A Combinatorial Approach to Extracting Rooted Tree Statistics from the Order Quasisymmetric Function

Lucy Cai* Espen Slettnes† Jeremy Zhou‡

March 3, 2020

Abstract

The chromatic symmetric function defined by Stanley is a power series that is symmetric in an infinite number of variables and generalizes the chromatic polynomial. Shareshian and Wachs defined the chromatic quasisymmetric function, and Awan and Bernardi defined an analog of it for digraphs.

Three decades ago, Stanley posed a question equivalent to "Does the chromatic symmetric function distinguish between all trees?" A similar question can be raised for rooted trees: "Does the chromatic quasisymmetric function distinguish between all rooted trees?". Hasebe and Tsujie showed algebraically the stronger statement that the *order* quasisymmetric function distinguishes rooted trees. Here, we aim to directly extract useful statistics about a tree given only its order quasisymmetric function. This approach emphasizes the combinatorics of trees over the the algebraic properties of quasisymmetric functions. We show that a rooted-tree-statistic we name the "co-height profile profile" is extractable, and that it distinguishes rooted 2-caterpillars.

*jcai1@exeter.edu
†slettnes@ohs.stanford.edu
‡jzhou21@andover.edu

1 Introduction

The *chromatic polynomial* $\chi_G(n)$ of a graph is a well-studied graph invariant, defined as the unique polynomial that counts proper colorings of a graph G using n colors.

Stanley [1] defined the *chromatic symmetric function* $X_G(\mathbf{x})$, a symmetric power series in an infinite number of variables that generalizes the chromatic polynomial. Specifically, the chromatic symmetric function is defined for a graph G and indeterminates $\mathbf{x} = (x_1, x_2, \ldots)$ by

$$X_G(\mathbf{x}) = \sum_f \prod_{v \in V(G)} x_{f(v)},$$

where the sum is over all proper colorings $f : V(G) \to \mathbb{N}$. This generalizes the chromatic polynomial by the following: $\chi_G(n) = X_G(\underbrace{1, \ldots, 1}_{n}, 0, \ldots)$.

There is an open question posed by Stanley equivalent to whether or not the chromatic symmetric function uniquely distinguishes all trees, which admitted some partial results but is still being actively studied [1, 4, 2, 8, 3].

Shareshian and Wachs [5] defined a refinement of the chromatic symmetric function for labeled graphs, which they called the *chromatic quasisymmetric function*, defined by

$$X_G(\mathbf{x}, t) = \sum_f \left(t^{\operatorname{asc}(f)} \prod_{v \in V(G)} x_{f(v)} \right),$$

where the sum is again over all proper colorings $f : V(G) \to \mathbb{N}$ and $\operatorname{asc}(f) := |\{\{i, j\} \in E(G) : i < j \text{ and } f(i) < f(j)\}|$. Awan and Bernardi [6, Equation 74] adopted this to define the chromatic quasisymmetric function of an unlabeled acyclic digraph D by setting $\operatorname{asc}(f) := |\{(u, v) \in A(D) : f(u) < f(v)\}|$, where $A(D)$ is the arc set of D.

Taking an analogue of Stanley's question for digraphs, we might ask if the chromatic quasisymmetric function distinguishes rooted trees.

Hasebe and Tsujie [7] showed with an algebraic method that rooted trees are distinguished by what they call the *strict order quasisymmetric function*, defined for a poset P by

$$\Gamma^{<}(P, \mathbf{x}) = \sum_{f \in \operatorname{Hom}^{<}(P, \mathbb{N})} \prod_{v \in P} x_{f(v)},$$

where $\operatorname{Hom}^{<}(P, \mathbb{N})$ is the set of strict homomorphisms from P to $\mathbb{N}$. This is a stronger claim than the question above; for posets, the strict order quasisymmetric function only takes the terms of the chromatic quasisymmetric function where the degree of t is $|E(G)|$. In this paper, we similarly deduce statistics of rooted trees from their order quasisymmetric function, but instead present a combinatorial approach.

In Section 2, we formulate our question of interest precisely and define relevant terminology. In Section 3, we show that a rooted tree statistic we call the "co-height profile profile" can be extracted from the order quasisymmetric function. Finally, in Section 4, we study the information extractable from the co-height profile profile. Among other results, we show that the co-height profile profile distinguishes rooted 2-caterpillars.

2 Terminology

We say that a digraph T is a *rooted tree* if its underlying graph is a tree, and all vertices except for one, the *root*, denoted v_T, have indegree 1. The *leaves* of a rooted tree are the vertices with outdegree 0. The *height* of a vertex v in a rooted tree, denoted H_v, is the length of (number of arcs in) the longest directed path from v to a leaf, and the *co-height*, which we denote h_v, is the length of the directed path from the root v_T to v. We define the *height* of T to be H_{v_T}, and denote it H_T. The i^{th} *layer* of the rooted tree, denoted L_i, is the set of all vertices with co-height $i-1$. See Figure 1 for an example rooted tree labeled with its layers. The subtree with root v, denoted S_v, is the induced subdigraph formed by every vertex w where the directed path from v_T to w goes through v. The *weight* of v is $|S_v|$, which we denote w_v.

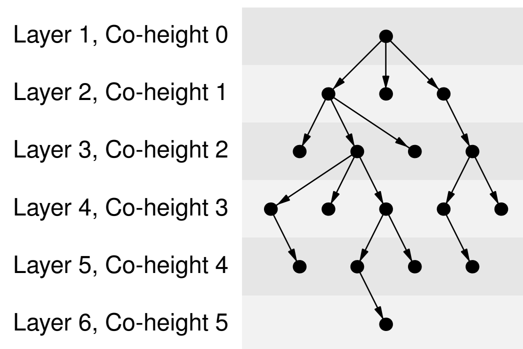

Figure 1:
Co-heights and layers of a rooted tree.

The strict order quasisymmetric function of a poset P and indeterminants $\mathbf{x} = (x_1, x_2, \ldots)$ is

$$\Gamma^{<}(P, \mathbf{x}) = \sum_{f \in \mathrm{Hom}(P,\mathbb{N})} \prod_{a \in P} x_{f(a)}$$

where $\mathrm{Hom}(P, \mathbb{N})$ is the set of homomorphisms from P to $\mathbb{N}$. For any acyclic digraph D, we define $\Gamma^{<}(D, \mathbf{x})$ to be $\Gamma^{<}(P_D, \mathbf{x})$ where P_D is the poset on the vertices of D with the relation $u < v$ iff there is a path from u to v.

We say that a *tree-statistic* is a function from the set of rooted trees to any set. A tree-statistic S is *extractable* if for all rooted trees T_1, T_2 where $\Gamma^{<}(T_1, \mathbf{x}) = \Gamma^{<}(T_2, \mathbf{x})$, $S(T_1) = S(T_2)$, i.e., if $S(T)$ can be extracted from $\Gamma^{<}(T, \mathbf{x})$.

If f is a formal power series in x, then we denote by $[x_i^a]f$ the coefficient of f on the term x_i . For multivariate formal power series, we similarly use e.g. $[x_i^a y_j^b]f$ for the coefficient of f on the monomial $x_i^a y_j^b$.

We denote multisets with double curly braces, e.g.,

$$\left\{\!\!\left\{ \left\{ 2j \,\middle|\, j \in \mathbb{N}, j < \frac{n}{2} \right\} \,\middle|\, n \in \{\!\{0, 2, 3, 3, 4, 5\}\!\} \right\}\!\!\right\} = \{\!\{\varnothing, \varnothing, \{2\}, \{2\}, \{2\}, \{2, 4\}\}\!\}.$$

Definition 2.1. For a rooted tree T, a vertex statistic $(a_u)_{u \in V(T)}$, and a vertex $v \in V(T)$, we define P_v^a to be the multiset $\{\!\{a_u \mid u \in S_v\}\!\}$ and P_T^a to be $P_{v_T}^a$.

Definition 2.2. We define the *height profile*, *co-height profile* and *weight profile* of a vertex or rooted tree X to be P_X^H, P_X^h, and P_X^w, respectively. We can also extend this to say e.g., that the weight profile profile profile of X is $P_X^{P^{P^w}}$.

Example 2.3. The co-height profile profile of the rooted tree T depicted in Figure 2 is

$$P_T^{P^h} = \{\!\!\{ \{\!\!\{2\}\!\!\}, \{\!\!\{3\}\!\!\}, \{\!\!\{3\}\!\!\}, \{\!\!\{3\}\!\!\}, \{\!\!\{4\}\!\!\}, \{\!\!\{4\}\!\!\}, \{\!\!\{4\}\!\!\},$$
$$\{\!\!\{2,3\}\!\!\}, \{\!\!\{3,4\}\!\!\}, \{\!\!\{3,4\}\!\!\}, \{\!\!\{3,4\}\!\!\},$$
$$\{\!\!\{2,3,4\}\!\!\}, \{\!\!\{2,3,3,4\}\!\!\}, \{\!\!\{2,3,3,4\}\!\!\},$$
$$\{\!\!\{1,2,3,3,4\}\!\!\}, \{\!\!\{1,2,2,3,3,4\}\!\!\}, \{\!\!\{1,2,2,3,3,4\}\!\!\},$$
$$\{\!\!\{0,1,1,1,2,2,2,2,2,3,3,3,3,3,3,4,4,4\}\!\!\} \}\!\!\}$$

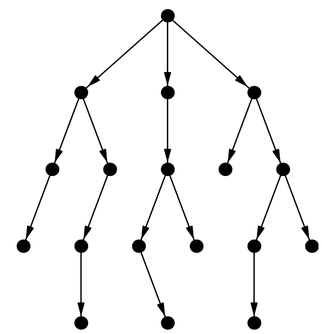

Figure 2: Rooted tree T.

3 Main Result

Throughout this section, we fix a rooted tree T. The main result of this paper is that the tree's co-height profile profile $P_T^{P^h}$ is extractable (Corollary 3.21).

We begin with some definitions.

Definition 3.1. A *coloring* is a function $f : V(T) \to \mathbb{N}$, where $V(T)$ is the vertex set of T. We think of the elements of $\mathbb{N}$ as being colors, so f assigns a color in $\mathbb{N}$ to each vertex in T.

Definition 3.2. A coloring $f : V(T) \to \mathbb{N}$ is *increasing* if for every edge (u, v), f satisfies $f(u) < f(v)$. Viewing T as a poset, this condition is equivalent to f being a homomorphism of posets. Notice that the strict order quasisymmetric function is defined as a sum over increasing colorings, so we only consider increasing colorings in this paper.

Definition 3.3. Given an increasing coloring f, we define

$$[f]\Gamma^<(T, \mathbf{x}) = \prod_{v \in T} x_{f(v)}.$$

In other words, $[f]\Gamma^<(T, \mathbf{x})$ is the term in the strict order quasisymmetric function corresponding to f.

Definition 3.4. Given an increasing coloring f, let $f^{-1}(i)$ be the set of vertices that are colored i in f. Notice that the exponent of x_i in $[f]\Gamma^<(T, \mathbf{x})$ is $|f^{-1}(i)|$.

Definition 3.5. For two infinite tuples $a = (a_1, a_2, \dots)$ and $b = (b_1, b_2, \dots)$, we say that a is *lexicographically less* than b if there exists an i such that $a_1 = b_1, \dots, a_i = b_i$ and $a_{i+1} < b_{i+1}$. Clearly, this defines a total order on infinite tuples.

Definition 3.6. For p a polynomial in $(x_i)_{i \in \mathbb{N}}$, let $\max(p)$ be its lexicographically greatest term, where each term $x_1^{e_1} x_2^{e_2} \cdots$ is considered as the tuple $(e_1, e_2, \dots)$.

To show that the co-height profile profile $P_T^{P^h}$ is extractable, we first extract a simpler tree-statistic, the co-height profile P_T^h.

To do this, we consider a particular increasing coloring $f_\varnothing$ that we can isolate from the strict order quasisymmetric polynomial. In the following lemma, we ascertain exactly what $f_\varnothing$ looks like. Then, in the final proof, we show that $f_\varnothing$ contains enough information to extract P_T^h.

Definition 3.7. Let $f_\varnothing$ be the increasing coloring such that

$$[f_\varnothing]\Gamma^<(T,\mathbf{x}) = \max\big(\Gamma^<(T,\mathbf{x})\big).$$

Equivalently, $f_\varnothing$ is the increasing coloring that maximizes $(|f^{-1}(1)|, |f^{-1}(2)|, \dots)$ with respect to lexicographic order (from now on, all order is lexicographic).

Lemma 3.8. *$f_\varnothing$ is the increasing coloring such that $f_\varnothing^{-1}(i) = L_i$ for all i. One example is shown in Figure 3.*

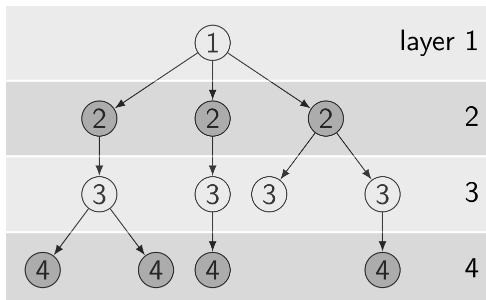

Figure 3: One example of the increasing coloring $f_\varnothing$.

Proof. We proceed by induction.

Base case: By definition, $f_\varnothing$ is the increasing coloring that maximizes $|f^{-1}(1)|$. Note that the only vertex that can be colored 1 is v_T: if any other vertex was colored 1, then the root could not be colored with a positive integer. Thus, the maximum value of $|f^{-1}(1)|$ is 1, attained by setting $f^{-1}(1) = \{v_T\} = L_1$.

Inductive step: Suppose that $f^{-1}(i) = L_i$ for all $i \le n$. Recall that $f_\varnothing$ is the increasing coloring that maximizes $|f^{-1}(n+1)|$. Pick any vertex $v \in L_{n+1}$, and let us restrict our attention to the subtree S_v. Notice that we have the same situation as the base case: the only vertex that can be colored $n+1$ is the root v. Because there are $|L_{n+1}|$ such subtrees S_v, the maximum value of $|f^{-1}(n+1)|$ is $|L_{n+1}|$, attained by setting $f^{-1}(n+1) = L_{n+1}$. This completes the induction. $\square$

Theorem 3.9. *The co-height profile P_T^h is extractable.*

Proof. Recall that if we write $[f_\varnothing]\Gamma^<(T,\mathbf{x})$ as $x_1^{e_1}x_2^{e_2}\cdots$, then by the definition of the strict order quasisymmetric function, $e_i = |f_\varnothing^{-1}(i)|$. We just found in Lemma 3.8 that $|f_\varnothing^{-1}(i)| = |L_i|$. From the $|L_i|$, we can deduce the co-height profile P_T^h because there are exactly $|L_i|$ vertices at co-height $i-1$. Thus, from $[f_\varnothing]\Gamma^<(T,\mathbf{x})$, we can deduce P_T^h. $\square$

To make the next step towards proving the main result, we will perturb $f_\varnothing$ to obtain a new increasing coloring f_n. This new increasing coloring may be thought of as $f_\varnothing$ with one extra "gap."

First, we show how to isolate f_n from the strict order quasisymmetric polynomial. In the following lemma, we ascertain exactly what f_n looks like. Then, in the final proof, we show that f_n contains enough information to extract a useful preliminary result (not yet $P_T^{P^h}$).

Definition 3.10. Given a positive integer n for which T has an nth layer, let f_n be the increasing coloring such that

$$[x_1^{|L_1|} \dots x_{n-1}^{|L_{n-1}|} x_n^{|L_n|-1}][f_n]\Gamma^{<}(T, \mathbf{x}) = \max([x_1^{|L_1|} \dots x_{n-1}^{|L_{n-1}|} x_n^{|L_n|-1}]\Gamma^{<}(T, \mathbf{x})).$$

Equivalently, f_n is the increasing coloring that maximizes $(|f^{-1}(1)|, |f^{-1}(2)|, \dots)$ given that $|f^{-1}(1)|, \dots, |f^{-1}(n-1)|, |f^{-1}(n)|$ are fixed at $|L_1|, \dots, |L_{n-1}|, |L_n| - 1$.

We need one quick definition to state the next lemma.

Definition 3.11. Let $L_j(v)$ be the j^{th} layer of S_v. If $j \le 0$, then we set $L_j(v) = \varnothing$.

Lemma 3.12. *f_n is the increasing coloring such that $f_n^{-1}(i) = L_{i-n}(g) \cup (L_i \setminus L_{i-n+1}(g))$, where $g \in L_n$ is the vertex having the lexicographically least coheight profile out of all of the vertices in L_n. One example is shown in Figure 4.*

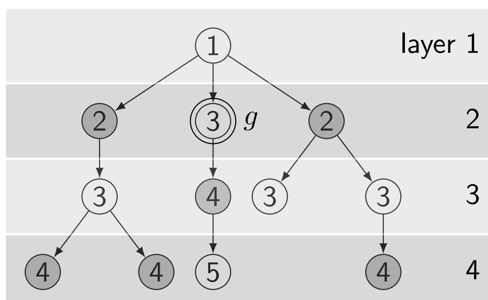

Figure 4: One example of the increasing coloring f_2. The gap g is circled.

Proof. We have that $|f^{-1}(1)| = |L_1| = 1$. As in the proof of Lemma 3.8, note that the only vertex that can be colored 1 is v_T. Thus, the only way to attain $|f^{-1}(1)| = 1$ is by setting $f^{-1}(1) = \{v_T\} = L_1$.

Using a similar induction argument as in Lemma 3.8, we can conclude that $f^{-1}(i) = L_i$ for $i < n$. We also notice that $|f^{-1}(n)|$ is maximized when $f^{-1}(n) = L_n$, but we have fixed $|f^{-1}(n)|$ at $|L_n| - 1$, so there must be one vertex in L_n that is not colored n: what we call a "gap." Let this vertex be g. Soon, we will show that g is the vertex with the lexicographically least co-height profile out of all the vertices in L_n.

Pick any vertex $v \in L_n$, and let us restrict our attention to the subtree S_v. Suppose that its root v is colored $f(v)$. The argument used in Lemma 3.8 gives that for all j, $L_j(v)$ must be completely colored $f(v) + j - 1$. This additionally implies that within S_v, for any color i, the only vertices colored i are $L_{i-f(v)+1}(v)$; in other words, $f^{-1}(i) \cap S_v = L_{i-f(v)+1}(v)$.

Now, remember that f_n is the increasing coloring that maximizes $|f^{-1}(n+1)|$.

First, we consider only the vertices in S_g. Notice that the only vertex in S_g that can be colored $n+1$ is g. Thus, in order to maximize $|f^{-1}(n+1)|$, we must have $f(g) = n + 1$.

Recall that most vertices in L_n are colored n, with the exception of g; in other words, for $v \neq g \in L_n$,

$f(v) = n$. Thus, we have that

$$\begin{aligned} f^{-1}(i) &= \bigcup_{v \in L_n} (f^{-1}(i) \cap S_v) \\ &= \bigcup_{v \in L_n} L_{i-f(v)+1}(v) \\ &= L_{i-n}(g) \cup \bigcup_{v \neq g \in L_n} L_{i-n+1}(v) \\ &= L_{i-n}(g) \cup (L_i \setminus L_{i-n+1}(g)), \end{aligned}$$

as desired.

It remains to show that g is the vertex with the lexicographically least co-height profile out of all the vertices in L_n.

The above formula for $f^{-1}(i)$ gives that

$$|f^{-1}(n+1)| = |L_1(g)| + |L_{n+1}| - |L_2(g)|$$

Given that $|L_1(g)|$ is always 1, the maximum value of $|f^{-1}(n+1)|$ is attained with the choice of g such that $|L_2(g)|$ is the smallest.

If there are multiple g with this minimum, then we maximize $|f^{-1}(n+2)|$. Again, we use the above formula for $f^{-1}(i)$ to obtain

$$|f^{-1}(n+2)| = |L_2(g)| + |L_{n+2}| - |L_3(g)|.$$

Remember that we are only considering g that attain the minimum value of $|L_2(g)|$. Thus, the maximum value of $|f^{-1}(n+2)|$ is attained with the choice of g that minimizes $|L_3(g)|$. Continuing in this fashion, we see that g is the vertex in L_n such that $(|L_2(g)|, |L_3(g)|, \dots)$ is lexicographically least, which is equivalent to P_g^h being lexicographically least. □

Now, to state the next theorem, we need one new piece of notation.

Definition 3.13. Given a multiset of profiles $\mathcal{S}$, we let $\min(\mathcal{S})$ be its lexicographically least element, where each profile is considered as the tuple

$$(\text{multiplicity of } 1, \text{multiplicity of } 2, \dots).$$

Theorem 3.14. *Given a positive integer n for which T has an nth layer, $\min\left(\{\!\{P_v^h \mid v \in L_n\}\!\}\right)$ (the least co-height profile of a vertex in layer n) is extractable.*

Proof. Recall that if we write $[f_n]\Gamma^{<}(T, \mathbf{x})$ as $x_1^{e_1} x_2^{e_2} \cdots$, then by the definition of the strict order quasisymmetric function, $e_i = |f_n^{-1}(i)|$. By the formula for $|f_n^{-1}(i)|$ that we found in Lemma 3.8, we can use the $|f_n^{-1}(i)|$ to find $|L_j(g)|$ for all j:

$$|L_1(g)| = 1,$$

and

$$|L_j(g)| = |L_{n+j-1}| - (|f_n^{-1}(n+j-1)| - |L_{j-1}(g)|).$$

(Remember that by Theorem 3.9, we already know all values of $|L_{n+j-1}|$.)

Thus, knowing $[f_n]\Gamma^<(T,\mathbf{x})$ gives us $(|L_1(g)|, |L_2(g)|, \dots)$. This is equivalent to P_g^h, which by definition of g is the same as $\min\left(\{\{P_v^h \mid v \in L_n\}\}\right)$. □

Remark 3.15. There is a possibility that there are multiple choices for g, i.e. that there are multiple vertices having the least co-height profile out of all the vertices in L_n. In this case, there would not be a unique choice of f_n. However, putting any choice of f_n through the algorithm in Theorem 3.14 will produce the same co-height profile P_g^h, since by definition the choices of g have the same co-height profile.

The above result is not enough to show that P_T^{Ph} is extractable. However, if we generalize the above result from 1 gap to m gaps, then we will be able to prove the main result.

This generalization uses another perturbation of $f_\varnothing$ that we call $f_{n,m}$. Again, we show how to isolate $f_{n,m}$ in the first definition; in the following lemma, we ascertain what $f_{n,m}$ looks like; and in the final proof, we show that $f_{n,m}$ contains enough information to extract the mth least co-height profile in layer n.

Definition 3.16. Given positive integers n, m for which L_n has at least m vertices, let $f_{n,m}$ be the increasing coloring such that

$$[x_1^{|L_1|} \dots x_{n-1}^{|L_{n-1}|} x_n^{|L_n|-m}][f_{n,m}]\Gamma^<(T,\mathbf{x}) = \max([x_1^{|L_1|} \dots x_{n-1}^{|L_{n-1}|} x_n^{|L_n|-m}]\Gamma^<(T,\mathbf{x})).$$

Equivalently, $f_{n,m}$ is the increasing coloring that maximizes $(|f^{-1}(1)|, |f^{-1}(2)|, \dots)$ given that

$|f^{-1}(1)|, \dots, |f^{-1}(n-1)|, |f^{-1}(n)|$ are fixed at $|L_1|, \dots, |L_{n-1}|, |L_n| - m$.

Lemma 3.17. *$f_{n,m}$ is the increasing coloring such that*

$$f_{n,m}^{-1}(i) = \bigcup_{1\le k\le m} L_{i-n}(g_k) \cup \left(L_i \setminus \bigcup_{1\le k\le m} L_{i-n+1}(g_k) \right),$$

where $g_1, \dots, g_m \in L_n$ are the m vertices having the lexicographically least coheight profiles out of all of the vertices in L_n. One example is shown in Figure 5.

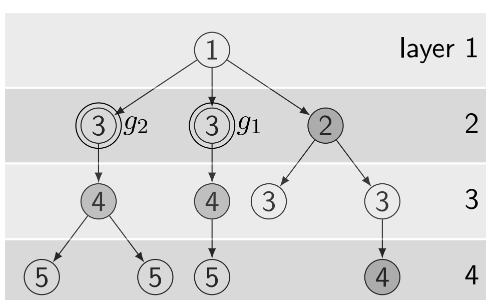

Figure 5: One example of the increasing coloring $f_{2,2}$. The gaps g_1 and g_2 are circled.

Proof. Let $g_1, \dots, g_m \in L_n$ be the m vertices with the lexicographically least coheight profiles out of all the vertices in L_n. Using the same reasoning as Lemma 3.12, $f_{n,m}$ is the coloring such that:

- L_i for $i < n$ is completely colored i,
- $L_j(g_k)$ for any j, k is colored $n + j$,
- and $L_j(v)$ for any $j, v \neq g_k \in L_n$ is colored $n + j - 1$.

This means that

$$f_{n,m}^{-1}(i) = \bigcup_{1 \le k \le m} L_{i-n}(g_k) \cup \left(L_i \setminus \bigcup_{1 \le k \le m} L_{i-n+1}(g_k) \right).$$

□

To state the next theorem, we generalize the $\min(\mathcal{S})$ notation:

Definition 3.18. We extend the $\min(\mathcal{S})$ notation to $\min_m(\mathcal{S})$. Recall that $\min(\mathcal{S})$ is the lexicographically least element of $\mathcal{S}$; we let $\min_m(\mathcal{S})$ be the mth lexicographically least element of $\mathcal{S}$.

Theorem 3.19. *Given positive integers n, m for which L_n has at least m vertices, $\min_m\left(\{\!\{P_v^h \mid v \in L_n\}\!\}\right)$ (the mth least co-height profile of a vertex in layer n) is extractable.*

Proof. Recall that if we write $[f_{n,m}]\Gamma^<(T, \mathbf{x})$ as $x_1^{e_1} x_2^{e_2} \cdots$, then by the definition of the strict order quasisymmetric function, $e_i = |f_{n,m}^{-1}(i)|$. Just as in Theorem 3.14, we can use the $|f_{n,m}^{-1}(i)|$ to find $\left|\bigcup_{1 \le k \le m} L_j(g_k)\right|$ for all j:

$$\left| \bigcup_{1 \le k \le m} L_1(g_k) \right| = m$$

and

$$\left| \bigcup_{1 \le k \le m} L_j(g_k) \right| = |L_{n+j-1}| - \left(|f_{n,m}^{-1}(n + j - 1)| - \left| \bigcup_{1 \le k \le m} L_{j-1}(g_k) \right| \right).$$

(Remember that by Theorem 3.9, we already know all values of $|L_{n+j-1}|$.)

If we do the same procedure for $[f_{n,m-1}]\Gamma^<(T, \mathbf{x})$, then we can find $\left|\bigcup_{1 \le k \le m-1} L_j(g_k)\right|$ for all j. Then, since the S_{g_k} are disjoint, we have

$$L_j(g_m) = \left| \bigcup_{1 \le k \le m} L_j(g_k) \right| - \left| \bigcup_{1 \le k \le m-1} L_j(g_k) \right|.$$

Thus, from $[f_{n,m}]\Gamma^<(T, \mathbf{x})$ and $[f_{n,m-1}]\Gamma^<(T, \mathbf{x})$ we can deduce $(L_1(g_m), L_2(g_m), \ldots)$. This is equivalent to $P_{g_m}^h$, which is by definition of g_m the same as $\min_m\left(\{\!\{P_v^h \mid v \in L_n\}\!\}\right)$. □

Remark 3.20. There is a possibility that the choices of $g_1, \ldots, g_m$ are not unique. In this case, just as in Remark 3.15, any choices of $f_{n,m}$ and $f_{n,m-1}$ will work: putting any choices of $f_{n,m}$ and $f_{n,m-1}$ through the algorithm in Theorem 3.19 will produce the same co-height profile.

Now, we finally have enough information to prove the main result.

Corollary 3.21. *The co-height profile profile $P_T^{P^h}$ is extractable.*

Proof. By Theorem 3.19, we know that $\min_m\left(\{\!\{P_v^h \mid v \in L_n\}\!\}\right)$ is extractable for all positive integers n, m such that L_n has at least m vertices.

Notice that for a fixed n, compiling $\min_m\left(\{\!\{P_v^h \mid v \in L_n\}\!\}\right)$ for every $1 \leq m \leq |L_n|$ gives the entire multiset $\{\!\{P_v^h \mid v \in L_n\}\!\}$. Then, compiling $\{\!\{P_v^h \mid v \in L_n\}\!\}$ for every $1 \leq n \leq H_T$ gives the coheight profile profile $P_T^{P^h}$. ☐

4 Consequences

We show that the coheight profile profile P^{P^h} allows us to extract a lot of information about rooted trees. In this section, we present four examples of tree statistics that can be derived from the coheight profile profile, but cannot be derived from just the coheight profile.

Corollary 4.1. *The weight profile (P^w) of layer L_i is extractable.*

Proof. For a given vertex v in layer L_i of tree T, its weight $|S_v|$ can be determined by counting the number of vertices in its coheight profile P_v^h. The weight profile of layer L_i is then obtained by taking the multiset of this result over all vertices in layer L_i. ☐

Corollary 4.2. *The height profile (P^H) of layer L_i is extractable.*

Proof. For a given vertex v in layer i of tree T, its height H_v can be determined by counting the number of layers in its coheight profile P_v^h. Taking the multiset of this result over all vertices in L_i gives us the height profile of layer i, and also of the entire tree T. ☐

Corollary 4.3. *The outdegree distribution of the vertices in L_i is extractable.*

Proof. For a vertex v in layer i, the outdegree of v can be determined by counting the number of vertices in $L_2(S_v)$. Taking the multiset of this result over all vertices in L_i gives us the outdegree distribution of layer i, and thus of the entire tree T. ☐

Corollary 4.4. *The number of leaves in L_i is extractable.*

Proof. The number of leaves in L_i can be determined by counting the number of vertices in layer i that have the same coheight profile as that of a leaf. ☐

Our main consequence is that the coheight profile profile can distinguish between rooted 2-caterpillars.

Definition 4.5. Define the *distance* between two vertices to be the length of the shortest path between them. Then, an *n-caterpillar tree* is rooted directed tree in which there exists a directed path of vertices $(v_1, v_2, \ldots v_k)$ such that for every vertex w there exists a vertex v_i in which the distance between w and v_i is at most n. The path $(v_1, v_2, \ldots v_k)$ is called the *spine* of the tree.
One example is shown in Figure 6.

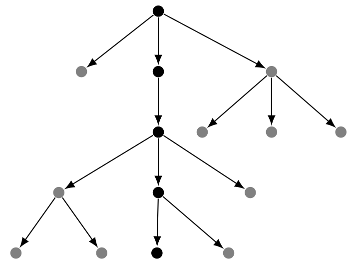

Figure 6: An example of a 2-caterpillar tree, in which the spine is the central column of vertices.

Corollary 4.6. *If T_1 and T_2 are two 2-caterpillar trees with the same coheight profile profile, then $T_1 = T_2$.*

Proof. Consider a 2-caterpillar tree C. Note that each vertex v in C that does not belong to the spine must belong to one of three types (which we denote types 1, 2, and 3):

1. a leaf that is a child of a vertex on the spine.
2. a leaf that is a grandchild of a vertex on the spine.
3. a vertex that is a child of a vertex on the spine, with S_v a tree of height 1.

We use an inductive argument. We know L_1 must consist of one vertex, the root v_C. Then, suppose C can be uniquely determined up to layer $n-1$. In L_n, there must be exactly one vertex on the spine. From $\bigcup_{v \in L_n} P_v^h$, the number of type 3 vertices and the number of leaves (of either type 1 or 2) in layer n can be determined. However, notice that all type 2 vertices in L_n are uniquely determined by the P_v^h of type 3 vertices in layer $n-1$. Thus, the remaining leaves in L_n must all be type 1. In this way, it is possible to uniquely determine the vertices in layer n for $n = 1, 2, \ldots, H_C$ using $\bigcup_{v \in L_n} P_v^h$ and $\bigcup_{v \in L_{n-1}} P_v^h$, both of which are derived from the coheight profile profile. C itself is then uniquely determined. □

Remark 4.7. It follows that the coheight profile profile can distinguish between all 1-caterpillars.

Remark 4.8. Since every tree of height 2 is expressible as a 2-caterpillar, the coheight profile profile can distinguish between all trees of height 2.

5 Limitations

Our methods are capable of distinguishing trees of height 2, but not all trees of height 3. For example, the two trees in Figure 7 have the same coheight profile profile, and are thus indistinguishable by our current method.

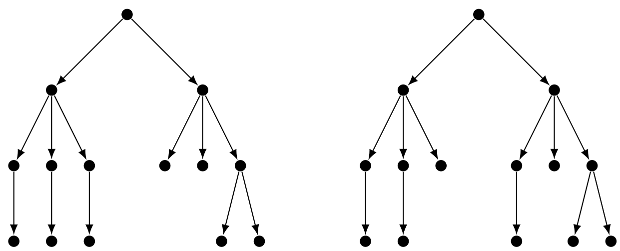

Figure 7: Two distinct trees with the same coheight profile profile.

It would be interesting to find a combinatorial method of proving trees of height 3 are distinguishable using a similar argument to the coheight profile profile.

Since the most powerful tree-statistic we have found so far is the coheight profile profile, it is natural to consider families of trees that share this statistic. We are interested, then, in the following question:

1. Under what conditions will two trees T_1, T_2 share the same coheight profile profile?

We attempt to make progress at this question by considering a narrower situation. We first introduce a definition.

Definition 5.1. In a rooted directed tree with root r, let the path from r to a vertex m be $(r = x_1, x_2, \dots x_{h_m+1} = m)$ and let the path from r to a vertex n in the same layer as m be $(r = y_1, y_2, \dots y_{h_n+1} = n)$. Then, vertices m and n are *path compatible* if $P^h_{x_i} = P^h_{y_i}$ for all $1 \leq i \leq h_m = h_n$.

Given a new rooted directed tree T_1, we choose two subtrees S_a, S_b of T_1 such that $h_a = h_b$. Create a second tree T_2 identical to T_1 except that the locations of S_a and S_b are swapped.

Proposition 5.2. *Two such trees T_1, T_2 share the same coheight profile profile if and only if at least one of the two following conditions are satisfied:*

1. *a, b share the same coheight profile.*
2. *Let the least common ancestor of a and b be vertex v. Let the parents of a, b be $\hat{a}, \hat{b}$. Let T be the tree obtained by deleting S_a and S_b from T_1. Then, $\hat{a}$ and $\hat{b}$ are path compatible in T.*

Proof. For the sake of simplicity, redefine T_1, T_2 to be their subtrees having v as the root, since these are the only vertices whose coheight profiles are affected by the swapping of S_a and S_b. To force T_1 and T_2 to have the same coheight profile profile, we use casework, starting from the root layer and traversing down. Let $T_i(m)$ be the vertex m as an element of tree T_i. In T_1, let the paths from v_{T_1} to $T_1(a)$ and $T_1(b)$ be $(v = T_1(p_1), T_1(p_2), \dots T_1(p_{h_a+1}) = a)$ and $(v = T_1(q_1), T_1(q_2), \dots T_1(q_{h_b+1}) = b)$, and define the paths from from v_{T_2} to $T_2(a)$ and $T_2(b)$ analogously. The idea is that in each layer L_i, either $T_1(p_i)$ and $T_2(p_i)$ will have the same coheight profile, or $T_1(p_i)$ and $T_2(q_i)$ will have the same coheight profile.

Layer 1: $P^h_{T_1} = P^h_{T_2}$. This is always true.

Layer 2: $\{P^h_{T_1(p_2)}, P^h_{T_1(q_2)}\} = \{P^h_{T_2(p_2)}, P^h_{T_2(q_2)}\}$. Here we have two cases:

Case 1: $P^h_{T_1(p_2)} = P^h_{T_2(p_2)}$ and $P^h_{T_1(q_2)} = P^h_{T_1(q_2)}$. Since $S_{T_1(p_2)} \backslash S_a$ and $S_{T_1(q_2)} \backslash S_b$ remain constant after swapping, this occurs only if $P^h_a = P^h_b$ (S_a and S_b share the same coheight profile, condition 1).

Case 2: $P^h_{T_1(p_2)} = P^h_{T_2(q_2)}$ and $P^h_{T_1(q_2)} = P^h_{T_2(p_2)}$. Since S_a and S_b remain constant, this occurs only if $P^h_{T_1(S_{p_2} \backslash S_a)} = P^h_{T_2(S_{p_2} \backslash S_a)}$ and $P^h_{T_1(S_{q_2} \backslash S_b)} = P^h_{T_2(S_{q_2} \backslash S_b)}$, so in general, $P^h_{S_{p_2} \backslash S_a} = P^h_{S_{q_2} \backslash S_b}$ (condition 2).

Continuing in this manner, we see that for each successive layer, it is always the case that at least one of the two conditions must be satisfied. The other direction is clear: if either one of the two conditions are satisfied, T_1 and T_2 must share the same coheight profile profile. This concludes our proof. □

Remark 5.3. In the scenario above, to create tree T_2 from T_1, we swapped S_a and S_b, the entire subtrees defined by a and b. There is a generalized version of proposition 5.2 where we swap only the subtrees defined by j children branching off each of a and b, where j is less than the outdegree of either a or b. This is clear by applying proposition 5.2 multiple times on the layer below a and b. In this case, analogous results occur, and the straightforward statement and proof is left to the reader.

6 Conclusion

In this paper, we were able to extract the co-height profile profile of T directly from its order quasisymmetric function $\Gamma^<(T, \mathbf{x})$. It follows from the work of Hasabe and Tsujie [7] that knowing $\Gamma^<(T, \mathbf{x})$ is enough to determine the co-height profile … profile for arbitrarily many iterations of "profile", but the method in [7] is not algorithmic. It would be interesting if a method were found to extract these arbitrary iterations of "profile" directly from $\Gamma^<(T, \mathbf{x})$. For example, although the two trees in Figure 7 have the same co-height profile profile, they do not have the same co-height profile profile profile. If a method were found to extract arbitrarily many iterations of "profile", it would completely determine any tree T directly from its $\Gamma^<(T, \mathbf{x})$.

7 Acknowledgments

Special thanks to our MIT PRIMES-USA mentor Duncan Levear and his advisor Olivier Bernardi of Brandeis University for providing us with helpful guidance. We are also grateful to Dr. Pavel Etingof, Dr. Slava Gerovitch, and Dr. Tanya Khovanova of MIT Mathematics for valuable feedback and for making the PRIMES research program available, thus providing us with the opportunity to work on this project.

References

[1] Richard. P. Stanley. "A Symmetric Function Generalization of the Chromatic Polynomial of a Graph". In: *Advances in Mathematics* 111.1 (1995), pp. 166–194. ISSN: 0001-8708. DOI: `https:`

//doi.org/10.1006/aima.1995.1020. URL: http://www.sciencedirect.com/science/article/pii/S0001870885710201.

[2] Jeremy L. Martin, Matthew Morin, and Jennifer D. Wagner. "On distinguishing trees by their chromatic symmetric functions". In: *Journal of Combinatorial Theory, Series A* 115.2, math/0609339 (Sept. 2006), pp. 237–253. ISSN: 0097-3165. DOI: https://doi.org/10.1016/j.jcta.2007.05.008. arXiv: math/0609339 [math.CO].

[3] José Aliste-Prieto and José Zamora. "Proper caterpillars are distinguished by their symmetric chromatic function". In: *Discrete Mathematics* 315-316.1, arXiv:1208.2267 (Jan. 2014), pp. 158–164. ISSN: 0012-365X. DOI: 10.1016/j.disc.2013.10.016. arXiv: 1208.2267 [math.CO].

[4] Martin Loebl and Jean-Sébastien Sereni. "Potts partition function and isomorphisms of trees". In: (May 2014). URL: https://iti.mff.cuni.cz/series/2014/611.pdf.

[5] J. Shareshian and M. L. Wachs. "Chromatic quasisymmetric functions". In: *arXiv e-prints*, arXiv:1405.4629 (May 2014), arXiv:1405.4629. arXiv: 1405.4629 [math.CO].

[6] Jordan Awan and Olivier Bernardi. "Tutte Polynomials for Directed Graphs". In: *arXiv e-prints*, arXiv:1610.01839 (Oct. 2016), arXiv:1610.01839. arXiv: 1610.01839 [math.CO].

[7] T. Hasebe and S. Tsujie. "Order Quasisymmetric Functions Distinguish Rooted Trees". In: *arXiv e-prints*, arXiv:1610.03908 (Oct. 2016), arXiv:1610.03908. arXiv: 1610.03908 [math.CO].

[8] José Aliste-Prieto, Anna de Mier, and José Zamora. "On trees with the same restricted U-polynomial and the Prouhet-Tarry-Escott problem". In: *Discrete Mathematics* 340.6, arXiv:1509.09210 (June 2017), pp. 1435–1441. ISSN: 0012-365X. DOI: 10.1016/j.disc.2016.09.019. arXiv: 1509.09210 [math.CO].