SUBSET COUNTING IN TREES

STEPHAN G. WAGNER

ABSTRACT. Various enumeration problems for classes of simply generated families of trees have been the object of investigation in the past. We mention the enumeration of independent subsets, connected subsets or matchings for instance. The aim of this paper is to show how combinatorial problems of this type can also be solved for rooted trees and trees, which enables us to take better account of isomorphisms. As an example, we will determine the average number of independent vertex subsets of trees and binary rooted trees (every node has outdegree ≤ 2).

1. INTRODUCTION

Simply generated families in the sense of Meir and Moon [11] have been investigated in a lot of papers, such as [5, 12, 13]. A simply generated family is determined by a sequence $c_0 = 1, c_1, c_2, \ldots$ of weights. The weight of a rooted ordered tree is then given by

$$c(T) = \prod c_i^{N_i(T)},$$

where $N_i(T)$ is the number of vertices in T with exactly i children. One can define a generating function for the total weight of all trees on n vertices via

$$Y(x) = \sum_T c(T) x^{|T|}.$$

It is easy to see now that $Y(x)$ must satisfy a functional equation of the form $Y(x) = x\Phi(Y(x))$, where $\Phi(t) = \sum_{i=0}^{\infty} c_i t^i$. Special cases include ordinary rooted ordered trees ($\Phi(t) = \frac{1}{1-t}$) and rooted labelled trees ($\Phi(t) = e^t$). Because of the simple functional equation for $Y(x)$, enumeration problems of various kind can be solved by an appropriate study of generating functions. For example, the average number of independent or maximal independent subsets, connected subsets or matchings have been studied by various authors [4, 8, 9, 10, 12, 13, 18]. None of them investigates the average behavior for rooted trees or trees. The number of trees was first determined by Otter [15] in 1948 using methods which go back to Cayley [3] and Pólya [16]. It is well known (s. [7]) that the generating function $T(x)$ for the number of rooted trees satisfies the following functional equation:

$$T(x) = x \exp\left(\sum_{m=1}^{\infty} \frac{1}{m} T(x^m)\right). \tag{1}$$

The generating function $\tilde{T}(x)$ for the number of trees is connected to $T(x)$ via

$$\tilde{T}(x) = T(x) - \frac{1}{2}\Big(T^2(x) - T(x^2)\Big). \tag{2}$$

Thus, rooted trees do not belong to the class of simply generated families of trees. This also complicates the analysis of enumeration problems. However, it seems desirable to obtain information on the average behaviour of certain combinatorial indices for trees with consideration of isomorphisms. Apart from purely combinatorial interest, some of them even play a role in theoretical chemistry, such as the so-called *Merrifield-Simmons-index*, the number of independent vertex subsets of a

Date: 29th August 2005.
2000 *Mathematics Subject Classification.* 05A15, 05A16, 05C30, 05C69.
Key words and phrases. Tree, rooted tree, binary tree, subset counting, independent subset, generating functions.
The author is supported by the project S-8307-MAT of the Austrian Science Fund FWF.

graph (cf. [14]). As an example, we will determine the average behaviour of this index for trees and binary rooted trees. However, our method works for other enumeration problems, for example the number of matchings or connected subsets, just as well with the appropriate modifications.

We introduce some notation first. Let $|T|$ be the size (number of vertices) of a tree, and let $\sigma(T)$ denote the number of independent vertex subsets (i.e. subsets which contain no pair of adjacent vertices) of a tree T. Furthermore, for a rooted tree T, let $\sigma_1(T)$ and $\sigma_2(T)$ denote the number of independent vertex subsets containing resp. not containing the root. The value $\sigma(T)$ was introduced by Prodinger and Tichy [17], who used the name "Fibonacci number" of a tree for it, since the number of independent vertex subsets of a single path P_n with n vertices is exactly the Fibonacci number F_{n+2}. Among other things, they were able to prove that

$$F_{n+2} = \sigma(P_n) \leq \sigma(T) \leq \sigma(S_n) = 2^{n-1} + 1 \tag{3}$$

for all trees T with n vertices, where S_n denotes the star tree.

Furthermore, for a group G, let $Z(G)$ be the cycle index of G, written as a polynomial in $s_1, s_2, \ldots$. For a function $f(x)$, let $Z(G, f(x))$ be the cycle index of G with $f(x^k)$ in place of s_k (cf. [7]). In this way, equation (1) can be written as

$$T(x) = x\left(1 + \sum_{m=1}^{\infty} Z(S_m, T(x))\right),$$

since it is a well-known identity that

$$1 + \sum_{m=1}^{\infty} Z(S_m, f(x)) = \exp\left(\sum_{m=1}^{\infty} \frac{1}{m} f(x^n)\right).$$

For the generating function $T^{(2)}(x)$ of rooted trees with maximal outdegree ≤ 2, we obtain

$$\begin{aligned} T^{(2)}(x) &= x\left(1 + Z(S_1, T^{(2)}(x)) + Z(S_2, T^{(2)}(x))\right) \\ &= x\left(1 + T^{(2)}(x) + \frac{1}{2}\left(T^{(2)}(x) + T^{(2)}(x^2)\right)\right). \end{aligned} \tag{4}$$

2. The average number of independent subsets of a rooted tree

For a simply generated family of trees as defined in the introduction, it is not difficult to determine functional equations for the generating functions $S_1(x)$ and $S_2(x)$ of σ_1 and σ_2. In fact, if $T_1, \ldots, T_r$ are the branches of a rooted tree T, it is easy to see that

$$\sigma_1(T) = \prod_{i=1}^{r} \sigma_2(T_i)$$

and

$$\sigma_2(T) = \prod_{i=1}^{r} (\sigma_1(T_i) + \sigma_2(T_i)),$$

so that we immediately obtain

$$S_1(x) = x\Phi(S_2(x)),\ S_2(x) = x\Phi(S_1(x) + S_2(x)).$$

For rooted trees, things are a little more difficult. Note that terms of type $T(x^k)$ appear in equation (1). These belong to k-tuples of isomorphic rooted trees among the branches. In the equations for σ_1 and σ_2, these give a contribution of the form $\sigma_1(T_i)^k$ resp. $(\sigma_1(T_i) + \sigma_2(T_i))^k$. Therefore, it is necessary to introduce some more generating functions of the form

$$S_{k,l}(x) = \sum_T \sigma_1(T)^k \sigma_2(T)^l x^{|T|},$$

where the sum is over all rooted trees T. Now, it is not difficult to see that

$$S_1(x) := S_{1,0}(x) = x \exp\left(\sum_{m=1}^{\infty} \frac{1}{m} S_{0,m}(x^m)\right) \tag{5}$$

and

$$S_2(x) := S_{0,1}(x) = x \exp\left(\sum_{m=1}^{\infty} \frac{1}{m} \sum_{k=0}^{m} \binom{m}{k} S_{k,m-k}(x^m)\right). \tag{6}$$

Observe in the latter equation that

$$\sum_{k=0}^{m} \binom{m}{k} S_{k,m-k}(x)$$

is, in fact, a generating function for $\sigma(T)^m = (\sigma_1(T) + \sigma_2(T))^m$. In order to find the asymptotic behavior of the average values of σ_1 and σ_2, we have to determine the dominating singularity of S_1 and S_2. For this purpose, we employ the same trick that is used in the asymptotic calculation of the number of trees (in fact, we will almost directly follow the proof of Otter's tree-counting theorem given in [7]): we observe that only the summands corresponding to $m = 1$ in the functional equations are not holomorphic around the singularity. To prove this, we need an a-priori estimate.

Let $s_{n,1}$ and $s_{n,2}$ be the coefficients of S_1 and S_2. Then, we have $s_{n-1,2} \leq s_{n,1} \leq s_{n,2}$. These relations follow easily from the recurrences, but can also be proved by a combinatorial argument: for the former inequality, note that a rooted tree T with a single branch T_1 satisfies $\sigma_1(T) = \sigma_2(T_1)$; for the latter, note that removing the root from an independent subset containing the root always results in another independent set.

Therefore, S_1 and S_2 have a common radius of convergence ρ_S; as the coefficients of S_1 and S_2 are positive, ρ_S is a singularity of both of them. Let us denote the radius of convergence of T by ρ. It is known (cf. [7]) that $\rho \approx 0.338322 < \frac{1}{2}$. Now, define $C_{m,n} := \sum_{|T|=n} \sigma(T)^m$. From estimate (3), we obtain $\sigma(T) \leq 2^{|T|}$ and thus

$$C_{m,n} \leq C_{1,n} 2^{(m-1)n} \ll \rho_S^{(-1-\epsilon)n} 2^{(m-1)n}$$

for any $\epsilon > 0$. On the other hand, $\frac{\rho}{2} \leq \rho_S \leq \rho < \frac{1}{2}$.

Now, we are ready to prove two auxiliary lemmas:

Lemma 1. *The series*

$$\sum_{m=2}^{\infty} \frac{1}{m} S_{0,m}(x^m)$$

and

$$\sum_{m=2}^{\infty} \frac{1}{m} \sum_{k=0}^{m} \binom{m}{k} S_{k,m-k}(x^m)$$

define analytic functions within a circle of radius $\eta_S > \rho_S$.

Proof. We have to proof that the convergence radius of both series is larger than ρ_S. In fact, this is only necessary for the second series, since the first is a partial sum of the second and all coefficients are positive. Now, let $\eta \in (\rho_S, \sqrt{\frac{\rho_S}{2}})$. Since $\rho_S < \frac{1}{2}$, this interval is nonempty and $2\eta < \sqrt{2\rho_S} < 1$. Furthermore, choose $\epsilon > 0$ in such a way that $\alpha = 2\eta^2 \rho_S^{-1-\epsilon} < 1$. There exists some constant

$A > 0$ such that $C_{1,n} \le A\rho_S^{(-1-\epsilon)n}$ for all n. Therefore, we have

$$\begin{aligned}\sum_{m=2}^{\infty}\frac{1}{m}\sum_{k=0}^{m}\binom{m}{k}S_{k,m-k}(\eta^m) &= \sum_{m=2}^{\infty}\frac{1}{m}\sum_{n=1}^{\infty}C_{m,n}\eta^{mn} \le \sum_{m=2}^{\infty}\frac{1}{m}\sum_{n=1}^{\infty}C_{1,n}2^{(m-1)n}\eta^{mn} \\ &\le \sum_{m=2}^{\infty}\frac{1}{m}\sum_{n=1}^{\infty}A\rho_S^{(-1-\epsilon)n}2^{(m-1)n}\eta^{mn} = \sum_{m=2}^{\infty}\frac{A}{m}\frac{\rho_S^{-1-\epsilon}2^{m-1}\eta^m}{1-\rho_S^{-1-\epsilon}2^{m-1}\eta^m} \\ &\le \sum_{m=2}^{\infty}\frac{A}{m}\rho_S^{-1-\epsilon}2^{m-1}\eta^m\frac{1}{1-2\rho_S^{-1-\epsilon}\eta^2} \le \frac{A\rho_S^{-1-\epsilon}}{4(1-2\rho_S^{-1-\epsilon}\eta^2)}\sum_{m=2}^{\infty}(2\eta)^m \\ &= \frac{A\rho_S^{-1-\epsilon}\eta^2}{(1-2\rho_S^{-1-\epsilon}\eta^2)(1-2\eta)} < \infty.\end{aligned}$$

Hence, the series converges (absoutely, since all summands are positive) for every $\eta < \sqrt{\frac{\rho_S}{2}}$, which means that its radius of convergence is $\eta_S \ge \sqrt{\frac{\rho_S}{2}} > \rho_S$. So it represents an analytic function within a circle of radius $\eta_S > \rho_S$ around the origin. ■

Lemma 2. *The limits $\lim_{x\to\rho_S-} S_1(x)$ and $\lim_{x\to\rho_S-} S_2(x)$ exist, and the power series for S_1 and S_2 converge at ρ_S (to the respective limits).*

Proof. Note that, for $0 \le x < \rho_S$, we have $S_1(x) \le S_2(x)$ and

$$\log\left(\frac{S_2(x)}{x}\right) = \sum_{m=1}^{\infty}\frac{1}{m}\sum_{k=0}^{m}\binom{m}{k}S_{k,m-k}(x^m) \ge S_{0,1}(x) = S_2(x).$$

Thus, it follows that

$$\frac{S_2(x)/x}{\log(S_2(x)/x)} \le \frac{1}{x},$$

which means that $S_2(x)$ (and thus $S_1(x)$) must be bounded on the interval $(0, \rho_S)$. Since $S_1(x)$ and $S_2(x)$ are monotonous functions on this interval, the left-hand limits must exist. It follows easily that the power series converge at ρ_S. ■

Next, we investigate the values of $S_1(x)$ and $S_2(x)$ at $x = \rho_S$:

Lemma 3. *ρ_S is the only singularity of S_1 and S_2 on their circle of convergence. The values $s_1 = S_1(\rho_S)$ and $s_2 = S_2(\rho_S)$ satisfy the equation*

$$s_2(1+s_1) = 1. \tag{7}$$

Proof. We write the functional equations for $S_1(x)$ and $S_2(x)$ in the following form:

$$\begin{aligned}F_1(S_1(x), S_2(x), x) &= x\exp(S_2(x) + R_1(x)) - S_1(x) = 0,\\ F_2(S_1(x), S_2(x), x) &= x\exp(S_1(x) + S_2(x) + R_2(x)) - S_2(x),\end{aligned}$$

where $R_1(x)$ and $R_2(x)$ are abbreviations for $\sum_{m=2}^{\infty}\frac{1}{m}S_{0,m}(x^m)$ and $\sum_{m=2}^{\infty}\frac{1}{m}\sum_{k=0}^{m}\binom{m}{k}S_{k,m-k}(x^m)$ respectively. We already know that R_1 and R_2 are analytic within a circle of radius $\eta_S > \rho_S$. The Jacobian determinant of these equations has to vanish at a singularity. Otherwise, by the implicit function theorem, they would have a unique analytic solution in a certain neighborhood. Therefore, we calculate the Jacobian matrix of $F_1(y_1, y_2, x)$ and $F_2(y_1, y_2, x)$:

$$\frac{\partial F}{\partial y} = \begin{pmatrix} -1 & F_1(y_1,y_2,x)+y_1 \\ F_2(y_1,y_2,x)+y_2 & F_2(y_1,y_2,x)+y_2-1\end{pmatrix} = \begin{pmatrix} -1 & y_1 \\ y_2 & y_2-1\end{pmatrix},$$

since both F_1 and F_2 must vanish. The determinant is thus given by

$$\left|\frac{\partial F}{\partial y}\right| = 1 - y_2 - y_1y_2,$$

which means that equation (7) must be satisfied. Now let $\xi \ne \rho_S$ be another point on the circle of convergence. Then, since all coefficients of S_1 and S_2 are positive real numbers, we have $|S_1(\xi)| < s_1$ and $|S_2(\xi)| < s_2$, so the equation $1 - S_2(\xi) - S_1(\xi)S_2(\xi)$ can clearly not be satisfied. ■

Therefore, we may make use of the following well-known theorem (cf. [1, 2, 7]):

Theorem 4. *Let $F(x,y)$ be analytic in each variable seperately in some neighborhood of (x_0,y_0) and suppose that the following conditions are satisfied:*

(1) $F(x_0,y_0)=0$,
(2) $y=f(x)$ *is analytic in* $|x|<|x_0|$ *and* x_0 *is the unique singularity on the circle of covergence,*
(3) *if* $f(x)=\sum_{n=0}^{\infty} f_n x^n$ *is the expansion of* f *at the origin, then* $y_0=\sum_{n=0}^{\infty} f_n x_0^n$,
(4) $F(x,f(x))=0$ *for* $|x|<|x_0|$,
(5) $\frac{\partial F}{\partial y}(x_0,y_0)=0$,
(6) $\frac{\partial^2 F}{\partial y^2}(x_0,y_0)\neq 0$.

Then $f(x)$ *may be expanded about* x_0*:*

$$f(x)=f(x_0)+\sum_{k=1}^{\infty} a_k (x_0-x)^{k/2},$$

and if $a_1\neq 0$,

$$f_n\sim \frac{-a_1}{2\sqrt{\pi}} x_0^{-n+1/2} n^{-3/2}.$$

If $a_1=0$ *and* $a_3\neq 0$,

$$f_n\sim \frac{3a_3}{4\sqrt{\pi}} x_0^{-n+3/2} n^{-5/2}.$$

Note that $S(x)=S_1(x)+S_2(x)$ satisfies the equation

$$S(x)=x\exp(S(x)+R_2(x))+x\exp(x\exp(S(x)+R_2(x))+R_1(x)),$$

so the conditions of the theorem are satisfied by the preliminary lemmas with $f(x)=S(x)$ and

$$F(x,y)=x\exp(y+R_2(x))+x\exp(x\exp(y+R_2(x))+R_1(x))-y.$$

They are also satisfied for $f(x)=S_2(x)$ and

$$F(x,y)=x\exp(x\exp(y+R_1(x))+y+R_2(x))-y,$$

so S_1,S_2 and S may be expanded around ρ_S in the way that is given by the theorem. We only have to care about the values of the implied constants and their calculation. First of all, ρ_S is uniquely defined by the equations

$$\begin{aligned} s_1 &= \rho_S\exp(s_2+R_1(\rho_S)),\\ s_2 &= \rho_S\exp(s_1+s_2+R_2(\rho_S)),\\ 1 &= s_2(s_1+1). \end{aligned} \tag{8}$$

Note that $R_1(x)$ and $R_2(x)$ are convergent series within a circle of radius $\eta_S>\rho_S$. Therefore, if we calculate the coefficients of R_1 and R_2 up to some power x^N, we obtain estimates $\overline{R}_1(x)$ and $\overline{R}_2(x)$ which can be uniformly bounded within a circle of radius $\eta_S-\epsilon$. Solving the system with $\overline{R}_i$ instead of R_i thus gives estimates for s_1, s_2 and ρ_S.
The error can even be quantified in the following way: clearly, we have $C_{1,n}\leq 2^n t_n$, where t_n is the number of rooted trees of size n. This shows, following the estimates of Lemma 1, that the error can be uniformly and explicitly bounded within the circle of radius $\frac{\sqrt{\rho}}{2}-\epsilon$. On the other hand, by the left-hand estimate in (3), we have $\rho_S\leq\frac{\rho(\sqrt{5}-1)}{2}<\frac{\sqrt{\rho}}{2}$, which means that the error can be estimated explicitly. Numerical computation shows that $\rho_S\approx 0.2020447686$, $s_1\approx 0.4202770330$ and $s_2\approx 0.7040879890$. Computational details will be discussed in section 4. Now, write

$$\begin{aligned} S_1(x) &= s_1-b_1\sqrt{\rho_S-x}+\dots,\\ S_2(x) &= s_2-b_2\sqrt{\rho_S-x}+\dots,\\ S(x) &= s-b\sqrt{\rho_S-x}+\dots. \end{aligned}$$

To determine b_1 and b_2 (and thus $b = b_1 + b_2$), we note first that

$$S_1'(x)(1 - S_2(x) - S_1(x)S_2(x)) = \frac{b_1}{2}(s_1b_2 + s_2b_1 + b_2) + O((\rho_S - x)^{1/2})$$

and

$$S_2'(x)(1 - S_2(x) - S_1(x)S_2(x)) = \frac{b_2}{2}(s_1b_2 + s_2b_1 + b_2) + O((\rho_S - x)^{1/2}),$$

so we have

$$\begin{aligned} \frac{b_1}{2}(s_1b_2 + s_2b_1 + b_2) &= \lim_{x \to \rho_S} S_1'(x)(1 - S_2(x) - S_1(x)S_2(x)) =: c_1, \\ \frac{b_2}{2}(s_1b_2 + s_2b_1 + b_2) &= \lim_{x \to \rho_S} S_2'(x)(1 - S_2(x) - S_1(x)S_2(x)) =: c_2. \end{aligned} \tag{9}$$

The values on the right can be calculated by differentiating the functional equations for S_1 and S_2 first:

$$S_1'(x) = \frac{S_1(x)}{x} + S_1(x)(S_2'(x) + R_1'(x))$$

and

$$S_2'(x) = \frac{S_2(x)}{x} + S_2(x)(S_1'(x) + S_2'(x) + R_2'(x)).$$

Solving this system for $S_1'(x)$ and $S_2'(x)$ yields

$$S_1'(x)(1 - S_2(x) - S_1(x)S_2(x)) = \frac{S_1(x)}{x} + S_1(x)R_1'(x) + S_1(x)S_2(x)(R_2'(x) - R_1'(x))$$

and

$$S_2'(x)(1 - S_2(x) - S_1(x)S_2(x)) = \frac{S_2(x)(1 + S_1(x))}{x} + S_1(x)S_2(x)R_1'(x) + S_2(x)R_2'(x).$$

Therefore,

$$c_1 = \frac{s_1}{\rho_S} + s_1 \sum_{m=2}^{\infty} S_{0,m}'(\rho_S^m)\rho_S^{m-1} + s_1s_2 \sum_{m=2}^{\infty} \sum_{k=1}^{m} \binom{m}{k} S_{k,m-k}'(\rho_S^m)\rho_S^{m-1}$$

and

$$c_2 = \frac{1}{\rho_S} + s_1s_2 \sum_{m=2}^{\infty} S_{0,m}'(\rho_S^m)\rho_S^{m-1} + s_2 \sum_{m=2}^{\infty} \sum_{k=0}^{m} \binom{m}{k} S_{k,m-k}'(\rho_S^m)\rho_S^{m-1},$$

which can be calculated numerically. Furthermore, solving the system (9) for b_1 and b_2 gives us

$$b_1 = \frac{\sqrt{2}c_1}{\sqrt{s_2c_1 + c_2 + s_1c_2}}, \; b_2 = \frac{\sqrt{2}c_2}{\sqrt{s_2c_1 + c_2 + s_1c_2}}$$

and thus

$$b = \frac{\sqrt{2}(c_1 + c_2)}{\sqrt{s_2c_1 + c_2 + s_1c_2}}. \tag{10}$$

Numerical calculations show that $b \approx 3.8130254771$. Noting that the number t_n of rooted trees of size n satisfies $t_n \sim A \cdot n^{-3/2}\rho^{-n}$ with $A \approx 0.4399240126$, we have obtained the following theorem:

Theorem 5. *The average number of independent vertex subsets in a rooted tree of size n is given by*

$$\mathrm{av}_n \sim (1.0990334536) \cdot (1.6744895662)^n.$$

3. The average number of independent subsets of a tree

Now, having established the asymptotics for rooted trees, we are also able to give them for trees. We will make use of Otter's theorem [15] which states that the number of different representations of a tree as a rooted tree equals 1 plus the number of representations as a pair of two unequal rooted trees (the order being irrelevant), with their roots joined by an edge (see also [7]). It is easy to see that, if two rooted trees T_1, T_2 are joined by an edge connecting their root, the resulting tree T has a total number of

$$\sigma(T) = \sigma_1(T_1)\sigma_2(T_2) + \sigma_2(T_1)\sigma_1(T_2) + \sigma_2(T_1)\sigma_2(T_2)$$

independent vertices. Thus, if we denote the generating function which counts independent subsets in all trees instead of rooted trees by $\tilde{S}$, we have

$$\tilde{S}(x) = S(x) - \frac{1}{2}\left(2S_1(x)S_2(x) + S_2(x)^2 - 2S_{1,1}(x^2) - S_{0,2}(x^2)\right). \tag{11}$$

$S_{1,1}(x^2)$ and $S_{0,2}(x^2)$ are holomorphic around ρ_S. Thus, we only have to determine the expansion of the remaining terms around ρ_S. Let

$$\tilde{S}(x) = a_0 - a_1\sqrt{\rho_S - x} + a_2(\rho_S - x) + a_3(\rho_S - x)^{3/2} + \dots.$$

We know that $s_2(s_1+1) = 1$. Furthermore, from the equation

$$S_1(x) = x\exp(S_2(x) + R_1(x)),$$

we obtain $b_1 = s_1 b_2$. Inserting the expansions of S_1 and S_2 in (11) and using these relations shows that $a_1 = 0$. To determine a_3, we differentiate twice:

$$\tilde{S}''(x) = \frac{3a_3}{4}(\rho_S - x)^{-1/2} + \dots.$$

On the other hand, we differentiate the functional equations for S_1' and S_2':

$$S_1''(x) = \frac{S_1'(x)}{x} - \frac{S_1(x)}{x^2} + S_1'(x)(S_2'(x) + R_1'(x)) + S_1(x)(S_2''(x) + R_1''(x))$$

and

$$S_2''(x) = \frac{S_2'(x)}{x} - \frac{S_2(x)}{x^2} + S_2'(x)(S_1'(x) + S_2'(x) + R_2'(x)) + S_2(x)(S_1''(x) + S_2''(x) + R_2''(x)).$$

We solve this system for S_1'' and S_2'' and insert it in

$$\tilde{S}''(x) = S_1''(x) + S_2''(x) - S_1''(x)S_2(x) - 2S_1'(x)S_2'(x) - S_1(x)S_2''(x) - S_2'(x)^2 - S_2(x)S_2''(x) + R''(x)$$

together with the expressions for S_1' and S_2'. Note that $R(x) = S_{1,1}(x^2) + \frac{1}{2}S_{0,2}(x^2)$ is holomorphic within the circle of radius $\eta_S > \rho_S$. Then, we use the expansions of S_1 and S_2 together with the relations for s_1, s_2, b_1, b_2 to obtain the final expression for a_3:

$$\frac{3a_3}{4} = \sqrt{\frac{c_2^3}{2s_2(1 + s_2 - s_2^2)}} \approx 11.7914747833.$$

This gives us the asymptotic behavior of the coefficients of $\tilde{S}$ and, together with the asymptotic formula for the number $\tilde{t}_n$ of trees of size n, which is $\tilde{t}_n \sim B \cdot n^{-5/2}\rho^{-n}$ with $B \approx 0.5349496061$, we have established the following theorem:

Theorem 6. *The average number of independent vertex subsets in a tree of size n is given by*

$$\tilde{\mathrm{av}}_n \sim (1.1294102715) \cdot (1.6744895662)^n.$$

Thus, interestingly, a tree contains more independent sets on average than a rooted tree.

4. Efficient computation of the auxiliary functions and numerical values

In the approximate solution of the system (8), it was necessary to compute a sufficient number of coefficients of the auxiliary functions $S_{k,l}$. For this purpose, it is possible, of course, to compute the number of independent subsets explicitly for all rooted trees of size $n \leq N$. However, this brute-force method is highly inefficient, so it is desirable to have a better method at hand. It is quite simple to achieve this: we can deduce functional equations for $S_{k,l}$ in the same manner as we did for $S_1 = S_{1,0}$ and $S_2 = S_{0,1}$. These are given by the general formula

$$S_{k,l}(x) = x \exp\left(\sum_{m=1}^{\infty} \frac{1}{m} \sum_{r=0}^{ml} \binom{ml}{r} S_{ml-r,mk+r}(x^m)\right), \tag{12}$$

which enables us to compute the coefficients of $S_{k,l}$ in a simple recursive manner. We give the initial values of $S_{1,0}$, $S_{0,1}$ and $S_{2,0}$ for instance:

$$S_{1,0}(x) = x + x^2 + 3x^3 + 10x^4 + 38x^5 + 143x^6 + 577x^7 + 2325x^8 + 9697x^9 + 40853x^{10} + \dots,$$
$$S_{0,1}(x) = x + 2x^2 + 7x^3 + 24x^4 + 91x^5 + 341x^6 + 1370x^7 + 5504x^8 + 22914x^9 + 96457x^{10} + \dots,$$
$$S_{2,0}(x) = x + x^2 + 5x^3 + 30x^4 + 196x^5 + 1267x^6 + 8615x^7 + 58613x^8 + 411209x^9 + 2909597x^{10} + \dots.$$

Note that the functional equation can also be used to calculate higher moments of the number of independent subsets of a random tree. We give some numerical instances of the average values for rooted trees resp. trees in the following table:

n	av_n	$\tilde{\mathrm{av}}_n$	n	av_n	$\tilde{\mathrm{av}}_n$
1	2	2	8	68.08	70.83
2	3	3	9	114.02	119.09
3	5	5	10	190.97	199.02
4	8.5	8.5	15	2512.81	2608.75
5	14.33	14.67	20	33063.90	34210.51
6	24.2	24.83	50	$1.719535 \cdot 10^{11}$	$1.771075 \cdot 10^{11}$
7	40.56	42.09	100	$2.687782 \cdot 10^{22}$	$2.765055 \cdot 10^{22}$

Table 1. Some values of av_n and $\tilde{\mathrm{av}}_n$.

5. Independent subsets in a degree-restricted tree

It is clear that the methods we established in section 2 are easily generalized to other classes of trees or tree-like structures. As an example, we will determine the asymptotic average number of independent subsets in binary rooted trees (maximal outdegree ≤ 2, cf. [15, 7]). The functional equation for $T^{(2)}$, the generating function for the number of such trees, has already been given in the introduction. Next, we define $S_{k,l}^{(2)}$ in the same manner as in section 2. The functional equation

$$S_{k,l}^{(2)}(x) = x\left(1 + \sum_{r=0}^{l}\binom{l}{r} S_{l-r,k+r}^{(2)}(x) + \frac{1}{2}\left(\sum_{r=0}^{l}\binom{l}{r} S_{l-r,k+r}^{(2)}(x)\right)^2 + \frac{1}{2}\left(\sum_{r=0}^{2l}\binom{2l}{r} S_{2l-r,2k+r}^{(2)}(x^2)\right)\right)$$

follows at once in the same way as for rooted trees. In particular, we have

$$\begin{aligned} S_1^{(2)}(x) &:= S_{1,0}^{(2)}(x) = x\left(1 + S_2^{(2)}(x) + \frac{1}{2}S_2^{(2)}(x)^2 + \frac{1}{2}S_{0,2}^{(2)}(x^2)\right), \\ S_2^{(2)}(x) &:= S_{0,1}^{(2)}(x) = x\Bigg(1 + S_1^{(2)}(x) + S_2^{(2)}(x) + \frac{1}{2}\left(S_1^{(2)}(x) + S_2^{(2)}(x)\right)^2 \\ &\qquad + \frac{1}{2}\left(S_{0,2}^{(2)}(x^2) + 2S_{1,1}^{(2)}(x^2) + S_{2,0}^{(2)}(x^2)\right)\Bigg). \end{aligned} \tag{13}$$

Analogously to Theorem 5, we achieve the following result:

Theorem 7. *The average number of independent vertex subsets in a rooted tree of size n with maximal outdegree ≤ 2 is given by*

$$\mathrm{av}_n^{(2)} \sim (1.1311298442) \cdot (1.6425223181)^n.$$

It is not surprising that the average number of independent subsets decreases by the degree restriction. Rooted trees with restricted outdegrees are typically more "path-like", so – in view of inequality (3) – the number of independent subsets is closer to the minimum. Again, we give some numerical values in the following table:

n	$\mathrm{av}_n^{(2)}$	n	$\mathrm{av}_n^{(2)}$
1	2	8	60.04
2	3	9	98.55
3	5	10	161.91
4	8.33	15	1934.40
5	13.5	20	23121.26
6	22.27	50	$6.748132 \cdot 10^{10}$
7	36.67	100	$4.024331 \cdot 10^{21}$

Table 2. Some values of $\mathrm{av}_n^{(2)}$.

References

1. Edward A. Bender, *Asymptotic methods in enumeration*, SIAM Rev. **16** (1974), 485–515.
2. E. Rodney Canfield, *Remarks on an asymptotic method in combinatorics*, J. Combin. Theory Ser. A **37** (1984), no. 3, 348–352.
3. Arthur Cayley, *A theorem on trees*, Quart. J. Math. Oxford Ser. **23** (1889), 376–378.
4. Michael Drmota and Peter Kirschenhofer, *On generalized independent subsets of trees*, Random Structures Algorithms **2** (1991), no. 2, 187–208.
5. R. C. Entringer, A. Meir, J. W. Moon, and L. A. Székely, *The Wiener index of trees from certain families*, Australas. J. Combin. **10** (1994), 211–224.
6. I. P. Goulden and D. M. Jackson, *Combinatorial enumeration*, A Wiley-Interscience Publication, John Wiley & Sons Inc., New York, 1983, With a foreword by Gian-Carlo Rota, Wiley-Interscience Series in Discrete Mathematics.
7. Frank Harary and Edgar M. Palmer, *Graphical enumeration*, Academic Press, New York, 1973.
8. Peter Kirschenhofer, Helmut Prodinger, and Robert F. Tichy, *Fibonacci numbers of graphs. II*, Fibonacci Quart. **21** (1983), no. 3, 219–229.
9. ______, *Fibonacci numbers of graphs. III. Planted plane trees*, Fibonacci numbers and their applications (Patras, 1984), Math. Appl., vol. 28, Reidel, Dordrecht, 1986, pp. 105–120.
10. Martin Klazar, *Twelve countings with rooted plane trees*, European J. Combin. **18** (1997), no. 2, 195–210.
11. A. Meir and J. W. Moon, *On the altitude of nodes in random trees*, Canad. J. Math. **30** (1978), no. 5, 997–1015.
12. ______, *On subtrees of certain families of rooted trees*, Ars Combin. **16** (1983), no. B, 305–318.
13. ______, *On maximal independent sets of nodes in trees*, J. Graph Theory **12** (1988), no. 2, 265–283.
14. R. E. Merrifield and H. E. Simmons, *Topological methods in chemistry*, Wiley, New York, 1989.
15. Richard Otter, *The number of trees*, Ann. of Math. (2) **49** (1948), 583–599.
16. George Pólya, *Kombinatorische Anzahlbestimmungen für Gruppen, Graphen und chemische Verbindungen*, Acta Math. **68** (1937), 145–254.
17. Helmut Prodinger and Robert F. Tichy, *Fibonacci numbers of graphs*, Fibonacci Quart. **20** (1982), no. 1, 16–21.
18. Frank Ruskey, *Listing and counting subtrees of a tree*, SIAM J. Comput. **10** (1981), no. 1, 141–150.

Stephan G. Wagner, Institut für Mathematik, Technische Universität Graz, Steyrergasse 30, 8010 Graz, Austria

E-mail address: `wagner@finanz.math.tugraz.at`