ON DISTINGUISHING TREES BY THEIR CHROMATIC SYMMETRIC FUNCTIONS

JEREMY L. MARTIN, MATTHEW MORIN, AND JENNIFER D. WAGNER

Abstract. Let T be an unrooted tree. The *chromatic symmetric function* $\mathbf{X}_T$, introduced by Stanley, is a sum of monomial symmetric functions corresponding to proper colorings of T. The *subtree polynomial* $\mathbf{S}_T$, first considered under a different name by Chaudhary and Gordon, is the bivariate generating function for subtrees of T by their numbers of edges and leaves. We prove that $\mathbf{S}_T = \langle \Phi, \mathbf{X}_T \rangle$, where $\langle \cdot, \cdot \rangle$ is the Hall inner product on symmetric functions and Φ is a certain symmetric function that does not depend on T. Thus the chromatic symmetric function is a stronger isomorphism invariant than the subtree polynomial. As a corollary, the path and degree sequences of a tree can be obtained from its chromatic symmetric function. As another application, we exhibit two infinite families of trees (*spiders* and some *caterpillars*), and one family of unicyclic graphs (*squids*) whose members are determined completely by their chromatic symmetric functions.

Introduction

Let G be a simple graph with vertices $V(G)$ and edges $E(G)$, and let $\mathbb{P}$ denote the positive integers. A *(proper) coloring* of G is a function $\kappa : V(G) \to \mathbb{P}$ such that $\kappa(v) \neq \kappa(w)$ whenever the vertices v, w are adjacent. Stanley ([16]; see also [17, pp. 462–464]) defined the *chromatic symmetric function* of G as

$$\mathbf{X}_G = \mathbf{X}_G(x_1, x_2, \dots) = \sum_{\kappa} \prod_{v \in V(G)} x_{\kappa(v)},$$

the sum over all colorings κ, where $x_1, x_2, \dots$ are countably infinitely many commuting indeterminates. This definition is invariant under permutations of $\{x_i\}$, so $\mathbf{X}_G$ is a symmetric function, homogeneous of degree $n = \#V(G)$.

The chromatic symmetric function is a much stronger isomorphism invariant than the well-known *chromatic polynomial* $\chi_G(k)$, a polynomial function of k that gives the number of colorings of G using at most k colors. Indeed, for any integer k, the number $\chi_G(k)$ can be obtained from $\mathbf{X}_G$ by setting $x_1 = \cdots = x_k = 1$ and $x_i = 0$ for all $i > k$.

It is natural to ask whether $\mathbf{X}_G$ is a complete isomorphism invariant; that is, whether two non-isomorphic graphs must have different chromatic symmetric functions. The answer is no; the smallest example, shown in Figure 1, was given by Stanley in [16]. Brylawski [2] introduced a graph invariant called the *polychromate* and constructed an infinite family of pairs of nonisomorphic graphs with the

2000 *Mathematics Subject Classification.* 05C05, 05C60, 05E05.

Key words and phrases. graph, tree, chromatic symmetric function.

The first author was supported in part by the University of Kansas New Faculty General Research Fund. The second author was supported in part by the National Sciences and Engineering Research Council of Canada.

FIGURE 1. Stanley's example [16] of two non-isomorphic graphs with the same chromatic symmetric function.

same polychromate. Sarmiento [15] proved that the polychromate is equivalent to the *U-polynomial* studied by Noble and Welsh [13], a stronger invariant than X_G; therefore, each pair in Brylawski's construction shares the same chromatic symmetric functions. (We thank Anna de Mier for directing our attention to these results.) In another direction, the non-commutative version of the chromatic symmetric function, studied by Gebhard and Sagan [7], is easily seen to be a complete invariant.

Stanley's question of whether $\mathbf{X}_G$ is a complete isomorphism invariant for *trees* remains open. This is equivalent to the problem of whether a tree is determined by its U-polynomial, since the formula for $\mathbf{X}_G$ in terms of the U-polynomial [13, Theorem 6.1] is easily seen to be reversible for trees. Stanley's question was answered in the affirmative for certain special kinds of trees by Fougere [6] and the second author [12], both of whom listed several other tree invariants that can be extracted from the chromatic symmetric function. Additionally, Tan [18] has verified computationally that the answer is "yes" for trees with 23 or fewer vertices. (In contrast, the chromatic polynomial is nearly useless for distinguishing trees, because $\chi_T(k) = k(k-1)^{n-1}$ for every tree T with n vertices.)

Our main tool is Stanley's expansion of the chromatic symmetric function in the basis of power-sum symmetric functions p_λ [16, Theorem 2.5]; see equation (5) below. When T is a (possibly trivial) tree, the coefficient $c_\lambda(T)$ of p_λ in $\mathbf{X}_T$ has a particularly simple combinatorial interpretation. For $A \subseteq E(T)$, define the *type* of A to be the partition whose parts are the sizes of the vertex sets of the graph with vertices $V(T)$ and edges A (see Figure 3 for an example). Then, up to a sign, $c_\lambda(T)$ is the number of edge sets A of type λ. As we will see, many other invariants of T can be recovered from $\mathbf{X}_T$.

Recall that the *degree* of a vertex is the number of edges incident to it. A *leaf* of a tree is a vertex of degree 1, and the unique incident edge is called a *leaf edge*. We define the *subtree polynomial* of T by

$$\mathbf{S}_T = \mathbf{S}_T(q,r) = \sum_{\text{subtrees } S} q^{\#S} r^{\#L(S)},$$

where the sum runs over all subtrees S of T with at least one edge, and $L(S)$ denotes the set of leaf edges of S. Setting $q = t(z+1)$ and $r = 1/(z+1)$ in $\mathbf{S}_T$ recovers the polynomial $f_E(T;t,z)$ studied by Chaudhary and Gordon in [3, Section 3]. Conversely, $f_E(T; qr, (1-r)/r) = \mathbf{S}_T(q,r)$, so the two polynomials provide identical information about T.

For every non-empty set $A \subseteq E(T)$, there is a unique minimal subset $K(A) \subseteq E(T) - A$, called the *connector* of A, such that $A \cup K(A)$ is a tree. (So $K(A) = \emptyset$ if and only if A is itself a subtree of T.) The *connector polynomial* of T is then defined as

$$\mathbf{K}_T = \mathbf{K}_T(x,y) = \sum_{\emptyset \neq A \subseteq E(T)} x^{\#A} y^{\#K(A)}.$$

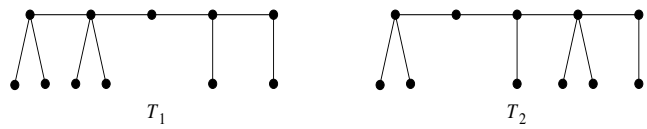

FIGURE 2. Two trees with the same subtree polynomial but different chromatic symmetric functions.

The polynomials $\mathbf{S}_T$ and $\mathbf{K}_T$ provide equivalent information about T; we will prove in Proposition 4 below that each of these invariants can be obtained from the other. Moreover, the path sequence and degree sequence of T can easily be recovered from $\mathbf{S}_T(q,r)$, as observed by Chaudhary and Gordon [3, Proposition 18].

For a partition $\lambda = (\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_\ell) \vdash n$, and integers a, b, i, j, define

$$\psi(\lambda, a, b) = (-1)^{a+b} \binom{\ell - 1}{\ell - n + a + b} \sum_{k=1}^{\ell} \binom{\lambda_k - 1}{a}, \tag{1}$$

$$\phi(\lambda, i, j) = (-1)^{i+j} \binom{\ell - 1}{\ell - n + i} \sum_{d=1}^{j} (-1)^d \binom{i-d}{j-d} \sum_{k=1}^{\ell} \binom{\lambda_k - 1}{d}. \tag{2}$$

We can now state our main theorem, which asserts that the subtree and connector polynomials can be recovered from the chromatic symmetric function $\mathbf{X}_T$.

Theorem 1. *For every $n \geq 1$, and for every tree T with n vertices,*

$$\mathbf{K}_T(x,y) \;=\; \sum_{a>0} \sum_{b \geq 0} x^a y^b \sum_{\lambda \vdash n} \psi(\lambda, a, b) c_\lambda(T) \tag{3}$$

and

$$\mathbf{S}_T(q,r) \;=\; \sum_{i=1}^{n-1} \sum_{j=1}^{i} q^i r^j \sum_{\lambda \vdash n} \phi(\lambda, i, j) c_\lambda(T). \tag{4}$$

It follows that the chromatic symmetric function is at least as strong an invariant as the subtree and connector polynomials. In particular, the path and degree sequences of T can be recovered from $\mathbf{X}_T$, as announced previously in [11]; this generalizes a earlier result of Fougere [6, Theorem 3.3.1]. Section 2 contains the proof of Theorem 1, as well as explicit formulas for the path and degree sequences, and a reinterpretation of (3) and (4) in terms of the usual scalar product on the space of symmetric functions.

Theorem 1 implies that $\mathbf{X}_T$ is a stronger invariant than $\mathbf{S}_T$. In fact, it is *strictly* stronger: the two trees shown in Figure 2 have different chromatic symmetric functions, but the same subtree polynomial. (Eisenstat and Gordon [4] constructed an infinite family of pairs of non-isomorphic trees with the same subtree polynomials, of which Figure 2 is the smallest example.) Thus Stanley's question remains open.

As another application of the combinatorial interpretation of the coefficients $c_\lambda(T)$, we identify some classes of trees for which the chromatic symmetric function is in fact a complete invariant. These trees include all *spiders* (trees having exactly one vertex of degree ≥ 3) and some *caterpillars* (trees from which deleting all leaves yields a path, such as those in Figure 2). We prove in Section 3 that every spider can be reconstructed from its subtree polynomial, hence from its chromatic symmetric function (generalizing results of Fougere [6]). The corresponding problem

for caterpillars is more difficult; however, certain special kinds of caterpillars can indeed be reconstructed from their chromatic symmetric functions, and the methods we use to prove this may be extendible to all caterpillars.

A *unicyclic graph* is a graph with one cycle. Connected unicyclic graphs can be recognized as such from their chromatic symmetric functions. While the combinatorial data provided by Stanley's expansion (5) is not as fine for unicyclic graphs as it is for trees, we can still obtain some uniqueness results for special unicyclic graphs by mimicking our results for spiders and caterpillars. In particular, we show in Section 5 that no two *squids* (unicyclic graphs with at most one vertex of degree two or more) can have the same chromatic symmetric function, although it is not clear whether membership in the class of squids can be determined from $\mathbf{X}_G$. An analogous result holds for *crabs* (unicyclic graphs in which every vertex not on the cycle is a leaf) satisfying an additional technical condition.

Acknowledgements. Our collaboration began at the Graduate Student Combinatorics Conference held at the University of Minnesota on April 16 and 17, 2005. We thank the organizers of the conference for their efforts, and we thank François Bergeron, Tom Enkosky, Gary Gordon, Brandon Humpert, Rosa Orellana, Victor Reiner, Bruce Sagan, and Stephanie van Willigenburg for many helpful conversations. The database of trees constructed by Piec, Malarz, and Kulakowski [14], and John Stembridge's freely available Maple package `SF`, were invaluable for calculating examples and formulating (and checking small cases of) Conjectures 6 and 7. Finally, we thank two anonymous referees for their suggestions, and in particular for making us aware of Fougere's undergraduate thesis [6].

1. Background

We assume that the reader is familiar with basic facts about graphs and trees (see, e.g., [1, Chapter I]). We denote a graph G by an ordered pair (V, E), where $V = V(G)$ is the set of vertices and $E = E(G)$ is the set of edges. All our graphs are *simple*; that is, we forbid loops and parallel edges. The *order* of a graph is its number of vertices. A *tree* is a graph G which is acyclic and connected and for which $\#V(G) = \#E(G) + 1$; any two of these conditions together imply the third. We consider the graph with one vertex and no edges to be a tree, the *trivial tree*; unless otherwise specified, all our statements about trees include this possibility. A *leaf* of a tree is a vertex of degree 1, that is, with exactly one incident edge. Every nontrivial tree has at least two leaves [1, p. 11]. It is often notationally convenient to ignore the distinction between a graph and its edge set.

We now review some facts about symmetric functions (for which the standard references are [10] and [17, Chapter 7]) and about the chromatic symmetric function (introduced by Stanley in [16]).

A *partition* is a sequence $\lambda = (\lambda_1, \ldots, \lambda_\ell)$ of positive integers in weakly decreasing order. The numbers λ_k are called the *parts* of λ. We say that λ is a partition of n, written $\lambda \vdash n$, if $\sum_k \lambda_k = n$. The number $\ell = \ell(\lambda)$ is called the *length* of λ.

Let $x_1, x_2, \ldots$ be a countably infinite set of commuting indeterminates. For $k \in \mathbb{P}$, the kth *power-sum symmetric function* is

$$p_k = \sum_{i \geq 1} x_i^k$$

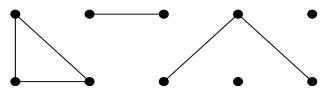

FIGURE 3. An edge set of type $(3,3,2,1,1)$.

and for a partition λ we define

$$p_\lambda = \prod_{k=1}^{\ell(\lambda)} p_{\lambda_k}.$$

It is well known that $\{p_\lambda \mid \lambda \vdash n\}$ is a basis for the $\mathbb{Q}$-vector space Λ_n consisting of all symmetric functions that are homogeneous of degree n.

Let G be a graph of order n. Stanley [16, Thm. 2.5] proved that

$$\mathbf{X}_G = \sum_{A \subseteq E(G)} (-1)^{\#A} p_{\mathrm{type}(A)}, \tag{5}$$

where $\mathrm{type}(A)$ is the partition whose parts are the orders of the connected components of the subgraph of G induced by A (see Figure 3 for an example). Note that $\mathrm{type}(A)$ depends upon A and $V(G)$, but not on $E(G)$. We write $c_\lambda(G)$ for the coefficient of p_λ in the expansion (5); that is,

$$\mathbf{X}_G = \sum_{\lambda \vdash n} c_\lambda(G) p_\lambda. \tag{6}$$

We will abbreviate $c_\lambda = c_\lambda(G)$ when no confusion can arise.

The chromatic symmetric function of a graph provides much more combinatorial information when that graph is a tree. In general, the coefficient c_λ does not count edge sets of type λ, because $(-1)^{\#A}$ is not constant for all such sets A. On the other hand, if $T = (V, E)$ is a tree of order n, then every $A \subseteq E$ is acyclic, so its induced subgraph has $n - \#A$ connected components. Hence $\ell(\mathrm{type}(A)) = n - \#A$, and we obtain a useful combinatorial interpretation for the numbers c_λ:

$$c_\lambda = (-1)^{n-\ell(\lambda)} \#\{A \subseteq E \mid \mathrm{type}(A) = \lambda\}. \tag{7}$$

The invariants c_λ are far from independent; in particular, (7) implies that

$$\sum_{\lambda:\ \ell(\lambda)=k} (-1)^{n-\ell(\lambda)} c_\lambda(T) = \binom{n-1}{k}. \tag{8}$$

We next list some basic invariants of graphs that can be recovered from its chromatic symmetric function. Several of these facts were previously observed by the second author in [12]. For notational simplicity, we shall often omit the parentheses and singleton parts of a partition, for instance, writing c_j rather than $c_{(j,1,1,\ldots,1)}$.

Proposition 2. *For every graph $G = (V, E)$,*

(i) *the symmetric function $\mathbf{X}_G$ is homogeneous of degree $\#V$;*
(ii) *$-c_2(G) = \#E$; and*
(iii) *the number k of connected components of G is $\min\{\ell(\lambda) \mid c_\lambda(G) \neq 0\}$.*

If G is a tree, then in addition

(iv) *for $k \geq 2$, $|c_j(G)|$ is the number of subtrees of G with j vertices; and*

(v) $c_{n-1}(G)$ *is the number of leaves of* G.

Proof. Assertion (i) is immediate from the definition of $\mathbf{X}_G$, and (ii) follows easily from (5). For (iii), (5) implies that $c_\lambda = 0$ whenever $\ell(\lambda) < k$. On the other hand, $c_\lambda = \sum_A (-1)^{\#A} = (-1)^{n-k} t_G(1,0)$, the sum over all $A \subseteq E(G)$ with $\lambda(A) = \lambda(E(G))$, where t_G is the Tutte polynomial of G (see [1, Ch. X]). Up to sign, this formula counts acyclic orientations of G with exactly one sink in each component (this follows from [1, Thm. 8, p. 348]). In particular, $c_\lambda \neq 0$.

Assertion (iv) holds because $A \subseteq E(G)$ is the set of edges of a j-vertex subtree if and only if $\mathrm{type}(A) = (j, 1, \ldots, 1)$, and (v) follows because every subtree of order $n-1$ is of the form $G - v$, where v is a leaf. □

By (i), (ii) and (iii) of Proposition 2, trees can be distinguished from non-trees by their chromatic symmetric functions. Moreover, part (v) implies that paths (trees with exactly two leaves) and stars (trees with exactly one non-leaf) are determined up to isomorphism by their chromatic symmetric functions.

The *girth* of a graph G is defined as the length of the smallest cycle in G, or ∞ if G is acyclic. With a little more work, we can compute the girth of G from $\mathbf{X}_G$. The idea is to find the smallest edge set for which (8) fails.

Proposition 3. *Let $G = (V, E)$ be a graph with n vertices and m edges. Let k be the largest number such that $\sum_{\lambda \vdash n,\, \ell(\lambda)=k} c_\lambda(G) \neq (-1)^{n-k}\binom{m}{n-k}$. Then the girth of G is $n-k+1$.*

Proof. Let g be the girth of G. Suppose first that $k > n-g+1$. Then $n-k < g-1$, so every subset $A \subseteq E$ with $n-k$ edges is acyclic and hence has k connected components. On the other hand, if $\ell(\mathrm{type}(A)) = k$, then the maximum size of a component of A is $n-(k-1) < g$, so A must be acyclic and hence must have $n-k$ edges. Therefore

$$\sum_{\substack{\lambda \vdash n \\ \ell(\lambda)=k}} c_\lambda(G) \;=\; \sum_{\substack{A \subseteq E \\ \ell(\mathrm{type}(A))=k}} (-1)^{\#A} \;=\; \sum_{\substack{A \subseteq E \\ \#A=n-k}} (-1)^{\#A} \;=\; (-1)^{n-k}\binom{m}{n-k}.$$

Now suppose that $k = n-g+1$. We claim that $A \subseteq E$ has k components if and only if it either has $n-k$ edges (hence is acyclic) or is precisely a cycle of length g. The "if" direction is evident. For the "only if" direction, suppose that A has k components and is not acyclic, hence contains a cycle C. By definition of G, the length of C cannot be less than g; on the other hand, there are at least $k-1$ vertices that do not belong to C (one for each other component of A), so $\#V(C) \leq n-(k-1) = g$. Thus C has length exactly g. Moreover, $A - C$ cannot contain any other edge with an endpoint outside C (because then it would have fewer than k components) or an edge joining two vertices of C (because then G would contain a cycle of length $< g$). Hence $A = C$ as desired. Denoting by Γ the set of g-cycles of G, we have

$$\begin{aligned}\sum_{\substack{\lambda \vdash n \\ \ell(\lambda)=k}} c_\lambda(G) \;&=\; \sum_{\substack{A \subseteq E \\ \#A=n-k}} (-1)^{\#A} + \sum_{A \in \Gamma} (-1)^{\#A} \\ &= (-1)^{n-k}\binom{m}{n-k} + (-1)^{n-k+1}\#\Gamma \;\neq\; (-1)^{n-k}\binom{m}{n-k}\end{aligned}$$

as desired. □

2. Proof of the main theorem

Theorem 1 expresses the subtree polynomial $\mathbf{S}_T$ and connector polynomial $\mathbf{K}_T$ of a tree T in terms of the chromatic symmetric function $\mathbf{X}_T$. The first step is to show that $\mathbf{S}_T$ and $\mathbf{K}_T$ are interchangeable. In what follows, we will often abuse notation by ignoring the distinction between a tree $T = (V, E)$ and its edge set E.

Proposition 4. *Let T be a tree. Then:*

(1) $\mathbf{S}_T(q, r) = \mathbf{K}_T(qr, q(1-r))$.
(2) $\mathbf{K}_T(x, y) = \mathbf{S}_T(x+y, x/(x+y))$.

Proof. For each nontrivial subtree $S \subset T$, write $L(S)$ for the set of leaf edges of S. Note that $\#L(S) \geq 1$, with equality if and only if S consists of a single edge. Moreover, observe that $A \cup K(A) = S$ if and only if $L(S) \subseteq A \subseteq S$. Hence

$$\begin{aligned}
\mathbf{K}_T(qr, q(1-r)) &= \sum_{A \subseteq T} (qr)^{\#A} (q(1-r))^{\#K(A)} \\
&= \sum_{A \subseteq T} q^{\#(A \cup K(A))} r^{\#A} (1-r)^{\#K(A)} \\
&= \sum_{\text{subtrees } S \subseteq T} q^{\#S} \sum_{\substack{A: \\ L(S) \subseteq A \subseteq S}} r^{\#A} (1-r)^{\#S - \#A} \\
&= \sum_{S} q^{\#S} r^{\#L(S)} \sum_{G \subseteq S - L(S)} r^{\#G} (1-r)^{\#(S - L(S)) - \#G} \\
&= \sum_{S} q^{\#S} r^{\#L(S)} \left(r + (1-r)\right)^{\#(S - L(S))} \;=\; \mathbf{S}_T(q, r),
\end{aligned}$$

giving the first equality. Meanwhile, solving the equations $x = qr, y = q(1-r)$ for q and r yields $q = x + y$, $r = x/(x+y)$, giving the second equality. □

We now prove the main theorem. To do so, we establish a formula for the connector polynomial of a tree in terms of its chromatic symmetric function, then apply Proposition 4 to obtain a formula for the subtree polynomial.

Proof of Theorem 1. By definition, the coefficient of $x^a y^b$ in $\mathbf{K}_T(x, y)$ is

$$\begin{aligned}
\#\{A \subseteq T \mid \#A = a,\ \#K(A) = b\} &= \sum_{\substack{A \subseteq T \\ \#A = a,\ \#K(A) = b}} 1 \\
&= (-1)^b \sum_{\substack{A \subseteq T \\ \#A = a}} \sum_{\substack{B \subseteq T - A \\ \#B = b}} \sum_{\substack{C \subseteq B \\ K(A) \subseteq C}} (-1)^{\#C},
\end{aligned}$$

because the innermost sum vanishes unless $B = K(A)$, when it is $(-1)^b$. Setting $D = B - C$, we may rewrite this expression as

$$(-1)^b \sum_{\substack{A \subseteq T \\ \#A = a}} \sum_{\substack{C \subseteq T - A \\ K(A) \subseteq C}} \sum_{\substack{D \subseteq T - A - C \\ \#D = b - \#C}} (-1)^{\#C}$$

and setting $F = C \cup A$ yields

$$\begin{aligned}
(-1)^b \sum_{\substack{A\subseteq T\\ \#A=a}} \sum_{\substack{F\subseteq T\\ A\cup K(A)\subseteq F}} \sum_{\substack{D\subseteq T-F\\ \#D=a+b-\#F}} (-1)^{\#F-a} \\
&= (-1)^{a+b} \sum_{F\subseteq T} \sum_{\substack{A:\ \#A=a,\\ A\cup K(A)\subseteq F}} \binom{\#(T-F)}{a+b-\#F} (-1)^{\#F} \\
&= (-1)^{a+b} \sum_{\lambda\vdash n} \sum_{\substack{F\subseteq T\\ \mathrm{type}(F)=\lambda}} \binom{\#(T-F)}{a+b-\#F} \sum_{\substack{A:\ \#A=a,\\ A\cup K(A)\subseteq F}} (-1)^{\#F} \\
&= (-1)^{a+b} \sum_{\lambda\vdash n} \binom{\ell(\lambda)-1}{\ell(\lambda)-n+a+b} (-1)^{n-\ell(\lambda)} \sum_{\substack{F\subseteq T\\ \mathrm{type}(F)=\lambda}} \alpha(F) \qquad (9)
\end{aligned}$$

where $\alpha(F) = \#\{A \mid \#A = a,\ A\cup K(A) \subseteq F\}$. The set $A\cup K(A)$ is connected, so if it is a subset of F then it must be a subset of some component of F. On the other hand, if F' is a (possibly trivial) component of F and $A \subseteq F'$, then $A\cup K(A) \subseteq F'$, because F' is a tree containing A and $A \cup K(A)$ is the unique minimal such tree. Thus if $\mathrm{type}(F) = \lambda$ then

$$\alpha(F) = \sum_{k=1}^{\ell(\lambda)} \binom{\lambda_k - 1}{a}. \qquad (10)$$

Note that this formula is valid only if $a > 0$. Substituting (10) into (9), we obtain

$$\begin{aligned}
&(-1)^{a+b} \sum_{\lambda\vdash n} \binom{\ell(\lambda)-1}{\ell(\lambda)-n+a+b} (-1)^{n-\ell(\lambda)} \sum_{\substack{F\subseteq T\\ \mathrm{type}(F)=\lambda}} \sum_{k=1}^{\ell(\lambda)} \binom{\lambda_k-1}{a} \\
&= (-1)^{a+b} \sum_{\lambda\vdash n} \binom{\ell(\lambda)-1}{\ell(\lambda)-n+a+b} \sum_{k=1}^{\ell(\lambda)} \binom{\lambda_k-1}{a} c_\lambda(T) \\
&= \sum_{\lambda\vdash n} \psi(\lambda,a,b) c_\lambda(T)
\end{aligned}$$

(where $\psi(\lambda, a, b)$ is defined by (1)), giving the desired formula (3).

We now turn to the proof of (4). By Proposition 4, we have

$$\begin{aligned}
\mathbf{S}_T(q,r) &= \mathbf{K}_T(qr, q(1-r)) \\
&= \sum_{a>0}\sum_{b\geq 0} (qr)^a (q(1-r))^b (-1)^{a+b} \sum_{\lambda\vdash n} \binom{\ell(\lambda)-1}{\ell(\lambda)-n+a+b} \sum_{k=1}^{\ell(\lambda)} \binom{\lambda_k-1}{a} c_\lambda(T).
\end{aligned}$$

Setting $i = a + b$, we may rewrite the last expression as

$$\sum_{i>0}\sum_{a=1}^{i} (-1)^i q^i r^a (1-r)^{i-a} \sum_{\lambda\vdash n} \binom{\ell(\lambda)-1}{\ell(\lambda)-n+i} \sum_{k=1}^{\ell(\lambda)} \binom{\lambda_k-1}{a} c_\lambda(T).$$

Applying the binomial expansion to $(1-r)^{i-a}$ yields

$$\sum_{i>0}\sum_{a=1}^{i}(-1)^i q^i r^a \sum_{h=0}^{i-a}\binom{i-a}{h}(-1)^h r^h \sum_{\lambda\vdash n}\binom{\ell(\lambda)-1}{\ell(\lambda)-n+i}\sum_{k=1}^{\ell(\lambda)}\binom{\lambda_k-1}{a}c_\lambda(T).$$

Now setting $h=j-a$ gives

$$\sum_{i>0}\sum_{a=1}^{i}(-1)^i q^i \sum_{j=a}^{i}\binom{i-a}{j-a}(-1)^{j-a} r^j \sum_{\lambda\vdash n}\binom{\ell(\lambda)-1}{\ell(\lambda)-n+i}\sum_{k=1}^{\ell(\lambda)}\binom{\lambda_k-1}{a}c_\lambda(T)$$

and setting $a=d$ and rearranging gives

$$\sum_{i>0}\sum_{j=1}^{i} q^i r^j \sum_{\substack{\lambda\vdash n\\ \ell(\lambda)=\ell}}\left((-1)^{i+j}\binom{\ell-1}{\ell-n+i}\sum_{d=1}^{j}(-1)^d\binom{i-d}{j-d}\sum_{k=1}^{\ell}\binom{\lambda_k-1}{d}\right)c_\lambda(T)$$

$$= \sum_{i>0}\sum_{j=1}^{i} q^i r^j \sum_{\substack{\lambda\vdash n\\ \ell(\lambda)=\ell}}\phi(\lambda,i,j)c_\lambda(T)$$

which is the desired formula (4). □

Two basic invariants of a tree are its *path sequence* and its *degree sequence*. The path sequence of T is defined as $(\pi_1,\pi_2,\dots)$, where $\pi_i=\pi_i(T)$ is the number of i-edge paths in T. The degree sequence of T is defined as $(\delta_1,\delta_2,\dots)$, where $\delta_j=\delta_j(T)$ is the number of degree-j vertices in T. Knowing the degree sequence is equivalent to knowing the *star sequence* $(\sigma_1,\sigma_2,\dots)$, where $\sigma_k=\sigma_k(T)$ is the number of k-edge stars in T. Indeed, it is not hard to see that

$$\sigma_k=\sum_{j\geq k}\binom{j}{k}\delta_j$$

for every $2\leq k\leq n-1$, and so

$$\delta_j=\sum_{k\geq j}\binom{k}{j}(-1)^{j+k}\sigma_k.$$

Corollary 5. *The degree and path sequences of a tree T can be recovered from its chromatic symmetric function.*

Proof. The key observation, due to Chaudhary and Gordon [3, Proposition 18], is that the path and star sequences of T can be recovered from $\mathbf{S}_T$. Indeed, π_1 is the number of edges of T, and for every $i\geq 2$, π_i is just the coefficient of q^ir^2 in $\mathbf{S}_T(q,r)$. Meanwhile, for every $k\geq 1$, σ_k is the coefficient of q^kr^k. □

We note that Fougere had proved [6, Theorem 3.3.1] that the sum of the squared vertex degrees, $\sum_j \delta_j j^2$, could be obtained from the coefficient of the monomial symmetric function $m_{(3,1,1,\dots)}$ in $\mathbf{X}_T$.

We can rephrase the formulas for $\mathbf{K}_T$ and $\mathbf{S}_T$ in terms of the usual scalar product $\langle\cdot,\cdot\rangle$ on the space Λ_n of degree-n symmetric functions (see [17, §7.9] or [10, §I.4]),

where n is the order of T. Define symmetric functions $\Psi_n(x,y)$ and $\Phi_n(q,r)$ by

$$\Psi_n(x,y) = \sum_{a>0}\sum_{b\geq 0} x^a y^b \sum_{\lambda\vdash n} \psi(\lambda,a,b)\frac{p_\lambda}{z_\lambda},$$

$$\Phi_n(q,r) = \sum_{i=1}^{n-1}\sum_{j=1}^{i} q^i r^j \sum_{\lambda\vdash n} \phi(\lambda,i,j)\frac{p_\lambda}{z_\lambda}.$$

Then the formulas (3) and (4) are respectively equivalent to

$$\mathbf{K}_T(x,y) = \big\langle \Psi_n(x,y),\ \mathbf{X}_T\big\rangle,$$

$$\mathbf{S}_T(q,r) = \big\langle \Phi_n(q,r),\ \mathbf{X}_T\big\rangle.$$

The symmetric function Ψ_n appears to have certain positivity and integrality properties, as we now explain. The ith *homogeneous symmetric function* h_i in indeterminates $x_1, x_2, \ldots$ is the sum of all monomials of degree i, and for a partition $\lambda = (\lambda_1, \ldots, \lambda_\ell)$ we define $h_\lambda = h_{\lambda_1}\cdots h_{\lambda_\ell}$. The symmetric functions $\{h_\lambda \mid \lambda \vdash n\}$ form a vector space basis for Λ_n [10], so there is a unique list of rational numbers $\xi(\lambda,i,j) \in \mathbb{Q}$ such that

$$\Psi_n(x,y) = \sum_{i,j}\sum_{\lambda\vdash n} \xi(\lambda,i,j) x^i y^j h_\lambda.$$

Conjecture 6 (Positivity)**.** *Let $\mu \vdash n$ be a partition, and let $\epsilon(\mu)$ be the number of parts of μ of even length. Then, for all integers i,j, $(-1)^{\epsilon(\mu)}\xi(\mu,i,j) \geq 0$.*

Conjecture 7 (z-Integrality)**.** *Let $\mu \vdash n$ be a partition. Then, for all integers i,j, the number $\xi(\mu,i,j)z_\mu$ is an integer.*

We have verified Conjectures 6 and 7 computationally[1] for all $n \leq 20$, which we think is strong evidence that they hold for all n. A formula for $\xi(\mu,i,j)$ can be written out explicitly using the known transition matrices between bases of symmetric functions (see [5]). However, we do not know a direct combinatorial interpretation for $\xi(\mu,i,j)$ or for $\xi(\mu,i,j)z_\mu$.

One might hope that for every two trees T, U with $\#V(T) > \#V(U) > 1$, the number of subtrees of T isomorphic to U might be given by a scalar product $\big\langle \zeta_U,\ \mathbf{X}_T\big\rangle$, where ζ_U is some symmetric function independent of T. Such a result would generalize Corollary 5 (which covers only the case that U is a path or a star) and, by a theorem of Harary and Palmer [9], would imply that every tree is distinguished by its chromatic symmetric function. In fact, it appears that such a function ζ_U exists *only if* U is a star or a path, as we have verified computationally for all U of order ≤ 8, with one trivial exception.[2]

Theorem 1 does not resolve Stanley's question, because $\mathbf{S}_T$ is not a complete isomorphism invariant. Indeed, the two trees T_1, T_2 shown in Figure 2 share the same subtree polynomial; this is a special case of a theorem of Eisenstat and Gordon [4]. On the other hand, $\mathbf{X}_{T_1} \neq \mathbf{X}_{T_2}$. This inequality follows from Tan's calculations [18], and also for the following elementary reason. Let $A \subset E(T_1)$ be the edge set obtained by deleting the two rightmost horizontal edges in Figure 2;

[1] A Maple worksheet containing the calculations is available at `http://math.ku.edu/~jmartin/sourcecode/`.

[2] Up to isomorphism, there are three four-edge trees: the star S_4, the path P_4, and another tree U. Since the number of four-edge subtrees of T is just $c_5(T)$, we have $\zeta_U = c_5 - \zeta_{S_4} - \zeta_{P_4}$.

then $\mathrm{type}(A) = (7,2,2)$. On the other hand, no subset of $E(T_2)$ has that type. Therefore $c_{(7,2,2)}(T_1) \neq 0$ and $c_{(7,2,2)}(T_2) = 0$.

The remainder of the article is devoted to identifying special classes of trees T for which the invariants c_λ suffice to reconstruct T up to isomorphism.

3. The Chromatic Symmetric Function Distinguishes Spiders

A tree is a *spider* (or *starlike tree*) if exactly one of its vertices has degree ≥ 3. By Corollary 5, whether or not a tree is a spider can be determined from its subtree polynomial. A spider may equivalently be defined as a collection of edge-disjoint paths (the *legs*) joined at a common endpoint t (the *torso*).

Up to isomorphism, every spider on n vertices can be described by a partition $\lambda \vdash n-1$ whose parts are the lengths of its legs (so $\ell(\lambda) \geq 3$). We denote the corresponding spider by T_λ. Note that $\ell(\lambda)$ equals both the number of leaves of T_λ and the degree of its torso.

We will show that the isomorphism type of a spider can be determined from its subtree polynomial, hence from its chromatic symmetric function. Fougere [6, Chapter 2] had previously shown that *forks* (spiders with exactly one leg of length > 1 and *extended stars* (spiders in which every leg has length k or $k+1$ for some k) could be reconstructed from their chromatic symmetric functions.

Before continuing, we describe a combinatorial problem whose solution will play a role in the proof. Let $m_1, \ldots, m_k$ be nonnegative integers with $\sum m_i = \ell$. Suppose that we have ℓ distinguishable boxes, of which m_i have capacity i for each $i \in [k]$. Let $\Omega(m_1, \ldots, m_k)$ be the number of ways of distributing k indistinguishable balls among these boxes so that no box is filled beyond its capacity. In general, it is not easy to write down a closed formula for $\Omega(m_1, \ldots, m_k)$, although individual instances can be computed using an inclusion-exclusion argument (for example).

Theorem 8. *Let $\lambda \vdash n-1$ be a partition with $\ell = \ell(\lambda) \geq 3$, and let $T = T_\lambda$ be the corresponding spider. Then T can be reconstructed from its subtree polynomial.*

Proof. First, suppose that $\ell = 3$. Then $\lambda_1 + \lambda_2$, $\lambda_1 + \lambda_2 + \lambda_3$, and $\lambda_1\lambda_2\lambda_3$ are respectively the diameter, number of edges, and number of three-leaf subtrees of T_λ. These invariants can be recovered from $\mathbf{S}_\lambda$, and together they determine λ.

Now suppose that $\ell > 3$. Let m_k denote the number of parts of size k in λ, and let $s(i,j)$ denote the number of subtrees of the spider T with i edges and j leaf edges (that is, the coefficient of $q^i r^j$ in $\mathbf{S}_T(q,r)$). We will show by induction on k that m_k can be calculated from the numbers $s(i,j)$.

First, suppose $k = 1$. Since there is a bijection between legs of T of length 1 (i.e., consisting of a single edge) and subtrees with $n-2$ edges and $\ell-1$ legs (which are formed by deleting such an edge). Hence $m_1 = s(n-2, \ell-1)$.

For $k > 1$, we can choose a subtree $S \subset T$ with $n-1-k$ edges and $\ell-1$ leaves as follows. First, fix $j \in [k]$ and delete a leg of length j; there are m_j ways to do this. If $j < k$, then we still need to delete $k-j$ more edges. It suffices to specify how many edges to delete from the end of each of the other $\ell-1$ legs, so the number of ways to do this is the solution to the balls-in-boxes problem described above, regarding the $k-j$ edges to be deleted as balls and each remaining leg of length i as a box of capacity $\min(i-1, k-j)$ (since deleting the entire leg will result in a tree with fewer than $\ell-1$ leaves). Therefore $s(n-k-1, \ell-1)$ is given by the

formula

$$m_k + \sum_{j=1}^{\lfloor k/2 \rfloor} m_j \Omega\big(m_2, \ldots, m_{j-1}, m_j - 1, m_{j+1}, \ldots, m_{k-j}, \ell - (m_1 + \cdots + m_{k-j})\big)$$
$$+ \sum_{j=\lfloor k/2 \rfloor + 1}^{k-1} m_j \Omega\big(m_2,\ \ldots,\ m_{k-j},\ \ell - 1 - (m_1 + \cdots + m_{k-j})\big). \quad (11)$$

By induction, the $s(i,j)$ determine $m_1, \ldots, m_{k-1}$, and equation (11) implies that they determine m_k as well. □

Another way of reconstructing a spider from its chromatic symmetric function will be useful in Section 5. Define the *two-part portion* $\tilde{\mathbf{X}}_G$ of $\mathbf{X}_G$ by

$$\tilde{\mathbf{X}}_G = \sum_{\ell(\lambda)=2} c_\lambda p_\lambda. \quad (12)$$

While $\tilde{\mathbf{X}}_G$ is evidently a much weaker invariant than $\mathbf{X}_G$, it contains enough information to distinguish among spiders.

Theorem 9. *Let T be a spider of order n. Let $m = \lfloor \frac{n}{2} \rfloor$, let $\varepsilon = n - 2m$, let $d_a = |c_{a,n-a}(T)|$ for $1 \le a \le m$, and let $d(T) = (d_1, \ldots, d_m)$. Then one of the following conditions holds:*

1. *The sequence $d(T)$ is a partition, that is, $d_1 \ge d_2 \ge \cdots \ge d_m \ge 0$.*
2. *There is a number $t \le m$ such that $d_1 \ge \cdots \ge d_t = 1$. Moreover, $d_{t+1} = d_{t+2} = \cdots = d_{m-1} = 2$, and $d_m \in \{1, 2\}$.*

In the first case, let $\mu = d(T)$. In the second case, define a partition μ from $d(T)$ by replacing all 2's after the tth place with two 1's.

Then $T = T_\lambda$, where λ is the conjugate partition to μ.

Proof. For each edge $e \in E(T)$, denote by $\|e\|$ the minimum of the orders of the components of $T - e$, so that $d_a = \#\{e \mid \|e\| = a\}$. If we label the legs of T as $L_1, \ldots, L_\ell$, where L_i has λ_i edges, and label the edges of each L_i as $e_{i,1}, e_{i,2}, \ldots, e_{i,\lambda_i}$, starting from the leaf and reading in toward the torso, then $\|e_{i,j}\| = \min(j, n-j)$.

Case 1: $\lambda_1 \le n - \lambda_1$. Then $j \le n - j$ for every edge $e_{i,j}$. In this case $d_a = \#\{i \mid \lambda_i \ge a\}$, and $d(T)$ is just the conjugate partition of λ.

Case 2: $\lambda_1 > n - \lambda_1$. If $i \ne 1$, then $\lambda_i \le (\lambda_2 + \cdots + \lambda_k) - 1 = n - \lambda_1 - 2$, and $\|e_{i,j}\| = j$ for every j. So we can give a formula for $d(T)$ in terms of λ:

$$d_a = \begin{cases} \#\{i \mid \lambda_i \ge a\} & \text{if } 1 \le a \le n - \lambda_1 - 1, \\ 2 & \text{if } n - \lambda_1 \le a \le m - 1, \\ 1 + \varepsilon & \text{if } a = m. \end{cases} \quad (13)$$

Note that $d_{n-\lambda_1-1} = 1$, because $e_{1,n-\lambda_1-1}$ is the unique edge whose deletion contains a component of order $n - \lambda_1 - 1$. On the other hand, $n - \lambda_1 = 2m + \varepsilon - \lambda_1 \le 2m + \varepsilon - (m+1) = m + \varepsilon - 1$, so $d_{n-\lambda_1} = 2$ whether n is odd or even. Therefore, $d(T)$ is not a partition, but has the form described in case (2) of the theorem. Formula (13) implies that if we "iron out" $d(T)$ by replacing every 2 after the $(n - \lambda_1)^{th}$ entry with two 1's, we will obtain the conjugate partition to λ.

We now see how to recover λ, and thus the isomorphism type of the spider T_λ, from the data $d(T_\lambda)$. Either $d(T_\lambda)$ is a partition, in which case it is the conjugate

of λ, or it has the form just described, in which case the "ironing-out" operation yields the conjugate of λ. This is precisely the statement of the theorem. □

4. Chromatic Symmetric Functions of Some Caterpillars

A *caterpillar* is a tree T with the property that the induced subgraph on the non-leaf vertices is a nontrivial path, called the *spine* of T. That is, every vertex of T either lies on the spine, or is a leaf whose unique neighbor lies on the spine. Since the spine is nontrivial, a caterpillar must have at least four vertices.

While Eisenstat and Gordon's result in [4] rules out the possibility of distinguishing caterpillars by their subtree polynomials, there is still reason to hope that the additional information provided by the chromatic symmetric function of a caterpillar may suffice to reconstruct it up to isomorphism.

Our first result is that the chromatic symmetric function distinguishes caterpillars from non-caterpillars. The number of leaves and the diameter of a tree T (the maximum length of a path in T) can be recovered from $\mathbf{X}_T$ by Corollary 5, so it suffices to prove the following fact.

Proposition 10. *Let T be a tree with $n \geq 4$ vertices. Then T is a caterpillar if and only if* $\operatorname{diam}(T) - 1 = n - \delta_1(T)$.

Proof. If T is a caterpillar, then every path of maximum length consists of the spine together with a leaf attached to each of its endpoints, hence contains all the non-leaf vertices and two other (leaf) vertices. In particular, the number of edges in such a path is one more than the number of non-leaf vertices. On the other hand, if P is a path of maximum length in T, then the internal vertices of P are not leaves of T (because each has two neighbors in P) but its endpoints are (otherwise P could be lengthened). If $\operatorname{diam}(T) - 1 = n - \delta_1(T)$, then all the vertices not lying on P must be leaves, which is to say that T is a caterpillar. □

Let T be a caterpillar with spine vertices $v_0, \ldots, v_s$. For each i, let $e_i = \deg(v_i) - 1$, where $\deg(v_i)$ denotes the degree of the vertex v_i. Gordon and McDonnell [8, Lemma 2] showed that the numbers e_i are almost determined by the path sequence of T, and are indeed determined by the path sequence when the caterpillar is *symmetric* (that is, $e_i = e_{s-i}$ for $0 \leq i \leq s$). Therefore, every symmetric caterpillar is determined up to isomorphism by its chromatic symmetric function, a result proved in another way by the second author [12, Theorem 4.3.1].

We now describe another class of caterpillars that can be reconstructed from their chromatic symmetric functions. We retain the labeling of the vertices of T as $v_0, \ldots, v_s$. Let f_i be the number of *leaves* adjacent to v_i, so that $f_i = \deg(v_i) - 1$ for $i = 0, s$ and $f_i = \deg(v_i) - 2$ for $0 < i < s$. Call f_i the ith *leaf number* of T. In addition, call a partition λ *singleton-free* if all its parts are at least 2.

Theorem 11. *Let T be a caterpillar whose leaf numbers f_i are strictly positive and distinct. Then T can be reconstructed from its chromatic symmetric function.*

Proof. Let L be the set of leaf edges of T. Since $f_i > 0$ for all i, every spine vertex is adjacent to at least one leaf. Therefore, the edge sets $A \subseteq T$ such that $\operatorname{type}(A)$ is singleton-free are precisely those that contain L. In particular, $\lambda = \operatorname{type}(L)$ is the unique singleton-free partition with $s+1$ parts whose coefficient $c_\lambda(T)$ is non-zero. Up to reordering, the parts of λ are the numbers $f_0 + 1, \ldots, f_s + 1$. Furthermore, for each spine edge $e_i = v_{i-1}v_i$, the edge set $L \cup \{e_i\}$ contributes $(-1)^{n-s}$ to $c_{\mu_i}(T)$,

where $\mu_i = \{f_0 + 1, \ldots, f_{i-2} + 1, f_{i-1} + f_i + 2, f_{i+1} + 1, \ldots, f_s + 1\}$. Note that the partitions μ_i are all distinct. Moreover, μ_i has $s - 1$ of its parts in common with λ; the remaining two parts of λ must be the leaf numbers of adjacent spine vertices of T. (This statement is valid even if the parts of μ_i are not all distinct.) In this way we can recover the leaf numbers of the endpoints of every edge of T, and this data specifies the caterpillar up to isomorphism. □

Using the same argument, we can relax the condition of the theorem slightly: we need only require that all leaf numbers are positive and that for each k, the set of spine vertices with leaf number k form a subpath of the spine.

5. Unicyclic Graphs: Squids and Crabs

Despite the title of this article, we devote the last section to a family of graphs G that are not trees, but rather *unicyclic*; that is, G contains a unique cycle. Equivalently, $c = n - e + 1$, where c, n, e are the numbers of components, vertices and edges respectively, so unicyclicity can be detected from X_G by Proposition 2. While we can no longer interpret the coefficients c_λ as in (7), we can use Stanley's expansion of $\mathbf{X}_G$ in terms of broken circuits [16, Thm. 2.9]. A special case of that result is the following: if G has a unique cycle C and e_0 is an edge of C, then

$$\mathbf{X}_G = \sum_{\substack{A \subseteq E(G) \\ C - e_0 \not\subseteq A}} (-1)^{\#A} p_{\mathrm{type}(A)}. \tag{14}$$

We do not know whether there exist two unicyclic graphs with the same chromatic symmetric functions.

A *squid* is a connected unicyclic graph with a single vertex v of degree greater than 2. Note that v must lie on the cycle. A squid is described up to isomorphism by the length of its cycle and of the *tentacles* (the paths from the leaf vertices to v). For example, if T_λ is the spider whose leg lengths are given by the parts of λ, then adding an edge between the leaves at the ends of the two longest legs produces a squid with cycle length $\lambda_1 + \lambda_2$ and tentacle lengths $\lambda_3, \lambda_4, \ldots$.

It is not clear how to determine from $\mathbf{X}_G$ whether or not a unicyclic graph G is a squid (for instance, we cannot recover the degree sequence of an arbitrary graph from its chromatic symmetric function as we can for a tree). Nevertheless, the following uniqueness result does hold.

Theorem 12. *No two non-isomorphic squids have the same chromatic symmetric function.*

Proof. Let S be a squid with unique cycle C. Let $k + 1$ be the length of C; this number can be recovered from $\mathbf{X}_S$ by Proposition 3. Let v be the unique vertex of S of degree > 2, and let μ be the partition whose parts are the edge lengths of the tentacles. Label the edges of C as $e_0, e_1, \ldots, e_k$, starting at v and proceeding around the cycle. By (14) and inclusion-exclusion, we obtain

$$\begin{aligned}\mathbf{X}_S &= \sum_{i=1}^{k} \left(\sum_{e_i \notin A} (-1)^{\#A} p_{\mathrm{type}(A)} \right) - \sum_{1 \le i < j \le k} \left(\sum_{e_i, e_j \notin A} (-1)^{\#A} p_{\mathrm{type}(A)} \right) + \ldots \\ &= \sum_{i=1}^{k} \mathbf{X}_{S - e_i} - \sum_{1 \le i < j \le k} \mathbf{X}_{S - e_i - e_j} + \ldots\end{aligned}$$

where the omitted terms involve edge sets A lacking three or more edges from $C-e_0$. Deleting three or more edges from S produces a graph with three or more connected components, so passing to two-part portions as in (12) yields $\tilde{\mathbf{X}}_S = \sum_{i=1}^{k} \tilde{\mathbf{X}}_{S-e_i} - \sum_{1\le i<j\le k} \tilde{\mathbf{X}}_{S-e_i-e_j}$. Each graph $S-e_i-e_j$ has exactly two components, of sizes $j-i$ and $n-j+i$. Removing additional edges will strictly increase the number of components, so

$$\tilde{\mathbf{X}}_S = \sum_{i=1}^{k} \tilde{\mathbf{X}}_{T_{\mu,i,k-i}} - \sum_{1\le i<j\le k} (-1)^{n-2} p_{(j-i,n-j+i)}$$

where $T_{\mu,i,k-i} = S - e_i$ is the spider with legs whose lengths are i, $k-i$, and the parts of μ. Therefore, $\mathbf{X}_S$ determines the quantity

$$\sum_{i=1}^{k} \tilde{\mathbf{X}}_{T_{\mu,i,k-i}} = \tilde{\mathbf{X}}_S + \sum_{1\le i<j\le k} (-1)^{n-2} p_{(j-i,n-j+i)}. \tag{15}$$

Leaving the foregoing calculations aside for the moment, we note that if T is a tree with $n-1$ edges, then to calculate $\tilde{\mathbf{X}}_T$ we need only consider the edge subsets of cardinality $n-2$. In particular, if $T = T_\lambda$ is a spider, then $\tilde{\mathbf{X}}_T = (-1)^{n-2} \sum_{i=1}^{\ell(\lambda)} \sum_{j=1}^{\lambda_i} p_{(j,n-j)}$. It follows that for any partition μ and numbers k, i, we have

$$\begin{aligned}\tilde{\mathbf{X}}_{T_{\mu,k}} &= \tilde{\mathbf{X}}_{T_{\mu,i,k-i}} + (-1)^{n-2}\left(\sum_{j=1}^{k} p_{(j,n-j)} - \sum_{j=1}^{i} p_{(j,n-j)} - \sum_{j=1}^{k-i} p_{(j,n-j)}\right)\\ &= \frac{1}{k}\sum_{i=1}^{k}\left(\tilde{\mathbf{X}}_{T_{\mu,i,k-i}} + (-1)^{n-2}\left(\sum_{j=1}^{k} p_{(j,n-j)} - \sum_{j=1}^{i} p_{(j,n-j)} - \sum_{j=1}^{k-i} p_{(j,n-j)}\right)\right)\end{aligned}$$

which can be computed from $\mathbf{X}_S$ using (15). Meanwhile, by Theorem 9, we can reconstruct the spider $T_{\mu,k}$ from $\tilde{\mathbf{X}}_{T_{\mu,k}}$. In particular, we can reconstruct the partition μ, which gives the tentacle lengths of the squid S. □

Just as squids can be regarded as the unicyclic analogues of spiders, the unicyclic analogues of caterpillars are *crabs*: connected unicyclic graphs in which every vertex not lying on the cycle is a leaf. In analogy to Theorem 12 and its proof, we ask whether is it possible to use the results of Section 4 to classify the chromatic symmetric functions of (some) crabs. As a starting point, we prove the following analogue of Theorem 11.

Proposition 13. *Let G be a crab such that the degrees of its non-leaf vertices are all distinct and all greater than 2. Then, if H is another crab with this property, and $\mathbf{X}_G = \mathbf{X}_H$, then $G \equiv H$.*

Proof. We will show that the coefficients of $\mathbf{X}_G$, together with the knowledge that G is a crab with the property just mentioned, determine G up to isomorphism.

The girth g of G can be recovered from $\mathbf{X}_G$ by Proposition 3. Let C be the unique cycle of G, and label its vertices in cyclic order as $v_1, \ldots, v_g$. Let $f_i = \deg(v_i) - 2$ be the number of leaves adjacent to v_i. Note that G can be specified up to isomorphism by the cyclically ordered list of numbers $f_1, \ldots, f_g$.

Let L denote the set of leaf edges of G. The subsets of $E(G)$ whose type is singleton-free are precisely those that contain L as a subset. In particular, L itself

is the unique edge set whose type is a singleton-free partition of length g. Thus $\mathrm{type}(L) = \{f_1 - 1, f_2 - 1, \ldots, f_g - 1\}$ can be recovered from $\mathbf{X}_G$. Moreover, there are precisely g edge sets whose type is a singleton-free partition of length $g - 1$; these edge sets are of the form $L \cup \{e\}$ for some $e \in C$. Just as in Theorem 11, the types of these edge sets are all distinct, and they specify which pairs of the f_i correspond to adjacent vertices of the cycle. □

References

[1] Béla Bollobás, *Modern Graph Theory*, Graduate Texts in Mathematics, 184, Springer-Verlag, New York, 1998.

[2] Tom Brylawski, *Intersection theory for graphs*, J. Combin. Theory Ser. B **30** (1981), 233–246.

[3] Sharad Chaudhary and Gary Gordon, *Tutte polynomials for trees*, J. Graph Theory **15** (1991), no. 3, 317–331.

[4] David Eisenstat and Gary Gordon, *Non-isomorphic caterpillars with identical subtree data*, Discrete Math. **306** (2006), 827–830.

[5] Ömer Eğecioğlu and Jeffrey B. Remmel, *Brick tabloids and the connection matrices between bases of symmetric functions*, Discrete Appl. Math. **34** (1991), 107–120.

[6] Joshua Fougere, *On symmetric chromatic polynomials of trees*, undergraduate thesis, Dartmouth College, 2003.

[7] David D. Gebhard and Bruce E. Sagan, *A chromatic symmetric function in noncommuting variables*, J. Alg. Combin. **2** (2001), 227–255.

[8] Gary Gordon and Eleanor McDonnell, *Trees with the same degree sequence and path number*, Discrete Math. **147** (1995), 297–300.

[9] Frank Harary and Ed Palmer, *The reconstruction of a tree from its maximal subtrees*, Canadian J. Math. 18 (1966), 803–810.

[10] I.G. Macdonald, *Symmetric Functions and Hall Polynomials*, 2nd edn., The Clarendon Press, Oxford University Press, New York, 1995.

[11] Jeremy L. Martin and Jennifer Wagner, *On the chromatic symmetric function of a tree*, extended abstract for 18th Annual Conference on Formal Power Series and Algebraic Combinatorics (San Diego, 2006).

[12] Matthew Morin, *Caterpillars, Ribbons, and the Chromatic Symmetric Function*, M.S. thesis, University of British Columbia, 2005.

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

[14] S. Piec, K. Malarz, and K. Kulakowski, *How to count trees?*, preprint, `arXiv:cond-mat/0501594`, 2005.

[15] Irasema Sarmiento, *The polychromate and a chord diagram polynomial*, Ann. Comb. **4** (2000), 227–236.

[16] Richard P. Stanley, *A symmetric function generalization of the chromatic polynomial of a graph*, Adv. Math. **111** (1995), 166–194.

[17] Richard P. Stanley, *Enumerative Combinatorics*, vol. 2, Cambridge University Press, Cambridge, 1999.

[18] Li-Yang Tan, *Computational verification that the chromatic symmetric function determines up to isomorphism trees of order at most 23*, `http://www.cs.wustl.edu/~lt1/csf.html`, retrieved on May 14, 2007.

Department of Mathematics, University of Kansas, 405 Snow Hall, 1460 Jayhawk Blvd., Lawrence, KS 66045, USA

E-mail address: `jmartin@math.ku.edu`

Department of Mathematics, University of British Columbia, Room 121, 1984 Mathematics Road, Vancouver, BC, Canada, V6T 1Z2

E-mail address: `mjmorin@math.ubc.ca`

Department of Mathematics and Statistics, Washburn University, 1700 SW College Ave., Topeka, KS 66621, USA

E-mail address: `jennifer.wagner1@washburn.edu`