A FEW MORE TREES THE SYMMETRIC CHROMATIC FUNCTION CAN DISTINGUISH

JAKE HURYN

Abstract. A well-known open problem in graph theory concerns whether the *symmetric chromatic function* introduced by Stanly, a generalization of the chromatic polynomial, distinguishes between any two non-isomorphic trees[1]. Previous work has proven the conjecture for a class of trees called *spiders*[2]. I generalize the class of spiders to *n-spiders*, where normal spiders correspond to $n = 1$, and prove the conjecture for $n = 2$.

1. The Symmetric Chromatic Function and its Coefficients

Let $G = (V, E)$ be a graph. Let $K(F)$ denote the set of connected components of the graph (V, F) for $F \subseteq E$. Then the symmetric chromatic function of a graph G can be defined as

$$\mathbf{X}_G = \sum_{F \subseteq E} \left((-1)^{|F|} \prod_{K \in K(F)} p_{|V(K)|} \right).$$

A *tree* is a connected acyclic graph. Every edge of a tree is a *bridge*, that is, its removal disconnects the graph. A *leaf* is a vertex of degree one.

Let $T = (V, E)$ be a tree. I discuss subsets $F \subseteq E$ by how many edges are removed from E, which is one less than the number of connected components of (V, F) and thus factors in each product $\prod p_{|V(K)|}$. Moreover, each $F \subseteq E$ defines a partition $\lambda = (\lambda_1, \ldots, \lambda_\ell)$ of $|V|$, whose parts are the $|V(K)|$ (written here in increasing order), since the connected components of (V, F) partition V itself. Given such a partition λ, denote the absolute value of the coefficient on $p_{\lambda_1} \cdots p_{\lambda_\ell}$ by $c_\lambda(T) = c_{\lambda_1, \ldots, \lambda_{\ell-1}}(T)$. I will write the subscripts in decreasing order, omitting the largest part, and often the tree being operated on.

We thus come to our first combinatorial interpretation for c_λ, which represents the number of ways we can remove $\ell - 1$ edges from T (an $(\ell-1)$-cut of T) to partition T, by the spanning subgraph, into connected components of order $\lambda_1, \ldots, \lambda_\ell$ (Figure 1). In particular, c_1 is the number of leaves of T.

Of importance in the following results is the use of *rooted trees*, that is, trees with one vertex designated as the root. Let each rooted tree isomorphism class be

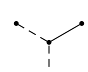

Figure 1. A 2-cut corresponding to the product $p_1^2 p_2$ and contributing to the coefficient $c_{1,1}$. Dashed lines represent deletions. Note that there are three different ways to make a 2-cut of this tree resulting in the same partition, so the tree's $\mathbf{X}_T$ has a $-3p_1^2 p_2$ term.

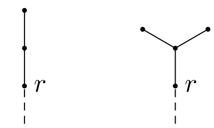

FIGURE 2. Rooted subtrees $R_{3,1}$ and $R_{4,2}$, respectively, with the root labeled. The dashed lines are the unique connection to the rest of the tree.

named $R_{n,i}$ (giving the same name to any rooted tree in this class), where n is the order of the tree and i is an arbitrary indexing of the rooted trees of order n. In particular, let $R_{n,1}$ denote a path with the root at a leaf, and let $R_{n+1,2}$ be the same, with the addition of a single vertex appended to the second outward-most vertex from the root.

I also define $c_\lambda(R)$ for R a rooted tree. Given a partition λ, define $c_\lambda(R)$ as the ways to cut R into connected components, such that the orders of those not containing the root correspond to the parts of λ (no parts of λ being omitted from the subscript, as the connected component containing the root is already omitted from λ). For example, the root of R may also be a leaf, but it will not contribute to $c_1(R)$. We also have $c_1(R_{n,1}) = 1$ and $c_1(R_{n,2}) = 2$.

In practice, these rooted trees will be used as subtrees of a tree T, where the root is the only vertex of the subtree connected to the rest of T, and is connected in that way by a single edge (Figure 2). Equivalently, a rooted subtree is a subtree which may be disconnected from the tree by removing a single edge. Let the number of rooted subtrees of T isomorphic to $R_{n,i}$ be denoted $\rho_{n,i}(T)$, or simply $\rho_{n,i}$.

Let $|V(T)| = d$, and let $r(n)$ be the number of rooted tree isomorphism classes having order n. It follows from our definitions that we have the equations

$$c_n = \sum_{i=1}^{r(n)} \rho_{n,i} = \sum_{i=1}^{r(d-n)} \rho_{d-n,i}$$

if $n \neq \frac{d}{2}$, and

$$c_{\frac{d}{2}} = \frac{1}{2} \sum_{i=1}^{r(\frac{d}{2})} \rho_{\frac{d}{2},i}.$$

The following result is a similar but more complicated equation for $c_{n,1}$. Note first that

$$c_{1,1} = \binom{c_1}{2} + c_2,$$

but this does not give us any more information about T.

Proposition 1. *Let $T = (V, E)$ be a tree of order d. Let $T_{n,i}$ be the tree obtained from $R_{n,i}$ by removing the distinction of the root. If $2 \leq n < \frac{d-1}{2}$, then*

$$c_{n,1} = \sum_{i=1}^{r(n)} (c_1 - c_1(R_{n,i}))\rho_{n,i} + \sum_{j=1}^{r(n+1)} c_1(T_{n+1,j})\rho_{n+1,j}. \tag{1}$$

Proof. The coefficient $c_{n,1}$ tells us the number of ways we can cut T in two places to get a connected component of order one and another of order n, and of course a

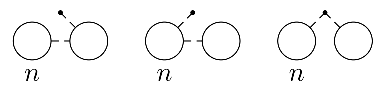

Figure 3. The three ways of joining the connected components of the partition contributing to $c_{n,1}$. The first picture corresponds to the first term of (1). The labeled components have order n.

third of order $d-n-1$. There are three ways of joining these components with two edges as illustrated by Figure 3.

The derivation of (1) should then be clear from the figure. In particular, observe that the second picture requires that the isolated vertex not originally be the root of the rooted subtree of order $n+1$, while the third requires exactly the opposite. Thus, both terms combined negate the distinction of the root. □

Finally, let $d \geq 6$. From Figure 4 it is easy to see that

$$c_{1,1,1,1} = \binom{c_1}{4} + c_2\binom{c_1-1}{2} + \binom{c_2}{2} + (c_1-1)\rho_{3,1} + (c_1-2)\rho_{3,2} + c_4. \tag{2}$$

Thus the coefficients $c_{1,1,1,1}$ and c_3 form a system of equations which allows us to solve for $\rho_{3,1}$ and $\rho_{3,2}$. We require $d \geq 6$ so that the circle drawn in Figure 4 is distinguished from a vertex.

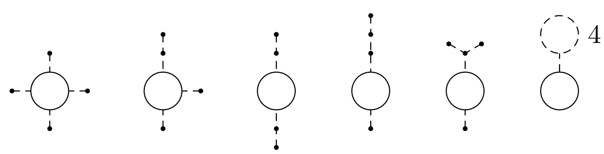

Figure 4. The ways of joining the connected components of the partition contributing to $c_{1,1,1,1}$. The pictures correspond to the respective terms in (2).

2. Distinguishing Spiders and their Generalizations

A spider is a tree with exactly one vertex of degree at least three. That vertex is called the *torso* of the spider, and the other vertices form paths extending from the torso called *legs*. It is known that $\mathbf{X}_G$ completely distinguishes spiders.

Call a tree a *2-spider* if it is a modification of a spider in which any leg may be appended with a single vertex joined to the second outward-most vertex of that leg, with respect to the torso. Call such modified legs the *2-legs*. The reasoning behind these names is that one can think of these modified legs which are copies of the rooted subtree $R_{n,2}$, while normal legs are $R_{n,1}$. We may thus call normal spiders *1-spiders* and normal legs *1-legs*.

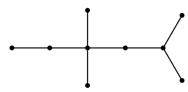

Figure 5. A 2-spider with one 2-leg, satisfying (ii) of Theorem 2, described by $(\lambda, \mu) = (1, 1, 2; 4)$.

We may describe these 2-spiders up to isomorphism by a pair of positive integer sequences $(\lambda, \mu) = (\lambda_1, \dots, \lambda_\ell; \mu_1, \dots, \mu_m)$, where λ lists the orders of the 1-legs and μ the orders of the 2-legs.

Theorem 2. *Let T be a 2-spider of order d. Then one can uniquely reconstruct T from $\mathbf{X}_T$.*

Proof. The smallest 2-spider which is not also a 1-spider has order six, so assume $d \geq 6$.

We would like to determine the number λ_n^* of 1-legs with order at least n for $1 \leq n \leq \frac{d}{2}$, and the number μ_n^* of 2-legs with order at least n for $3 \leq n \leq \frac{d}{2}$. If we consider λ and μ to be partitions, then λ^* and μ^* are their respective conjugate partitions, so with this information we can easily reconstruct T.

The proof is separated based on the following cases:

- **i:** T has only three legs, two of order one.
- **ii:** If a leg of T has length k, then $k \leq \frac{d}{2}$.

First, suppose T satisfies (i). We find that either $\rho_{3,2} = 1$ and $c_1 = 3$, or $\rho_{3,2} = 2$ and $c_2 = 4$. This is sufficient to distinguish this case, and we also have $|\mu| = \rho_{3,2} - 1$. From this it is easy to reconstruct T since we know d.

Now suppose T satisfies (ii) but not (i). Suppose we cut an edge of some leg L to split T into two rooted subtrees. The subtree containing the torso will not be of the form $R_{n,1}$ or $R_{n,2}$ for any n (note that this is false if T satisfies (i)). However, this second subtree will be of the form $R_{n,1}$ if L is a 1-leg or $R_{n,2}$ if L is a 2-leg, and L must have order at least n. Thus $\lambda_n^* = \rho_{n,1}$ and $\mu_n^* = \rho_{n,2}$. This does not quite hold, however, if $n = 1$, since $\rho_{1,1}$ also counts both leaves on each 2-leg. To correct for this, we use

$$\lambda_1^* = \rho_{1,1} - \frac{1}{2}\rho_{3,2}.$$

We know $\rho_{1,1} = c_1$ and $\rho_{2,1} = c_2$. Let $4 \leq n < \frac{d}{2}$, so it follows from the length restriction that if $i \neq 1, 2$, $\rho_{n,i} = 0$. Then

$$c_n = \rho_{n,1} + \rho_{n,2},$$

and by Proposition 1,

$$c_{n-1,1} = (c_1 - 1)\rho_{n-1,1} + (c_1 - 2)\rho_{n-1,2} + 2\rho_{n,1} + 3\rho_{n,2}.$$

Thus we may solve inductively for all such $\rho_{n,1}$ and $\rho_{n,2}$ since we know $\rho_{3,1}$ and $\rho_{3,2}$.

Now suppose $n = \frac{d}{2}$; there can only be one leg of such order. From (1) we can easily tell if there is a leg of order n by examining $c_{n-1,1}$, since if there is not, the second sum in the equation would be zero. Note then that exactly one of $\rho_{n-1,1}$ and $\rho_{n-1,2}$ can be nonzero, since a 2-spider cannot have a leg of order $\frac{d}{2} - 1$ if it also has one of order $\frac{d}{2}$. Thus we know the type of that largest leg.

Note that in case (ii), $c_i \leq c_j$ if $3 \leq i < j < \frac{d}{2}$.

Finally, suppose T satisfies none of the cases. Observe that the smallest such 2-spider has order 9, so suppose $d \geq 9$. Let the length of the (unique) largest leg have order $k > \frac{d}{2}$. A quick calculation verifies that we cannot also have a leg of order $d-k-1$ or larger, and $3 \leq d-k-1 < \frac{d}{2}-1$. Thus $c_{d-k-1} = 1$, but $c_{d-k} = 2$. This distinguishes from the second case. Moreover, we only need to find λ_n^* and μ_n^* up to $n = d-k-1$, which will give the types of all legs, and their respective lengths, excluding the length of the longest leg. To do this, we simply solve for $\rho_{3,1}$ and $\rho_{3,2}$ and do the same process as in the second case up to $c_{d-k-2,1}$. Then we can find the length of the longest leg, thus reconstructing T, by looking at d. □

Remark. It turns out that solving Stanley's conjecture can be done simply by solving for all $\rho_{n,i}$, for $n \leq \frac{d}{2}$, in terms of the coefficients of $\mathbf{X}_T$, as we did with c_3 and $c_{1,1,1,1}$ for $n = 3$. While there are some interesting heuristics for determining which c_λ can give what coefficients on the desired $\rho_{n,i}$ in its expansion, even for $n = 4$ this is difficult to do.

Even worse, if we were attempting to distinguish trees of order 20, we would need to solve for $\rho_{10,i}$, of which there would be more than coefficients of $\mathbf{X}_T$! What this implies then is that there must be some "hidden" information about the values $\rho_{n,i}$ can take implicit in the definition of a tree that cannot be found in $\mathbf{X}_T$.

References

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

[2] J. L. Martin, M. Morin, and J. D. Wagner. "On distinguishing trees by their chromatic symmetric functions". In: *ArXiv Mathematics e-prints* (2006). eprint: `math/0609339`.

The Ohio State University.
E-mail address: `huryn.5@osu.edu`