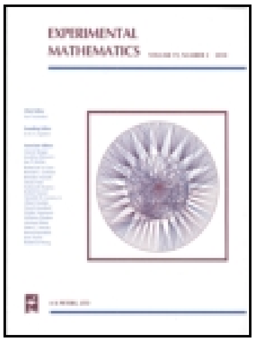

Experimental Mathematics

ISSN: 1058-6458 (Print) 1944-950X (Online) Journal homepage: http://www.tandfonline.com/loi/uexm20

On the Asymptotic Probability of Forbidden Motifs on the Fringe of Recursive Trees

Mohan Gopaladesikan, Stephan Wagner & Mark Daniel Ward

EXPERIMENTAL MATHEMATICS
2016, VOL. 25, NO. 3, 237–245
http://dx.doi.org/10.1080/10586458.2015.1065525

Taylor & Francis
Taylor & Francis Group

On the Asymptotic Probability of Forbidden Motifs on the Fringe of Recursive Trees

Mohan Gopaladesikan[a], Stephan Wagner[b], and Mark Daniel Ward[a]

[a]Department of Statistics, Purdue University, West Lafayette, Indiana, USA; [b]Department of Mathematical Sciences, Stellenbosch University, Stellenbosch, South Africa

ABSTRACT
We use analytic methods to study the probability of a family of motifs not occurring on the fringe of a random recursive tree. We obtain an asymptotic formula for this probability by means of singularity analysis. Two regimes are treated in particular: the case that a fixed proportion of motifs of size γ is forbidden, and the case that a fixed number of motifs of size γ is forbidden. In both cases, we observe phase transitions as the size of the random tree and the size of the motif tend to infinity. The required asymptotic expansions of the dominant singularities were first found by computer experiments and only later made rigorous.

KEYWORDS
analytic combinatorics; asymptotic analysis; forbidden patterns; generating functions; recursive trees

2000 AMS SUBJECT CLASSIFICATION
60C05; 05C05

1. Introduction

A recursive tree is a randomly generated, rooted, non-planar tree that is constructed by the insertion of nodes labeled 1, 2, 3, ... Node 1 is the root. Every subsequent node is inserted (uniformly, at random) as a child of one of the earlier nodes. Under this probability model, we have $(n-1)!$ equally likely recursive trees that contain n nodes.

This stochastic model has been used to study the growth of pyramid schemes, spread of chain letters, recruitment schemes, and the evolution of the Union-Find Algorithm. A detailed survey about the various studies and applications of recursive trees is given in [Smythe and Mahmoud 95]. Some of the other studies on recursive trees were conducted in [Dondajewski and Szymański 82], [Drmota 09], [Fuchs et al. 06], [Kuba and Panholzer 08], [Mahmoud and Smythe 92], [Panholzer and Prodinger 04], and [Pittel 94].

A motif is a (fixed) rooted, *unlabeled*, nonplanar tree of size γ. Figure 1 shows all of the motifs of size 5. The numbers of motifs of sizes 1, 2, 3, 4, 5, … are 1, 1, 2, 4, 9, …. The On-Line Encyclopedia of Integer Sequences entry A000081 gives a great deal more information about this sequence of data structures and related objects and enumerations (see http://oeis.org/A000081). A subtree of a recursive tree, rooted at a given node, includes that given node and all of its descendants.

A motif is said to *occur on the fringe* of a tree if any rooted subtree of the recursive tree takes the shape of the motif. For example, in Figure 2 we have a given motif of size 3 that appears on the fringe of some recursive trees of size 5 but does not appear on the fringe of some other recursive trees of size 5. We emphasize that the occurrence (or nonoccurrence) of a motif on the fringe of a recursive tree is not influenced by the labeling scheme. We have only included labels in Figure 2 on the recursive trees to emphasize that motifs are unlabeled shapes, and recursive trees are labeled.

Patterns in random unrooted trees have been studied in detail in [Chyzak et al. 08]. Forbidden patterns in binary search trees have been studied in [Flajolet et al. 97]. Patterns on the fringe of recursive trees have been studied in [Feng and Mahmoud 10] and [Gopaladesikan et al. 14]. In this paper, we study the asymptotic properties of the number of recursive trees (say, of size n) that do not have any members of a family Γ of motifs, each of size γ, occurring on the fringe. Dividing by $(n-1)!$, this enumeration immediately yields the probability p_n of a random recursive tree of size n not having any members of a family of motifs on the fringe.

We study this probability asymptotically by means of generating functions and singularity analysis. In particular, the location of the dominant singularity depending on γ will be analyzed in detail. Two specific scenarios will be of particular interest: a fixed ratio of all motifs of size γ is forbidden, or a fixed number of motifs of size γ is forbidden. The asymptotic expansions for the dominant

CONTACT Mark Daniel Ward mdw@purdue.edu Department of Statistics, Purdue University, 150 North University Street, West Lafayette, Indiana, 47907, USA.
Dedicated to the memory of Philippe Flajolet.

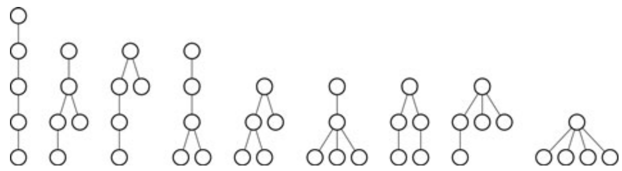

Figure 1. Plot of all (rooted, unlabeled, nonplanar) motifs of size 5.

singularities (see Corollaries 4.1 and 4.2) were first determined from an explicit expression involving the WhittakerM function by means of a bootstrapping approach that will be explained briefly in Section 3. Only later the authors found a rigorous direct method of locating the singularities; see Section 4.

If γ is fixed or grows only very slowly with n, it is reasonable to expect the probability that a large recursive tree avoids the motifs in Γ (on the fringe of the randomly generated tree) will tend to 0, i.e., $p_n \to 0$ as $n \to \infty$. We will precisely measure the rate at which this happens. As γ grows faster with n, we will observe phase transitions in both aforementioned scenarios.

2. Recursion for the probability

We apply a decomposition, similar to the one from [van der Hofstad et al. 02], in which we remove the edge that connects nodes 1 and 2. As a result, we have a subtree rooted at 2, which we refer to as the *special subtree*, and we let U_n denote the size of the special subtree. The remainder of the tree is rooted at 1, and we refer to this tree, of size $n - U_n$, as the *nonspecial subtree*. It is well known that U_n is a discrete uniform random variable on $\{1, \ldots, n-1\}$, and that (given U_n) the shape of the special subtree is (conditionally) independent of the shape of the nonspecial subtree. Moreover both of these subtrees are random recursive trees themselves. Let A_n be the event that a recursive tree of size n does not have any member of the family Γ occurring on the fringe.

Remark 2.1. For $n > \gamma$, for A_n to occur, there are three possibilities:

1. The nonspecial subtree has size strictly larger than γ, and neither the special subtree nor the nonspecial tree have any member of Γ on the fringe.
2. The nonspecial subtree has size strictly smaller than γ, and the special subtree does not have any member of Γ on the fringe.
3. The nonspecial subtree has size equal to γ, and the special subtree does not have any member of Γ on the fringe. (In this latter case, even if the nonspecial subtree is a member of Γ, we observe that the original recursive tree does not have any member of Γ on the fringe, once the special subtree is brought under consideration.)

With this remark in mind (using the three respective cases), we can condition on the value of U_n, and we derive the following, for $n > \gamma$:

$$\begin{aligned} P(A_n) = &\sum_{k=1}^{n-\gamma-1} P(U_n = k) P(\widehat{A}_{U_n} \cap \widetilde{A}_{n-U_n} \mid U_n = k) \\ &+ \sum_{k=n-\gamma+1}^{n-1} P(U_n = k) P(\widehat{A}_{U_n} \mid U_n = k) \\ &+ P(U_n = n-\gamma) P(\widehat{A}_{U_n} \mid U_n = n-\gamma), \end{aligned}$$

where $\widehat{A}_{U_n}$ is the event that the special subtree does not have any member of the family Γ occurring on the fringe, and $\widetilde{A}_{n-U_n}$ is the event that the nonspecial subtree does not have any member of Γ occurring on the fringe. We use independence of $\widehat{A}_{U_n}$ and $\widetilde{A}_{n-U_n}$ in the first case. We

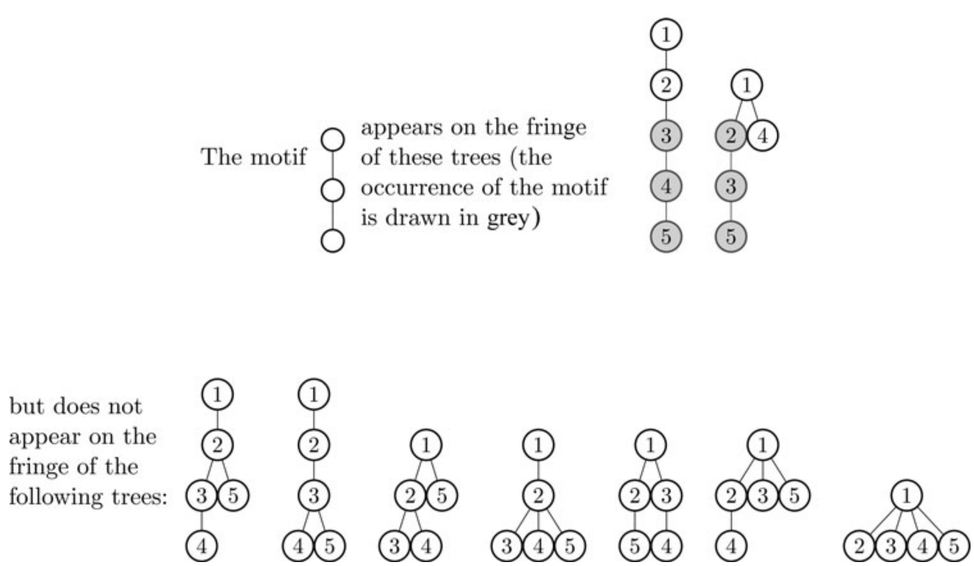

Figure 2. Examples of recursive trees in which a given motif does, or does not, appear on the fringe.

use the fact that $\widetilde{A}_{n-U_n}$ always occurs in the second case, because the special subtree has size $n-k$, which is strictly less than γ, in the second case. Thus, for $n > \gamma$,

$$P(A_n) = \frac{1}{n-1}\left[\left(\sum_{k=1}^{n-\gamma-1} P(\widehat{A}_k)P(\widetilde{A}_{n-k})\right) + \left(\sum_{k=n-\gamma+1}^{n-1} P(\widehat{A}_k)P(\widetilde{A}_{n-k})\right) + P(\widehat{A}_{n-\gamma})\right],$$

which simplifies to

$$P(A_n) = \frac{1}{n-1}\left[\left(\sum_{k=1}^{n-1} P(\widehat{A}_k)P(\widetilde{A}_{n-k})\right) + P(\widehat{A}_{n-\gamma})(1-P(\widetilde{A}_\gamma))\right]. \tag{2–1}$$

3. Generating function

We use $p_n := P(A_n)$ to denote the probability that a recursive tree of size n does not have any member of the family Γ occurring on the fringe. (Recall: all members of the family Γ have size γ.) We use $\mathcal{C}(\Gamma)$ to denote the probability that a recursive tree of size γ takes the shape of a motif in the family Γ. ([Feng and Mahmoud 10] define $\mathcal{C}(\Gamma)$ in a similar way, but only in the special case where Γ consists of just one motif; our definition is more general.) Thus $p_j = 1$ for all $j < \gamma$, and $p_\gamma = 1 - \mathcal{C}(\Gamma)$. Next we define the probability generating function

$$f(z) = \sum_{n=0}^{\infty} p_n z^n.$$

From (2–1), we obtain, for $n > \gamma$,

$$(n-1)p_n = \left(\sum_{k=1}^{n-1} p_k p_{n-k}\right) + p_{n-\gamma}(1-p_\gamma).$$

Multiplying by z^n and summing over $n > \gamma$, we obtain

$$\sum_{n>\gamma}(n-1)p_n z^n = \sum_{n>\gamma}\left[\left(\sum_{k=1}^{n-1} p_k p_{n-k}\right) + p_{n-\gamma}(1-p_\gamma)\right] z^n.$$

We add

$$\sum_{n=1}^{\gamma}(n-1)p_n z^n = \frac{z^{\gamma+1}(\gamma-1) - z^\gamma(\gamma-1) - z^{\gamma+1} + z^2}{(z-1)^2} + (\gamma-1)(1-\mathcal{C}(\Gamma))z^\gamma$$

on both sides of the previous equation, and then we simplify. This yields the Riccati differential equation

$$zf'(z) = (f(z))^2 + (\mathcal{C}(\Gamma)z^\gamma - 1)f(z) - \gamma\mathcal{C}(\Gamma)z^\gamma. \tag{3–2}$$

Next we solve the equation, which is an excellent example for both the strengths and limitations of computer algebra. Maple immediately provides the solution

$$f(z) = 1 - \frac{kz^\gamma}{(\gamma-1)!} + \frac{\alpha(z)}{\beta(z)}, \tag{3–3}$$

where

$$\alpha(z) = z\exp\left(\frac{-kz^\gamma}{2\gamma!}\right)$$

and

$$\beta(z) = \exp\left(\frac{kz^\gamma}{2\gamma!}\right) - z\exp\left(\frac{-kz^\gamma}{2\gamma!}\right) - \frac{\gamma\sqrt{z}}{\gamma+1}\left(\frac{\gamma!}{k}\right)^{\frac{1}{2\gamma}} M_{\frac{1}{2\gamma},\frac{\gamma+1}{2\gamma}}\left(\frac{kz^\gamma}{\gamma!}\right).$$

Here, M denotes the WhittakerM function, a classical special function, discussed at length in [Andrews et al. 99]. They define this function in equation (4.3.2) on page 195, as follows:

$$M_{k,m}(x) = e^{-x/2}x^{\frac{1}{2}+m}{}_1F_1\left[\begin{matrix}\frac{1}{2}+m-k \\ 1+2mx\end{matrix}; x\right],$$

where ${}_1F_1$ is the hypergeometric function, which can be defined in several ways, including

$${}_1F_1\left[\begin{matrix}a \\ b\end{matrix}; z\right] := \sum_{\ell=0}^{\infty} \frac{\prod_{i=0}^{\ell-1}(a+i)}{\prod_{j=0}^{\ell-1}(b+j)}\frac{z^\ell}{\ell!}.$$

Hypergeometric functions are a key object of study in the theory of special functions; [Andrews et al. 99] serves as an excellent reference to this beautiful subject.

While (3–3) is explicit, it is not particularly useful for further rigorous analysis. In particular, we need the singularities of $f(z)$, which are the zeros of $\beta(z)$, and those are quite difficult to find from the given expression (especially if one does not only want to consider fixed values of γ). However, it is possible to discover an asymptotic expansion by means of bootstrapping and heavy use of computer algebra, which was in fact our first approach. Let us briefly present this approach in one specific case, which we believe to be useful on its own right.

It will be shown later (Corollary 4.1) that in the case where $\mathcal{C}(\Gamma) = q$ is fixed, the dominant singularity z_0 (closest to the origin) of $f(z)$ satisfies

$$z_0 = 1 + \frac{q}{\gamma^2} + \frac{3q^2-4q}{4\gamma^3} + \frac{58q^3-63q^2+72q}{72\gamma^4} + O(\gamma^{-5}).$$

This was originally found as follows: z_0 is the smallest zero of

$$\beta(z) = \exp\left(\frac{kz^\gamma}{2\gamma!}\right) - z\exp\left(\frac{-kz^\gamma}{2\gamma!}\right) - \frac{\gamma\sqrt{z}}{\gamma+1}\left(\frac{\gamma!}{k}\right)^{\frac{1}{2\gamma}} M_{\frac{1}{2\gamma},\frac{\gamma+1}{2\gamma}}\left(\frac{kz^\gamma}{\gamma!}\right).$$

To find this singularity z_0, we can use Maple to assist. We first define

```
mybeta := exp(k*z^g/2/g!)
          -z*exp(-k*z^g/2/g!)
          - g*sqrt(z)/(g+1)
          *(g!/k)^(1/(2*g))
          *WhittakerM(1/(2*g),(g+1)/
          (2*g),k*z^g/g!);
```

We suspect the singularity z_0 is near 1. A plot confirms this suspicion:

```
Digits := 50:
q := 1/10:
g := 'g':
z0 := solve( subs( k = q*
     (g-1)!, mybeta), z):
plot( z0, g=1..50 );
```

(Notice that we are using $q = 1/10$ in this example, but we emphasize that any value of q will give comparable results. Maple is unable to yield any such insights if we do not specify a value for q. So we demonstrate the analysis for $q = 1/10$ and emphasize that the analogous proceeds in an similar way for other values of q.)

We check if it is, say, approximately $1/\gamma$ beyond 1; but it is not, because the plot of $(z_0 - 1)/\gamma$ decreases to 0.

```
plot( (z0-1)*g, g=1..50 );
```

A quick check, however, confirms that z_0 is actually on the order of $1/\gamma^2$ beyond 1:

```
plot( (z0-1)*g^2, g=1..50 );
```

and the leading order constant looks like q, so now we have

$$z_0 = 1 + \frac{q}{\gamma^2} + o(\gamma^{-2}).$$

In fact, we can ascertain that the next term in the asymptotic expansion is a constant multiple of $1/\gamma^3$ with the following:

```
plot( (z0-1-q/g^2)*g^3, g=1..50 );
```

but now we are faced with finding that constant multiple for the $1/\gamma^3$ term. There are various ways to do so. For instance, we can increase the number of digits of accuracy in the computation, and plot the value of this constant over all possible q values:

```
Digits := 1000;
q := 'q': g := 300:
plot( (evalf(solve( subs( k = q*
(g-1)!, mybeta), z)) - 1 -
q/g^2)*g^3, q=0..1 );
```

This allows us to see that the next term is quadratic in q. Indeed, if we change the range above from $0 \le q \le 1$ to using $0 \le q \le 2$, we can see that this parabola crosses the x-axis at $q = 0$ and $q = 4/3$, and has minimum at the point $(2/3, -1/3)$, so the next constant must be $(3q^2 - 4q)/4$. This yields

$$z_0 = 1 + \frac{q}{\gamma^2} + \frac{3q^2 - 4q}{4\gamma^3} + o(\gamma^{-3}).$$

Returning to our earlier series of calculations, if we compute:

```
Digits := 500;
q := 1/10:
g := 'g':
plot( (z0-1-q/g^2-q*((3*q-
4)/(4*g^3))*g^4, g=1..120 );
```

We can see that the order of the error term is actually $\Theta(\gamma^{-4})$.

By similar reasoning, we see that the next term in the asymptotic expansion is $(58q^3 - 63q^2 + 72q)/(72\gamma^4)$, and the error term after incorporating this term is then $\Theta(\gamma^{-5})$.

While not rigorous, it was very useful to have an asymptotic formula like this to guide one's intuition. This was our first approach, and since it is both instructive and in the spirit of the journal, we decided to keep this discussion in the paper.

We now continue with the more formal approach, which no longer uses the explicit expression (3–3). Instead, we apply a substitution to the original differential equation (3–2), a technique that was also applied successfully in a similar context in a paper of [Flajolet et al. 97]. Setting $f(z) = 1 - \frac{zg'(z)}{g(z)}$ yields (after some simplifications) the linear differential equation

$$g''(z) - \mathcal{C}(\Gamma)z^{\gamma-1}g'(z) - \mathcal{C}(\Gamma)(\gamma-1)z^{\gamma-2}g(z) = 0.$$

for the auxiliary function g. Notice that this can also be written as

$$g''(z) - \mathcal{C}(\Gamma)\frac{d}{dz}(z^{\gamma-1}g(z)) = 0,$$

so $g'(z) - \mathcal{C}(\Gamma)z^{\gamma-1}g(z)$ must be constant. Thus we are left with a first-order linear differential equation that can be solved by standard methods. It turns out that the equation for g has the two linearly independent solutions

$$g_1(z) = e^{\mathcal{C}(\Gamma)z^\gamma/\gamma} \quad \text{and} \quad g_2(z) = e^{\mathcal{C}(\Gamma)z^\gamma/\gamma}\int_0^z e^{-\mathcal{C}(\Gamma)t^\gamma/\gamma}\,dt.$$

Thus we get

$$g(z) = e^{\mathcal{C}(\Gamma)z^\gamma/\gamma}\left(A + B\int_0^z e^{-\mathcal{C}(\Gamma)t^\gamma/\gamma}\,dt\right).$$

Next, since we know that the expansion of $f(z)$ starts $1+z+\cdots$, we must have $g'(0)/g(0) = -1$. Since $g(0) = A$ and $g'(0) = B$, this gives us $B = -A$, so

$$g(z) = Ae^{\mathcal{C}(\Gamma)z^\gamma/\gamma}\left(1 - \int_0^z e^{-\mathcal{C}(\Gamma)t^\gamma/\gamma}\,dt\right),$$

and without loss of generality we may take $A = 1$ (since we are only interested in the quotient $\frac{g'(z)}{g(z)}$). This finally gives us

$$f(z) = 1 - \mathcal{C}(\Gamma)z^\gamma + \frac{z}{g(z)} = 1 - \mathcal{C}(\Gamma)z^\gamma + \frac{ze^{-\mathcal{C}(\Gamma)z^\gamma/\gamma}}{1-\int_0^z e^{-\mathcal{C}(\Gamma)t^\gamma/\gamma}\,dt}, \tag{3–4}$$

an expression that is much easier to work with than the solution in terms of the WhittakerM-function.

We see that f is a meromorphic function with poles at all solutions of the equation

$$I(z) = \int_0^z e^{-\mathcal{C}(\Gamma)t^\gamma/\gamma}\,dt = 1. \tag{3–5}$$

All these poles are simple, since the derivative of $I(z)$ with respect to z is $e^{-\mathcal{C}(\Gamma)z^\gamma/\gamma}$, which is never 0. The integral can be expressed in terms of an incomplete Gamma function, but this will also be immaterial for us.

4. Asymptotic analysis

In order to carry out the Flajolet–Odlyzko singularity analysis, we need information on the location of the poles of the generating function f, i.e., the solutions to (3–5). There is a unique positive real solution z_0, since $I(x)$ is increasing for positive real x, with $I(0) = 0$ and

$$\lim_{x\to\infty} I(x) = \int_0^\infty e^{-\mathcal{C}(\Gamma)t^\gamma/\gamma}\,dt = (\gamma/\mathcal{C}(\Gamma))^{1/\gamma} \times\Gamma(1+1/\gamma) \geq \gamma^{1/\gamma}\Gamma(1+1/\gamma) > 1.$$

In the following, we prove that there is no other singularity of f whose absolute value is z_0 (by Pringsheim's Theorem, we know immediately that there cannot be any singularities whose absolute value is less), and we provide an asymptotic expansion for z_0 in terms of $\mathcal{C}(\Gamma)$ and γ.

Proposition 4.1. *For every $\gamma \geq 2$ and every possible value of $\mathcal{C}(\Gamma)$, the unique positive real solution z_0 of (3–5) is the only solution whose modulus is less than $1 + 1/\gamma$. Setting $a = \mathcal{C}(\Gamma)/\gamma$, we have*

$$z_0 = 1 + \frac{a}{\gamma+1} + \frac{(3\gamma+1)a^2}{2(\gamma+1)(2\gamma+1)} + \frac{(29\gamma^3+32\gamma^2+10\gamma+1)a^3}{6(\gamma+1)^2(2\gamma+1)(3\gamma+1)} + O\left(\frac{a^4}{\gamma}\right),$$

where the constant implied in the O-term is absolute (independent of a and γ). Moreover, we have

$$|f(z)| = O(\gamma)$$

for $|z| = 1 + 1/\gamma$, where the O-constant does not depend on a (or equivalently $\mathcal{C}(\Gamma)$).

Table 1. Numerical values for small γ.

Case	Numerical value of z_0	Smallest nonreal solutions of $I(z) = 1$
$\gamma = 2, \mathcal{C}(\Gamma) = 1$	1.27555	$2.39632 \pm 2.33408i$
$\gamma = 3, \mathcal{C}(\Gamma) = 1$	1.11259	$-2.34252 \pm 0.87577i$
$\gamma = 3, \mathcal{C}(\Gamma) = 1/2$	1.04755	$-2.94931 \pm 1.12578i$

Remark 4.1. As can be seen from the proof that follows, it is possible to continue the asymptotic expansion of z_0 arbitrarily far.

Proof. We are looking for solutions to the equation

$$I(z) = \int_0^z e^{-at^\gamma}\,dt = 1.$$

The cases $\gamma = 2$ and $\gamma = 3$ are listed in Table 1. It is noteworthy that the singularities for $\gamma = 2, \mathcal{C}(\Gamma) = 1$ and for $\gamma = 3, \mathcal{C}(\Gamma) = 1/2$ also occur in the enumeration of permutations avoiding a consecutive 132- or 1342-pattern; see [Elizalde and Noy 03]. A formal proof that there are no nonreal solutions with $|z| \leq 1 + 1/\gamma$ in these cases can be given by applying the argument principle: to this end, we numerically compute the zero-counting integral

$$\frac{1}{2\pi i}\oint_{|z|=1+1/\gamma} \frac{I'(z)}{I(z)-1}\,dz.$$

Since we know that its value must be an integer, the numerical integration does not even have to be particularly accurate to show that it is in fact equal to 1. It would also be possible to give a proof in the cases $\gamma = 2$ and $\gamma = 3$ by further strengthening the arguments that we use in the following for $\gamma \geq 4$, but this is somewhat tedious. Moreover, the aforementioned numerical proof seemed more in the spirit of this journal. For more information about the numerical computation of complex solutions, we refer to [Kravanja and Van Barel 00].

In the following, we assume that $\gamma \geq 4$. Note first that $\mathcal{C}(\Gamma) \leq 1$, so that $0 < a \leq 1/\gamma$. Thus if $|t| \leq T = 1 + 1/\gamma$, we have

$$|-at^\gamma| \leq \frac{T^\gamma}{\gamma}$$

and consequently

$$|1-e^{-at^{\gamma}}| \le \frac{e^{T^{\gamma}/\gamma}-1}{T^{\gamma}/\gamma}\cdot a|t|^{\gamma}.$$

Integrating this estimate gives us, for $|z| \le T = 1+1/\gamma$,

$$\begin{aligned}|z-I(z)| &= \left|\int_0^z \left(1-e^{-at^{\gamma}}\right)dt\right| \\ &\le \int_0^{|z|} \frac{e^{T^{\gamma}/\gamma}-1}{T^{\gamma}/\gamma}\cdot au^{\gamma}\,du \\ &\le \frac{e^{T^{\gamma}/\gamma}-1}{T^{\gamma}/\gamma}\cdot\frac{1}{\gamma}\cdot\frac{T^{\gamma+1}}{\gamma+1} \\ &= (e^{T^{\gamma}/\gamma}-1)\cdot\frac{T}{\gamma+1}.\end{aligned}$$

By our choice of T, we have $T^{\gamma}/\gamma = (1+1/\gamma)^{\gamma}/\gamma \le e/\gamma \le e/4$ and $T/(\gamma+1) = 1/\gamma$. It follows that

$$\begin{aligned}|(z-1)-(I(z)-1)| = |z-I(z)| &\le \frac{e^{e/\gamma}-1}{\gamma} \\ &\le \frac{e^{e/4}-1}{\gamma} < \frac{1}{\gamma}\end{aligned} \tag{4–6}$$

for $|z| \le T$. Therefore, any solution of the equation $I(z) = 1$ with $|z| \le T = 1+1/\gamma$ must also satisfy $|z-1| < 1/\gamma$. Now we can apply Rouché's theorem to the circle given by $|z-1| = 1/\gamma$: For any z on this circle, we have $|z| \le T = 1+1/\gamma$ and thus also

$$|(z-1)-(I(z)-1)| < \frac{1}{\gamma} = |z-1|$$

by the estimate (4–6) above. It follows that the equation $I(z) = 1$ has equally many solutions as the equation $z = 1$ inside this circle, i.e., exactly one. This solution has to be real: we trivially have $I(1) < 1$, and since $|I(T) - T| < 1/\gamma$ by (4–6), we also have $I(T) > T - 1/\gamma = 1$. So by the intermediate value theorem, the (unique) positive real solution lies between 1 and $T = 1+1/\gamma$.

Next we prove the estimate for $|f(z)|$. If $|z| = T = 1 + 1/\gamma$, we have

$$|1-\mathcal{C}(\Gamma)z^{\gamma}| \le 1+T^{\gamma} < 1+e$$

and (in view of (4–6))

$$\begin{aligned}&\left|\frac{ze^{-az^{\gamma}}}{1-I(z)}\right| \le \frac{|z|e^{a|z|^{\gamma}}}{|1-I(z)|} \\ &\le \frac{Te^{aT^{\gamma}}}{|1-z|-|z-I(z)|} \le \frac{\frac{5}{4}\cdot e^{e/4}}{1/\gamma-(e^{e/4}-1)/\gamma} = O(\gamma),\end{aligned}$$

thus

$$f(z) = 1-\mathcal{C}(\Gamma)z^{\gamma}+\frac{ze^{-az^{\gamma}}}{1-I(z)} = O(\gamma),$$

which is what we wanted to prove. It remains to justify the asymptotic formula for z_0. To this end, we use the power series expansion

$$\begin{aligned}I(z) &= \int_0^z e^{-at^{\gamma}}\,dt = \int_0^z \sum_{k=0}^{\infty}\frac{(-1)^k}{k!}a^k t^{\gamma k}\,dt \\ &= \sum_{k=0}^{\infty}\frac{(-1)^k}{(\gamma k+1)k!}a^k z^{\gamma k+1}.\end{aligned}$$

We already know that we only have to consider values of z between 1 and $1+1/\gamma$, for which z^{γ} is bounded. In this region,

$$I(z) = z-\frac{a}{\gamma+1}z^{\gamma+1}+O\Big(\frac{a^2}{\gamma}\Big),$$

and since $a \le 1/\gamma$, then $I(z) = z + O(a/\gamma)$. Hence, we conclude

$$1 = I(z_0) = z_0 + O\Big(\frac{a}{\gamma}\Big),$$

which gives us a first approximation $z_0 = 1 + O(a/\gamma)$. This can be refined by bootstrapping: we know now that $z_0^{\gamma+1} = 1 + O(a)$, thus

$$1 = I(z_0) = z_0 - \frac{a}{\gamma+1} + O\Big(\frac{a^2}{\gamma}\Big).$$

This readily gives us the first two terms:

$$z_0 = 1+\frac{a}{\gamma+1}+O\Big(\frac{a^2}{\gamma}\Big).$$

Now we can refine further:

$$z_0^{\gamma+1} = 1+a+O(a^2)$$

and consequently

$$\begin{aligned}1 = I(z_0) &= z_0-\frac{a}{\gamma+1}z_0^{\gamma+1}+\frac{a^2}{2(2\gamma+1)}z_0^{2\gamma+1}+O\Big(\frac{a^3}{\gamma}\Big) \\ &= z_0-\frac{a}{\gamma+1}-\frac{a^2}{\gamma+1}+\frac{a^2}{2(2\gamma+1)}+O\Big(\frac{a^3}{\gamma}\Big) \\ &= z_0-\frac{a}{\gamma+1}-\frac{(3\gamma+1)a^2}{2(\gamma+1)(2\gamma+1)}+O\Big(\frac{a^3}{\gamma}\Big).\end{aligned}$$

This yields the third term, and further terms are obtained by continuing the process. □

Now we are ready to apply singularity analysis. The result is summarized in the following theorem:

Theorem 4.1. *With z_0 as described in Proposition 4.1, the probability p_n that a recursive tree of size n does not contain a motif in Γ is asymptotically given by*

$$p_n = z_0^{-n}+O\left(\gamma\left(1+\frac{1}{\gamma}\right)^{-n}\right),$$

where the O-constant is independent of γ or $\mathcal{C}(\Gamma)$.

Proof. We apply singularity analysis in the meromorphic setting;

see Chapters IV and V of [Flajolet and Sedgewick 09]. Cauchy's integral formula yields

$$p_n = \frac{1}{2\pi i}\oint_{|z|=1}\frac{f(z)}{z^{n+1}}\,dz = -\operatorname*{Res}_{z=z_0}\frac{f(z)}{z^{n+1}} + \frac{1}{2\pi i}\oint_{|z|=1+1/\gamma}\frac{f(z)}{z^{n+1}}\,dz.$$

The only term in $f(z)$ that contributes to the residue is

$$\frac{ze^{-\mathcal{C}(\Gamma)z^\gamma/\gamma}}{1-\int_0^z e^{-\mathcal{C}(\Gamma)t^\gamma/\gamma}\,dt} = \frac{zI'(z)}{1-I(z)},$$

and the asymptotic behavior at the pole z_0 is given by

$$f(z) \sim \frac{zI'(z)}{1-I(z)} \sim \frac{z_0 I'(z_0)}{-I'(z_0)(z-z_0)} \sim \frac{1}{1-z/z_0}$$

as $z \to z_0$, so

$$\operatorname*{Res}_{z=z_0}\frac{f(z)}{z^{n+1}} = -z_0^{-n}.$$

Thus it only remains to estimate the integral. However, we already know from Proposition 4.1 that $|f(z)| = O(\gamma)$ for $|z| = 1 + 1/\gamma$. Thus

$$\left|\frac{1}{2\pi i}\oint_{|z|=1+1/\gamma}\frac{f(z)}{z^{n+1}}\right| \le \left(1+\frac{1}{\gamma}\right)^{-n}\max_{|z|=1+1/\gamma}|f(z)| = O\Big(\gamma\Big(1+\frac{1}{\gamma}\Big)^{-n}\Big),$$

which completes the proof. □

Remark 4.2. We obtain that

$$p_n \sim z_0^{-n} \tag{4–7}$$

as $n \to \infty$ for *all* $\gamma < n$, which may potentially go to ∞ as well. For $\gamma = o(n/\log n)$, this follows from Theorem 4.1, since the O-term is indeed dominated by z_0^{-n} in this case. For larger $\gamma < n$ (in fact, as soon as $\gamma/\sqrt{n} \to \infty$, see also Theorem 4.2), the asymptotic approximation in (4–7) is trivial, since both sides go to 1. Let us provide a simple probabilistic proof for this fact. To this end, we determine the expected total number of occurrences of motifs of size γ. Thus we count recursive trees with a motif of size γ appearing on the fringe with root node k. There are $\binom{n-k}{\gamma-1}$ ways to pick the labels of the other nodes, $\mathcal{C}(\Gamma)(\gamma-1)!$ choices for the fringe subtree itself, $(n-\gamma-1)!$ possibilities for the shape of the remaining tree, and $k-1$ possible nodes to which node k can be attached. Thus we end up with a total expected value (consistent with [Feng and Mahmoud 10] and [Gopaladesikan et al. 14]) of

$$\frac{1}{(n-1)!}\sum_{k=2}^{n}\binom{n-k}{\gamma-1}\mathcal{C}(\Gamma)(\gamma-1)!(n-\gamma-1)!(k-1) = \frac{\mathcal{C}(\Gamma)n}{\gamma(\gamma+1)}.$$

If $\gamma/\sqrt{n} \to \infty$, then this expected value goes to 0, and thus (by the Markov inequality) the probability that any motif of size γ (be it in Γ or not) appears also goes to 0.

If $\gamma/\sqrt{n} \to \infty$ and $\gamma = o(n/\log n)$, Theorem 4.1 also provides the following asymptotic formula for the probability that a motif of size γ actually appears:

$$1 - p_n \sim \frac{\mathcal{C}(\Gamma)n}{\gamma^2}. \tag{4–8}$$

The condition that $\gamma = o(n/\log n)$ can be dropped again: a similar counting argument as before shows that the expected number of pairs of distinct fringe subtrees both showing a motif in Γ is $O(\mathcal{C}(\Gamma)^2 n^2/\gamma^4)$. This also means that trees where at least two disjoint motifs in Γ appear only contribute $O(\mathcal{C}(\Gamma)^2 n^2/\gamma^4)$ to the mean number of appearances, and that there are only $O(\mathcal{C}(\Gamma)^2 n^2/\gamma^4)$ such trees. Since this goes faster to 0 than $\mathcal{C}(\Gamma)n/\gamma^2$, the main term of the expected number of appearances must be asymptotically equal to the probability that a motif in Γ appears, i.e., (4–8) remains correct for arbitrary $\gamma < n$, provided that $\gamma/\sqrt{n} \to \infty$.

4.1. *Fixed ratio of motifs are forbidden*

Recall that $\mathcal{C}(\Gamma)$ denotes the probability that a recursive tree of size γ takes the shapes in Γ. Since there are altogether $(\gamma-1)!$ recursive trees of size γ, we define k as the number of recursive trees of size γ that have the shapes in Γ. One possible scenario is to examine the family of forbidden motifs in which k and $(\gamma-1)!$ have a fixed ratio. In other words, in this situation we are studying a scenario in which

$$\mathcal{C}(\Gamma) = \frac{k}{(\gamma-1)!} = q$$

is constant, a particular instance being the case that all motifs of size γ are forbidden (i.e., $q = 1$). From Proposition 4.1, we immediately obtain

Corollary 4.1. *In the scenario that $\mathcal{C}(\Gamma) = q$ is fixed, the dominant singularity of $f(z)$ is*

$$z_0 = 1 + \frac{q}{\gamma^2} + \frac{3q^2-4q}{4\gamma^3} + \frac{58q^3-63q^2+72q}{72\gamma^4} + O(\gamma^{-5}).$$

Now Theorem 4.1 gives us the following result, which exhibits a phase transition:

Theorem 4.2. *If $\mathcal{C}(\Gamma) = q$ is fixed, the probability p_n that a recursive tree of size n does not have any member of the family Γ occurring on the fringe satisfies*

$$p_n \to \begin{cases} 0 & \text{if } \gamma/\sqrt{n} \to 0, \\ e^{-q/a^2} & \text{if } \gamma/\sqrt{n} \to a \in (0,\infty), \\ 1 & \text{if } \gamma/\sqrt{n} \to \infty. \end{cases}$$

Figure 3. Example in which Γ consists of all straight paths of γ nodes.

4.2. *Family of forbidden motifs is fixed*

Another possibility is to consider the case where the number k of recursive trees of size γ that have the shapes in Γ is held constant. In such a case, we have the following:

Corollary 4.2. *In the scenario that $k = \mathcal{C}(\Gamma)(\gamma-1)!$ is fixed, the dominant singularity of f(z) is*

$$\begin{aligned} z_0 = 1 &+ \frac{k}{(\gamma+1)!} + \frac{(3\gamma+1)k^2}{2(2\gamma+1)\gamma!(\gamma+1)!} \\ &+ \frac{(29\gamma^3+32\gamma^2+10\gamma+1)k^3}{6(2\gamma+1)(3\gamma+1)\gamma!(\gamma+1)!^2} \\ &+ O(\gamma^{-1}\gamma!^{-4}). \end{aligned}$$

We give three examples of families in which the value of $k = (q)(\gamma-1)!$ is fixed. These examples illustrate the kinds of families analyzed in Corollary 4.2. Besides the three examples we give here, we can combine any of these examples to get more. They are quite simple, but more complicated examples are (of course) possible as well.

Example 4.1. As a first example, we consider the family Γ consisting of straight paths. Each such path of length γ has probability $1/(\gamma-1)!$ (note that the labels have to increase from the root down and are therefore unique). Thus, this corresponds to the case $k = 1$. See Figure 3.

In this example, we obtain a very strong limit law by combining the information of Theorem 4.1 and Corollary 4.2. Specifically:

Theorem 4.3. *Define $\lambda = \lambda(n)$ as the positive integer for which $\Gamma(\lambda - \frac{1}{2}) \le n < \Gamma(\lambda + \frac{1}{2})$. In a random recursive tree of order n, the length of the longest straight path ending in a leaf is either $\lambda - 3$ or $\lambda - 2$ with probability going to 1 as n goes to infinity.*

Proof. In view of Theorem 4.1 and Corollary 4.2, as n goes to infinity, the probability that a recursive tree of size n does not have a straight path of $\lambda - 3$ nodes occurring on

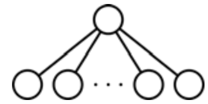

Figure 4. Example in which Γ consists of all stars, each of which is a parent and $\gamma - 1$ children.

the fringe satisfies

$$p_n \sim \exp\left(-\frac{n}{(\lambda-2)!}\right) \le \exp\left(-\frac{\Gamma(\lambda-1/2)}{(\lambda-2)!}\right) \to 0.$$

Likewise, the probability that there is no straight path of $\lambda - 1$ nodes occurring on the fringe satisfies

$$p_n \sim \exp\left(-\frac{n}{\lambda!}\right) \ge \exp\left(-\frac{\Gamma(\lambda+1/2)}{\lambda!}\right) \to 1.$$

The theorem follows immediately. □

An analogous theorem holds for stars as well, which are our second example:

Example 4.2. As a second example, consider the family Γ consisting of stars, each of which is a parent and $\gamma - 1$ children. Each such star has size γ and therefore probability $1/(\gamma-1)!$, so each of these also corresponds to the case $k = 1$ (again there is only one possible labeling once the set of labels is fixed). See Figure 4.

Example 4.3. As a final example, consider the family Γ consisting of a long string of length $\gamma - 1$, with one node to the side, so that there are γ nodes altogether. The branching node has a string of $r - 1$ descendants on one side and 1 child on the other side. We fix the value of r throughout the family Γ. With this setup, each such structure has probability $(r-1)/(\gamma-1)!$, which corresponds to the case $k = r - 1$. See Figure 5.

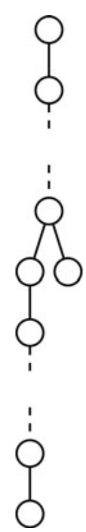

Figure 5. Example in which Γ consists of structures with 1 branching node that has a string of $r - 1$ descendants on one side and 1 child on the other side.

Acknowledgments

The authors thank Dr. Hosam Mahmoud for guidance and suggestions during the course of this investigation, as well as for his hospitality during M.G.'s and M.D.W.'s extended visit to The George Washington University in 2013–14.

Funding

M.G. and M.D.W. were supported by the NSF Science & Technology Center for Science of Information Grant CCF-0939370. S.W. was supported by the National Research Foundation of South Africa, grant number 96236.

References

[Andrews et al. 99] G. E. Andrews, R. Askey, and R. Roy. *Special Functions, Vol. 71 of Encyclopedia of Mathematics and Its Applications*. Cambridge: Cambridge University Press, 1999.

[Chyzak et al. 08] F. Chyzak, M. Drmota, T. Klausner, and G. Kok. "The Distribution of Patterns in Random Trees." *Comb. Probab. Comput.* 17(2008), 21–59.

[Dondajewski and Szymański 82] M. Dondajewski and J. Szymański. "On the Distribution of Vertex-degrees in a Strata of a Random Recursive Tree." *Bulletin de l'Académie Polonaise des Sci.* 30(1982), 205–209.

[Drmota 09] M. Drmota. "The Height of Increasing Trees." *Ann. Comb.* 12(2009), 373–402.

[Elizalde and Noy 03] S. Elizalde and M. Noy. "Consecutive Patterns in Permutations." *Adv. Appl. Math.* 30: 1–2 (2003), 110–125. Formal Power Series and Algebraic Combinatorics (Scottsdale, AZ, 2001).

[Feng and Mahmoud 10] Q. Feng and H. M. Mahmoud. "On the Variety of Shapes on the Fringe of a Random Recursive Tree." *J. Appl. Prob.* 47(2010), 191–200.

[Flajolet and Sedgewick 09] P. Flajolet and R. Sedgewick. *Analytic Combinatorics*. Cambridge: Cambridge University Press, 2009.

[Flajolet et al. 97] P. Flajolet, X. Gourdon, and C. Martínez. "Patterns in Random Binary Search Trees." *Random Struct. Algorithms* 11(1997), 223–244.

[Fuchs et al. 06] M. Fuchs, H.-K. Hwang, and R. Neininger. "Profiles of Random Trees: Limit Theorems for Random Recursive Trees and Binary Search Trees." *Algorithmica* 46: 3–4 (2006), 367–407.

[Gopaladesikan et al. 14] M. Gopaladesikan, H. Mahmoud, and M. D. Ward. "Asymptotic Joint Normality of Counts of Uncorrelated Motifs in Recursive Trees." *Method. Comput. Appl. Prob.* 2014. In press.

[Kravanja and Van Barel 00] P. Kravanja and M. Van Barel. *Computing the Zeros of Analytic Functions, Volume 1727 of Lecture Notes in Mathematics*. Berlin: Springer-Verlag, 2000.

[Kuba and Panholzer 08] M. Kuba and A. Panholzer. "Isolating Nodes in Recursive Trees." *Aequationes Mathematicae* 76(2008), 258–280.

[Mahmoud and Smythe 92] H. M. Mahmoud and R. T. Smythe. "Asymptotic Joint Normality of Outdegrees of Nodes in Random Recursive Trees." *Random Struct. Algorithms* 3(1992), 255–266.

[Panholzer and Prodinger 04] A. Panholzer and H. Prodinger. "Analysis of Some Statistics for Increasing Tree Families." *Discrete Math. Theor. Comput. Sci.* 6(2004), 437–460.

[Pittel 94] B. Pittel. "Note on the Heights of Random Recursive Trees and Random m-ary Search Trees." *Random Struct. Algorithms* 5(1994), 337–347.

[Smythe and Mahmoud 95] R. T. Smythe and H. Mahmoud. "A Survey of Recursive Trees." *Theory Prob. Math. Stat.* 51(1995), 1–27.

[van der Hofstad et al. 02] R. van der Hofstad, G. Hooghiemstra, and P. Van Mieghem. "On the Covariance of the Level Sizes in Random Recursive Trees." *Random Struct. Algorithms* 20(2002), 519–539.