Formal Power Series and Algebraic Combinatorics
Séries Formelles et Combinatoire Algébrique
San Diego, California 2006

On the chromatic symmetric function of a tree

Jeremy L. Martin and Jennifer D. Wagner

Abstract. Stanley defined the *chromatic symmetric function* $X(G)$ of a graph G as a sum of monomial symmetric functions corresponding to proper colorings of G, and asked whether a tree is determined up to isomorphism by its chromatic symmetric function. We approach Stanley's question by asking what invariants of a tree T can be recovered from its chromatic symmetric function $X(T)$. We prove that the *degree sequence* $(\delta_1, \dots)$, where δ_j is the number of vertices of T of degree j, and the *path sequence* $(\pi_1, \dots)$, where π_k is the number of k-edge paths in T, are given by explicit linear combinations of the coefficients of $X(T)$. These results are consistent with an affirmative answer to Stanley's question. We briefly present some applications of these results to classifying certain special classes of trees by their chromatic symmetric functions.

Résumé. Stanley a défini la *fonction symétrique chromatique* $X(G)$ d'un graphe G par une somme de fonctions symétriques monomials qui correspondent aux colorations propres de G, et il a demandé si un arbre est déterminé jusqu'à l'isomorphisme par sa fonction symétrique chromatique. Nous approchons la question de Stanley en demandant quels invariants d'un arbre T peut être récupéré de sa fonction symétrique chromatique $X(T)$. Nous prouvons que le *suite des degrés* $(\delta_1, \dots)$, où δ_j est le nombre des sommets de T de degré j, et le *suite des chemins* $(\pi_1, \dots)$, où π_k est le nombre de chemins de longueur k, sont données par des combinaisons lineaires explicites des coefficients $X(T)$. Ces résultats sont conformés à une réponse affirmative à la question de Stanley. Nous présentons brièvement quelques applications de ces résultats à classifier certaines classes spéciales des arbres par ses fonctions symétriques chromatiques.

Introduction

Let G be a simple graph with vertices $V(G)$ and edges $E(G)$, and let $n = \#V(G)$ (the *order* of G). We assume familiarity with standard facts about graphs and trees, as set forth in, e.g., [**11**, Chapters 1–2]. In particular, a *coloring* of G is a function $\kappa : V(G) \to \{1, 2, \dots\}$ such that $\kappa(v) \neq \kappa(w)$ whenever the vertices v, w are adjacent. Stanley [**7**] defined the *chromatic symmetric function* of G as

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

the sum over all proper colorings κ, where $x_1, x_2, \dots$ are countably infinitely many commuting indeterminates. Note that $X(G)$ is homogeneous of degree n, and is invariant under permuting the x_i, so that $X(G)$ is a symmetric function. Moreover, the usual chromatic function $\chi(G; k)$, the number of colorings of G using at most k colors [**11**, Chapter 5], may be obtained from $X(G)$ by setting

$$x_i = \begin{cases} 1 & \text{for } i \leq k, \\ 0 & \text{for } i > k. \end{cases}$$

Our work is an attempt to resolve the following question.

Question (Stanley [**7**]): Is $X(G)$ a complete isomorphism invariant for trees? That is, must two nonisomorphic trees have different chromatic symmetric functions?

2000 *Mathematics Subject Classification.* Primary 05C05,05C60; Secondary 05E05.
Key words and phrases. tree, chromatic symmetric function, degree sequence, path sequence.

The answer to the question is "no" for arbitrary graphs; Stanley [**7**] exhibited two nonisomorphic graphs G, G' on 5 vertices such that $X(G) = X(G')$. For trees, however, the problem remains open. We note that Gebhard and Sagan [**2**] studied a chromatic symmetric function in *noncommuting* variables $x_1, x_2, \dots$; this is easily seen to be a complete invariant of G. On the other hand, it is well-known (and elementary) that the chromatic function $\chi(G; k)$ is the same, namely $k(k-1)^{n-1}$, for all trees G on n vertices. Thus Stanley's question asks where $X(G)$ falls between these two extremes. Li-Yang Tan [**10**] has verified computationally[1] that $X(T)$ determines T up to isomorphism for all trees T of order ≤ 23.

Stanley showed that when $X(G)$ is expanded in the basis of power-sum symmetric functions p_λ (indexed by partitions λ), the coefficients c_λ enumerate the edge-selected subgraphs of G by the sizes of their components (see equations (1), (2) (3) below). With the additional assumption that G is a tree, this expansion is a powerful tool with which to recover the structure of G from $X(G)$. The first steps in this direction are due to Matthew Morin, who studied the chromatic symmetric functions of caterpillars (trees in which deleting all the leaves yields a path) in [**4, 5**].

We now summarize our results.

The *degree* $\deg_T(v)$ of a vertex v in a graph T is the number of edges having v as an endpoint, and the *degree sequence* of G is $(\delta_1, \delta_2, \dots)$, where δ_j is the number of vertices having degree k. Our first main result is that the numbers δ_j are given by explicit linear combinations of the power-sum coefficients $c_\lambda(T)$.

Theorem 1. *For every tree T, we have $\delta_1(T) = c_{n-1}(T)$, and for all $j \geq 2$,*

$$\delta_j(T) \;=\; \sum_{\lambda \vdash n} \left(\ell(\tilde{\lambda}) \sum_{k \geq j} (-1)^{j+k-1} \binom{k}{j} \binom{\ell(\lambda)-1}{k+\ell-n} \right) c_\lambda(T).$$

It is easier to compute directly the number s_k of subgraphs of T that are *k-edge stars*, or trees with one central vertex and k leaves. It is easily seen that the sequences $(s_1, s_2, \dots)$ and $(\delta_1, \delta_2, \dots)$ are linearly equivalent.

The *distance* between two vertices of T is the number of edges in the unique path connecting them. The *path sequence* of G is $(\pi_1, \pi_2, \dots)$, where π_k is the number of vertex pairs at distance k, or equivalently the number of k-edge paths occurring as subgraphs of G. Our second main result, Theorem 2, asserts that the numbers π_k are again given by certain linear combinations of the coefficients $c_\lambda(T)$, as follows.

Theorem 2. *For every tree T, we have $\pi_1(T) = c_2(T)$ and $\pi_2(T) = c_3(T)$, and for all $k \geq 3$,*

$$\pi_k(T) \;=\; \sum_{\lambda \vdash n} \left((-1)^{n+k+1-\ell(\lambda)} \binom{\ell(\lambda)-1}{k-n+\ell(\lambda)} m(\lambda) \right) c_\lambda(T),$$

where

$$m(\lambda) = \binom{n-\ell(\lambda)}{2} - \sum_{i=1}^{s} \binom{\lambda_i - 1}{2}.$$

To prove each of these theorems, we interpret the desired linear combination of the coefficients of $X(T)$ as generating functions for certain subgraphs of G, using Stanley's characterization of those coefficients. We then show that these labeled subgraphs admit a sign-reversing involution. The ensuing cancellation permits us to recognize the surviving terms as enumerating either stars or paths in G, as appropriate.

This extended abstract is organized as follows. Section 1 contains the elements of the theory of chromatic symmetric functions, as developed by Stanley in [**7**]. Sections 2 and 3 contain sketches of the proofs of the degree and path sequence theorems, respectively.

The final section contains some brief remarks about other isomorphism invariants that can be extracted from $X(T)$, and about some special classes of trees that can be distinguished up to isomorphism by their path and/or degree sequences (hence by their chromatic symmetric functions).

More details and applications will be found in a future paper written jointly by the present authors and Matthew Morin.

[1]In an earlier version of this extended abstract, it was mentioned that the present authors have checked this for trees of order ≤ 14, using the database of trees available online at `http://www.zis.agh.edu.pl/trees/`, generated by Piec, Malarz, and Kulakowski as described in [**6**]. Evidently Tan's result is a substantial improvement!

1. Basic properties of $X(G)$

We begin by reviewing some of the theory of chromatic symmetric functions developed by Stanley in [**7**]. A *partition* is a sequence of positive integers $\lambda = (\lambda_1, \dots, \lambda_s)$ with $\lambda_1 \geq \cdots \geq \lambda_s > 0$; the number $s = \ell(\lambda)$ is the *length* of λ. The corresponding *power-sum symmetric function* $p_\lambda = p_\lambda(x_1, x_2, \dots)$ is defined by $p_\lambda = p_{\lambda_1} p_{\lambda_2} \cdots p_{\lambda_s}$, where $p_k = x_1^k + x_2^k + \cdots$.

One can obtain a family of useful invariants of G by expanding $X(G)$ in terms of the power sum symmetric functions p_λ. For each $S \subseteq E$, let $\lambda(S)$ be the partition of n whose parts are the orders of the components of the edge-induced subgraph $G|_S = (V, S)$. Stanley [**7**, Theorem 2.5] proved that

$$X(G) = \sum_{S \subset E} (-1)^{\#S} p_{\lambda(S)}. \tag{1}$$

In particular, the number of components of $G|_S$ is $\ell(\lambda(S))$. When $G = T$ is a tree, every subgraph S is a forest, so $\ell(\lambda(S)) = n - \#S$. Therefore, we may rewrite (1) as

$$X(T) = \sum_{\lambda \vdash n} (-1)^{n - \ell(\lambda)} c_\lambda p_\lambda, \tag{2}$$

where

$$c_\lambda = c_\lambda(T) = \#\{S \subset E \ : \ \lambda(S) = \lambda\}. \tag{3}$$

The coefficients $c_\lambda(T)$ are concrete combinatorial invariants of T that can be extracted from the chromatic symmetric function $X(T)$. Note that the c_λ are themselves not independent. For instance, it is immediate from (3) that

$$\sum_{\lambda:\ \ell(\lambda) = k} c_\lambda = \binom{n-1}{k}, \tag{4}$$

and there are several invariants of T that can be expressed in more than one distinct way in terms of the c_λ.

For notational simplicity, we shall often omit the parentheses and singleton parts when giving the index of one of these coefficients; for example, we abbreviate $c_{(h,1,1,\dots,1)}$ by c_h. (This raises the question of how we are going to denote the partition $\lambda = (1, 1, \dots, 1) = 1^n$. In fact, we won't need to do so, because 1^n is the only partition of length n, so (4) implies that $c_{1^n}(G) = 1$ for all G.)

For future reference, we list some properties of graphs and trees that can easily be read off its chromatic symmetric function. Several of these facts have already been noted by Morin [**4, 5**], and all of them are easy to deduce from (1) or (for trees) (2) and (3).

Proposition 3. *Let $G = (V, E)$ be a graph of order $n = \#V$.*

(i) *The number of vertices of G is the degree of $X(G)$.*
(ii) *The number of edges of G is c_2.*
(iii) *The number of components of G is $\min\{\ell(\lambda) \mid c_\lambda(G) \neq 0\}$.*
(iv) *If T is a tree, then the number of subtrees of T with k vertices is $c_k(T)$.*
(v) *If T is a tree, then the number of leaves (vertices of degree 1) in G is $c_{n-1}(T)$.*

Recall that a graph G is a tree if and only if it is connected and $\#E(G) = \#V(G) - 1$. Therefore, by (i), (ii) and (iii) of Proposition 3, the trees can be distinguished from other graphs by their chromatic symmetric functions. Moreover, part (v) implies that paths (trees with exactly two leaves) and stars (trees with exactly one nonleaf) are determined up to isomorphism by their chromatic symmetric functions.

2. The degree sequence

Let T be a tree with n vertices (and hence $n - 1$ edges). Recall that the degree $\deg_T(v)$ of a vertex $v \in V(T)$ is defined as the number of edges having v as an endpoint; a vertex of degree one is called a *leaf* of T.

Definition 4. The *degree sequence* of T is $(\delta_1, \delta_2, \dots, \delta_{n-1})$, where

$$\delta_j = \delta_j(T) = \#\{v \in V(T) \ : \ \deg_T(v) = j\}.$$

Notice that $\delta_j = 0$ whenever $j < 1$ or $j \geq n$. Moreover, it is a standard fact that $\sum \delta_j = 2n-2$.

For $k \geq 1$, let S_k be the tree with vertices $\{0,1,\dots,k\}$ in which 0 is adjacent to every other vertex. Any graph that is isomorphic to S_k is called a *k-star*. If $k \geq 2$, then every k-star has a unique non-leaf vertex, called its *center*.

Definition 5. The *star sequence* of T is defined to be $(s_1, s_2, \dots, s_{n-1})$, where

$$s_k = s_k(T) = \#\{U \subset T \;:\; U \equiv S_k\}.$$

Notice that $s_1 = n-1$ (the number of edges of T), and that $s_k = 0$ whenever $k < 1$ or $k \geq n$.

Knowing the degree sequence of T is equivalent to knowing the number of substars of T of each possible order; it is straightforward to show that

$$s_k = \sum_{j \geq k} \binom{j}{k} \delta_j \tag{5a}$$

and that

$$\delta_j = \sum_{k \geq j} \binom{k}{j} (-1)^{j+k} s_k. \tag{5b}$$

It is more straightforward to recover the star sequence from the power-sum coefficients c_λ than it is to recover the degree sequence directly. For $\lambda \vdash n$, define $\ell(\lambda)$ to be the number of parts of λ, and let $\tilde{\lambda}$ be the partition obtained by deleting all the singleton parts of λ.

Theorem 6. *Let T be a tree with n vertices, and let $2 \leq k < n$. Then*

$$s_k(T) \;=\; -\sum_{\lambda \vdash n} \ell(\tilde{\lambda}) \binom{\ell(\lambda)-1}{k+\ell(\lambda)-n} c_\lambda(T).$$

We sketch the proof, omitting many of the calculations and technical details. First, we obtain by straightforward calculation the identity

$$\sum_{\lambda \vdash n} \ell(\tilde{\lambda}) \binom{\ell(\lambda)-1}{k+\ell(\lambda)-n} c_\lambda(T) \;=\; \sum_{\substack{F \subseteq T:\\ \#F = k}} \; \sum_{G \subseteq F} \; \sum_{\substack{\text{nontrivial}\\ \text{components}\\ C \text{ of } G}} (-1)^{\#G}. \tag{6}$$

For each subforest $F \subset T$, denote by $\Sigma(F)$ the summand indexed by F on the right-hand side of (6). The second step in the proof is to analyze $\Sigma(F)$. When F is a star, it is not hard to see that this summand reduces to

$$\left(\sum_{G \subseteq F} (-1)^{\#G} \right) - (-1)^{\#\emptyset} \;=\; -1.$$

Now, suppose that F is not a star; we wish to show that $\Sigma(F) = 0$. The expression $\Sigma(F)$ may be regarded as counting ordered pairs (G, C), where $G \subset F$ is a subforest and C is a nontrivial component of G, assigning to each such pair the weight $(-1)^{\#G}$. We construct an involution ψ on the set of such pairs (G, C). Whenever (G, C) and (G', C') are paired by ψ, we have $\#G' = \#G \pm 1$; in particular, the summands in $\Sigma(F)$ corresponding to (G, C) and (G', C') cancel. We conclude that $\Sigma(F) = 0$ as desired.

Theorem 6 now follows immediately from (6) together with the calculation of $\Sigma(F)$. The degree sequence formula, Theorem 1, follows in turn from Theorem 6 together with (5b).

3. The path sequence

Let $T = (V, E)$ be a tree with $\#V = n$. For any two vertices $v, w \in V$, their *distance* $d(v, w) = d_T(v, w)$ is defined as the number of edges in the unique path joining v and w. Define

$$\pi_k(T) \;:=\; \#\left\{\{v, w\} \subseteq V \;:\; d(v, w) = k\right\}.$$

Equivalently, $\pi_k(T)$ is the number of paths with exactly k edges that occur as subgraphs of T. It is easy to see that $\pi_k(T) = 0$ if $k \leq 0$ or $k \geq n$, and that $\sum_k \pi_k(T) = \binom{n}{2}$. Moreover, we have $\pi_1(T) = \#E = n-1$ and $\pi_2(T) = s_2(T)$ (because a two-edge path is identical to a two-edge star). As we already know, these quantities can be recovered from $X(T)$.

Suppose that $k \geq 3$. We now recall Theorem 2, which describes the path numbers $\pi_k(T)$ as linear combinations of the coefficients of $X(T)$. For a partition $\lambda = (\lambda_1, \dots, \lambda_\ell) \vdash n$, define

$$m(\lambda) = \binom{n-\ell(\lambda)}{2} - \sum_{i=1}^{s} \binom{\lambda_i - 1}{2}. \tag{7}$$

THEOREM 2. For every tree T, we have $\pi_1(T) = c_2(T)$ and $\pi_2(T) = c_3(T)$, and for all $k \geq 3$,

$$\pi_k(T) \;=\; \sum_{\lambda \vdash n} (-1)^{n+k+1-\ell(\lambda)} \binom{\ell(\lambda)-1}{k-n+\ell(\lambda)} m(\lambda) c_\lambda(T).$$

Again, we give just a sketch of the proof. Using Stanley's interpretation for $c_\lambda(T)$, we can rewrite the right-hand side of the desired equality as

$$\sum_{A \subseteq E} (-1)^{k+1+\#A} \binom{n-\#A-1}{k-\#A} m(\lambda(A)).$$

We interpret the binomial coefficient $\binom{n-a-1}{k-a}$ as counting the subsets of $E - A$ of cardinality $k - a$, and interpret $m(\lambda(A))$ as the number of pairs of distinct edges $e, f \in A$ that belong to different components of the induced subgraph (V, A); call such a pair of edges A-*okay*. Thus the last expression becomes

$$(-1)^{k+1} \sum_{A \subseteq E} \; \sum_{\substack{B \subseteq E-A \\ \#B = k-\#A}} \; \sum_{A\text{-okay pairs } e,f} (-1)^{\#A}.$$

For $e, f \in E$, let $P = P(e, f)$ be the unique shortest path between an endpoint of e and an endpoint of f. Then e, f is an A-okay pair if and only if $e, f \in A$ and $A \not\supseteq P$. In particular, e, f have no common endpoint (we abbreviate this condition as $e \cap f = \emptyset$), and $P \neq \emptyset$. Changing the order of summation and letting $A' = A - e - f$ and $C = A' \cup B$, we can rewrite the last expression as

$$(-1)^{k+1} \sum_{e \cap f = \emptyset} \; \sum_{\substack{C \subseteq E-e-f \\ \#C = k-2}} \left(\sum_{\substack{A' \subseteq C \\ A' \not\supseteq P(e,f)}} (-1)^{\#A'} \right). \tag{8}$$

If we remove the condition $A' \not\supseteq P(e, f)$ from the last summation, then the parenthesized expression becomes zero (since $\#C = k - 2 > 0$). Therefore (8) can be rewritten as

$$(-1)^{k} \sum_{e \cap f = \emptyset} \; \sum_{\substack{C \subseteq E-e-f \\ \#C = k-2}} \; \sum_{\substack{A': \\ P(e,f) \subseteq A \subseteq C}} (-1)^{\#A'}. \tag{9}$$

The last sum is zero unless $C = P(e, f)$. So (9) collapses to

$$(-1)^k \sum_{e \cap f = \emptyset} \chi\Big[\#P(e,f) = k-2\Big](-1)^k \;=\; \sum_{e \cap f = \emptyset} \chi\Big[\#(e \cup f \cup P(e,f)) = k\Big] \;=\; \pi_k(T),$$

(where χ is the "Garsia chi": $\chi[S] = 1$ if the sentence S is true, or 0 if S is false). This completes the proof of Theorem 2.

4. Further remarks

4.1. Other invariants recoverable from $X(T)$. Theorems 1 and 2 imply that any isomorphism invariant of a tree T that can be derived from its path and degree sequences can be recovered from $X(T)$. Examples of such invariants include the *diameter* (the number of edges in a longest path) and the *Wiener index* (the quantity $\sigma(T) = \sum_{v,w} d(v, w)$, where v, w range over all pairs of vertices of T. The Wiener index can be obtained from the chromatic symmetric function in other ways. For example, when $X(T)$ is expanded as a sum of *elementary* symmetric functions, Stanley has interpreted the coefficients as counting sinks in acyclic orientations; this observation gives rise to a different expression for $\sigma(T)$. Note that the Wiener index is far from distinguishing trees up to isomorphism; see [**1**, §13].

One might ask whether the methods of Theorems 1 and 2 can be used to count other kinds of subtrees of a tree T (that is, other than stars and paths) by appropriate linear combinations of the coefficients of

$X(T)$. Such a class may be quite subtle; our empirical computations seem to rule out, for instance, spiders and double-stars (i.e., caterpillars with two branch vertices).

4.2. Spiders. Let T be a tree. A vertex $v \in V(T)$ is called a *branch vertex* if $\deg_T(v) \geq 3$. A *spider* (or *starlike tree*) is a tree with exactly one branch vertex (to avoid trivialities, we do not consider paths to be spiders). Since the definition of a spider relies only on the degree sequence, Theorem 1 implies that membership in the class of spiders can be deduced from $X(T)$. In fact, much more is true: one can show that every spider is determined up to isomorphism by its chromatic symmetric function.

We sketch the proof briefly. A spider may be regarded as a collection of edge-disjoint paths (the *legs*) joined at a common endpoint t (the *torso*). The torso is the unique branch vertex, and the lengths of the legs determine the spider up to isomorphism. That is, the isomorphism classes of spiders with n edges correspond to the partitions $\mu \vdash n$ with $\ell(\mu) \geq 3$. The partition μ can then be recovered from the coefficients $c_\lambda(T)$, where $\lambda \vdash n$ has exactly two parts. For example, when no single leg of the spider contains as many as half the edges, the sequence

$$(c_{1,n-1},\ c_{2,n-2},\ \ldots);$$

is a partition whose conjugate is precisely μ. (The case of a spider with one "giant leg" is only slightly more complicated.)

4.3. Caterpillars. A *caterpillar* is a tree such that deleting all the leaves yields a path (called the *spine* of the caterpillar). It is not hard to see that this is equivalent to the condition that the diameter of T is one more than the number of nonleaf vertices; therefore, whether or not T is a caterpillar can be deduced from $X(T)$. When T is a *symmetric* caterpillar (i.e., it has an automorphism reversing the spine), it is determined up to isomorphism by $X(T)$. This fact was proved by Morin[**4**], and can also be recovered from Theorem 2. However, the corresponding statement for arbitrary caterpillars remains unknown. Gordon and McDonnell [**3**] showed that there exist arbitrarily large families of nonisomorphic caterpillars with the same path and degree sequences; however, we suspect that the additional information furnished by the chromatic symmetric function of a caterpillar T will be enough to reconstruct it up to isomorphism.

Acknowledgements

Chromatic symmetric functions came to the first author's attention thanks to Matthew Morin's talk at the Graduate Student Combinatorics Conference, held at the University of Minnesota on April 16 and 17, 2005. The authors are grateful to Matthew Morin, Victor Reiner, Bruce Sagan, and Stephanie van Willigenburg for numerous helpful discussions and suggestions.

References

[1] Andrey A. Dobrynin, Roger Entringer, and Ivan Gutman, Wiener index of trees: theory and applications, *Acta Appl. Math.* **66** (2001), no. 3, 211–249.
[2] David D. Gebhard and Bruce E. Sagan, A chromatic symmetric function in noncommuting variables, *J. Alg. Combin.* **2** (2001), 227–255.
[3] Gary Gordon and Eleanor McDonnell, Trees with the same degree sequence and path number, *Discrete Math.* **147** (1995), 297–300.
[4] Matthew Morin, Caterpillars, ribbons, and the chromatic symmetric function, M.S. thesis, University of British Columbia, 2005.
[5] Matthew Morin, The chromatic symmetric function of symmetric caterpillars, preprint, 2005.
[6] S. Piec, K. Malarz, and K. Kulakowski, How to count trees?, preprint, arXiv:cond-mat/0501594.
[7] Richard P. Stanley, A symmetric function generalization of the chromatic polynomial of a graph, *Adv. Math.* **111** (1995), 166–194.
[8] Richard P. Stanley, *Enumerative Combinatorics*, Volume 2. Cambridge Univ. Press, Cambridge, 1997.
[9] Richard P. Stanley, Graph colorings and related symmetric functions: ideas and applications, *Discrete Math.* **193** (1998), 267–286.
[10] Li-Yang Tan, personal communication. Details of computation to be posted at `http://www.cs.wustl.edu/~lt1`.
[11] Douglas B. West, *Introduction to Graph Theory*, 2nd edn. Prentice Hall, Upper Saddle River, NJ, 2001.

Department of Mathematics, University of Kansas, Lawrence, KS 66045
E-mail address: `jmartin@math.ku.edu`
URL: `http://www.math.ku/edu/~jmartin`

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