Common edges in rooted trees
and polygonal triangulations

Sean Cleary [*]

Department of Mathematics
The City College of New York, NY, USA

cleary@sci.ccny.cuny.edu

Andrew Rechnitzer [†]

Department of Mathematics
University of British Columbia, Vancouver, Canada

andrewr@math.ubc.ca

Thomas Wong

Department of Mathematics
University of British Columbia, Vancouver, Canada

twong@math.ubc.ca

Abstract

Rotation distance between rooted binary trees measures the degree of similarity of two trees with ordered leaves and is equivalent to edge-flip distance between triangular subdivisions of regular polygons. There are no known polynomial-time algorithms for computing rotation distance. Existence of common edges makes computing rotation distance more manageable by breaking the problem into smaller subproblems. Here we describe the distribution of common edges between randomly-selected triangulations and measure the sizes of the remaining pieces into which the common edges separate the polygons. We find that asymptotically there is a large component remaining after sectioning off numerous small polygons which gives some insight into the distribution of such distances and the difficulty of such distance computations, and we analyze the distributions of the sizes of the largest and smaller resulting polygons.

1 Introduction

A binary search tree is a rooted binary tree where the leaves are ordered "left to right". The efficiency of binary search trees depends on their height and balance. A rotation at a node is a simple mechanism for changing the shape of the tree while preserving order on leaves. The general schematic for rotations is shown in Figure 1 and Figure 2 shows an example of this rotation.
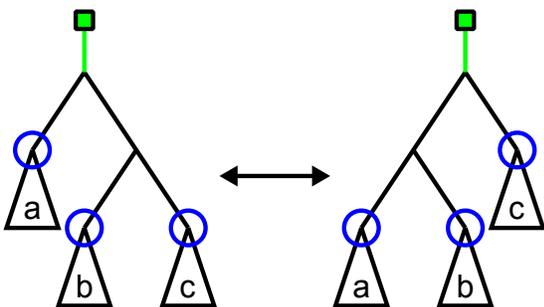
Figure 1: Rotation at a node, where the lettered triangles denote either leaves or subtrees of the tree.
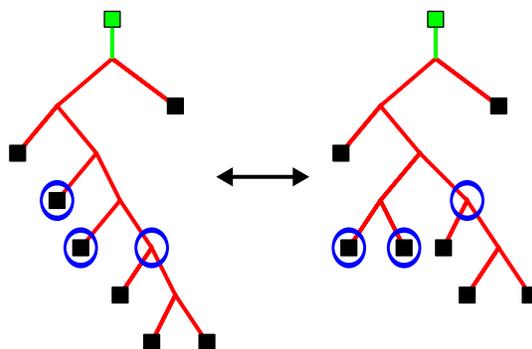
Figure 2: Rotation on the circled nodes of a rooted binary tree.

Culik and Wood [9] showed this rotation operation is transitive as each binary tree on n internal nodes can be transformed to and from a right caterpillar tree. There may be more a more direct sequence of transformations, giving a notion of rotation distance between two rooted binary trees of the same size. There is a simple upper bound of $2n - 2$ for rotation distance for two binary trees with n internal nodes, which was not only improved to $2n-6$ for values of $n \geqslant 12$ but furthermore shown to be optimal for large n through the remarkable work of Sleator, Tarjan and Thurston [15]. Recent work of Pournin [13] shows that the bound is optimal for $n \geqslant 11$. There are no known polynomial time algorithms for computing rotation distance, though there are polynomial- and linear-time approximation algorithms of Baril and Pallo [2] and Cleary and St. John [6].

Rooted binary trees with n internal nodes and thus $n + 1$ leaves are Catalan objects, as are triangulations of regular polygons with $n + 2$ sides. Rotations in rooted binary trees map to flipping of diagonal edges of quadrilaterals in polygonal subdivisions, where the natural metric is known as edge-flipping distance. Computing rotation distance between rooted trees is equivalent to computing edge-flip distance between triangulations of polygons. Here, we use the polygonal description for understanding the distance and the relevant algorithms.

Definition 1.1. *The Catalan numbers are defined on the integers ($k \geqslant 0$) as $c_k = \frac{(2k)!}{k!(k+1)!} = \frac{1}{k+1}\binom{2k}{k}$.*

Definition 1.2. *A triangulation of a convex $n + 1$-gon is a set of $n - 2$ non-intersecting chords that partition the polygon into $n - 1$ triangles.*

Figure 3 illustrates an example of the bijection between rooted binary trees and triangulations. Under this bijection, each leaf of the rooted binary tree corresponds to an edge in the triangulation. We consider the edge of the polygon corresponding to the root of the tree to be "marked," giving rise to a marked triangulation.
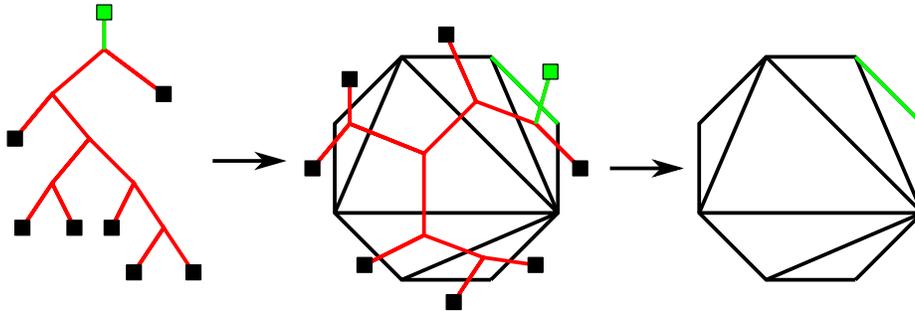
Figure 3: Converting a rooted binary tree with 7 leaves into its corresponding triangulation of an octagon with a marked edge.

We will count triangulations by the number of non-rooted edges (or by leaves in the corresponding rooted binary tree). Thus, a triangulation of size n means a triangulation on an $n + 1$-gon or with n leaves in the underlying binary tree.

Using the fact that rooted binary trees and triangulations of polygons are both Catalan objects with a straightforward bijection between them, we can easily convert from rotation distance to edge-flip distance. Specifically, each chord in a triangulation defines a unique quadrilateral by taking its adjacent two triangles. A edge-flip move replaces the selected chord with the other diagonal in that quadrilateral. Figure 4 gives an example of an edge-flip move on a 12-gon and the result rotation on the corresponding tree.

Definition 1.3. *The associahedron on the set of triangulations of size n is the graph $G = (V, E)$ where V is the set of all triangulations of size n (that is, of $n + 1$-gons) and E is set of unordered pairs of triangulations that differ by exactly one edge flip.*

The associahedron of the set of triangulations of size 5 is shown in Figure 5. The connectedness of the associahedron follows from considering edge-flips to fan triangulations (subdivisions with triangles all incident on a selected vertex– one of the fan triangulations corresponds to the right caterpillar tree.) Further, the flip distance between a pair of triangulations (T_1, T_2) is defined to be the minimum number of edge flips required to convert T_1 into T_2 and vice versa. This translates directly into finding the length of the shortest path in the associahedron. This minimal distance between T_1 and T_2 is realized by a path from T_1 to T_2 of minimal length, and any such path of minimal length connecting T_1 and T_2 is known as a geodesic path.

We let $\mathscr{P}_n$ denote the combinatorial class of all pairs of triangulations of size n. Further, we denote $\mathscr{P} = \bigcup_{n \geqslant 0} \mathscr{P}_n$ to be the combinatorial class of all pairs of triangulations of all sizes. Since there are c_n triangulations of size $n + 1$, there are c_n^2 pairs of triangulations of size $n + 1$.
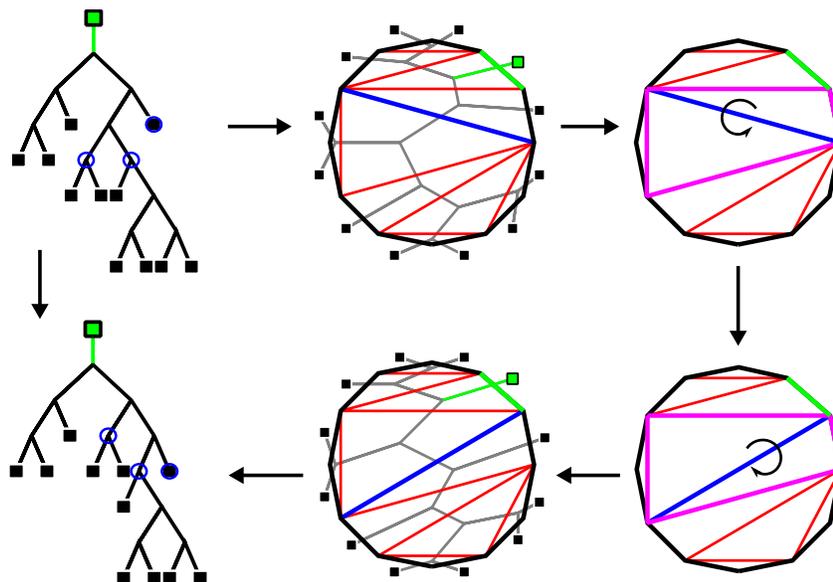
Figure 4: Example of an edge flip in 12-gon. After selecting the edge to be flipped, we highlight the quadrilateral it spans and we replace the selected edge with the opposite diagonal. The corresponding rooted binary trees and rotation move is also shown.
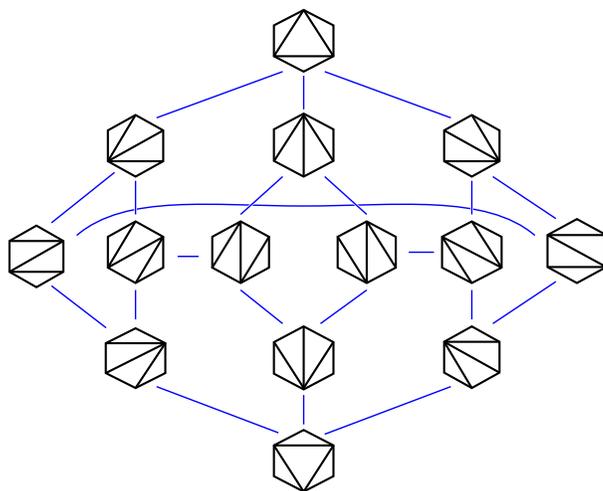
Figure 5: The associahedron for triangulations of the hexagon.

Definition 1.4. *We denote the generating function for the combinatorial class $\mathscr{P}$ by $P(z)$, where z is conjugate to the number of non-rooted edges in the polygon between a pair of triangulations. Thus,*

$$P(z) = \sum_{n=0}^{\infty} c_n^2 z^{n+1}. \tag{1}$$

For each element of $\mathscr{P}$, we consider superimposing the pair of triangulations. The chords that overlap are chords that are common to both triangulations. These common chords (or matched edges) play a central role for the rest of the article and we study how these matched edges partition pairs of triangulations into collections of pairs of smaller triangulations. Sleator, Tarjan and Thurston [15] showed that for any pair of triangulations, there is a geodesic between them that does not involve flipping any matched edges. This reduces the problem of finding geodesics between a pair of large triangulations into finding geodesics between the sub-components defined by partitioning along the matched edges. If this partition separates the triangulations into $O(n)$ small sub-components, then each piece can be dealt with individually, say via lookup tables. However, if this partition leaves any large $O(n)$ pieces, then while the problem is simplified, the problem of finding geodesics between $O(n)$ triangulations remains. To understand the effectiveness of partitioning large triangulations as a method of finding geodesics, the following two statistics about pairs of triangulations related to flip distances will be of interest:

1. On average, how many matched edges are there between two random triangulations?

2. On average, what does the resulting partition look like?

Pairs of rooted binary trees of the same size are a way of representing elements of Thompson's group, a well-studied group with a number of perplexing properties. Cannon, Floyd and Parry [3] give background on Thompson's group and its description via tree pair diagrams. In work investigating asymptotic properties of subgroups of Thompson's group F, Cleary, Elder, Rechnitzer and Taback [5] showed that on average, there is at least one common edge of a particular peripheral type and count effectively the number of tree pairs with no peripheral common edges. Here we address more general questions about the number of common edges and how their distribution partitions the tree pairs into collections of smaller tree pairs.

Since common chords will overlap if the two triangulations are superimposed, partitioning each element of $\mathscr{P}$ along the common edges is a well-defined operation. Each resulting sub-piece will also be an element in $\mathscr{P}$ (that is, a pair of triangulations of the same size). These sub-pieces will inherit the special property that they do not possess any common chords. An example of superimposing two triangulations and partitioning according to common chords is shown in Figure 8.

This special subclass of $\mathscr{P}$ is of particular importance. We define the combinatorial class $\mathscr{R}$ (with the corresponding generating function $R(z)$) to be the set of all pairs of triangulations (of the same size), with no common chords between them. The class $\mathscr{R}$ can be interpreted as the "primitive pieces" or the "reduced pieces" of $\mathscr{P}$ since each element

in $\mathscr{P}$ can be partitioned into a sequence of these reduced pieces by splitting along the matched edges. Furthermore, the class $\mathscr{R}$ can also be interpreted as the "root component" by identifying the primitive piece in each element of $\mathscr{P}$ that contains the root edge. For simplicity, a pair of triangulations with no common chords will be described as an $\mathscr{R}$-piece and a pair of triangulations with no such restrictions will be described a as $\mathscr{P}$-piece using the same convention.

Thus, the first two questions of interest can be rephrased as follows.

1. On average, how many $\mathscr{R}$-pieces result from partitioning a random $\mathscr{P}$-piece?

2. On average, what are the sizes of these $\mathscr{R}$-pieces?

We also investigate the distribution of another useful method of reducing complexity for instances of edge-flipping distance or rotation distance calculations. This is via identifying and flipping "one-off" edges as follows:

Definition 1.5. *Given two triangulations of the same size P and Q, an edge e in P is called a* one-off *edge with respect to Q if a single diagonal flip operation in P for e results in a new edge f which lies in Q.*

Note that the notion of being a one-off edge is not necessarily symmetric- the edge f in Q to which e in P may be flipped in a single move in P may not be capable of being flipped in Q to e there, as illustrated in Figure 6.
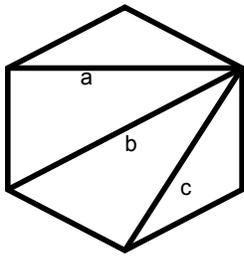
Figure 6: Edges a and c in the left triangulation are one-off edges with respect to the right one (mapping to d and f respectively). But there is just one one-off edge e in the right one with respect to the left one (mapping to b).
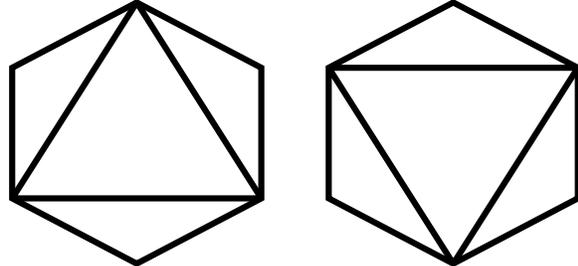
Figure 7: Two triangulations of the hexagon, with no one-off edges present in either triangulation.

As shown in Sleator, Tarjan, and Thurston [15], any time there is a one-off edge in P with respect to Q, there is a geodesic path which begins by flipping that edge. So each one-off edge will, after a single required flip, give a new common edge. Knowing how many one-off edges are typically initially present also gives some insight in to the typical difficulty of calculating flip-distance instances. Note that flipping a one-off edge may create additional one-off edges, so a natural means of making progress in finding a

geodesic path from S to T is to identify common edges to split the problem into smaller subproblems, then in each smaller problem of converting S_i to T_i, identify all one-off edges in S_i with respect to T_i and flip them to obtain S_i'. If in S_i', there are one-off edges with respect to T_i, we iterate to obtain eventually pair (S_i'', T_i). If not, we look for one-off edges in T_i with respect to the resulting S_i'', again iterating if possible, and then switching to work from the S direction once more, continuing to seek one-off edges in either direction until no longer possible. Each flip that we perform is indeed essential, and the process will either terminate in a geodesic path from the original S to T, or it will end at a collection of problems where there are no common edges or one-off edges in each pair. Such pairs do exist, see Figure 7; we discuss these further at the end of the paper.

2   Triangular subdivisions of polygons
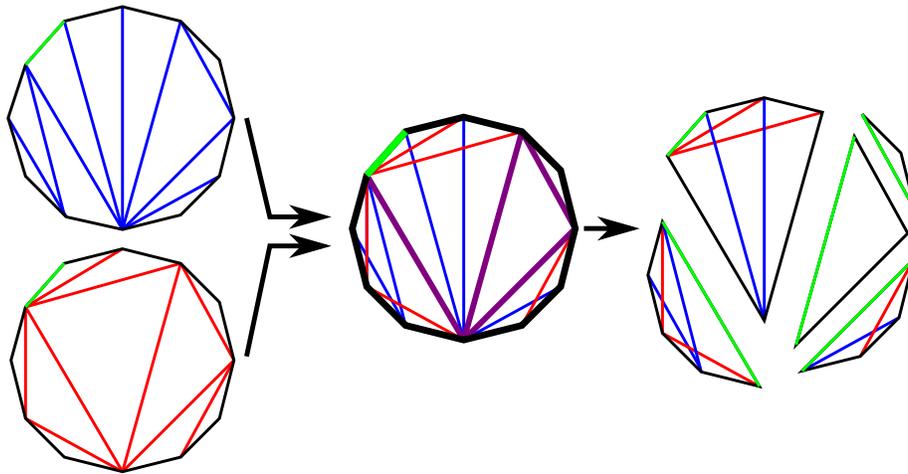
Figure 8: Superimposing and partitioning a pair of triangulations by its matched edges.

As described in Cleary and St. John [7], finding a geodesic between two triangulations is equivalent to finding a geodesic in each of the $\mathscr{R}$-pieces. To aid the calculation, we define the auxiliary variables m and q on the generating function $P(z)$ with m conjugate to the number of matched edges and q conjugate to the size of the $\mathscr{R}$-piece containing the root edge. Further, we define the bivariate generating functions $M(z, m)$ and $Q(z, q)$. This can be combined to form the multivariate generating function $P(z; m, q)$ which counts both the number of matched edges and root component size. Note that $P(z; m, 1) = M(z, m)$, $P(z; 1, q) = Q(z, q)$ and $M(z, 1) = Q(z, 1) = P(z; 1, 1) = P(z)$. While the function $P(z; m, q)$ provides the most generality, the parameters m and q will be studied independently and thus the functions $M(z, m)$ and $Q(z, m)$ will be used as specializations of $P(z; m, q)$ to simplify notation.

The relationship between these generating functions can be summarized by:

Theorem 2.1. *The two functions $P(z; m, q)$ and $R(z)$ are related by:*

$$P(z; m, q) = R\left(q(z + mP(z; m, 1))\right). \tag{2}$$

Proof. The combinatorial class for all pairs of triangulations $\mathscr{P}$ can be constructed recursively. Each pair of triangulations can be constructed by first picking the corresponding root component in $\mathscr{R}$ (with size counted by q) and then making the following substitution. We consider each edge of the root component in turn:

1. If that edge is an edge in the superimposed triangulation, do nothing;

2. If that edge is a common chord, replace it with the corresponding $\mathscr{P}$-piece. This substitution step does not contribute to the root component size q, but it does increase the number of matched edges by 1 since the replaced edge is now a common chord.

Thus $\mathscr{P} = \mathscr{R} \circ (\mathscr{P} + \mathscr{Z})$ where $\mathscr{Z}$ is the atomic class denoting an edge in the root component. □

By specializing to $q = 1$ and $m = 1$, the following relationships are obtained.

Corollary 2.2. *The relationship between $R(z)$ and $M(z, m)$ is given by*

$$M(z, m) = R(z + mM(z, m)), \tag{3}$$

and the relationship between $R(z)$ and $Q(z, q)$ is given by

$$Q(z, q) = R(q(z + Q(z, 1))). \tag{4}$$

Corollary 2.3. *The two functions $P(z)$ and $R(z)$ are related by:*

$$P(z) = \sum_{k=0}^{\infty} \frac{P(z)^k}{k!} \left(\frac{\mathrm{d}}{\mathrm{d}z}\right)^k R(z). \tag{5}$$

This can be extended to $Q(z, q)$ as follows:

$$Q(z, q) = \sum_{k=0}^{\infty} \frac{P(z)^k}{k!} \left(\frac{\mathrm{d}}{\mathrm{d}z}\right)^k R(qz). \tag{6}$$

Proof. We expand the series R from Theorem 2.1 as a Taylor series about a point α,

$$R(t) = \sum_{k=0}^{\infty} \frac{(t - \alpha)^k}{k!} \left(\frac{\mathrm{d}}{\mathrm{d}z}\right)^k R(\alpha). \tag{7}$$

Then set $\alpha = z$ and $t = z + P$ to obtain the result. The extension follows from the observation that the only contribution to the size of the root component comes from R. □

A similar result can be obtained for $M(z, m)$.

Corollary 2.4. *The two functions $P(z)$ and $R(z)$ are related by:*

$$P(z) = \sum_{k=0}^{\infty} \frac{1}{(k+1)!} \left(\frac{\mathrm{d}}{\mathrm{d}z}\right)^k R^{k+1}(z). \tag{8}$$

This can be extended to give:

$$M(z, m) = \sum_{k=0}^{\infty} \frac{m^k}{(k+1)!} \left(\frac{\mathrm{d}}{\mathrm{d}z}\right)^k R^{k+1}(z). \tag{9}$$

Proof. This is a direct consequence of the Lagrange Reversion Theorem. For additional details, we refer the reader to Section 7.32 of Whittaker and Watson [18] □

Corollary 2.4 and Corollary 2.3 both admit combinatorial proofs. For Corollary 2.4, we consider a sequence of $k + 1$ $\mathscr{R}$-pieces and select k non-root edges from the sequence. The first $\mathscr{R}$-piece is also a $\mathscr{P}$-piece. Proceeding by induction gives one of two possible cases at each step.

- If at least one of the edges on the $\mathscr{P}$-piece is selected, then we attach the next unused $\mathscr{R}$-piece in the sequence to the first selected edge on the $\mathscr{P}$-piece counter clockwise from the root edge on the $\mathscr{P}$-piece.

- If there are no selected edges in the $\mathscr{P}$-piece, then we select the next $\mathscr{R}$-piece in the sequence that contains at least one selected edge. We attach the $\mathscr{P}$-piece to the first selected edge of the $\mathscr{R}$-piece counter clockwise from the root edge of the $\mathscr{R}$-piece. The existence of the $\mathscr{R}$-piece is an immediate consequence of selecting k edges over $k + 1$ $\mathscr{R}$-pieces.

This process associates the root edges of k $\mathscr{R}$-pieces to the k selected edges leaving one $\mathscr{R}$-piece as as root component. Once this association is determined, the order in which the $\mathscr{R}$-pieces are connected does not affect the result. This results in the factor of $\frac{1}{(k+1)!}$. An example for $k = 7$ is shown in Figure 9. Conversely, each element in $\mathscr{P}$ can be obtained in this manner by partitioning the element into $\mathscr{R}$-pieces and keeping track of where the splitting takes places and which $\mathscr{R}$-pieces associated to those edges.

Corollary 2.3 follows immediately from Corollary 2.4 by considering the root component and amalgamating each branch of $\mathscr{R}$-pieces into a $\mathscr{P}$-piece. This is equivalent to selecting k elements in $\mathscr{P}$ and selecting k non-root edges in the root component to associated them to. As with the previous case, the order in which they are combined does not alter the result and hence, each resulting in a weight of $\frac{1}{k!}$. The example in Figure 9 is modified according to this process to give Figure 10.

From this point, it is trivial to see the combinatorial derivation of Theorem 2.1. Each edge of the root component is simply left alone or replaced by an element in $\mathscr{P}$ resulting in a substitution operation (Theorem I.4, Flajolet and Sedgewick [11]).
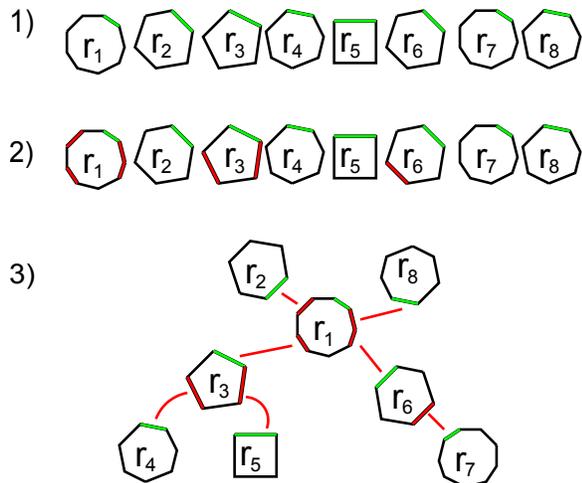
Figure 9: A sequence of 8 $\mathscr{R}$-pieces (1) can be combined to form a pair of triangulations by selecting 7 non-root edges (2) and associating each $\mathscr{R}$-piece to a selected edge (3).
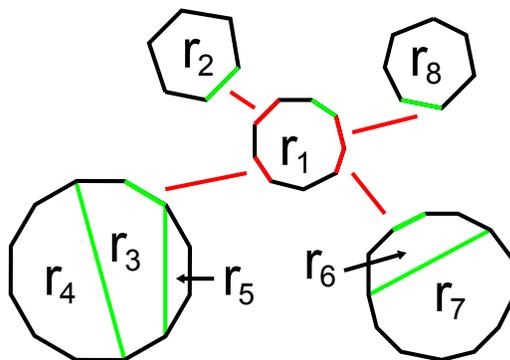
Figure 10: Each branch off the root piece can be thought of as an element in $\mathscr{P}$ by amalgamating the primitive pieces in the previous diagram.

2.1 Calculating Statistics

The following expression will simplify calculations later on. We consider the relation $P(z) = R(z + P(z))$ and set $t = T(z) = z + P(z)$. After substitution and rearranging, this results in:

$$R(t) = t - T^{-1}(t). \tag{10}$$

Taking the derivative results in

$$R'(t) = 1 - \frac{1}{T'(t - R(t))}. \tag{11}$$

The second derivative can be found using a similar manner

$$R''(t) = -\frac{T''(t - R(t))}{T'(t - R(t))^3}. \tag{12}$$

Note that setting $t = z + P(z)$ into the above two expressions gives

$$R'(z + P(z)) = 1 - \frac{1}{T'(z)}, \tag{13}$$

$$R''(z + P(z)) = -\frac{T''(z)}{T'(z)^3}. \tag{14}$$

Using a similar logic, we write $M(z, m)$ in a similar fashion. Since $M(z, m)$ is defined implicitly, this rewriting rule will allow the numerical extraction of coefficients of $M(z, m)$ more efficiently. Since

$$M(z, m) = R(z + mM(z, m)), \tag{15}$$

we set $u = U(z,m) = z + mM(z,m)$ and rearrange to give

$$u - mR(u) = z. \tag{16}$$

Since R is a known function, this equation can be solved for $u = U(z,m)$. This can then be converted into an expression of $M(z,m)$.

The mean number of edges for a random pair of triangulations on $n + 1$ non-rooted edges can then be found using the techniques described in Flajolet and Sedgewick [11]. We define the discrete random variable ρ_n to be the size of the root component in a pair of triangulations of size n. Similarly, we define the discrete random variable τ_n to be the number of matched edges between a pair of triangulations of size n.

Differentiating $M(z,m)$ with respect to m gives

$$\frac{\partial M}{\partial m}(z,m) = R'(z + M(z,m)) \cdot \left(M(z,m) + m\frac{\partial M}{\partial m}(z,m) \right). \tag{17}$$

Setting $m = 1$ and writing $M(z,1) = P(z)$ we have

$$\frac{\partial M}{\partial m}(z,1) = R'(z + P(z)) \cdot \left(P(z) + \frac{\partial M}{\partial m}(z,1) \right) \tag{18}$$

$$= \frac{R'(z + P(z))}{1 - R'(z + P(z))} = P(z) \cdot P'(z). \tag{19}$$

We can interpret (or derive) this expression directly as follows. Differentiating with respect to m and setting $m = 1$ is equivalent to marking a single common edge. Cutting along this edge divides the $\mathscr{P}$-piece into two $\mathscr{P}$-pieces one of which has a marked edge on its boundary. This pair is counted by the generating function $P'(z)P(z)$.

Thus, the expected number of matched edges for the pair of triangulations of size n is given by

$$\mathbb{E}(\tau_n) = \frac{[z^n]\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}z}P(z)^2}{[z^n]P(z)}. \tag{20}$$

This process can be repeated to find the mean size of the root component by using Corollary 2.2 which results in:

$$\mathbb{E}(\rho_n) = \frac{[z^n]\frac{P'(z)(z+P(z))}{1+P'(z)}}{[z^n]P(z)}. \tag{21}$$

A natural extension to computing the mean is to consider the variance of the two statistics. To obtain the variance, we use the second derivatives $\frac{\partial^2}{\partial m^2}M(z,m)\Big|_{m=1}$ and $\frac{\partial^2}{\partial q^2}Q(z,q)\Big|_{q=1}$. A quick Maple exercise shows that:

$$\frac{\partial^2}{\partial m^2}M(z,m)\Big|_{m=1} = \frac{1}{3}\frac{\partial^2}{\partial z^2}P(z)^3. \tag{22}$$

Similarly, the variance for the size of the root piece can also be calculated by making use of the second derivative of $Q(z, q)$:

$$\left. \frac{\partial^2}{\partial q^2} Q(z, q) \right|_{q=1} = \frac{P''(z)(z + P(z))^2}{(P'(z) + 1)^3}. \tag{23}$$

We note here that we can also compute the mean number of one-off edges, but not their full distribution. We consider a $\mathscr{P}$-piece with a marked one-off edge. Flipping that edge creates a common edge and so we can split the $\mathscr{P}$-piece into two $\mathscr{P}$-pieces one of which has a marked boundary edge. Conversely, we can take any two $\mathscr{P}$-pieces, mark one edge of the boundary on the first, glue the root of the second to this marked edge. This creates a common edge, which becomes a marked one-off when flipped. Hence

$$\text{genarating function of } \mathscr{P}\text{-pieces with 1 marked one-off} = P'(z) \cdot P(z). \tag{24}$$

Since this generating function is the same equation (19), it follows that the mean number of one-off edges is the same as the mean number of common edges. Their distributions are different, however. This confirms experimental evidence of Cleary and Chu [4] that the mean number of common edges and the mean number of one-off edges are the same, with significantly smaller standard deviation observed for one-off edges.

A similar argument gives the generating function of $\mathscr{R}$-pieces with 1 marked one-off edge:

$$\text{generating function of } \mathscr{R}\text{-pieces with 1 marked one-off} = R'(z) \cdot R(z). \tag{25}$$

3 Asymptotics

3.1 Growth of $\mathscr{R}$-pieces

The asymptotic expansion of the Catalan numbers is very well understood. Hence, the properties and asymptotics of $P(z)$ are also well understood. Asymptotically $c_n \sim \frac{4^n}{n^{\frac{3}{2}}\sqrt{\pi}}$ and so $p_n \sim \frac{16^n}{\pi n^3}$. We note that the n^{-3} term indicates the presence of logarithmic singularities and so $P(z)$ is not algebraic (see VII.7 of [11] for a general discussion of the asymptotics of coefficients of algebraic functions), but it does satisfy a linear differential equation with polynomial coefficients — and so is D-finite, defined below:

Definition 3.1. *A formal power series $f(x)$ is said to be differentiably finite (or D-finite) if there exist finitely many polynomials $d_0(x), d_1(x), \ldots, d_m(x)$ (with $d_m(x)$ non-zero) such that:*

$$d_0(x)f(x) + d_1(x)f'(x) + \ldots + d_m(x)f^{(m)}(x) = 0.$$

For more information about differentiably finite functions, the reader is referred to Stanley [17]. The other functions of interest, $R(z), M(z, m)$ and $Q(z, q)$ are not D-finite — the proof of this hinges on the location of the dominant singularity of $R(z)$. Using equation (10) we obtain

Lemma 3.2. *The dominant singularity of $R(z)$ occurs at $z_r = \frac{1}{\pi} - \frac{1}{4}$.*

So the exponential growth rate of $\mathscr{R}$-pieces is $\frac{1}{z_r} \sim 14.6392$ and so they are exponentially rare in the set of $\mathscr{P}$-pieces. Extending this analysis further, it is possible to obtain the dominant asymptotic behavior of $R(z)$.

$$[z^n]R(z) = \frac{(4-\pi)^2}{64} \left(\frac{4\pi}{4-\pi}\right)^n n^{-3} \left(1 - \frac{3(4-\pi)}{2}\frac{\log(n)}{n} + O\left(\frac{1}{n}\right)\right). \tag{26}$$

Lemma 3.3. *The functions $R(z)$, $M(z,m)$ and $Q(z,q)$ are not D-finite as series in z.*

Proof. Using standard techniques from ordinary differential equations, if $R(z)$ were D-finite, then its singularities must correspond to zeros of the polynomial $d_m(z)$ multiplying the highest order derivative (in the definition of D-finite above). Since the dominant singularity of $R(z)$ is not an algebraic number, it cannot be the root of a polynomial and hence $R(z)$ cannot be D-finite.

The specialization of any D-finite function must also be D-finite, see Lipshitz [12]. Since we can write $R(z) = M(z,0)$, it follows that $M(z,m)$ is not D-finite. A similar argument shows that $Q(z,q)$ is not D-finite. $\qquad\square$

We note that the functions $R(z)$, $M(z,m)$ and $Q(z,q)$ do satisfy algebraically differentiable equations (ADE).

3.2   Number of matched edges and size of root component

Even though the functions $M(z,m)$ and $Q(z,q)$ are not D-finite, their asymptotic behavior can be extracted by making use of the asymptotic expansion of $P(z)$. Using the well known expansion of the Catalan numbers, the asymptotic expansion of the coefficients of $P(z)$ is given by:

$$[z^n]P(z) = \frac{16^n}{\pi n^3} \left(\frac{1}{16} + \frac{3}{64n} + O\left(\frac{1}{n^2}\right)\right). \tag{27}$$

In addition, there is closed form for the series $P(z)$ which plays a central role determining the asymptotic expansions of the mean and variance.

$$P(z) = \sum_{k=0}^{\infty}(c_k)^2 z^{k+1} = \frac{1}{256\pi}\left(-\pi(256z+64) + (2048z-128)\mathrm{K}(4\sqrt{z}) + \mathrm{E}(4\sqrt{z})\right), \tag{28}$$

where K and E are the complete elliptic integral of the first and second kind respectively. These functions are defined by:

$$K(k) = \int_0^1 \frac{1}{\sqrt{1-t^2}\sqrt{1-k^2t^2}}dt \qquad \text{and} \qquad E(k) = \int_0^1 \frac{\sqrt{1-k^2t^2}}{\sqrt{1-t^2}}dt. \tag{29}$$

This expansion forms the denominator in all of the mean and variance computations. The numerator is the asymptotic expansion for the mean number of matched edges Equation (20). Combined, this gives:

$$\mathbb{E}(\tau_n) = \frac{[z^n]\frac{\mathrm{d}}{\mathrm{d}z}\left(\frac{1}{2}P(z)^2\right)}{[z^n]P(z)} = \frac{16-5\pi}{\pi}n + \frac{7\pi-20}{\pi} + O\left(\frac{\log(n)}{n}\right), \tag{30}$$

which evaluates to about $0.092958n + 0.633802 + O\left(\frac{\log(n)}{n}\right)$. Since the mean number of matched edges is exactly the mean number of one-off edges, the above also gives the mean number of one-off edges in a $\mathscr{P}$-piece. As noted above, we can also find the mean number of one-off edges in a $\mathscr{R}$-piece, and this gives

$$\mathbb{E}(\#\text{one offs}) = \frac{[z^n]\frac{\mathrm{d}}{\mathrm{d}z}\left(\frac{1}{2}R(z)^2\right)}{[z^n]R(z)} = \left(\frac{5}{4} - \frac{1}{(4-\pi)}\right)n + O(1), \tag{31}$$

where this leading constant is approximately 0.08505. This result on matched edges can be viewed as a generalization of previous work by Cleary, Elder, Rechnitzer and Taback [5]. They considered the fraction of tree pairs which had at least one "common caret". Mapping these result to the current context, common carets correspond to a collection of completely matched components glued to the boundary of the $\mathscr{P}$-piece. They showed that probability that a tree pair has at least one common caret converges exponentially fast to 1.

A similar method can be applied to find the mean size of the root component. In this case, the asymptotic expansion required is given by that of Equation 21, which results in

$$\mathbb{E}(\rho_n) = \frac{[z^n]\frac{P'(z)(z+P(z))}{1+P'(z)}}{[z^n]P(z)} \simeq 0.674195n - 0.674195\log(n) + 1.212467223 + O\left(\frac{\log^2(n)}{n}\right), \tag{32}$$

where these first two constants are $\frac{\pi(4-\pi)}{4}$.
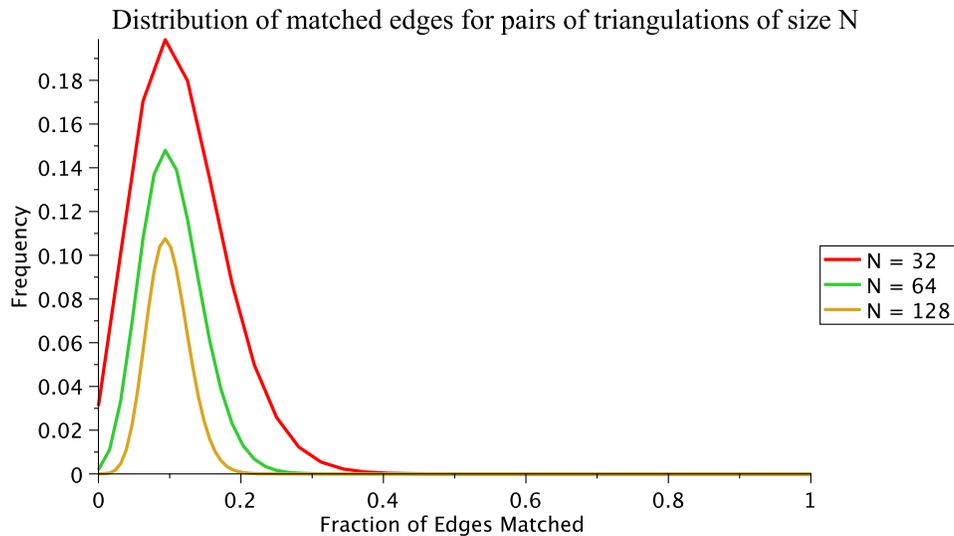
Figure 11: Fractional matched edge distribution plot of different values of N (from expansion of the generating function). Asymptotically, the mean is about $0.09n$ with variance about $0.2n$.

These methods can be extended to determine the asymptotic form for the variance of the number of matched edges ($\mathbb{V}(\tau_n)$) and the root component ($\mathbb{V}(\rho_n)$). The asymptotic

expansions of the variances are given by:

$$\mathbb{V}(\tau_n) \simeq 0.1015994012n + O(1) \tag{33}$$

$$\mathbb{V}(\rho_n) \simeq 0.12419674n^2 + O(n\log(n)). \tag{34}$$
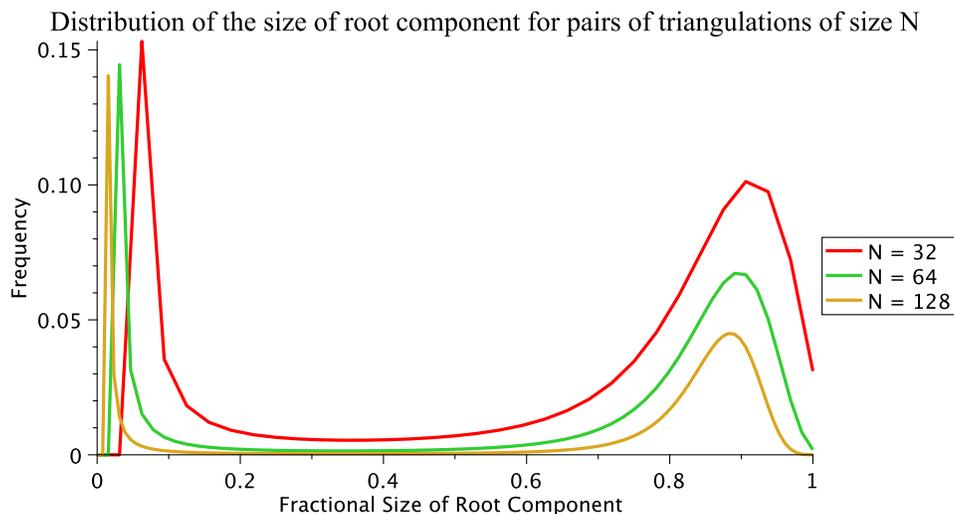
Figure 12: Fractional root component size distribution plot of different values of N (from the expansion of generating function). Asymptotically, the mean is about $0.67n$ with variance on the order of n^2. The spike on the left is the positive incidence of the root component being small, predominantly triangular or quadrilateral.

Since the variance of the number of matched edges grows linearly with the mean, we expect the distribution of matched edges (and so number of components) to be tight about its mean value. On the other hand, we see that the variance of the size of the root component grows quadratically with size and so the distribution must be very wide. Numerical experiments can be used to confirm the asymptotic expansions of the mean and variance, as well as study the rate of convergence. Figure 11 and Figure 12 show the distributions of the number of matched edges and the size of the root components as a fraction of the triangulations for various sizes and show quick convergence to the limiting probabilities computed above.

We note that these plots share some similarities with the results obtained in Banderier, Flajolet, Schaeffer and Soria [1] on decomposition of planar maps into 2-connected components. They find that a typical decomposition contains a linear number of components of which one is size $O(n)$. Their work extends to the composition of quite general generating functions similar to those in Theorem 2.1. However, the singular forms presented here fall into one of the few cases their work does not cover; unfortunately we cannot simply apply their results. The presence of logarithmic terms considerably increases the difficulty of the analysis and an analogously complete understanding of this double peaked distribution would be a significant undertaking.

3.3 Size of components

The positive density of matched edges implies that a typical $\mathscr{P}$-piece decomposes into a linear number of $\mathscr{R}$-pieces. While the root component can be "large" (of similar size to the whole $\mathscr{P}$-piece), it must be the case that there are many small components; indeed there is a positive probability that the root component is "small".
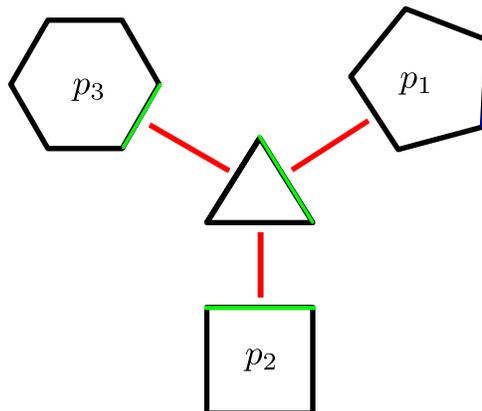
Figure 13: Attaching 3 $\mathscr{P}$-pieces to a triangle with one edge marked on one of the pieces. For each of the attached pieces which is empty, the triangle will have a boundary edge.

We can compute the mean number of the smallest possible (that is, triangular) $\mathscr{R}$-pieces inside a $\mathscr{P}$-piece in much the same manner by which we computed the mean number of edges above. In particular, if we mark any such triangular $\mathscr{R}$-piece within a $\mathscr{P}$-piece, this gives a natural decomposition into a triangle and 3 (possibly trivial) $\mathscr{P}$-pieces; one edge on the boundary of one of the $\mathscr{P}$-pieces is marked, as shown in Figure 13. At the level of generating functions this gives

$$\text{generating function of marked triangles} = (z + P(z))^2 \frac{\mathrm{d}}{\mathrm{d}z}\left(z + P(z)\right). \qquad (35)$$

Repeating the above analysis (and taking liberties with notation) gives

$$\mathbb{E}(\# \text{ triangles}) = \left(\frac{4 - \pi}{\pi}\right)^2 n + \cdots \qquad (36)$$

Hence the majority of the components are triangles (about 80%). We can go further and determine the expected number of triangles with 0, 1 or 2 edges on the boundary of the $\mathscr{P}$-piece, obtaining:

$$\mathbb{E}(\# \text{ triangles with 0 boundary edges}) = \left(\frac{16 - 5\pi}{4\pi}\right)^2 n + \cdots \qquad (37)$$

$$\mathbb{E}(\# \text{ triangles with 1 boundary edges}) = \frac{(16 - 5\pi)}{8\pi} n + \cdots \qquad (38)$$

$$\mathbb{E}(\# \text{ triangles with 2 boundary edges}) = \frac{1}{16} n + \cdots \qquad (39)$$

Thus about 84% of the common triangles are peripheral with two boundary edges, about 15.5% with one boundary edge, and less than 1% internal triangles. A similar analysis shows that about 13% of the components are quadrilaterals, and only about 7% are pentagons or larger.

For further analysis of the size of the root piece, we follow Example IX.42 from [11]. The probability that a $\mathscr{P}$-piece of size n has root component size k is given by

$$\mathbb{P}(\rho_n = k) = \frac{[z^n q^k] R(qT(z))}{p_n} = \frac{r_k [z^n] T(z)^k}{p_n} \qquad (40)$$

When the root component is "small", we keep k fixed and let $n \to \infty$; singularity analysis then shows that $[z^n] T(z)^k \underset{n\to\infty}{\sim} kT(\frac{1}{16})^{k-1} p_n$. Since $T(1/16) = z_r$, in this limit the probability that the root component is size k is $k z_r^{k-1} r_k$. Summing over k gives the probability that the root component is small, namely $R'(z_r) = 1 - 1/T'(1/16) = 1 - \pi/4 \approx 0.215$. Thus the probability that the root component is "large" is 0.785.

Since we know the mean size of the root component, and the probability that the root component is "small" we can estimate the mean size of the root component when it is "large" — this should also be the mean size of the largest component since it is well above $n/2$.

$$\text{Mean size of root component when large} = \frac{\frac{\pi}{4}(4 - \pi)n + o(n)}{\pi/4} = (4 - \pi)n + o(n) \qquad (41)$$

Unfortunately we cannot apply the general scheme detailed in Theorem IX.16 and Proposition IX.24 in [11], since the scaling of $P(z)$ falls precisely between two of the cases there. The dominant logarithmic singularity in $P(z)$ makes generalizing those results quite difficult and we have not pursued it here. We note, however, that if we naively apply case (iii) from Theorem IX.16, then we arrive at the same estimate $(4 - \pi)n$.

To bolster this argument we also examined the distribution numerically. We generated 2^{20} pairs of random binary trees of sizes up to 2^{14} using Remy's bijection [14]; we then mapped them to pairs of triangulations and computed histograms of the number of matched edges, root component size, largest component size and second-largest component size. The distribution of matched edges quickly converges to the expected normal distribution, as shown in Figure 14. The distribution of the size of the largest component also appears to converge to a distribution with the mean of $(4-\pi)n$. In Figure 15, we show the distribution of sizes of the largest component for sizes 1024 and 4096, superimposed with an appropriately scaled version of the distribution of sizes of the root component, and we see close agreement between those as well the appearance of convergence toward a normal distribution.

We also consider the size of the second largest component in $\mathscr{P}$-pieces. Banderier, Flajolet, Schaeffer and Soria [1] found that the average size of the second largest component of planar maps' decomposition into 2-connected components to be of size $n^{\frac{2}{3}}$. Numerical investigation suggests that in our case of triangulations, the average size of the second largest component grows about as $n^{\frac{1}{2}}$. It appears that there may be some

complicating corrections, and as shown in Figure 17, the agreement is not as complete as for the statistics we computed earlier.
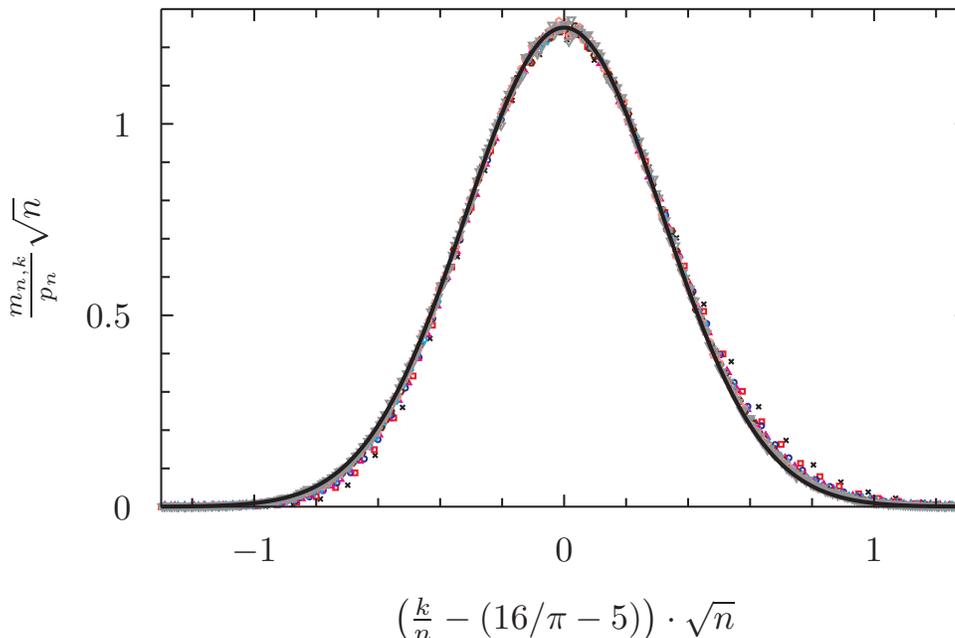
Figure 14: Let $m_{n,k}$ be the number of $\mathscr{P}$-pieces of size n and k matched edges. In this plot we show the scaled distribution of $m_{n,k}/p_n$ for lengths $128, 256, \ldots, 16384$ (from random sampling data), which as expected, collapse to the normal distribution with mean zero and variance 0.1015994012 (solid line).

4   Discussion

4.1   Distances in the associahedron

The distributions described in Figures 11 and 12 provide insight into the difficulty finding geodesics in the associahedron, and thus equivalently into finding rotation distance between rooted trees or edge-flip distance between triangulations. The task of finding a geodesic between a pair of large triangulations can be reduced to finding geodesics for smaller triangulations by partitioning along the common edges. Figure 11 shows that on average, a pair of large triangulations can be split into a linear number of parts. This plot agrees with the results obtained asymptotically. That is, the mean grows linearly with n and the standard deviation is $O\left(\sqrt{n}\right)$ and we observe a Gaussian distribution. Were it the case that each part had roughly the same size, this would imply that each part must be constant size on average (independent of the size of initial pair of triangulations). This would on average break the large problem into many much more tractable small problems and give a feasible way of computing rotation distance.
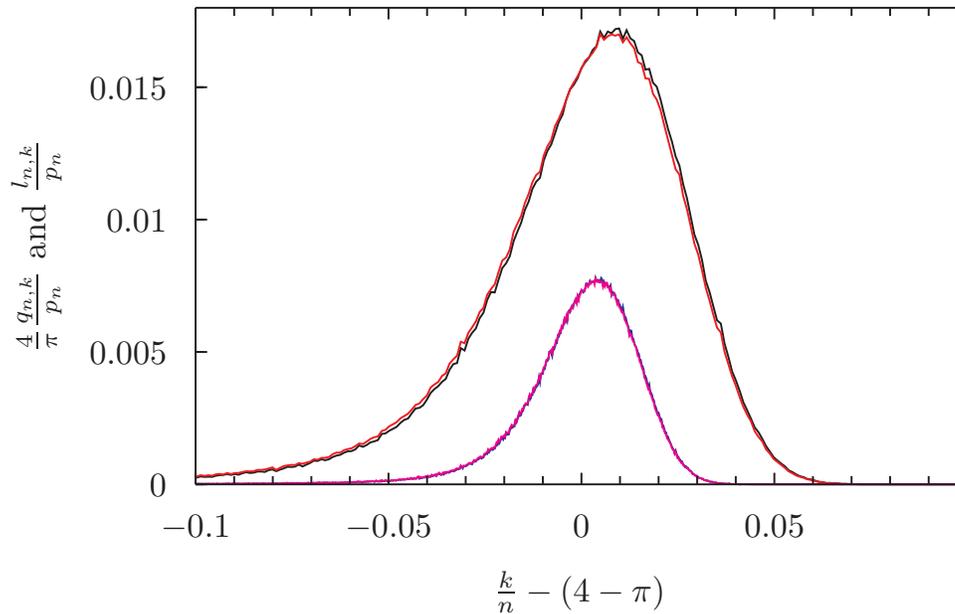
Figure 15: Let $q_{n,k}$ (resp. $l_{n,k}$) be the number of $\mathscr{P}$-pieces of size n and root component size k (resp. largest component size k). In this plot we show the distributions of fractional size of the root component and largest component for lengths 1024 (upper) and 4096 (lower). We scale the height of the distribution of the size of the root component by $\frac{4}{\pi}$ since this is the reciprocal of the probability that the root component is "large" — this gives excellent agreement. Additionally, it appears that these distributions are converging to a normal distribution, though very slowly with n.

However, Figure 12 shows that it is not the case that the components are roughly the same size. The plot agrees with the asymptotic results that show both the mean and standard deviation both grow linearly with n. That is, there will always be a non-zero probability of finding a primitive piece that is a constant fraction of the size of the original triangulation, typically quite large and by virtue of it being often of size more than $\frac{n}{2}$, necessarily the largest component of all components. There are also positive probabilities that the root component will be quite small, predominantly triangles and quadrilaterals with those probabilities declining initially sharply with size.

In terms of finding a geodesic effectively, a random pair of triangulations of size n will contain a number of matched edges which grows linearly with n. Those matched edges will partition the pair of triangulations into many small parts and generally one large component that is of size about $0.86n$. There will also be a number of one-off edges which grows linearly with n, and again we can partition and expect many small parts and a remaining large piece of considerable size. At this point, the obvious techniques of making progress on finding a geodesic fail, with the remaining large piece having no common edges and no one-off edges. This essentially gives a "kernelization" of the problem to the difficult instances — these large primitive pieces of the pairs of triangulations with no one-off edges.
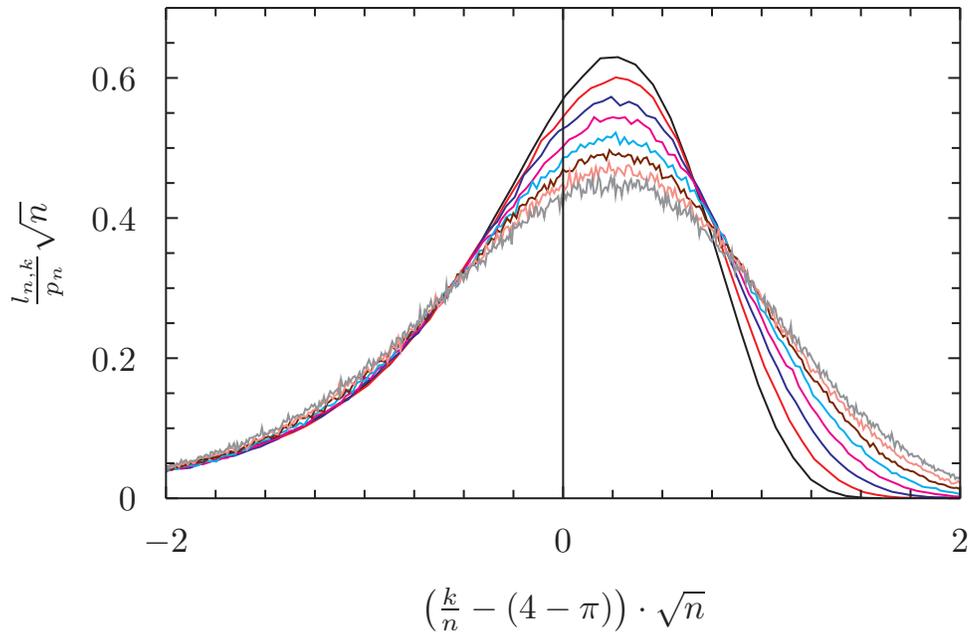
Figure 16: Let $q_{n,k}$ be the number of $\mathscr{P}$-pieces of size n and second largest component size k. In this plot we show the scaled distribution of $q_{n,k}/p_n$ for lengths $128, 256, \ldots, 16384$ (highest peak to lowest peak). It appears that the data is converging very slowly to a Gaussian distribution. We expect the presence of logarithmic corrections but we have not been able establish the nature of these corrections and so have unable to collapse our data to a single curve.

There are examples of triangulations which have no common edges, no one-off or two-off edges, and it is possible to construct examples of increasingly large triangulations with no k-off edges for k in $0 \ldots r$ for arbitrarily large r, described by St. John [16]. It seems from preliminary computational searches, though, that such examples are increasingly sparse in the set of all triangulations.

There are two distinct notions of difficulty of converting one triangulation into another. One notion of difficulty is the edge-flip distance itself; that is, a difficult transformation is one which requires a large number of edge flips. Another notion of difficulty is in terms of the computational difficulty of finding such a geodesic — that is, the expense of searching to find the optimal path, even if that path is not particularly long relative to the extremal case. The known cases of St. John [16] which have no small k-off edges are not close to maximally distant as they involve vertices with high valence. Indeed, the upper bound of $2n - 6$ generally means that after finally finding a common edge from an initial configuration with no small k-off edges, there is a cascade of obvious flips creating matched edges. Furthermore, it appears that many difficult instances (in the sense of large edge-flip distance) may still have relatively small k-off instances. For example the family of instances proposed in Dehornoy [10] as potential extremal $2n - 6$ cases all have 3-off edges.
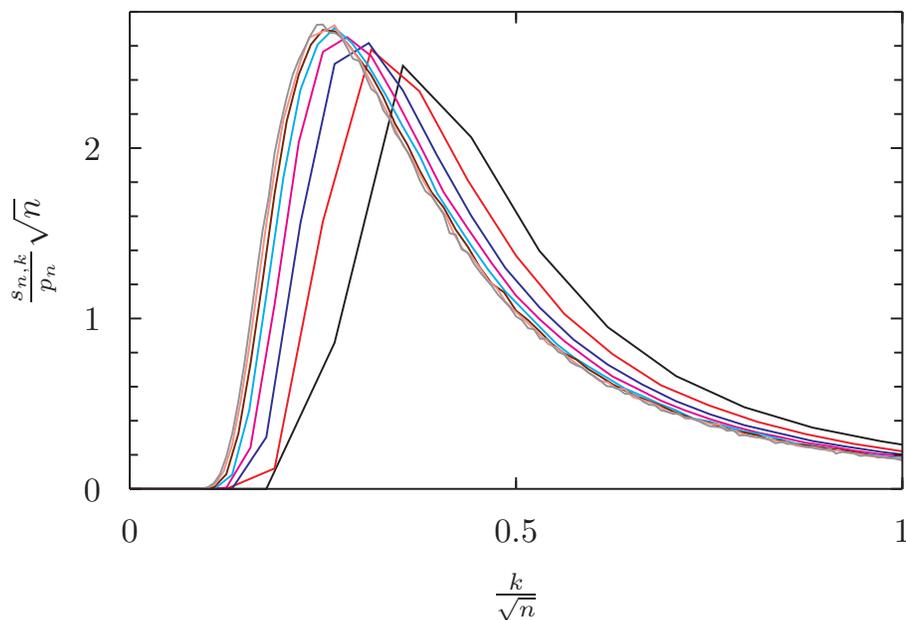
Figure 17: Let $s_{n,k}$ be the number of $\mathscr{P}$-pieces of size n and second largest component size k. In this plot we show the scaled distribution of $s_{n,k}/p_n$ for lengths $128, 256, \ldots, 16384$ (rightmost peak to leftmost peak). This suggests that the mean size of the second largest component is approximately $\sqrt{n}$ with logarithmic corrections — we have not been able to establish the nature of these corrections and so have unable to collapse our data to a single curve.

4.2   Thompson's group

The understanding of the asymptotic properties of these pairs of triangulations also gives some insight into the structure of Thompson's group F. Thompson's group F contains numerous copies of itself in a fractal-like manner, and many of these copies of F commute with each other (and thus essentially do not interact), as described in Cleary and Taback [8]. Elements of Thompson's group F whose corresponding polygonal triangulations have a common edge can be factored into products of three group elements of simpler complexity, where the central factor lies in a "clone" subgroup isomorphic to F itself. As with the triangulations, this analysis shows that a typical element of F selected from tree pair diagrams of a fixed size does factor into many such pieces, most of which are small but with generally one large piece which does not factor further in this way. This is relevant to the question of finding geodesic words representing elements of Thompson's group. The difficulty of such questions depends heavily upon the generating set used. With respect to the standard finite generating set and standard $\{x_0, x_1, \ldots\}$ infinite generating sets, the algorithms are linear-time and well-known. But in the case of the infinite generating set $\{y_\emptyset, y_0, y_1, y_{00}, \ldots\}$ finding geodesics is exactly analogous to finding edge-flip distance of corresponding triangulations. So with respect to the y_α generators, we see the behavior described above, where a typical large instance splits into many small instances which are

straightforward, together with a large piece which would require more significant analysis and for which no efficient algorithm is currently known.

References

[1] C. Banderier, P. Flajolet, G. Schaeffer, and M. Soria. Random maps, coalescing saddles, singularity analysis, and airy phenomena. *Random Structures & Algorithms*, 19(3-4):194–246, 2001.

[2] J. Baril and J. Pallo. Efficient lower and upper bounds of the diagonal-flip distance between triangulations. *Information Processing Letters*, 100(4):131–136, 2006.

[3] J. Cannon, W. Floyd, and W. Parry. Introductory notes on Richard Thompson's groups. *Enseign. Math. (2)*, 42(3-4):215–256, 1996.

[4] T. Chu and S. Cleary. Expected conflicts in pairs of rooted binary trees. *Involve*, to appear.

[5] S. Cleary, M. Elder, A. Rechnitzer, and J. Taback. Random subgroups of Thompson's group F. *Groups Geom. Dyn.*, 4(1):91–126, 2010.

[6] S. Cleary and K. St. John. Rotation distance is fixed-parameter tractable. *Inform. Process. Lett.*, 109(16):918–922, 2009.

[7] S. Cleary and K. St. John. A linear-time approximation for rotation distance. *J. Graph Algorithms Appl.*, 14(2):385–390, 2010.

[8] S. Cleary and J. Taback. Geometric quasi-isometric embeddings into Thompson's group F. *New York J. Math.*, 9:141–148 (electronic), 2003.

[9] K. Culik and D. Wood. A note on some tree similarity measures. *Information Processing Letters*, 15(1):39–42, 1982.

[10] P. Dehornoy. On the rotation distance between binary trees. *Adv. Math.*, 223(4):1316–1355, 2010.

[11] P. Flajolet and R. Sedgewick. *Analytic combinatorics*. Cambridge Univ Pr, 2009.

[12] L. Lipshitz. D-finite power series. *J. Algebra*, 122(2):353–373, 1989.

[13] L. Pournin. The diameters of associahedra, 2012. arXiv:1207.6296.

[14] J. Rémy. Un procédé itératif de dénombrement d'arbres binaires et son application à leur génération aléatoire. *RAIRO Inform. Théor.*, 19(2):179–195, 1985.

[15] D. Sleator, R. Tarjan, and W. Thurston. Rotation distance, triangulations, and hyperbolic geometry. *American Mathematical Society*, 1(3), 1988.

[16] K. St. John. personal communication, 2012.

[17] R. Stanley. Differentiably finite power series. *European J. Combin*, 1(2):175–188, 1980.

[18] E. Whittaker and G. Watson. *A course of modern analysis*. Cambridge University Press, 4th edition, 1996.