AofA'12 DMTCS proc. **AQ**, 2012, 141–154

On the number of transversals in random trees

Bernhard Gittenberger[1†] and Veronika Kraus[2†‡]

[1]*Institute of Discrete Mathematics and Geometry, TU Wien, Wiedner Hauptstr. 8-10/104, A-1040 Wien, Austria.*
[2]*Institute of Bioinformatics and Translational Research, UMIT, Eduard-Wallnoefer Zentrum 1, 6020 Hall in Tyrol, Austria.*

We study transversals in random trees with n vertices asymptotically as n tends to infinity. Our investigation treats the average number of transversals of fixed size, the size of a random transversal as well as the probability that a random subset of the vertex set of a tree is a transversal for the class of simply generated trees and for Pólya trees. The last parameter was already studied by Devroye for simply generated trees. We offer an alternative proof based on generating functions and singularity analysis and extend the result to Pólya trees.

Keywords: simply generated trees, Pólya trees, singularity analysis, functional equations

1 Introduction

A transversal of a rooted tree is a subset A of the set of nodes such that every path from the root to a leaf contains at least one node of A. Farley [5] asked for a tight lower bound of the number of transversals in a binary tree with n nodes. The question was answered by Campos et al. [1] in the more general setting of trees where each node may have at most d children ($d \geq 2$ fixed). They identified the so-called caterpillars (*cf.* [8]) as those trees where the number of transversals is minimal. Furthermore, they studied the limiting distribution of the probability that a (not necessarily uniform) random subset of the set of nodes is a transversal. The authors established bounds for this probability and asymptotic formulas for the case where the number of children of a any node is exactly d under certain restrictions on the depth of the leaves. They used clever counting arguments and probabilistic methods to achieve the results. This question was also pursued by Devroye [2] who determined the asymptotic average of this probability on the set of all simply generated trees of a given size, if the size tends to infinity.

In this paper we want to address related questions, but utilize a completely different method. We will use generating functions and singularity analysis in the style of [7] in order to study various aspects concerning the number of transversals of simply generated random trees. In particular, we offer an alternative proof of the main result in [2]. In the final section we extend the result to the class of Pólya trees. Interestingly, the average size of a transversal in a simply generated tree of size n is asymptotically $n/2$, no matter which

†This research has been supported by *FWF (Austrian Science Foundation), National Research Network S9600, grant S9604*.
‡This research has been supported by *FWF (Austrian Science Foundation), grant P22029*.

particular class of simply generated trees is taken, whereas in a Pólya tree this size is asymptotically cn where $c \approx 0.505903$. This is another proof of the fact that Pólya trees are definitely not simply generated. A simple proof of this was presented in [4], but this is to our knowledge the first tree property where the difference between these two tree classes can be actually “seen”. A precise characterization of the difference is the topic of the forthcoming paper [10].

Our method to obtain enumerative and distributional results relies on extracting coefficients of implicitly defined power series asymptotically. It should be mentioned that there is a well-developed theory on this topic, see [3, Ch. 2] and [7, Ch. VII]. This theory provides rather explicit formulae for implicitly defined functions having a singularity because the implicit function theorem fails. When counting transversals we are faced with functional equations which are themselves singular at the singularity of the functions they describe. Therefore the classical theorems are not directly applicable and some care is required during the analysis.

2 Preliminaries and Main Results

Let $\mathcal{T}$ denote a family of trees, $\mathcal{T}_n$ the subsets of those trees in $\mathcal{T}$ which have n nodes, and t_n be the number of trees in $\mathcal{T}_n$. If $T \in \mathcal{T}_n$, then let $X_{nk}(T)$ denote the number of transversals of size k in T. Moreover, set

$$t_{nk} = \sum_{T \in \mathcal{T}_n} X_{nk}(T).$$

When considering random trees, we get the random variable X_{nk}.

Some basic parameters of interest are for instance the average number of transversals in a tree of size n,

$$\mathbf{E} \sum_{k=0}^{n} X_{nk} = \frac{\sum_{k=0}^{n} t_{nk}}{t_n}$$

and the average number of transversals of size k in $\mathcal{T}_n$ which is $\mathbf{E}X_{nk} = t_{nk}/t_n$. Another interesting random variable is the size of a random transversal: Let Y_n denote this size if the random transversal is chosen at random from the set of all transversals of all trees in $\mathcal{T}_n$ where the probability of each transversal is proportional to the probability of the tree where it belongs to. We then have

$$\mathbf{P}\{Y_n = k\} = \frac{t_{nk}}{\sum_{k=0}^{n} t_{nk}}.$$

Devroye [2] studied the following problem: Take a tree T and construct a subset of its vertex set by choosing each vertex independently with probability p (where $0 < p < 1$). What is the probability that this set is a transversal? Let us call such a subset a *random p-set* and let $Z_n(T)$ be the event that a random p-set of T is a transversal. Then clearly

$$\mathbf{P}\{Z_n(T)\} = \sum_{k=0}^{n} X_{nk}(T) p^k (1-p)^{n-k}. \tag{1}$$

Since it is not easy to study this directly for a given tree, Devroye [2] computed the asymptotic average probability of Z_n in $\mathcal{T}_n$.

In this note we will focus on simply generated trees and on Pólya trees. Simply generated trees were introduced by Meir and Moon [12] and are the trees associated to a (critical) Galton-Watson branching processes with an offspring distribution which has a finite exponential moment (and therefore in particular a finite offspring variance) and which is conditioned on the size. These trees can be characterized as the tree class where the generating function $t(z) = \sum_{n\ge 0} t_n z^n$ satisfies a functional equation of the form

$$t(z) = z\varphi(t(z)) \tag{2}$$

where $\varphi(t) = \sum_{\ell\ge 0} \varphi_\ell t^\ell$ is an analytic function with the following properties: The radius of convergence is sufficiently large such that the equation

$$\tau\varphi'(\tau) = \varphi(\tau) \tag{3}$$

has a (unique) positive solution, $\varphi_\ell \ge 0$ for all $\ell \ge 0$ and $\varphi_0 > 0$. The offspring distribution of the GW-process is then given by

$$\mathbf{P}\{\xi = k\} = \tau^k \varphi_k/\varphi(\tau) \tag{4}$$

and, in particular, the offspring variance is

$$\sigma^2 = \tau^2\varphi''(\tau)/\varphi(\tau). \tag{5}$$

For technical simplicity we will moreover assume that there are no periodicities, i.e. that there is no power series $a(z)$ such that $t(z) = z^r a(z^s)$ for some integers $r \ge 0$ and $s \ge 2$. This guarantees that $t(z)$ has only one singularity on its circle of convergence.

Theorem 1. *The average size of a random transversal in trees of size n is asymptotically normally distributed. Mean and variance satisfy*

$$\mathbf{E}Y_n = \frac{\sum_{k=0}^n k t_{nk}}{\sum_{k=0}^n t_{nk}} \sim \frac{1}{2}n,$$

as $n \to \infty$, and

$$\mathbf{Var}Y_n = \frac{\sum_{k=0}^n k^2 t_{nk}}{\sum_{k=0}^n t_{nk}} - (\mathbf{E}Y_n)^2 \sim \frac{1}{4}n,$$

as $n \to \infty$.

Theorem 2 (Devroye [2])**.** *The average probability that a random p-set of a tree of size n is a transversal satisfies*

$$\lim_{n\to\infty} \mathbf{E}(\mathbf{P}\{Z_n\}) = \lim_{n\to\infty} \frac{\sum_{T\in\mathcal{T}_n} \mathbf{P}\{Z_n(T)\}}{t_n} = \frac{p}{1-(1-p)z_0\varphi'(T_0)} \tag{6}$$

where z_0 is the unique dominant singularity of the tree generating function $t(z)$ and T_0 is the solution of the equation

$$T_0 = p\tau + (1-p)\tau\left(\frac{\varphi(T_0)}{\varphi(\tau)} - \frac{\varphi_0}{\varphi(\tau)}\right) \tag{7}$$

Remark 1. Note that at the first glance (6) looks different from Devroye's [2] result. But if we set $r = T_0/\tau$, $g(r) = \varphi(\tau r)/\varphi(\tau)$ and $p_0 = \varphi_0/\varphi(\tau)$ (*cf.* Eq. (4)) then we get precisely his form.

Remark 2. Since the proof we offer is based on singular expansions of generating functions, it enables us to determine the speed of convergence as well. In fact, we have

$$\mathbf{E}(\mathbf{P}\{Z_n\}) = \frac{p}{1-(1-p)z_0\varphi'(T_0)} + \mathcal{O}\left(\frac{1}{\sqrt{n}}\right)$$

Corollary. *The average number of transversals in a tree of size n is*

$$\mathbf{E}\sum_{k=0}^{n} X_{nk} \sim \frac{2^n}{2-z_0\varphi'(T_0)}$$

Proof: Set $p = 1/2$ in Eq. (1) and then apply the theorem. □

Theorem 3. *If $k = \lambda n$, then the average number of transversals of size k in trees of size n is asymptotically given by*

$$\mathbf{E}X_{nk} \sim \frac{z_0(1-\lambda)}{1-z_0(1-\lambda)\varphi'(T_0)} \cdot \sqrt{\frac{1-\lambda}{2\pi\lambda n}} \cdot \frac{1}{\lambda^{\lambda n}(1-\lambda)^{(1-\lambda)n}}$$

where T_0 is the solution of (7) *when p is replaced by λ and the estimate is uniform for $\lambda \in [\varepsilon, 1-\varepsilon]$*

3 Proofs

Our proofs are based on generating functions and therefore we will need the tree generating function $t(x)$ and the transversal generating function

$$T(z,u) = \sum_{n\geq 0}\sum_{k=0}^{n} t_{nk} z^n u^k.$$

First we recall simple lower and upper bounds for $X_{nk}(T)$ which were already mentioned in [1]: Every subset of the vertex set which contains the root is a transversal. On the other hand every transversal is a subset of the vertex set. Thus

$$\binom{n-1}{k-1} \leq X_{nk}(T) \leq \binom{n}{k}. \tag{8}$$

Due to the observation leading to the lower bound we partition the set of transversals of $\mathcal{T}_n$ into two subsets, the set of transversals containing the root and its complement. There are $r_{nk} = t_n\binom{n-1}{k-1}$ transversals in the first set and $s_{nk} = t_{nk} - r_{nk}$ in the second one. Thus

$$R(z,u) = \sum_{n\geq 0}\sum_{k=0}^{n} r_{nk} z^n u^k = u\sum_{n\geq 0} t_n(1+u)^{n-1} z^n = \frac{u}{1+u} t(z(1+u)).$$

To proceed let us introduce the "bivariate" combinatorial structures $\mathcal{R}$ and $\mathcal{S}$ of simply generated trees with some distinguished nodes forming a transversal containing the root or not, respectively. Then an element of $\mathcal{S}$ is a tree together with a transversal such that the root is not belonging to the transversal. The root cannot be a leaf because the empty vertex set is not a transversal. The subtrees of the root can be either in $\mathcal{R}$ or in $\mathcal{S}$. Slightly abusing the notation, we denote by $\mathcal{T}$ the set of all trees with a distinguished transversal. ¿From the context it should always be clear whether $\mathcal{T}$ is just the class of tree or there is in addition a distinguished transversal associated to each tree. We obtain therefore the specification

$$\begin{aligned}\mathcal{T} &= \mathcal{R} \cup \mathcal{S}\\ \mathcal{S} &= \{\circ\} \times \mathtt{set}_{\geq 1}(\mathcal{R} \cup \mathcal{S}).\end{aligned}$$

Here $\mathtt{set}_{\geq 1}$ denotes a generalized set construction where the sets are non-empty and weighted according to the weight function $\varphi(t)$ given in (2). Substituting the first into the second equation and translating into generating functions yields

$$S(z,u) = z\varphi(T(z,u)) - z\varphi_0$$

and hence

$$T(z,u) = uz\varphi(t(z(1+u))) + z\varphi(T(z,u)) - z\varphi_0. \tag{9}$$

Proof of Theorem 2. The quantity we are interested in is

$$\mathbf{E}(\mathbf{P}\{Z_n\}) = \frac{\sum_{T\in\mathcal{T}_n} \mathbf{P}\{Z_n(T)\}}{t_n} = \frac{[z^n]T\left((1-p)z, \frac{p}{1-p}\right)}{t_n}.$$

The denominator is well known and was first computed in [12]. We have, as n tends to ∞,

$$t_n \sim \frac{\tau}{\sigma z_0^n \sqrt{2\pi n^3}}. \tag{10}$$

In order to cope with the numerator, the usual procedure would be determining the singularity of $z \mapsto T(z,u)$ by using the theory of implicit functions (*cf.* [3, Th. 2.19]). This means that we have to solve the system

$$\begin{aligned}T(z,u) &= uz\varphi(t(z(1+u))) + z\varphi(T(z,u)) - z\varphi_0,\\ 1 &= z\varphi'(T(z,u)),\end{aligned} \tag{11}$$

with respect to z and T where u is considered a parameter. Let us consider the case $u > 0$. Then the singularity $\rho_0(u)$ of the function $z \mapsto T(z,u)$ can be easily determined. In fact, using the inequalities (8) we obtain

$$\frac{u}{1+u}t(z(1+u)) = \sum_{n\geq 0}\sum_{k=0}^{n}\binom{n-1}{k-1}t_n u^k z^n \leq T(z,u) \leq \sum_{n\geq 0}\sum_{k=0}^{n}\binom{n}{k}t_n u^k z^n = t(z(1+u)).$$

Thus $\rho_0(u) = z_0/(1+u)$ which is precisely the singularity of the first term on the right-hand side of the first equation of the system (11). Let us write this functional equation in the form $F(z,u,T) = 0$. The considerations above show that at the singularity $\rho_0(u)$ of the solution $T(z,u)$ the funcion F is itself singular. Hence [3, Th. 2.19] is not directly applicable and we have to take some care when determining the nature of the singularity of $T(z,u)$. Usually, the location of the dominant singularity of an implicitly defined generating function is given by the nearest point to the origin where the assumptions of the implicit function theorem are violated. The first lemma shows that there is no such point within the domain of convergence of the solution $T(z,u)$. Therefore, roughly speaking, the singularity does not originate from the functional equation but rather from the singularity of $t(z)$.

Lemma 1. *Let $u > 0$ and set $\rho_0(u) := \frac{z_0}{1+u}$ and $T_0(u) := T(\rho_0(u), u)$. Then*

$$T_0(u) < \tau \text{ and } \rho_0(u)\varphi'(T_0(u)) < 1. \tag{12}$$

Consequently, the system (11) *does not have a solution $(z, T) = (\rho(u), T(\rho(u), u))$ with $|z| \le \rho_0(u)$.*

Proof: Observe that (8) implies

$$\begin{aligned} T_0(u) &= \sum_{n\ge 0}\sum_{k=0}^{n} t_{nk}\left(\frac{u}{1+u}\right)^k\left(\frac{1}{1+u}\right)^{n-k} z_0^n \\ &< \sum_{n\ge 0}\sum_{k=0}^{n}\binom{n}{k} t_n\left(\frac{u}{1+u}\right)^k\left(\frac{1}{1+u}\right)^{n-k} z_0^n. \end{aligned} \tag{13}$$

Since $\sum_{k=0}^{n}\binom{n}{k}\left(\frac{u}{1+u}\right)^k\left(\frac{1}{1+u}\right)^{n-k} = 1$ the expression in Eq. (13) equals $t(z_0) = \tau$, so we obtain $T_0(u) < \tau$.

The function $\varphi(t)$ has only non-negative coefficients and is therefore monotone on the positive real axis. The same applies to $\varphi'(t)$. Thus $\varphi'(T_0(u)) < \varphi'(\tau)$ and since $u > 0$ we have $\rho_0(u) = z_0/(1+u) < z_0$. The relation $z_0\varphi'(\tau) = 1$ now completes the proof of (12).

The second assertion easily follows, since $\rho_0(u)$ is the singularity of the first term of the first equation of (11). $\square$

Now turning to Theorem 2 let $u_0 = p/(1-p)$. Further, set $\rho_0 := \rho_0(u_0) = (1-p)z_0$ and $T_0 := T_0(u_0)$. The functional equation (9) becomes then

$$\begin{aligned} T_0 &= u_0\rho_0\varphi(t(\rho_0(1+u_0))) + \rho_0(\varphi(T_0) - \varphi_0) \\ &= pz_0\varphi(t(z_0)) + (1-p)z_0(\varphi(T_0) - \varphi_0) \\ &= p\tau + (1-p)\frac{\tau}{\varphi(\tau)}(\varphi(T_0) - \varphi_0). \end{aligned}$$

This is precisely the equation defining T_0, Eq. (7).

Next we determine the nature of the singularity of $T(z, u)$. We know that for $z \to \rho_0 = z_0/(1+u_0)$ the following singular expansion holds:

$$t(z(1+u_0)) = \tau - \frac{\tau\sqrt{2}}{\sigma}\sqrt{1 - \frac{z(1+u_0)}{z_0}} + \mathcal{O}\left(\left|1 - \frac{z(1+u_0)}{z_0}\right|\right). \tag{14}$$

Hence, locally around $z = \rho_0$ we have

$$\begin{aligned} T(z,u_0) = T_0 + (T(z,u_0) - T_0) &= pt(z(1+u_0)) + z\varphi(T(z,u_0)) - \rho_0\varphi_0 \\ &= \underbrace{p\tau + \rho_0\varphi(T_0) - \rho_0\varphi_0}_{=T_0} - \frac{p\tau\sqrt{2}}{\sigma}\sqrt{1 - \frac{z}{\rho_0}} \\ &\quad + \rho_0\varphi'(T_0)(T(z,u_0) - T_0) + \varphi(T_0)(z - \rho_0) \\ &\quad + \mathcal{O}\left(|z-\rho_0|^2\right) + \mathcal{O}\left(|T(z,u_0) - T_0|^2\right) \end{aligned}$$

Solving this equation w.r.t. $T(z,u_0)-T_0$ we obtain the singular expansion

$$T(z,u_0)=T_0-\frac{\tau\sqrt{2}}{\sigma}\cdot\frac{p}{1-\rho_0\varphi'(T_0)}\sqrt{1-\frac{z}{\rho_0}}+\mathcal{O}\left(|z-\rho_0|\right) \tag{15}$$

Comparing this with the singular expansion of the tree function (*cf.* Eq. (14)) and applying a transfer lemma of Flajolet and Odlyzko [6] (*cf.* also [7, Ch. VI]) we obtain (6) and the proof of Theorem 2 is complete. □

Proof of Theorem 1. In the proof of Theorem 2 we needed only that u is a real-valued parameter. In order to get asymptotic normality we need uniform error estimates when u is in a complex vicinity of 1. We turn back to the functional equation (9). The tree function appearing in the first term on the right-hand side admits the local representation

$$t(z(1+u))=g(z,u)-h(z,u)\sqrt{1-\frac{z}{\rho_0(u)}}$$

where $g(z,u)$ and $h(z,u)$ are analytic around $(\rho_0(u),u)$. In particular, this implies that the error term in (14) is uniform for $|u-1|<\varepsilon$ and $z\to\rho_0(u)$ within the domain $\frac{1}{1+u}\cdot\Delta$ where

$$\Delta=\Delta(z_0,\eta,\vartheta)=\{z\,|\,|z|\le z_0+\eta,|\arg(z-z_0)|\ge\vartheta\}$$

is the classical pacman shaped domain used in [6]. The system of functional equations (11) can then be rewritten to the form

$$T=F(z,u,T)+g(z,u)-h(z,u)\sqrt{1-\frac{z}{\rho_0(u)}},\qquad 1=F_T(z,u,T).$$

Then $1>F_T(\rho_0(u),u,T_0(u))$ for $|u-1|<\varepsilon$ and $z\in\frac{1}{1+u}\cdot\Delta$. Hence the implicit function theorem implies that the functional equation

$$T=\tilde{F}(z,u,v,T)=F(z,u,T)+v$$

has an analytic solution $T(z,u,v)$ for $|u-1|<\varepsilon$, $z\in\frac{1}{1+u}\cdot\Delta$ and $|v|\le|g(\rho_0(u),u)|+\varepsilon$. Plugging in $v=g(z,u)-h(z,u)\sqrt{1-\frac{z}{\rho_0(u)}}$ we obtain the local representation

$$\begin{aligned}T(z,u)&=T_1(z,u)-T_2(z,u)\sqrt{1-\frac{z}{\rho_0(u)}}\\&=T_0(u)-\frac{\tau\sqrt{2}}{(1+u)\sigma(1-\rho_0(u)\varphi'(T_0(u)))}\sqrt{1-\frac{z}{\rho_0(u)}}+\mathcal{O}\left(|z-\rho_0(u)|\right)\end{aligned}$$

with analytic functions $T_1(z,u)$ and $T_2(z,u)$. Hence the error term in the last equation is uniform. Now by [3, Lemma 2.18] we obtain

$$[z^n]T(z,u)=\frac{\tau}{(1+u)\sigma(1-\rho_0(u)\varphi'(T_0(u)))}\cdot\frac{\rho_0(u)^{-n}}{\sqrt{2\pi n^3}}+\mathcal{O}\left(\frac{\rho_0(u)^{-n}}{n^{5/2}}\right), \tag{16}$$

uniformly for $|u-1|<\varepsilon$. Thus

$$\mathbf{E}u^{Y_n} = \frac{2u(1-\rho_0(u)\varphi'(T_0(1)))}{(1+u)(1-\rho_0(u)\varphi'(T_0(u)))}\left(\frac{\rho_0(1)}{\rho_0(u)}\right)^n (1+\mathcal{O}\left(n^{-1}\right)).$$

Finally, we can apply Hwang's quasi-power theorem [9] to obtain asymptotic normality.

Let $\mu = \frac{\rho_0'(1)}{\rho_0(1)}$. By a closer look at the theorem, the above representation implies that

$$\mathbf{E} = \mu n + \mathcal{O}(1)$$

$$\mathbf{Var} = (\mu + \mu^2 - \frac{\rho_0''(1)}{\rho_0(1)})n + \mathcal{O}(1)$$

Hence with $\rho_0(u) = \frac{z_0}{1+u}$ it is easy to compute $\mathbf{E}Y_n = \frac{1}{2}n$ and $\mathbf{Var}Y_n = \frac{1}{4}n$. □

Proof of Theorem 3. We have $\mathbf{E}X_{nk} = t_{nk}/t_n$. The denominator is estimated in (10). For the numerator observe that the shape of the coefficients in (16) is of the form $A(u)B(u)^n$ with

$$A(u) = \frac{\tau}{(1+u)\sigma\sqrt{2\pi n^3}(1-\rho_0(u)\varphi'(T_0(u)))} \quad \text{and} \quad B(u) = \frac{1}{\rho_0(u)} = \frac{1+u}{z_0}.$$

This function satisfies the assumptions of [7, Proposition VIII.8]. Hence a straight forward saddle point method is applicable: The saddle point is given by $\zeta B'(\zeta)/B(\zeta) = \lambda$ and therefore $\zeta = \lambda/(1-\lambda)$. Now a direct application leads to the assertion and the proof is complete. □

4 Pólya Trees

Pólya trees are non-plane rooted trees without any further restrictions and where the set of all trees of the same size is equipped with the uniform distribution. The methods we were using to analyze simply generated trees allow an analysis of Pólya trees as well. Of course, the specification is similar. The only difference is that we have to take into account symmetries since isomorphic trees are considered as the same tree. The tree function is defined by

$$t(z) = z\exp\left(\sum_{i\geq 1}\frac{t(z^i)}{i}\right).$$

and it is well known that $t(z)$ has a unique singularity ρ on its circle of convergence and near it admits a local expansion of the form

$$t(z) = 1 - b\sqrt{\rho - z} + \mathcal{O}\left(|\rho - z|\right), \text{ as } z\to\rho \text{ in } \Delta = \Delta(\rho,\eta,\vartheta).$$

This was already shown by Pólya [14] who also computed $\rho\approx 0.33832$ numerically. Otter [13] computed further coefficients of the singular expansion numerically, among them $b\approx 2.681266$. Applying a transfer lemma [6] yields

$$t_n = [z^n]t(z) \sim \frac{b\rho^{-n+\frac{1}{2}}}{2\sqrt{\pi n^3}}, \text{ as } n\to\infty.$$

As before we introduce the random variables X_{nk} (number of transversals of size k in a random tree of size n) and Y_n (size of a random transversal chosen uniformly from all transversals in all size n trees) as well as the event Z_n (random p-set on a size n tree is a transversal). Note that (1) is likewise true for Pólya trees since the symmetries are already taken into account by the random variable X_{nk}.

As in the previous section we work with the bivariate combinatorial structure $\mathcal{T}$ denoting the set of pairs (T, S) where T is a tree and S one of its transversals as well as $\mathcal{R}$ and $\mathcal{S}$ denoting the subsets of $\mathcal{T}$ where the transversal contains the root or not, respectively. Then we get

$$\mathcal{T} = \mathcal{R} \cup \mathcal{S} \tag{17}$$

$$\mathcal{S} = \{\circ\} \times \texttt{multiset}_{\geq 1}(\mathcal{R} \cup \mathcal{S}). \tag{18}$$

As before, any subset of the set of vertices which contains the root is a transversal. Nethertheless, due to the non-planarity we do not get such a tight lower bound for the number of transversals as in (8), but the best possible lower bound is 1. Indeed, the star graph (root with $n-1$ leaves) contains exactly one transversal of size k for $k = 1, \ldots, n$, since the children of the root are not distinguishable. Therefore, we introduce the combinatorial structure $\tilde{\mathcal{T}}$, the set of all rooted trees with an arbitrary number of distinguished vertices. Then we obtain the following specifications:

$$\tilde{\mathcal{T}} = \{\circ\} \times \texttt{multiset}(\tilde{\mathcal{T}}) \tag{19}$$

$$\mathcal{R} = \{\bullet\} \times \texttt{multiset}(\tilde{\mathcal{T}}) \tag{20}$$

In terms of generating functions (17)–(20) translate to

$$\tilde{T}(z,u) = z(1+u)\exp\left(\sum_{i\geq 1} \frac{\tilde{T}(z^i, u^i)}{i}\right) \tag{21}$$

and

$$\begin{aligned} T(z,u) &= zu\exp\left(\sum_{i\geq 1} \frac{\tilde{T}(z^i, u^i)}{i}\right) + z\left(\exp\left(\sum_{i\geq 1} \frac{T(z^i, u^i)}{i}\right) - 1\right) \\ &= \frac{u}{1+u}\tilde{T}(z,u) + z\left(\exp\left(\sum_{i\geq 1} \frac{T(z^i, u^i)}{i}\right) - 1\right) \end{aligned} \tag{22}$$

We will show the following theorems:

Theorem 4. *The average probability that a random p-set of a Pólya tree of size n is a transversal satisfies*

$$\lim_{n\to\infty} \mathbf{E}(\mathbf{P}\{Z_n\}) = p\left(1 - \rho_0 \exp\left(\sum_{i\geq 1} \frac{T(\rho_0^i, (p/(1-p))^i)}{i}\right)\right)^{-1}$$

where ρ_0 is the unique positive singularity of $\tilde{T}(z, p/(1-p))$ where $\tilde{T}(z,u)$ is defined implicitly by (21) *and $T(z,u)$ is defined implicitly by* (22)*.*

Corollary. *The average number of transversals in a tree of size* n *is*

$$\mathbf{E}\sum_{k=0}^{n} X_{nk} \sim \frac{2^n}{2-2\tilde{\rho}\exp\left(\sum_{i\geq 1}\frac{T(\tilde{\rho}^i,1)}{i}\right)},$$

where $\tilde{\rho} \approx 0.180343$ *is the dominant singularity of* $\tilde{t}(z)$*, which is defined implicitly by*

$$\tilde{t}(z) = 2z\exp\left(\sum_{i\geq 1}\frac{\tilde{t}(z^i)}{i}\right).$$

Numerically, we have the approximation

$$\frac{1}{2-2\tilde{\rho}\exp\left(\sum_{i\geq 1}\frac{T(\tilde{\rho}^i,1)}{i}\right)} \approx 0.6126998.$$

Remark. The function $\tilde{t}(z)$ satisfies $\tilde{t}(z) = \tilde{T}(z,1)$ and is interesting in its own right: In fact the sequence of coefficients

$$\tilde{t}(z) = 2z + 4z^2 + 14z^3 + 52z^4 + 214z^6 + 916z^7 + 4116z^8 + 18996z^9 + \cdots \tag{23}$$

is precisely sequence A000515 in Sloane online encyclopedia [15] multiplied by a factor 2. This sequence appeared already in various context of enumerative combinatorics, see [11, 16].

Theorem 5. *The average size of a random transversal in Pólya trees of size* n *is asymptotically normally distributed. Mean and variance satisfy*

$$\mathbf{E}Y_n \sim an,$$

as $n \to \infty$*, where* $a \approx 0.505903$ *and*

$$\mathbf{Var}Y_n \sim bn,$$

as $n \to \infty$*, for some explicitly (numerically) computable constant.*

Remark. Note that we observe here an actual difference between Pólya trees and simply generated trees. While for every simply generated family the asymptotic average size of a transversal is precisely $n/2$, for Pólya trees it is slightly larger. Thus the family of Pólya trees cannot be realized by a conditioned Galton-Watson process. A proof of this fact was already given in [4], but this proof, though quite simple, was an argument for properties of the generating function. The above theorem makes the difference between those two classes visible on the level of the actual trees.

As in the simply generated case, we will show that the origin of the dominant singularity of $T(z,u)$ is the singularity of $\tilde{T}(z,u)$ and not a point (z,T) where the implicit function theorem cannot be applied to (22).

Lemma 2. *Let* $u > 0$*. Then within the domain of convergence of* $z \mapsto T(z,u)$ *the implicit function theorem is always applicable to* (22)*. Thus the mappings* $z \mapsto T(z,u)$ *and* $z \mapsto \tilde{T}(z,u)$ *have the same dominant singularity.*

Proof: Let $\tilde{\rho}(u)$ denote the dominant singularity of the function $z \mapsto \tilde{T}(z,u)$ and $\bar{\rho}(u)$ be the smallest positive value on the positive real axis such that the functional equation (22) does not fulfill the assumptions of the implicit function theorem at the point $(\bar{\rho}, T) = (\bar{\rho}(u), T(\bar{\rho}(u), u))$. Moreover, set

$$\tilde{Q}(z,u) := \exp\left(\sum_{i\geq 2} \frac{\tilde{T}(z^i,u^i)}{i}\right) \text{ and } Q(z,u) := \exp\left(\sum_{i\geq 2} \frac{T(z^i,u^i)}{i}\right) \tag{24}$$

Then $(\tilde{\rho}(u), \tilde{T}(\tilde{\rho}(u), u)$ satisfies (21) as well as

$$1 = (1+u)\tilde{\rho}(u)e^{\tilde{T}(\tilde{\rho}(u),u)}\tilde{Q}(\tilde{\rho}(u), u).$$

Hence $\tilde{T}(\tilde{\rho}(u), u) = 1$. Furthermore, $\bar{\rho}(u)$ must fulfill

$$1 = \bar{\rho}(u)e^{T(\bar{\rho}(u),u)}Q(\bar{\rho}(u), u).$$

Let us assume that $\bar{\rho}(u) \leq \tilde{\rho}(u)$, i.e., $\bar{\rho}(u)$ lies within the domain of convergence of the power series of the mapping $z \mapsto \tilde{(}z,u)$. Now recall that $[z^n u^k]\tilde{T}(z,u)$ is the number of all k-element subsets of vertices in all trees of size n (counted up to symmetries) whereas $[z^n u^k]T(z,u)$ is the number of those of these subsets which are transversals. Since $u > 0$ this implies $\tilde{T}(z,u) > T(z,u)$ for positive z. Thus

$$\bar{\rho}(u)e^{\tilde{T}(\bar{\rho}(u),u)}\tilde{Q}(\bar{\rho}(u),u) > 1 = (1+u)\tilde{\rho}(u)e^{\tilde{T}(\tilde{\rho}(u),u)}\tilde{Q}(\tilde{\rho}(u),u) > \tilde{\rho}(u)e^{\tilde{T}(\tilde{\rho}(u),u)}\tilde{Q}(\tilde{\rho}(u),u).$$

Since the function $z \mapsto ze^{\tilde{T}(z,u)}\tilde{Q}(z,u)$ is a power series with non-negative coefficients, it is strictly increasing which contradicts the assumption $\bar{\rho}(u) \leq \tilde{\rho}(u)$. □

Proof of Theorem 4. In view of the preceding lemma again we are faced with a situation where [3, Th. 2.19] is not applicable and therefore we have to analyze the nature of the singularity by going back directly to multivariate Taylor expansions as in the previous section. Set

$$u_0 := \frac{p}{1-p}, \quad \rho_0 := \tilde{\rho}(u_0), \quad T_0 := T(\rho_0, u_0), \quad Q_0 := Q(\rho_0, u_0).$$

Then, using (22) as well as the singular expansion

$$\tilde{T}(z,u) = g(z,u) - h(z,u)\sqrt{1 - \frac{z}{\rho_0}},$$

we obtain (omitting the arguments of $T = T(z, u_0)$)

$$\begin{aligned} T = T_0 + (T - T_0) &= p\underbrace{\tilde{T}(\rho_0, u_0)}_{=1} + z\left(e^T Q_0 - 1\right) \\ &= \underbrace{pg(\rho_0,u_0) + \rho_0(e^{T_0}Q_0 - 1)}_{=T_0} - ph(\rho_0,u_0)\sqrt{1 - \frac{z}{\rho_0}} \\ &\quad + \rho_0 e^{T_0} Q_0 \cdot (T - T_0) + \mathcal{O}\left(|z - \rho_0|\right) + \mathcal{O}\left(|T - T_0|^2\right), \end{aligned}$$

and thus

$$T(z,u_0) \sim T_0 - \frac{ph(\rho_0,u_0)}{1-\rho_0 e^{T_0} Q_0}\sqrt{1-\frac{z}{\rho_0}}. \tag{25}$$

Comparison with the singular expansion of $\tilde{T}(z,u)$ completes the proof. □

Proof of Theorem 5. Asymptotic normality can be shown in a similar way as in the proof of Theorem 1: We need to prove that we have uniform error terms for complex u close to 1 and that we can thus apply the quasi-power theorem. This then guarantees that the moment generating function admits a representation of the form

$$\mathbf{E}u^{Y_n} F(u)\left(\frac{\rho_0(1)}{\rho_0(u)}\right)^n$$

which is uniform. Thus we have asymptotic normality as well as an expectation and variance which are asymptotically proportional to n, as $n \to \infty$. We omit the details here and are left with the calculation of the proportionality factors. We can compute expectation and variance of Y_n by

$$\begin{aligned}\mathbf{E}Y_n &= \frac{[z^n]\frac{\partial}{\partial u}T(z,u)\big|_{u=1}}{[z^n]T(z,1)},\\ \mathbf{Var}Y_n &= \frac{[z^n]\frac{\partial}{\partial u}\left(u\frac{\partial}{\partial u}T(z,u)\right)\big|_{u=1}}{[z^n]T(z,1)} - (\mathbf{E}Y_n)^2.\end{aligned}$$

Here we will confine ourselves to the mean, since the computation of the variance would require determining third order asymptotics of the mean as well as the partial derivatives of $T(z,u)$ and would lead to rather messy expressions involving lots of constants which can only be computed numerically.

In the following, we indicate partial derivatives by the use of subscripts, e.g. $T_u(z,u) = \frac{\partial}{\partial u}T(z,u)$. Furthermore, we use the notations from (24) and let $\tilde{\rho} := \tilde{\rho}(1)$. Differentiation of equation (22) with respect to u and setting $u=1$ gives

$$\begin{aligned}T_u(z,1) &= \frac{1}{2}\tilde{T}_u(z,1) + \frac{1}{4}\tilde{T}(z,1) + ze^{T(z,1)}Q_u(z,1) + ze^{T(z,1)}Q(z,1)T_u(z,1)\\ T_u(z,1) &= \frac{\frac{1}{2}\tilde{T}_u(z,1) + \frac{1}{4}\tilde{T}(z,1) + ze^{T(z,1)}Q_u(z,1)}{1-ze^{T(z,1)}Q(z,1)}\end{aligned}$$

where $\tilde{T}_u(z,u)$ can be determined by differentiation of (21):

$$\begin{aligned}\tilde{T}_u(z,1) &= ze^{\tilde{T}(z,1)}\tilde{Q}(z,1) + 2ze^{\tilde{T}(z,1)}\tilde{Q}_u(z,1) + 2ze^{\tilde{T}(z,1)}\tilde{Q}(z,1)\tilde{T}_u(z,1)\\ \tilde{T}_u(z,1) &= \frac{ze^{\tilde{T}(z,1)}\tilde{Q}(z,1) + 2ze^{\tilde{T}(z,1)}\tilde{Q}_u(z,1)}{1-\tilde{T}(z,1)}\\ &\sim \frac{\frac{1}{2} + 2\tilde{\rho}eQ_u(\tilde{\rho},1)}{1-\tilde{T}(z,1)}, \quad as \quad z \to \tilde{\rho}.\end{aligned}$$

The function $\tilde{t}(z) = \tilde{T}(z,1)$ satisfies the functional equation (23). With the help of MAPLE we can get a numerical approximation for the singularity $\tilde{\rho}$ and a square-root type Puiseux expansion locally around $\tilde{\rho}$ from there:

$$\tilde{t}(z) = 1 - h(z)\sqrt{1-\frac{x}{\tilde{\rho}}} \sim 1 - h(\tilde{\rho})\sqrt{1-\frac{x}{\tilde{\rho}}},$$

where $h(\tilde{\rho}) = 2\sqrt{\tilde{\rho}e(\tilde{Q}(\tilde{\rho},1) + \tilde{\rho}\tilde{Q}_z(\tilde{\rho},1)} \approx 1.41504$. This implies

$$T_u(z,1) \sim \frac{1}{2}\tilde{T}_u(z,1) \sim \frac{1 + 4\tilde{\rho}e\tilde{Q}_u(\tilde{\rho},1)}{4h(\tilde{\rho})\sqrt{1-\frac{x}{\tilde{\rho}}}}.$$

The normalization factor $[z^n]T(z,1)$ can be computed by means of (25), which finally yields

$$\mathbf{E}Y_n \sim \frac{1 + 4\tilde{\rho}e\tilde{Q}_u(\tilde{\rho},1)}{h(\tilde{\rho})^2}n$$

with

$$\frac{1 + 4\tilde{\rho}e\tilde{Q}_u(\tilde{\rho},1)}{h(\tilde{\rho})^2} \approx 0.505903.$$

□

An analogue of Theorem 3 for Pólya trees can be derived in a similar way as that for simply generated trees, but since we are lacking an explicit representation of $\rho_0(u)$ we will not get any nice form. Thus we decided to skip this.

Acknowledgments

The authors thank Luc Devroye for pointing out reference [2] to us.

References

[1] Victor Campos, Vašek Chvátal, Luc Devroye, and Perouz Taslakian. Transversals in trees. *J. Graph Theory*, 2011. to appear.

[2] Luc Devroye. A note on the probability of cutting a Galton-Watson tree. *Electron. J. Probab.*, 16:2001–2019, 2011.

[3] Michael Drmota. *Random Trees*. Springer, Wien, NewYork, 2009. An interplay between combinatorics and probability.

[4] Michael Drmota and Bernhard Gittenberger. The shape of unlabeled rooted random trees. *European J. Combin.*, 31(8):2028–2063, 2010.

[5] J.D. Farley. Breaking Al Qaeda cells: a mathematical analysis of counterterrorism operations (a guide for risk assessment and decision making). *Studies in Conflict and Terrorism*, 26:399–411, 2003.

[6] Ph. Flajolet and A. M. Odlyzko. Singularity analysis of generating functions. *SIAM Journal on Discrete Mathematics*, 3:216–240, 1990.

[7] Philippe Flajolet and Robert Sedgewick. *Analytic combinatorics*. Cambridge University Press, Cambridge, 2009.

[8] Frank Harary and Allen J. Schwenk. The number of caterpillars. *Discrete Math.*, 6:359–365, 1973.

[9] Hsien-Kuei Hwang. On convergence rates in the central limit theorems for combinatorial structures. *European J. Combin.*, 19(3):329–343, 1998.

[10] Veronika Kraus and Konstantinos Panagiotou. Random Labeled Trees in Random Unlabeled Trees, 2012. manuscript.

[11] Pierre Leroux and Brahim Miloudi. Généralisations de la formule d'Otter. *Ann. Sci. Math. Québec*, 16(1):53–80, 1992.

[12] A. Meir and J. W. Moon. On the altitude of nodes in random trees. *Canadian Journal of Mathematics*, 30:997–1015, 1978.

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

[14] George Pólya. Kombinatorische Anzahlbestimmungen für Gruppen, Graphen und chemische Verbindungen. *Acta Math.*, 68:145–254, 1937.

[15] N. J. A. Sloane. *The on-line Encyclopedia of Integer Sequences*. Published electronically at `http://www.research.att.com/~njas/sequences/`., 2006.

[16] Stephan G. Wagner. On an identity for the cycle indices of rooted tree automorphism groups. *Electron. J. Combin.*, 13(1):Note 14, 7, 2006.