DISCRETE MATHEMATICS 6 (1973) 359–365. © North-Holland Publishing Company

THE NUMBER OF CATERPILLARS *

Frank HARARY ** and Allen J. SCHWENK **

Department of Mathematics, University of Michigan, Ann Arbor, Mich. USA

Received 3 January 1973

Abstract. A caterpillar is a tree which metamorphoses into a path when its cocoon of endpoints is removed. The number of nonisomorphic caterpillars with $n + 4$ points is $2^n + 2^{[n/2]}$. This neat formula is proved in two ways: first, as a special case of an application of Pólya's enumeration theorem which counts graphs with integer-weighted points; secondly, by an appropriate labeling of the lines of the caterpillar.

1. Weighted graphs

A *graph with weighted points* is obtained from a given graph G by assigning a positive integer n_i to each point v_i of G. If H is such a graph with weighted points obtained from G, then H may be regarded as an ordinary (unweighted) graph by taking each point v_i of G and adding $(n_i - 1)$ new points and $(n_i - 1)$ new lines joining them to v_i. This process is shown in Fig. 1.

The next unary operation on graphs apparently was introduced in [5, p. 63] for trees, but can be defined for an arbitrary connected graph G with at least three points. The *derivative* G' of G is obtained on deleting all the endpoints of G. Thus the underlying graph G of a graph with weighted points is the same as the derivative of its associated ordinary graph (illustrated in Fig. 1) if and only if the weight at each endpoint of G is at least 2, and if G happens to be K_1, the weight at its point is at least 3.

We need to define some concepts involving graphs and groups (see [1, Chapter 14]). The *group of a graph* G, written $\Gamma(G)$, is the permutation group consisting of all its automorphisms, each being a permu-

* Research supported in part by a grant from the Air Force Office of Scientific Research.

** Present address: Mathematical Institute, University of Oxford, 24–29 St. Giles, Oxford, England.

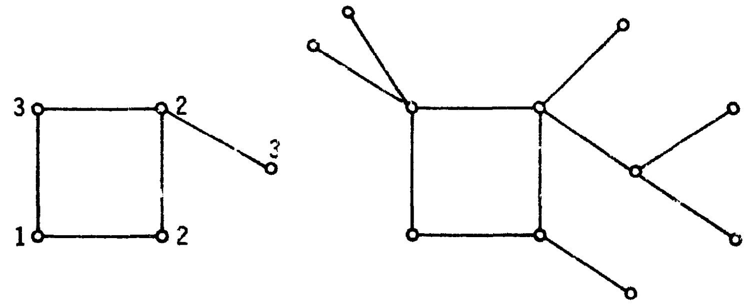

Fig. 1. A graph with weighted points, and its associated ordinary graph.

tation on the set of points of G. The *cycle index* of a permutation group A in which each permutation α has $j_k(\alpha)$ cycles of length k is the following polynomial in d variables y_i, where d is the degree of A:

$$(1) \qquad Z(A) = \frac{1}{|A|} \sum_{\alpha \in A} \prod_{k=1}^{d} y_k^{j_k(\alpha)} .$$

The symbol $Z(A; f(x))$ is the expression obtained from (1) when each variable y_k is replaced by $f(x^k)$.

By an application of Pólya's classical enumeration theorem (see [2, Chapter 2]), it follows that the coefficient of x^r in $Z(\Gamma(G); 1+x)$ gives the number of inequivalent r-subsets of points of G with respect to the group of G. This result can be viewed as the number of weighted graphs of weight r with underlying graph G where each point has weight 0 or 1. Once we realize this, it is easy to see how to generalize this result by substituting a "bigger" series into $Z(\Gamma(G))$.

Theorem 1.1. *The number of weighted graphs with underlying graph G whose points all have positive weight is given by*

$$(2) \qquad \omega(x) = Z(\Gamma(G);\ x/(1-x)) .$$

Proof. Since $x/(1-x) = x + x^2 + x^3 + \ldots$, this is the appropriate series to provide possible point weights of 1, 2, 3, ... Thus we may apply Pólya's theorem with $x/(1-x)$ as the figure counting series and $\Gamma(G)$ as the configuration group to obtain equation (2).

Corollary 1.2. *Let G be a connected graph with $a_i(G)$ points of degree i for $i = 0$ and 1. Let $h(x) = \sum_{i=3}^{\infty} h_i x^i$, where h_i is the number of connected* *graphs H with i points such that $H' = G$. Then*

(3) $$h(x) = x^{2a_0 + a_1} Z(\Gamma(G); x/(1-x)) .$$

Proof. As we observed above, the number of weighted graphs with G as the underlying graph is the same as the number of graphs H having $G = H'$ if and only if the weight at each endpoint of G is at least 2, but if G is K_1, then the weight at its point is at least 3. Thus if we subtract 1 from the weight of each endpoint, and if $G = K_1$, we subtract 2 from the weight of its point, then we obtain a weighted graph with arbitrary positive weights. We now apply Theorem 1.1 and include a factor of $x^{2a_0+a_1}$ to correct for the weights we have altered, and obtain equation (3).

2. Caterpillars

We may now use this result to count a certain species of trees which we studied in [3, 4], called *caterpillars* by A. Hobbs. These are the trees T with at least three points whose derived tree T' is a path. Let $c(x) = \Sigma_{p=3}^{\infty} c_p x^p$ be the generating function for the number of caterpillars.

Theorem 2.1. *The number of caterpillars on $p \geq 3$ points is given by*

(4) $$c(x) = \frac{x^3(1-3x^2)}{(1-2x)(1-2x^2)} ,$$

or equivalently, writing $p = n + 4$,

(5) $$c_{n+4} = 2^n + 2^{[n/2]} .$$

Proof. We wish to count those p-point caterpillars whose derived tree is the path P_k for each positive k. We observe that when $k = 1$, we have $a_0 = 1$ and $a_1 = 0$, and when $k > 1$, we see that $a_0 = 0$ and $a_1 = 2$. Thus, in either case, we have $x^{2a_0+a_1} = x^2$, so applying Corollary 1.2, we have

(6) $$c(x) = \sum_{k=1}^{\infty} x^2 Z(\Gamma(P_k); x/(1-x)) .$$

We split this sum into even and odd terms and write

(7) $$c(x) = x^2 \sum_{i=1}^{\infty} Z\left(\Gamma(P_{2i});\ \frac{x}{1-x}\right) + x^2 \sum_{i=0}^{\infty} Z\left(\Gamma(P_{2i+1});\ \frac{x}{1-x}\right) .$$

Now it follows from the definition of the cycle index that

(8) $$Z(\Gamma(P_{2i})) = Z(S_2[E_i]) = \tfrac{1}{2}(y_1^{2i} + y_2^i) ,$$

(9) $$Z(\Gamma(P_{2i+1})) = Z(E_1 \cdot S_2[E_i]) = \tfrac{1}{2} y_1 (y_1^{2i} + y_2^i) .$$

We proceed to substitute $x/(1-x)$ into (8) and (9) and substitute the resulting expression into (7) to obtain

(10) $$c(x) = x^2 \sum_{i=1}^{\infty} \tfrac{1}{2}\left(\left(\frac{x}{(1-x)}\right)^{2i} + \left(\frac{x^2}{(1-x^2)}\right)^{i}\right) + x^2 \sum_{i=0}^{\infty} \frac{x}{2(1-x)}\left(\left(\frac{x}{(1-x)}\right)^{2i} + \left(\frac{x^2}{(1-x^2)}\right)^{i}\right) .$$

This simplifies routinely to equation (4).

Now to verify equation (5), we use partial fractions to express this as

(11) $$c(x) = \frac{Ax^3}{1-2x} + \frac{Bx^3 + Cx^4}{1-2x^2} .$$

This leads to the system of simultaneous equations $A + B = 1$, $-2B + C = 0$ and $-2A - 2C = -3$, which we solve to find $A = B = \frac{1}{2}$ and $C = 1$, so that

(12) $$c(x) = \frac{x^3/2}{1-2x} + \frac{x^3/2 + x^4}{1-2x^2} .$$

From (12), it is easy to see that the coefficient of x^p is given by $c_p = c_{n+4} = 2^n + 2^{[n/2]}$, proving equation (5).

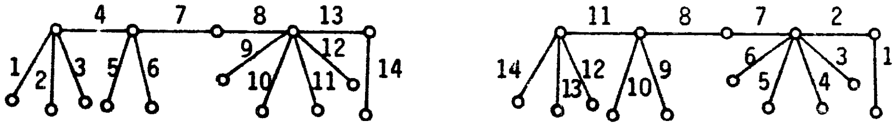

Fig. 2. The two labelings of a caterpillar.

3. Alternate proof

The simplicity of this final result suggests that there may be a more natural and direct approach to obtain this formula, as occurs so often in combinatorial mathematics, and in fact we have found one.

We start by attempting to define a unique labeling for the lines of each caterpillar T. Let x_1 be an endline of a longest path in T. We proceed to label the remaining lines inductively. Assume i lines have been labeled $x_1, ..., x_i$ and assign x_{i+1} to an endline adjacent to x_i which has not yet been labeled, provided one exists, but if there is no such endline, then by the structure of a caterpillar, there remains only one unlabeled line incident with x_i, and it lies on the path T'. We label it x_{i+1}. Fig. 2 displays two possible labelings of a particular caterpillar.

In fact, since at each step in the labeling procedure all the unlabeled endlines adjacent to a given line are equivalent, it is clear that the only way we obtain different labelings of a caterpillar is by choosing as x_1 an endline incident with a different non-endpoint. Thus there are at most two labelings of a caterpillar, and these arise by starting at opposite ends of a longest path. Note, however, that these two labelings are identical if the caterpillar has an automorphism α interchanging the endpoints of a longest path. These caterpillars which have just one labeling we call *symmetric*.

We now examine the number of "caterpillar labelings" which can occur on p points. Lines x_1 and x_2 can be adjacent in just one way, but for each of the $p - 3$ remaining lines, there are two ways for x_{i+1} to be incident with x_i, namely, at either endpoint of x_i. Thus we have 2^{p-3} labelings.

Now most caterpillars are counted twice among these labelings, but the symmetric ones are just counted once. In order to correct for this, we must determine the number of symmetric caterpillars. If $p = 2k + 1$ is odd, then the automorphism α interchanging the ends of a longest path must fix the central point v_0 of T. Since T has an even number of lines, v_0 must have even degree, so we may use half of the lines in T to

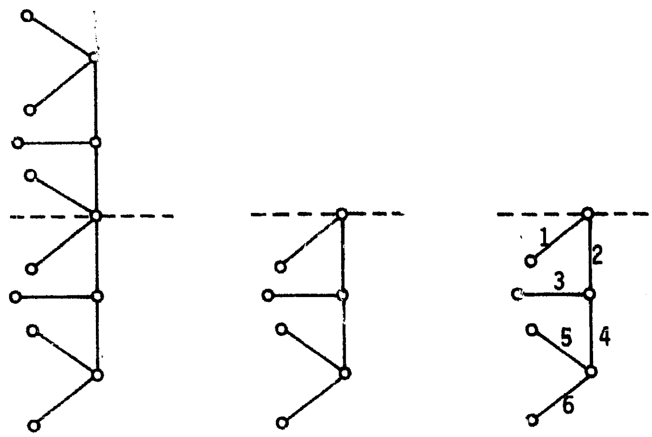

Fig. 3. A 13-point symmetric tree, its subtree, and the labeled subtree.

induce a subtree whose structure determines that of the entire tree. We label this subtree which is also a caterpillar by the procedure given above. This process is depicted in Fig. 3.

Now there are 2^{k-1} possible labeled subtrees since each line after the first has two ways to be adjacent to its predecessor. Thus there are $2^{k-1} = 2^{[(p-2)/2]}$ symmetric caterpillars on $p = 2k + 1$ points.

We find a similar result when $p = 2k$ is even, except that two cases must be considered. If T' has an odd number of points, the central point v_0 must have odd degree, and we obtain a subtree to label as depicted in Fig. 4(a). There are 2^{k-2} such labelings. Finally, if T' has an even num-

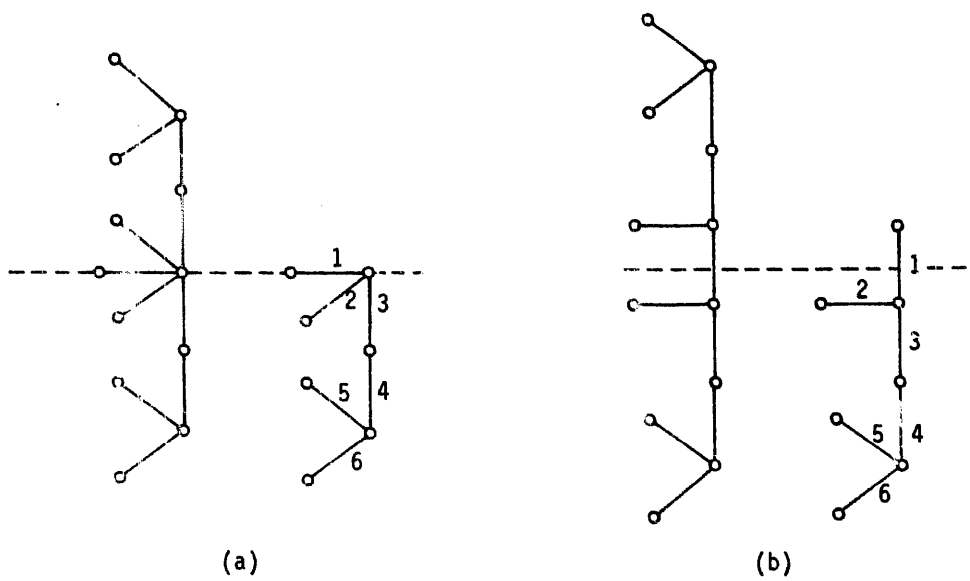

Fig. 4. Two 12-point symmetric caterpillars formed from the same labeled subtree.

ber of points, then T has a "symmetry line" (see [1, p. 189]) and again we obtain a subtree to label (Fig. 4(b)). Here too there are 2^{k-2} labelings, so we add these two to get $2^{k-1} = 2^{[(p-2)/2]}$, which happens to be the same number as obtained in the case p odd. Thus in either case, we may correct our original count by adding $2^{[(p-2)/2]}$ and dividing by 2 to get equation (5).

Observe that in Fig. 4, the same labeled subtree has been used in two different ways to produce two distinct symmetric caterpillars. In one of these, T' has even length, and in the other, it has odd length. In this way, 2^{k-2} labeled subtrees yield 2^{k-1} symmetric caterpillars.

References

[1] F. Harary, Graph theory (Addison–Wesley, Reading, Mass., 1969).
[2] F. Harary and E.M. Palmer, Graphical enumeration (Academic Press, New York, 1973).
[3] F. Harary and A.J. Schwenk, Trees with hamiltonian square, Mathematika 18 (1971) 138–140.
[4] F. Harary and A.J. Schwenk, A new crossing number for bipartite graphs, Utilitas Math. 1 (1972) 203–209.
[5] D. König, Theorie der endlichen und unendlichen Graphen (Teubner, Leipzig, 1936; Chelsea, New York, 1950).