Reducing Simply Generated Trees by Iterative Leaf Cutting*

Benjamin Hackl[†] Clemens Heuberger[‡] Stephan Wagner[§]

Abstract

We consider a procedure to reduce simply generated trees by iteratively removing all leaves. In the context of this reduction, we study the number of vertices that are deleted after applying this procedure a fixed number of times by using an additive tree parameter model combined with a recursive characterization.

Our results include asymptotic formulas for mean and variance of this quantity as well as a central limit theorem.

1 Introduction

Trees are one of the most fundamental combinatorial structures with a plethora of applications not only in mathematics, but also in, e.g., computer science or biology. A matter of recent interest in the study of trees is the question of how a given tree family behaves when applying a fixed number of iterations of some given deterministic reduction procedure to it. In [8] four different iterative reduction procedures acting on plane trees are investigated and distributional results for corresponding tree parameters are proved. A similar analysis is carried out in [10], where the combinatorial objects of interest are from a special subclass of plane trees. Furthermore, see [9] for the investigation of an iterative reduction procedure acting on binary trees that gives rise to an alternative interpretation of the well-known register function.

每次删去所有叶子节点

In the scope of this extended abstract we focus on the, in a sense, most natural reduction procedure: we reduce a given rooted tree by cutting off all leaves so that only internal nodes remain. This process is illustrated in Figure 1. While in this extended abstract we are mainly interested in the class of simply generated trees, further classes of rooted trees will be investigated in the full version.

It is easy to see that the number of steps it takes to reduce the tree so that only the root remains is precisely the height of the tree, i.e., the greatest distance from the root to a leaf. A more delicate question—the one in the center of this article—is to ask for a precise analysis of the number of vertices deleted when applying the "cutting leaves" reduction a fixed number of times.

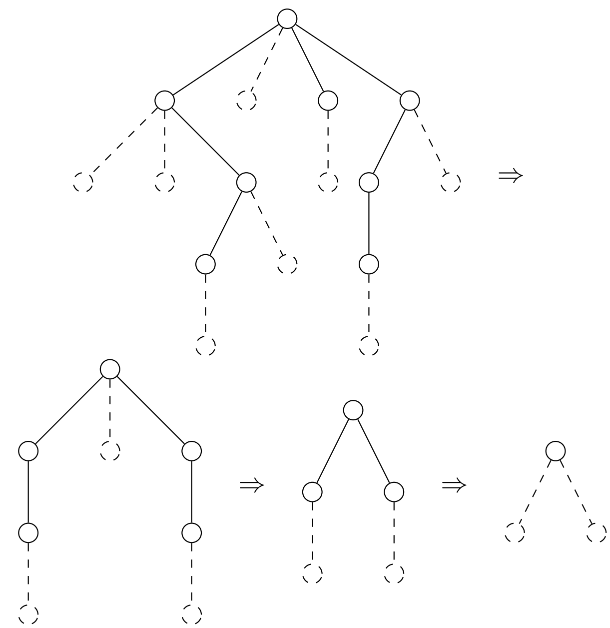


Figure 1: Multiple applications of the "cutting leaves" process to a given rooted tree

The key concepts behind our analysis are a recursive characterization and bivariate generating functions. Details on our model are given in Section 2. The asymptotic analysis is then carried out in Section 3, with our main result given in Theorem 3.1. It includes precise asymptotic formulas for the mean and variance of the number of removed vertices when applying the reduction a fixed number of times. Furthermore, we also prove a central limit theorem.

Finally, in Section 4 we give an outlook on the analysis of the "cutting leaves" reduction in the context of other classes of rooted trees. Qualitative results for these classes are given in Theorem 4.1. The corresponding details will be published in the full version of this extended abstract.

The computational aspects in this extended abstract were carried out using the module for manipulating

*B. Hackl and C. Heuberger are supported by the Austrian Science Fund (FWF): P 28466-N35. S. Wagner is supported by the National Research Foundation of South Africa, grant 96236.

[†]Institut für Mathematik, Alpen-Adria-Universität Klagenfurt, Austria, benjamin.hackl@aau.at

[‡]Institut für Mathematik, Alpen-Adria-Universität Klagenfurt, Austria, clemens.heuberger@aau.at

[§]Department of Mathematical Sciences, Stellenbosch University, South Africa, swagner@sun.ac.za

这个使用 SageMath 计算近似的材料可以看一下

asymptotic expansions [7] in the free open-source mathematics software system SageMath [16]. A notebook containing our calculations can be found at

`https://arxiv.org/src/1808.00363/anc/simply-generated-trees.ipynb`.

2 Preliminaries

So-called *additive tree parameters* play an important role in our analysis of the number of removed nodes.

Definition 2.1. *A* fringe subtree *of a rooted tree is a subtree that consists of a vertex and all its descendants. An* additive tree parameter *is a functional G satisfying a recursion of the form*

$$G(T) = \sum_{j=1}^{k} G(T^j) + g(T), \tag{2.1}$$

where T is some rooted tree, T^1, T^2, …, T^k are the branches of the root of T, i.e., the fringe subtrees rooted at the children of the root of T, and g is a so-called toll function.

There are several recent articles on distributional results for additive tree parameters, see for example [17], [14], and [15].

It is easy to see that such an additive tree parameter can be computed by summing the toll function over all fringe subtrees, i.e., if $T^{(v)}$ denotes the fringe subtree rooted at the vertex v of T, then we have

$$G(T) = \sum_{v\in T} g(T^{(v)}).$$

In particular, the parameter is fully determined by specifying the toll function f.

Additive tree parameters play an important role in our analysis because our quantity of interest can be seen as such a parameter.

Proposition 2.1. *Let $r \geq 0$ be an integer and let $a_r(T)$ denote the number of removed vertices when applying the "cutting leaves" reduction r times for a given rooted tree T. Then $a_r(T)$ is an additive tree parameter, and the toll function belonging to $a_r(T)$ is given by*

$$g_r(T) = \begin{cases} 1 & \textit{if the height of } T \textit{ is less than } r, \\ 0 & \textit{else.} \end{cases} \tag{2.2}$$

In other words, if $\mathcal{T}_r$ denotes the family of rooted trees of height less than r, the toll function can be written in Iverson notation[1] *as $g_r(T) = [\![T \in \mathcal{T}_r]\!]$.*

[1] The Iverson notation, as popularized in [6], is defined as follows: $[\![\mathrm{expr}]\!]$ evaluates to 1 if expr is true, and to 0 otherwise.

Proof. It is easy to see that the number of removed vertices satisfies this additive property—the number of deleted nodes in some tree T is precisely the sum of all deleted nodes in the branches of T in case the root is not deleted. Otherwise, the sum has to be increased by one to account for the root node. Thus, the toll function determines whether or not the root node of T is deleted.

The fact that the root node is deleted if and only if the number of reductions r is greater than the height of the tree is easy to see; cf. Figure 1.

Basically, our strategy to analyze the quantity $a_r(T)$ for simply generated families of trees uses the recursive structure of (2.1) together with the structure of the family itself to derive a functional equation for a suitable bivariate generating function $A_r(x, u)$. In this context, the trees T in the family $\mathcal{T}$ are enumerated with respect to their size (corresponding to the variable x) and the value of the parameter $a_r(T)$ (corresponding to the variable u).

Throughout the remainder of this extended abstract, $\mathcal{T}$ denotes the family of trees under investigation (which will be the family of simply generated trees except in Section 4), and for all $r \in \mathbb{Z}_{\geq 1}$, $\mathcal{T}_r \subset \mathcal{T}$ denotes the class of trees of height less than r.

Furthermore, from now on, T_n denotes a random[2] tree of size n (i.e., a tree that consists of n vertices) from $\mathcal{T}$. This means that formally, the quantity we are interested in analyzing is the random variable $a_r(T_n)$ for large n.

3 Reducing Simply Generated Trees

3.1 Recursive Characterization Let us begin by recalling the definition of simply generated trees. A simply generated family of trees $\mathcal{T}$ can be defined by imposing a weight function on plane trees. For a sequence of nonnegative weights $(w_k)_{k\geq 0}$ (we will make the customary assumption that $w_0 = 1$ without loss of generality; cf. [13, Section 4]), one defines the weight of a plane tree T as the product 习惯的

$$w(T) := \prod_{j\geq 0} w_j^{N_j(T)},$$

where $N_j(T)$ is the number of vertices in T with precisely j children. The weight generating function

$$F(x) = \sum_{T\in\mathcal{T}} w(T) x^{|T|}, \tag{3.3}$$

where $|T|$ denotes the size of T (i.e., the number of vertices), is easily seen to satisfy a functional equation.

[2] The underlying probability distribution will always be clear from context.

By setting $\Phi(t) = \sum_{j\geq 0} w_j t^j$ and applying the symbolic method (see [5, Chapter I]) to decompose a simply generated tree as the root node with some simply generated trees attached, we have

$$F(x) = x\Phi(F(x)). \tag{3.4}$$

We define a probability measure on the set of all plane trees with n vertices by assigning a probability proportional to $w(T)$ to every tree T. In particular, this means that for a random simply generated tree of size n denoted by T_n, we have

$$\mathbb{P}(T_n = T) = \frac{w(T)}{Z_n} \tag{3.5}$$

where T is some plane tree with n vertices and the normalizing factor Z_n (the *partition function*) is given by

$$Z_n := \sum_{\substack{T\in\mathcal{T}\\ |T|=n}} w(T),$$

see [13, Section 2.3] for more details. Also, note that simply generated trees are a generalization of the well-known Galton–Watson trees conditioned on their size.

Remark 3.1. *Several important families of trees are covered by suitable choices of weights:*

- *plane trees are obtained from the weight sequence with $w_j = 1$ for all j,*
- *labelled trees correspond to weights given by $w_j = \frac{1}{j!}$,*
- *and d-ary trees (where every vertex has either d or no children) are obtained by setting $w_0 = w_d = 1$ and $w_j = 0$ for all other j.*

In the context of simply generated trees, it is natural to define the bivariate generating function $A_r(x, u)$ to be a weight generating function, i.e.,

$$A_r(x,u) = \sum_{T\in\mathcal{T}} w(T)x^{|T|}u^{a_r(T)}.$$

As the toll function of the additive tree parameter a_r depends on the combinatorial class $\mathcal{T}_r$ by Proposition 2.1, we need its weight generating function

$$F_r(x) = \sum_{T\in\mathcal{T}_r} w(T)x^{|T|}$$

to derive a functional equation for $A_r(x, u)$.

Clearly, $F_1(x) = x$, since there is only one rooted tree of height 0, which only consists of the root. Moreover, via the decomposition mentioned in the interpretation of (3.4), we have

$$F_r(x) = x\Phi(F_{r-1}(x)) \tag{3.6}$$

for every $r > 1$.

Now we are prepared to derive the aforementioned functional equation.

Proposition 3.1. *The bivariate weight generating function $A_r(x, u)$ satisfies the functional equation*

$$\boxed{A_r(x,u) = x\Phi(A_r(x,u)) + \Big(1 - \frac{1}{u}\Big)F_r(xu). \tag{3.7}}$$

Proof. By definition of $A_r(x, u)$, we have

$$\begin{aligned} A_r(x,u) &= \sum_{T\in\mathcal{T}} w(T)x^{|T|}u^{a_r(T)} \\ &= \sum_{T\in\mathcal{T}} w(T)x^{|T|}u^{a_r(T)-[\![T\in\mathcal{T}_r]\!]} \\ &\qquad + \sum_{T\in\mathcal{T}_r} w(T)x^{|T|}(u^{a_r(T)} - u^{a_r(T)-1}). \end{aligned}$$

For the first sum, we represent T by a root with k branches $T^1, \ldots, T^k$ for some $k \geq 0$ and use (2.2). For the second sum, we use the fact that $a_r(T) = |T|$ for $T \in \mathcal{T}_r$. This results in

$$\begin{aligned} A_r(x,u) &= \sum_{k\geq 0} w_k \sum_{T^1\in\mathcal{T}} \cdots \sum_{T^k\in\mathcal{T}} \Big(\prod_{j=1}^k w(T^j)\Big) \\ &\qquad\qquad \times x^{1+|T^1|+\cdots+|T^k|}u^{a_r(T^1)+\cdots+a_r(T^k)} \\ &\qquad + \Big(1-\frac{1}{u}\Big)\sum_{T\in\mathcal{T}_r} w(T)x^{|T|}u^{|T|} \\ &= x\sum_{k\geq 0} w_k\Big(\sum_{T\in\mathcal{T}} w(T)x^{|T|}u^{a_r(T)}\Big)^k \\ &\qquad + \Big(1-\frac{1}{u}\Big)\sum_{T\in\mathcal{T}_r} w(T)(xu)^{|T|} \\ &= x\Phi(A_r(x,u)) + \Big(1-\frac{1}{u}\Big)F_r(xu). \end{aligned}$$

Remark 3.2. *By definition, we have $A_r(x, 1) = F(x)$. To verify consistency, we observe that (3.7) reduces to (3.4) when $u = 1$.*

The functional equation (3.7) provides enough leverage to carry out a full asymptotic analysis of the behavior of $a_r(T)$ for simply generated trees.

3.2 Parameter Analysis Now we use the functional equation in Proposition 3.1 to determine mean and variance of $a_r(T_n)$, which are obtained from the partial derivatives of $A_r(x, u)$ with respect to u, evaluated at $u = 1$. To be more precise, if T_n denotes a random (with respect to the probability distribution determined by the given weight sequence as mentioned in Section 3.1

and particularly in (3.5)) simply generated tree of size n, then after normalization, the factorial moments

$$\mathbb{E}a_r(T_n)^{\underline{k}} := \mathbb{E}(a_r(T_n)(a_r(T_n)-1)\cdots(a_r(T_n)-k+1))$$

can be extracted as the coefficient of x^n in the partial derivative $\frac{\partial^k}{\partial u^k}A_r(x,u)\big|_{u=1}$. And from there, expectation and variance can be obtained in a straightforward way.

From this point on, we make some reasonable assumptions on the weight sequence $(w_k)_{k\geq 0}$:

- In addition to $w_0 = 1$, we assume that there is a $k > 1$ with $w_k > 0$ to avoid trivial cases.
- Furthermore, we require that if $R > 0$ is the radius of convergence of the weight generating function $\Phi(t) = \sum_{k\geq 0} w_k t^k$, there is a unique positive τ (the *fundamental constant*) with $0 < \tau < R$ such that $\Phi(\tau) - \tau\Phi'(\tau) = 0$.

The second constraint is to ensure that the singular behavior of $F(x)$ can be fully characterized (see, e.g., [5, Section VI.7]).

The following theorem presents the results of our investigation of the additive tree parameter $a_r(T_n)$.

主要
结论

THEOREM 3.1. *Let $r \in \mathbb{Z}_{\geq 1}$ be fixed and $\mathcal{T}$ be a simply generated family of trees with weight generating function Φ and fundamental constant τ, and set $\rho = \tau/\Phi(\tau)$. Furthermore, let $\mathcal{T}_r \subset \mathcal{T}$ denote the set of trees with height less than r and write $F_r(x)$ for the corresponding weight generating function.*

1. *If T_n denotes a random tree from $\mathcal{T}$ of size n, then for $n \to \infty$ the expected number of removed nodes when applying the "cutting leaves" procedure r times to T_n and the corresponding variance satisfy*

$$(3.8)\quad \mathbb{E}a_r(T_n) = \mu_r n + \frac{\rho\tau^2 F_r'(\rho) + 3\beta\tau F_r(\rho) - \alpha^2 F_r(\rho)}{2\tau^3} + O(n^{-1}),$$

and

$$(3.9)\quad \mathbb{V}a_r(T_n) = \sigma_r^2 n + O(1).$$

The constants μ_r and σ_r^2 are given by

$$(3.10)\quad \mu_r = \frac{F_r(\rho)}{\tau}$$

and

$$(3.11)\quad \sigma_r^2 = \frac{1}{2\tau^4}\big(4\rho\tau^3 F_r'(\rho) - 4\rho\tau^2 F_r(\rho)F_r'(\rho) + (2\tau^2-\alpha^2)F_r(\rho)^2 - 2\tau^3 F_r(\rho)\big),$$

respectively. The constants α and β are given by

$$(3.12)\quad \alpha = \sqrt{\frac{2\tau}{\rho\Phi''(\tau)}}, \qquad \beta = \frac{1}{\rho\Phi''(\tau)} - \frac{\tau\Phi'''(\tau)}{3\rho\Phi''(\tau)^2}.$$

2. *For $r \to \infty$ the constants μ_r and σ_r^2 behave as follows:*

$$\mu_r = 1 - \frac{2}{\rho\tau\Phi''(\tau)}r^{-1} + o(r^{-1})$$

and

$$\sigma_r^2 = \frac{1}{3\rho\tau\Phi''(\tau)} + o(1).$$

3. *Finally, if $r \geq 2$ or $\mathcal{T}$ is not a family of d-ary trees, then $a_r(T_n)$ is asymptotically normally distributed, meaning that for $x \in \mathbb{R}$ we have*

$$\mathbb{P}\Big(\frac{a_r(T_n) - \mu_r n}{\sqrt{\sigma_r^2 n}} \leq x\Big) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-t^2/2}\,dt + O(n^{-1/2}).$$

In case $r = 1$ and $\mathcal{T}$ is a family of d-ary trees, we have the deterministic relation $a_1(T_n) = \frac{n(d-1)+1}{d}$.

We will prove this theorem step by step in the following sections.

3.3 Asymptotic Expansions for Mean and Variance

REMARK 3.3. *For the sake of technical convenience, we are going to assume that $\Phi(t)$ is an aperiodic power series, meaning that the period p, i.e., the greatest common divisor of all indices j for which $w_j \neq 0$, is 1. This implies (see [5, Theorem VI.6]) that $F(x)$ has a unique square root singularity located at $\rho = \tau/\Phi(\tau)$, which makes some of our computations less tedious. However, all of our results also apply (mutatis mutandis) if this aperiodicity condition is not satisfied—with the restriction that then, $n-1$ has to be a multiple of the period p. One possible approach for this purpose is to perform the substitution $F(x) = xG(x^p)$ (as well as $F_r(x) = xG_r(x^p)$, etc.), so that the new function G has only one dominant singularity (of square root type, cf. (3.16) below) on the circle of convergence. Then singularity analysis applies in the same way as in the aperiodic case.*

We begin by proving Part 1 of Theorem 3.1.

Proof. First, we use implicit differentiation with respect to u on the functional equation (3.7) of Proposition 3.1

套路是比较熟悉的，先得到生成函数的方程，然后想得到数学期望就是求一阶（偏）导，
想得到方差就是二阶，然后应用 AC 中的定理

to find

$$\frac{\partial}{\partial u}A_r(x,u) = x\Phi'(A_r(x,u))\frac{\partial}{\partial u}A_r(x,u) + \frac{1}{u^2}F_r(xu) + x\Big(1-\frac{1}{u}\Big)F_r'(xu). \tag{3.13}$$

This results in

$$\frac{\partial}{\partial u}A_r(x,u)\Big|_{u=1} = \frac{F_r(x)}{1-x\Phi'(A_r(x,1))} = \frac{F_r(x)}{1-x\Phi'(F(x))} \tag{3.14}$$

where $A_r(x,1)=F(x)$ has been used in the last step. In the same vein, differentiating (3.13) a second time with respect to u and setting $u=1$ yields

$$\frac{\partial^2}{\partial u^2}A_r(x,u)\Big|_{u=1} = \frac{2(xF_r'(x)-F_r(x))}{1-x\Phi'(F(x))} + \frac{xF_r(x)^2\Phi''(F(x))}{(1-x\Phi'(F(x)))^3}. \tag{3.15}$$

Writing (3.4) in the form $\Phi(F(x)) = F(x)/x$ and taking its derivatives allows to eliminate all occurrences of $\Phi(F(x))$ and its derivatives from (3.14) and (3.15). This results in

$$\frac{\partial}{\partial u}A_r(x,u)\Big|_{u=1} = \frac{xF'(x)F_r(x)}{F(x)}$$

and

$$\frac{\partial^2}{\partial u^2}A_r(x,u)\Big|_{u=1} = \Big(\frac{2F_r(x)^2}{F(x)^2} - \frac{2F_r(x)}{F(x)} + \frac{2xF_r'(x)}{F(x)}\Big)xF'(x) + \frac{x^2F_r(x)^2F''(x)}{F(x)^2} - \frac{2x^2F_r(x)^2F'(x)^2}{F(x)^3}.$$

By the assumptions listed before Theorem 3.1, there is a positive real number τ that is smaller than the radius of convergence of Φ and satisfies the equation $\tau\Phi'(\tau) = \Phi(\tau)$. It is well known (see [5, Section VI.7]) that in this case, $F(x)$ has a square root singularity at $\rho = \tau/\Phi(\tau) = 1/\Phi'(\tau)$, with singular expansion

$$F(x) = \tau - \alpha\sqrt{1-x/\rho} + \beta(1-x/\rho) + O((1-x/\rho)^{3/2}). \tag{3.16}$$

Here, the coefficients α and β are given as specified in (3.12). The expressions for these coefficients are computed in our corresponding SageMath file (see end of Section 1). If more precise asymptotics are desired, further terms of the singular expansion can be computed easily.

Due to our aperiodicity assumption (see beginning of Section 3.3), ρ is the only singularity on F's circle of convergence (cf. [5, Theorem VI.6]), and the conditions of singularity analysis (see [4] or [5, Chapter VI], for example) are satisfied.

Next we note that F_r has greater radius of convergence than F. This follows from (3.6) by induction on r: it is clear for $r=1$, and if F_{r-1} is analytic at ρ, then so is F_r, since $|F_{r-1}(\rho)| < F(\rho) = \tau$ is smaller than the radius of convergence of Φ. So F_r has greater radius of convergence than F. This implies that F_r has a Taylor expansion around ρ:

$$F_r(x) = F_r(\rho) - \rho F_r'(\rho)(1-x/\rho) + O((1-x/\rho)^2). \tag{3.17}$$

Note that for the moment, we consider the function $F(x)$ (which does not depend on r) as a known function; we only expand $F_r(x)$ and its derivatives around ρ. By combining the previous expansion of $F(x)$ in (3.16) and the expansion of $F_r(x)$ in (3.17), we find that

$$\begin{aligned}\frac{\partial}{\partial u}A_r(x,u)\Big|_{u=1} &= \frac{xF'(x)F_r(x)}{F(x)}\\ &= \frac{F_r(\rho)}{\tau}\cdot(xF'(x))\\ &\quad + \frac{\rho\tau^2F_r'(\rho) + 3\beta\tau F_r(\rho) - \alpha^2F_r(\rho)}{2\tau^3}\cdot F(x)\\ &\quad + C_1 + C_2(1-x/\rho) + O((1-x/\rho)^{3/2})\end{aligned}$$

for certain real constants C_1 and C_2.

The nth coefficient of the derivative $[z^n]\frac{\partial}{\partial u}A_r(x,u)\big|_{u=1}$ can now be extracted by means of singularity analysis. Normalizing the result by dividing by $[z^n]A_r(x,1) = [z^n]F(x)$ (again extracted by means of singularity analysis; the corresponding expansion is given in (3.16)) yields an asymptotic expansion for $\mathbb{E}a_r(T_n)$. We find

$$\mathbb{E}a_r(T_n) = \frac{F_r(\rho)}{\tau}\cdot n + \frac{\rho\tau^2F_r'(\rho) + 3\beta\tau F_r(\rho) - \alpha^2F_r(\rho)}{2\tau^3} + O(n^{-1}). \tag{3.18}$$

Similarly, from

$$\frac{\partial^2}{\partial u^2}A_r(x,u)\Big|_{u=1} = \Big(\frac{2F_r(x)^2}{F(x)^2} - \frac{2F_r(x)}{F(x)} + \frac{2xF_r'(x)}{F(x)}\Big)xF'(x) + \frac{x^2F_r(x)^2F''(x)}{F(x)^2} - \frac{2x^2F_r(x)^2F'(x)^2}{F(x)^3}$$

这里的技巧好像类似数学分析里的不等式技巧

$$
\begin{aligned}
&= \left(\frac{F_r(\rho)}{\tau}\right)^2 \cdot (x^2F''(x) + xF'(x)) \\
&\quad + \frac{xF'(x)}{2\tau^4}\big(4\rho\tau^3F_r'(\rho) - 2\rho\tau^2F_r(\rho)F_r'(\rho) \\
&\qquad + (2\tau^2 + 6\beta\tau - 3\alpha^2)F_r(\rho)^2 - 4\tau^3F_r(\rho)\big) \\
&\quad + C_3 + O((1 - x/\rho)^{1/2})
\end{aligned}
$$

for some real constant C_3, we can use singularity analysis to find an asymptotic expansion for the second factorial moment $\mathbb{E}a_r(T_n)^{\underline{2}}$. Plugging the result and the expansion for the mean from (3.18) into the well-known identity

$$
\mathbb{V}a_r(T_n) = \mathbb{E}a_r(T_n)^{\underline{2}} + \mathbb{E}a_r(T_n) - (\mathbb{E}a_r(T_n))^2
$$

then yields

$$
\begin{aligned}
\mathbb{V}a_r(T_n) = \frac{n}{2\tau^4}\big(&4\rho\tau^3F_r'(\rho) - 4\rho\tau^2F_r(\rho)F_r'(\rho) \\
&+ (2\tau^2 - \alpha^2)F_r(\rho)^2 - 2\tau^3F_r(\rho)\big) + O(1).
\end{aligned}
$$

This proves Part 1 of Theorem 3.1.

3.4 Asymptotic Behavior for large r

Within the first part of Theorem 3.1 we obtained a precise characterization for the mean and the variance of the number of deleted vertices. However, it would be interesting to have more information on how these quantities behave for a very large number of iterated reductions. Furthermore, the constants μ_r and σ_r^2 of Theorem 3.1 depend on the generating function $F_r(x)$ and its derivatives evaluated at ρ. In order to better understand these constants, we consider their asymptotic behavior for $r \to \infty$.

We proceed by proving Part 2 of Theorem 3.1.

Proof. In order to obtain the behavior of μ_r and σ_r^2 for $r \to \infty$, we have to study the behavior of $c_r = F_r(\rho)$ and $d_r = F_r'(\rho)$ as $r \to \infty$. First, from (3.6) we get the recursion

$$
c_r = \rho\Phi(c_{r-1}).
$$

We note that c_r is increasing in r: For every j and $r \geq 0$, we have $[x^j]F_r(x) \leq [x^j]F_{r+1}(x) \leq [x^j]F(x)$ in view of the combinatorial interpretation of these generating functions. This implies that for every fixed positive real x, $F_r(x)$ is increasing in r. This, in turn, proves that c_r is increasing as well. Together with the fact that $F(\rho) < \infty$, this implies that c_r is a convergent sequence. Then, as the coefficients of $F_r(x)$ converge to those of $F(x)$ by the combinatorial interpretation of $F_r(x)$, we have $\lim_{r\to\infty} c_r = \lim_{r\to\infty} F_r(\rho) = F(\rho) = \tau$ by the assumptions listed before Theorem 3.1. Note that limit and summation can be exchanged due to the non-negativity of the coefficients and of ρ.

By Taylor expansion around τ, we obtain

$$
\begin{aligned}
\tau - c_r &= \rho\Phi(\tau) - \rho\Phi(c_{r-1}) \\
&= \rho\Phi'(\tau)(\tau - c_{r-1}) - \frac{\rho\Phi''(\tau)}{2}(\tau - c_{r-1})^2 \\
&\quad + O((\tau - c_{r-1})^3),
\end{aligned}
$$

and since $\rho\Phi'(\tau) = 1$, it follows that

$$
\frac{1}{\tau - c_r} = \frac{1}{\tau - c_{r-1}} + \frac{\rho\Phi''(\tau)}{2} + O(\tau - c_{r-1}).
$$

Now choose $K > 0$ such that

$$
\left|\frac{1}{\tau - c_r} - \frac{1}{\tau - c_{r-1}} - \frac{\rho\Phi''(\tau)}{2}\right| \leq K|\tau - c_{r-1}|
$$

holds for all $r \geq 0$. As $c_r \to \tau$ for $r \to \infty$, for any $\delta > 0$ we can find some $r_0 \geq 0$ such that we have $|\tau - c_{r-1}| < \delta$ for all $r \geq r_0$. From this, we find

$$
\begin{aligned}
&\left|\frac{1}{\tau - c_r} - \frac{1}{\tau - c_{r_0}} - (r - r_0)\frac{\rho\Phi''(\tau)}{2}\right| \\
&\qquad = \left|\sum_{j=r_0+1}^{r}\left(\frac{1}{\tau - c_j} - \frac{1}{\tau - c_{j-1}} - \frac{\rho\Phi''(\tau)}{2}\right)\right| \\
&\qquad \leq \sum_{j=r_0+1}^{r}\left|\frac{1}{\tau - c_j} - \frac{1}{\tau - c_{j-1}} - \frac{\rho\Phi''(\tau)}{2}\right| \\
&\qquad < (r - r_0)K\delta \\
&\qquad \leq rK\delta,
\end{aligned}
$$

such that we can conclude

$$
\frac{1}{\tau - c_r} = \frac{\rho\Phi''(\tau)}{2}r + o(r),
$$

and therefore

$$
\mu_r = \frac{c_r}{\tau} = 1 - \frac{2}{\rho\tau\Phi''(\tau)}r^{-1} + o(r^{-1}).
$$

Further terms could be derived by means of bootstrapping. Similarly, differentiating the identity $F_r(x) = x\Phi(F_{r-1}(x))$ gives us the recursion

$$
d_r = \rho\Phi'(c_{r-1})d_{r-1} + \frac{c_r}{\rho}.
$$

Since $\rho\Phi'(c_{r-1}) < \rho\Phi'(\tau) = 1$ and c_r is bounded, it follows from the recursion that $d_r = O(r)$. Now, we use Taylor expansion again to obtain

$$
\begin{aligned}
d_r &= \rho\big(\Phi'(\tau) - \Phi''(\tau)(\tau - c_{r-1}) \\
&\qquad + O((\tau - c_{r-1})^2)\big)d_{r-1} + \frac{c_r}{\rho} \\
&= \big(\rho\Phi'(\tau) - \rho\Phi''(\tau)(\tau - c_{r-1}) \\
&\qquad + O((\tau - c_{r-1})^2)\big)d_{r-1} + \frac{c_r}{\rho} \\
&= \Big(1 - \frac{2}{r} + o(r^{-1})\Big)d_{r-1} + \frac{\tau}{\rho} + o(1).
\end{aligned}
$$

Using the fact that $d_{r-1} = O(r)$, this can be rewritten as $r^2 d_r = (r-1)^2 d_{r-1} + \frac{\tau}{\rho} r^2 + o(r^2)$. For the sake of simplicity, set $e_r := r^2 d_r$. For any $\varepsilon > 0$ we can find an $r_0 \geq 0$ such that

$$\left| e_r - e_{r-1} - \frac{\tau}{\rho} r^2 \right| < \varepsilon r^2$$

holds for all $r \geq r_0$. Therefore, we find

$$\begin{aligned} &\left| e_r - e_{r_0} - \frac{\tau}{\rho}\Big(\frac{r(r+1)(2r+1)}{6} - \frac{r_0(r_0+1)(2r_0+1)}{6} \Big) \right| \\ &= \Big| \sum_{j=r_0+1}^{r} \Big(e_j - e_{j-1} - \frac{\tau}{\rho} j^2 \Big) \Big| \\ &\leq \sum_{j=r_0+1}^{r} \Big| e_j - e_{j-1} - \frac{\tau}{\rho} j^2 \Big| \\ &< \varepsilon \sum_{j=r_0+1}^{r} j^2 \leq \varepsilon \frac{r(r+1)(2r+1)}{6}, \end{aligned}$$

which gives us $e_r = \frac{\tau}{3\rho} r^3 + o(r^3)$ and allows us to conclude that

$$d_r = \frac{\tau}{3\rho} r + o(r).$$

Plugging the formulas for c_r and d_r into (3.10) and (3.11), we find that

$$\sigma_r^2 = \frac{1}{3\rho\tau\Phi''(\tau)} + o(1).$$

This concludes the proof of the second part of Theorem 3.1.

3.5 Asymptotic Normal Distribution As a consequence of our results in Section 3.4, we can also observe that for sufficiently large r, the constant σ_r^2 is strictly positive. This is necessary for a proof that $a_r(T_n)$ is asymptotically normally distributed in these cases. However, we can do even better: we can prove that $a_r(T_n)$ always admits a Gaussian limit law, except for an—in
病态的
some sense—pathological case.

Our strategy for proving a linear lower bound for the variance relies on the following lemma.

LEMMA 3.1. *Let $r \geq 2$ or $r = 1$, and suppose that $\mathcal{T}$ is not a family of d-ary trees. Then there exist two trees T^1 and $T^2 \in \mathcal{T}$ with $|T^1| = |T^2|$ such that $a_r(T^1) \neq a_r(T^2)$, and $a_r(T^1), a_r(T^2) < |T^1|$.*

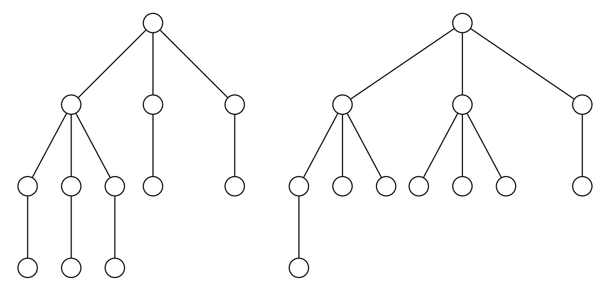
In other words, neither of the two trees is completely reduced after r steps, and the number of vertices removed after r steps differs between T^1 and T^2.

Note that the assumption in the case $r = 1$ cannot be removed: In a d-ary tree T of size n, the number of leaves is deterministically given by $\frac{n(d-1)+1}{d}$, which implies that $a_1(T) = \frac{n(d-1)+1}{d}$.

Let us now prove Lemma 3.1.

Proof. If $r = 1$ and $\mathcal{T}$ is not a d-ary family of trees, there have to be at least three different possibilities for the number of children, namely 0, d, and e. Let k_d and k_e be nonnegative integers. Then it is always possible to construct a tree with k_d and k_e inner nodes of out-degree d and e, respectively, as follows: We start at the root and add e children. Then we traverse (in breadth-first order) further $k_e - 1$ nodes and add e children each; then k_d nodes and add d children each, see Figure 2. The number of leaves is then equal to $k_d(d-1) + k_e(e-1) + 1$ by the handshaking lemma, and the total number of vertices is equal to $k_d d + k_e e + 1$.

We choose any such tree T^1 with $k_d > e$. Then T^2 is constructed by decreasing k_d by e and increasing k_e by d. This ensures that the total number of vertices remains the same, but the number of leaves does not. This is illustrated in Figure 2.



(a) $k_1 = 5$, $k_3 = 2$ (b) $k_1 = 2$, $k_3 = 3$

Figure 2: Construction of trees T^1 and T^2 with $d = 1$, $e = 3$

We now consider the case $r \geq 2$. Let d be a positive integer for which the weight w_d is positive (i.e., a node in a tree in $\mathcal{T}$ can have d children). We choose T^1 to be the complete d-ary tree of height r. A second d-ary tree T^2 is then constructed by arranging the same number of internal vertices as a path and by attaching suitably many leaves. The handshaking lemma then guarantees that both trees have the same size; but $a_r(T^1)$ is obviously larger than $a_r(T^2)$.

Now, with this result at hand we are able to prove Part 3 of Theorem 3.1.

Proof. Observe that as soon as we are able to prove that the variance of $a_r(T_n)$ is actually linear with respect to n, i.e., $\sigma_r^2 \neq 0$, all conditions of [2, Theorem 2.23] hold and are checked easily, thus proving our Gaussian limit law.

We choose two trees T^1 and T^2 with the properties as in Lemma 3.1.

It is well-known (see e.g. [1] or [14] for stronger results) that large trees (except for a negligible proportion) contain a linear (with respect to the size of the tree) number of copies of T^1 and T^2 as fringe subtrees. To be more precise, this means that there is a positive constant $c > 0$ such that the probability that a tree of size n contains at least cn copies of the patterns T^1 and T^2 is greater than 1/2.

Now, consider a large random tree T_n in $\mathcal{T}$ and replace all occurrences of T^1 and T^2 by marked vertices. We call the resulting tree (with the marked vertices) the shape S of the original tree T_n. Write $m(S)$ for the number of marked vertices in S. Note that our considerations from above can be written as $\mathbb{P}(m(S) \geq cn) \geq 1/2$ and yield the lower bound

$$\mathbb{E}m(S) \geq \frac{cn}{2} \tag{3.19}$$

for the expected number of marked vertices in a random shape S.

Given a shape S, the random variable C counting the number of occurrences of T^1 in the original (random) tree T_n follows a binomial distribution with size parameter $m(S)$ and probability $p = w(T^1)/(w(T^1) + w(T^2)) \in (0, 1)$ that only depends on the weights of the patterns. The number of occurrences of T^2 then equals $m(S) - C$.

Then we have

$$a_r(T_n) = Ca_r(T^1) + (m(S) - C)a_r(T^2) + A,$$

for some A which only depends on the shape S. Therefore, the conditional variance of $a_r(T_n)$ given the shape S of T_n can be written as

$$\begin{aligned} \mathbb{V}(a_r(T_n)|S) &= \mathbb{V}(m(S)a_r(T^2) + C(a_r(T^1) - a_r(T^2))|S) \\ &= (a_r(T^1) - a_r(T^2))^2\mathbb{V}(C|S) \\ &= (a_r(T^1) - a_r(T^2))^2 p(1-p)m(S). \end{aligned}$$

Applying the law of total variance then yields

$$\begin{aligned} \mathbb{V}a_r(T_n) &\geq \mathbb{E}(\mathbb{V}(a_r(T_n)|S)) \\ &= (a_r(T^1) - a_r(T^2))^2 p(1-p)\mathbb{E}m(S) \\ &\geq (a_r(T^1) - a_r(T^2))^2 p(1-p)\frac{c}{2}n, \end{aligned}$$

where the last inequality is justified by (3.19). As $a_r(T^1) \neq a_r(T^2)$ by construction, the variance of $a_r(T_n)$ has to be of linear order.

Finally, in order to prove that the speed of convergence is $O(n^{-1/2})$, we replace the formulation of Hwang's Quasi-Power Theorem without quantification of the speed of convergence (cf. [2, Theorem 2.22]) in the proof of [2, Theorem 2.23] with a quantified version (see [12] or [11] for a generalization to higher dimensions).

This completes our proof of Theorem 3.1.

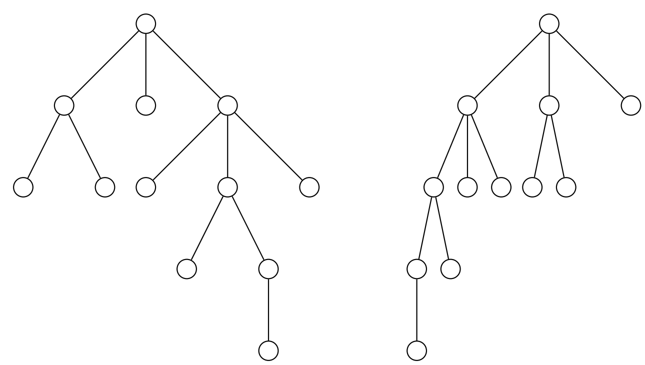
4 Outlook

Our approach for analyzing the "cutting leaves" reduction procedure on simply generated families of trees can be adapted to work for other families of trees as well. In this section, we describe two additional classes of rooted trees to which our approach is applicable and give qualitative results. Details on the analysis for these classes as well as quantitative results will be given in the full version of this extended abstract.

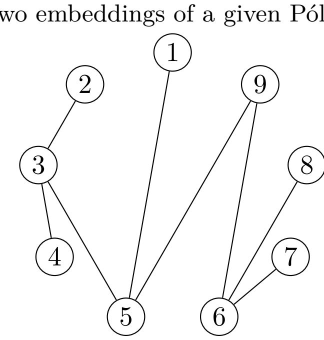
The two additional classes of trees are *Pólya trees* and *noncrossing trees*. Pólya trees are unlabeled rooted trees where the ordering of the children is not relevant. Noncrossing trees, on the other hand, are special labeled trees that satisfy two conditions:

- the root node has label 1,
- when arranging the vertices in a circle such that the labels are sequentially ordered, none of the edges of the tree are crossing.

Obviously, noncrossing trees have their name from the second property. Both classes of trees, Pólya trees as well as noncrossing trees, are illustrated in Figure 3.



(a) Two embeddings of a given Pólya tree



(b) A noncrossing tree of size 9

Figure 3: Pólya trees and noncrossing trees

The basic principle in the analysis of both of these tree classes is the same: we leverage the recursive nature of the respective family of trees to derive a functional equation for $A_r(x, u)$. From there, similar techniques as in Section 3.2 (i.e., implicit differentiation and propagation of the singular expansion of the basic generating function $F(x)$) can be used to obtain (arbitrarily precise) asymptotic expansions for the mean and the variance of the number of deleted nodes when cutting the tree r times.

Qualitatively, in both of these cases we can prove a theorem of the following nature.

THEOREM 4.1. *Let $r \in \mathbb{Z}_{\geq 1}$ be fixed and $\mathcal{T}$ be either the family of Pólya trees or the family of noncrossing trees. If T_n denotes a uniform random tree from $\mathcal{T}$ of size n, then for $n \to \infty$ the expected number of removed nodes when applying the "cutting leaves" procedure r times to T_n and the corresponding variance satisfy*

$$\mathbb{E}a_r(T_n) = \mu_r n + O(1), \quad \text{and} \quad \mathbb{V}a_r(T_n) = \sigma_r^2 n + O(1),$$

for explicitly known constants μ_r and σ_r^2. Furthermore, the number of deleted nodes $a_r(T_n)$ admits a Gaussian limit law.

Note that more precise asymptotic expansions for the mean and the variance (with explicitly known constants) can also be computed.

Another way to extend our results is to allow r to grow with n. Recall that the height of a tree T with n vertices is precisely the value of h for which $a_h(T) = n-1$, i.e., all vertices except for the root are removed in the first h steps. It is well known that the height is of asymptotic order $\sqrt{n}$ in simply generated trees, with a limiting theta distribution (see [3]), so we can expect $\sqrt{n}$ to be a natural threshold. Indeed, it seems plausible that the central limit theorem for $a_r(T_n)$ remains true provided that $r = o(\sqrt{n})$.

References

[1] David Aldous, *Asymptotic fringe distributions for general families of random trees*, Ann. Appl. Probab. **1** (1991), no. 2, 228–266.

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

[3] Philippe Flajolet, Zhicheng Gao, Andrew Odlyzko, and Bruce Richmond, *The distribution of heights of binary trees and other simple trees*, Combin. Probab. Comput. **2** (1993), no. 2, 145–156.

[4] Philippe Flajolet and Andrew Odlyzko, *Singularity analysis of generating functions*, SIAM J. Discrete Math. **3** (1990), 216–240.

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

[6] Ronald L. Graham, Donald E. Knuth, and Oren Patashnik, *Concrete mathematics. A foundation for computer science*, second ed., Addison-Wesley, 1994.

[7] Benjamin Hackl, Clemens Heuberger, and Daniel Krenn, *Asymptotic expansions in SageMath*, `http://trac.sagemath.org/17601`, 2015, module in SageMath 6.10.

[8] Benjamin Hackl, Clemens Heuberger, Sara Kropf, and Helmut Prodinger, *Fringe analysis of plane trees related to cutting and pruning*, Aequationes Math. **92** (2018), 311–353.

[9] Benjamin Hackl, Clemens Heuberger, and Helmut Prodinger, *Reductions of binary trees and lattice paths induced by the register function*, Theoret. Comput. Sci. **705** (2018), 31–57.

[10] Benjamin Hackl and Helmut Prodinger, *Growing and destroying Catalan–Stanley trees*, Discrete Math. Theor. Comput. Sci. **20** (2018), no. 1.

[11] Clemens Heuberger and Sara Kropf, *Higher dimensional quasi-power theorem and Berry–Esseen inequality*, Monatsh. Math. (2018).

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

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

[14] ———, *Asymptotic normality of fringe subtrees and additive functionals in conditioned Galton-Watson trees*, Random Structures Algorithms **48** (2016), no. 1, 57–101.

[15] Dimbinaina Ralaivaosaona and Stephan Wagner, *Additive functionals of d-ary increasing trees*, Preprint.

[16] The SageMath Developers, *SageMath Mathematics Software (Version 8.2)*, 2018, `http://www.sagemath.org`.

[17] Stephan Wagner, *Central limit theorems for additive tree parameters with small toll functions*, Combin. Probab. Comput. **24** (2015), 329–353.