CHROMATIC SYMMETRIC FUNCTIONS AND POLYNOMIAL INVARIANTS OF TREES

JOSÉ ALISTE-PRIETO, JEREMY L. MARTIN, JENNIFER D. WAGNER, AND JOSÉ ZAMORA

ABSTRACT. Stanley asked whether a tree is determined up to isomorphism by its chromatic symmetric function. We approach Stanley's problem by studying the relationship between the chromatic symmetric function and other invariants. First, we prove Crew's conjecture that the chromatic symmetric function of a tree determines its generalized degree sequence, which enumerates vertex subsets by cardinality and the numbers of internal and external edges. Second, we prove that the restriction of the generalized degree sequence to subtrees contains exactly the same information as the subtree polynomial, which enumerates subtrees by cardinality and number of leaves. Third, we construct arbitrarily large families of trees sharing the same subtree polynomial, proving and generalizing a conjecture of Eisenstat and Gordon.

1. INTRODUCTION

The *chromatic symmetric function* $\mathbf{X}_G$ of a graph G enumerates proper colorings of G by their distributions of colors. Introduced by Stanley in [Sta95], the chromatic symmetric function (henceforth CSF) is a far-reaching generalization of the classical chromatic polynomial introduced by Birkhoff [Bir13]. It is closely related to other important invariants in algebraic combinatorics, including Noble and Welsh's *U-polynomial* [NW99], whose definition is motivated by knot theory, and Billera, Jia and Reiner's *quasisymmetric function of matroids* [BJR09]. The CSF plays a key role in the theory of combinatorial Hopf algebras [ABS06], arising as the canonical morphism from graphs to quasisymmetric functions. It has natural analogues in noncommutative symmetric functions [GS01] and quasisymmetric functions [SW16], with applications including the cohomology of Hessenberg subvarieties of flag manifolds. Variations of the CSF have been developed for directed graphs, rooted trees, etc.: e.g., [APdMZ17, Ell17, AWvW21, ADM22, Paw22, LW24].

Stanley's original article posed the question of whether the CSF is a complete isomorphism invariant for trees, i.e., whether $\mathbf{X}_T = \mathbf{X}_{T'}$ implies $T \cong T'$. The problem remains unsolved despite considerable attention. There is neither an easy way to construct two nonisomorphic trees with the same CSF, nor to extract sufficient local information to reconstruct a tree from the global data encoded in the CSF. The distinguishing power of the CSF remains mysterious, in sharp contrast to weaker invariants like the chromatic polynomial (which provides no information about a tree other than the number of vertices) or versions of the CSF for labeled graphs (e.g., the noncommutative CSF of labeled trees studied by Gebhard and Sagan [GS01] is easily seen to be a complete invariant). The uniqueness problem is also well understood for non-tree graphs: Stanley's original paper gave two five-vertex graphs with the same CSF, and Orellana and Scott [OS14] constructed an infinite family of pairs of unicyclic graphs (i.e., graphs that can be made into trees by deleting one edge) with the same CSF. Additional families of graphs with the same CSFs are constructed in [APCSZ21].

Partial progress has been made on Stanley's question. Heil and Ji [HJ19] have verified that the CSF is a complete invariant for trees with up to 29 vertices. The conjecture is also known for certain special classes of trees, including spiders [MMW08], caterpillars [APZ14, LS19], 2-spiders [HC20], and proper trees with diameter at most five [APdMOZ23].

One approach to Stanley's problem is by comparing the CSF to other polynomial graph invariants. Martin, Morin and Wagner [MMW08] proved that the CSF determines the *subtree polynomial* $\mathbf{S}_T$, first studied by

Date: February 15, 2024.

2020 *Mathematics Subject Classification.* 05C05, 05C31, 05C60, 05E05.

Key words and phrases. Chromatic symmetric function, generalized degree sequence, subtree polynomial, Crew's conjecture, Eisenstat-Gordon conjecture.

The second and third authors thank their home institutions for granting sabbatical leave in Fall 2023, and Universidad Andres Bello for hosting them during that time, when this research was carried out.

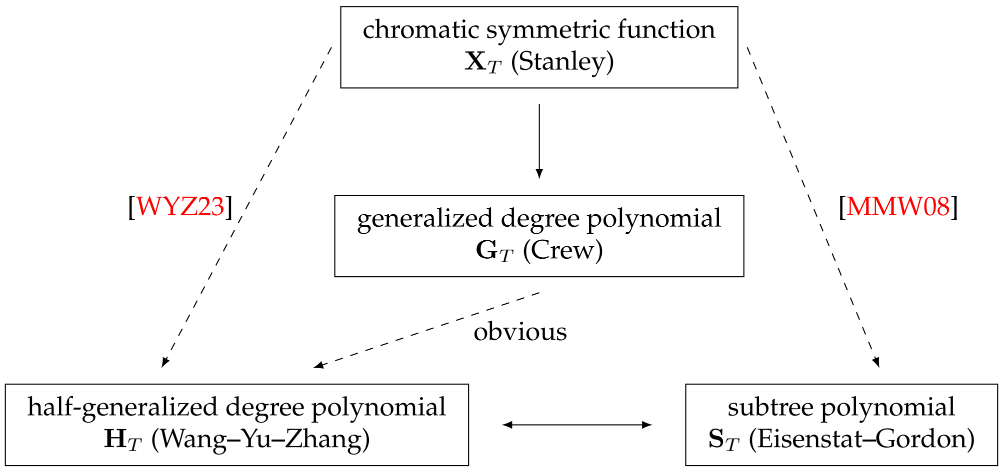


FIGURE 1. Relationships between isomorphism invariants of trees. The invariant at the tail of an arrow determines the invariant at the head. Solid arrows are results of this paper; dotted arrows indicate previously known results.

Eisenstat and Gordon [EG06], which enumerates subtrees of T by size and number of leaves. In particular, this result implies that the degree and path sequences of a tree are recoverable from its CSF. In this paper, we build on the results of [MMW08] by describing the hierarchy among several polynomial invariants of trees, focusing on the three-variable *generalized degree polynomial* $\mathbf{G}_T$ introduced by Crew [Cre22]; the textithalf-generalized degree polynomial, a two-variable specialization $\mathbf{H}_T$ of $\mathbf{G}_T$, studied by Wang, Yu and Zhang [WYZ23]; and the two-variable Eisenstat-Gordon *subtree polynomial* $\mathbf{S}_T$. We will give precise definitions of all these invariants in Section 2.

Our results are as follows:

First, we prove (Theorem 2) that the CSF of a tree determines its generalized degree polynomial linearly. This result was conjectured by Crew in [Cre22], and strengthens the result of Wang, Yu and Zhang [WYZ23] that $\mathbf{X}_T$ determines $\mathbf{H}_T$. Our proof of Theorem 2 is completely explicit: we exhibit an integer matrix that transforms the vector of coefficients of $\mathbf{X}_T$, written in the power-sum basis, into the vector of coefficients of $\mathbf{G}_T$. The entries of the matrix are purely combinatorial and depend only on the number of vertices of T. The proof is given in Section 3.

Second, we prove (Theorem 4) that the polynomials $\mathbf{S}_T$ and $\mathbf{H}_T$ are related by an invertible linear transformation, hence contain the same information. To prove this result, we construct square integer matrices M, N, invertible over $\mathbb{Q}$, such that $Mf = Ng$, where f, g are the vectors of coefficients of $\mathbf{H}_T$ and $\mathbf{S}_T$ respectively. This proof is somewhat less combinatorial than that of Theorem 2 in the sense that there are no direct combinatorial interpretations for the entries of the transition matrices $N^{-1}M$ and $M^{-1}N$ (indeed, N is not invertible over $\mathbb{Z}$). The proof is given in Section 4.

Third, we show (Theorem 9) how to construct arbitrarily large families of non-isomorphic trees sharing the same half-generalized degree polynomial, or, equivalently, the same subtree polynomial. The trees in question are all *caterpillars*, which are indexed by integer compositions. Billera, Thomas and van Willigenburg [BTvW06, Theorem 3.6] showed that every composition α admits a certain unique irreducible factorization of the form $\alpha_1 \circ \cdots \circ \alpha_k$ (the operation $\circ$ is defined in Section 6 below). We show that replacing any of the α_i's with their reversals produces a caterpillar with the same half-generalized degree polynomial. This result proves and strengthens a conjecture of Eisenstat and Gordon [EG06, Conjecture 2.8] on constructing pairs of caterpillars with the same subtree polynomial. The main tool in the proof is a recurrence for the half-generalized degree polynomial of a tree, using the operation of *near-contraction* introduced in [APdMOZ23]. The recurrence is proved in Section 5 and the construction of half-generalized degree polynomial equivalence classes in Section 6.

The relationships between these invariants are depicted in Figure 1.

In the final section of the paper, we discuss possible directions for further research, including extending the constructions to non-caterpillar trees; comparison of the distinguishing power of the generalized and half-generalized degree polynomials (surprisingly, we cannot certify that the former is a strictly stronger invariant); and another polynomial that simultaneously generalizes the half-generalized degree polynomial and the subtree polynomial.

2. Background

We make use of standard definitions and basic facts from graph theory. A **tree** is a connected acyclic graph $T = (V, E)$ with at least one vertex. A **subtree** S of a tree T is a connected subgraph whose vertex set $V(S)$ is nonempty. The set of subtrees of T is denoted $\mathcal{S}(T)$.

2.1. **The chromatic symmetric function.** Let $G = (V, E)$ be a simple graph with $|V| = n$. The **chromatic symmetric function**, introduced by Stanley [Sta95] is the formal power series

$$\mathbf{X}_G = \sum_{\substack{\text{proper colorings} \\ \kappa: V \to \mathbb{N}_{>0}}} \prod_{v \in V} x_{\kappa(v)}$$

which is a symmetric function in commuting variables $x_1, x_2, \ldots$, homogeneous of degree n.

Consider the expansion of the chromatic symmetric function of a graph G in the power-sum basis:

$$\mathbf{X}_G = \sum_{\lambda \vdash n} c_\lambda(G) p_\lambda.$$

Stanley gave a combinatorial description of the numbers $c_\lambda(G)$ [Sta95, Thm. 2.5], which is cancellation-free if and only if $G = T$ is a tree. In that case, as observed in [MMW08, eqn. (7)], one has

$$c_\lambda(T) = (-1)^{n-\ell(\lambda)} |\{F \subseteq E : \operatorname{type}(F) = \lambda\}|. \tag{1}$$

Here $\operatorname{type}(F)$ is the partition whose parts are the sizes of the connected components of the graph (V, F). Thus we can regard the numbers $c_\lambda(T)$ as graph invariants derivable from the chromatic symmetric function.

2.2. **The generalized and half-generalized degree polynomials.** Let $T = (V, E)$ be a tree with $|V| = n$. For a vertex set $A \subseteq V$, define

$$E(A) = \{\text{edges of } E \text{ with both endpoints in } A\}, \qquad e(A) = |E(A)|,$$
$$D(A) = \{\text{edges of } E \text{ with exactly one endpoint in } A\}, \qquad d(A) = |D(A)|,$$

and let

$$g_T(a, b, c) = |\{A \subseteq V(T) : |A| = a,\ d(A) = b,\ e(A) = c\}|. \tag{2}$$

The **generalized degree polynomial** (or GDP) of T, introduced by Crew [Cre20], is

$$\mathbf{G}_T = \mathbf{G}_T(x, y, z) = \sum_{A \subseteq V} x^{|A|} y^{d(A)} z^{e(A)} = \sum_{a,b,c} g_T(a, b, c) x^a y^b z^c. \tag{3}$$

The reason for the name is that when $A = \{v\}$, the number $d(A)$ is just the degree of vertex v. (Crew [Cre20, §4.3] introduced the **generalized degree sequence** of T as the multiset of triples $(|A|, d(A), e(A))$ for $A \subseteq V$; our generating function is equivalent.)

For $A \subseteq V$, let $T|_A$ denote the subgraph induced by A, and define

$$h_T(b, c) = |\{A \subseteq V : T|_A \text{ connected},\ d(A) = b,\ e(A) = c\}| = g_T(c+1, b, c). \tag{4}$$

The **half-generalized degree polynomial** (or HDP) of T is defined as

$$\mathbf{H}_T = \mathbf{H}_T(y, z) = \sum_{\substack{\emptyset \neq A \subseteq V \\ T|_A \text{ connected}}} y^{d(A)} z^{e(A)} = \sum_{S \in \mathcal{S}(T)} y^{d(S)} z^{e(S)} = \sum_{b,c} h_T(b, c) y^b z^c. \tag{5}$$

Evidently, if $|A| = 0$, then $(d(A), e(A)) = (0, 0)$, and if $|A| = n$, then $(d(A), e(A)) = (0, n-1)$. Otherwise, if $1 \le |A| \le n - 1$, then:

- $1 \le d(A) \le n - 1$;
- $0 \le e(A) < |A|$; and

- $|A| \le d(A) + e(A) \le n - 1$.

The only inequality that is not immediate is $d(A) + e(A) \ge |A|$. To see this, observe that the forest $T|_A$ has $e(A)$ edges and $|A| - e(A)$ components. Each component K has at least one edge with one endpoint in K and one in $V(T) \setminus A$. These edges are all distinct external edges, so $|A| - e(A) \le d(A)$. Accordingly, we can write the generalized and half-generalized degree polynomials as

$$\mathbf{G}_T(x,\, y,\, z) = 1 + x^n z^{n-1} + \sum_{a=1}^{n-1} \sum_{c=0}^{a-1} \sum_{b=1}^{n-1-c} g_T(a,b,c) x^a y^b z^c,$$

$$\mathbf{H}_T(y,\, z) = 1 + z^{n-1} + \sum_{c=0}^{n-1} \sum_{b=1}^{n-1-c} h_T(b,c) y^b z^c.$$

Crew [Cre22] conjectured that $\mathbf{G}_T$ can be recovered from $\mathbf{X}_T$, and Wang, Yu, and Zhang [WYZ23, Thm. 5.3] proved that $\mathbf{H}_T$ can be recovered linearly from $\mathbf{X}_T$; that is, the coefficients $h_T(b,c)$ of the HDP are linear functions of the coefficients $c_\lambda(T)$ of (1).

As observed by Crew [Cre20, pp. 83–84], the generalized degree polynomial is not a complete invariant for trees. The two smallest trees with the same GDP are shown in Figure 2. These are also the smallest trees with the same HDP.

FIGURE 2. The smallest trees with the same generalized degree polynomial and the same subtree polynomial.

2.3. **The subtree polynomial.** Let $T = (V, E)$ be a tree with $|V| = n$, and let $S \in \mathcal{S}(T)$. An edge of S is called a *leaf edge* if at least one of its endpoints is a leaf of S. Let

$$\begin{aligned} E(S) &= E(V(S)) & e(S) &= |E(S)|, \\ D(S) &= D(V(S)), & d(S) &= |D(S)|, \\ L(S) &= \{\text{leaf edges of } S\}, & \ell(S) &= |L(S)|. \end{aligned}$$

and define

$$s_T(i,j) = |\{U \in \mathcal{S}(T) \colon e(U) = i,\ \ell(U) = j\}|. \tag{6}$$

The **subtree polynomial** (or STP) of T, introduced by Eisenstat and Gordon [EG06] (and equivalent to the *greedoid Tutte polynomial* introduced in [GM89]) is defined as

$$\mathbf{S}_T(q,r) = \sum_{S \in \mathcal{S}(T)} q^{e(S)} r^{\ell(S)} = \sum_{i,j} s_T(i,j) q^i r^j. \tag{7}$$

Martin, Morin and Wagner [MMW08] proved that the chromatic symmetric function determines the subtree polynomial linearly. The STP is not a complete tree invariant; as for the GDP and HDP, the two trees shown in Figure 2 are the smallest pair with the same STP.

3. THE CHROMATIC SYMMETRIC FUNCTION DETERMINES THE GENERALIZED DEGREE POLYNOMIAL

In this section we prove that the coefficients of the generalized degree polynomial of a tree are determined linearly by the coefficients of the power-sum expansion of its chromatic symmetric function. This result proves Crew's conjecture from [Cre22].

Throughout this section, let $T = (V, E)$ be a tree with $|V| = n$. For $F \subseteq E$ and $A \subseteq V$, write $F(A) = F \cap E(A)$ and $F(\bar{A}) = F \cap E(\bar{A})$. Say that A is ***F*-pure** if A is a union of vertex sets of components of F. Equivalent conditions are $F \cap D(A) = \emptyset$ and $F \subseteq E(A) \cup E(\bar{A})$. More specifically, say that A is ***F*-pure of type μ** if A is a union of vertex sets of components of F whose sizes are the parts of μ (so, in particular, $|A| = |\mu|$).

For partitions λ and μ define

$$\binom{\lambda}{\mu} := \prod_{i=1}^{n} \binom{m_i(\lambda)}{m_i(\mu)}$$

where $m_i(\lambda)$ is the number of occurrences of i as a part of λ. Observe that if $\mathrm{type}(F) = \lambda$, then the number of F-pure sets of type μ is $\binom{\lambda}{\mu}$.

We will need the following simple combinatorial identity:

Lemma 1. *For all sets P and numbers q, we have*

$$\sum_{F \subseteq P} (-1)^{|F|+q} \binom{|F|}{q} = \begin{cases} 1 & \text{if } |P| = q, \\ 0 & \text{if } |P| \neq q. \end{cases}$$

Proof. Let $p = |P|$. Using a standard binomial identity [GKP94, equation (5.21), p.167], we have

$$\sum_{k=0}^{p} (-1)^k \binom{p}{k}\binom{k}{q} = \sum_{k=0}^{p} (-1)^k \binom{p}{q}\binom{p-q}{k-q} \quad = \binom{p}{q} \sum_{k=q}^{p} (-1)^k \binom{p-q}{k-q} = (-1)^q \binom{p}{q} \sum_{j=0}^{p-q} (-1)^j \binom{p-q}{j}.$$

The sum is the binomial expansion of $(1-1)^{p-q}$. In particular it vanishes unless $p = q$, in which case the entire expression is 1. □

For $\lambda \vdash n$ and numbers a, b, c, define

$$\omega(\lambda,\, a,\, b,\, c) = (-1)^{n-b-1} \sum_{\mu \vdash a} \binom{a-\ell(\mu)}{c} \binom{\lambda}{\mu} \binom{n-\ell(\lambda)+\ell(\mu)-a}{n-b-c-1}.$$

Theorem 2. *The chromatic symmetric function determines the generalized degree polynomial linearly.*

Proof. We will prove that

$$g(a,b,c) = \sum_{\lambda \vdash n} c_\lambda(T)\omega(\lambda,a,b,c) \tag{8}$$

where $g_T(a,b,c)$ is defined as in (2).

Let R be the right-hand side of (8). We start by plugging in the definitions of $c_\lambda(T)$ and $\omega(\lambda,a,b,c)$:

$$\begin{aligned}
R &= \sum_{\lambda \vdash n} (-1)^{n-\ell(\lambda)} |\{F \subseteq E :\ \mathrm{type}(F) = \lambda\}| (-1)^{n-b-1} \sum_{\mu \vdash a} \binom{a-\ell(\mu)}{c} \binom{\lambda}{\mu} \binom{n-\ell(\lambda)+\ell(\mu)-a}{n-b-c-1} \\
&= \sum_{F \subseteq E} (-1)^{n-\ell(\mathrm{type}(F))} (-1)^{n-b-1} \sum_{\mu \vdash a} \binom{a-\ell(\mu)}{c} \binom{\mathrm{type}(F)}{\mu} \binom{n-\ell(\mathrm{type}(F))+\ell(\mu)-a}{n-b-c-1} \\
&= \sum_{F \subseteq E} (-1)^{|F|+n-b-1} \sum_{\mu \vdash a} \binom{a-\ell(\mu)}{c} \binom{\mathrm{type}(F)}{\mu} \binom{|F|+\ell(\mu)-a}{n-b-c-1} \\
&= \sum_{F \subseteq E} (-1)^{|F|+n-b-1} \sum_{\mu \vdash a} \sum_{\substack{A \in \binom{V}{a} \\ F\text{-pure of type } \mu}} \binom{a-\ell(\mu)}{c} \binom{|F|+\ell(\mu)-a}{n-b-c-1} \\
&= \sum_{F \subseteq E} (-1)^{|F|+n-b-1} \sum_{\substack{A \in \binom{V}{a} \\ F\text{-pure}}} \binom{a-\ell(\mathrm{type}(F(A)))}{c} \binom{|F|+\ell(\mathrm{type}(F(A)))-a}{n-b-c-1}
\end{aligned}$$

where $F(A)$ is regarded as an edge set on A, so that $\mathrm{type}(F(A)) \vdash |A| = a$. Now recall that A is F-pure if and only if $F \subseteq E(A) \cup E(\bar{A})$. So we may switch the order of summation to obtain

$$R = \sum_{A \in \binom{V}{a}} \sum_{F \subseteq E(A) \cup E(\bar{A})} (-1)^{|F|+n-b-1} \binom{a-\ell(\mathrm{type}(F(A)))}{c} \binom{|F|+\ell(\mathrm{type}(F(A)))-a}{n-b-c-1}.$$

Every set $F \subseteq E(A) \cup E(\bar{A})$ can be written uniquely as $F = F(A) \cup F(\bar{A})$, with $F(A) \subseteq E(A)$ and $F(\bar{A}) \subseteq E(\bar{A})$. Moreover, $a - \ell(\text{type}(F(A))) = |F(A)|$. So now we get

$$R = \sum_{A \in \binom{V}{a}} \sum_{F(A) \subseteq E(A)} \sum_{F(\bar{A}) \subseteq E(\bar{A})} (-1)^{|F(A)|+|F(\bar{A})|+n-b-1} \binom{|F(A)|}{c} \binom{|F(\bar{A})|}{n-b-c-1}$$
$$= \sum_{A \in \binom{V}{a}} \left(\sum_{F(A) \subseteq E(A)} (-1)^{|F(A)|+c} \binom{|F(A)|}{c} \right) \left(\sum_{F(\bar{A}) \subseteq E(\bar{A})} (-1)^{|F(\bar{A})|+n-b-c-1} \binom{|F(\bar{A})|}{n-b-c-1} \right).$$

Now, applying Lemma 1 twice, we get

$$R = \left| \{ A \subseteq [n] : \ |A| = a, \ e(A) = c, \ e(\bar{A}) = n-b-c-1 \} \right|$$

and the theorem follows since $d(A) = n - 1 - e(A) - e(\bar{A})$. ☐

Corollary 3. *The number of vertices of degree b in T is*

$$|\{A \subseteq [n] : \ |A| = 1, \ d(A) = b, \ e(A) = 0\}| = \sum_{\lambda \vdash n} c_\lambda(T)(-1)^{n-b-1} \sum_{\mu \vdash 1} \binom{1-\ell(\mu)}{0} \binom{\lambda}{\mu} \binom{n-\ell(\lambda)+\ell(\mu)-1}{n-b-0-1}$$
$$= \sum_{\lambda \vdash n} c_\lambda(T)(-1)^{n-b-1} m_1(\lambda) \binom{n-\ell(\lambda)}{n-b-1}.$$

By contrast, the results of [MMW08] imply that for $b \geq 2$, the number of vertices of degree b is also given by the formula

$$\sum_{k \geq b} \binom{k}{b} (-1)^{k+b} \sigma_k = \sum_{k \geq b} \binom{k}{b} (-1)^{k+b} [q^k r^k] \mathbf{S}_T(q,r)$$
$$= \sum_{k \geq b} \binom{k}{b} (-1)^{k+b} \sum_{\lambda \vdash n} c_\lambda(T) \binom{\ell(\lambda)-1}{\ell(\lambda)-n+k} \sum_{d=1}^{k} (-1)^d \sum_{j=1}^{\ell(\lambda)} \binom{\lambda_j - 1}{d}.$$

We take this opportunity to point out a minor error in [MMW08]: the last line of the proof of Corollary 5 therein should read "for every $k \geq 2$", not "for every $k \geq 1$". The mistake does not affect the proof since the number of leaves of T can easily be recovered from $\mathbf{X}_T$, for instance as $|c_{(n-1,1)}|$.

4. The half-generalized degree polynomial and the subtree polynomial are equivalent

In this section we prove that the coefficients of the half-generalized degree polynomial of a tree, and those of its subtree polynomial, determine each other linearly.

Throughout this section, let $T = (V, E)$ be a tree with $|V| = n$. Let $h(a,b) = h_T(a,b)$ and $s(i,j) = s_T(i,j)$ be the coefficients of $\mathbf{H}_T$ and $\mathbf{S}_T$ defined in (5) and (7) respectively. Define column vectors

$$\mathsf{H}_1 = [h(0,b)]_{b=0}^{n-1}, \qquad \mathsf{S}_1 = [s(k,k)]_{k=0}^{n-1},$$
$$\mathsf{H}_2 = [h(a,b)]_{1 \leq a,b;\ a+b \leq n-1}, \qquad \mathsf{S}_2 = [s(i,j)]_{2 \leq j \leq i \leq n-1}.$$

Observe that $\mathbf{G}_T$ is determined by the entries of H_1 and H_2 together, because $h(a,b) = 0$ for all other (a,b), except $h(n-1,0) = 1$. Similarly, $\mathbf{S}_T$ is determined by the numbers $s(0,0) = n$, $s(1,1) = n-1$, and $s(i,j)$ for $2 \leq j \leq i \leq n-1$.

Theorem 4. *The half-generalized degree polynomial and the subtree polynomial determine each other linearly. Specifically, there exist nonsingular integer matrices $\mathsf{P}, \mathsf{M}, \mathsf{N}$ such that $\mathsf{P}\mathsf{H}_1 = \mathsf{S}_1$ and $\mathsf{M}\mathsf{H}_2 = \mathsf{N}\mathsf{S}_2$.*

Proof. For a subtree $S \in \mathcal{S}(T)$ and a set $K \subseteq D(S)$, let $U = S \cup K$. Then U is a tree and every edge in K is a leaf edge of U. Since S can be recovered from the pair (U, K), we have a bijection

$$\{(S,K) : S \in \mathcal{S}(T), \ K \subseteq D(S)\} \xrightarrow{\xi} \{(U,K) : U \in \mathcal{S}(T), \ K \subseteq L(U)\}$$
$$(S,K) \mapsto (S \cup K, K)$$

with $\xi^{-1}(U,K) = (U \setminus K, K)$. For nonnegative integers a,k with $a+k \leq n-1$, the map ξ restricts to a bijection

$\{(S,K)\colon S \in \mathcal{S}(T),\ K \subseteq D(S),\ |S|=a,\ |K|=k\} \to \{(U,K)\colon U \in \mathcal{S}(T),\ K \subseteq L(U),\ |U|=a+k,\ |K|=k\}.$

Fixing a and k and summing over the possibilities for $b=d(S)$ and $j=\ell(U)$, we obtain equalities

$$\sum_{b=k}^{n-1-a} \binom{b}{k} h(a,b) = \sum_{j=k}^{n-1} \binom{j}{k} s(a+k,j) \tag{9}$$

for every a,k.

Claim 1: The vectors $\mathsf{H}_1 = [h(0,b)]_{b=0}^{n-1}$ and $\mathsf{S}_1 = [s(k,k)]_{k=0}^{n-1}$ determine each other.

Indeed, consider the n equations (9) when $a=0$ and $0 \leq k \leq n-1$: they are

$$\sum_{b=k}^{n-1} \binom{b}{k} h(0,b) = \sum_{j=k}^{n-1} \binom{j}{k} s(k,j) = s(k,k)$$

(because if $j>k$ then $s(k,j)=0$). In matrix form, this system of equations can be written as $\mathsf{P}\mathsf{H}_1 = \mathsf{S}_1$, where the matrix

$$\mathsf{P} = \left[\binom{j}{i}\right]_{i,j=0}^{n-1}$$

is evidently (uni)triangular, proving Claim 1.

Claim 2: The vectors $\mathsf{H}_2 = [h(a,b)]_{1\leq a,b;\ a+b\leq n-1}$ and $\mathsf{S}_2 = [s(i,j)]_{2\leq j\leq i\leq n-1}$ determine each other.

This time, consider the $\binom{n-1}{2}$ equations (9) when $a,k>0$ and $a+k \leq n-1$. Then the data sets in the claim are exactly the variables appearing in the equations. (Note that $s(a+k,1)=0$, because $a+k\geq 2$ and every tree with at least two edges has at least two leaf edges.) Therefore, the equations we are considering can be written in matrix form as $\mathsf{M}\mathsf{H}_2 = \mathsf{N}\mathsf{S}_2$, where M and N are $(n-1)\times(n-1)$ square matrices whose rows are indexed by the pairs (a,k).

If we list the rows of M in increasing order by a, then in increasing order by k, then M has the block diagonal form $\mathsf{M}_1 \oplus \cdots \oplus \mathsf{M}_{n-1}$, where

$$\mathsf{M}_a = \left[\binom{b}{k}\right]_{1\leq b,k\leq n-a}.$$

In particular, M_a is unitriangular, hence nonsingular, and so M is nonsingular as well (in fact invertible over $\mathbb{Z}$).

On the other hand, if we list the rows of N in increasing order by $a+k$, then in increasing order by a, then N has the block diagonal form $\mathsf{N}_2 \oplus \cdots \oplus \mathsf{N}_n$, where

$$\mathsf{N}_i = \left[\binom{j}{k}\right]_{2\leq j\leq i,\ 1\leq k\leq i-1}.$$

In particular, $\det \mathsf{N}_i$ is a binomial determinant in the sense of Gessel and Viennot [GV85], namely $\binom{2,3,\ldots,i}{1,2,\ldots,i-1}$ in the notation of that paper, and in particular [GV85, Corollary 2] guarantees that $\det(\mathsf{N}_i)>0$. In fact, a short calculation using [GV85, Lemmas 8 and 9] shows that $\det \mathsf{N}_i = i$ for each i, so that $\det \mathsf{N} = n!$. Thus we have shown that the coefficients of $\mathbf{H}_T$ and $\mathbf{S}_T$ determine each other linearly. $\square$

To illustrate the proof, for $n=6$, the matrix equation $\mathsf{M}\mathsf{H}_2 = \mathsf{N}\mathsf{S}_2$ can be written either as

$$\left[\begin{array}{ccc|cc|c} 1 & 2 & 3 & 0 & 0 & 0 \\ 0 & 1 & 3 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 1 \end{array}\right] \begin{bmatrix} h(1,1) \\ h(1,2) \\ h(2,1) \\ h(1,3) \\ h(2,2) \\ h(3,1) \end{bmatrix} = \begin{bmatrix} 2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 3 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 4 \\ 0 & 2 & 3 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 3 & 6 \\ 0 & 0 & 0 & 2 & 3 & 4 \end{bmatrix} \begin{bmatrix} s(2,2) \\ s(3,2) \\ s(3,3) \\ s(4,2) \\ s(4,3) \\ s(4,4) \end{bmatrix}$$

illustrating the block diagonal form of M, or as

$$\begin{bmatrix} 1 & 2 & 3 & 0 & 0 & 0 \\ 0 & 1 & 3 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 2 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} h(1,1) \\ h(1,2) \\ h(2,1) \\ h(1,3) \\ h(2,2) \\ h(3,1) \end{bmatrix} = \left[\begin{array}{c|cc|ccc} 2 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 1 & 3 & 0 & 0 & 0 \\ 0 & 2 & 3 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 1 & 4 \\ 0 & 0 & 0 & 1 & 3 & 6 \\ 0 & 0 & 0 & 2 & 3 & 4 \end{array}\right] \begin{bmatrix} s(2,2) \\ s(3,2) \\ s(3,3) \\ s(4,2) \\ s(4,3) \\ s(4,4) \end{bmatrix}$$

illustrating the block diagonal form of N.

As another remark, the explicit linear transformations mapping $\mathbf{H}_T$ and $\mathbf{S}_T$ to each other are given by the matrices $\mathsf{M}^{-1}\mathsf{N}$ and $\mathsf{N}^{-1}\mathsf{M}$. We do not expect the entries of either of these matrices to have combinatorially meaningful formulas, particularly for the matrix $\mathsf{N}^{-1}\mathsf{M}$, whose entries are not integers.

5. A RECURRENCE FOR THE HALF-GENERALIZED DEGREE POLYNOMIAL

Let T be a tree with $n \geq 3$ vertices, and let $\mathcal{S}'(T)$ denote the family of subtrees of T that contain at least one non-leaf vertex of T. It is convenient to work with a slight modification of the half-generalized degree polynomial, defined by

$$\bar{\mathbf{H}}_T = \sum_{S \in \mathcal{S}'(T)} y^{d(S)} z^{e(S)} = \mathbf{H}_T - (n - \ell(T))y. \tag{10}$$

The polynomial $\bar{\mathbf{H}}_T$ contains the same information as $\mathbf{H}_T$, but is more convenient to work with for our present purposes.

Let $e = vw$ be a non-leaf edge of T. The **near-contraction** of e [APdMOZ23, §3] is the tree $T \odot e$ with edge set

$$E(T \odot e) = E(T) \setminus \{wx \colon x \notin \{v,w\}\} \cup \{vx \colon x \notin \{v,w\}\}. \tag{11}$$

Equivalently, contract the edge e, retaining the name v for the resulting vertex, and introduce a new edge $e' = vw$, so that w is a leaf. See Figure 3 for an example.

FIGURE 3. Near-contraction of a non-leaf edge e in a tree T.

Proposition 5. *Let T_1, T_2 be trees, let v, w be vertices of T_1 and T_2, respectively, and let e be the edge vw. Let $T_1' = T_1 \cup \{e\}$, $T_2' = T_2 \cup \{e\}$, and $T = T_1 \cup T_2 \cup \{e\}$. Then*

$$\bar{\mathbf{H}}_T = \frac{y}{y+z}\left(\bar{\mathbf{H}}_{T_1'} + \bar{\mathbf{H}}_{T_2'}\right) + \frac{z}{y+z}\bar{\mathbf{H}}_{T \odot e}.$$

Before giving the proof we need to set up notation. Given any vertex v in T, let

$$\mathcal{S}'_{+v}(T) = \{S \in \mathcal{S}'(T) \colon v \in S\}, \qquad \mathcal{S}'_{-v}(T) = \{S \in \mathcal{S}'(T) \colon v \notin S\}.$$

In addition, if v, w are two distinct vertices in T, then define

$$\mathcal{S}'_{\pm v, \pm w}(T) = \mathcal{S}'_{\pm v}(T) \cap \mathcal{S}'_{\pm w}(T).$$

Using this notation we also define

$$\bar{\mathbf{H}}_T^{\pm v} = \sum_{S \in \mathcal{S}'_{\pm v}(T)} y^{d(S)} z^{e(S)}, \qquad \bar{\mathbf{H}}_T^{\pm v, \pm w} = \sum_{S \in \mathcal{S}'_{\pm v, \pm w}(T)} y^{d(S)} z^{e(S)}.$$

Proof of Proposition 5. We start with the identity

$$\bar{\mathbf{H}}_T = \bar{\mathbf{H}}_T^{+v,+w} + \bar{\mathbf{H}}_T^{+v,-w} + \bar{\mathbf{H}}_T^{-v,+w} + \bar{\mathbf{H}}_T^{-v,-w}, \tag{12}$$

which is immediate from the definitions.

First, there are bijections $\mathcal{S}'_{+v,+w}(T) \to \mathcal{S}'_{+v,+w}(T \odot e)$ and $\mathcal{S}'_{+v,+w}(T \odot e) \to \mathcal{S}'_{+v,-w}(T \odot e)$, since w is a leaf in $T \odot e$. Thus,

$$\bar{\mathbf{H}}_T^{+v,+w} = \bar{\mathbf{H}}_{T\odot e}^{+v,+w} = \frac{z}{y}\,\bar{\mathbf{H}}_{T\odot e}^{+v,-w}.$$

So

$$\bar{\mathbf{H}}_{T\odot e}^{+v} = \bar{\mathbf{H}}_{T\odot e}^{+v,+w} + \bar{\mathbf{H}}_{T\odot e}^{+v,-w} = \left(1 + \frac{y}{z}\right)\bar{\mathbf{H}}_{T\odot e}^{+v,+w}.$$

Combining these two equations yields

$$\bar{\mathbf{H}}_T^{+v,+w} = \frac{z}{y+z}\bar{\mathbf{H}}_{T\odot e}^{+v}. \tag{13}$$

Second, there is a bijection from $\mathcal{S}'_{-v,-w}(T)$ to $\mathcal{S}'_{-v}(T \odot e)$ since a subtree of $T \odot e$ that does not contain v cannot contain the leaf w, which is attached to v, so that

$$\begin{aligned}\bar{\mathbf{H}}_T^{-v,-w} &= \frac{y}{y+z}\bar{\mathbf{H}}_T^{-v,-w} + \frac{z}{y+z}\bar{\mathbf{H}}_T^{-v,-w} \\ &= \frac{y}{y+z}\bar{\mathbf{H}}_T^{-v,-w} + \frac{z}{y+z}\bar{\mathbf{H}}_{T\odot e}^{-v}\end{aligned} \tag{14}$$

where the first step is algebra and the second step uses the bijection.

Combining (12), (13) and (14) yields

$$\begin{aligned}\bar{\mathbf{H}}_T &= \frac{z}{y+z}\bar{\mathbf{H}}_{T\odot e}^{+v} + \bar{\mathbf{H}}_T^{+v,-w} + \bar{\mathbf{H}}_T^{-v,+w} + \frac{y}{y+z}\bar{\mathbf{H}}_T^{-v,-w} + \frac{z}{y+z}\bar{\mathbf{H}}_{T\odot e}^{-v} \\ &= \bar{\mathbf{H}}_T^{+v,-w} + \bar{\mathbf{H}}_T^{-v,+w} + \frac{y}{y+z}\bar{\mathbf{H}}_T^{-v,-w} + \frac{z}{y+z}\bar{\mathbf{H}}_{T\odot e}.\end{aligned} \tag{15}$$

It remains to show that

$$\bar{\mathbf{H}}_T^{+v,-w} + \bar{\mathbf{H}}_T^{-v,+w} + \frac{y}{y+z}\bar{\mathbf{H}}_T^{-v,-w} = \frac{y}{y+z}\left(\bar{\mathbf{H}}_{T_1'} + \bar{\mathbf{H}}_{T_2'}\right) \tag{16}$$

which when combined with (15) will give the desired result.

First, there is a bijection $\phi : \mathcal{S}'_{+v,-w}(T) \to \mathcal{S}'_{+v}(T_1)$; observe that $\phi(S)$ has one fewer external edge than S. Thus,

$$\bar{\mathbf{H}}_T^{+v,-w} = y\bar{\mathbf{H}}_{T_1}^{+v} = \frac{y}{y+z}\bar{\mathbf{H}}_{T_1'}^{+v},$$

and similarly,

$$\bar{\mathbf{H}}_T^{-v,+w} = y\bar{\mathbf{H}}_{T_2}^{+w} = \frac{y}{y+z}\bar{\mathbf{H}}_{T_2'}^{+w}.$$

Finally, each subtree of T containing neither v nor w is either a subtree of T_1' that does not contain v, or a subtree of T_2' that does not contain w. Hence,

$$\bar{\mathbf{H}}_T^{-v,-w} = \bar{\mathbf{H}}_{T_1'}^{-v} + \bar{\mathbf{H}}_{T_2'}^{-w}.$$

Combining the last three equations yields (16), completing the proof. ☐

Proposition 5 can be used to show that two non-isomorphic trees have the same (modified) HDP. See Figure 4 for an example. We will exploit this idea further in the next section.

As a remark, we have not been able to obtain recurrences for $\mathbf{G}_T$ similar to those for $\mathbf{H}_T$.

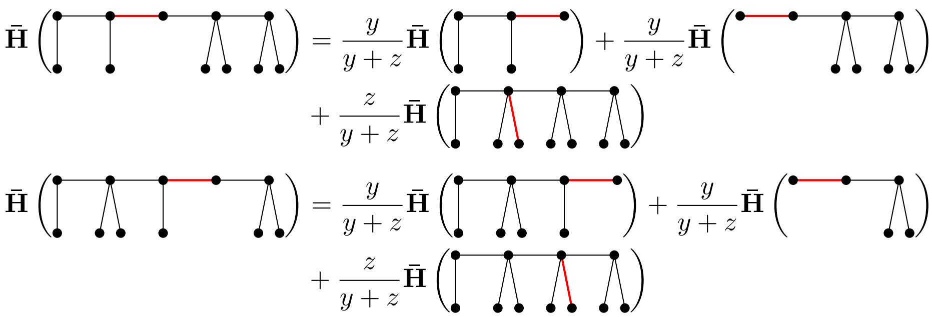


FIGURE 4. Application of Proposition 5. Note that the right-hand sides of both equations are equal, proving that the two trees on the left-hand sides have the same HDP.

6. FAMILIES OF TREES WITH THE SAME HALF-GENERALIZED DEGREE POLYNOMIAL

In this section we will construct arbitrarily large families of non-isomorphic trees with the same half-generalized degree polynomial by exploiting the recurrence given by Proposition 5.

Recall that a **composition** of an integer n is an ordered list of positive integers $\alpha = (a_1, \dots, a_k)$ that add up to n. The **length** of α is $\ell(\alpha) = k$. The **reverse** of α is $\alpha^* := (\alpha_k, \alpha_{k-1}, \dots, a_1)$. If $\alpha = (a_1, \dots, a_k)$ and $\beta = (b_1, \dots, b_m)$ with $|\alpha| = |\beta|$, we say that β is a **coarsening** of α, written $\beta \geq \alpha$, if every partial sum of β (i.e., every number $b_1 + \cdots + b_j$ for some j) is also a partial sum of α. Coarsening is a partial order on compositions of n. The **concatenation** of α and β is

$$\alpha \cdot \beta = (a_1, \dots, a_\ell, b_1, \dots, b_m),$$

and the **near-concatenation** is

$$\alpha \odot \beta = (a_1, \dots, a_{\ell-1}, a_\ell + b_1, b_2, \dots, b_m).$$

A **caterpillar** is a tree with the property that deleting all its leaves produces a path $\mathrm{sp}(T)$, called the **spine** of the caterpillar. We write $\mathrm{Cat}(a_1, \dots, a_k)$ for the caterpillar with k spine vertices $v_1, \dots, v_k$, adjacent to $a_1 - 1, \dots, a_k - 1$ leaves respectively, so that the total number of vertices is $n = a_1 + \cdots + a_k$. (See Figure 5 for an example.) The composition $\alpha = (a_1, \dots, a_k)$ is called the **signature** of the caterpillar. The signature is well-defined, and a complete invariant, up to reversal. Moreover, $a_1, a_k \geq 2$ since each end vertex of the spine is adjacent to at least one leaf; the other numbers a_i are unconstrained. Let $\mathcal{C}$ denote the set of compositions where the first and last parts are larger than 1, i.e., compositions that are signatures of caterpillars.

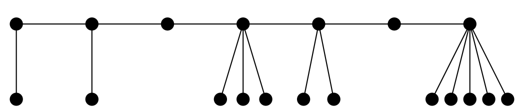


FIGURE 5. The caterpillar $\mathrm{Cat}(2, 2, 1, 4, 3, 1, 6)$.

Observe that deleting the edge $e = v_\ell v_{\ell+1}$ from $\mathrm{Cat}(\beta \cdot \gamma)$ produces the forest with components $\mathrm{Cat}(\beta)$ and $\mathrm{Cat}(\gamma)$. Moreover, $\mathrm{Cat}(\beta \odot \gamma)$ is just the near-contraction $\mathrm{Cat}(\beta \cdot \gamma) \odot e$ (see (11)).

For convenience, given a composition α, define

$$\eta(\alpha) = \bar{\mathbf{H}}_{\mathrm{Cat}(1 \odot \alpha \odot 1)}.$$

In this notation, Proposition 5 becomes

$$\eta(\alpha \cdot \beta) = \frac{y}{y+z}(\eta(\alpha) + \eta(\beta)) + \frac{z}{y+z}\eta(\alpha \odot \beta). \tag{17}$$

Define

$$\zeta(\alpha)=\zeta(\alpha)(y,z,x_1,x_2,\dots)=\sum_{\gamma\geq\alpha}\frac{y^{\ell(\gamma)-1}z^{\ell(\alpha)-\ell(\gamma)}}{(y+z)^{\ell(\alpha)-1}}\sum_{i=1}^{\ell(\gamma)}x_{\gamma_i}. \tag{18}$$

Our next step is to show that the power series $\zeta(\alpha)$ satisfies the same recurrence as the $\eta(\alpha)$, which will enable us to obtain a closed form for the HDP of a caterpillar using the right-hand side of (18).

Proposition 6. *For all compositions α and β,*

$$\zeta(\alpha\cdot\beta)=\frac{y}{y+z}\left(\zeta(\alpha)+\zeta(\beta)\right)+\frac{z}{y+z}\zeta(\alpha\odot\beta). \tag{19}$$

Proof. In each coarsening $\gamma\geq\alpha\cdot\beta$, the last part of α and the first part of β are either merged or kept separate. In the first case, γ coarsens $\alpha\odot\beta$. In the second, γ is the concatenation of a coarsening of α with a coarsening of β. Therefore, we may split up the expression for $\zeta(\alpha\cdot\beta)$ given by (18) as

$$\zeta(\alpha\cdot\beta)=\underbrace{\sum_{\gamma\geq\alpha\odot\beta}\frac{y^{\ell(\gamma)-1}z^{\ell(\alpha\cdot\beta)-\ell(\gamma)}}{(y+z)^{\ell(\alpha\cdot\beta)-2}}\sum_{i=1}^{\ell(\gamma)}x_{\gamma_i}}_{A}+\underbrace{\sum_{\gamma'\geq\alpha}\sum_{\gamma''\geq\beta}\frac{y^{\ell(\gamma'\cdot\gamma'')-1}z^{\ell(\alpha\cdot\beta)-\ell(\gamma'\cdot\gamma'')}}{(y+z)^{\ell(\alpha\cdot\beta)-1}}\sum_{i=1}^{\ell(\gamma'\cdot\gamma'')}x_{(\gamma'\cdot\gamma'')_i}}_{B}. \tag{20}$$

First, observe that

$$A=\frac{z}{y+z}\sum_{\gamma\geq\alpha\odot\beta}\frac{y^{\ell(\gamma)-1}z^{\ell(\alpha\odot\beta)-\ell(\gamma)}}{(y+z)^{\ell(\alpha\odot\beta)-1}}\sum_{i=1}^{\ell(\gamma)}x_{\gamma_i}=\frac{z}{y+z}\zeta(\alpha\odot\beta). \tag{21}$$

Second,

$$\begin{aligned}B&=\frac{y}{y+z}\sum_{\gamma'\geq\alpha}\sum_{\gamma''\geq\beta}\frac{y^{\ell(\gamma'\cdot\gamma'')-2}z^{\ell(\alpha\cdot\beta)-\ell(\gamma'\cdot\gamma'')}}{(y+z)^{\ell(\alpha\cdot\beta)-2}}\sum_{i=1}^{\ell(\gamma'\cdot\gamma'')}x_{(\gamma'\cdot\gamma'')_i}\\&=\frac{y}{y+z}\sum_{\gamma'\geq\alpha}\sum_{\gamma''\geq\beta}\frac{y^{\ell(\gamma')-1}z^{\ell(\alpha)-\ell(\gamma')}}{(y+z)^{\ell(\alpha)-1}}\frac{y^{\ell(\gamma'')-1}z^{\ell(\beta)-\ell(\gamma'')}}{(y+z)^{\ell(\beta)-1}}\left(\sum_{i=1}^{\ell(\gamma')}x_{\gamma'_i}+\sum_{i=1}^{\ell(\gamma'')}x_{\gamma''_i}\right)\\&=\frac{y}{y+z}\left(\left(\sum_{\gamma''\geq\beta}\frac{y^{\ell(\gamma'')-1}z^{\ell(\beta)-\ell(\gamma'')}}{(y+z)^{\ell(\beta)-1}}\right)\zeta(\alpha)+\left(\sum_{\gamma'\geq\alpha}\frac{y^{\ell(\gamma')-1}z^{\ell(\alpha)-\ell(\gamma')}}{(y+z)^{\ell(\alpha)-1}}\right)\zeta(\beta)\right).\end{aligned} \tag{22}$$

On the other hand,

$$\sum_{\gamma'\geq\alpha}y^{\ell(\gamma')-1}z^{\ell(\alpha)-\ell(\gamma')}=\sum_{k=1}^{\ell(\alpha)}\binom{\ell(\alpha)-1}{k-1}y^{k-1}z^{\ell(\alpha)-k}=(y+z)^{\ell(\alpha)-1}, \tag{23}$$

so the parenthesized sums in (22) may be dropped, and then substituting (22) and (21) into (20) yields the desired equality. □

Proposition 7. *For every composition α, we have*

$$\begin{aligned}\bar{\mathbf{H}}_{\mathrm{Cat}(1\odot\alpha\odot1)}&=\zeta(\alpha)(y,z,x_1,x_2,\dots)|_{x_1=(y+z)^2,\ x_2=(y+z)^3,\ \dots,\ x_i=(y+z)^{i+1},\ \dots}\\&=\sum_{\gamma\geq\alpha}\frac{y^{\ell(\gamma)-1}z^{\ell(\alpha)-\ell(\gamma)}}{(y+z)^{\ell(\alpha)-1}}\sum_{i=1}^{\ell(\gamma)}(y+z)^{\gamma_i+1}.\end{aligned} \tag{24}$$

Proof. We induct on the length of α. For the base case, say $\alpha=(a)$. Then $1\odot\alpha\odot1=(a+2)$; that is, $\mathrm{Cat}(a+2)$ is the star with $a+1$ edges. So

$$\eta(\alpha)=\bar{\mathbf{H}}_{\mathrm{Cat}(a+2)}=\frac{1}{(y+z)^{-1}}(y+z)^a=(y+z)^{a+1},$$

which is the right-hand side of (24). Meanwhile, the inductive step follows directly from (17), (19) and the induction hypothesis. □

The preceding machinery can be used to construct caterpillars with the same HDP. The procedure follows the construction of compositions with the same ribbon Schur function [BTvW06] or $\mathcal{L}$-polynomial [APZ14]. Given two compositions $\alpha = (a_1, \ldots, a_\ell)$ and β, define

$$\alpha \circ \beta = \beta^{\odot a_1} \cdot \beta^{\odot a_2} \cdots \beta^{\odot a_\ell}$$

where $\beta^{\odot i} = \beta \odot \cdots \odot \beta$ (i times). This operation satisfies the identities

$$(\alpha \cdot \gamma) \circ \beta = (\alpha \circ \beta) \cdot (\gamma \circ \beta), \tag{25}$$

$$(\alpha \odot \gamma) \circ \beta = (\alpha \circ \beta) \odot (\gamma \circ \beta), \tag{26}$$

$$(\alpha \circ \beta)^* = \alpha^* \circ \beta^*. \tag{27}$$

Proposition 8. *For all compositions α and β, we have*

$$\zeta(\alpha \circ \beta) = \zeta(\alpha)\big|_{x_i = \zeta(\beta^{\odot i})}.$$

Proof. We induct on the length of α. For the base case, suppose $\alpha = (a)$, so by definition $\zeta(\alpha \circ \beta) = \zeta(\beta^{\odot a})$ and $\zeta(a) = x_a$. Thus $\zeta(a)|_{x_i = \zeta(\beta^{\odot i})} = \zeta(\beta^{\odot a})$ as desired. For the inductive step, we calculate

$$\begin{aligned}
\zeta(\alpha \circ \beta) &= \zeta((\alpha_{1,\ell-1} \circ \beta) \cdot (\alpha_\ell \circ \beta)) \\
&= \frac{y}{y+z}\left(\zeta(\alpha_{1,\ell-1} \circ \beta) + \zeta(\alpha_\ell \circ \beta)\right) + \frac{z}{y+z}\zeta((\alpha_{1,\ell-1} \odot \alpha_\ell) \circ \beta) \\
&= \frac{y}{y+z}\left(\zeta(\alpha_{1,\ell-1}) + \zeta(\alpha_\ell)\right)|_{x_i = \zeta(\beta^{\odot i})} + \frac{z}{y+z}\zeta(\alpha_{1,\ell-1} \odot \alpha_\ell)|_{x_i = \zeta(\beta^{\odot i})} \\
&= \zeta(\alpha)|_{x_i = \zeta(\beta^{\odot i})}.
\end{aligned}$$

□

A **factorization** of a composition α is an equality $\alpha = \alpha_1 \circ \alpha_2 \circ \cdots \circ \alpha_k$. The factorization is *nontrivial* if (i) no α_i is the composition 1; (ii) no two consecutive factors both have length 1; and (iii) no two consecutive factors both have all parts equal to 1. The factorization is **irreducible** if it is nontrivial and no α_i admits a nontrivial factorization. In fact every composition admits a unique irreducible factorization [BTvW06, Theorem 3.6].

Let α be a composition with irreducible factorization $\alpha_1 \circ \cdots \circ \alpha_k$. A **switch** of α is a composition β of the form $\beta_1 \circ \cdots \circ \beta_k$, where $\beta_i \in \{\alpha_i, \alpha_i^*\}$ for each i. Unique factorization implies that switching is an equivalence relation on compositions.

Theorem 9. *(1) Suppose that α and β are related by switching. Then $\zeta(\alpha) = \zeta(\beta)$, and consequently*

$$\mathbf{H}_{\mathrm{Cat}(1 \odot \alpha \odot 1)} = \mathbf{H}_{\mathrm{Cat}(1 \odot \beta \odot 1)}.$$

(2) In particular, if α has ℓ irreducible factors that are not palindromes, then the ζ-equivalence class of α contains at least $2^{\ell-1}$ non-equivalent compositions. In this case, $\mathrm{Cat}(1 \odot \alpha \odot 1)$ is one of at least $2^{\ell-1}$ pairwise non-isomorphic caterpillars with the same half-generalized degree polynomial (equivalently, by Theorem 4, with the same subtree polynomial).

In particular, there exist arbitrarily large sets of non-isomorphic caterpillars with the same HDP and STP.

Proof. First, observe that $\zeta(\alpha) = \zeta(\alpha^*)$, because $\gamma \geq \alpha$ if and only if $\gamma^* \geq \alpha^*$, and γ and γ^* make the same contribution to the right-hand side of (18). Second, observe that

$$\zeta(\alpha^* \circ \beta^*) = \zeta(\alpha \circ \beta) = \zeta(\alpha^* \circ \beta) = \zeta(\alpha \circ \beta^*)$$

The first and third equalities follows from (27), while the second follows from the first identity of Proposition 8. The desired corollary now follows by induction. □

Example 10. Consider the irreducible factorizations

$$\alpha = (1,2) \circ (1,2) = (1,2)^{\odot 1} \cdot (1,2)^{\odot 2} = (1,2) \cdot (1,3,2) = (1,2,1,3,2),$$

$$\beta = (2,1) \circ (1,2) = (1,2)^{\odot 2} \cdot (1,2)^{\odot 1} = (1,3,2) \cdot (1,2) = (1,3,2,1,2).$$

Then $\zeta(\alpha) = \zeta(\beta)$ by Theorem 9, and the corresponding caterpillars

$$\mathrm{Cat}(1 \odot \alpha \odot 1) = \mathrm{Cat}(2,2,1,3,3), \qquad \mathrm{Cat}(1 \odot \beta \odot 1) = \mathrm{Cat}(2,3,2,1,3)$$

have the same half-generalized and subtree polynomials. In fact, these are the two smallest such trees, shown above in Figure 2. ◀

The Eisenstat-Gordon conjecture. Let $p(x)$ be a polynomial whose coefficients are all 0's and 1's, with no two consecutive 0's; we may as well assume that the leading and trailing coefficients are 1. Call such a polynomial *gap-free*. Let $a < b$ be positive integers. Expand the polynomials $(a + bx)p(x)$ and $(b + ax)p(x)$, read off the coefficient lists, and add 1 to the first and last terms to obtain lists $L_{p,1}, L_{p,2}$, which we regard as the signatures of caterpillars $C_{p,1}, C_{p,2}$. For example, if $p(x) = 1 + x + x^3$ then

$$(a + bx)p(x) = a + (a + b)x + bx^2 + ax^3 + bx^4 \qquad L_{p,1} = (a + 1, a + b, b, a, b + 1) \qquad C_{p,1} = \mathrm{Cat}(L_{p,1}),$$
$$(b + ax)p(x) = b + (a + b)x + ax^2 + bx^3 + ax^4 \qquad L_{p,2} = (b + 1, a + b, a, b, a + 1) \qquad C_{p,2} = \mathrm{Cat}(L_{p,2}).$$

For example, when $(a, b) = (1, 2)$, this construction produces the caterpillars in Figure 2.

Eisenstat and Gordon conjectured that for any gap-free $p(x)$ and any positive integers a and b, the caterpillars $C_{p,1}, C_{p,2}$ have the same subtree polynomial [EG06, Conjecture 2.8]. In fact, this statement is a special case of Theorem 9, as follows. Write $0 = i_0 < i_1 < i_2 < \ldots < i_\ell = \deg p$, where $\{i_1, \ldots, i_{\ell-1}\}$ are all the indices of the coefficients of p equal to 0. Let $\beta = (i_1 - i_0, i_2 - i_1, \ldots, i_\ell - i_{\ell-1})$ and $\alpha = (a, b)$. Then

$$C_{p,1} = \mathrm{Cat}(1 \odot (\beta \circ \alpha) \odot 1) \quad \text{and} \quad C_{p,2} = \mathrm{Cat}(1 \odot (\beta \circ \alpha^*) \odot 1),$$

so indeed $C_{p,1}$ and $C_{p,2}$ have the same HDP and thus also the same STP.

7. Further remarks

7.1. **Beyond caterpillars.** The near-concatenation operation $\odot$, and thus the construction of Theorem 9 may be extended to trees that are not caterpillars. Say that a **polarized tree** is a tree T with two distinguished vertices, the **left end** and **right end**. Given two polarized trees T, T' with left ends L, L' and right ends R, R' respectively, the **concatenation** is the polarized tree $T \cdot T' = (T + T') \cup \{RL'\}$ (where $+$ denotes disjoint union), with left end u and right end R'. The **near-concatenation** is $T \odot T' = (T \cdot T') \odot RL'$, with the same left and right ends.

Let $\beta = (\beta_1, \ldots, \beta_\ell)$ be a composition and T a polarized tree. Define

$$\beta \circ T = T^{\odot\beta_1} \cdot T^{\odot\beta_2} \cdot \ldots \cdot T^{\odot\beta_\ell}$$

where

$$T^{\odot i} = \underbrace{T \odot T \odot \cdots \odot T}_{i \text{ times}}.$$

Examples of these constructions are given in Figure 6.

By extending the arguments of Section 6 from caterpillars to polarized trees, we can generalize Proposition 8 and Theorem 9 to arbitrary trees, as follows.

Theorem 11. *Let T be a tree, and α a composition. Then*

$$\bar{\mathbf{H}}_{1\odot(\alpha\circ T)\odot 1} = \zeta(\alpha)|_{x_i = \bar{\mathbf{H}}_{1\odot T^{\odot i}\odot 1}}.$$

Moreover, if α and β are related by switching, then

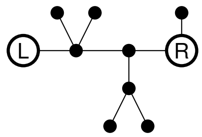
$$\mathbf{H}_{\mathrm{Cat}(1\odot(\alpha\circ T)\odot 1)} = \mathbf{H}_{\mathrm{Cat}(1\odot(\beta\circ T)\odot 1)}.$$

Not all pairs of trees with the same HDP arise from the construction of Theorem 11. The unique smallest example, obtained by computer experimentation, is shown in Figure 7.

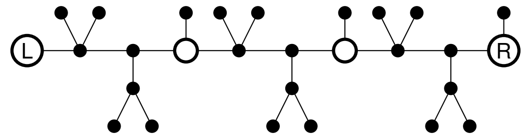
7.2. **Comparing the GDP and the HDP.** For trees T and T', we know that $\mathbf{G}_T = \mathbf{G}_{T'}$ implies $\mathbf{H}_T = \mathbf{H}_{T'}$. Does the reverse implication hold?

Intuitively, the three-variable polynomial $\mathbf{G}_T$ should contain more information than its two-variable specialization $\mathbf{H}_T$. Somewhat surprisingly, explicit computation indicates that the equivalence relations induced by $\mathbf{G}$ and $\mathbf{H}$ are identical for all $n \leq 18$. For $n \leq 10$, the equivalence classes are all singletons; that is, both the GDS and the HDP are complete invariants. For $11 \leq n \leq 18$, the non-singleton equivalence classes are all of size two, and are enumerated as follows:

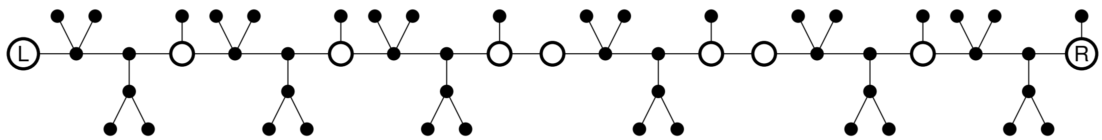
n	11	12	13	14	15	16	17	18
Number of size-two equivalence classes	1	1	1	5	1	7	17	15



T



$T^{\odot 3}$



$(3, 1, 2) \circ T$

FIGURE 6. Operations on polarized trees.

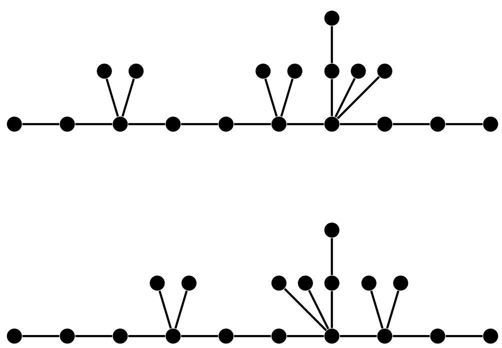


FIGURE 7. Trees with the same HDP that cannot be obtained as $\alpha \circ T$

7.3. **The $\mathcal{L}$-polynomial.** The $\mathcal{L}$-polynomial of a composition α is defined as

$$\mathcal{L}(\alpha) = \sum_{\gamma \geq \alpha} \prod_i x_{\gamma_i}.$$

It is a specialization of the U-polynomial and hence also a specialization of the chromatic symmetric function [APZ14]. It is clear that $\zeta(\alpha)$ can be computed from $\mathcal{L}(\alpha)$. In [BTvW06, APZ14], it was shown that two caterpillars have the same $\mathcal{L}$-polynomial if and only if their signatures are related by switching.

Question 1. If $\zeta(\alpha) = \zeta(\beta)$, are α and β necessarily related by switching?

We have verified by explicit computation that the answer to the latter question is affirmative for caterpillars up to 18 vertices.

7.4. **A closed formula for the gdp of a caterpillar.** The generalized degree polynomial of a caterpillar can be expressed compactly by grouping the sets $A \subseteq V(T)$ by $\mathrm{sp}(A) := A \cap \mathrm{sp}(T)$, since the contribution of each leaf to $\mathbf{G}_T$ depends only on whether its neighbor belongs to A. Thus we obtain

$$\begin{aligned}\mathbf{G}_{\mathrm{Cat}(\alpha)} &= \sum_{U\subseteq[k]} x^{|U|} y^{\tilde{d}(U)} z^{\tilde{e}(U)} (xz+y)^{\sum_{u\in U}(a_u-1)} (xy+1)^{\sum_{u\in[k]\setminus U}(a_u-1)} \\ &= (xy+1)^{n-k} \sum_{U\subseteq[k]} x^{|U|} y^{d(U)} z^{e(U)} \left(\frac{xz+y}{xy+1}\right)^{\sum_{u\in U}(a_u-1)}\end{aligned} \tag{28}$$

where $\tilde{d}(U)$ and $\tilde{e}(U)$ are the numbers of boundary edges and internal edges in the spine, i.e.,

$$\begin{aligned}\tilde{d}(U) &= |\{i \in [k-1] :\ i \in U \text{ xor } i+1 \in U\}|, \\ \tilde{e}(U) &= |\{i \in [k-1] :\ i \in U \text{ and } i+1 \in U\}|.\end{aligned}$$

There is a similar formula for the half-generalized degree polynomial of a caterpillar, since $A \subseteq V(T)$ induces a subtree if and only if either (i) $A = \{v\}$ for some $v \in L(T)$, or (ii) $\mathrm{sp}(A)$ is a nontrivial path and every non-spine vertex in A has a neighbor in $\mathrm{sp}(A)$. Thus we obtain

$$\mathbf{H}_{\mathrm{Cat}(\alpha)} = (n-k)y + \sum_{1\leq i\leq j\leq k} y^{i>1} y^{j<k} z^{j-i} (z+y)^{\sum_{u=i}^{j}(a_u-1)} \tag{29}$$

where, e.g., the symbol $y^{i>1}$ is interpreted as "y if $i > 1$, else 1." Formulas (28) and (29) are useful for explicit computation; for instance, the number of terms in (28) is exponential in the size of the spine rather than the size of the tree, a significant savings. On the other hand, we do not know how to obtain usable recurrences for these expressions.

7.5. **Combining the HDP and the STP.** We propose another tree invariant for further study. Define the **souped-up subtree polynomial** as the following common generalization of the HDP and STP:

$$\mathbf{Q}_T = \mathbf{Q}_T(x,\, y,\, z) = \sum_{S\in\mathcal{S}(T)} x^{e(S)} y^{d(S)} z^{\ell(S)}$$

so that $\mathbf{H}_T = \mathbf{Q}_T(x,\, y,\, 1)$ and $\mathbf{S}_T = \mathbf{Q}_T(x,\, 1,\, z)$.

Question 2. What can be said about the ability of $\mathbf{Q}_T$ to distinguish trees?

The souped-up subtree polynomial contains strictly more information than either the half-generalized degree polynomial or the subtree polynomial, because it distinguishes the two 11-vertex caterpillars in Figure 2. Indeed, T_1 has exactly one subgraph with three edges, two leaves, and one external edge (induced by the four leftmost vertices), but T_2 has none. Therefore, the coefficients of x^3yz^2 differ in $\mathbf{Q}_{T_1}$ and $\mathbf{Q}_{T_2}$. In particular, $\mathbf{Q}_T$ cannot be recovered from the generalized degree polynomial.

We have not found a pair of non-isomorphic trees with the same souped-up subtree polynomial, and we have checked by explicit computation that $\mathbf{Q}_T$ is a complete invariant for trees with 18 or fewer vertices. We have also checked computationally that neither $\mathbf{Q}_T$ nor $\mathbf{X}_T$ can be obtained from the other linearly for $n = 8$.

References

[ABS06] Marcelo Aguiar, Nantel Bergeron, and Frank Sottile, *Combinatorial Hopf algebras and generalized Dehn-Sommerville relations*, Compos. Math. **142** (2006), no. 1, 1–30. MR 2196760

[ADM22] Jean-Christophe Aval, Karimatou Djenabou, and Peter RW McNamara, *Quasisymmetric functions distinguishing trees*, preprint, arXiv:2201.11763, 2022.

[APCSZ21] José Aliste-Prieto, Logan Crew, Sophie Spirkl, and José Zamora, *A vertex-weighted Tutte symmetric function, and constructing graphs with equal chromatic symmetric function*, Electron. J. Combin. **28** (2021), no. 2, Paper No. 2.1, 33. MR 4245294

[APdMOZ23] José Aliste-Prieto, Anna de Mier, Rosa Orellana, and José Zamora, *Marked graphs and the chromatic symmetric function*, SIAM J. Discrete Math. **37** (2023), no. 3, 1881–1919. MR 4632385

[APdMZ17] José Aliste-Prieto, Anna de Mier, and José Zamora, *On trees with the same restricted u-polynomial and the prouhet–tarry–escott problem*, Discrete Math. **340** (2017), no. 6, 1435–1441.

[APZ14] José Aliste-Prieto and José Zamora, *Proper caterpillars are distinguished by their chromatic symmetric function*, Discrete Math. **315** (2014), 158–164. MR 3130368

[AWvW21] Farid Aliniaeifard, Victor Wang, and Stephanie van Willigenburg, *The chromatic symmetric function of a graph centred at a vertex*, arXiv:2108.04850, 2021.

[Bir13] George D. Birkhoff, *A determinant formula for the number of ways of coloring a map*, Ann. of Math. (2) **14** (1912/13), no. 1-4, 42–46. MR 1502436

[BJR09] Louis J. Billera, Ning Jia, and Victor Reiner, *A quasisymmetric function for matroids*, European J. Combin. **30** (2009), no. 8, 1727–1757. MR 2552658

[BTvW06] Louis J. Billera, Hugh Thomas, and Stephanie van Willigenburg, *Decomposable compositions, symmetric quasisymmetric functions and equality of ribbon Schur functions*, Adv. Math. **204** (2006), no. 1, 204–240. MR 2233132

[Cre20] Logan Crew, *Vertex-Weighted Generalizations of Chromatic Symmetric Functions*, Ph.D. thesis, University of Pennsylvania, 2020.

[Cre22] ______, *A note on distinguishing trees with the chromatic symmetric function*, Discrete Math. **345** (2022), no. 2, Paper No. 112682, 4 pp. MR 4327391

[EG06] David Eisenstat and Gary Gordon, *Non-isomorphic caterpillars with identical subtree data*, Discrete Math. **306** (2006), no. 8-9, 827–830. MR 2234989

[Ell17] Brittney Ellzey, *Chromatic quasisymmetric functions of directed graphs*, Sém. Lothar. Combin. **78B** (2017), Art. 74, 12. MR 3678656

[GKP94] Ronald L. Graham, Donald E. Knuth, and Oren Patashnik, *Concrete Mathematics*, second ed., Addison-Wesley Publishing Company, Reading, MA, 1994. MR 1397498

[GM89] Gary Gordon and Elizabeth McMahon, *A greedoid polynomial which distinguishes rooted arborescences*, Proc. Amer. Math. Soc. **107** (1989), no. 2, 287–298. MR 967486

[GS01] David D. Gebhard and Bruce E. Sagan, *A chromatic symmetric function in noncommuting variables*, J. Algebraic Combin. **13** (2001), no. 3, 227–255. MR 1836903

[GV85] Ira Gessel and Gérard Viennot, *Binomial determinants, paths, and hook length formulae*, Adv. in Math. **58** (1985), no. 3, 300–321. MR 815360

[HC20] Jake Huryn and Sergei Chmutov, *A few more trees the chromatic symmetric function can distinguish*, Involve, a Journal of Mathematics **13** (2020), no. 1, 109–116.

[HJ19] S. Heil and C. Ji, *On an algorithm for comparing the chromatic symmetric functions of trees*, Australas. J. Combin. **75** (2019), 210–222. MR 4008247

[LS19] Martin Loebl and Jean-Sébastien Sereni, *Isomorphism of weighted trees and Stanley's isomorphism conjecture for caterpillars*, Ann. Inst. Henri Poincaré D **6** (2019), no. 3, 357–384. MR 4002670

[LW24] Nicholas A. Loehr and Gregory S. Warrington, *A rooted variant of Stanley's chromatic symmetric function*, Discrete Math. **347** (2024), no. 3, Paper No. 113805, 18. MR 4670484

[MMW08] Jeremy L. Martin, Matthew Morin, and Jennifer D. Wagner, *On distinguishing trees by their chromatic symmetric functions*, J. Combin. Theory Ser. A **115** (2008), no. 2, 237–253. MR 2382514

[NW99] S. D. Noble and D. J. A. Welsh, *A weighted graph polynomial from chromatic invariants of knots*, Ann. Inst. Fourier (Grenoble) **49** (1999), no. 3, 1057–1087. MR 1703438

[OS14] Rosa Orellana and Geoffrey Scott, *Graphs with equal chromatic symmetric functions*, Discrete Math. **320** (2014), 1–14. MR 3147202

[Paw22] Brendan Pawlowski, *Chromatic symmetric functions via the group algebra of S_n*, Algebr. Comb. **5** (2022), no. 1, 1–20. MR 4389958

[Sta95] Richard P. Stanley, *A symmetric function generalization of the chromatic polynomial of a graph*, Adv. Math. **111** (1995), no. 1, 166–194. MR 1317387

[SW16] John Shareshian and Michelle L. Wachs, *Chromatic quasisymmetric functions*, Adv. Math. **295** (2016), 497–551. MR 3488041

[WYZ23] Yuzhenni Wang, Xingxing Yu, and Xiao-Dong Zhang, *A class of trees determined by their chromatic symmetric functions*, preprint, arXiv:2308.03980, 2023.

Departamento de Matemáticas, Facultad de Ciencias Exactas, Universidad Andres Bello, Santiago, Chile
Email address: jose.aliste@unab.cl

Department of Mathematics, University of Kansas, Lawrence, KS, United States
Email address: jlmartin@ku.edu

Department of Mathematics and Statistics, Washburn University, Topeka, KS, United States
Email address: jennifer.wagner1@washburn.edu

Departamento de Matemáticas, Facultad de Ciencias Exactas, Universidad Andres Bello, Santiago, Chile
Email address: josezamora@unab.cl