this description allows a detailed analysis of the decorations and leads to some extensions of Panagiotou and Stufler's [13] results on the decorations.

1.3. **Outline of the paper.** The paper is organized as follows. In Section 2 we present the combinatorial setup and the main results. Section 3 is devoted to the study of the size of $D$-forests and the size of the $C$-tree $C_n$ in a random Pólya tree $T_n$. We also offer new proofs of some results from [13] using the analytic combinatorics framework. In Section 4 we prove Theorems 2 and 4. A detailed study of $D$-forest is the topic of Section 4. There we study the distribution of the size a randomly chosen $D$-forest within a Pólya tree as well as the distribution of the number of components. All the results can be generalized to further Pólya structures, albeit we do not obtain as explicit expressions as in the case of Pólya trees. This will be the topic of Section 5. We conclude in Section 6 with some final remarks.

## 2. Basic structures and main results

The generating function $C(z)$ defined in (1.5) counts several combinatorial objects. In this section we comment on the different interpretations and answer the question of what $C$-trees really are.

For a set $\mathcal{T}$ we define two functions: a size function $|\cdot| : \mathcal{T} \to \mathbb{N}$ (normally the number of its nodes) and a weight function $w : \mathcal{T} \to \mathbb{R}$. We call $\mathcal{T}$ a combinatorial class if the number of elements $T \in \mathcal{T}$ of any given size is finite. The (weighted) generating function $T(z)$ of $\mathcal{T}$ is given by

$$T(z) = \sum_{T\in\mathcal{T}} w(T) z^{|T|},$$

and the (weighted) exponential generating function $\hat{T}(z)$ of $\mathcal{T}$ is given by

$$\hat{T}(z) = \sum_{T\in\mathcal{T}} w(T)\frac{z^{|T|}}{|T|!} = \sum_{n\geq 0}\Big(\sum_{\substack{T\in\mathcal{T}\\ |T|=n}} w(T)\Big)\frac{z^n}{n!}.$$

First, let $\mathcal{C}_1$ be the combinatorial class of Cayley trees. These are labeled trees with the constant weight function $w(T) = 1$ for all $T \in \mathcal{C}_1$, see [14]. Then, $C(z)$ is the corresponding exponential generating function.

Second, let $\mathcal{C}_2$ be the combinatorial class of simply generated trees with $\Phi(z) = \exp(z)$. These are unlabeled plane trees with weight function

$$w(T) = \prod_{k\geq 0}\left(\frac{1}{k!}\right)^{n_k(T)} \tag{2.1}$$

where $n_k(T)$ is the number of nodes of $T$ with outdegree $k$, see [9, 10]. Then, $C(z)$ is the generating function of these trees.

Third, we can define the class $\mathcal{C}_3$ of $C$-trees after all.

**Definition 1.** A $C$-tree is a rooted non-plane tree $T$ with weight

$$w(T) = \mathrm{e}(T)\prod_{k\geq 0}\left(\frac{1}{k!}\right)^{n_k(T)} \tag{2.2}$$

where $\mathrm{e}(T)$ is the number of embeddings of $T$ into the plane.

*Remark* 1. The class $\mathcal{C}_3$ is the non-plane version of the class $\mathcal{C}_2$.

**Lemma 1.** *The generating function of $\mathcal{C}_3$ is $C(z)$.*

*Proof.* For a given $C$-tree $T$ of size $n$ and with root of outdegree $d$, let $T_1, T_2, \dots, T_k$ be the distinct subtrees of the root, appearing with multiplicities $m_1, \dots, m_k$, respectively. Then, $d = m_1 + m_2 + \cdots + m_k$ with $m_i \geq 1$. We define

$$\delta(T) := |\{\text{all permutations of } (\underbrace{T_1, \dots, T_1}_{m_1}, \underbrace{T_2, \dots, T_2}_{m_2}, \dots, \underbrace{T_k, \dots, T_k}_{m_k})\}| = \binom{d}{m_1, m_2, \dots, m_k}.$$

Hence, we get

$$w(T) = \frac{\delta(T)}{d!} \prod_{i=1}^{k} w(T_i)^{m_i} = \left(\frac{w(T_1)^{m_1}}{m_1!}\right)\left(\frac{w(T_2)^{m_2}}{m_2!}\right)\cdots\left(\frac{w(T_k)^{m_k}}{m_k!}\right).$$

This implies that the generating function of the class $\mathcal{C}_3$ satisfies $C(z) = z\exp(C(z))$. □

Now, we turn to the introduction of $D$-forests. We begin with some preliminary observations: In order to analyze the dominant singularity of $T(z)$, we follow [12, 14], see also [6, Chapter VII.5], and we rewrite (1.2) into

$$T(z) = ze^{T(z)}D(z), \qquad \text{where} \qquad D(z) = \sum_{n\geq 0} d_n z^n = \exp\left(\sum_{i=2}^{\infty} \frac{T(z^i)}{i}\right). \tag{2.3}$$

We observe that $D(z)$ is analytic for $|z| < \sqrt{\rho} < 1$ and that $\sqrt{\rho} > \rho$. From (2.3) it follows that $T(z)$ can be expressed in terms of the generating function of $C$-trees: Indeed, assume that $T(z)$ is a function $H(zD(z))$ depending on $zD(z)$. By (2.3) this is equivalent to $H(x) = x\exp(H(x))$. Yet, this is the functional equation for the generating function of $C$-trees. As this functional equation has a unique power series solution we have $H(x) = C(x)$, and we just proved

$$T(z) = C(zD(z)). \tag{2.4}$$

Note that $T(z) = C(zD(z))$ is a case of a super-critical composition schema which is characterized by the fact that the dominant singularity of $T(z)$ is strictly smaller than that of $D(z)$. In other words, the dominant singularity $\rho$ of $T(z)$ is determined by the outer function $C(z)$. Indeed, $\rho\, D(\rho) = e^{-1}$, because $e^{-1}$ is the unique dominant singularity of $C(z)$. Moreover, the composition schema corresponds to the substitution construction of combinatorial structures [6]. Indeed, (2.4) shows that a Pólya tree is a $C$-tree where to each vertex there has been attached a combinatorial object from a combinatorial class associated with the generating function $D(z)$. As Pólya trees are trees, the $D$-structures must be forests. Inspecting the generating function $D(z)$ more closely will show that there is a natural way of defining these $D$-forests.

**Definition 2.** Let $\mathrm{MSET}^{(\geq 2)}(\mathcal{T})$ denote the class of all multisets (or forests) of Pólya trees where each of its distinct components appears at least in duplicate (see Figure 2.1). For $F \in \mathrm{MSET}^{(\geq 2)}(\mathcal{T})$ let $\mathrm{Aut}(F)$ denote the automorphism group of $F$. Moreover, let $\sigma_i$ denote the number of cycles of length $i$ in the automorphism $\sigma \in \mathrm{Aut}(F)$. To each $F \in \mathrm{MSET}^{(\geq 2)}(\mathcal{T})$ we assign the weight

$$w(F) = \frac{|\{\sigma \in \mathrm{Aut}(F) \mid \sigma_1 = 0\}|}{|\mathrm{Aut}(F)|}.$$

Then the class $(\mathrm{MSET}^{(\geq 2)}(\mathcal{T}), w)$ is called the class of $D$-forests.

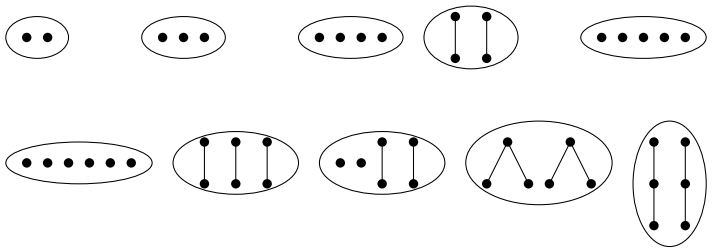

FIGURE 2.1. All $D$-forests of size $2, 3, 4, 5, 6$.

Our first main result is that this is indeed the combinatorially natural weighting for the $D$-forests satisfying (2.4) as well as relating the weights of $C$-trees in terms of automorphisms associated to a given Pólya tree. In particular, we will show that the cumulative weight $d_n$ (defined in (2.5)) of all such forests of size $n$ satisfies

$$d_n = \sum_{\substack{F \in \mathrm{MSET}^{(\geq 2)}(\mathcal{T}) \\ |F|=n}} \frac{|\{\sigma \in \mathrm{Aut}(F) \mid \sigma_1 = 0\}|}{|\mathrm{Aut}(F)|}$$

From (1.2) and (2.3) one gets the first values of this sequence:

$$D(z) = \sum_{n=0}^{\infty} d_n z^n = 1 + \frac{1}{2}z^2 + \frac{1}{3}z^3 + \frac{7}{8}z^4 + \frac{11}{30}z^5 + \frac{281}{144}z^6 + \frac{449}{840}z^7 + \cdots . \tag{2.5}$$

From (2.3) we can derive a recursion of $d_n$ as well. We get

$$d_n = \frac{1}{n}\sum_{i=2}^{n} d_{n-i} \sum_{\substack{m|i \\ m\neq i}} m t_m, \qquad \text{for } \ n \geq 2,$$

as well as $d_0 = 1$, and $d_1 = 0$.

Before we can formulate the first main result, we have to introduce further generating functions. Let $c_{n,k}$ denote the cumulative weight of all $C$-trees of size $k$ that are contained in Pólya trees of size $n$. By $t_{c,n}(u)$ and $T_c(z,u)$ we denote the corresponding generating function and the bivariate generating function of $(c_{n,k})_{n,k\geq 0}$, respectively, that is,

$$t_{c,n}(u) = \sum_{k=1}^{n} c_{n,k} u^k \quad \text{and} \quad T_c(z,u) = \sum_{n\geq 0} t_{c,n}(u) z^n.$$

Note that $c_{n,k}$ is in general not an integer. By marking the nodes of all $C$-trees in Pólya trees we find a functional equation for the bivariate generating function $T_c(z,u)$, which is

$$T_c(z,u) = zu \exp\left(T_c(z,u)\right) \exp\left(\sum_{i=2}^{\infty} \frac{T(z^i)}{i}\right) = zu \exp\left(T_c(z,u)\right) D(z). \tag{2.6}$$

Now we are ready to state the first main result:

**Theorem 2.** *Let $\mathcal{T}$ be the set of all Pólya trees, and $\mathrm{MSET}^{(\geq 2)}(\mathcal{T})$ be the multiset (or forest) of Pólya trees where each tree appears at least twice if it appears at all. Then the cumulative*

*weight $d_n$ (defined in (2.5)) of all such forests of size $n$ satisfies*

$$d_n = \sum_{\substack{F\in \mathrm{MSET}^{(\geq 2)}(\mathcal{T})\\ |F|=n}} \frac{|\{\sigma \in \mathrm{Aut}(F) \mid \sigma_1 = 0\}|}{|\mathrm{Aut}(F)|}$$

*where $\mathrm{Aut}(F)$ is the automorphism group of $F$. Furthermore, the polynomial associated to $C$-trees in Pólya trees of size $n$ is given by*

$$t_{c,n}(u) = \sum_{T\in\mathcal{T},\, |T|=n} t_T(u), \qquad \text{where } \; t_T(u) = \frac{1}{|\mathrm{Aut}(T)|} \sum_{\sigma\in\mathrm{Aut}(T)} u^{\sigma_1}.$$

*In particular, for all $T \in \mathcal{T}$, we have $t_T'(1) = |\mathcal{P}(T)|$ where $\mathcal{P}(T)$ is the set of all trees which are obtained by pointing (or coloring) one single node in $T$.*

Note that the decomposition of a Pólya tree into a $C$-tree and $D$-forests is in general not unique, because the $D$-forests consist of Pólya trees and their vertices are not distinguishable from those of the $C$-tree, see Figure 2.2. However, for a given Pólya tree $T$ the polynomial

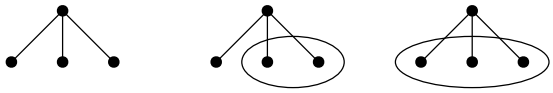

FIGURE 2.2. The decomposition of a Pólya tree with 4 nodes into a $C$-tree (non-circled nodes) and $D$-forests (circled nodes). For this Pólya tree there are 3 different decompositions.

$t_T(u)$ gives rise to a probabilistic interpretation of the composition scheme (2.4) (see also Example 2 in Section 4): Let $T_n$ denote a uniform random Pólya tree of size $n$. Then select one automorphism $\sigma$ of $T_n$ uniformly at random and let $C_n$ be the (random) subtree of $T_n$ consisting of all fixed points of $\sigma$. The tree $C_n$ is indeed a $C$-tree, *i.e.*, the remaining vertices in $T_n$ form a set of $D$-forests. So, fixing a Pólya tree together with one of its automorphisms uniquely defines a decomposition into a $C$-tree and a set of $D$-forests.

The coefficient of $u^k$ can then be interpreted as the probability that the underlying $C$-tree is of size $k$. In other words, $t_T(u)$ is the probability generating function of the random variable $C_T$ of the number of $C$-tree nodes in the tree $T$ defined by

$$\mathbb{P}(C_T = k) := [u^k]t_T(u). \tag{2.7}$$

The random variable $C_T$ is a refinement of $T_n$ in the following sense:

$$\mathbb{P}(C_T = k) = \mathbb{P}\left(|C_n| = k \mid T_n = T\right)$$

Now, let us turn to the second main result. Select a random Pólya tree $T_n$ and one of its automorhisms and consider the random $C$-tree $C_n$ given by this choice. For every vertex $v$ of $C_n$, we use $F_n(v)$ to denote the $D$-forest that is attached to the vertex $v$ in $T_n$, see Figure 2.3.

Let $L_n$ be the maximal size of a $D$-forest contained in $T_n$, that is, $|F_n(v)| \leq L_n$ holds for all $v$ of $C_n$ and the inequality is sharp.

**Theorem 3.** *For $0 < s < 1$,*

$$(1-(\log n)^{-s})\left(\frac{-2\log n}{\log \rho}\right) \leq L_n \leq (1+(\log n)^{-s})\left(\frac{-2\log n}{\log \rho}\right) \tag{2.8}$$

*holds with probability $1 - o(1)$.*

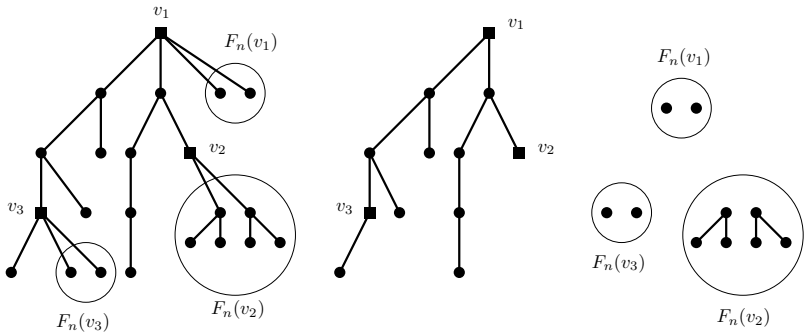


FIGURE 2.3. A random Pólya tree $T_n$ (left), a (possible) $C$-tree $C_n$ (middle) that is contained in $T_n$ where all $D$-forests $F_n(v)$, except $F_n(v_1), F_n(v_2), F_n(v_3)$ (right), are empty.

Our first main result is a proof of Theorem 3 by applying the unified framework of Gourdon [8]. A big-$O$ result for the upper bound was given by Panagiotou and Stufler [13, Eq. (5.5)].

Finally, we derive the limiting probability that for a random node $v$ the attached forest $F_n(v)$ is of a given size. This result is consistent with the Boltzmann sampler from [13]. The precise statement of our third main result is the following:

**Theorem 4.** *The generating function $T^{[m]}(z,u)$ of Pólya trees, where each vertex is marked by $z$, and each weighted $D$-forest of size $m$ is marked by $u$, is given by*

$$T^{[m]}(z,u) = C\left(uzd_m z^m + z\left(D(z) - d_m z^m\right)\right), \tag{2.9}$$

*where $d_m = [z^m]D(z)$. The probability that the $D$-forest $F_n(v)$ attached to a random $C$-tree node $v$ is of size $m$ is given by*

$$\mathbb{P}\left(|F_n(v)| = m\right) = \frac{d_m \rho^m}{D(\rho)}\left(1 + \mathcal{O}\left(n^{-1}\right)\right).$$

## 3. The maximal size of a $D$-forest

We will use the generating function approach from [8] to analyze the maximal size $L_n$ of $D$-forests in a random Pólya tree $T_n$, which provides a new proof of Theorem 3. Following the same approach, we can establish a central limit theorem for the random variable $|C_n|$, which has been done in [16] for the more general random $\mathcal{R}$-enriched trees.

*Proof of Theorem 3.* In (5.5) of [13], only an upper bound of $L_n$ is given. By directly applying Gourdon's results (Theorem 4 and Corollary 3 of [8]) for the super-critical composition schema, we find that for any positive $m$,

$$\mathbb{P}[L_n \le m] = \exp\left(-\frac{c_1 n}{m^{3/2}}\rho^{m/2}\right)\left(1 + \mathcal{O}(\exp(-m\varepsilon))\right),$$

where

$$c_1 \sim \frac{b}{2\sqrt{\pi}(1-\sqrt{\rho})(D(\rho) + \rho D'(\rho))},$$

as $n \to \infty$. Moreover, the maximal size $L_n$ satisfies asymptotically, as $n \to \infty$,

$$\mathbb{E}L_n = -\frac{2\log n}{\log \rho} - \frac{3}{2}\frac{2}{\log \rho}\log\log n + \mathcal{O}(1) \quad \text{and} \quad \mathbb{V}\mathrm{ar}\, L_n = \mathcal{O}(1).$$

By using Chebyshev's inequality, one can prove that $L_n$ is highly concentrated around the mean $\mathbb{E}L_n$. We set $\varepsilon_n = (\log n)^{-s}$ where $0 < s < 1$ and we get

$$\mathbb{P}(|L_n - \mathbb{E}L_n| \geq \varepsilon_n \cdot \mathbb{E}L_n) \leq \frac{\mathbb{V}\mathrm{ar}\, L_n}{\varepsilon_n^2 \cdot (\mathbb{E}L_n)^2} = o(1),$$

which means that (2.8) holds with probability $1 - o(1)$. □

It was shown in [16] that the size $|C_n|$ of the $C$-tree $C_n$ in $T_n$ satisfies a central limit theorem and $|C_n| = \Theta(n)$ holds with probability $1 - o(1)$. The precise statement is the following.

**Theorem 5** ([16, Eq. (3.9) and (3.10)], [13, Eq. (5.6)]). *The size of the $C$-tree $|C_n|$ in a random Pólya tree $T_n$ of size $n$ satisfies a central limit theorem where the expected value $\mathbb{E}|C_n|$ and the variance $\mathbb{V}ar|C_n|$ are asymptotically*

$$\mathbb{E}|C_n| = \frac{2n}{b^2\rho}(1 + \mathcal{O}(n^{-1})), \qquad \text{and} \qquad \mathbb{V}ar|C_n| = \frac{11n}{12b^2\rho}(1 + \mathcal{O}(n^{-1})).$$

*Furthermore, for any $s$ such that $0 < s < 1/2$, with probability $1 - o(1)$ we have*

$$(1 - n^{-s})\frac{2n}{b^2\rho} \leq |C_n| \leq (1 + n^{-s})\frac{2n}{b^2\rho}. \tag{3.1}$$

Random Pólya trees belong to the class of random $\mathcal{R}$-enriched trees and we refer the readers to [16] for the proof of Theorem 5 in the general setting. Here we provide a proof of Theorem 5 to show the connection between a bivariate generating function and the normal distribution and to emphasize the simplifications for the concrete values of the expected value and variance in this case.

*Proof of Theorem 5 (see also [16] for a probabilistic proof).* It follows from [4, Th. 2.23] that the random variable $|C_n|$ satisfies a central limit theorem. In the present case, we set $F(z, y, u) = zu\exp(y)D(z)$. It is easy to verify that $F(z, y, u)$ is an analytic function when $z$ and $y$ are near 0 and that $F(0, y, u) \equiv 0$, $F(x, 0, u) \not\equiv 0$ and all coefficients $[z^n y^m]F(z, y, 1)$ are real and nonnegative. From [4, Th. 2.23] we know that $T_c(z, u)$ is the unique solution of the functional identity $y = F(z, y, u)$. Since all coefficients of $F_y(z, y, 1)$ are nonnegative and the coefficients of $T(z)$ are positive as well as monotonically increasing, this implies that $(\rho, T(\rho), 1)$ is the unique solution of $F_y(z, y, 1) = 1$, which leads to the fact that $T(\rho) = 1$. Moreover, the expected value is

$$\begin{aligned}
\mathbb{E}|C_n| &= \frac{nF_u(z, y, u)}{\rho F_z(z, y, u)} \\
&= \frac{[z^n]\partial_u T_c(z, u)|_{u=1}}{[z^n]T(z)} \\
&= \left([z^n]\frac{T(z)}{1 - T(z)}\right)([z^n]T(z))^{-1} \\
&= \frac{2n}{b^2\rho}(1 + \mathcal{O}(n^{-1})).
\end{aligned}$$

The asymptotics are directly derived from (1.3). Likewise, we can compute the variance

$$\mathbb{V}\mathrm{ar}\,|C_n| = \frac{[z^n]T(z)(1 - T(z))^{-3}}{[z^n]T(z)} - (\mathbb{E}\,|C_n|)^2 = \frac{11n}{12b^2\rho}(1 + \mathcal{O}(n^{-1})).$$

Furthermore, $|C_n|$ is highly concentrated around $\mathbb{E}\,|C_n|$, which can be proved again by using Chebyshev's inequality. We set $\varepsilon_n = n^{-s}$ where $0 < s < 1/2$ and get

$$\mathbb{P}(\big||C_n| - \mathbb{E}|C_n|\big| \geq \varepsilon_n \cdot \mathbb{E}|C_n|) \leq \frac{\mathbb{V}\mathrm{ar}|C_n|}{\varepsilon_n^2 \cdot (\mathbb{E}|C_n|)^2} = \mathcal{O}(n^{2s-1}) = o(1),$$

which yields (3.1). □

As a simple corollary, we also get the total size of all weighted $D$-forests in $T_n$. Let $\mathcal{D}_n$ denote the union of all $D$-forests in a random Pólya tree $T_n$ of size $n$.

**Corollary 6.** *The size of weighted $D$-forests in a random Pólya tree of size $n$ satisfies a central limit theorem where the expected value $\mathbb{E}|\mathcal{D}_n|$ and the variance $\mathbb{V}ar|\mathcal{D}_n|$ are asymptotically*

$$\mathbb{E}|\mathcal{D}_n| = n\left(1 - \frac{2}{b^2\rho}\right)(1 + \mathcal{O}(n^{-1})), \quad \text{and} \quad \mathbb{V}ar|\mathcal{D}_n| = \frac{11n}{12b^2\rho}(1 + \mathcal{O}(n^{-1})).$$

Theorem 5 and Corollary 6 tell us that a random Pólya tree $T_n$ consists mostly of a $C$-tree (proportion $\frac{2}{b^2\rho}$ comprising $\approx 82.2\%$ of the nodes) and to a small part of $D$-forests (proportion $1 - \frac{2}{b^2\rho}$ comprising $\approx 17.8\%$ of the nodes). Furthermore, the average size of a $D$-forest $F_n(v)$ attached to a random $C$-tree vertex in $T_n$ is $\frac{b^2\rho}{2} - 1 \approx 0.216$, which indicates that on average the $D$-forest $F_n(v)$ is very small, although the maximal size of all $D$-forests in a random Pólya tree $T_n$ reaches $\Theta(\log n)$.

*Remark* 2. Let us describe the connection of (2.4) to the Boltzmann sampler from [13]. We know that $F(z, y, 1) = z\Phi(y)D(z)$ where $\Phi(x) = \exp(x)$ and $y = T(z)$. By dividing both sides of this equation by $y = T(z)$, one obtains from (2.3) that

$$1 = \frac{zD(z)}{T(z)}\exp(T(z)) = \exp(-T(z))\sum_{k\geq 0}\frac{T^k(z)}{k!},$$

which implies that in the Boltzmann sampler $\Gamma T(x)$, the number of offspring contained in the $C$-tree $C_n$ is Poisson distributed with parameter $T(x)$. As an immediate result, the random $C$-tree contained in the Pólya tree generated by the Boltzmann sampler $\Gamma T(\rho)$ corresponds to a critical Galton-Watson tree since the expected number of offspring is $F_y(z, y, 1) = 1$ for $(z, y) = (\rho, 1)$.

## 4. $D$-forests and $C$-trees

In order to get a better understanding of $D$-forests and $C$-trees, we need to return to the original proof of Pólya on the number of Pólya trees [14]. The important step is the treatment of tree automorphisms by the cycle index.

As before, we denote by $\sigma_i$ the number of cycles of length $i$ of a permutation $\sigma$. Let $S_k$ be the symmetric group of order $k$. The *type* of a permutation $\sigma \in S_k$ is the $k$-tuple $(\sigma_1, \sigma_2, \ldots, \sigma_k)$. Note that $k = \sum_{i=1}^{k} i\sigma_i$.

**Definition 3** (Cycle index)**.** Let $G$ be a subgroup of the symmetric group $S_k$. Then the *cycle index* is

$$Z(G; s_1, s_2, \ldots, s_k) = \frac{1}{|G|}\sum_{\sigma\in G} s_1^{\sigma_1}s_2^{\sigma_2}\cdots s_k^{\sigma_k}.$$

Now we are ready to prove Theorem 2.

4.1. **Proof of Theorem 2.** By Pólya's enumeration theory [14], the generating function $T(z)$ satisfies the functional equation

$$\begin{aligned} T(z) &= z\sum_{k\geq 0} Z(S_k; T(z), T(z^2), \dots, T(z^k)) \\ &= z\sum_{k\geq 0}\frac{1}{k!}\sum_{\sigma\in S_k}(T(z))^{\sigma_1}(T(z^2))^{\sigma_2}\cdots(T(z^k))^{\sigma_k}, \end{aligned}$$

which can be simplified to (1.1), the starting point of our research, by a simple calculation. However, this shows that the generating function of $D$-forests from (2.3) is given by

$$\begin{aligned} D(z) &= \exp\left(\sum_{i=2}^{\infty}\frac{T(z^i)}{i}\right) \\ &= \sum_{k\geq 0} Z(S_k; 0, T(z^2), \dots, T(z^k)) = \sum_{k\geq 0}\frac{1}{k!}\sum_{\sigma\in S_k \text{ such that } \sigma_1=0}(T(z^2))^{\sigma_2}\cdots(T(z^k))^{\sigma_k}. \end{aligned} \tag{4.1}$$

The weight of a $D$-forest of size $n$ comprising $k$ trees is given by the ratio of fixed point free automorphisms over the total number of automorphisms. This quotient equals the number of fixed point free permutations $\sigma\in S_k$ of the trees which the forest consists of divided by the total number of orderings $k!$, since the automorphisms of the subtrees of the root contribute to both the number of all and the number of fixed point free automorphisms of the forest. But this last quotient is precisely the coefficient of $z^n$ in the $k$th summand of (4.1). Thus

$$d_n = [z^n]D(z) = \sum_{\substack{F\in \mathrm{MSET}^{(\geq 2)}(\mathcal{T})\\ |F|=n}} \frac{|\{\sigma\in \mathrm{Aut}(F)\mid \sigma_1=0\}|}{|\mathrm{Aut}(F)|}.$$

This proves the first assertion of Theorem 2.

*Example* 1. The smallest $D$-forest is of size 2, and it consists of a pair of single nodes, see Figure 2.1. This forest has only one fixed point free automorphism, thus $d_2 = 1/2$. For $n=3$ the forest consists of 3 single nodes. The fixed point free permutations are the 3-cycles, thus $d_3 = 2/6 = 1/3$. The case $n=4$ is more interesting. A forest consists either of 4 single nodes, or of 2 identical trees, each consisting of 2 nodes and one edge. In the first case we have 6 4-cycles and 3 pairs of transpositions. In the second case we have 1 transposition swapping the two trees. Thus, $d_4 = \frac{6+3}{24}+\frac{1}{2} = \frac{7}{8}$. ♢

These results also yield a natural interpretation of $C$-trees. We recall that by definition

$$T_c(z,u) = \sum_{n\geq 0} t_{c,n}(u) z^n,$$

where $t_{c,n}(u) = \sum_k c_{n,k}u^k$ is the polynomial marking the $C$-trees in Pólya trees of size $n$. From the decompositions (2.4) and (2.6) we get the first few terms:

$$t_{c,1}(u) = u, \qquad t_{c,2}(u) = u^2, \qquad t_{c,3}(u) = \frac{3}{2}u^3+\frac{1}{2}u, \qquad t_{c,4}(u) = \frac{8}{3}u^4+u^2+\frac{1}{3}u.$$

Evaluating these polynomials at $u=1$ obviously returns $t_{c,n}(1) = t_n$, which is the number of Pólya trees of size $n$. Their coefficients, however, are weighted sums depending on the number of $C$-tree nodes. For a given Pólya tree there are in general several ways to decide

what is a $C$-tree node and what is a $D$-forest node. The possible choices are encoded in the automorphisms of the tree, and these are responsible for the above weights as well.

Let $T$ be a Pólya tree and $\mathrm{Aut}(T)$ its automorphism group. For an automorphism $\sigma \in \mathrm{Aut}(T)$ the nodes which are fixed points of $\sigma$ are $C$-tree nodes. All other nodes are part of $D$-forests. Summing over all automorphisms and normalizing by the total number gives the $C$-tree generating polynomial for $T$:

$$t_T(u) = Z(\mathrm{Aut}(T); u, 1, \dots, 1) = \frac{1}{|\mathrm{Aut}(T)|} \sum_{\sigma \in \mathrm{Aut}(T)} u^{\sigma_1}. \tag{4.2}$$

The polynomial of $C$-trees in Pólya trees of size $n$ is then given by

$$t_{c,n}(u) = \sum_{T \in \mathcal{T},\, |T|=n} t_T(u),$$

which completes the proof of the second assertion of Theorem 2.

*Example* 2. For $n = 3$ we have 2 Pólya trees, namely the chain $T_1$ and the cherry $T_2$. Thus, $\mathrm{Aut}(T_1) = \{\mathrm{id}\}$, and $\mathrm{Aut}(T_2) = \{\mathrm{id}, \sigma\}$, where $\sigma$ swaps the two leaves but the root is unchanged. Thus,

$$t_{T_1}(u) = u^3, \qquad t_{T_2}(u) = \frac{1}{2}(u^3 + u).$$

For $n = 4$ we have 4 Pólya trees shown in Figure 4.1. Their automorphism groups are given by $\mathrm{Aut}(T_1) = \mathrm{Aut}(T_2) = \{\mathrm{id}\}$, $\mathrm{Aut}(T_3) = \{\mathrm{id}, (v_3\, v_4)\} \cong S_2$, and

$$\mathrm{Aut}(T_4) = \{\mathrm{id}, (v_2\, v_3), (v_3\, v_4), (v_2\, v_4), (v_2\, v_3\, v_4), (v_2\, v_4\, v_3)\} \cong S_3.$$

This gives

$$t_{T_1}(u) = t_{T_2}(u) = u^4, \qquad t_{T_3}(u) = \frac{1}{2}(u^4 + u^2), \qquad t_{T_4}(u) = \frac{1}{6}(u^4 + 3u^2 + 2u).$$

This enables us to give a probabilistic interpretation of the composition scheme (2.4). For a given tree the weight of $u^k$ is the probability that the underlying $C$-tree is of size $k$. In particular, $T_1$ and $T_2$ do not have $D$-forests. The tree $T_3$ consists of a $C$-tree with 4 or with 2 nodes, each case with probability 1/2. In the second case, as there is only one possibility for the $D$-forest, it consists of the pair of single nodes which are the leaves. Finally, the tree $T_4$ has either 4 $C$-tree nodes with probability 1/6, 2 with probability 1/2, or only one with probability 1/3. These decompositions are shown in Figure 2.2. ♢

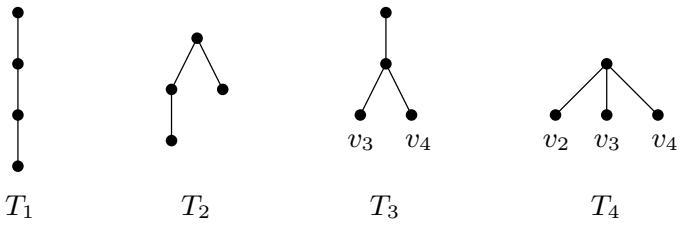


FIGURE 4.1. All Pólya trees of size 4.

In the same way as we got the composition scheme in (2.4), we can rewrite $T_c(z,u)$ from (2.6) into $T_c(z,u) = C(uzD(z))$. The expected total weight of all $C$-trees contained in all Pólya trees of size $n$ is the $n$-th coefficient of $T_c(z)$, which is

$$T_c(z) := \frac{\partial}{\partial u} T_c(z,u)\Big|_{u=1} = \frac{T(z)}{1-T(z)} = z + 2z^2 + 5z^3 + 13z^4 + 35z^5 + 95z^6 + \cdots . \tag{4.3}$$

Let us explain why these numbers are integers, although the coefficients of $t_{c,n}(u)$ are in general not. We will show an even stronger result.

**Lemma 7.** *Let $T$ be a tree and $\mathcal{P}(T)$ be the set of all trees with one single pointed (or colored) node which can be generated from $T$. Then for all $T \in \mathcal{T}$ we have $t_T'(1) = |\mathcal{P}(T)|$.*

*Proof.* From (4.2) we get $t_T'(1) = \sum_{\sigma\in \mathrm{Aut}(T)} \frac{\sigma_1}{|\mathrm{Aut}(T)|}$ is the expected number of fixed points in a uniformly at random chosen automorphism of $T$. The associated random variable $C_T$ is defined in (2.7). We will prove $\mathbb{E}(C_T) = |\mathcal{P}(T)|$ by induction on the size of $T$.

The most important observation is that only if the root of a subtree is a fixed point, its children can also be fixed points. Obviously, the root of the tree is always a fixed point.

For $|T| = 1$, the claim holds as $\mathbb{E}(C_T) = 1$ and there is just one tree with a single node and a marker on it. For larger $T$ consider the construction of Pólya trees. A Pólya tree consists of a root $T_0$ and its children, which are a multiset of smaller trees. Thus, the set of children is of the form

$$\{T_{1,1}, \dots, T_{1,k_1}, T_{2,1}, \dots, T_{2,k_2}, \dots, T_{r,1}, \dots, T_{r,k_r}\}, \quad \text{with} \quad T_{i,j} \in \mathcal{T},$$

and where trees with the same first index are isomorphic. On the level of children, the possible behaviors of automorphisms are permutations within the same class of trees. In other words, an automorphism may interchange the trees $T_{1,1}, \dots, T_{1,k_1}$ in $k_1!$ many ways, etc. Here the main observation comes into play: only subtrees of which the root is a fixed point might also have other fixed points. Thus, the expected number of fixed points is given by the expected number of fixed points in a random permutation of $S_{k_i}$ times the expected number of fixed points in $T_{k_i}$. By linearity of expectation we get

$$\mathbb{E}(C_T) = \mathbb{E}(C_{T_0}) + \sum_{i=0}^{r} \underbrace{\mathbb{E}(\text{number of fixed points in } S_{k_i})}_{=1} \mathbb{E}(C_{T_i}),$$

where $\mathbb{E}(C_{T_i}) = \mathbb{E}(C_{T_{i,j}})$ for all $1 \le j \le k_i$ and $\mathbb{E}(C_{T_0}) = 1$ because the root is a fixed point of any automorphism. Since the expected number of fixed points for each permutation is 1, we get on average 1 representative for each class of trees. This is exactly the operation of labeling one tree among each equivalence class. Finally, by induction the claim holds. □

This completes the proof of Theorem 2. □

As an immediate consequence of Lemma 7, $t_{c,n}'(1)$ counts the number of Pólya trees with $n$ nodes and a single labeled node (see OEIS A000107, [15]). This also explains the construction of non-empty sequences of trees in (4.3): Following the connection of [2, pp. 61–62] one can draw a path from the root to each labeled node. The nodes on that path are the roots of a sequence of Pólya trees.

*Remark* 3. Note that Lemma 7 also implies that the total number of fixed points in all automorphisms of a tree is a multiple of the number of automorphisms.

*Remark* 4. Lemma 7 can also be proved by considering cycle-pointed Pólya trees; see [3, Section 3.2] for a full description. Let $(T, c)$ be a cycle-pointed structure considered up to symmetry where $T$ is a Pólya tree and $c$ is a cycle of an automorphism $\sigma \in \mathrm{Aut}(T)$. Then, the number of such cycle-pointed structures $(T, c)$ where $c$ has length 1 is exactly the number $t'_T(1)$.

Let us analyze the $D$-forests in $T_n$ more carefully. We want to count the number of $D$-forests that have size $m$ in a random Pólya tree $T_n$. Therefore, we label such $D$-forests with an additional parameter $u$ in (2.4). From the bivariate generating function (2.9) we can recover the probability $\mathbb{P}[|F_n(v)| = m]$ to generate a $D$-forest of size $m$ in the Boltzmann sampler from [13].

4.2. **Proof of Theorem 4.** The first result is a direct consequence of (2.4), where only vertices with weighted $D$-forests of size $m$ are marked. For the second result we differentiate both sides of (2.9) and get

$$T_u^{[m]}(z,1) = \frac{T(z)}{1-T(z)}\frac{d_m z^m}{D(z)} = T_c(z)\frac{d_m z^m}{D(z)}.$$

Then, the sought probability is given by

$$\mathbb{P}\left[|F_n(v)| = m\right] = \frac{[z^n]T_u^{[m]}(z,1)}{[z^n]T_c(z)} = \frac{d_m \rho^m}{D(\rho)}\left(1 + \mathcal{O}\left(n^{-1}\right)\right).$$

For the last equality we used the fact that $D(z)$ is analytic in a neighborhood of $z = \rho$.

Let $P_n(u)$ be the probability generating function for the size of a weighted $D$-forest $F_n(v)$ attached to a vertex $v$ of $C_n$ in a random Pólya tree $T_n$. From the previous theorem it follows that

$$P_n(u) = \sum_{m\geq 0}\frac{[z^n]T_u^{[m]}(z,1)}{[z^n]T_c(z)}u^m = \frac{[z^n]T_c(z)\frac{D(zu)}{D(z)}}{[z^n]T_c(z)} = \frac{D(\rho u)}{D(\rho)}\left(1 + \mathcal{O}\left(n^{-1}\right)\right).$$

This is exactly [13, Eq. (5.2)]. □

Summarizing, we state in Table 1 the asymptotic probabilities that a weighted $D$-forest $F_n(v)$ in $T_n$ has size equal to or greater than $m$.

| $m$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|---|
| $\mathbb{P}[\lvert F_n(v)\rvert = m] \approx$ | 0.9197 | 0.0000 | 0.0526 | 0.0119 | 0.0105 | 0.0015 | 0.0027 | 0.0003 |
| $\mathbb{P}[\lvert F_n(v)\rvert \geq m] \approx$ | 1.0000 | 0.0803 | 0.0803 | 0.0277 | 0.0161 | 0.0060 | 0.0041 | 0.0014 |

TABLE 1. The probability that a weighted $D$-forest $F_n(v)$ has size equal to or greater than $m$ when $0 \leq m \leq 7$.

As most of the vertices in $C_n$ have an empty $D$-forest, it is also interesting to condition on the non-empty ones only. Its generating function is given by

$$Q_n(u) = \sum_{n\geq 2}\mathbb{P}[|F_n(v)| = m \;\big|\; |F_n(v)| > 0]u^m = \frac{D(\rho u) - 1}{D(\rho) - 1}\left(1 + \mathcal{O}\left(n^{-1}\right)\right).$$

Its first values are listed in Table 2. It is interesting to see in these tables that the sequence of probabilities is not decreasing in $m$. Additionally, we deduce that more than 80% of the used $D$-forests are $D$-forests of size 2 and 3. These are two or three copies of a single node.

| $m$ | | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|
| $\mathbb{P}[\lvert F_n(v)\rvert = m$ | $\lvert F_n(v)\rvert > 0] \approx$ | 0.656 | 0.148 | 0.131 | 0.019 | 0.034 | 0.003 | 0.007 | 0.001 |

TABLE 2. The probability that a weighted $D$-forest $F_n(v)$ has size equal to or greater than $m$ when $0 \le m \le 7$.

4.3. **Properties of $D$-forests.** Let us start with a short analysis of $D(z)$.

**Lemma 8.** *The generating function $D(z)$ of $D$-forests has radius of convergence $\sqrt{\rho}$. It has two dominant singularities at $z = \pm\sqrt{\rho}$. Let $\xi(z) = e^{\frac{T(z^3)}{3}+\frac{T(z^4)}{4}+\cdots}$, which is analytic for $|z| < \rho^{1/3}$. Then,*

$$d_n = \left(\xi(\sqrt{\rho}) + (-1)^n \xi(-\sqrt{\rho})\right) b\sqrt{\frac{\rho e}{8\pi}} \frac{\rho^{-n/2}}{\sqrt{n^3}} \left(1 + \mathcal{O}\left(\frac{1}{n}\right)\right). \tag{4.4}$$

*Furthermore, we have $D(\rho) = \frac{1}{e\rho}$ and $D'(\rho) = \frac{1}{e\rho^2}\left(\frac{b^2\rho}{2} - 1\right)$.*

*Proof.* The key essence to this result is the elementarily checked fact that if $T(z)$ has radius of convergence $\rho$, then $T(z^2)$ will have radius of convergence $\sqrt{\rho}$. Therefore, $\pm\sqrt{\rho}$ are the dominant singularities, as $T(z)$ has a unique singularity at $z = \rho$.

The asymptotic expansions are then derived from (1.3) as

$$T(z^2) = 1 - b\sqrt{2\rho}\left(1 \mp \frac{z}{\sqrt{\rho}}\right)^{1/2} + \mathcal{O}\left(1 \mp \frac{z}{\sqrt{\rho}}\right), \qquad \text{for } z \to \pm\sqrt{\rho}.$$

Next, note that $\xi(z)$ is analytic for $|z| < \rho^{1/3}$ due to the same reasoning as above. Thus, the asymptotic expansion of $D(z) = e^{\frac{T(z^2)}{2}}\xi(z)$ is derived by combining the contributions on the two dominant singularities.

Finally, the values for $D(\rho)$ and $D'(\rho)$ are derived from (2.4). □

We want to determine the number of trees in a given $D$-forest. Let $d_{n,k}$ be the weight of $D$-forests of size $n$ consisting of $k$ trees. Then, by (2.3) the bivariate generating function satisfies

$$D(z, v) = \sum_{n,k\ge 0} d_{n,k} z^n v^k = \exp\left(\sum_{i=2}^{\infty} v^i \frac{T(z^i)}{i}\right).$$

**Theorem 9.** *Let $X_n$ be the random variable for the number of trees in a $D$-forest of size $n$,* i.e., *$\mathbb{P}[X_n = k] := \frac{d_{n,k}}{d_n}$. Then we have*

$$\mathbb{E}X_n = \begin{cases} 3 + \mu_0 + \mathcal{O}\left(n^{-1}\right) \approx 3.2715 + \mathcal{O}\left(n^{-1}\right), & \text{for } n \text{ even}, \\ 3 + \mu_1 + \mathcal{O}\left(n^{-1}\right) \approx 6.7852 + \mathcal{O}\left(n^{-1}\right), & \text{for } n \text{ odd}, \end{cases}$$

*with*

$$\mu_0 := \frac{\xi(\sqrt{\rho})\gamma(\sqrt{\rho}) + \xi(-\sqrt{\rho})\gamma(-\sqrt{\rho})}{\xi(\sqrt{\rho}) + \xi(-\sqrt{\rho})}, \qquad \mu_1 := \frac{\xi(\sqrt{\rho})\gamma(\sqrt{\rho}) - \xi(-\sqrt{\rho})\gamma(-\sqrt{\rho})}{\xi(\sqrt{\rho}) - \xi(-\sqrt{\rho})},$$

$$\xi(z) = \exp\Big(\sum_{i\ge 3} \frac{T(z^i)}{i}\Big), \qquad \gamma(z) = \sum_{i\ge 2} T(z^i).$$

*Proof.* The correspondence with generating functions gives

$$\mathbb{E}X_n = \frac{[z^n]D_v(z,1)}{[z^n]D(z)} = \frac{[z^n]D(z)\sum_{i\geq 2}T(z^i)}{[z^n]D(z)}.$$

As $T(z)$ has a unique singularity at $\rho$, $T(z^k)$ is singular at $\omega^k\rho^{1/k}$ where $\omega = \exp(2\pi i/k)$ is a $k$-th root of unity. By linearity of the coefficient extraction operator all that remains is to consider $D(z)T(z^2)$. By Lemma 4.4 we get

$$D(z)T(z^2) = -3b\sqrt{\frac{e\rho}{2}}\left(1 \mp \frac{z}{\sqrt{\rho}}\right)^{1/2} + \mathcal{O}\left(1 \mp \frac{z}{\sqrt{\rho}}\right), \qquad \text{for } z \to \pm\sqrt{\rho}.$$

Therefore, by standard techniques of singularity analysis [6] we get

$$\frac{[z^n]D(z)T(z^2)}{[z^n]D(z)} = 3 + \mathcal{O}\left(n^{-1}\right).$$

The fluctuating constant $\mu_n$ arises from the second part $\frac{[z^n]D(z)\gamma(z)}{[z^n]D(z)}$. Due to the reasoning above $\gamma(z)$ is analytic for $|z| < \rho^{1/3}$. Thus, we can again use Lemma 4.4 to combine the singular expansions of $D(z)$ at $\pm\sqrt{\rho}$ with the analytic expansion of $\gamma(z)$.

For the computations of the approximate values we used Maple. □

As a next step we investigate the number of trees of a random $D$-forest in a random Pólya tree of size $n$. Let $t_{n,k}$ be the weight of Pólya trees of size $n$ with having $k$ trees in their $D$-forests. Then, by (2.4) the bivariate generating function satisfies

$$T(z,v) = \sum_{n,k\geq 0} t_{n,k} z^n v^k = C(zD(z,v)) = zD(z,v)e^{T(z,v)}.$$

**Theorem 10.** *The total number $Y_n$ of trees of all $D$-forests in a random Pólya tree $T_n$ of size $n$ satisfies a central limit theorem where the expected value $\mathbb{E}Y_n$ and the variance $\mathbb{V}arY_n$ are asymptotically*

$$\mathbb{E}Y_n = \frac{2\gamma(\rho)}{b^2\rho}n(1+\mathcal{O}(n^{-1})), \qquad \text{and} \qquad \mathbb{V}arY_n = \sigma^2 n(1+\mathcal{O}(n^{-1})),$$

*with* $\sigma^2 = \frac{2}{b^2\rho}\left(\frac{(2b^3\rho+72d\rho+18b)\gamma(\rho)^2}{9b^3\rho} - \frac{4\gamma'(\rho)\gamma(\rho)}{b^2} + \gamma_2(\rho)\right) \approx 0.26718$, *and* $\gamma_2(z) = \sum_{i\geq 2} iT(z^i)$. *Furthermore, for any $s$ such that $0 < s < 1/2$, with probability $1 - o(1)$ we have*

$$(1-n^{-s})\frac{2\gamma(\rho)}{b^2\rho}n \leq Y_n \leq (1+n^{-s})\frac{2\gamma(\rho)}{b^2\rho}n. \tag{4.5}$$

*Proof.* The proof uses the same techniques as the one of Theorem 5. In particular, it follows again from [4, Th. 2.23] that $Y_n$ satisfies a central limit theorem. Here, we set $F(z,y,v) = ze^yD(z,v)$. The technical conditions are easy to check, and we know that $T(z,v)$ is the uniuqe solution of $y = F(z,y,v)$. □

Theorems 5 and 10 tell us that in a Pólya trees of size $n$ there are on average $\frac{2\gamma(\rho)}{b^2\rho}n \approx 0.15776n$ trees in $D$-forests, and $\gamma(\rho) \approx 0.191837$ trees in the $D$-forest of a $C$-tree node. Comparing this number with the average size of the $D$-forest $\frac{b^2\rho}{2} - 1 \approx 0.216$ of a $C$-tree node, we conclude that most trees consist of only one node. In particular, as every tree consists of at least a root node, a component of a $D$-forest has on average approximately

0.024167 non-root nodes. This implies, that on average only every 42nd $D$-tree has more than one node.

## 5. Other Pólya structures

5.1. **Pólya trees with outdegree restriction.** Our work can be extended to $\Omega$-Pólya trees in the same way, so we omit the proof. For any $\Omega \subseteq \mathbb{N}_0 = \{0, 1, \ldots\}$ such that $0 \in \Omega$ and $\{0, 1\} \neq \Omega$, *an $\Omega$-Pólya tree* is a rooted unlabeled tree considered up to symmetry and with outdegree set $\Omega$. When $\Omega = \mathbb{N}_0$, a $\mathbb{N}_0$-Pólya tree is a Pólya tree.

Let $a_n$ be the number of Pólya trees of size $n$ with outdegree set $\Omega$, and $\mathcal{A}(z)$ the associated generating function. That is, $a_n = [z^n]\mathcal{A}(z)$. From Pólya enumeration theory [14] and Burnside's Lemma, the generating function $\mathcal{A}(z)$ satisfies the functional equation

$$\begin{aligned}\mathcal{A}(z) &= z \cdot \sum_{k\in\Omega} Z(S_k; \mathcal{A}(z), \mathcal{A}(z^2), \ldots, \mathcal{A}(z^k)) \\ &= z \sum_{k\in\Omega} \frac{1}{k!} \sum_{\sigma\in S_k} (\mathcal{A}(z))^{\sigma_1} (\mathcal{A}(z^2))^{\sigma_2} \cdots (\mathcal{A}(z^k))^{\sigma_k}. \qquad (5.1)\end{aligned}$$

Proposition 11 has been also used in [13]. It was actually implicitly stated in [1] and fits into the general theorem on implicit functions in [4, 6].

**Proposition 11.** *Let $\tau$ be the unique dominant singularity of $\mathcal{A}(z)$. Then $0 < \tau < 1$ and $0 < \mathcal{A}(\tau) < \infty$. Furthermore, $\tau$ is the unique real solution of*

$$\sum_{k\in\Omega} \frac{\partial}{\partial x} Z(S_k; x, \mathcal{A}(\tau^2), \ldots, \mathcal{A}(\tau^k))\big|_{x=\mathcal{A}(\tau)} = \frac{1}{\tau}. \qquad (5.2)$$

*and $\mathcal{A}(z)$ has a local expansion of the form*

$$\mathcal{A}(z) = \mathcal{A}(\tau) - b_1 (\tau - z)^{1/2} + c_1(\tau - z) + \mathcal{O}\left((\tau - z)^{3/2}\right) \qquad (5.3)$$

*where $b_1 > 0$ is a constant and $a_n = [z^n]\mathcal{A}(z)$ is asymptotically*

$$a_n = \frac{b_1\sqrt{\tau}}{2\sqrt{\pi}} n^{-3/2} \tau^{-n} (1 + \mathcal{O}(n^{-1})). \qquad (5.4)$$

Similar to $T_n$, we consider a random Pólya tree of size $n$ with outdegree set $\Omega$, denoted by $T_n^{(d)}$, which is a tree that is selected uniformly at random from all Pólya trees of $n$ vertices and with outdegree set $\Omega$. Similar to $L_n$, let $L_n^{(d)}$ be the maximal size of a $D$-forest contained in $T_n^{(d)}$. In the same way as Theorem 3, we have

**Theorem 12.** *For $0 < s < 1$,*

$$(1 - (\log n)^{-s}) \left(\frac{-2\log n}{\log \tau}\right) \leq L_n^{(d)} \leq (1 + (\log n)^{-s}) \left(\frac{-2\log n}{\log \tau}\right)$$

*holds with probability $1 - o(1)$.*

Note that $\tau$ is determined by (5.2). From (5.1) we see that every Pólya tree with outdegree restriction $\Omega$ is a multiset of small Pólya trees with outdegree restriction $\Omega$. We first consider the bivariate generating function

$$\mathcal{A}(z, u) = uz \cdot \sum_{k\in\Omega} Z(S_k; \mathcal{A}(z, u), \mathcal{A}_\Omega(z^2), \ldots, \mathcal{A}_\Omega(z^k)) \qquad (5.5)$$

For a random Pólya tree $T_n^{(d)}$ with outdegree restriction $\Omega$, we then select one automorphism of $T_n^{(d)}$ uniformly at random and all fixed points of such an automorphism form a random $C$-tree, denoted by $C_n^{(d)}$. It was also shown in [13] that the size $|C_n^{(d)}|$ in $T_n^{(d)}$ satisfies a central limit theorem and $|C_n^{(d)}| = \Theta(n)$ holds with probability $1 - o(1)$.

**Theorem 13** ( [16, Eq. (3.9) and (3.10)], [13, Eq. (5.6)]). *The size of the $C$-tree $|C_n^{(d)}|$ in a random Pólya tree $T_n^{(d)}$ of size $n$ satisfies a central limit theorem where the expected value $\mathbb{E}|C_n^{(d)}|$ and the variance $\mathbb{V}ar|C_n^{(d)}|$ are asymptotically*

$$\mathbb{E}\,|C_n^{(d)}| = \frac{n}{1+\mu}(1+\mathcal{O}(n^{-1})) \ \text{ where } \mu = \frac{\tau^2}{\mathcal{A}(\tau)}\sum_{k\in\Omega}\frac{\partial}{\partial x}Z(S_k;\mathcal{A}(\tau),\mathcal{A}(x^2),\dots,\mathcal{A}(x^k))\big|_{x=\tau}$$

*and the variance is $\mathbb{V}ar|C_n^{(d)}| = \sigma^2 n$ where $\sigma > 0$. Furthermore, for any $s$ such that $0 < s < 1/2$, with high probability we have*

$$(1-n^{-s})\frac{n}{1+\mu} \le |C_n^{(d)}| \le (1+n^{-s})\frac{n}{1+\mu}.$$

*Example* 3. For $\Omega = \mathbb{N}_0 - \{1\}$, any $\Omega$-Pólya tree is a Pólya tree without nodes of degree 1, which is also called *a hierarchy*. Let $T^*(z)$ be the ordinary generating function of hierarchies. Then if we remove the root of a hierarchy, we are left with a multiset of smaller hierarchies and the number of such subtrees is at least 2. That is,

$$\begin{aligned} T^*(z) &= z\exp\left(\sum_{i=1}^{\infty}\frac{T^*(z^i)}{i}\right) - zT^*(z) \\ &= \frac{z}{1+z}\exp\left(\sum_{i=1}^{\infty}\frac{T^*(z^i)}{i}\right) = z + z^3 + z^4 + 2z^5 + 3z^6 + \cdots. \end{aligned} \tag{5.6}$$

The generating function of hierarchies was also derived in [7] where the size of a hierarchy is defined as the number of leaves, instead of the number of nodes. From (5.2) we find that $\tau$ is the unique solution of

$$\begin{aligned} &\sum_{k\in\mathbb{N}_0}\frac{\partial}{\partial x}Z(S_k;x,T^*(\tau^2),\dots,T^*(\tau^k))|_{x=T^*(\tau)} - \frac{\partial}{\partial x}Z(S_1;x)|_{x=T^*(\tau)} \\ &= \exp\left(\sum_{i=1}^{\infty}\frac{T^*(z^i)}{i}\right) - T^*(\tau) = \tau^{-1}T^*(\tau) = \tau^{-1}. \end{aligned}$$

This yields $T^*(\tau) = 1$. If we again differentiate both sides of (5.6) and take the $n$-th coefficient of $z$, we get a recursion of hierarchies, namely, let $t^*(n) = [z^n]T^*(z)$. Then we have $t^*(1) = 1$, $t^*(2) = 0$ and for $n \ge 3$,

$$t^*(n) = \frac{1}{n-1}\sum_{i=1}^{n-2}(t^*(n-i) + t^*(n-i-1))\sum_{m|i} m t^*(m) + \frac{1}{n-1}\sum_{\substack{m|(n-1)\\ m\neq(n-1)}} m t^*(m).$$

With the help of this recursion, we can use Maple to generate the numbers of hierarchies, by which we can find the approximate solution $\tau \approx 0.4580838$ of $T^*(\tau) = 1$. Furthermore, for

the case of hierarchies, we compute $\mu$ in Theorem 13, which is

$$\mu = \frac{\tau^2}{T^*(\tau)} \exp(T^*(\tau)) \left( \exp(\sum_{i=2}^{\infty} \frac{T^*(z^i)}{i}) \right)'_{z=\tau} \approx 0.6701252,$$

where we used $T^*(\tau) = 1$. ♢

*Example* 4. For $\Omega = \{0, 2\}$, any $\Omega$-Pólya tree is a binary Pólya tree. Let $T_2(z)$ be the ordinary generating function of binary Pólya trees. Then we have

$$T_2(z) = z + \frac{1}{2} z (T_2(z))^2 + \frac{1}{2} z T_2(z^2). \tag{5.7}$$

From (5.2) we find that $\tau$ is the unique solution of

$$\frac{\partial}{\partial x}(Z(S_0) + Z(S_2; x, T_2(\tau^2)))|_{x=T_2(\tau)} = T_2(\tau) = \tau^{-1}, \tag{5.8}$$

and as before we can derive a recursion from (5.7), namely, let $t_2(n) = [z^n]T_2(z)$. Note that every binary Pólya trees has an even number of nodes, that is, for even $n$, $t_2(n) = 0$. For odd $n$, $n \geq 3$, we have

$$t_2(n) = \frac{1}{2} \sum_{i=1}^{n-2} t_2(i) t_2(n-1-i) + \frac{1}{2} t_{\lfloor \frac{n-1}{2} \rfloor},$$

and $t_2(1) = 1$. With the help of this recursion, we can use Maple to generate the numbers of binary Pólya trees, by which we can find the approximate solution $\tau \approx 0.6348553$ of $T_2(\tau) = \tau^{-1}$. Furthermore, for the case of binary Pólya trees, we compute $\mu$ in Theorem 13, which is

$$\mu = \frac{\tau^2}{T_2(\tau)} (1 + \frac{1}{2} \frac{\partial}{\partial x} T_2(x^2)|_{x=\tau}) = \tau^3 (1 + \tau T_2'(\tau^2)) \approx 0.5330644,$$

where we used $T_2(\tau) = \tau^{-1}$ from (5.8). ♢

5.2. **Rooted identity trees.** Rooted identity trees are a further Pólya structure which has been listed and treated in [7]. They do not fit into the framework of $\Omega$-Pólya trees. Nevertheless, there are some analogies to Pólya trees when our framework is applied. This section presents a discussion of what happens when we use our framework on rooted identity trees, which will eventually lead to a combinatorial interpretation of OEIS sequence A052806.

A rooted identity tree is a Pólya tree whose automorphism group is the identity group. Let $R(z)$ denote the ordinary generating function of rooted identity trees. Then we can identify every rooted identity tree as a powerset of smaller rooted identity trees, which is a multiset of rooted identity trees that involves no repetition; see [6]. This gives

$$R(z) = z \exp\left( \sum_{i=1}^{\infty} (-1)^{i-1} \frac{R(z^i)}{i} \right) = z \exp(R(z)) \exp\left( \sum_{i=2}^{\infty} (-1)^{i-1} \frac{R(z^i)}{i} \right). \tag{5.9}$$

The first few terms of $R(z)$ are

$$R(z) = z + z^2 + z^3 + 2z^4 + 3z^5 + 6z^6 + 12z^7 + 25z^8 + 52z^9 + \cdots.$$

Note that the sets of rooted identity trees can be generated as multiset of rooted identity trees with signed weights, realizing an inclusion-exclusion process. Let $\mathcal{R}$ be the set of all rooted identity trees. We consider the multiset of the elements in $\mathcal{R}$, which is denoted by $\mathrm{MSET}(\mathcal{R})$, and as before we use $\mathrm{MSET}^{(\geq 2)}(\mathcal{R})$ to denote the multiset of rooted identity trees

where each tree appears at least twice if it appears at all. Here a $D^*$-forest of size $n$ is an element of $\mathrm{MSET}^{(\geq 2)}(\mathcal{R})$. The generating function for the $D^*$-forests of rooted identity trees is

$$D^*(z) = \exp\left(\sum_{i=2}^{\infty}(-1)^{i-1}\frac{R(z^i)}{i}\right) = \sum_{k\geq 0} Z(S_k; 0, -R(z^2), \cdots, (-1)^{k-1}R(z^k))$$
$$= \sum_{k\geq 0}\frac{1}{k!}\sum_{\substack{\sigma\in S_k\\ \sigma_1=0}}(-1)^{\sigma_2+\sigma_4+\cdots}(R(z^2))^{\sigma_2}\cdots(R(z^k))^{\sigma_k}.$$

Then their cumulative weights are given by

$$d_n^* = [z^n]D^*(z) = \sum_{\substack{F\in\mathrm{MSET}^{(\geq 2)}(\mathcal{R})\\ |F|=n}}\frac{1}{|\operatorname{Aut}(F)|}\sum_{\sigma\in\operatorname{Aut}(F)\text{such that }\sigma_1=0}(-1)^{\sigma_2+\sigma_4+\cdots}$$

and a single term of this sum is the (signed) weight of a $D^*$-forest $F$. The first few terms of $D^*(z)$ are

$$D^*(z) = 1 - \frac{1}{2}z^2 + \frac{1}{3}z^3 - \frac{5}{8}z^4 + \frac{1}{30}z^5 + \frac{11}{144}z^6 - \frac{139}{840}z^7 + \cdots.$$

*Example* 5. The smallest $D^*$-forest is of size 2, and it consists of a pair of single nodes. The only fixed point free automorphism is a transposition, thus $d_2^* = -1/2$. For $n = 3$, the $D^*$ consists of 3 single nodes. The only fixed point free automorphisms are the 3-cycles, thus $d_3^* = 2/6 = 1/3$. For $n = 4$, a $D^*$-forest consists either of 4 single nodes, or of 2 identical trees, each consisting of 2 nodes and one edge. In the first case we have 6 4-cycles and 3 pairs of transpositions. In the second case we have 1 transposition swapping the two trees. Thus, $d_4^* = (-6+3)/24 - 1/2 = -5/8$. ♢

Now we define bivariate generating function in analogy to what we did for Pólya trees. Define a function via the functional equation

$$R_c(z,u) = zu\exp(R_c(z,u))\exp\left(\sum_{i=2}^{\infty}(-1)^{i-1}\frac{R(z^i)}{i}\right).$$

and set

$$R_c(z,u) = \sum_{n\geq 0} r_{c,n}(u)z^n \quad \text{where} \quad r_{c,n}(u) = \sum_{\substack{T\in\mathrm{MSET}(\mathcal{R})\\ |T|=n}} r_T(u). \tag{5.10}$$

If we set $u = 1$ then we get back the generating function of rooted identity trees. Note that the coefficients $[u^k]r_{c,n}(u)$ do not have the nice interpretation as cumulative weight of all $C$-trees identity trees of size $k$ contained in rooted identity trees of size $n$. This is because a rooted identity tree has only the trivial automorphism which means that every vertex is a fixed point and thus the whole tree is its $C$-tree. But this is in contradiction to $R(z) \neq C(z)$.

On the other hand, we have $R(z) = C(zD^*(z))$ meaning that a rooted identity tree is a $C$-tree to which $D^*$-forests have been attached. But due to the signed weights the cumulative weight of all decompositions of Pólya tree $T$ into a $C$-tree and a set of $D^*$-forests is zero if

$T$ is not a rooted identity tree and 1 otherwise, as the following computation shows: In the same way as by Theorem 2 it follows that for $T \in \mathrm{MSET}(\mathcal{R})$ and $|T| = n$,

$$r_T(u) = Z(\mathrm{Aut}(T); u, -1, 1, \ldots) = \frac{1}{|\mathrm{Aut}(T)|} \sum_{\sigma \in \mathrm{Aut}(T)} (-1)^{\sigma_2+\sigma_4+\cdots} u^{\sigma_1}.$$

Clearly, if $T \in \mathcal{R}$, then $|\mathrm{Aut}(T)| = 1$ and $\sigma_1 = n$, thus $r_T(u) = u^n$. It should be noted that if $T$ is not a rooted identity tree, *i.e.*, $T \notin \mathcal{R}$ and $|T| = n$, we have

$$Z(\mathrm{Aut}(T); 1, -1, 1, \ldots) = \frac{1}{|\mathrm{Aut}(T)|} \sum_{\sigma \in \mathrm{Aut}(T)} (-1)^{\sigma_2+\sigma_4+\cdots} = 0,$$

which implies that

$$[z^n]R(z) = \sum_{\substack{T \in \mathcal{R} \\ |T|=n}} r_T(1) = \sum_{\substack{T \in \mathrm{MSET}(\mathcal{R}) \\ |T|=n}} r_T(1).$$

*Example* 6. For $n = 3$ we have 2 Pólya trees, namely the chain $T_1$ and the cherry $T_2$. Both belong to the multiset of rooted identity trees. Obviously, $\mathrm{Aut}(T_1) = \{\mathrm{id}\}$, and $\mathrm{Aut}(T_2) = \{\mathrm{id}, \sigma\}$, where $\sigma$ swaps the two leaves but the root is unchanged. This contributes a minus sign to $r_{T_2}(u)$. Thus,

$$r_{T_1}(u) = u^3, \qquad r_{T_2}(u) = \frac{1}{2}(u^3 - u).$$

Note that the cherry $T_2$ is not a rooted identity tree, so $r_{T_2}(1) = 0$, while the chain $T_1$ is a rooted identity tree, so $r_{T_1}(u) = u^3$. For $n = 4$ we have 4 Pólya trees shown in Figure 4.1. Other than $T_3$, the other three Pólya trees belong to the multiset of rooted identity trees. Their automorphism groups are given by $\mathrm{Aut}(T_1) = \mathrm{Aut}(T_2) = \{\mathrm{id}\}$, and

$$\mathrm{Aut}(T_4) = \{\mathrm{id}, (v_2\, v_3), (v_3\, v_4), (v_2\, v_4), (v_2\, v_3\, v_4), (v_2\, v_4\, v_3)\} \cong S_3.$$

See Figure 2.2. This gives

$$r_{T_1}(u) = p_{T_2}(u) = u^4, \qquad r_{T_4}(u) = \frac{1}{6}(u^4 - 3u^2 + 2u).$$

Both $T_1$ and $T_2$ are rooted identity trees, while $T_4$ is not. ♢

In the same way as we got the composition scheme in (2.4), we can rewrite $R_c(z, u)$ as $R_c(z, u) = C(uzD^*(z))$. The expected total weight of all $C$-trees contained in all Pólya trees of size $n$, according to the weights of their decompositions into a $C$-tree and a set of $D^*$-forests is the $n$-th coefficient of $R_c(z)$, which is

$$R_c(z) := \frac{\partial}{\partial u} R_c(z, u)\Big|_{u=1} = \frac{R(z)}{1 - R(z)} = z + 2z^2 + 4z^3 + 9z^4 + 20z^5 + 46z^6 + \cdots.$$

By construction, recall (5.10), these numbers count the number of points which are fixed points in all automorphisms of Pólya trees that are generated by a root to which rooted identity trees are attached. For example consider the Pólya trees of size 4 shown in Figure 4.1. The trees $T_1$, $T_2$, and $T_4$ are constructed in this way. In these 3 trees there are in total 9 points which are always fixed points. Yet, $T_4$ is no rooted identity tree. Note that these numbers also count a simple grammar, see OEIS A052806 [15].

*Remark* 5. It would be desirable to have a similar relation between rooted identity trees and $C$-trees as we have between $C$-trees and Pólya trees. However, when setting $C(z) = R(zE(z))$ we obtain $E(z) = 1 + \frac{1}{2}z^2 - \frac{1}{3}z^3 + \frac{11}{8}z^4 - \frac{6}{5}z^5 + \frac{629}{144}z^6 + \cdots$, a power series with not only nonnegative coefficients. Thus there is no straight-forward interpretation in the desired form.

## 6. Conclusion

In this paper we develop a combinatorial framework to describe the relation between Pólya trees and simply generating trees. Since we kept the framework light, it is not strong enough to reprove the functional limit theorem presented by Panagiotou and Stufler [13], but it yields a description to this limit theorem which is to our opinion more elementary and more easily accessible to combinatorialists. In addition, we provide not only an alternative proof of the known big-$O$ result on the maximal size of $D$-forests in a random Pólya tree, but are able to extend this result. We provide a lower bound of the same order and also precise constants in both bounds. By interpreting all weights on $D$-forests and $C$-trees in terms of automorphisms associated to a Pólya tree, we derive the limiting probability that for a random node $v$ the attached $D$-forest $F_n(v)$ is of a given size as well as some structural properties.

In view of the connection between Boltzmann samplers and generating functions, it comes as no surprise that the "colored" Boltzmann sampler from [13] is closely related to a bivariate generating function. But the unified framework in analyzing the (bivariate) generating functions offers stronger results on the limiting distributions of the size of the $C$-trees and the maximal size of $D$-forests as well as more detailed knowledge on the $D$-forests within a random Pólya tree.

## References


[1] J.P. Bell, S.N. Burris and K.A. Yeats, *Counting rooted trees: the universal law* $t(n) \sim C\rho^{-n}n^{-3/2}$, Electron. J. Combin., 13(1) (2006), Research paper R63, 64p.

[2] F. Bergeron, G. Labelle and P. Leroux. *Combinatorial Species and Tree-Like Structures*. Cambridge, 1998.

[3] M. Bodirsky, É. Fusy, M. Kang and S. Vigerske. Boltzmann samplers, Pólya theory, and cycle pointing. *SIAM J. Comput.*, 40(3):721–769, 2011.

[4] M. Drmota. *Random Trees. An Interplay Between Combinatorics and Probability.* Springer Verlag, 2008.

[5] M. Drmota and B. Gittenberger. The shape of unlabeled rooted random trees. *European Journal of Combinatorics*, 31(8):2028–2063, 2010.

[6] P. Flajolet and R. Sedgewick. *Analytic Combinatorics*. Cambridge University Press, 2009.

[7] A. Genitrini. *Full asymptotic expansion for Pólya structures*. Proceedings of AofA 2016.

[8] X. Gourdon. Largest components in random combinatorial structures. *Discrete Mathematics*, 180:185–209, 1998.

[9] Svante Janson. Simply generated trees, conditioned Galton-Watson trees, random allocations and condensation. *Probab. Surv.*, 9:103–252, 2012.

[10] A. Meir and J.W. Moon. On the altitude of nodes in random trees. *Canad. J. Math.*, 30(5):997–1015, 1978.

[11] A. Nijenhuis and H.S. Wilf. *Combinatorial Algorithms.* Academic Press, Inc. [Harcourt Brace Jovanovich, Publishers], New York-London, second edition, 1978. For computers and calculators, Computer Science and Applied Mathematics.

[12] R. Otter. The number of trees. *Ann. of Math.*, 49(2):583–599, 1948.

[13] K. Panagiotou and B. Stufler. Scaling limits of random Pólya trees. *Probability Theory and Related Fields*, to appear. Preprint: arXiv:1502.07180v2, 2015.

[14] G. Pólya. Kombinatorische Anzahlbestimmungen für Gruppen, Graphen und chemische Verbindungen. *Acta Mathematica* 68(1):145–254, 1937.

[15] N.J.A. Sloane, *The On-line Encyclopedia of Integer Sequences (OEIS).*

[16] B. Stufler. Random enriched trees with applications to random graphs. *Preprint:* arXiv:1504.02006v6, 2015.

Institut für Diskrete Mathematik und Geometrie, Technische Universität Wien, Wiedner Hauptstr. 8–10/104, 1040 Vienna, Austria

*E-mail address*: gittenberger@dmg.tuwien.ac.at

*E-mail address*: yu.jin@tuwien.ac.at

*E-mail address*: michael.wallner@tuwien.ac.at