

k -Protected Vertices in Unlabeled Rooted Plane Trees

Keith Copenhaver¹ 

Received: 14 June 2016 / Revised: 22 December 2016 / Published online: 7 February 2017
© Springer Japan 2017

Abstract We find a simple, closed formula for the proportion of vertices which are k -protected in all unlabeled rooted plane trees on n vertices. We also find that, as n goes to infinity, the average rank of a random vertex in a tree of size n approaches 0.727649, and the average rank of the root of a tree of size n approaches 1.62297.

Keywords Tree · Enumeration · Asymptotics

Mathematics Subject Classification 05A15 · 05A16 · 05C05

1 Introduction

Unlabeled rooted plane trees are one of the simplest tree structures. They are used to model any network with a point of origin. We will ask the question, if you start at a leaf and take only upward steps, how far might we expect to go to reach a given vertex?

The modern world is full of different kinds of networks. In each network it could be advantageous or disadvantageous to have highly protected vertices. For example, in many online communities, people must be invited in order to participate. If we were to create a tree of users where each person is connected to the person who invited them, then each time a user adds a member they once again become 1-protected. Thus a high level of protection most likely indicates that someone is not regularly bringing in new members.

In a network that must remain secure, it is desirable to make it difficult to reach the root, so it is preferable for the root to be highly protected (another question we

✉ Keith Copenhaver
keithc@ufl.edu

¹ Department of Mathematics, University of Florida, Gainesville, FL 32612, USA

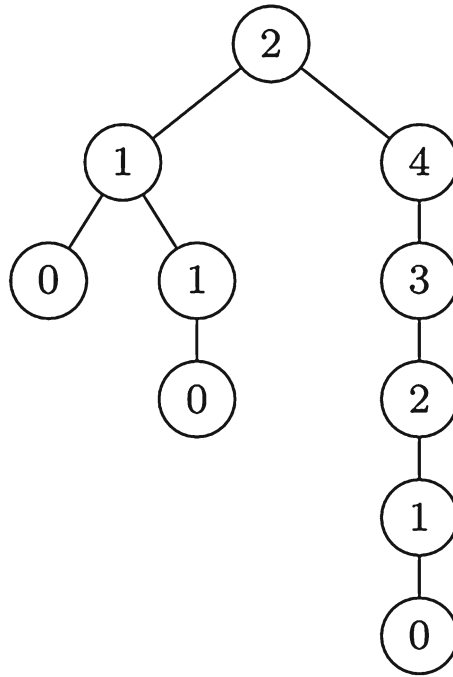


Fig. 1 A tree on 10 vertices with each vertex labeled with the greatest k for which it is k -protected

will address directly), but we also would like nodes in general to have a high level of protection. For example, in a computer network with different levels of access, even if the network is very tall (in other words, the lowest entry point is many steps from the root), if it has a leaf at a high level, the root will still be far more accessible than the height of the network would imply.

A vertex is k -protected if it is at least k upward steps from any leaf. In Fig. 1, each vertex is labeled by the highest k for which it is k -protected, this k is also known as the rank. For example, a vertex that is of rank four is 4-protected, 3-protected, etc. The number of k -protected vertices was explored for unlabeled rooted plane trees with $k = 2$ by Cheon and Shapiro [4]. The topic of protected vertices has been examined for random recursive trees by Mahmoud and Ward [8], k -ary trees by Mansour [9], binary search trees by Bóna [1] and Bóna and Pittel [3], random phylogenetic trees by Bóna and Flajolet [2], several types of random trees by Devroye and Janson [5], and digital search trees by Du and Prodinger [6].

Throughout this paper any reference to a tree will specifically mean an unlabeled rooted plane tree.

We will also freely use a few facts: the number of such trees on n vertices, denoted as $t(n) = \frac{1}{n} \binom{2n-2}{n-1}$ is the $(n - 1)$ st Catalan number, and the ordinary generating function of the number of such trees on n vertices is $\sum_{n=1}^{\infty} t(n)x^n = T(x) = \frac{1-\sqrt{1-4x}}{2}$. The number of all vertices of trees of size n , denoted as $v(n) = \binom{2n-2}{n-1}$, is the $(n - 1)$ st central binomial coefficient, and they have the ordinary generating function $\sum_{n=1}^{\infty} v(n)x^n = V(x) = \frac{x}{\sqrt{1-4x}}$.

2 Grafting

Grafting is a process in horticulture where a branch from one tree is removed and a branch from a different tree is inserted in its place. The following lemma does something very similar.

Lemma 1 *Let $R_k(x) = \sum_{n=1}^{\infty} r_k(n) x^n$ be the ordinary generating function for $r_k(n)$, the number of rooted plane trees on n vertices whose root is k -protected. Let $L(x) = \sum_{n=0}^{\infty} l(n) x^n$ be the ordinary generating function for $l(n)$, the number of leaves in all rooted plane trees of size $n + 1$ (or, equivalently, all rooted plane trees with n edges). Let $T_k(x) = \sum_{n=1}^{\infty} t_k(n) x^n$ be the ordinary generating function for $t_k(n)$, the number of k -protected vertices in all trees of size n . Then $T_k(x) = L(x) \cdot R_k(x)$.*

Proof Let $0 < m \leq n$. The number of k -protected vertices on a tree with n vertices can be counted by choosing a tree on $n - m + 1$ vertices, removing a specific leaf, then replacing that leaf with a tree whose root is k -protected. A leaf can be chosen in one of $l(n - m)$ ways and the tree in $r_k(m)$ ways, so the result follows by the product formula. □

To apply this, of course, we need $L(x)$. Let $T(x, y) = \sum_{n,m} t_{n,m} x^n y^m$ where $t_{n,m}$ is the number of trees on n vertices with m leaves. The generating function $T(x, y)$ satisfies the functional equation $T(x, y) = xy + x \left(\frac{T(x,y)}{1-T(x,y)} \right)$, and solving for $T(x, y)$ yields $T(x, y) = \frac{1}{2} (1 - x + xy - \sqrt{-4xy + (-1 + x - xy)^2})$. The function we want, $L(x)$, will be equal to $\left. \frac{\partial}{\partial y} T(x, y) \right|_{y=1}$. Thus

$$L(x) = \frac{1}{2} \left(1 + \frac{1}{\sqrt{1-4x}} \right). \tag{1}$$

To find expressions for $R_k(x)$, we first observe that since a tree with a 1-protected root is simply a non-empty sequence of trees, then $R_1(x) = \frac{xT(x)}{1-T(x)}$ where $T(x) = \frac{1-\sqrt{1-4x}}{2}$ giving $R_1(x) = \frac{1-2x-\sqrt{1-4x}}{2}$. Further, this can be iterated since a root is k -protected if and only if it is a non-empty sequence of trees whose roots are $(k - 1)$ -protected. Thus the recursion $R_k(x) = x \cdot \frac{R_{k-1}(x)}{1-R_{k-1}(x)}$ holds.

Theorem 1 *For all $k \geq 2$,*

$$R_k(x) = \frac{x^{k-2} (n_k(x) - \sqrt{1-4x})}{2d_k(x)}, \tag{2}$$

where $n_k(x)$ and $d_k(x)$ are polynomials defined as follows: for all $k \geq 2$, $n_k(x) = 1 - 2x - 2x^2 - \dots - 2x^k$, $d_2 = 2 + x$, and for all $k \geq 3$

$$d_k(x) = \sum_{i=0}^{k-3} (i + 1)x^i + \sum_{i=k-2}^{2k-3} (2k + 2 - i)x^i. \tag{3}$$

For some numerical justification, this gives the series expansions $R_2(x) = x^3 + 2x^4 + 6x^5 + 18x^6 + \dots$ and $R_3(x) = x^4 + 2x^5 + 6x^6 + \dots$, each of which are accurate up to trees of size 6 by examination. To prove this theorem we will need some purely computational lemmas.

Lemma 2 For all $k \geq 2$,

$$d_{k+1}(x) = d_k(x) - x^{k-2}n_k(x) + x^{2k-1}. \tag{4}$$

Proof We will proceed by induction on k .

For the base case, if $k = 2$, then

$$\begin{aligned} d_3(x) &= 1 + 3x + 2x^2 + x^3 = 2 + x - (1 - 2x - 2x^2) + x^3 \\ &= d_2(x) - x^{2-2}n_2(x) + x^{2(2)-1}. \end{aligned}$$

For the induction step, we assume that the statement holds for $d_k(x)$, so from (3),

$$\begin{aligned} d_{k+1}(x) &= \sum_{i=0}^{(k+1)-3} (i+1)x^i + \sum_{i=k-1}^{2(k+1)-3} (2(k+1) + 2 - i)x^i \\ &= \sum_{i=0}^{k-3} (i+1)x^i + \sum_{i=k-2}^{2k-3} (2k + 2 - i)x^i - x^{k-2} \left(1 - 2 \sum_{i=1}^k x^i \right) + x^{2k-1} \\ &= d_k(x) - x^{k-2}n_k(x) + x^{2k-1}. \end{aligned}$$

□

Lemma 3 If $k \geq 2$, then

$$n_k^2(x) - (1 - 4x) = 4x^3d_k(x). \tag{5}$$

Proof The expansion of $n_k^2(x)$ splits nicely into 3 parts as follows:

The first two terms will be $1 - 4x$ for all $k \geq 1$.

For all $1 < i \leq k$ we will have a two copies of $-2x^i$ and $i - 1$ copies of $4x^i$ giving a net total of $(i - 2)x^i$. This means there will be no x^2 term.

For all $k < i \leq 2k$ we will have no negative terms, and for each term we will have $2k - 1 - i$ copies of $4x^i$ giving us

$$\begin{aligned} n_k^2(x) &= 1 - 4x + \sum_{i=3}^k (i - 2)4x^i + \sum_{i=k+1}^{2k} (2k - 1 - i)4x^i \\ &= 1 - 4x + 4x^3 \left(\sum_{i=0}^{k-3} (i+1)x^i + \sum_{i=k-2}^{2k-3} (2k + 2 - i)x^i \right). \end{aligned}$$

□

Proof (of Theorem 1) We once again proceed by induction on k .

For the base case, if $k = 2$, then, after clearing denominators and multiplying by the conjugate of the denominator,

$$R_2(x) = \frac{x \cdot R_1(x)}{1 - R_1(x)} = \frac{2x(1 - 2x - 2x^2 - \sqrt{1 - 4x})}{4x(2 + x)} = \frac{x^0(n_2(x) - \sqrt{1 - 4x})}{2d_2(x)}.$$

Now, assuming that the statement holds for $R_k(x)$, and after clearing denominators and multiplying by the conjugate of the denominator,

$$\begin{aligned} R_{k+1}(x) &= \frac{x \cdot R_k}{1 - R_k} \\ &= \frac{x^{k-1}(2d_k(x)n_k(x) - x^{k-2}n_k^2(x) - 2d_k(x)\sqrt{1 - 4x} + x^{n-2}(1 + 4x))}{4d_k^2(x) - 4x^{k-2}d_k(x)n_k(x) + x^{2k-4}n_k^2(x) - x^{2k-4}(1 - 4x)} \\ &= \frac{x^{k-1}(2d_k(x)n_k(x) - x^{k-2}(4x^3d_k(x)) - 2d_k(x)\sqrt{1 - 4x})}{4d_k^2(x) - 4x^{k-2}d_k(x)n_k(x) + x^{2k-4}(4x^3d_k(x))} \\ &= \frac{x^{k-1}(2n_k(x) - 4x^{k+1} - 2\sqrt{1 - 4x})}{4d_k(x) - 4x^{k-2}n_k(x) + 4x^{2k-1}} = \frac{x^{k-1}(n_{k+1}(x) - \sqrt{1 - 4x})}{2d_{k+1}(x)}. \end{aligned}$$

□

3 Asymptotics

We will say that $a_n \sim b_n$ if $\lim_{n \rightarrow \infty} \frac{a_n}{b_n} = 1$.

Theorem 2 (Bender’s Lemma [7]) *Suppose that $A(z) = \sum a_n z^n$ and $B(z) = \sum b_n z^n$ are power series with radii of convergence $\alpha > \beta \geq 0$, respectively. Suppose b_{n-1}/b_n approaches a limit b as n approaches infinity. If $A(b) \neq 0$, then $c_n \sim A(b)b_n$, where $\sum c_n z^n = A(z)B(z)$.*

Lemma 4 *For all $k \geq 1$, $d_k(x)$ is never zero on the closed disk of radius $\frac{4}{15}$.*

Proof Let $|x| \leq \frac{4}{15}$. Then $|d_1(x)| = 1 > 0$, $|d_2(x)| \geq 2 - |x| > 0$, and

$$|d_3(x)| \geq 1 - |3x| - |2x^2| - |x^3| \geq 1 - \frac{4}{5} - \frac{32}{225} - \frac{192}{3375} = \frac{131}{3375}.$$

Observe that, for $k \geq 2$, $|d_k(x) - 1| < |x^{k-2} + \sum_{k=1}^{\infty} (k + 1)x^k|$. Thus

$$\begin{aligned} |d_k(x) - 1| &< \left| x^{k-2} + \sum_{k=1}^{\infty} (k + 1)x^k \right| \\ &\leq \left(\frac{4}{15}\right)^{k-2} + \sum_{k=1}^{\infty} (k + 1) \left(\frac{4}{15}\right)^k = \left(\frac{4}{15}\right)^{k-2} + \frac{165}{196}. \end{aligned}$$

If $k \geq 4$, then $(\frac{4}{15})^{k-2} < \frac{31}{196}$, so that $|d_k(x) - 1| < 1$. Hence $|d_k(x)| \geq 1 - |d_k(x) - 1| > 0$ for all $k \geq 4$. \square

Theorem 3 *Let $t_k(n)$ be the number of vertices which are k -protected in all rooted plane trees of size of n . Then $t_k(n) \sim \frac{3v(n)}{(4^k+2)} = \frac{3\binom{2n-2}{n-1}}{(4^k+2)}$.*

Proof Note that $\frac{x}{\sqrt{1-4x}}$ is the generating function for the number of all vertices of all trees of size n . We have that

$$\begin{aligned} T_k(x) &= L(x) \cdot R_k(x) = \left(\frac{1 + \sqrt{1 - 4x}}{2\sqrt{1 - 4x}} \right) \cdot \left(\frac{x^{k-2} (n_k(x) - \sqrt{1 - 4x})}{2d_k(x)} \right) \\ &= x^{k-2} \cdot \frac{n_k(x) - (1 - 4x) + (n_k(x) - 1)\sqrt{1 - 4x}}{4d_k(x)\sqrt{1 - 4x}} \\ &= \frac{x}{\sqrt{1 - 4x}} \cdot \frac{x^{k-3}(n_k(x) - 1 + 4x)}{4d_k(x)} - \frac{x^{k-2}(1 - n_k(x))}{4d_k(x)} \end{aligned}$$

Since $d_k(x)$ is non-zero on the closed disk of radius $4/15$, it follows that the term on the right is asymptotically irrelevant and that $\frac{x^{k-3}(n_k(x)-1+4x)}{4d_k(x)}$ has a radius of convergence larger than $1/4$ for all k . Thus we can apply Bender’s Lemma to the term on the left for any k , giving

$$[x^n]T_k(x) \sim \binom{2n - 2}{n - 1} \frac{(1/4)^{k-3}(n_k(1/4) - 1 + 4(1/4))}{4d_k(1/4)}. \tag{6}$$

Since $n_{k+1}(1/4) = n_k(1/4) - 2(1/4)^{k+1}$ and $n_1(1/4) = 3/4$, we have $n_k(1/4) = \frac{2+4^k}{3 \cdot 4^k}$. Applying Lemma 3,

$$n_k^2(1/4) - (1 - 4(1/4)) = 4(1/4)^3 d_k(1/4),$$

so that $d_k(1/4) = 16n_k^2(1/4)$. Thus

$$\frac{(1/4)^{k-3}(n_k(1/4) - 1 + 4(1/4))}{4d_k(1/4)} = \frac{1}{4^k \cdot \frac{2+4^k}{3 \cdot 4^k}} = \frac{3}{4^k + 2}.$$

\square

Corollary 1 *Let $p_k(n)$ be the probability that a random vertex in a random rooted plane tree of size n is k -protected. Then $p_k(n) \sim \frac{3}{4^k+2}$.*

Proof To find the probability, we divide by $\binom{2n-2}{n-1}$. \square

This gives the sequence of values $1, 1/2, 1/6, 1/22, 1/86, \dots$, and also shows that as we progress to a higher level of protection, we lose about $1/4$ of the vertices each time. Letting $k = 2$, we re-obtain the result by Cheon and Shapiro [4].

For numerical justification, the proportion of 3-protected vertices in trees of size 50 is

$$\frac{[x^{50}]T_3(x)}{[x^{50}]\frac{x}{\sqrt{1-4x}}} = \frac{88,972,411,304,864,387,146,864,997}{1,959,816,327,613,912,069,440,802,200} \approx 0.0453986$$

and $\frac{1}{22} = 0.045\bar{5}$.

There are some other interesting questions which can be answered using $R_k(x)$. The height of a tree is defined as the longest path from the root to a leaf. The function $R_k(x)$ enumerates instead by the *shortest* path from the root to a leaf.

Theorem 4 *Let $r_k(n)$ be the number of trees on n vertices whose root is k -protected. Then*

$$r_k(n) \sim \frac{9}{4^{1-k} + 4 + 4^k} \cdot t(n)$$

where $t(n)$ denotes the number of trees on n vertices, which is the $(n - 1)$ st Catalan number.

Proof We have that

$$R_k(x) = \frac{x^{k-2}(1 - \sqrt{1 - 4x})}{2d_k(x)} - \frac{\sum_{i=1}^k x^{i+k-2}}{d_k(x)}.$$

The term on the right is once again asymptotically irrelevant, so we have

$$R_k(x) = \frac{x^{k-2}}{d_k(x)} \cdot T(x) - \frac{\sum_{i=1}^k x^{i+k-2}}{d_k(x)}.$$

Thus we have

$$\begin{aligned} \frac{(1/4)^{k-2}}{d_k(1/4)} \cdot t(n) &= \frac{(1/4)^{k-2}}{16n_k^2(1/4)} \cdot t(n) = \frac{1}{4^k \left(\frac{2+4k}{3 \cdot 4^k}\right)^2} \cdot t(n) \\ &= \frac{9}{4^{-k}(4 + 4^{k+1} + 4^{2k})} \cdot t(n) = \frac{9}{4^{1-k} + 4 + 4^k} \cdot t(n). \end{aligned}$$

□

This gives the sequenced of values 1, 1, 4/9, 16/121, 64/1849, ..., and once again shows that we lose about 1/4 of the trees each time we progress to a higher level of protection.

Recall that the *rank* of a vertex is the greatest k for which it is k -protected, or alternatively, the length of the shortest downward path from a vertex to a leaf.

Corollary 2 *The number of vertices of rank k in trees of size n approaches the value*

$$\frac{9\binom{2n-2}{n-1}}{10+4^{1-k}+4^{1+k}}.$$

Proof It follows immediately from the definition that a vertex has rank k if and only if it is k -protected but not $(k + 1)$ -protected, so by Theorem 3, the number of vertices of rank k approaches

$$\begin{aligned} & \frac{3\binom{2n-2}{n-1}}{(4^k + 2)} - \frac{3\binom{2n-2}{n-1}}{(4^{k+1} + 2)} \\ &= 3\binom{2n-2}{n-1} \cdot \left(\frac{4^{k+1} + 2 - 4^k - 2}{4^{2k+1} + 2(4^k) + 2(4^{k+1}) + 4} \right) \\ &= 3\binom{2n-2}{n-1} \cdot \frac{4^k(4-1)}{4^k(4^{k+1} + 2 + 2(4) + 4^{1-k})} = \frac{9\binom{2n-2}{n-1}}{10 + 4^{1-k} + 4^{1+k}}. \end{aligned}$$

□

4 Expectations

Theorem 5 Let $X_{r,n}$ be the random variable whose value is the rank of the root of a tree and whose sample space is the set of trees of size n . Then

$$\mathbb{E}[X_{r,n}] \sim \sum_{k=1}^{\infty} \frac{9}{4^{1-k} + 4 + 4^k} \approx 1.62297. \tag{7}$$

Theorem 6 Let $X_{v,n}$ be the random variable whose value is the rank of a vertex and whose sample space is the set of vertices in all trees of size n . Then

$$\mathbb{E}[X_{v,n}] \sim \sum_{k=1}^{\infty} \frac{3}{4^k + 2} \approx 0.727649. \tag{8}$$

Proof (of Theorem 5) Let $r(n)$ be the sum of the ranks of the roots of all trees of size n . Then $\mathbb{E}[X_{v,n}] = \lim_{n \rightarrow \infty} \frac{r(n)}{i(n)}$. To find $r(n)$, we need only consider ranks of up to $n - 1$, since there are no vertices of rank n or more in trees of size n . Thus $r(n) = \sum_{k=1}^{n-1} [x^n]R_k(x)$, since the trees with roots of rank one are counted once by $R_1(x)$, the trees with roots of rank two are counted once by $R_1(x)$ and once by $R_2(x)$, and similarly the trees with roots of rank n are counted once by each $R_k(x)$ for $1 \leq k \leq n - 1$. Let $E_R(x) = \sum_{n=1}^{\infty} r(n)x^n$. Since the $R_k(x)$ have no terms of degree lower than k , their sum converges as a formal power series. It follows that

$$E_R(x) = \sum_{k=1}^{\infty} R_k(x) = \sum_{k=1}^{\infty} \left(\frac{x^{k-2}(1 - \sqrt{1-4x})}{2d_k(x)} - \frac{\sum_{i=1}^{k-2} x^{i+k-2}}{d_k(x)} \right).$$

The sums $\sum_{k=1}^{\infty} \frac{x^{k-2}}{d_k(x)}$ and $\sum_{k=1}^{\infty} \frac{\sum_{i=1}^{k-2} x^{i+k-2}}{d_k(x)}$ both converge absolutely on the closed disk of radius $1/4$, so we can split the sum as follows:

$$\begin{aligned} \sum_{k=1}^{\infty} R_k(x) &= \sum_{k=1}^{\infty} \left(\frac{x^{k-2}}{d_k(x)} \cdot T(x) - \frac{\sum_{i=1}^{k-2} x^{i+k-2}}{d_k(x)} \right) \\ &= \sum_{k=1}^{\infty} \frac{x^{k-2}}{d_k(x)} \cdot T(x) - \sum_{k=1}^{\infty} \frac{\sum_{i=1}^{k-2} x^{i+k-2}}{d_k(x)}. \end{aligned}$$

Since the coefficient of $T(x)$ and the double sum on the right both converge absolutely to bounded, analytic functions on the closed disk of radius $5/14$, it follows that the term on the right remains asymptotically irrelevant, and that we may apply Bender’s Lemma to the term on the left, and the result follows. \square

Proof (of Theorem 6) The proof is similar to that of Theorem 5 with the roles of $R_k(x)$ and $T(x)$ replaced by $T_k(x)$ and $V(x)$, respectively. \square

For numerical justification, we may compute the n th coefficient using only the $(n - 1)$ st partial sum, since there are no vertices or roots which are n -protected in a tree of size n . We have that

$$\frac{[x^{50}] \sum_{k=1}^{49} R_k(x)}{[x^{50}] T(x)} = \frac{1,874,097,069,430,998,779,470,999}{1,152,833,133,890,536,511,435,766} \approx 1.62564$$

and

$$\frac{[x^{50}] \sum_{k=1}^{49} T_k(x)}{[x^{50}] V(x)} = \frac{4,630,522,930,774,422,812,075,437,903}{6,369,403,064,745,214,225,682,607,150} \approx 0.726995.$$

Acknowledgements I am grateful to Miklos Bóna for guidance on the ins and outs of the submission process, as well as careful reading of the manuscript; my wife, Jaelyn van Wingerden for her many readings and constant support; Arnold Knopfmacher and Aubrey Blecher for their very helpful comments; and the referees for helpful suggestions.

References

1. Bóna, M.: k -Protected vertices in binary search trees. *Adv. Appl. Math.* **53**, 1–11 (2014)
2. Bóna, M., Flajolet, P.: Isomorphism and symmetries in random phylogenetic trees. *J. Appl. Probab.* **46**, 1005–1019 (2009)
3. Bona, M., Pittel, B.: On a random search tree: asymptotic enumeration of vertices by distance from leaves. *J. Appl. Probab.* (to appear)
4. Cheon, G.-S., Shapiro, L.W.: Protected points in ordered trees. *Appl. Math. Lett.* **21**, 516–520 (2008)
5. Devroye, L., Janson, S.: Protected nodes and fringe subtrees in some random trees. *Electron. Commun. Probab.* **19**, 1–10 (2014)
6. Du, R.R.X., Prodinger, H.: On protected nodes in digital search trees. *Appl. Math. Lett.* **25**, 1025–1028 (2012)
7. Flajolet, P., Sedgewick, R.: *Analytic combinatorics*. Cambridge University Press, Cambridge (2009)
8. Mahmoud, H.M., Ward, M.D.: Asymptotic properties of protected nodes in random recursive trees. *J. Appl. Probab.* **52**, 290–297 (2015)
9. Mansour, T.: Protected points in k -ary trees. *Appl. Math. Lett.* **24**, 478–480 (2011)