Applicable Analysis and Discrete Mathematics
available online at http://pefmath.etf.rs

Appl. Anal. Discrete Math. **17** (2023), 001–024.
https://doi:10.2298/AADM190329010G

PROTECTION NUMBERS IN SIMPLY GENERATED TREES AND PÓLYA TREES

Bernhard Gittenberger, Zbigniew Gołębiewski,
Isabella Larcher and Małgorzata Sulkowska

We determine the limit of the expected value and the variance of the protection number of the root in simply generated trees, in Pólya trees, and in unlabelled non-plane binary trees, when the number of vertices tends to infinity. Moreover, we compute expectation and variance of the protection number of a randomly chosen vertex in all those tree classes. We obtain exact formulas as sum representations, where the obtained sums are rapidly converging thus allowing an efficient numerical computation of high accuracy.

1. INTRODUCTION

The protection number of a tree is the length of the shortest path from the root to a leaf. It is interchangeably called *the protection number of the root.* We define *the protection number of a vertex* v in tree T as the protection number of a maximal subtree of T having v as its root. We say that a vertex is *k-protected* if k does not exceed its protection number.

Previous research concerning protection numbers has been conducted in two closely related directions: (i) the number of k-protected vertices in a tree of size n, and (ii) the protection number of a root or a random vertex.

Cheon and Shapiro [**3**] were the first ones to investigate the number of 2-protected nodes in trees. They stated the results for unlabelled ordered trees and Motzkin trees. Later on Mansour [**22**] complemented their work by solving

2020 Mathematics Subject Classification. 05C05, 05A15, 05A16.
Keywords and Phrases. Protection number, Simply generated trees, Pólya trees, Non-plane binary trees.

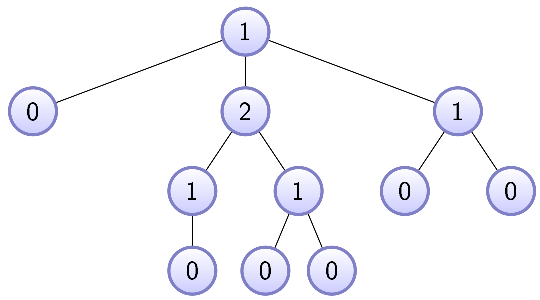

Figure 1: Tree with vertices holding their protection numbers.

the case of k-ary trees. Over the next several years these results were followed by a series of papers examining the number of k-protected nodes (usually for small values of k) in various models of random trees. To mention just a few, Du and Prodinger [**11**] analysed the average number of 2-protected nodes in random digital search trees, Mahmoud and Ward [**20**] presented a central limit theorem as well as exact moments of all orders for the number of 2-protected nodes in binary search trees and three years later they found the number of 2-protected nodes in recursive trees (consult [**21**]). The family of binary search trees was investigated also by Bóna and Pittel [**2**] who showed that the number of its k-protected nodes decays exponentially in k.

In 2015 Holmgren and Janson [**18**] went for more general results. Using probabilistic methods, they derived a normal limit law for the number of k-protected nodes in a binary search tree and a random recursive tree.

Soon after, two particular parameters attracted the attention of the algorithmic community. These were (as already mentioned earlier) the protection number of the root and the protection number of a random vertex. In 2017 Copenhaver [**4**] found that in a random unlabelled plane tree the expected value of the protection number of the root and the expected value of the protection number of a random vertex approach 1.62297 and 0.727649, respectively, as the size of the tree tends to infinity. These results were extended by Heuberger and Prodinger [**17**]. They showed the exact formulas for the first terms of the expectation, the variance and the probability of the respective protection numbers.

The protection number of the root is closely related to parameters called *minimal fill-up level* and *saturation level*. These were studied previously by, among others, Devroye [**6**] and Drmota [**8, 9**].

The aim of this paper is to generalize the protection number results to a larger class of rooted trees. We study both the root protection number as well as a random vertex protection number for the family of simply generated trees (introduced by Meir and Moon [**23**]) and their non-plane counterparts: unlabelled non-plane rooted trees, also called Pólya trees due to their first extensive treatment by Pólya [**26**], examined further by Otter [**24**] including numerical results and the binary case. The present paper broadens the results from [**17**], but maintaining the emphasis on as concrete formulas as possible.

For simply generated trees a general theory of asymptotics of additive cost functionals was developed recently in [**5**], but this theory, which is based on embeddings into Brownian excursion and weak limit theorems, does not cover functionals in the local regime (*i.e.*, functional with small toll functions), such as the number of protected nodes. Devroye and Janson [**7**] presented a unified approach to obtaining the number of k-protected nodes in various classes or random trees by putting them in the general context of fringe subtrees introduced by Aldous in [**1**]. We have obtained analogous results for simply generated trees, but employing a different methodology. This allows an efficient numerical treatment and may serve as a basis for random generation in the framework of Boltzmann sampling [**12**]. Parts of our investigations fall into the general framework of additive functionals treated in [**27**], but our focus on concrete expressions allows an easy access to numerical evaluation of the considered parameters.

Plan of the paper

In Sections 2, 3, and 4 we consider simply generated trees, Pólya trees and non-plane binary trees, respectively. In each section the expected value and the variance of the protection number of the root and the protection number of a random vertex are computed. All these quantities tend to constants when the tree size tends to infinity. The emphasis is on deriving exact expressions for these constants in terms of characteristic parameters of the considered tree class. We obtain them in terms of sums that converge at an exponential rate and therefore enable us to compute efficiently accurate numerical values. We provide numerical values for several well-known simply generated tree classes as well as for two non-plane classes studied in Sections 3 and 4.

2. SIMPLY GENERATED TREES

2.1 Protection number of the root

The class $\mathcal{T}$ of simply generated trees was introduced in [**23**] and can be described as the class of plane rooted trees whose generating function satisfies a functional equation of particular type: If t_n denotes the sum of the weights of all trees with n vertices, then the generating function $T(z) = \sum_{n\geq 0} t_n z^n$ satisfies

$$T(z) = z\phi(T(z)),$$

where the power series $\phi(t) = \sum_{j\geq 0} \phi_j t^j$ has only non-negative coefficients, $\phi_0 > 0$, and there is a $j \geq 2$ such that $\phi_j > 0$. Moreover, it is required that the equation $\tau\phi'(\tau) = \phi(\tau)$ has a unique positive solution.

We are interested in the asymptotic protection number of a random simply generated tree, sampled according to the weights from all simply generated trees with n vertices, where n tends to infinity.

Remark. For the sake of simplicity we assume that ϕ is non-periodic, meaning that $\gcd\{j > 0 : \phi_j > 0\} = 1$. The periodic case can be dealt with in the very same way, but the calculations leading to the desired number have to be done repeatedly (for analogous situations) in order to collect several contributions to the final value.

Within this paper the primary tool that is used will be singularity analysis (see [**14, 15**]), which provides a direct connection between the singularities of a generating function and the asymptotic behaviour of its coefficients. By Pringsheim's theorem [**15**, p. 240] we know that a generating function with non-negative coefficients must have a singularity at $z = R$, if R denotes the radius of convergence. Our assumption that ϕ is non-periodic guarantees furthermore that this is the only singularity on the circle of convergence. Throughout this paper we will call $z = R$ the *dominant singularity* of the generating function. In particular, we denote the dominant singularity of $T(z)$ by ρ. Furthermore, we say that a function f has an algebraic singularity of type α at s, if there is a constant C such that $f(z) \sim f(s) + C \cdot (1 - \frac{z}{s})^{\alpha}$ as z tends to s in such a way that $z - s \notin \mathbb{R}^+$. If $f(z) \sim \sum_{k\geq 0} f_i \left(1 - \frac{z}{\rho}\right)^{\alpha k}$ then we say f admits a Puiseux expansion. For instance, it is well known that the generating function $T(z)$ associated to some class of simply generated trees has an algebraic singularity ρ of type $1/2$ (for obvious reasons also called square root singularity) the location of which is determined by the system $T(\rho) = \rho\phi(T(\rho))$, $1 = \rho\phi'(T(\rho))$ and that it admits a Puiseux expansion there, *cf.* [**9**]. For further information on this theory we refer the reader to [**15**] and [**14**].

Let $T_k(z)$ denote the generating function of the class of simply generated trees that have protection number at least k, where z marks the total number of nodes. Then, $T_k(z)$ can be defined by

$$T_k(z) = z\left(\phi(T_{k-1}(z)) - \phi_0\right). \tag{1}$$

Note that $T_0(z) = T(z)$.

Lemma 1. *All generating functions $T_k(z)$ have the same dominant singularity as $T(z)$, and it is a square root singularity.*

Proof. First let us consider that the generating function $T_k(z)$ reads as

$$T_k(z) = \Omega^k(T(z))$$

where $\Omega(t) = z\phi(t) - z\phi_0$ and $\Omega^k(\cdot)$ denotes the k-fold composition. Since $\Omega(t)$ is analytic at $T(\rho)$, inserting a function admitting a Puiseux expansion $t(z) = \alpha_0 + \alpha_1\sqrt{1 - \frac{z}{\rho}} + \dots$ results in

$$\Omega(t(z)) = \Omega(\alpha_0) + \Omega'(\alpha_0)\alpha_1\sqrt{1 - \frac{z}{\rho}} + \dots,$$

again being a Puiseux expansion at $z = \rho$. It is well known that $T(z)$ admits a Puiseux expansion $\tau_0 + \tau_1\sqrt{1 - \frac{z}{\rho}} + \ldots$ with nonzero numbers τ_0 and τ_1. Moreover, we always insert one of the functions $T_k(z)$, thus α_0 attains the positive values $T_k(\rho)$, $k = 0, 1, 2, \ldots$, implying that $\Omega'(\alpha_0)$ is always positive, as $\Omega(t)$ is a power series with only non-negative coefficients. By induction it is guaranteed that α_1 is always negative and thus all the function $T_k(z)$ have a unique dominant singularity of square root type at $z = \rho$. □

In order to derive the expected value of the protection number X_n of a random simply generated tree of size n (*i.e.* with n nodes) asymptotically, we use the well known formula

$$\mathbb{E}X_n = \sum_{k \geq 1} \mathbb{P}(X_n \geq k). \tag{2}$$

Thus, we need to calculate the probability $\mathbb{P}(X_n \geq k)$, which is given by

$$\mathbb{P}(X_n \geq k) = \frac{[z^n]T_k(z)}{[z^n]T(z)}.$$

However, first we show an auxiliary result that will be needed in the following.

Lemma 2. *Let $T_k(z)$ denote the generating function of a class of simply generated trees with protection number at least k defined as in (1) and let ρ denote its dominant singularity. Then the sum $\sum_{k\geq 1}\prod_{i=1}^{k-1}(\rho\phi'(T_i(\rho)))$ converges.*

Proof. It is easy to see that the sequence $(T_i(\rho))_{i\geq 0}$ is monotonically decreasing, since the number of trees with protection number at least i is always greater than the number of trees that have an $(i+1)$-protected root, *i.e.* protection number at least $i + 1$. Since ϕ' is monotonically increasing on the positive real axis, this implies that $\rho\phi'(T_i(\rho)) \leq \rho\phi'(T_1(\rho)) < \rho\phi'(T(\rho)) = 1$. Thus, we can estimate the sum

$$\sum_{k\geq 1}\prod_{i=1}^{k-1}(\rho\phi'(T_i(\rho))) < \sum_{k\geq 1}(\rho\phi'(T_1(\rho)))^{k-1}, \tag{3}$$

which converges, since $\rho\phi'(T_1(\rho)) < 1$. □

Theorem 1. *Let X_n be the protection number of a random simply generated tree of size n. Then the expected value $\mathbb{E}X_n$ and the variance $\mathbb{V}X_n$ satisfy*

$$\lim_{n\to\infty}\mathbb{E}X_n = \sum_{k\geq 1}\rho^{k-1}\prod_{i=1}^{k-1}\phi'(T_i(\rho)),$$

and

$$\lim_{n\to\infty}\mathbb{V}X_n = \sum_{k\geq 1}(2k-1)\rho^{k-1}\prod_{i=1}^{k-1}\phi'(T_i(\rho)) - \Big(\lim_{n\to\infty}\mathbb{E}X_n\Big)^2$$

with ρ denoting the dominant singularity of the generating function $T(z) = z\phi(T(z))$ of the class of simply generated trees.

Proof. We know that the asymptotic behaviour of the generating function, namely $T(z) = \tau_0 + \tau_1\sqrt{1-\frac{z}{\rho}} + \tau_2(1-\frac{z}{\rho}) + \ldots$, implies

$$[z^n]T(z) \sim -\tau_1 \frac{n^{-3/2}}{\Gamma(-1/2)}\rho^{-n}, \tag{4}$$

as n tends to infinity. In order to derive the asymptotics of the n-th coefficient of $T_k(z)$, observe that we know from Lemma 1 that all generating functions $T_i(z)$ have the same dominant singularity ρ of type $\frac{1}{2}$. Setting $\eta = \sqrt{1-\frac{z}{\rho}}$, the Puiseux expansions of $T_k(z)$ and $T_{k-1}(z)$ read as

$$T_k(z) = \tau_{0,k} + \tau_{1,k}\eta + \tau_{2,k}\eta^2 + \ldots.$$

and

$$T_{k-1}(z) = \tau_{0,k-1} + \tau_{1,k-1}\eta + \tau_{2,k-1}\eta^2 + \ldots.$$

Plugging these expansions into (1) and using $z = \rho(1-\eta^2)$ we get

$$\tau_{0,k} + \tau_{1,k}\eta + \tau_{2,k}\eta^2 + \ldots = \rho(1-\eta^2)\left(\sum_{j\geq 0}\phi_j\left(\tau_{0,k-1} + \tau_{1,k-1}\eta + \tau_{2,k-1}\eta^2 + \ldots\right)^j - \phi_0\right).$$

Expanding and comparing coefficients of η^0 and η^1 yields

$$\begin{aligned}
[\eta^0] &: \tau_{0,k} = \rho\phi(\tau_{0.k-1}) - \rho\phi_0,\\
[\eta^1] &: \tau_{1,k} = \rho\sum_{j\geq 0}\phi_j j\tau_{1,k-1}\tau_{0,k-1}^{j-1}.
\end{aligned}$$

Obviously, the $\tau_{0,i}$'s match exactly the $T_i(\rho)$, $i \geq 0$, as they are the constant terms in the Puiseux expansions of the functions $T_i(z)$, with $0 \leq i \leq k$. Thus, the equation for $\tau_{1,k}$ can be rewritten as $\tau_{1,k} = \rho\tau_{1,k-1}\phi'(T_{k-1}(\rho))$.

As $\tau_{1,0} = \tau_1$, we get

$$\tau_{1,k} = \tau_1\rho^{k-1}\prod_{i=1}^{k-1}\phi'(T_i(\rho)).$$

Applying a transfer lemma [**14**] directly gives the asymptotics of the coefficients of $T_k(z)$ and plugging them in conjunction with (4) into Equation (2) yields the asymptotic value for the mean. In order to derive the formula for the asymptotic variance we use the equation

$$\mathbb{V}X_n = \mathbb{E}(X_n^2) - (\mathbb{E}X_n)^2 \quad \text{and} \quad \mathbb{E}(X_n^2) = \sum_{k\geq 1}(2k-1)\mathbb{P}(Y_n \geq k)$$

and immediately get the asserted result. The convergence of the obtained sums follows from Lemma 2. □

As the sums in Theorem 1 are majorized by convergent geometric series (*i.e.*, the inequality (3) even holds term-wise), we can calculate efficiently the asymptotic mean and variance for all classes of simply generated trees with arbitrary accuracy. We will now exemplify this by calculating the limits of mean and variance of the protection number of some prominent classes of simply generated trees.

Example (Plane trees)**.** The generating function $C(z)$ of plane trees is the unique power series solution of

$$C(z) = z\frac{1}{1-C(z)},$$

which yields

$$C(z) = \frac{1}{2} - \sqrt{\frac{1}{4} - z}. \tag{5}$$

Thus, its dominant singularity is $\rho = \frac{1}{4}$, and $C(\rho) = \frac{1}{2}$.

The recursion for the $T_i(\rho)$'s reads as

$$T_1(\rho) = \frac{1}{4}, \qquad T_i(\rho) = \frac{1}{4 - 4T_{i-1}(\rho)} - \frac{1}{4}.$$

In the case of plane trees the recursion can be solved explicitly, leading to

$$T_i(\rho) = \frac{3}{2(4^i+2)}.$$

The limits of expected value and variance are therefore given by

$$\lim_{n\to\infty} \mathbb{E}X_n = \sum_{k\geq 1} \frac{1}{4^{k-1}} \prod_{i=1}^{k-1} \frac{1}{\left(1 - \frac{3}{2(4^i+2)}\right)^2} \approx 1.622971384715353,$$

and

$$\lim_{n\to\infty} \mathbb{V}X_n = \sum_{k\geq 1} (2k-1)\frac{1}{4^{k-1}} \prod_{i=1}^{k-1} \frac{1}{\left(1 - \frac{3}{2(4^i+2)}\right)^2} - \left(\lim_{n\to\infty} \mathbb{E}X_n\right)^2 \approx 0.7156950717833327,$$

which has already been calculated by Heuberger and Prodinger in [**17**].

Example (Motzkin trees)**.** The generating function $M(z)$ of Motzkin trees is defined by

$$M(z) = z\left(1 + M(z) + M(z)^2\right),$$

which can be solved to result in

$$M(z) = \frac{1 - z - \sqrt{1-2z-3z^2}}{2z}.$$

Thus, its dominant singularity is $\rho = \frac{1}{3}$ and $M(\rho) = 1$.

The recursion for the $T_i(\rho)$'s reads as

$$T_1(\rho) = \frac{2}{3}, \qquad T_i(\rho) = \frac{1}{3}\left(T_{i-1}(\rho)^2 + T_{i-1}(\rho)\right)$$

We are not aware of a method to solve this recursion explicitly, but from the equivalent equation

$$3^i T_i(\rho) = 2\prod_{j=1}^{i-1}(1 + T_j(\rho)),$$

we directly get $T_i(\rho) \sim C \cdot 3^{-i}$ for a constant $C > 0$. Due to this exponential decrease estimates are easily obtained and we can calculate the limits of mean and variance for the protection number numerically with arbitrary accuracy:

$$\lim_{n\to\infty} \mathbb{E}X_n \approx 2.546378248338912, \qquad \lim_{n\to\infty} \mathbb{V}X_n \approx 1.679348871220563.$$

Example (Incomplete binary trees)**.** The generating function $I(z)$ of incomplete binary trees is defined by

$$I(z) = z\left(1 + 2I(z) + I(z)^2\right),$$

which gives

$$I(z) = \frac{1 - 2z - \sqrt{1-4z}}{2z}.$$

The dominant singularity is therefore at $\rho = \frac{1}{4}$ and $I(\rho) = 1$.

The recursion for the $T_i(\rho)$'s reads as

$$T_1(\rho) = \frac{3}{4}, \qquad T_i(\rho) = \frac{1}{4}(T_{i-1}(\rho)^2 + 2T_{i-1}(\rho)).$$

This recursion cannot be solved explicitly, but the numerical values can be easily computed: They are

$$\lim_{n\to\infty} \mathbb{E}X_n \approx 3.536472483525321, \qquad \lim_{n\to\infty} \mathbb{V}X_n \approx 3.763883442795153.$$

Example (Cayley trees)**.** Though, in a strict sense, Cayley trees do not belong to the class of simply generated trees (*cf.* the discussions in [**19**] and [**16**]), they are usually listed as an example for that class. In fact, they are closely related (see [**25**] for a thorough analysis and [**16**] for an analysis of the differences) and in many contexts (like the one considered here), quotients of coefficients are computed which makes the fact that in this case the generating functions are exponential ones irrelevant.

The (exponential) generating function $C(z)$ of Cayley trees is defined by

$$C(z) = ze^{C(z)},$$

which has its dominant singularity at $\rho = \frac{1}{e}$. Moreover, we have $C(\rho) = 1$.

The recursion for the $T_i(\rho)$'s reads as

$$T_1(\rho) = 1 - \frac{1}{e}, \qquad T_i(\rho) = \frac{1}{e}(e^{T_{i-1}(\rho)} - 1).$$

As in the two previous examples the recursion for the $T_i(\rho)$'s cannot be solved explicitly, but the numerical values are

$$\lim_{n\to\infty} \mathbb{E}X_n \approx 2.286198316708012, \qquad \lim_{n\to\infty} \mathbb{V}X_n \approx 1.598472890455086.$$

Example (Binary trees)**.** This is the class of complete binary trees with only internal vertices contributing to the size. The generating function is then defined by the functional equation $B(z) = 1 + zB(z)^2$ with $B(z) = C(z)/z$ where $C(z)$ is the function displayed in (5). Though this class does not strictly fall into the simply generated framework, the functional equation is of the form $B(z)-1 = z\phi(B(z)-1)$, which reflects the fact that incomplete binary trees with all nodes counted are in bijection to complete binary trees with only internal vertices counted. For the protection number this causes some shifts within the tree. But the methodology presented above works here as well. We get $T_0(z) = B(z)$ and $T_k(z) = zT_{k-1}(z)^2$. Since $\rho = 1/4$ we have $T_k(\rho) = 2^{2-2^k}$, for all $k \geq 0$, and then finally $\mathbb{P}(X_n \geq k) \to 2^{k+1-2^k}$, as n tends to infinity. Thus we obtain

$$\lim_{n\to\infty} \mathbb{E}X_n \approx 1.562988296151161, \qquad \lim_{n\to\infty} \mathbb{V}X_n \approx 0.372985688954940.$$

2.2 Protection number of a random vertex

In the first part of this section we studied the average protection number of a simply generated tree, that is the protection number of the root of the simply generated tree. Now we are interested in the average protection number of a randomly chosen vertex in a simply generated tree of size n. We denote this sequence of random variables by Y_n.

As in the previous section we calculate the mean via $\mathbb{E}Y_n = \sum_{k\geq 1} \mathbb{P}(Y_n \geq k)$. In order to do so we proceed analogously to Heuberger and Prodinger in [**17**] and define $S_k(z)$ to be the generating function of the sequence $(s_{n,k})_{n\geq 0}$ of k-protected vertices summed over all trees of size n. As in [**17**] this generating function can be calculated by

$$S_k(z) = z^{-1} T_k(z) \frac{\partial}{\partial u} T(z,1), \tag{6}$$

by means of the bivariate generating function $T(z,u)$ of simply generated trees, where z marks the size and u the number of leaves, and the generating function $T_k(z)$ of simply generated trees with protection number at least k. The formula for $S_k(z)$ arises from considering a k-protected vertex in the following way: First point at a leaf in a simply generated tree (which yields the factor $\frac{\partial}{\partial u}T(z,1)$), then remove this leaf (which explains the z^{-1}) and finally attach a tree with protection number at least k (giving the factor $T_k(z)$).

Remark. The procedure works also for complete binary trees, where only internal vertices contribute to the tree size. The only difference is that for complete binary trees the factor z^{-1} in (6) must be removed, because removing a leaf does not change the size.

Using the generating function $S_k(z)$ we can express the probability $\mathbb{P}(Y_n \geq k)$ by

$$\mathbb{P}(Y_n \geq k) = \frac{[z^n]S_k(z)}{n[z^n]T(z)}. \tag{7}$$

Before proving the main theorem of this subsection, we show the following lemma concerning the decay of $T_k(\rho)$.

Lemma 3. *Let $T_k(z)$ denote the generating function of a class of simply generated trees with protection number at least k defined as in* (1) *and let ρ denote its dominant singularity. Then the sequence $(T_k(\rho))_{k\geq 0}$ tends to zero exponentially fast.*

Proof. From (1) we have

$$T_k(\rho) = \rho\phi(T_{k-1}(\rho)) - \rho\phi_0 = \rho \sum_{i\geq 1} \phi_i T_{k-1}(\rho)^i = \rho T_{k-1}(\rho) \sum_{i\geq 1} \phi_i T_{k-1}(\rho)^{i-1}.$$

Since we know that $T_k(\rho)$ is monotonically decreasing in k and $T_1(\rho) < T_0(\rho) = T(\rho)$, we directly obtain

$$\frac{T_k(\rho)}{T_{k-1}(\rho)} = \rho \sum_{i\geq 1} \phi_i T_{k-1}(\rho)^{i-1} \leq \rho \sum_{i\geq 1} \phi_i T_1(\rho)^{i-1} < \rho \sum_{i\geq 1} \phi_i T(\rho)^{i-1} = \rho\phi'(T(\rho)) = 1.$$

Hence we found a constant $q := \rho \sum_{i\geq 1} \phi_i T_1(\rho)^{i-1}$ such that

$$\frac{T_k(\rho)}{T_{k-1}(\rho)} \leq q < 1,$$

which proves $T_k(\rho) \leq T_1(\rho) q^{k-1}$ and thus the exponential decay. ∎

Theorem 2. *Let Y_n be the protection number of a randomly chosen vertex in a random simply generated tree of size n. Then,*

$$\lim_{n\to\infty} \mathbb{E}Y_n = \frac{\phi_0}{T(\rho)} \sum_{k\geq 1} T_k(\rho),$$

and

$$\lim_{n\to\infty} \mathbb{V}Y_n = \frac{\phi_0}{T(\rho)} \sum_{k\geq 1} (2k-1)T_k(\rho) - \Big(\lim_{n\to\infty} \mathbb{E}Y_n\Big)^2.$$

Proof. First we need to determine the n-th coefficient of $S_k(z)$. We have

$$\frac{\partial}{\partial u} T(z,1) = \frac{z\phi_0}{1 - z\phi'(T(z))}. \tag{8}$$

Using $T'(z) = z\phi'(T(z))T'(z) + \phi(T(z))$ and $\phi(T(z)) = \frac{T(z)}{z}$ we get

$$z\phi'(T(z)) = \frac{T'(z) - \frac{T(z)}{z}}{T'(z)}.$$

Therefore (8) transforms to

$$\frac{\partial}{\partial u} T(z,1) = \frac{T'(z) z^2 \phi_0}{T(z)}.$$

Thus, altogether we have

$$[z^n]S_k(z) = [z^n] z^{-1} T_k(z) \frac{T'(z) z^2 \phi_0}{T(z)},$$

which gives

$$[z^n]S_k(z) \sim \frac{-\tau_{0,k}\tau_1\phi_0}{2\tau_0} \frac{n^{-1/2}}{\Gamma(1/2)} \rho^{-n}.$$

Finally, we get

$$\mathbb{E}Y_n = \sum_{k\geq 1} \mathbb{P}(Y_n \geq k) = \sum_{k\geq 1} \frac{[z^n]S_k(z)}{n[z^n]T(z)} \overset{n\to\infty}{\to} \sum_{k\geq 1} \frac{T_k(\rho)\phi_0}{T(\rho)}.$$

For the variance we use again the formula $\mathbb{V}Y_n = \sum_{k\geq 1}(2k-1)\mathbb{P}(Y_n \geq k) - \mathbb{E}(Y_n)^2$ and (7). The convergence of the obtained sums follows from Lemma 3. $\square$

	$\lim_{n\to\infty} \mathbb{E}Y_n$	$\lim_{n\to\infty} \mathbb{V}Y_n$
Plane trees	0.7276492769137261	0.8168993794836289
Motzkin trees	1.307604625963334	1.730614214799486
Incomplete binary trees	1.991819588602741	3.638259051495130
Cayley trees	1.186522661652180	1.632206223956926
Complete binary trees	1.265686036087572	0.226591112528581

Table 1: The approximate values for the limits of mean and variance of the protection number of a random vertex in different classes of simply generated trees. In the case of complete binary trees only internal vertices are considered.

3. PÓLYA TREES

3.1 Protection number of the root

Let $T(z)$ be the generating function of Pólya trees, which reads as

$$T(z) = ze^{T(z)} \exp\left(\sum_{i\geq 2} \frac{T(z^i)}{i}\right),$$

and in correspondence to the previous section let us denote by $T_k(z)$ the generating function of the class of Pólya trees that have protection number at least k. This generating function can be specified by

$$T_k(z) = ze^{T_{k-1}(z)} \exp\left(\sum_{i\geq 2} \frac{T_{k-1}(z^i)}{i}\right) - z, \tag{9}$$

with $T_0(z) = T(z)$. From the classical results of Pólya [**26**] we know that $T(z)$ has a unique dominant singularity ρ of type $1/2$ and admits a Puiseux series expansion there, which starts as

$$T(z) \sim 1 - b\sqrt{1-\frac{z}{\rho}} + \frac{b^2}{3}\left(1-\frac{z}{\rho}\right) + d\left(1-\frac{z}{\rho}\right)^{3/2} + \cdots. \tag{10}$$

Numerical approximations for the constants have been first computed by Otter [**24**]. This was also topic in the book of Finch [**13**, Section 5.6] and the book of Flajolet and Sedgewick [**15**, p. 477] where we find approximations up to 25 digits: $\rho \approx 0.3383218568992076951961126$ and $b \approx 1.559490020374640885542260$.

Lemma 4. *All the generating functions $T_k(z)$ have their (unique) dominant singularity at ρ, and the singularity is a square root singularity.*

Proof. First let us recall that $T_0(z) = T(z)$. Thus, for $k = 0$ the lemma is trivial. For $k \geq 1$ we proceed by induction. Therefore let us assume that $T_{k-1}(z)$ has the dominant singularity ρ which is of type $\frac{1}{2}$. Then the dominant singularity of $T_k(z)$, satifying the recurrence relation (9), comes from $e^{T_{k-1}(z)}$, since $\exp\left(\sum_{i\geq 2} \frac{T_{k-1}(z^i)}{i}\right)$ is analytic in $|z| < \rho + \epsilon$ with $\epsilon > 0$ sufficiently small. Applying the exponential function to a function having an algebraic singularity does neither change the location nor the type of the singularity, which proves the assertion after all. $\square$

The goal of this section is to derive an asymptotic value for the average protection number of Pólya trees. We use again the formula $\mathbb{E}X_n = \sum_{k\geq 1}\mathbb{P}(X_n \geq k)$, but rewrite this equation as

$$\mathbb{E}X_n = \sum_{k\geq 1}\prod_{i=1}^{k}\mathbb{P}(X_n \geq i | X_n \geq i-1),$$

where the conditional probabilities can be obtained by

$$\mathbb{P}(X_n \geq k | X_n \geq k-1) = \frac{[z^n]T_k(z)}{[z^n]T_{k-1}(z)}. \tag{11}$$

Let us again start by showing the exponential decay of the sequence $(T_k(\rho))_{k\geq 1}$.

Lemma 5. *Let $T_k(z)$ denote the generating function of the class of Pólya trees with protection number at least k defined as in* (9) *and let ρ denote its dominant singularity. Then the sequence $(T_k(\rho))_{k\geq 0}$ tends to zero exponentially fast.*

Proof. Let us denote $C_k := \exp\left(\sum_{i\geq 2} \frac{T_k(\rho^i)}{i}\right)$. Then Equation (9) directly yields

$$\begin{aligned} T_k(\rho) &= \rho e^{T_{k-1}(\rho)} C_{k-1} - \rho < \rho e^{T_{k-1}(\rho)} C_{k-1} \\ &= \rho C_{k-1} \sum_{i\geq 0} \frac{T_{k-1}(\rho)^i}{i!} = \rho C_{k-1} T_{k-1} \sum_{i\geq 1} \frac{T_{k-1}(\rho)^{i-1}}{i!}, \end{aligned}$$

which implies

$$\frac{T_k(\rho)}{T_{k-1}(\rho)} < \rho C_{k-1} \sum_{i\geq 1} \frac{T_{k-1}(\rho)^{i-1}}{i!} \leq \rho C_{k-1} \frac{1}{T_1(\rho)} \sum_{i\geq 1} \frac{T_1(\rho)^i}{i!} = \rho C_{k-1} \frac{e^{1-\rho}-1}{1-\rho},$$

since $T_k(\rho) \leq T_1(\rho)$ for $k \geq 1$ and $T_1(\rho) = 1 - \rho$. Estimating C_k yields

$$C_k \leq C_1 < \exp\left(\sum_{i\geq 2} \frac{T(\rho^i)}{i}\right) = \frac{1}{e\rho}.$$

Putting all together gives

$$\frac{T_k(\rho)}{T_{k-1}(\rho)} < \frac{e^{1-\rho}-1}{e(1-\rho)} \approx 0.52153,$$

which finally yields the exponential decay. $\square$

Lemma 6. *The asymptotic expansions of the n-th coefficients of $T_k(z)$ and $T_{k-1}(z)$ read as*

$$[z^n]T_{k-1}(z) = \frac{\gamma_k \rho^{-n} n^{-\frac{3}{2}}}{\Gamma(-1/2)} \left(1 + \mathcal{O}\left(\frac{1}{n}\right)\right),$$
$$[z^n]T_k(z) = \frac{(T_k(\rho)+\rho)\gamma_k \rho^{-n} n^{-\frac{3}{2}}}{\Gamma(-1/2)} \left(1 + \mathcal{O}\left(\frac{1}{n}\right)\right),$$

as $n \to \infty$, with a constant $\gamma_k > 0$.

Proof. Let the Puiseux expansion of $T_{k-1}(z)$ be given by

$$T_{k-1}(z) = T_{k-1}(\rho) - \gamma_k \sqrt{1 - \frac{z}{\rho}} + \dots.$$

Then $T_k(z)$ behaves asymptotically as $T_k(z) \sim \rho e^{T_{k-1}(\rho)} C_{k-1} e^{-\gamma_k \sqrt{1-\frac{z}{\rho}}}$, where $C_k = \exp\left(\sum_{i\geq 2} \frac{T_k(\rho^i)}{i}\right)$. Applying the asymptotic relation $e^{-\gamma_k \sqrt{1-\frac{z}{\rho}}} \sim 1 - \gamma_k \sqrt{1-\frac{z}{\rho}}$ and using the equation $\rho e^{T_{k-1}(\rho)} C_{k-1} = T_k(\rho) + \rho$ completes the proof. □

Plugging the expansions obtained in Lemma 6 into Equation (11) gives

$$\mathbb{P}(X_n \geq k | X_n \geq k-1) = T_k(\rho) + \rho,$$

which directly yields the following theorem.

Theorem 3. *Let X_n be the protection number of a random Pólya tree of size n. Then*

$$\lim_{n\to\infty} \mathbb{E}X_n = \sum_{k\geq 1} \prod_{i=1}^{k} (T_i(\rho) + \rho) \approx 2.154889671973873, \tag{12}$$

and $\lim_{n\to\infty} \mathbb{V}X_n \approx 1.369993017502652$.

Proof. The proof for the asymptotic mean follows directly by Lemma 6. In order to determine the variance we use the representation

$$\lim_{n\to\infty} \mathbb{V}X_n = \sum_{k\geq 1} (2k-1) \prod_{i=1}^{k} (T_i(\rho) + \rho) - \mathbb{E}(X_n)^2.$$

The convergence of the sums for the expected mean and variance follows by Lemma 5. □

Remark. Note that in order to get accurate numerical values, we must not compute $T_k(\rho)$ by insertion into a (truncated) series expansion for $T_k(z)$. The reason is that ρ lies on the circle of convergence and thus the convergence is very slow at $z = \rho$. Instead, $T_k(\rho)$ can be directly computed using the recurrence relation (9). The values $T_k(\rho^i)$ for $i \geq 2$, which appear in that recurrence relation, can be computed with the help of the series expansion of $T_k(z)$, because ρ^i then lies in the interior of region of convergence where the series converges at an exponential rate.

Remark. We could also have used the same approach as for simply generated trees in order to get the asymptotic mean. Then the resulting formula looks like

$$\lim_{n\to\infty} \mathbb{E}X_n = \sum_{k\geq 1} \rho^{k-1} \prod_{i=1}^{k-1} C_i e^{T_i(\rho)}, \tag{13}$$

where $C_j = \exp\left(\sum_{i\geq 2} \frac{T_j(\rho^i)}{i}\right)$. Since $T_i(\rho)$ tends to 0 exponentially fast and hence C_i tends to 1, we immediately get the constant given in Theorem 3. However, since this approach requires more technical calculations, we decided to switch to the more direct strategy using the conditional probabilities. Moreover note that the equivalence of (12) and (13) is immediate from (9).

3.2 Protection number of a random vertex

The method of marking a leaf and replacing it by a tree with protection number k does not work here. Due to possible symmetries in non-plane trees, this would result in wrong counting: Indeed, if there are k-protected vertices $x_1, \ldots, x_\ell$ which can be mapped to each other by some automorphisms of the tree (*i.e.*, they lie in the same vertex class), then only one of them is counted. Though this is counterbalanced by trees having ℓ leaves in the same vertex class one of which is replaced by a tree with protection number k (the root of this tree is then counted ℓ times), there are further overcounts: As all leaves are marked, trees having several leaves in the same vertex class are counted several times, and so are their k-protected vertices.

Thus we appeal to the proof of [**27**, Theorem 3.1] here: For a tree T let

$$f(T) = \begin{cases} 1 & \text{if } T \text{ has protection number at least } k, \\ 0 & \text{otherwise.} \end{cases}$$

Moreover, we define $F(T)$ to be the number of k-protected nodes in T. Then the generating function $R_k(z,u) = \sum_T z^{|T|} u^{F(T)}$ satisfies (*cf.* [**27**, Equ. (3.1)])

$$z \exp\left(\sum_{i\geq 1} \frac{R_k(z^i, u^i)}{i}\right) = \sum_{n\geq 1} z^n \sum_{T:|T|=n} u^{F(T)-f(T)} \tag{14}$$

As in Section 2.2 we utilize the formula $\mathbb{E}Y_n = \sum_{k\geq 1} \mathbb{P}(Y_n \geq k)$ and express the occurring probabilities as $\mathbb{P}(Y_n \geq k) = [z^n]S_k(z)/(n[z^n]T(z))$ with $S_k(z)$ being the generating function whose nth coefficient is the cumulative number of k-protected nodes in all trees of size n. Obviously, $((\partial/\partial u)R_k)(z,1) = S_k(z)$ and thus by differentiating (14) with respect to u and inserting $u = 1$ we obtain

$$T(z) \sum_{i\geq 1} S_k(z^i) = S_k(z) - T_k(z). \tag{15}$$

This implies

$$S_k(z) = \frac{T(z)\sum_{i\geq 2} S_k(z^i) + T_k(z)}{1-T(z)} \sim \frac{\sum_{i\geq 2} S_k(\rho^i) + T_k(\rho)}{b\sqrt{1-\frac{z}{\rho}}} \tag{16}$$

where b is the constant appearing in (10). Standard transfer theorems applied to (10) give

$$[z^n]T(z) \sim \frac{-bn^{-3/2}\rho^{-n}}{\Gamma(-1/2)} = \frac{bn^{-3/2}\rho^{-n}}{2\sqrt{\pi}},$$

and from (16) we get

$$[z^n]S_k(z) \sim \frac{\left(\sum_{i\geq 2} S_k(\rho^i) + T_k(\rho)\right) n^{-1/2}\rho^{-n}}{b\sqrt{\pi}}$$

and thus

$$\mathbb{P}(Y_n \geq k) \sim \frac{2}{b^2}\left(\sum_{i\geq 2} S_k(\rho^i) + T_k(\rho)\right). \tag{17}$$

Now we prove that $\sum_{i\geq 2} S_k(\rho^i)$ decreases exponentially.

Lemma 7. *Let $S_k(z)$ denote the generating function of the cumulative number of k-protected vertices in all Pólya trees of size n, and let ρ denote the dominant singularity of the generating function of Pólya trees. Then $\sum_{i\geq 2} S_k(\rho^i)$ decreases exponentially fast with growing k.*

Proof. First, let $C_k(z)$ be the generating function of plane trees with protection number at least k and $C(z,u)$ denote the bivariate generating function of plane trees, where z marks the size and u the number of leaves.

Now, observe that the cumulative number of k-protected vertices in all Pólya trees of size n is smaller than the respective number in all planted plane trees of size n, since in the latter case different plane embeddings of the same Pólya tree are counted separately (and so are their respective protected nodes). This in conjunction with (6) implies that for $0 < z \leq \rho^2$ (note that $\rho^2 < 1/4$) we have

$$S_k(z) < z^{-1}C_k(z)\frac{\partial}{\partial u}C(z,1) < 4C_k\left(\frac{1}{4}\right)\frac{\partial}{\partial u}C\left(\frac{1}{4},1\right). \tag{18}$$

From Lemma 3 we know that $C_k\left(\frac{1}{4}\right)$ tends to zero exponentially fast and, in view of (18), so does $S_k(z)$. $\square$

By means of Lemma 5 and Lemma 7 we know that the series for $\mathbb{E}Y_n$, namely

$$\mathbb{E}Y_n = \sum_{k\geq 1}\mathbb{P}(Y_n \geq k),$$

converges rapidly. But (17) still bears a secret, because we do not have an explicit expression for $S_k(z)$ and we cannot solve the functional equation (15).

For numerical purposes, however, it is not necessary to have an explicit expression for $S_k(z)$. If we write $S_k(z) = \Psi(S_k(z))$ with Ψ being the operator on the ring of formal power series defined by

$$\Psi(f(z)) = \frac{T(z)\sum_{i\geq 2} f(z^i) + T_k(z)}{1 - T(z)},$$

then Ψ is a contraction on the metric space $\mathbb{R}[[z]]$ equipped with the formal topology (*cf.* [**15**, Appendix A.5]). Indeed, if $f(z)$ and $g(z)$ coincide up to their ℓth coefficient, then the first $2\ell + 2$ coefficients of $\Psi(f(z))$ and$\Psi(g(z))$ coincide.

As there is exactly one tree with $k+1$ vertices which possesses k-protected vertices at all (namely the path of length k has a k-protected root) whereas all smaller trees do not possess any k-protected vertices, we know that the (one-term) series z^{k+1} coincides with $S_k(z) = z^{k+1} + \cdots$ in its first $k+2$ coefficients. Applying Ψ to z^{k+1} a few times, with each application more than doubling the number of known coefficients of $S_k(z)$, gives quickly a fairly accurate expression for $S_k(z)$. We obtain the following theorem:

Theorem 4. *Let Y_n be the protection number of a random vertex in a random Pólya tree of size n. Then*

$$\lim_{n\to\infty} \mathbb{E}Y_n = \sum_{k\geq 1} \frac{2}{b^2}\left(\sum_{i\geq 2} S_k(\rho^i) + T_k(\rho)\right) \approx 0.9953254987$$

and $\lim_{n\to\infty} \mathbb{V}Y_n \approx 1.3818769746$.

4. NON-PLANE BINARY TREES

4.1 Protection number of the root

We denote by $T(z)$ the generating function of non-plane binary trees, where z marks the number of internal nodes. Then $T(z)$ satisfies

(19) $$T(z) = 1 + z\left(\frac{1}{2}T(z)^2 + \frac{1}{2}T(z^2)\right).$$

The generating function $T_k(z)$ of non-plane binary trees with protection number at least k satisfies

(20) $$T_k(z) = z\left(\frac{1}{2}T_{k-1}(z)^2 + \frac{1}{2}T_{k-1}(z^2)\right),$$

and $T_0(z) = T(z)$. As in the previous sections, the decay rate of $T_k(z)$ guarantees the convergence of all the series appearing in the subsequent results. For combinatorial reasons the function $T(z)$ has a singularity at some $0 < \rho < 1$ and from the functional equation (19) we can deduce that $T(\rho) = 1/\rho < 1$. Inserting this into (20) and using the estimate $T_{k-1}(z^2) < T_k(z)$ gives $T_1(z) < 1$, for $0 \leq z \leq \rho$. Using the estimate again yields $T_k(z) < \rho T_{k-1}(z^2)$ and therefore $T_k(z)$ decays even super-exponentially, as k tends to infinity.

In order to obtain the asymptotic mean and variance for the protection number of a random non-plane binary tree of size n we proceed analogously as in the previous section for Pólya trees. Thus, we use

$$\mathbb{E}X_n = \sum_{k\geq 1}\prod_{i=1}^{k} \mathbb{P}(X_n \geq i | X_n \geq i-1) = \sum_{k\geq 1}\prod_{i=1}^{k} \frac{[z^n]T_i(z)}{[z^n]T_{i-1}(z)}.$$

Theorem 5. *Let X_n be the protection number of a random non-plane binary tree of size n. Then*

$$\lim_{n\to\infty} \mathbb{E}X_n = \sum_{k\geq 1}\prod_{i=1}^{k-1} (\rho T_i(\rho)) \approx 1.707603060723366$$

and $\lim_{n\to\infty} \mathbb{V}X_n \approx 0.431102549825064$.

Proof. Let the Puiseux expansion of $T_k(z)$ and $T_{k+1}(z)$ read as

$$T_{k-1}(z) = T_{k-1}(\rho) - \gamma_k \sqrt{1-\frac{z}{\rho}} + \mathcal{O}\left(1-\frac{z}{\rho}\right),$$

and

$$T_k(z) = \rho\left(\frac{1}{2}T_{k-1}(\rho)^2 + \frac{1}{2}T_{k-1}(\rho^2)\right) + \rho T_{k-1}(\rho)\gamma_k\sqrt{1-\frac{z}{\rho}} + \mathcal{O}\left(1-\frac{z}{\rho}\right).$$

Using singularity analysis yields the desired result for the mean. For the variance we use again the formula $\mathbb{V}X_n = \sum_{k\geq 1}(2k-1)\mathbb{P}(X_n \geq k) - \mathbb{E}(X_n)^2$. $\square$

4.2 Protection number of a random internal vertex

The asymptotic mean and variance for the protection number of a randomly chosen internal vertex in a random non-plane binary tree can be obtained in the same way as in the previous section for Pólya trees.

Thus, we again set up an equation for the generating function $R_k(z,u)$ where the coefficients $[z^n u^l]R_k(z,u)$ count the number of non-plane binary trees of size n with l k-protected vertices:

$$\frac{z}{2}\left(R_k(z,u)^2 + R_k(z^2,u^2)\right) = \sum_{n\geq 1} z^n \sum_{T:|T|=n} u^{F(T)-f(T)}$$

Differentiating this equation with respect to u and setting $u = 1$ yields

$$zT(z)S_k(z) + zS_k(z^2) = S_k(z) - T_k(z).$$

Therefore we get

$$S_k(z) = \frac{zS_k(z^2) + T_k(z)}{1 - zT(z)}.$$

The asymptotic expansion of $T(z)$ is given by

$$T(z) \sim \frac{1}{\rho} - a\sqrt{1 - \frac{z}{\rho}}.$$

In [**15**, p. 477] we find the numerical values of the constants ρ and a. (*Caveat*: The scaling is different, so [**15**, p. 477] in fact lists $a \cdot \rho$, not a.) We have $\rho \approx 0.4026975036714412909690453$ and $a \approx 2.8061602222420538943722824$. Using this expansion we get

$$\mathbb{P}(Y_n \geq k) = \frac{[z^n]S_k(z)}{n[z^n]T(z)} \sim \frac{2}{a^2\rho}\left(\rho S_k(\rho^2) + T_k(\rho)\right).$$

By denoting $\Psi(f(z)) = \frac{zf(z^2)+T_k(z)}{1-zT(z)}$ we can use the same arguments as in the Pólya case to efficiently obtain numerical values for the probabilities $\mathbb{P}(Y_n \geq k)$. Finally, we are able to calculate the asymptotic mean and variance for the protection number of a random node in non-plane binary trees.

Theorem 6. *Let Y_n be the protection number of a random internal vertex in a random non-plane binary tree of size n. Then*

$$\lim_{n\to\infty} \mathbb{E}Y_n = \frac{2}{a^2\rho}\sum_{k\geq 1}\left(\rho S_k(\rho^2) + T_k(\rho)\right) \approx 1.3124128299,$$

and $\lim_{n\to\infty} \mathbb{V}Y_n \approx 0.2676338724$.

Remark. As in the Pólya case, we can compare $S_k(z)$ with the corresponding function of plane binary trees and argue as in the proof of Lemma 7. This shows that $S_k(z)$ converges to zero at an even super-exponential rate.

5. CONCLUSION

In this paper we generalized the work of Heuberger and Prodinger, who derived the average protection number of plane trees, to a more general framework. We obtained the average protection number for all simply generated trees, as well as for Pólya trees and non-plane binary trees. We did not include Pólya trees with general degree restrictions, since the general expressions will look clumsy and only numerical results for specific classes may be of interest. But it is immediate that the asymptotic mean and variance of the protection number for Pólya-trees with any kind of degree restriction can be calculated in the very same way. As we saw in some of the examples, there are classes of trees, for which the obtained formulas involve a recurrence that might not be solvable explicitly. However, using these equations it is possible to calculate the asymptotic mean and variance in an arbitrarily accurate way with a fairly low computational effort. In Table 2 we summarize the obtained results for some specific tree classes.

Tree model	$\lim_{n\to\infty} \mathbb{E}X_n$	$\lim_{n\to\infty} \mathbb{E}Y_n$
Simply generated trees		
Plane trees	1.62297	0.72765
Motzkin trees	2.54638	1.30760
Incomplete binary trees	3.53647	1.99182
Cayley trees	2.28620	1.18652
Complete binary trees	1.56298	1.26568
Non-plane trees		
Pólya trees	2.15489	0.99532
Non-plane binary trees	1.70760	1.31241

Table 2: Summary of the obtained mean values for the protection numbers.

It is well known that Cayley trees and Pólya trees are very similar, but the latter are not simply generated, as the simple proof presented in [**10**] shows. A detailed analysis of the structural differences was done in [**16, 25**]: Roughly speaking, Pólya trees are Cayley trees (more precisely, the simply generated class whose ordinary generating function is the exponential generating function of Cayley trees) with small forests attached to each vertex. Comparing the resulting values (from Table 2) for Cayley trees and Pólya trees shows the quantitative effect of those forests, which have on average less than one vertex. As expected, these additional forests decrease the protection numbers.

For complete binary trees the correspondence between plane and non-plane is a bit different due to the strict degree constraint. The small forests are not attached anywhere, but they always consist of two identical trees and attachment is done by replacing a leaf. The effect of the presence or absence of symmetries seems stronger than the possible increase of the protection number by adding forests, because the larger number of plane structures gives some bias to lower protection numbers.

Furthermore, since we have explicit formulas for the asymptotic probabilities $\mathbb{P}(X_n \geq k)$ and $\mathbb{P}(Y_n \geq k)$, we have direct access to the discrete probability distributions of the random variables X_n and Y_n, summarized in Table 3. Some concrete approximate values of the limiting distribution are given in Table 4.

Tree model	$\lim_{n\to\infty} \mathbb{P}(X_n = k)$
Simply generated trees	$\rho^{k-1} \prod_{i=1}^{k-1} \phi'(T_i(\rho))(1-\rho\phi'(T_k(\rho)))$
Pólya trees	$\prod_{i=1}^{k}(T_i(\rho)+\rho)(1-\rho-T_{k+1}(\rho))$
Non-plane binary trees	$\prod_{i=1}^{k-1}(\rho T_i(\rho))(1-\rho T_k(\rho))$

Tree model	$\lim_{n\to\infty} \mathbb{P}(Y_n = k)$
Simply generated trees	$\frac{\phi_0}{T(\rho)}(T_k(\rho)-T_{k+1}(\rho))$
Pólya trees	$\frac{2}{b^2}(S_k(\rho)-S_{k+1}(\rho))$
Non-plane binary trees	$\frac{2}{a^2\rho}(\rho S_k(\rho^2)+T_k(\rho)-\rho S_{k+1}(\rho^2)-T_{k+1}(\rho))$

Table 3: Discrete asymptotic distributions of the protection number of the root (X_n), and of a random vertex (Y_n), respectively.

	plane trees	binary trees	Pólya trees	non-plane binary trees
$k=1$	0.55556	0.5	0.33832	0.4026975
$k=2$	0.31221	0.4375	0.35205	0.4888019
$k=3$	0.09762	0.06201171875	0.19036	0.1067009
$k=4$	0.02585	0.000488266	0.07698	0.0017995
$k=5$	0.00655	0.000000015	0.02775	0.0000002
$k=6$	0.00165	$6.939 \cdot 10^{-15}$	0.00959	$2.346 \cdot 10^{-15}$

Table 4: The first few (rounded) values for $\lim_{n\to\infty} \mathbb{P}(X_n = k)$.

Table 5 lists the first few probabilities of four tree classes. As in the binary cases the leaves do not contribute to the tree size, they are also disregarded with respect to protection numbers. If we take them into account, the probabilities would only have to be multiplied by $1/2$. In this case, the values for expectation and variance would change as well. For the plane case, we would have

$$\lim_{n\to\infty} \mathbb{E}(Y_n = k) \approx 0.6328430180 \quad \text{and} \quad \lim_{n\to\infty} \mathbb{V}(Y_n = k) \approx 0.5137858418,$$

instead of the values listed in Table 1 and in the non-plane case

$$\lim_{n\to\infty} \mathbb{E}(Y_n = k) \approx 0.6562064150 \quad \text{and} \quad \lim_{n\to\infty} \mathbb{V}(Y_n = k) \approx 0.5644237952.$$

instead of the values listed in Theorem 6.

	plane trees	binary trees	Pólya trees	non-plane binary trees
$k=1$	0.33333	0.75	0.2939995	0.7137829891
$k=2$	0.12121	0.23438	0.1610251	0.2602367674
$k=3$	0.03383	0.01556	0.0682504	0.0257646827
$k=4$	0.00870	0.00006	0.0252114	0.0002155459
$k=5$	0.00219	$9.314 \cdot 10^{-10}$	0.0088016	0.0000000148
$k=6$	0.00055	$2.168 \cdot 10^{-19}$	0.0030102	$7.025 \cdot 10^{-17}$

Table 5: The first few (rounded) values for $\lim_{n\to\infty} \mathbb{P}(Y_n = k)$. In the binary cases, only internal nodes are considered.

Acknowledgements. This research has been supported by the ÖAD, grant PL04-2018, and Wrocław Univeristy of Science and Technology grant 0401/0052/18. The first and third author have also been supported by the Austrian Science Fund (FWF), grant SFB F50-03.

REFERENCES

1. D. Aldous: *Asymptotic fringe distributions for general families of random trees.* Ann. Appl. Probab., **1(2)** (1991), 228–266.
2. M. Bóna, B. Pittel: *On a random search tree: asymptotic enumeration of vertices by distance from leaves.* Adv. Appl. Probab., **49(3)** (2017), 850–876.
3. Gi-Sang Cheon, L. W. Shapiro: *Protected points in ordered trees.* Appl. Math. Lett., **21(5)** (2008), 516–520.
4. K. Copenhaver: *k-protected vertices in unlabeled rooted plane trees.* Graph. Comb., **33(2)** (2017), 347–355.
5. J.-F. Delmas, J.-S. Dhersin, M. Sciauveau: *Cost functionals for large (uniform and simply generated) random trees.* Electron. J. Probab., **23** (2018), Paper No. 87.
6. L. Devroye: *A note on the probabilistic analysis of Patricia trees.* Random Struct. Algor., **3(2)** (1992), 203–214.
7. L. Devroye, S. Janson: *Protected nodes and fringe subtrees in some random trees.* Electron. Commun. Probab., **19** (2014), Paper No. 6.
8. M. Drmota: *The saturation level in binary search tree.* In Danièle Gardy and Abdelkader Mokkadem, editors, Mathematics and Computer Science, Trends. Math., Birkhäuser Basel (2000), 41–51.
9. M. Drmota: *Random Trees. An interplay between combinatorics and probability.* Springer, Wien, NewYork, 2009.

10. M. Drmota, B. Gittenberger: *The shape of unlabeled rooted trees.* European J. Combinat., **31** (2010), 2028–2063.

11. R. R. X. Du, H. Prodinger: *Notes on protected nodes in digital search trees.* Appl. Math. Lett., **25(6)** (2012), 1025–1028.

12. P. Duchon, P. Flajolet, G. Louchard, G. Schaeffer: *Boltzmann samplers for the random generation of combinatorial structures.* Combin. Probab. Comput., **13(4-5)** (2004), 577–625.

13. S. R. Finch: *Mathematical Constants*, volume 94 of Encyclopedia of Mathematics and its Applications. Cambridge University Press, Cambridge, 2003.

14. P. Flajolet, A. Odlyzko: *Singularity analysis of generating functions.* SIAM J. Discrete Math., **3(2)** (1990), 216–240.

15. P. Flajolet, R. Sedgewick: *Analytic Combinatorics.* Cambridge University Press, Cambridge, 2009.

16. B. Gittenberger, E.Y. Jin, M. Wallner: *On the shape of random Pólya structures.* Discrete Math., **341(4)** (2018), 896–911.

17. C. Heuberger, H. Prodinger: *Protection number in plane trees.* Appl. Anal. and Discr. Math., **11(2)** (2017), 314–326.

18. C. Holmgren, S. Janson: *Limit laws for functions of fringe trees for binary search trees and random recursive trees.* Electron. J. Probab., **20** (2015), Paper No. 4

19. S. Janson: *Simply generated trees, conditioned Galton-Watson trees, random allocations and condensation.* Probab. Surv., **9** (2012), 103–252.

20. H. M. Mahmoud, M. D. Ward: *Asymptotic distribution of two-protected nodes in random binary search trees.* Appl. Math. Lett., **25(12)** (2012), 2218–2222.

21. H. M. Mahmoud, M. D. Ward: *Asymptotic properties of protected nodes in random recursive trees.* J. Applied Probability, **52(1)** (2015), 290–297.

22. T. Mansour: *Protected points in k-ary trees.* Appl. Math. Lett., **24(4)** (2011), 478–480.

23. A. Meir, J. W. Moon: *On the altitude of nodes in random trees.* Canadian J. Math., **30** (1978), 997–1015.

24. R. Otter: *The number of trees.* Ann. Math., **49(3)** (1948), 583–599.

25. K. Panagiotou, B. Stufler: *Scaling limits of random Pólya trees.* Probab. Theory Rel., **170(3-4)** (2018), 801–820.

26. G. Pólya: *Kombinatorische Anzahlbestimmungen für Gruppen, Graphen und chemische Verbindungen.* Acta Math., **68(1)** (1937), 145–254.

27. S. Wagner: *Central limit theorems for additive tree parameters with small toll functions.* Combin. Probab. Comput., **24(1)** (2015), 329–353.

Bernhard Gittenberger,
Department of Discrete Mathematics and Geometry,
Technische Universität Wien,
Wiedner Hauptstraße 8-10/104, A-1040
Wien, Austria
E-mail: *gittenberger@dmg.tuwien.ac.at*

(Received 29. 03. 2019.)
(Revised 31. 03. 2021.)

Zbigniew Gołębiewski,
Department of Fundamentals of Computer Science,
Faculty of Fundamental Problems of Technology,
Wrocław University of Science and Technology,
ul. Wybrzeże Wyspiańskiego 27, 50-370,
Wrocław, Poland
E-mail: *zbigniew.golebiewski@pwr.edu.pl*

Isabella Larcher,
Department of Discrete Mathematics and Geometry,
Technische Universität Wien,
Wiedner Hauptstraße 8-10/104, A-1040
Wien, Austria
E-mail: *isabella.larcher@tuwien.ac.at*

Małgorzata Sulkowska,
Department of Fundamentals of Computer Science,
Faculty of Fundamental Problems of Technology,
Wrocław University of Science and Technology,
ul. Wybrzeże Wyspiańskiego 27, 50-370,
Wrocław, Poland
E-mail: *malgorzata.sulkowska@pwr.edu.pl*