Iterative Cutting and Pruning of Planar Trees

Benjamin Hackl* Sara Kropf† Helmut Prodinger‡

Abstract

Rooted plane trees are reduced by four different operations on the fringe. The number of surviving nodes after reducing the tree repeatedly for a fixed number of times is asymptotically analyzed. The four different operations include cutting all or only the leftmost leaves or maximal paths. This generalizes the concept of pruning a tree.

The results include exact expressions and asymptotic expansions for the expected value and the variance as well as central limit theorems.

1 Introduction

Rooted plane trees are among the most interesting elementary combinatorial objects; they appear in the literature under many different names such as ordered trees, planar trees, planted plane trees, etc. They have been analyzed under various aspects, especially due to their relevance in Computer Science. Two particularly well-known quantities are the height, since it is equivalent to the stack size needed to explore binary (expression) trees, and the pruning number (pruning index), since it is equivalent to the register function (Horton-Strahler number) of binary trees. Several results for the height of rooted plane trees can be found in [2, 6, 12], for the register function, we refer to [3, 7, 9], and for the pruning number to [3, 16].

Reducing (cutting-down) trees has also been a popular research theme during the last decades [8, 10, 11]: According to a certain probabilistic model, a given tree is reduced until a certain condition is satisfied (usually, the root is isolated).

In the present paper, the point of view is slightly different, as we reduce the trees in a completely deterministic fashion at the leaves until the tree has no more edges. All these reductions take place on the fringe. We consider four different models:

- In one round, all leaves together with the corresponding edges are removed (see Section 2).
- In one round, all maximal paths (linear graphs), with the leaves on one end, are removed (see Section 3). This process is called pruning.
- A leaf is called an old leaf, if it is the leftmost sibling of its parents. This concept was introduced in [1]. In one round, only old leaves are removed (see Section 4).
- The last model deals with pruning of old paths. There might be several interesting models related to this; the one we have chosen here is that in one round maximal paths are removed, under the condition that each of their nodes is the leftmost child of their parent node (see Section 5).

The first model is clearly related to the height of the plane tree, and the second one to the Horton-Strahler number via the pruning index [16]. While there are no surprises here about the number of rounds that the process takes, we are interested in how the fringe develops. The number of leaves and nodes altogether in the remaining tree after a fixed number of reduction rounds is the main parameter analyzed in this extended abstract.

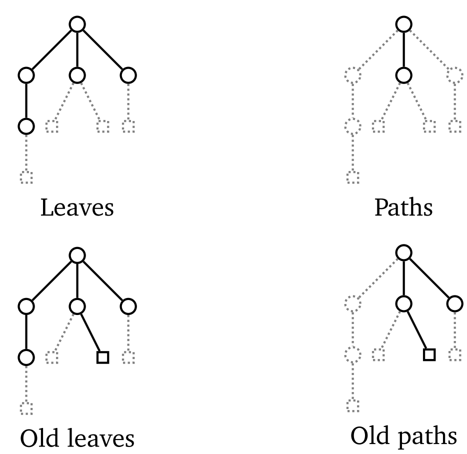

Figure 1: Removal of (old) leaves / paths.

The four tree reductions are illustrated in Figure 1. We describe these reductions more formally in the corresponding sections.

Proofs and additional results will appear in a full version of this extended abstract.

*Institut für Mathematik, Alpen-Adria-Universität Klagenfurt, 9020 Klagenfurt, Austria. benjamin.hackl@aau.at

†Institute of Statistical Science, Academia Sinica, 115 Taipei, Taiwan. sarakropf@stat.sinica.edu.tw

‡Department of Mathematical Sciences, Stellenbosch University, 7602 Stellenbosch, South Africa. hproding@sun.ac.za

Observe that some trees do not "survive" a fixed number of rounds. However, since we are interested in the asymptotics for a random tree with n nodes, where $n \to \infty$, and since we can show that the proportion of the trees that die out is exponentially small, the contribution of these trees is asymptotically irrelevant.

The random variable $X_{n,r}$ models the general tree size after reducing a rooted plane tree of size n (that is chosen uniformly at random among all trees with n nodes) r-times iteratively. In particular, for $r = 0$, the given plane tree is not changed and $X_{n,0} = n$.

As we will see later, a key aspect of the analysis of the four tree reductions mentioned above is the translation of the algorithmic description of the reduction into an operator that acts on the corresponding generating functions. Then, in order to analyze the r-fold application of these operators to the appropriate generating functions, some prerequisites and auxiliary concepts are required. In particular, we need generating functions for counting all nodes and leaves and also for nodes and old leaves. The first enumeration (nodes and leaves) is intimately linked to Narayana numbers. In this paper, it is easier to deal with non-leaf nodes and leaf nodes, which is just a small modification.

PROPOSITION 1.1. *The generating function $T(z,t)$ which enumerates rooted plane trees with respect to their internal nodes (marked by the variable z) and leaves (marked by t) is given explicitly by*

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

This follows from the functional equation

$$T(z,t) = t + \frac{zT(z,t)}{1-T(z,t)}.$$

DEFINITION 1.1. *The* Narayana numbers *are defined as*

$$N_{n,k} = \frac{1}{n}\binom{n}{k-1}\binom{n}{k}$$

for $1 \le n$, $1 \le k \le n$ and $N_{0,0} = 1$. All other indices give $N_{n,k} = 0$. The Narayana polynomials *are defined as*

$$N_n(x) = \sum_{k=1}^{n} N_{n,k} x^{k-1}$$

for $n \ge 1$ and $N_0(x) = 1$, and the associated Narayana polynomials *are defined as $\tilde{N}_n(x) = x \cdot N_n(x)$ for $n \ge 0$. Note that*

$$N_n(1) = \tilde{N}_n(1) = C_n = \frac{1}{n+1}\binom{2n}{n}$$

is the n-th Catalan number.

REMARK 1.1. *The generating function*

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

enumerates Catalan numbers, see [4, Thm. 3.2], and the generating function $T(z,tz)$ enumerates Narayana numbers

$$T(z,tz) = zt + \sum_{n\ge 2}\sum_{k=1}^{n-1} N_{n-1,k} z^n t^k = \sum_{n\ge 1} z^n \tilde{N}_{n-1}(t).$$

PROPOSITION 1.2. *The generating function $L(z,w)$ enumerating rooted plane trees with respect to old leaves together with their predecessor (marked by the variable w) and all other nodes (marked by z) is given by*

$$L(z,w) = \frac{1-\sqrt{1-4z-4w+4z^2}}{2}.$$

There are $C_{k-1}\binom{n-2}{n-2k}2^{n-2k}$ rooted plane trees of size n with k old leaves.

For example, the original tree in Figure 1 corresponds to z^3w^3 because it has three old leaves and three nodes which are neither old leaves nor parents of old leaves.

This proposition follows from the functional equation

$$L(z,w) = z + \frac{w + z(L(z,w)-z)}{1-L(z,w)}$$

and a careful application of the Lagrange inversion formula.

DEFINITION 1.2. *The* Fibonacci polynomials *are recursively defined by*

$$F_r(z) = F_{r-1}(z) + zF_{r-2}(z)$$

for $r \ge 2$ and $F_0(z) = 0$, $F_1(z) = 1$. They satisfy the identities

$$F_r(z)F_{r+2}(z) + (-z)^r = F_{r+1}(z)^2,$$
$$F_r(z)F_{2r-1}(z) - F_r(z)(-z)^{r-1} = F_{r-1}(z)F_{2r}(z).$$

See [5] for the corresponding results for Fibonacci numbers.

Substituting $z = u/(1+u)^2$, we obtain

$$F_r(-z) = \frac{1-u^r}{(1-u)(1+u)^{r-1}}.$$

Note that $zF_r(-z)/F_{r+1}(-z)$ is the generating function of rooted plane trees with height $\le r$ (see [2]).

We would like to emphasize the importance of the substitution $z = u/(1+u)^2$, which makes our expressions manageable.

DEFINITION 1.3. *The polynomials $B_r(z)$ are the generating functions of binary trees with height $\leq r$ satisfying*

$$B_r(z) = 1 + zB_{r-1}(z)^2$$

for $r \geq 1$ and $B_0(z) = 1$.

They play a role in the third model, where we cut old leaves.

2 Cutting Leaves

In this section we want to analyze the tree reduction that acts on a given tree by removing all of its leaves. In order to obtain suitable generating functions, we study the corresponding inverse operation, which is by no means unique. This tree expansion is described by the following operations:

All leaves of a given tree are expanded by appending a nonempty sequence of leaves to each of them. Additionally, each inner node of the original tree is expanded by appending (possibly empty) sequences of leaves between two of its children as well as before the first and after the last one.

For a tree marked by $z^n t^k$, i.e. consisting of n inner nodes and k leaves, this means that the expansion of the leaves replaces t by $zt/(1-t)$. Additionally, there are $2n+k-1$ positions for attaching sequences of leaves to inner nodes: This can be seen by observing that every inner node has exactly one available position more than it has children.

A linear operator Φ that describes this expansion in terms of generating functions can now be constructed as we know that the relation

$$\Phi(z^n t^k) = z^n \frac{z^k t^k}{(1-t)^k} \frac{1}{(1-t)^{2n+k-1}}$$

holds. Extending this operator linearly to formal power series, we obtain

$$\Phi(f(z,t)) := (1-t) f\left(\frac{z}{(1-t)^2}, \frac{zt}{(1-t)^2}\right).$$

Applying Φ to a generating function enumerating a family of rooted plane trees gives a generating function counting all possible tree expansions of this family.

Note that the generating function for rooted plane trees T satisfies the functional equation

$$T(z,t) = t + \Phi(T(z,t)),$$

which expresses the fact that all trees except for the one consisting only of the root (denoted by the summand z) can be obtained by expanding a smaller tree.

The iterated action of Φ to a typical monomial is this:

LEMMA 2.1. *Let $M_0(z,t) = z^n t^k$ for integers n, $k \geq 0$ and $M_r(z,t) = \Phi M_{r-1}(z,t)$ for $r \geq 1$. Then we have for $r \geq 0$*

$$\begin{aligned} M_r(z,z) &= \frac{z^{n+k(r+1)} F_{r+1}(-z)^{2n-1}}{F_{r+2}(-z)^{2n+2k-1}} \\ &= \frac{1-u^{r+2}}{(1-u^{r+1})(1+u)} \left(\frac{u(1-u^{r+1})^2}{(1-u^{r+2})^2}\right)^n \left(\frac{u^{r+1}(1-u)^2}{(1-u^{r+2})^2}\right)^k. \end{aligned}$$

COROLLARY 2.1. *The bivariate generating function $G_r(z,v)$ enumerating rooted plane trees whose leaves can be cut at least r-times, where z marks the tree size and v marks the size of the r-fold cut tree, is given by*

$$G_r(z,v) = \Phi^r T(zv, tv)|_{t=z}$$

and, equivalently, by

$$\begin{aligned} G_r(z,v) = &\frac{1-u^{r+2}}{(1-u^{r+1})(1+u)} \\ &\times T\left(\frac{u(1-u^{r+1})^2}{(1-u^{r+2})^2} v, \frac{u^{r+1}(1-u)^2}{(1-u^{r+2})^2} v\right), \end{aligned}$$

where $z = u/(1+u)^2$.

This generating function tells us how many nodes (marked by v) are still in the tree after r reductions. It is completely described in terms of the function $T(z,t)$, although in a non-trivial way. Results about moments and the limiting distribution can be extracted from this explicit form.

We investigate the behavior of the random variable $X_{n,r}$ that is the number of nodes which are left after reducing a random tree T_n with n nodes r-times. The tree T_n is chosen uniformly among all trees of size n. The probability generating function of this random variable is

$$\mathbb{E} v^{X_{n,r}} = \frac{a_{n,r} + [z^n] G_r(z,v)}{C_{n-1}} \tag{2.1}$$

where $a_{n,r}$ is the number of trees of size n which are empty after reducing r-times. We have $a_{n,r} = C_{n-1} - [z^n] G_r(z,1)$. In this way, the random variable is well defined for all trees of size n.

Furthermore, we consider the random variables $I_{n,r}$ and $L_{n,r}$ that are the number of inner nodes and leaves, respectively, which remain after reducing a random tree of size n exactly r-times. The probability generating functions of these random variables are as in (2.1) with $G_r(z,v)$ replaced by $\Phi^r T(zv,t)|_{t=z}$ and $\Phi^r T(z,tv)|_{t=z}$, respectively. The relations $X_{n,r} \stackrel{d}{=} I_{n,r} + L_{n,r}$ and $I_{n,r} \stackrel{d}{=} X_{n+1,r}$ hold by the combinatorial interpretation of the operator Φ.

We find explicit generating functions for the factorial moments using the notion of falling factorials

$$x^{\underline{d}} = x(x-1)\cdots(x-d+1).$$

LEMMA 2.2. *The d-th factorial moments of $X_{n,r}$, $I_{n,r}$ and $L_{n,r}$ are given by*

$$\begin{aligned}\mathbb{E}X_{n,r}^{\underline{d}} &= \mathbb{E}I_{n-1,r}^{\underline{d}}\\ &= \frac{1}{C_{n-1}}[z^n]\frac{u^d d!}{(1+u)(1-u^{r+1})^d(1-u)^{d-1}}\tilde{N}_{d-1}(u^r)\end{aligned}$$

and

$$\mathbb{E}L_{n,r}^{\underline{d}} = \frac{1}{C_{n-1}}[z^n]\frac{u^{dr+d}(1-u)d!}{(1+u)(1-u^{r+2})^d(1-u^{r+1})^d}\tilde{N}_{d-1}(u)$$

where $z = u/(1+u)^2$ for $d \in \mathbb{Z}_{\geq 1}$.

The proof of this requires heavy machinery from Computer Algebra, in particular the Mathematica package Sigma [14] (for a survey of Sigma's capabilities see, e.g., [13]), and will appear in the full version.

COROLLARY 2.2. *The expected value of $X_{n+1,r}$ is explicitly given by*

$$\mathbb{E}X_{n+1,r} = \frac{1}{C_n}\sum_{\ell\geq 1}\binom{2n}{n+1-\ell(r+1)} - \binom{2n}{n-\ell(r+1)}.$$

Having determined a closed form for this generating function allows us to analyze the asymptotic behavior of $X_{n,r}$ in a relatively straightforward way.

THEOREM 2.1. *Let $r \in \mathbb{N}_0$ be fixed and n go to infinity. Then the expected size and the corresponding variance of an r-fold cut rooted plane tree are given by*

$$\mathbb{E}X_{n,r} = \frac{n}{r+1} - \frac{r(r-1)}{6(r+1)} + O(n^{-1}),$$

and

$$\mathbb{V}X_{n,r} = \frac{r(r+2)}{6(r+1)^2}n + O(1).$$

The factorial moments are asymptotically given by

$$\begin{aligned}\mathbb{E}X_{n,r}^{\underline{d}} &= \frac{1}{(r+1)^d}n^d\\ &+ \frac{d}{24(r+1)^d}(2dr^2 - 6r^2 - 8dr + 12r - 21d + 30)n^{d-1}\\ &\qquad + O(n^{d-3/2})\end{aligned}$$

for $d \geq 2$.

THEOREM 2.2. *The size $X_{n,r}$ of the tree obtained from a random rooted plane tree with n nodes by cutting it r-times is, after standardization, asymptotically normally distributed for n going to infinity and fixed r, i.e.,*

$$\frac{X_{n,r} - \dfrac{n}{r+1}}{\sqrt{\dfrac{r(r+2)}{6(r+1)^2}n}} \xrightarrow{d} \mathcal{N}(0,1).$$

As $I_{n,r-1} \overset{d}{=} X_{n,r}$, the same also holds for this random variable.

The proof uses a general central limit theorem for additive tree parameters [15] and can be found in the full version of this extended abstract.

REMARK 2.1. *Note that for $r = 1$, this result means that the Narayana numbers are asymptotically normally distributed, see for example [4, Thm. 3.13].*

3 Cutting Paths

As in the previous section, we work with a linear operator: Applying it to a generating function of plane trees produces the generating function of all trees that exactly leads to the given trees after one reduction (pruning).

The corresponding expansion operator can be constructed similarly to the one from the previous section: Instead of just appending leaves, we append paths that are denoted by $P = t/(1-z)$ (as they consist of a possibly empty sequence of inner nodes and one leaf). Also, observe that leaves in the original tree correspond to nodes from which at least two separate paths branch out in an expanded tree. Hence, the operator has to satisfy

$$\Phi(z^n t^k) = z^n \frac{z^k P^{2k}}{(1-P)^k}\frac{1}{(1-P)^{2n+k-1}}.$$

By linearly expanding Φ to formal power series, we find that Φ is defined as

$$\Phi(f(z,t)) = (1-P)f\left(\frac{z}{(1-P)^2}, \frac{zP^2}{(1-P)^2}\right)$$

with $P = t/(1-z)$. Observe that it is not possible to construct a path by expanding any other tree, as leaves in the original tree become nodes from which at least two paths branch out in the extended one. This fact leads to the valid functional equation

$$T(z,t) = P + \Phi(T(z,t)),$$

meaning that all trees can either be obtained by expanding some other tree, or are paths.

LEMMA 3.1. *Let $M_0(z,t) = z^n t^k$ for integers n, $k \geq 0$ and $M_r(z,t) = \Phi M_{r-1}(z,t)$ for $r \geq 1$. Then*

$$\begin{aligned}M_r(z,z) &= \frac{z^{n+(2^r-1)k}F_{2^{r+1}-1}(-z)^{2n-1}}{F_{2^{r+1}}(-z)^{2n+2k-1}} =\\ &\frac{1-u^{2^{r+1}}}{(1-u^{2^{r+1}-1})(1+u)}\\ &\qquad\times\left(\frac{u(1-u^{2^{r+1}-1})^2}{(1-u^{2^{r+1}})^2}\right)^n\left(\frac{u^{2^{r+1}-1}(1-u)^2}{(1-u^{2^{r+1}})^2}\right)^k\end{aligned}$$

holds for $r \geq 0$.

COROLLARY 3.1. *The bivariate generating function $G_r(z,v)$ enumerating rooted plane trees whose paths can be cut at least r-times, where z marks the tree size and v marks the size of the r-fold cut tree, is given by* $G_r(z,v) = \Phi^r T(zv,tv)|_{t=z}$ *and, equivalently, by*

$$G_r(z,v) = \frac{1-u^{2^{r+1}}}{(1-u^{2^{r+1}-1})(1+u)} \times T\left(\frac{u(1-u^{2^{r+1}-1})^2}{(1-u^{2^{r+1}})^2}v, \frac{u^{2^{r+1}-1}(1-u)^2}{(1-u^{2^{r+1}})^2}v\right)$$

where $z = u/(1+u)^2$.

We investigate the behavior of the random variable $Y_{n,r}$ that is the number of nodes which are left after reducing a random tree T_n with n nodes r-times. The tree T_n is chosen uniformly among all trees of size n. The probability generating function of this random variable is

$$\mathbb{E}v^{Y_{n,r}} = \frac{a_{n,r} + [z^n]G_r(z,v)}{C_{n-1}}$$

where $a_{n,r}$ is the number of trees of size n which are empty after reducing r-times. We have $a_{n,r} = C_{n-1} - [z^n]G_r(z,1)$.

By comparing Corollaries 2.1 and 3.1, we see that $Y_{n,r} \stackrel{d}{=} X_{n,2^{r+1}-2}$ because they have the same probability generating function. Thus all the results from Section 2 follow in this case, too. We just have to write $2^{r+1}-2$ instead of r.

THEOREM 3.1. *Let $r \in \mathbb{N}_0$ be fixed and n go to infinity. Then*

$$\mathbb{E}Y_{n,r} = \frac{n}{2^{r+1}-1} - \frac{(2^r-1)(2^{r+1}-3)}{3(2^{r+1}-1)} + O(n^{-1}),$$

and

$$\mathbb{V}Y_{n,r} = \frac{2^{r+1}(2^r-1)}{3(2^{r+1}-1)^2}n + O(1).$$

The factorial moments are asymptotically given by

$$\mathbb{E}Y_{n,r}^{\underline{d}} = \frac{n^d}{(2^{r+1}-1)^d} + \frac{d}{24(2^{r+1}-1)^d}(8\cdot 4^r(d-3) - 8\cdot 2^r(4d-9) + 3d - 18) + O(n^{d-3/2}).$$

Furthermore, the random variable $Y_{n,r}$ is asymptotically normally distributed, i.e.,

$$\frac{Y_{n,r} - \dfrac{n}{2^{r+1}-1}}{\sqrt{\dfrac{2^{r+1}(2^{r+1}-1)}{3(2^{r+1}-1)^2}n}} \xrightarrow{d} \mathcal{N}(0,1).$$

4 Cutting Old Leaves

We now turn to the analysis of the reduction procedure that deletes the old leaves in every step. Following [1], we say that a leaf is an old leaf if it is the leftmost child of its parent. Recall that in the context of these old leaves we mark an old leaf together with its parent by w in the generating functions, and all other nodes by z.

The corresponding expansion can be explained easily: For trees with at least two nodes all old leaves together with their respective parents have to be expanded by either adding another old leaf to the original one (contributing zw), by adding an old leaf to the parent (contributing zw as well), or by adding an old leaf to both the parent and the original leaf (contributing w^2). Thus, w has to be replaced by $(2z+w)w$. All other nodes can either remain as they are (which does not change z), or we can add an old leaf to the node (which gives w), meaning that z has to be replaced by $z+w$.

Therefore, the expansion operator Φ has to be defined as

$$\Phi(f(z,w)) = f(z+w,(2z+w)w),$$

and it is linear and multiplicative. The following lemma describes the result for applying Φ to a monomial repeatedly.

LEMMA 4.1. *Let $M_0(z,w) = z^n w^k$ for integers n, $k \geq 0$ and $M_r(z,w) = \Phi M_{r-1}(z,w)$ for $r \geq 1$. Then*

$$M_r(z,z^2) = z^{n+k}B_{r-1}(z)^n(B_r(z) - B_{r-1}(z))^k$$

holds for $r \geq 0$ where $B_r(z)$ are defined in Definition 1.3.

Note that as far as the combinatorial interpretation of Φ as an expansion operator goes, we have to be more careful in this case: We already mentioned that the expansion only works for trees with at least two nodes—the tree with one node (represented by z) is expanded incorrectly: $\Phi(z) = z + w$, meaning that the tree might also stay the same. In comparison to the previous sections, this correction is also the reason for the simple functional equation

$$L(z,w) = \Phi(L(z,w)).$$

In order to obtain the correct generating function we subtract this linear term from $L(z,w)$ and add those trees that reduce to ○ after r reductions separately. These trees are exactly those that reduce to the tree of size two (which is one old leaf plus its parent) after $r-1$ reductions.

COROLLARY 4.1. *The bivariate generating function $G_r(z,v)$ enumerating rooted plane trees whose paths can be cut at least r-times, where z marks the tree size and v marks the size of the r-fold cut tree, is given by*

$$G_r(z,v) = \Phi^r\big(L(zv,wv^2) - zv\big)\Big|_{w=z^2} + v\Phi^{r-1}(w)\Big|_{w=z^2}$$

and, equivalently, by

$$G_r(z,v) = L(zB_{r-1}(z)v, z(B_r(z) - B_{r-1}(z))v^2) - vzB_{r-2}(z).$$

We are now interested in the random variable $X_{n,r}$ which models the size of the tree that results from reducing a random tree T_n with n nodes r-times. The tree T_n is chosen uniformly at random among all trees of size n. Similar to the previous sections, the probability generating function of this random variable is

$$\mathbb{E}v^{X_{n,r}} = \frac{a_{n,r} + [z^n]G_r(z,v)}{C_{n-1}}$$

where $a_{n,r}$ is the number of trees of size n which are empty after reducing r-times. The relation $a_{n,r} = C_{n-1} - [z^n]G_r(z,1)$ holds.

While the height polynomials $B_r(z)$ make it very difficult to obtain general results for the factorial moments of $X_{n,r}$ or a central limit theorem, special moments like expectation and variance are no problem.

THEOREM 4.1. *Let $r \in \mathbb{N}_0$ be fixed and n go to infinity. Then the expected tree size after deleting the old leaves of a tree with n nodes r-times and the corresponding variance are given by*

$$\mathbb{E}X_{n,r} = (2 - B_{r-1}(1/4))n - \frac{B'_{r-1}(1/4)}{8} + O(n^{-1}),$$

and

$$\mathbb{V}X_{n,r} = \Bigg(B_{r-1}(1/4) - B_{r-1}(1/4)^2 + \frac{(2 - B_{r-1}(1/4))B'_{r-1}(1/4)}{2}\Bigg)n + O(1).$$

REMARK 4.1. *In [6], the asymptotic behavior of a sequence strongly related to $B_r(1/4)$ was studied. With the help of these results, the asymptotic behavior of $B_r(1/4)$ can be described as*

$$B_r(1/4) = 2 - \frac{4}{r} - \frac{4\log r}{r^2} + O(r^{-3}) \quad \textit{for } r \to \infty.$$

5 Cutting Old Paths

We are now interested in a variant of the concept of tree pruning within the context of old nodes. We choose to investigate the following pruning strategy: We only delete a maximal path if all of its nodes are leftmost children of their parent nodes. In particular, this means that in order for some path to be removed, it needs to branch out from the leftmost position.

In fact, the corresponding expansion can be constructed analogously to the expansion related to cutting old leaves—with the sole difference that we do not just append leaves, but entire paths. With $P = w/(1-z)$ denoting the generating function of these paths including the base node, the expansion operator Φ is given by

$$\Phi(f(z,w)) = f(z+P, (z+P)P).$$

LEMMA 5.1. *Let $M_0(z,w) = z^n w^k$ for integers $n, k \geq 0$ and $M_r(z,w) = \Phi M_{r-1}(z,w)$ for $r \geq 1$. Then*

$$M_r(z,z^2) = \frac{z^{n+(r+2)k}F_{r+1}(-z)^n}{F_{r+2}(-z)^{n+2k}} = \left(\frac{u(1-u^{r+1})}{(1+u)(1-u^{r+2})}\right)^n \left(\frac{u^{r+2}(1-u)^2}{(1+u)^2(1-u^{r+2})^2}\right)^k$$

holds for $r \geq 0$ where $F_r(z)$ are the Fibonacci polynomial.

As in the previous section, we have to be careful about the combinatorial interpretation of the expansion operator Φ. In particular, this implies the relation

$$L(z,w) = \Phi(L(z,w)).$$

In order to obtain the correct generating function, we subtract the linear term from $L(z,w)$ and add those trees that reduce to ○ after r reductions separately. These trees are exactly those that reduce to one path, denoted by P, after $r-1$ reductions.

COROLLARY 5.1. *The bivariate generating function $G_r(z,v)$ enumerating rooted plane trees whose paths can be cut at least r-times, where z marks the tree size and v marks the size of the r-fold cut tree, is given by*

$$G_r(z,v) = \Phi^r\big(L(zv, wv^2) - zv\big)\big|_{w=z^2} + v\Phi^{r-1}(P)\big|_{w=z^2}$$

and, equivalently, by

$$G_r(z,v) = L\left(\frac{u(1-u^{r+1})}{(1+u)(1-u^{r+2})}v, \frac{u^{r+2}(1-u)^2}{(1+u)^2(1-u^{r+2})^2}v^2\right) - \frac{vu(1-u^r)}{(1+u)(1-u^{r+1})}$$

where $z = u/(1+u)^2$.

We investigate the behavior of the random variable $X_{n,r}$ which is the number of nodes which are left after reducing a random tree T_n with n nodes r-times. The tree T_n is chosen uniformly among all trees of size n. The probability generating function of this random variable is

$$\mathbb{E}v^{X_{n,r}} = \frac{a_{n,r} + [z^n]G_r(z,v)}{C_{n-1}}$$

where $a_{n,r}$ is the number of trees of size n which are empty after reducing r-times. We have $a_{n,r} = C_{n-1} - [z^n]G_r(z,1)$.

LEMMA 5.2. *The first and second factorial moments of $X_{n,r}$ are*

$$\mathbb{E}X_{n,r} = \frac{1}{C_{n-1}}[z^n]\frac{u^{r+1}(1+u^2-2u^{r+2})}{(1+u)(1-u^{r+1})(1-u^{r+2})}$$

and

$$\mathbb{E}X_{n,r}(X_{n,r}-1) = \frac{1}{C_{n-1}}[z^n]\frac{2u^{r+2}(1+u)}{(1-u)(1-u^{r+2})^2},$$

respectively.

By expanding the expressions in Lemma 5.2 and using singularity analysis, we obtain the asymptotic growth of the expected value and the variance. Further factorial moments or a central limit theorem seem to be more involved.

THEOREM 5.1. *Let $r \in \mathbb{N}$ be fixed and n go to infinity. Then the expected size and the corresponding variance of an r-fold cut rooted plane tree are given by*

$$\mathbb{E}X_{n,r} = \frac{2n}{r+2} - \frac{r(r+1)}{3(r+2)} + O(n^{-1}),$$

and

$$\mathbb{V}X_{n,r} = \frac{2r(r+1)}{3(r+2)^2}n + O(1).$$

Acknowledgment We thank Carsten Schneider for showing us how to simplify double sums using Sigma.

References

[1] William Y. C. Chen, Emeric Deutsch, and Sergi Elizalde. Old and young leaves on plane trees. *European J. Combin.*, 27(3):414–427, 2006.

[2] Nicolaas G. de Bruijn, Donald E. Knuth, and Stephen O. Rice. The average height of planted plane trees. In *Graph theory and computing*, pages 15–22. Academic Press, New York, 1972.

[3] Mireille Vauchaussade de Chaumont. *Nombre de Strahler des arbres, languages algébrique et dénombrement de structures secondaires en Biologie moléculaire*. 1985. Doctoral thesis, Université de Bordeaux I.

[4] Michael Drmota. *Random trees*. SpringerWienNewYork, 2009.

[5] Dustan Everman, A. E. Danese, K. Venkanayah, and E. M. Scheuer. Some properties of the Fibonacci numbers. *Amer. Math. Monthly*, 67(7):694, 1960. Problem E1396.

[6] Philippe Flajolet and Andrew Odlyzko. The average height of binary trees and other simple trees. *J. Comput. System Sci.*, 25(2):171–213, 1982.

[7] Philippe Flajolet, Jean-Claude Raoult, and Jean Vuillemin. The number of registers required for evaluating arithmetic expressions. *Theoret. Comput. Sci.*, 9(1):99–125, 1979.

[8] Svante Janson. Random cutting and records in deterministic and random trees. *Random Structures Algorithms*, 29(2):139–179, 2006.

[9] Rainer Kemp. A note on the stack size of regularly distributed binary trees. *BIT*, 20(2):157–162, 1980.

[10] Amram Meir and John W. Moon. Cutting down random trees. *J. Austral. Math. Soc.*, 11:313–324, 1970.

[11] Alois Panholzer. Cutting down very simple trees. *Quaest. Math.*, 29(2):211–227, 2006.

[12] Helmut Prodinger. The height of planted plane trees revisited. *Ars Combin.*, 16(B):51–55, 1983.

[13] Carsten Schneider. Symbolic summation assists combinatorics. *Sém. Lothar. Combin.*, 56:1–36, 2007.

[14] Carsten Schneider. *Sigma V1.81*. RISC, 2015. `http://www.risc.jku.at/research/combinatorics/Sigma/`.

[15] Stephan Wagner. Central limit theorems for additive tree parameters with small toll functions. *Combinatorics, Probability and Computing*, 24:329–353, 2015.

[16] Doron Zeilberger. A bijection from ordered trees to binary trees that sends the pruning order to the Strahler number. *Discrete Math.*, 82(1):89–92, 1990.