Journal of Integer Sequences, Vol. 18 (2015),
Article 15.5.8

The Boundary of Ordered Trees

Dennis E. Davenport and Louis W. Shapiro
Mathematics Department
Howard University
Washington, DC 20059
USA
dennis.davenport@howard.edu
lshapiro@howard.edu

Lara K. Pudwell
Department of Mathematics and Statistics
Valparaiso University
Valparaiso, IN 46383
USA
lara.pudwell@valpo.edu

Leon C. Woodson
Department of Mathematics
Morgan State University
Baltimore, MD 21251
USA
leon.woodson@morgan.edu

Abstract

In this paper we compute the distribution of several statistics on the set of rooted ordered trees. In particular, we determine the number of boundary edges, the number of singleton boundary edges, and the analogous values when edges may take on one of k colors.

1 Introduction

In this paper we consider properties of trees. A *rooted ordered tree* is defined recursively as having a root and an ordered set of subtrees. A tree with n edges necessarily has $n+1$ vertices, so we may count trees according to either edges or vertices and obtain the same results. Rooted ordered trees appear naturally as data structures in computer science, but they may also appear as models for real-world objects such as river networks or family trees.

In our investigation several sequences were found which were already in the Online Encyclopedia of Integer Sequences (OEIS) [13]; the A-numbers in this paper are from this source.

Much enumerative work has already been done with ordered trees. The number of ordered trees with n edges is $C_n = \frac{1}{n+1}\binom{2n}{n}$, the n^{th} Catalan number. Recall that the generating function for the Catalan numbers is $C(z) = \sum_{n=0}^{\infty} C_n z^n$. It is well known that $C(z) = 1 + zC^2(z) = \frac{1}{1-zC(z)}$ and that

$$C(z) = \frac{1-\sqrt{1-4z}}{2z} = 1 + 1z + 2z^2 + 5z^3 + 14z^4 + 42z^5 + 132z^6 + \cdots (\text{A000108}).$$

We will denote $C(z)$ by C when this causes no confusion. A companion generating function is $B(z) = B = \sum_{n=0}^{\infty} \binom{2n}{n} z^n = 1 + 2zBC = \frac{1}{1-2zC} = \frac{C}{1-zC^2}$. Here,

$$B(z) = \frac{1}{\sqrt{1-4z}} = 1 + 2z + 6z^2 + 20z^3 + 70z^4 + 252z^5 + 924z^6 + \cdots (\text{A000984}).$$

It turns out that B is the generating function for ordered trees with one marked vertex. This is immediate since a tree with n edges has $n+1$ vertices and $\sum_{n=0}^{\infty}(n+1)C_n z^n = \sum_{n=0}^{\infty}(n+1)\frac{1}{n+1}\binom{2n}{n} z^n = \sum_{n=0}^{\infty}\binom{2n}{n} z^n = B$. A related result is that the number of trees with a marked leaf (outdegree 0) has the generating function $\frac{B}{C} = \frac{B+1}{2}$. A second companion generating function is the one for the Fine numbers, defined by $F(z) = F = \frac{C}{1+zC} = \frac{1}{1-z^2C^2} = 1 + 1z^2 + 2z^3 + 6z^4 + 18z^5 + 57z^6 + \cdots$ (A000957) or equivalently $C = \frac{F}{1-zF}$. These three sequences are known to count a number of different combinatorial objects. For example, the Catalan numbers have over 200 different combinatorial interpretations [12]. The Fine numbers also count many things, including ordered trees with even root degree and Dyck paths with no hills [4]. These three sequences appear in relation to one another in several contexts [3, 6, 7, 11], and they appear in results throughout this paper as we consider specific characteristics of trees.

In particular, we are concerned with the boundaries of trees. The *right boundary* of an ordered tree is the rightmost path that emanates from the root and terminates at the rightmost leaf. The *left boundary* is defined similarly. The *boundary* of an ordered tree is the union of the left and right boundaries. We measure the length of a boundary in terms of its number of distinct edges. To illustrate the boundary, consider the ordered tree in Figure 1. The vertices are labeled to make the description easier. The path a, b, e, h (the leftmost path emanating from the root, a, and terminating at the leftmost leaf, h) is the left boundary and the path a, d, g is the right boundary. Together they make up the boundary of the tree. The left boundary has length 3, the right boundary has length 2, and the total boundary has length 5.

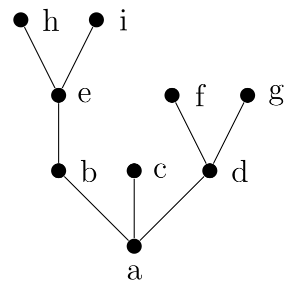

Figure 1: An ordered tree with 8 edges

In applications, the tree boundary may contain important information about the rest of a system. If the ordered tree represents a family tree, the left boundary carries information about the chain of oldest descendants. If the ordered tree represents a physical river system, the boundary may be the most useful region to take measures preventing pollution from outside sources. It is also natural to color boundary edges. For example, if the ordered tree represents a river network a boundary edge might be clean, polluted, or very polluted. If we are looking at supply lines, we might consider an edge as safe or vulnerable.

Another interesting subcase is to restrict such considerations to edges without siblings, i.e., edges above a vertex of updegree 1. We call such edges *singletons* or *single edges*. In Figure 1, edge (b, e) is the only single edge. All trunks are composed of single edges. In Figure 2, tree (i) has three single edges, tree (ii) one, tree (iii) none, and trees (iv) and (v) one. In applications, single edges may carry extra importance for the system. For example, a single supply line or river branch may require extra fortification. We also consider the case where only singleton edges on the boundary may be k-colored.

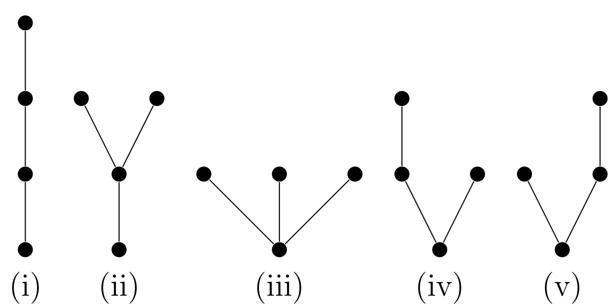

Figure 2: The 5 trees with 3 edges

The *trunk* of a tree is a path from the root to the first vertex with updegree greater than

1 or to a leaf. For example, if we consider the 5 trees with 3 edges given in Figure 2, the first two trees have trunks, while the others do not. Of course a trunk would be included in the boundary. A canopy is the portion of the tree that is above the trunk. In Figure 2, tree (i) has no canopy, tree (ii) has both a trunk and a canopy, while trees (iii), (iv), and (v) are canopies with no trunk. We call a tree with no trunk a bush.

In Section 2 we focus on the left boundary. We compute the total number of left boundary edges in the set of n-edge rooted ordered trees as well as the number of singleton edges on left boundaries. In Section 3, we generalize our results to the total number of edges on the (right and left) boundary. In Section 4 we specialize our results to when the number of available colors is either $k = 1$ or $k = 2$. Our computations enable us to determine the average values of such statistics in an arbitrarily chosen tree with n edges. Finally, in Section 5 we present bijections between trees and enriched Dyck paths that give alternate explanations for some of the results of Section 2.

2 Left boundary edges

We start by examining the left boundary edges. Two questions occur naturally. If we allow k colors for all left boundary edges how many trees do we have? Also how many edges are on the left boundary of all such trees with n edges? The answers to these questions will allow us to determine the average length of the left boundary and the average number of left boundary singleton edges among all n-edge trees with k-colored left boundaries.

Theorem 1. *Let T_k be the generating function counting all ordered trees with each left boundary edge one of k colors. Then $T_k = 1 + kzT_kC = \frac{1}{1-kzC}$*

Proof. In Figure 3, either we have the trivial tree, or there is a leftmost edge connected to the root. C represents an arbitrary subtree emanating from the root. Converting the picture to generating functions gives the proof. $\square$

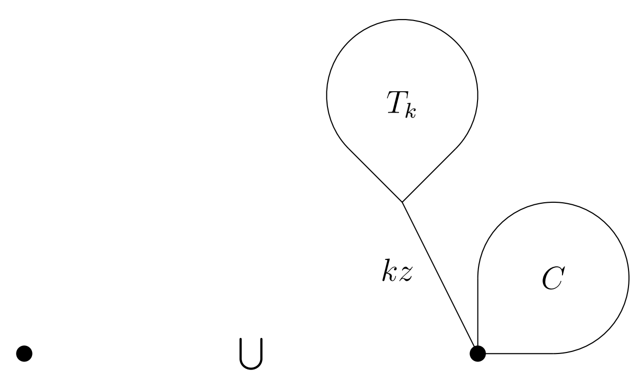

Figure 3: Marking trees with a k-colored left boundary

If $k = 0$, then $T_0 = 1$, counting the trivial tree. If $k = 1$, then we are just counting ordered trees and indeed $T_1 = \frac{1}{1-zC} = C$. If $k = 2$, then $T_2 = \frac{1}{1-2zC} = B$.

Now that we have counted trees with k-colored left boundary edges, we count the number of edges on the left boundary.

Theorem 2. *If $\mathcal{E}_k$ is the generating function for the number of boundary edges on the left boundary, then $\mathcal{E}_k = kzCT_k^2$.*

We present two approaches to compute $\mathcal{E}_k$.

Proof 1. The first proof is to mark one of the left boundary edges with kz and use Figure 4 to determine the generating function. □

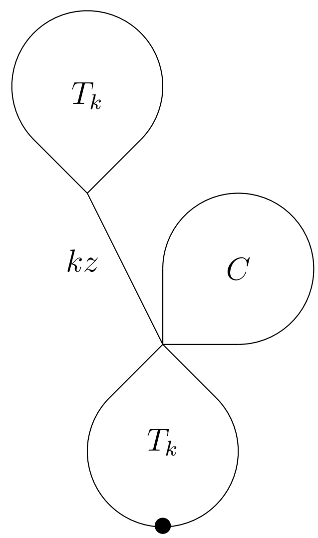

Figure 4: Marking left boundary edges

Proof 2. A second proof is to mark each edge on the left boundary with a ku. Then let $E(z, u) = E = \sum_t u^{m_t} z^{n_t}$, where the sum is taken over all rooted trees t, m_t is the number of edges on the left boundary of t, and n_t is the the number of edges of t. We get $E = 1 + kzuEC = \frac{1}{1-kzuC}$. Note that $E(z, 1) = T_k$.

If there are m edges on the left boundary, we want a contribution of m so we differentiate with respect to u to get

$$E_u = \frac{\partial E}{\partial u} = \frac{kzC}{(1 - kzuC)^2} = kzCE^2$$

Note that $\mathcal{E}_k = E_u|_{u=1}$. This gives a second proof of the result.

□

By choosing particular values for k, we obtain the following corollaries.

Corollary 3. *The generating function for the number of left boundary edges for uncolored ordered trees is* $zC^3 = z + 3z^2 + 9z^3 + 28z^4 + 90z^5 + 297x^6 + \cdots = \sum_{n\geq 1} \frac{3}{2n+1}\binom{2n+1}{n-1} z^n = \sum_{n\geq 1} \frac{3n}{n+2} C_n z^n$ *(A000245) and the average length of the left boundary is* $\frac{3n}{n+2}$*, which approaches* 3 *as* n *goes to* ∞.

Proof. Let $k = 1$. □

Corollary 4. *The generating function for the number of left boundary edges of ordered trees with bicolored edges on the left boundary is* $2zB^2C = B(2zBC) = B(B-1) = B^2 - B = \sum_{n\geq 1}(4^n - \binom{2n}{n})z^n = 2z + 10z^2 + 44z^3 + 186z^4 + 772z^5 + 3172z^6 + \cdots$ *(A068551). Thus the average length of the left boundary is* $\frac{4^n - \binom{2n}{n}}{\binom{2n}{n}} \sim \sqrt{\pi}n - 1 \sim \sqrt{\pi}n$.

Proof. Let $k = 2$ and observe that $\binom{2n}{n} \sim \frac{4^n}{\sqrt{\pi n}}$. □

Next, we modify our work to count trees where the only edges that are k-colored are singleton edges on the left boundary.

Theorem 5. *Let* $\widetilde{T}_k$ *be the generating function for the number of trees where the only edges that are* k*-colored are singleton edges on the left boundary. Then,*

$$\widetilde{T}_k = 1 + kz\widetilde{T}_k + z\widetilde{T}_k(C-1) = \frac{1}{1 - kz - z(C-1)} = \frac{1}{1 - kz - z^2C^2}.$$

Proof. There are three possible cases, the trivial tree, a tree with a trunk, or a bush. We mark each singleton edge with a kz.

Using Figure 5 we get $\widetilde{T}_k = 1 + kz\widetilde{T}_k + z\widetilde{T}_k(C-1) = \frac{1}{1-kz-z(C-1)} = \frac{1}{1-kz-z^2C^2}$. As before, 1 is for the trivial tree, while $C - 1$ gives us a nontrivial subtree on the right so the leftmost edge at the root is not a singleton. □

Here the case when $k = 0$ is nontrivial. $\widetilde{T}_0$ counts trees where there are no singleton edges on the left boundary. The generating function is $\widetilde{T}_0 = \frac{1}{1-z^2C^2} = F = 1 + z^2 + 2z^3 + 6z^4 + 18z^5 + 57z^6 + \cdots$ and we have the Fine numbers. This occurrence of the Fine numbers appears to be new to the literature.

For other values of k, we obtain other nice generating functions.

When $k = 1$,

$$\widetilde{T}_1 = \frac{1}{1 - z - z^2C^2} = C,$$

which counts ordered trees, as expected.

When $k = 2$,

$$\widetilde{T}_2 = \frac{1}{1 - 2z - z^2C^2} = C^2 = zC,$$

which is the sequence of Catalan numbers with shifted indexing.

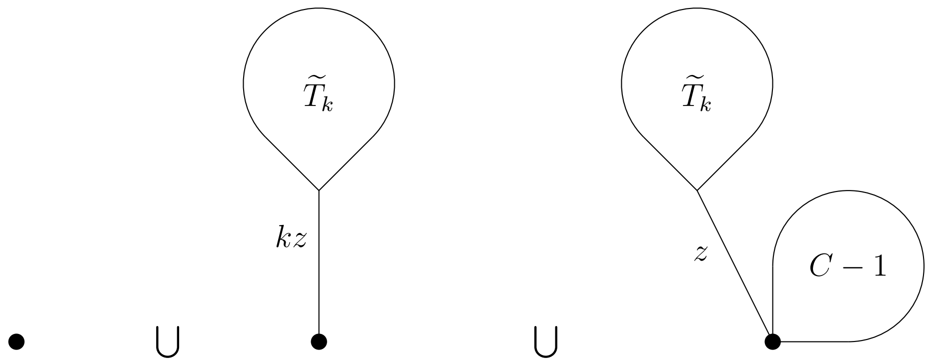

Figure 5: Marking k-colored singleton boundary edges

When $k = 3$,

$$\widetilde{T}_3 = BC = \sum_{n\geq 0} \binom{2n+1}{n+1} z^n = 1+3z+10z^2+35z^3+126z^4+462z^5+1716z^6+\cdots \text{ (A001700)},$$

which is known the enumerate a variety of combinatorial objects ranging from ordered trees to lattice paths [5, 8].

We give bijective explanations of these three results in Section 5.

Finally, we count the total number of singleton left boundary edges in n-edge trees counted by $\widetilde{T}_k$. If we also mark each singleton edge with a u, we get

$$\widetilde{E}(z,u) = \frac{1}{1-uz-z^2C^2} = 1+uz+(u^2+1)z^2+(u^3+2u+2)z^3+(u^4+3u^2+4u+6)z^4+\cdots.$$

Arranging these coefficients in matrix form gives

$$\begin{bmatrix} 1 & & & & & \\ 0 & 1 & & & & \\ 1 & 0 & 1 & & & \\ 2 & 2 & 0 & 1 & & \\ 6 & 4 & 3 & 0 & 1 & \\ 18 & 13 & 6 & 4 & 0 & 1 \\ & & \cdots & & & \end{bmatrix}.$$

The generating function for the k^{th} column ($k = 0, 1, 2, \ldots$) is $z^k F^{k+1}$, counting the number of trees with k singleton edges on the left branch. The row sums are the Catalan numbers since every tree has some number of left edge singletons. We could multiply the matrix by the column vector $[1, 1, 1, \ldots]^T$ to obtain the row sums. Also the total number of left boundary singletons is C_n, for $n \geq 1$ (i.e., multiply by $[0, 1, 2, 3, \ldots]^T$). In Riordan group terminology

$$(F, zF) * \frac{z}{(1-z)^2} = \frac{FzF}{(1-zF)^2} = zC^2 = C - 1.$$

If all ordered trees with n edges are equally likely, then there is one left boundary singleton per tree on average.

As before we can compute $\widetilde{\mathcal{E}}_k = kz\widetilde{T}_k^2$ by marking a singleton edge on the left boundary and then looking at Figure 6.

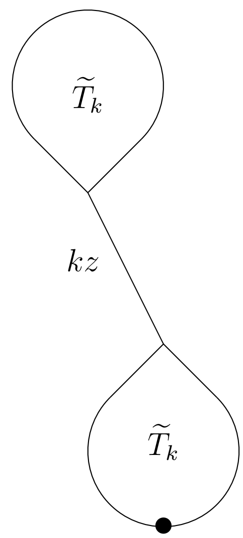

Figure 6: Marking a singleton left boundary edge

For a second approach we mark each singleton boundary edge by kuz, then differentiate with respect to u to find the number of singleton left boundary edges. Note that

$$\widetilde{E}_u = \frac{\partial \widetilde{E}}{\partial u} = -(1 - kuz - z(C-1))^{-2}(-kz), \text{ and } \widetilde{E}_u|_{u=1} = kz\widetilde{T}_k^2.$$

For a third approach we note, briefly, that the corresponding Riordan group array would be

$$\begin{bmatrix} 1 & & & & & \\ 0 & k & & & & \\ 1 & 0 & k^2 & & & \\ 2 & 2k & 0 & k^3 & & \\ 6 & 4k & 3k^2 & 0 & k^4 & \\ 18 & 13k & 6k^2 & 4k^3 & 0 & k^5 \\ & & \cdots & & & \end{bmatrix} = (F, zkF)$$

If $k = 3$ we have $(F, 3zF) * \frac{1}{1-z} = \frac{F}{1-3zF} = BC$.

Thus, we have three different explanations of the following result:

Theorem 6. *Let $\widetilde{\mathcal{E}}_k$ be the generating function for the number of singleton edges on the left boundary tree with k-colored singleton edges on the left boundary. Then $\widetilde{\mathcal{E}}_k = kz\widetilde{T}_k^2$.*

3 Total number of boundary edges

We now generalize from left boundaries to total boundaries of trees. Consider the set of ordered trees where the boundary edges can be one of k colors. We denote the generating function for the number of such trees according to number of edges as H_k. The companion generating function where only the singleton boundary edges are colored will be denoted $\widetilde{H}_k$.

Using Figure 7 we get the following defining equation for H_k.

$$H_k = 1 + kzH_k + k^2z^2T_k^2C = \frac{1}{1-kz}(1 + k^2z^2T_k^2C).$$

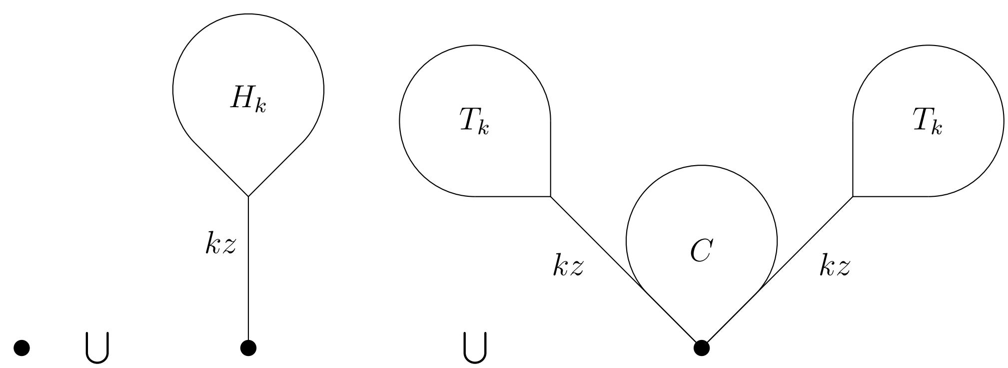

Figure 7: Marking left and right boundary edges

Using similar methods to find $\widetilde{H}_k$ we get,

$$\widetilde{H}_k = 1 + kz\widetilde{H}_k + z^2\widetilde{T}_k^2C = \frac{1}{1-kz}(1 + z^2\widetilde{T}_k^2C).$$

The factor $\frac{1}{1-kz}$ is the generating function for the stem. On top of the stem is either the trivial tree or a branch point. Note that $H_1 = \widetilde{H}_1 = C$, as expected.

We now find a generating function counting the boundary edges. To do so, we can add the left boundary edges to the dual right boundary edges. Since the stem would be counted twice, we subtract the stem edges. We denote the generating function that counts k colored boundary edges by M_k and the generating function counting k colored singleton boundary edges by $\widetilde{M_k}$. Using Theorem 6, we get the following.

$$M_k = 2kzCT_k^2 - \frac{kz}{(1-kz)^2}\left(1 + (kz)^2 T_k^2 C\right) \tag{1}$$

To find $\widetilde{M_k}$ it is simpler to first find the stem contribution, $\widetilde{S}_k$, and the canopy contribution, $\widetilde{V}_k$, and use the fact that $\widetilde{M_k} = \widetilde{S}_k + \widetilde{V}_k$.

Now

$$\widetilde{S}_k = \frac{kz}{(1-kz)^2}\left(1 + 2z^2 \widetilde{E}_k C \widetilde{T}_k^2\right) \quad \text{and} \quad \widetilde{V}_k = \frac{1}{1-kz}\left(2z^2 \widetilde{E}_k C \widetilde{T}_k^2\right).$$

Hence

$$\begin{aligned}\widetilde{M_k} &= \widetilde{S}_k + \widetilde{V}_k \\ &= \frac{kz}{(1-kz)^2}\left(1 + 2z^2 \widetilde{E}_k C \widetilde{T}_k^2\right) + \frac{1}{1-kz}\left(2z^2 \widetilde{E}_k C \widetilde{T}_k^2\right).\end{aligned}$$

Going back to Figure 2, we note that only one edge is not a boundary edge. So, an obvious question is, what proportion of edges are on the boundary in larger trees? We first consider the case when $k = 1$. Lemma 7 can be proved using the Equation 1 and the fact that $T_1 = C$.

Lemma 7. $M_1 = \frac{zC}{1-z} + \frac{2z^2C^4}{1-z}$

For some of our asymptotic estimates we use the following result of Bender.

Lemma 8 (Bender's lemma [1, 9])**.** *Let $A(z)$ and $B(z)$ be generating functions with radii α and β respectively, with $\alpha > \beta$. Let $C(z) = A(z)B(z)$ and suppose further $\frac{b_{n-1}}{b_n}$ approaches β as $n \to \infty$. If $A(\beta) \neq 0$, then $c_n \sim A(\beta) b_n$.*

Theorem 9. *When k=1, let R_n be the proportion of boundary edges among all edges for trees with n edges. Then*

$$R_n \sim \frac{17}{3n}.$$

Proof. The total number of edges in the set of all trees with n edges is nC_n. Using Lemma 7, the total number of boundary edges is

$$[z^n]\left(\frac{zC}{1-z} + \frac{2z^2C^4}{1-z}\right).$$

Using Bender's lemma, we obtain

$$[z^n]\left(\frac{zC}{1-z} + \frac{2z^2C^4}{1-z}\right) \sim \frac{\frac{1}{4}C_n}{1-\frac{1}{4}} + \frac{2\cdot\left(\frac{1}{4}\right)^2}{1-\frac{1}{4}}[z^n]C^4 = \frac{1}{3}C_n + \frac{1}{6}[z^n]C^4.$$

Now

$$[z^n]C^4 = \quad \frac{4}{2n+4}\binom{2n+4}{n} = \quad \frac{2}{n+2}\frac{(2n+4)(2n+3)(2n+2)(2n+1)(2n)!}{(n+4)(n+3)(n+2)(n+1)!n!} \sim 32C_n$$

Thus we get

$$[z^n]\left(\frac{zC}{1-z}+\frac{2z^2C^4}{1-z}\right) \sim \left[\frac{1}{3}+\frac{32}{6}\right]C_n = \frac{17}{3}C_n$$

Hence $R_n \sim \frac{17}{3n}$. □

For numerical reassurance we compare the two ratios, boundary edges to all edges, and $\frac{17}{3n}$ when $n = 100$. The ratio of these two ratios is approximately 0.97805.

We can also count boundary edges exactly. The average length of the left (resp. right) boundary is 3, as shown in Corollary 3. Similarly, to count edges in the stem, consider the trees in Figure 8. Then we obtain

$$(C - zC)\left(0 + z + 2z^2 + 3z^3 + \cdots\right) = C(1-z)\frac{z}{(1-z)^2} = \frac{zC}{1-C}$$

as the generating function counting the total number of stem edges in the set of ordered trees with n edges. A quick computation shows that

$$[z^n]\frac{zC}{1-C} = \sum_{i=0}^{n-1} C_i \text{ (A014137)},$$

and therefore the average number of stem edges as n approaches infinity is

$$\lim_{n\to\infty}\frac{\sum_{i=0}^{n-1} C_i}{C_n} = \frac{1}{3}.$$

Putting these two enumerations together, the average size of the boundary approaches $3 + 3 - \frac{1}{3} = \frac{17}{3}$ as n grows arbitrarily large.

4 Special cases

In this section we look at the special cases when $k = 1$ and $k = 2$ and we find the proportion of singleton boundary edges in these cases.

Theorem 10. *Let $\widetilde{M}_1$ be the generating function for the number of singleton boundary edges in all n-edge trees. Then,*

(a) $\widetilde{M}_1 = \frac{zC}{1-z} + \frac{2z^3C^4}{1-z} = z + 2z^2 + 6z^3 + 19z^4 + 61z^5 + 199z^6 + 661z^7 + 2234z^8 + \cdots$ (A228180).

(b) Let n be a positive integer, then the proportion of single edges that are on the boundary among all single edges is asymptotic to $\frac{5}{3n}$.

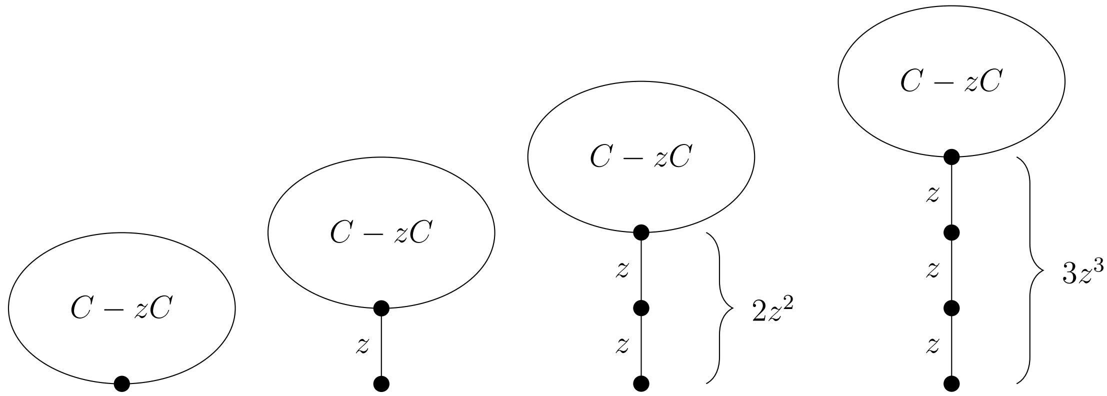

Figure 8: Counting stem edges

Proof. For (a), we consider the number of single boundary edges on a bush. The generating function for a single edge is simply z. The generating function counting boundary edges of a bush is $2z^2C^4$. Hence, the number of single edges on the boundary of a bush has generating function $2z^3C^4$. If the tree has only a trunk, then the generating function is $\frac{z}{(1-z)^2}$. Hence

$$\begin{aligned}\widetilde{M}_1 &= \frac{1}{1-z}\left(2z^3C^4\right) + \frac{z}{(1-z)^2}\left(1+z^2C^3\right)\\ &= \frac{2z^3C^4}{1-z} + \frac{zC}{1-z}.\end{aligned}$$

For (b), there is a simple bijection for counting single edges. Any single edge can be contracted to a vertex and conversely any marked vertex can be replaced by a single edge. Thus the generating function counting all single edges is $zB = \sum_{n=0}^{\infty} \binom{2n-2}{n-1} z^n$. The proportion $\frac{[z^n]\widetilde{M}_1}{[z^n]zB}$, by (a) and Bender's lemma, is asymptotic to $\frac{5}{3n}$. For related results see [2]. □

The proof of the following theorem is an application of Bender's lemma and is similar to the proof of Theorem 10, hence it is omitted.

Theorem 11. *The proportion of single boundary edges among all boundary edges in an n-edge tree approaches 5/17 as n goes to ∞.*

We now consider trees whose boundary edges are bicolored, i.e., $k = 2$. Recall that $T_2 = B = \sum_{n=0}^{\infty} \binom{2n}{n} z^n = \frac{1}{\sqrt{1-4z}} = 1 + 2zCB = \frac{1}{1-2zC}$.

This is the generating function for the central binomial coefficients. Thus we have the easily proven, yet possibly novel, result that the number of ordered trees with bicolored boundary edges allowed on the left branch is $\binom{2n}{n}$ for such trees with n edges.

Consider next trees where the boundary edges are bicolored and the root degree is at least two. In other words, bushes.

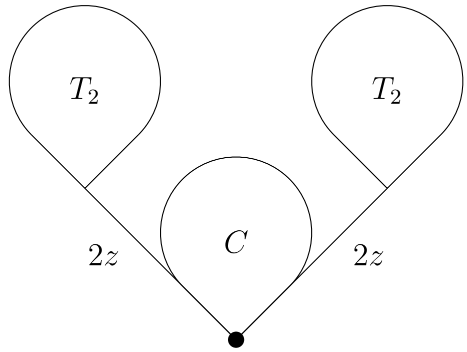

Figure 9: Marking edges in bushes

Using Figure 9, the generating function for the number of bushes with bicolored boundary edges is $(2z)^2 T_2^2 C = 4z^2 B^2 C = 4z^2 + 20z^3 + 88z^4 + 372z^5 + 1544z^6 + 6344z^7 + 25904z^8 + \cdots$.

In the case that the root degree is one we have a stem, a maximal chain of vertices, each of updegree 1, starting at the root. For this case the generating function is just $\frac{1}{1-2z} = \sum_{n=0}^{\infty} 2^n z^n$. Using the generating functions counting boundary edges on bushes and stems we get the following theorem.

Theorem 12. *Let H_2 be the generating function for the number of trees where the boundary edges are bicolored. Then,*

$$\begin{aligned} H_2 &= \frac{1}{1-2z}(1 + 4z^2 B^2 C) \\ &= \frac{1}{1-2z}(1 - 2zB + 2zB^2) \\ &= 1 + 2z + 8z^2 + 36z^3 + 160z^4 + 692z^5 + \cdots \text{ (A228197)}. \end{aligned}$$

Finally, consider the case where only boundary singleton edges can be bicolored. The computations in this case are slightly subtler but follow the same pattern. This situation is illustrated in Figure 10. We have

$$\widetilde{T}_2 = 1 + 2z\widetilde{T}_2 + z\widetilde{T}_2(C-1) = \frac{1}{1-2z-z(C-1)} = \frac{1}{1-z-zC} = C^2.$$

Thus using the same techniques as before, the bushes have generating function

$$z^2(\widetilde{T}_2)^2 C = z^2 C^5 = z^2 + 5z^3 + 20z^4 + 75z^5 + 275z^6 + 1001z^7 + \cdots \text{ (A000344)}.$$

The generating function for the stem is again $\frac{1}{1-2z}$. Using these generating functions we get the following theorem.

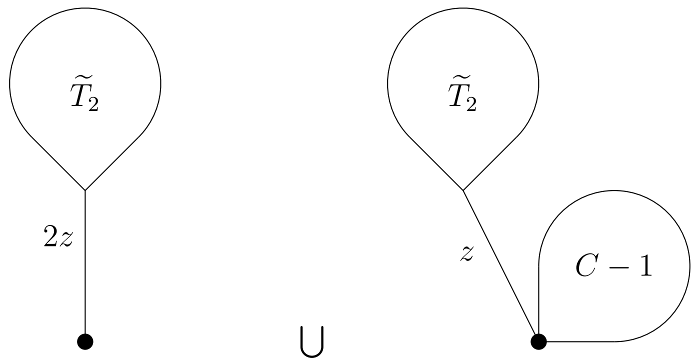

Figure 10: Marking singleton boundary edges

Theorem 13. *Let $\widetilde{H}_2$ be the generating function for the number of trees where the singleton boundary edges are bicolored. Then,*

$$\widetilde{H}_2 = \frac{1}{1-2z}(1+z^2C^5) = 1 + 2z + 5z^2 + 15z^3 + 50z^4 + 165z^5 + \cdots \text{(A228343)}.$$

There is much more that could be done along these lines. We could look at other related classes of trees where the updegree of each vertex is 0, 1, or 2 (Motzkin trees) or updegree must be even (even trees). We could change from left boundary to leftmost principal branch. However, we conclude this section with a curiosity, answering a question that may not have been bothering you. Where does the generating function C^7 occur in a "natural" setting? If we consider bushes (root degree ≥ 2) with bicolored singleton edges on the boundary, then the generating function counting the number of singleton boundary edges is $4z^3C^7$. The factor z^2C sets up the base of a bush. To count the number of left boundary edges, we have

$$\widetilde{\mathcal{E}}_k = 2z\widetilde{T}_k^2 = 2zC^4$$

and

$$\widetilde{T}_k = C^2$$

possible trees on the right branch. This gives us

$$z^2C(2zC^4)C^2$$

edges on the left side. And doubling this gives us

$$4z^3C^7 = 4z^3 + 28z^4 + 140z^5 + 616z^6 + \cdots = 4z^3\sum_{n=0}^{\infty}\frac{7}{2n+7}\binom{2n+7}{n}z^n.$$

5 Bijections

In this final section, we provide bijections to illustrate some of our previous results, namely that $\widetilde{T}_1 = C, \widetilde{T}_2 = C^2$, and $\widetilde{T}_3 = BC$. Recall from Section 2 that $\widetilde{T}_k$ is the generating function for the number of trees where the only edges that are k-colored are singleton edges on the left boundary.

The first bijection relates the edges of the left boundary to subtrees connected to the root, as shown in Figure 11. Here, for each edge E in the left boundary, let E^* be the subtree whose root is the lower vertex of E but not including edge E. If the original tree has ℓ edges in the left boundary, we obtain a new tree with ℓ edges emanating from the root. Each of these edges E is the stem for a subtree corresponding to E^*. Thus, $\widetilde{T}_1 = C$.

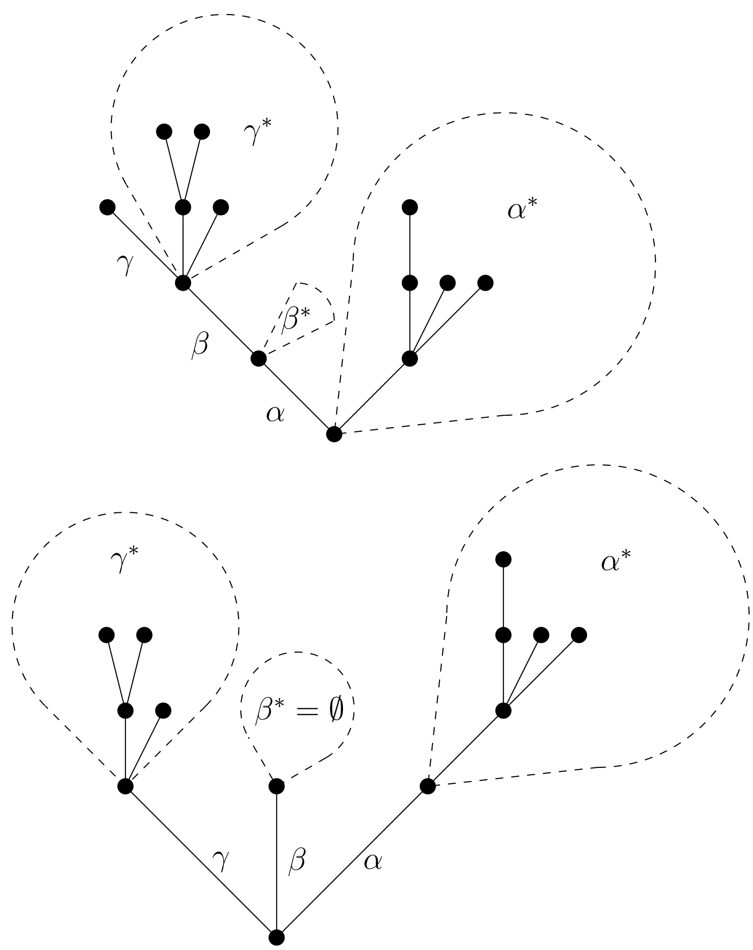

Figure 11: A bijection between left boundary edges and edges containing the root

The second bijection is well-known, called such things as "the worm climbing the tree", "the glove bijection", or "preorder transversal" [12]. As in the first bijection, edges on the left boundary are converted the edges emanating from the root, but then we take the transformation one step further, as shown in Figure 12. Now, edges emanating from the root are mapped to steps touching the x-axis in a Dyck path. Singletons edges on the left boundary in Figure 11 correspond to UD subpaths starting and ending on the x-axis in Figure 12. We call such UD subpaths *hills*.

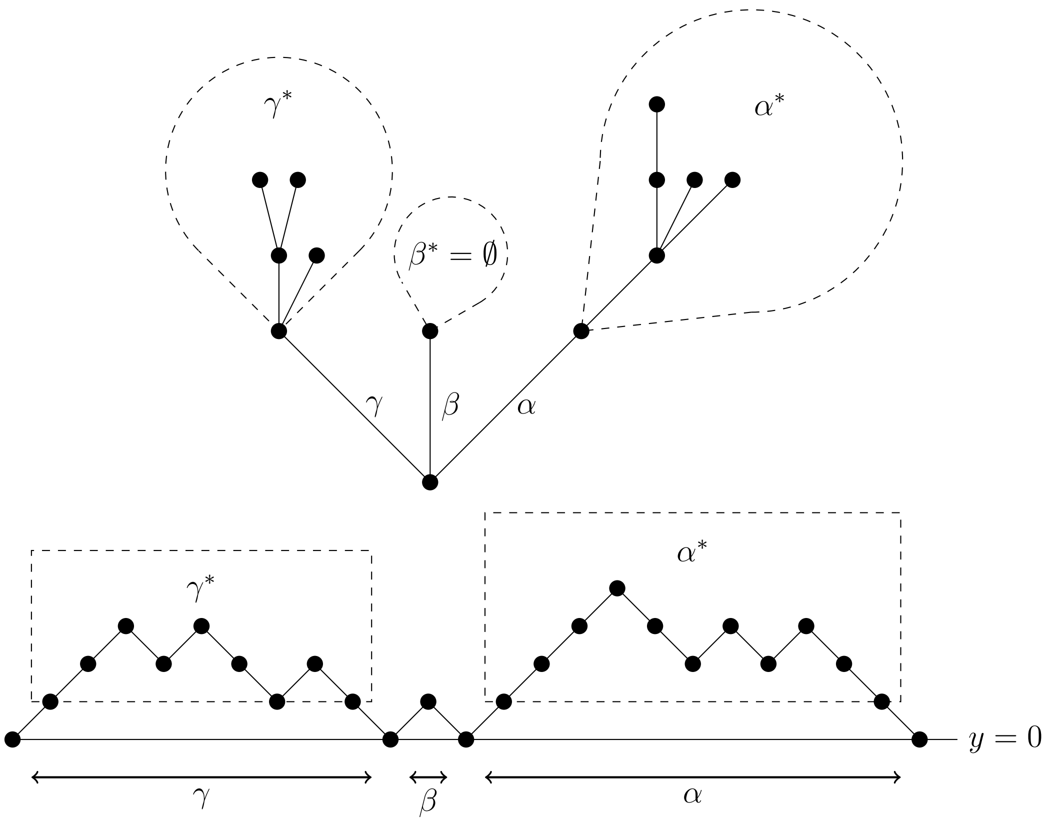

Figure 12: The worm climbing the tree

Now, towards understanding $\widetilde{T}_2$, suppose that singletons on the left boundary of a tree are bicolored. This is equivalent to having bicolored UD hills in our Dyck paths. Suppose each such UD hill is either green or red, and replace the red hills with DU, as in Figure 13. We now have an "enriched" Dyck path that has the freedom to go as low as the line $y = -1$. The number of such "enriched" Dyck paths has the generating function C^2.

Finally, keeping this framework, we tricolor singleton boundary edges so that the hills can now be green, red, or purple. Each time a purple hill is encountered we change UD to UU. If there are ℓ purple hills, the final height of the path will be 2ℓ. In Figure 14, the purple hills are denoted by the double red lines.

Suppose now we have ordered trees where the singleton edges can be either green, red, or purple. Using the first two bijections, these trees correspond to Dyck paths with tricolored

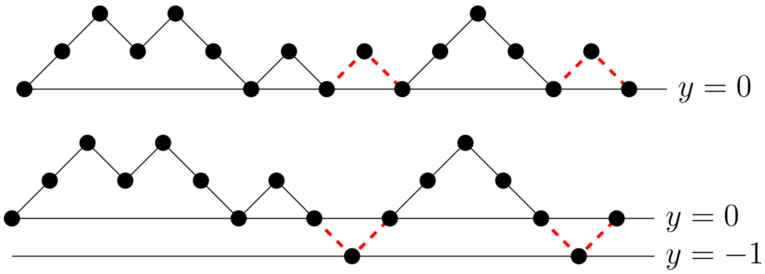

Figure 13: A bijection from bicolored left singletons to enriched Dyck paths

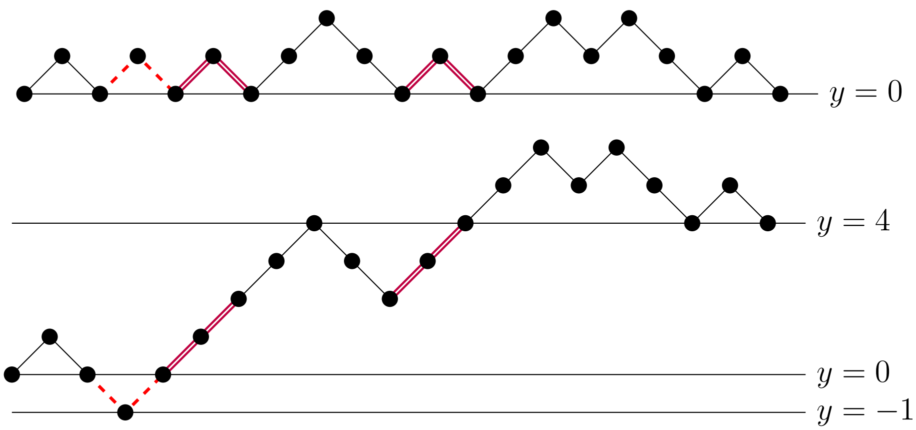

Figure 14: A bijection from tricolored left singletons to paths

hills. We traverse the path from left to right. When we encounter a green hill we do nothing. When we encounter a red hill, we change the steps UD to DU. This gives us enriched Dyck paths as in Figure 13. When we encounter a purple hill, we change UD to UU. Each purple hill now gives a path that ends 2 units higher. If there are ℓ purple hills, the path ends at the point $(2n, 2\ell)$ and the number of such paths is $\frac{2\ell+2}{n+\ell+2}\binom{2n+1}{n-\ell} = \binom{2n+1}{n-\ell} - \binom{2n+1}{n-\ell-1}$. Here, the last equality is found by using the reflection principle of "good paths come from all paths minus bad paths". Now summing over ℓ gives the telescoping sum $\sum_{\ell=0}^{n}\binom{2n+1}{n-\ell} - \binom{2n+1}{n-\ell-1} = \binom{2n+1}{n}$ and $\sum_{n\geq 0}\binom{2n+1}{n}z^n = BC$. Thus, $\widetilde{T}_3 = BC$.

An alternate proof using the Riordan group (see [10]) we have

$$\left(C^2, zC^2\right) = \begin{bmatrix} 1 & 0 & 0 & 0 & \cdots \\ 2 & 1 & 0 & 0 & \\ 5 & 4 & 1 & 0 & \\ 14 & 14 & 6 & 1 & \\ & & \cdots & & 1 \end{bmatrix}.$$

So by the fundamental theorem of Riordan arrays, we have

$$\left(C^2, zC^2\right) * \frac{1}{1-z} = \frac{C^2}{1-zC^2} = C\left(\frac{C}{1-zC^2}\right) = BC.$$

6 Acknowledgment

The authors would like to thank an anonymous referee for suggestions that improved the presentation of this paper.

References

[1] E. Bender, Asymptotic methods in enumeration, *SIAM Rev.* **16** (1974), 485–515; errata, **18** (1976), 292.

[2] G. Cheon and L. Shapiro, The uplift principle for ordered trees, *Appl. Math. Lett.* **25** (2012), 1010–1015.

[3] E. Deutsch and L. Shapiro, Seventeen Catalan identities, *Bull. Inst. Combin. Appl.* **31** (2001), 31–38.

[4] E. Deutsch and L. Shapiro, A survey of the Fine numbers, *Discrete Math.* **241** (2001), 241–265.

[5] A. Frosini, R. Pinzani, and S. Rinaldi, About half the middle binomial coefficient, *Pure Math. Appl.* **11** (2000), 497–508.

[6] R. Graham, D. Knuth, and O. Patashnik, *Concrete Mathematics, 2nd ed.*, Addison-Wesley, 1994.

[7] M. Jani and M. Zeleke, k-Trees and Catalan identities, *Congr. Numer* **165** (2003), 39–49.

[8] W. Lang, On polynomials related to powers of the generating function of Catalan's numbers, *Fibonacci Quart.* **38** (2000), 408–419.

[9] H. Mahmoud, *Evolution of Random Search Trees*, Wiley, 1992.

[10] L. Shapiro, S. Getu, W. J. Woan, and L. Woodson, The Riordan group, *Discrete Appl. Math.* **34** (1991), 229–239.

[11] L. Shapiro, 42 Catalan identities and why you should care, *Bull. Inst. Combin. Appl.* **71** (2014), 94–102.

[12] R. Stanley, *Catalan Numbers*, Cambridge University Press, 2015.

[13] The On-Line Encyclopedia of Integer Sequences, published electronically at http://oeis.org, 2015.

2010 *Mathematics Subject Classification*: Primary 05C30; Secondary 05C05.
Keywords: ordered tree, boundary, colored edge, singleton edge.

(Concerned with sequences A000108, A000245, A000344, A000957, A000984, A001700, A068551, A014137, A228180, A228197, and A228343.)

Received October 2 2013; revised versions received September 11 2014; May 27 2015. Published in *Journal of Integer Sequences*, May 28 2015.

Return to Journal of Integer Sequences home page.