ADVANCES IN MATHEMATICS 29, 60–88 (1978)

Distance Matrix Polynomials of Trees

R. L. GRAHAM

Bell Laboratories, Murray Hill, New Jersey

AND

L. LOVÁSZ

JATE Bolyai Intézet, Szeged, Hungary

Let G be a finite connected graph. If x and y are vertices of G, one may define a distance function d_G on G by letting $d_G(x, y)$ be the minimal length of any path between x and y in G (with $d_G(x, x) = 0$). Thus, for example, $d_G(x, y) = 1$ if and only if $\{x, y\}$ is an edge of G. Furthermore, we define the distance matrix $D(G)$ for G to be the square matrix with rows and columns indexed by the vertex set of G which has $d_G(x, y)$ as its (x, y) entry. In this paper we are concerned with properties of $D(G)$ for the case in which G is a tree (i.e., G is acyclic). In particular, we precisely determine the coefficients of the characteristic polynomial of $D(G)$. This determination is made by deriving surprisingly simple expressions for these coefficients as certain fixed linear combinations of the numbers of various subgraphs of G.

INTRODUCTION

If G is a finite connected graph,[1] one can define a metric d_G on G as follows:

> For vertices x, y of G, $d_G(x, y)$ is taken to be the length of the shortest path in G between x and y (where $d_G(x, x)$ is defined to be 0).

Thus, for example, $d_G(x, y) = 1$ if and only if $\{x, y\}$ is an edge of G.

In this paper we are concerned with the *distance matrix* $D(G)$ of G. This is formed by indexing the rows and columns with the vertex set of G and defining the (x, y) entry of $D(G)$ to be $d_G(x, y)$. Specifically, we investigate in detail the characteristic polynomial $\Delta_G(\lambda) = \det(D(G) - \lambda I)$ of $D(G)$ in the special case, G is a tree (i.e., G is connected and acyclic). There are several reasons why it is of interest to do this.

[1] We generally follow the terminology of [9]. In particular, G has no loops or multiple edges.

0001-8408/78/0291–0060$05.00/0

In the first place, $\Delta_G(\lambda)$ arose (rather unexpectedly) from a data communication problem studied by one of the authors [6] in the following way. Suppose one wishes to label each vertex v of G with an N-tuple $A(v) = (a_1, a_2, \ldots, a_N)$, where $a_i \in \{0, 1, *\}$, so that A "preserves distances." Precisely, this means that if we define the "distance"[2] $d(A(v), A(v'))$ between $A(v) = (a_1, \ldots, a_N)$ and $A(v') = (a'_1, \ldots, a'_N)$ by

$$d(A(v), A(v')) = |\{i: \{a_i, a'_i\} = \{0, 1\}\}|$$

then we require

$$d(A(v), A(v')) = d_G(v, v') \qquad \text{for all} \quad v, v' \in G. \tag{1}$$

For example, the labeling

$$\begin{aligned} A(x) &= (0, 0), \\ A(y) &= (0, 1), \\ A(z) &= (1, *), \end{aligned}$$

satisfies (1) when x, y, and z are the vertices of a triangle. It is not hard to show that such a labeling is always possible for any connected graph G provided N is chosen large enough. The problem is to determine the minimum $N = N(G)$ for which a labeling satisfying (1) exists. In [6] the following theorem is proved.

THEOREM.

$$N(G) \geqslant \max\{n_+(G), n_-(G)\}, \tag{2}$$

where $n_+(G)$ and $n_-(G)$ denote the number of positive and negative eigenvalues,[3] *respectively, of $D(G)$.*

In fact, equality in (2) seems to occur quite often, and the smallest known case of inequality is for the complete bipartite graph $K_{2,3}$.

Thus, the eigenvalues of $D(G)$, i.e., the roots of $\Delta_G(\lambda)$, are intimately connected with the embeddability of G into "squashed cubes" (the use of $*$ indicates some coordinate identification; see [7] for details). However, not much is known about $\Delta_G(\lambda)$ at present. For example, it is not even known which graphs G have $n_-(G) = |G| - 1$ or whether there is a graph for which $n_+(G) > n_-(G)$. It should be remarked that (2) provides the only known proof that the complete graph on n vertices cannot be decomposed into fewer than $n - 1$ edge-disjoint complete bipartite subgraphs (see [7]).

[2] d is not actually a metric.

[3] Since $D(G)$ is real and symmetric, it has real eigenvalues. Also, since the trace of $D(G)$ is zero, $n_+(G)$ and $n_-(G)$ are bounded above by $|G| - 1$, where $|G|$ denotes the number of vertices of G.

In the other direction, it is suspected[4] that the following inequality holds.

Conjecture.

$$N(G) \leqslant |G| - 1. \tag{3}$$

It is known that (3) holds with equality for complete graphs, odd cycles, trees, and most complete bipartite graphs (these have $n_-(G) = |G| - 1$) [6], and, in any case, (3) holds for all cycles, complete bipartite graphs, graphs with $d_G \leqslant 2$, and large classes of graphs with "spines" [1], [2].

In the second place, it is known [6] that for a *tree* T with n vertices, the determinant of $D(T)$ is always equal to $(-1)^{n-1}(n-1)\,2^{n-2}$, *independent of the structure of* T. Since $\det(D(T))$ is just the constant term $\delta_0(T)$ of the polynomial

$$\Delta_T(\lambda) = \sum_{k=0}^{n} \delta_k(T)\,\lambda^k,$$

it was natural to inquire about the other coefficients of $\Delta_T(\lambda)$ as well. It was hoped that by analogy with the corresponding results for the adjacency matrix of T, the $\delta_k(T)$ would have a natural interpretation in terms of the structure of certain subgraphs of T. What we mean by this is the following.

The adjacency matrix $A(G)$ of G is defined by setting the (x, y) entry of $A(G)$ equal to 1 if $\{x, y\}$ is an edge of G and 0 otherwise. Let us write the characteristic polynomial of $A(G)$ as

$$A_T(\lambda) = \det(A(G) - \lambda I) = \sum_{k=0}^{n} \alpha_k(G)\,\lambda^k.$$

It is well known (see [11–13]) that for a tree T, the $\alpha_k(T)$ depend only on the numbers of occurrences of subgraphs of T consisting of disjoint edges. Specifically, if $N_H(G)$ denotes the number of occurrences[5] of a graph H in a graph G, i.e., the number of subgraphs of G which are isomorphic to H, and mP_1 denotes the graph consisting of m disjoint paths of length 1, then

$$\begin{aligned} \alpha_{n-k}(T) &= (-1)^{n+k/2} N_{(k/2)P_1}(T) \qquad &&\text{if } k \text{ is even,} \\ &= 0 &&\text{otherwise.} \end{aligned} \tag{4}$$

In [5] it was shown that at least for $0 \leqslant k \leqslant 3$, $\delta_k(T)$ can also be expressed in a form similar to (4). For example, we have already stated

$$\delta_0(T) = (-1)^{n-1}(n-1)\,2^{n-2} = (-1)^{n-1}2^{n-2}N_{P_1}(T). \tag{5}$$

[4] In fact, one of the authors is currently offering U.S. $100 for the first proof or counterexample.

[5] By convention, if H is the empty graph, we take $N_H(G) = 1$.

It is also true that

$$\delta_1(T) = (-1)^{n-1}2^{n-3}(4N_{2P_1}(T) + 2N_{P_2}(T) + 4N_{P_1}(T) - 4), \tag{6}$$

where P_2 denotes the path of length 2, i.e., the unique tree with three vertices. However, with each new value of k, new and increasingly complicated arguments were required in [5] to obtain the expansion for $\delta_k(T)$.

The authors of [5] conjectured that similar expressions existed for all k, although they admitted that an attack on this problem by the same techniques seemed hopeless.

In this paper we settle the conjecture by showing that *all* the coefficients $\delta_k(T)$ can be expressed in the form

$$\delta_k(T) = (-1)^{n-1}2^{n-k-2} \sum_F A_F^{(k)} N_F(T), \tag{7}$$

valid for all trees T where $n = |T|$, the number of vertices of T, and F ranges over all acyclic graphs (i.e., *forests*) having $k-1$, k, or $k+1$ edges and no isolated vertices. Furthermore, the integer coefficients $A_F^{(k)}$ in (7) are unique and rather well behaved. We give explicit and surprisingly simple expressions for them (see Facts 5, 6, and 7). They turn out to depend only on the number of occurrences of various P_i in the connected components of F.

Still another motivating force for this study was to attack the conjecture [1] that T is uniquely determined by $\Delta_T(\lambda)$, in sharp contrast to the situation for $A_T(\lambda)$. Because of the simplicity of the expression for $\alpha_k(T)$ in (4), it is not surprising that the $\alpha_k(T)$ do not uniquely determine T. The smallest example [7] of two nonisomorphic trees having the same "spectrum," i.e., set of adjacency matrix eigenvalues, is shown in Fig. 1. For these trees,

$$A_{T_1}(\lambda) = A_{T_2}(\lambda) = \lambda^8 - 7\lambda^6 + 9\lambda^4.$$

Intuitively, since the entries of $D(T)$ are generally much larger than those of $A(T)$, the coefficients of $\Delta_T(\lambda)$ tend to be much larger than those of $A_T(\lambda)$ and consequently, $\Delta_T(\lambda)$ has a better chance of distinguishing nonisomorphic trees. However, this turns out not to be the case (see the discussion at the end of the paper.

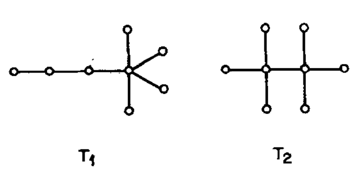


FIGURE 1

PRELIMINARIES

We consider an arbitrary fixed tree T with vertex set $V = \{v_1, ..., v_n\}$, where $n > 1$. We abbreviate $d_T(v_i, v_j)$ by d_{ij} and we let d_i denote the degree of v_i, i.e., the number of edges of T incident to v_i.

Basically, the plan for determining the coefficients $\delta_i(T)$ consists of three steps:

I. The inverse D^{-1} of $D = D(T)$ is found.

II. The individual terms of the determinant expansion for the characteristic polynomial $\Delta^*_T(\lambda)$ of D^{-1} are interpreted as enumerating the occurrences of certain *marked* subforests in T.

III. The contribution each forest makes in II is determined.

Because of the simple relationship between $\Delta^*_T(\lambda)$ and $\Delta_T(\lambda)$, the expansion of $\delta_k(T)$ into the form of (7) is immediate.

FORMATION OF D^{-1}

Let us define the $n \times n$ matrix B by

$$B = \begin{pmatrix} 0 & 1 & 1 & 1 & \cdots & 1 \\ 1 & -2 & & & & \\ 1 & & -2 & & 0 & \\ 1 & & & -2 & & \\ \vdots & & 0 & & \ddots & \\ 1 & & & & & -2 \end{pmatrix}.$$

Let $N = (n_{ij})$ be the $n \times n$ matrix defined by

$$\begin{aligned} n_{ij} &= 1 \qquad \text{if} \quad d_{1j} = d_{1i} + d_{ij}, \\ &= 0 \qquad \text{otherwise.} \end{aligned}$$

Thus, $n_{ij} = 1$ if and only if the unique path from v_1 to v_j contains v_i.

FACT 1. $D = N^T B N$.

Proof. First note that

$$BN = \begin{pmatrix} d_{11} & d_{12} & \cdots & d_{1n} \\ \multicolumn{4}{c}{\boxed{\quad 1 - 2n_{ij} \quad}} \end{pmatrix} = (c_{ij}).$$

Writing $N^TBN = (c'_{ij})$, we obtain

$$\begin{aligned} c'_{ij} &= \sum_{k=1}^{n} n_{ki}c_{kj} \\ &= d_{1j} + \sum_{k=2}^{n} n_{ki}(1 - 2n_{kj}). \end{aligned} \tag{8}$$

It now follows at once from the definition of n_{ij} that $c'_{ij} = d_{ij}$ and the fact is proved. ▮

FACT 2. *The inverse of N is given by $N^{-1} = (n^*_{ij})$, where*

$$\begin{aligned} n^*_{ij} &= 1 && \textit{if} \quad i = j, \\ &= -1 && \textit{if} \quad d_{ij} = 1 \quad \textit{and} \quad d_{1j} = 1 + d_{1i}, \\ &= 0 && \textit{otherwise.} \end{aligned}$$

Proof. Note that $n^*_{ij} = 1$ if and only if $n_{ij} = 1$ and i is adjacent to j. Since the (i, j) entry in the product NN^* is $\sum_{k=1}^{n} n_{ik}n^*_{kj}$, the only terms in the sum which have a nonzero contribution come from those k with both $n_{ik} \neq 0$ and $n^*_{kj} \neq 0$. If $i = j$ then we must have $k = i$ and the entry is 1. If $i \neq j$ then the only nonzero terms are for $k = j$ with $n_{ij}n^*_{jj} = 1 \cdot 1 = 1$ and $k = k'$, where $d_{k'j} = 1$ and $d_{1j} = 1 + d_{1k'}$ with $n_{ik'}n^*_{k'j} = 1(-1) = -1$. Thus, for $i \neq j$ the entry is 0. Hence $NN^* = I$ and the fact is proved. ▮

FACT 3.

$$N^*N^{*T} = \begin{pmatrix} d_1 + 1 & & & 0 \\ & d_2 & & \\ & & \ddots & \\ 0 & & & d_n \end{pmatrix} - A(T),$$

where $A(T) = (a_{ij})$ is the adjacency matrix of T.

Proof. The (i, j) entry in the product, namely,

$$\sum_{k=1}^{n} n^*_{ik}n^*_{jk},$$

has the values

(i) $d_1 + 1$ if $i = j = 1$, since all k with either $k = 1$ or v_k adjacent to v_1 contribute 1 to sum;

(ii) d_i if $i = j > 1$, since now we cannot take $k = 1$;

(iii) $-a_{ij}$ if $i \neq j$, for we cannot have $n^*_{ik} = n^*_{jk} \neq 0$.

Thus, the only nonzero contribution can occur if $n^*_{ik} = -n^*_{jk}$, v_i is adjacent to v_j, and so $a_{ij} = 1$. This proves the fact. ▮

Since the inverse of B is given by

$$B^{-1} = \frac{1}{2(n-1)} \begin{pmatrix} 4 & 2 & 2 & \cdots & 2 \\ 2 & -(n-2) & & 1 & \\ 2 & & -(n-2) & & \\ \vdots & & 1 & \ddots & \\ 2 & & & & -(n-2) \end{pmatrix}, \tag{9}$$

a straightforward calculation of $D^{-1} = (N^T B N)^{-1} = N^* B^{-1} N^{*T}$ using the preceding facts proves the following result.

LEMMA 1. *$D^{-1} = (d^*_{ij})$ is given by*

$$d^*_{ij} = \frac{(2-d_i)(2-d_j)}{2(n-1)} + \begin{cases} -d_i/2 & \text{if } i = j, \\ a_{ij}/2 & \text{if } i \neq j. \end{cases}$$

At this point the reader may wonder if any progress has been made, since the structure of D^{-1} is apparently not much nicer than that of D. However, it should be kept in mind that it is not D^{-1} we are primarily interested in but rather its characteristic polynomial

$$\Delta_T^*(\lambda) = \det(D^{-1} - \lambda I) = \sum_{k=0}^{n} \delta_k^*(T)\, \lambda^k.$$

Let $D'(\lambda)$ denote the $(n+1) \times (n+1)$ matrix defined by

$$D'(\lambda) = \left(\begin{array}{c|cccc} -(n-1) & 2-d_1 & \cdots & & 2-d_n \\ \hline 2-d_1 & -2\lambda - d_1 & & a_{ij} & \\ \vdots & & \ddots & & \\ & a_{ij} & & \ddots & \\ 2-d_n & & & & -2\lambda - d_n \end{array}\right). \tag{10}$$

By performing elementary row and column transformations on $D'(\lambda)$, the following result is readily obtained.

FACT 4.

$$\det D'(\lambda) = -(n-1)\, 2^n \Delta^*_T(\lambda). \tag{11}$$

Let us write

$$\det D'(\lambda) = \sum_{k=0}^{n} \delta_k'(T)\, \lambda^k.$$

Then from elementary linear algebra we have

$$\begin{aligned} \Delta_T(\lambda) &= (-\lambda)^n (\det D)\, \Delta_T{}^*(\lambda^{-1}) \\ &= -(n-1)\, 2^{n-2} \lambda^n \Delta_T{}^*(\lambda^{-1}) \qquad \text{by (6).} \end{aligned} \tag{12}$$

Hence, by (10) *and* (11),

$$\begin{aligned} \delta_k(T) &= -(n-1)\, 2^{n-2} \delta^*_{n-k}(T) \\ &= \tfrac{1}{4}\delta'_{n-k}(T). \end{aligned} \tag{13}$$

Since for large n most of the entries in $D'(\lambda)$ are 0, *it is reasonable to expect that the calculation of the $\delta_k(T)$ by means of* (13) *will be simpler than a direct expansion of* $\det(D - \lambda I)$. *If we expand* $\det D'(\lambda)$ *and collect the terms which contribute to the λ^{n-k} term, we find*

$$\delta'_{n-k}(T) = (-2)^{n-k} \sum_{i_1,\ldots,i_k} \det \begin{pmatrix} -(n-1) & 2-d_{i_1} & \cdots & 2-d_{i_k} \\ 2-d_{i_1} & -d_{i_1} & & a_{ij} \\ \vdots & & \ddots & \\ 2-d_{i_k} & a_{ij} & & -d_{i_k} \end{pmatrix}, \tag{14}$$

where the sum ranges over all choices of $1 \leqslant i_1 < \cdots < i_k \leqslant n$.

Marked Subforests of T

The next step is to interpret the individual terms in the expansion of the determinants in (14) as enumerating certain subforests F of T in which the vertices and edges of F have been marked in various ways. Before giving the general construction, we illustrate the ideas with a simple example.

Example. $k = 1$. From (14) we have

$$\delta'_{n-1}(T) = (-2)^{n-1} \sum_i \det \begin{pmatrix} -(n-1) & 2-d_i \\ 2-d_i & -d_i \end{pmatrix}, \tag{15}$$

where i ranges over $1 \leqslant i \leqslant n$. Since

$$\det \begin{pmatrix} -(n-1) & 2-d_i \\ 2-d_i & -d_i \end{pmatrix} = (n-1)\, d_i - (d_i - 2)(d_i - 2),$$

we can rewrite (15) as

$$\delta'_{n-1}(T) = (-2)^{n-1} \left\{ \sum_i (n-1)\, d_i - \sum_i d_i(d_i - 1) + 3 \sum_i d_i - \sum_i 4 \right\}. \quad (15')$$

Now, the term $(n-1)\, d_i$ can be thought of as counting the number of ways in which we can select an edge incident to v_i together with an arbitrary edge e (possibly the *same one* we chose incident to v_i). Since we sum over all i, the sum $\sum_i (n-1)\, d_i$ just represents the number of ways of selecting an edge incident to *some* v_i together with an arbitrary edge e.

Consider a fixed subforest $2P_1$ of T consisting of two disjoint edges shown in Fig. 2a. We claim it is counted four times in the sum $\sum_i (n-1)\, d_i$, for it contributes to the terms which correspond to choosing:

(i) an edge from v_{j_1} and $e = \{v_{j_3}, v_{j_4}\}$,
(ii) an edge from v_{j_2} and $e = \{v_{j_3}, v_{j_4}\}$,
(iii) an edge from v_{j_3} and $e = \{v_{i_1}, v_{i_2}\}$,
(iv) an edge from v_{j_4} and $e = \{v_{i_1}, v_{i_2}\}$.

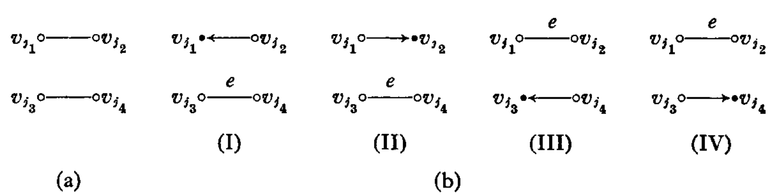


FIGURE 2.

We can indicate this by "marking" the edges of $2P_1$ as shown in Fig. 2b. The arrow on an edge indicates the vertex "responsible" for that edge. On the other hand a subforest P_1 of T consisting of a single edge $\{v_{j_1}, v_{j_2}\}$ is counted *twice* in the sum, corresponding to the choices $i = j_1$, $e = \{v_{j_1}, v_{j_2}\}$ and $i = j_2$, $e = \{v_{j_1}, v_{j_2}\}$ (see Figs. 3a and 3b).

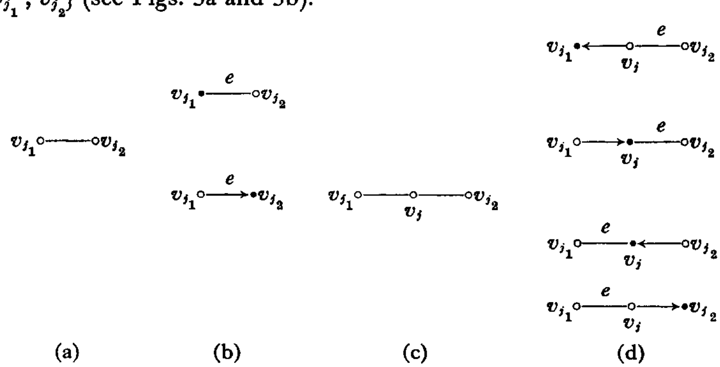


FIGURE 3.

Finally, a subforest P_2, a path of length 2, is counted four times, as shown in Figs. 3c and d. Thus, we must have

$$\sum_i (n-1)\, d_i = 4N_{2P_1}(T) + 2N_{P_1}(T) + 4N_{P_2}(T). \tag{16}$$

Next, we interpret $\sum_i d_i(d_i - 1)$ as an *ordered* choice of two distinct edges incident to v_i. The resulting subforest of T must be isomorphic to P_2. We show the corresponding marking convention in Fig. 4b. Hence,

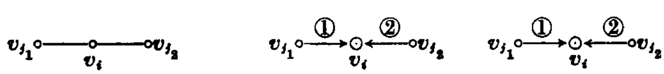
$$\sum_i d_i(d_i - 1) = 2N_{P_2}(T).$$

In a similar way, we obtain

$$3\sum_i d_i = 6N_{P_1}(T),$$

$$\sum_i 4 = 4n = 4N_{P_1}(T) + 4,$$

since T has $n-1$ edges.



(a) (b)

FIGURE 4.

Combining the preceding expressions, we obtain by (13)

$$\delta_1(T) = \tfrac{1}{4}\delta'_{n-1}(T) = (-1)^{n-1}2^{n-3}(4N_{2P_1}(T) + 2N_{P_2}(T) + 4N_{P_1}(T) - 4),$$

which is just (6).

Let us now examine the expansion of the general determinant in (14), namely,

$$\det \begin{bmatrix} -(n-1) & 2-d_{i_1} & 2-d_{i_2} & \cdots & 2-d_{i_k} \\ 2-d_{i_1} & -d_{i_1} & a_{i_1,i_2} & \cdots & a_{i_1,i_k} \\ 2-d_{i_2} & a_{i_2,i_1} & -d_{i_2} & & \vdots \\ \vdots & \vdots & & \ddots & \\ 2-d_{i_k} & a_{i_k,i_1} & \cdots & & -d_{i_k} \end{bmatrix}, \tag{17}$$

where we label the rows and columns with $\{0, 1,\ldots, k\}$. An important observation is that the only permutation choices from the above matrix which can contribute *nonzero* terms to the determinant correspond to permutations π with the cycle structure

$$\pi = (0 j_1 j_2 \cdots j_l)(j_{l+1} j_{l+2})(j_{l+3} j_{l+4}) \cdots$$
$$\cdots (j_{l+2m-1} j_{l+2m})(j_{l+2m+1}) \cdots (j_{l_k}). \tag{18}$$

This follows at once from the fact that since T contains *no cycles*, the only nontrivial cycles π can have either involve row 0 and column 0 (i.e., $(0 j_1 j_2 \cdots j_l)$) or have length 2 (i.e., $(j_s j_{s+1})$). Furthermore, all the terms $a_{j_1, j_2}, a_{j_2, j_3}, \ldots, a_{j_{l-1}, j_l}$ must be 1.

The permutation π in (18) corresponds to the term

$$(2 - d_{j_1})\, a_{j_1, j_2} \cdots a_{j_{l-1}, j_l} (2 - d_{j_l})(a_{j_{l+1}, j_{l+2}} a_{j_{l+2}, j_{l+1}})$$
$$\cdots (-d_{j_{l+2m+1}}) \cdots (-d_{j_{l_k}}) \tag{19}$$

in the expansion of (17). When $l = 0$, this has the slightly different form

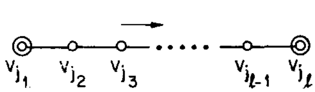
$$-(n-1)(a_{j_{l+1}, j_{l+2}} a_{j_{l+2}, j_{l+1}}) \cdots (-d_{j_{l+2m+1}}) \cdots (-d_{j_{l_k}}). \tag{19'}$$

We may expand the term in (19) into three similar terms formed by replacing $(2 - d_{j_1})(2 - d_{j_l})$ by

$$d_{j_1} d_{j_l} - 2(d_{j_1} + d_{j_l}) + 4.$$

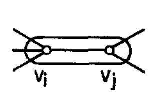
Next, we describe how various edges and vertices of T are to be marked in order to correspond to contributions from the terms (19) and (19').

(i) For the factor $a_{j_1, j_2} a_{j_2, j_3} \cdots a_{j_{l-1}, j_l}$ we distinguish the endpoints and the direction of the path in T from v_{j_1} to v_{j_l} (if $l \geq 2$) as follows:

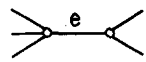


(ii) For the factor d_j, we mark an edge of T incident to v_j with an arrow pointing to the shaded vertex v_j:

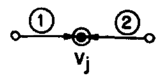
(iii) For the factor $(a_{i,j} a_{j,i})$ we distinguish the edge $\{v_i, v_j\}$ in T:



(iv) For the factor $(n - 1)$, we mark an edge with the symbol e:

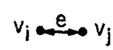


(v) For the factor $d_j d_j$ (which will occur only when $l = 1$), we mark one edge incident to v_j with an arrow to v_j and a symbol ① and we mark one edge (possibly the same edge) with an arrow to v_j and a symbol ②; also, we circle and shade v_j.

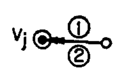


The terms (19) or (19′) now correspond exactly to the number of ways T can be marked according to rules just given. Of course, one must keep in mind the fact that degeneracies may occur; e.g., some edges of T may receive several marks.

For example, from (ii) and (iv) we may have



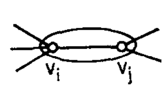
and from (v) we may have



The value of determinant (17) is now given by enumerating all possible ways of marking T according to (i) to (v) and summing the appropriate (signed) expressions over all choices of $1 \leqslant j_1 < \cdots < j_k \leqslant n$.

Of course, the terms of determinant (17) also have signs attached to them. Specifically, each term with the cycle structure of π in (18) has an additional sign factor of $(-1)^{l+m}$.

A considerable simplification now results from the following observation. For each marking of T which contains an edge marked by (iii), i.e.,



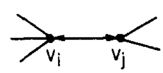
(because of a factor $(a_{i,j} a_{j,i})$), there is another marking of T which is identical except for the edge $\{v_i, v_j\}$, now (degenerately) marked by (ii) as

Furthermore, the corresponding terms in the expansion of (17) from which the two markings come have *opposite* signs. This is obvious, since the two permutations differ only in that the factor $(a_{i,j}a_{j,i})$ in one is replaced by $(-d_i)(-d_j)$ in the other and such a change certainly changes the sign of the permutation. Hence, all the contributions from the markings of the type in (iii) are canceled out by all the markings in which the edge $\{v_i, v_j\}$ has two arrows, one to v_i and one to v_j, and for which d_i and d_j have been selected from the diagonal of (17) (i.e., d_i and d_j do not come from the cycle $(0j_1 \cdots j_l) = (0i \cdots j)$ of a permutation).

Thus, we may henceforth restrict our consideration to permutations in (18) for which $m = 0$, i.e., those π of the form

$$\pi = (0j_1 \cdots j_l)(j_{l+1}) \cdots (j_k), \tag{20}$$

provided the edges $\{v_i, v_j\}$ in T marked as



come only from the factors $(2 - d_i)(2 - d_j)$ of a term. However, since for any permutation π in (20) there is at most one cycle $(0j_1 \cdots j_l)$ containing 0, in the corresponding markings of T, at most one edge can have arrows at each of its endpoints.

The specific terms of the determinant which come from the π in (20) are

I. $l = 0$: $-(n-1)(-d_{j_1}) \cdots (-d_{j_k})$, sign $\pi = 1$.

II. $l = 1$: $(2 - d_{j_1})(2 - d_{j_1})(-d_{j_2}) \cdots (-d_{j_k})$, sign $\pi = -1$.

We split this into the sum of the three terms:

(i) $d_{j_1}d_{j_1}(-d_{j_2}) \cdots (-d_{j_k})$,

(ii) $-4d_{j_1}(-d_{j_2}) \cdots (-d_{j_k})$,

(iii) $4(-d_{j_2}) \cdots (-d_{j_k})$.

III. $l \geqslant 2$: $(2 - d_{j_1})\, a_{j_1,j_2} \cdots a_{j_{l-1},j_l}(2 - d_{j_l})(-d_{j_{l+1}}) \cdots (-d_{j_k})$, sign $\pi = (-1)^l$.

We also split this into the sum of three terms:

(i) $(d_{j_1} - 1)\, a_{j_1,j_2} \cdots a_{j_{l-1},j_l}(d_{j_l} - 1)(-d_{j_{l+1}}) \cdots (-d_{j_k})$,

(ii) $-((d_{j_1} - 1) + (d_{j_l} - 1))\, a_{j_1,j_2} \cdots a_{j_{l-1},j_l}(-d_{j_{l+1}}) \cdots (-d_{j_k})$,

(iii) $a_{j_1,j_2} \cdots a_{j_{l-1},j_l}(-d_{j_{l+1}}) \cdots (-d_{j_k})$.

Our next task is to examine the number of ways a given subforest F of T can be marked so as to contribute to the nonzero terms in I and II. If F is a forest with connected components $C_1, \ldots, C_t$ (which of course are trees), we define

$$|F| = \text{the number of vertices of } F,$$

$$\|F\| = \text{the number of edges of } F,$$

$$\pi(F) = \prod_{i=1}^{t} |C_i|.$$

We let $\mathscr{F}_k$ denote the set of all forests having no isolated vertices and exactly k edges. For the empty forest F^*, we set $\pi(F^*) = 1$.

Since each factor d_j and $a_{i,j}$ corresponds to the marking of a unique edge of T, with the exception of d_{j_1} and d_{j_t} in II which may degenerate, it follows that we must have $\|F\| = k + 1$, k or $k - 1$. We remark that at this point, it follows readily that $\delta_k(T)$ can be expressed in the form (7). Furthermore, since the F have no isolated vertices, it follows from results of Szemerédi and one of the authors [8] that the coefficients $A_F^{(k)}$ will be ***unique***. These coefficients we now proceed to determine.

The Number of Markings of F

Let F be an arbitrary fixed subforest of T with components $C_1, \ldots, C_t$ and no isolated points. We wish to determine in how many ways T may be marked according to the conditions of the preceding section so that the marked edges are exactly the edges of F. Because of the restrictions on marking T, it follows that all C_i except possibly one, which we denote by C^*, have all edges marked according to (ii), i.e., as

We say that C_i is marked ***normally*** in this case. The number of ways C_i can be marked normally is just $|C_i|$, the number of vertices of C_i. This is because each vertex of such a C_i, except for exactly one vertex v, must have an incoming arrow. All other vertices have exactly one incoming arrow and all other incident edges with outgoing arrow. Thus, v serves as a "source" and all other arrows are determined (see Fig. 5). Hence, it suffices to determine the number of ways the exceptional component C^* can be marked.

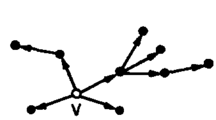


Fig. 5. A normally marked component.

As we have noted, F can have only $k+1$, k, or $k-1$ edges. We treat the three cases separately. We first interpret the absolute value of the terms and then determine the appropriate signs.

$\|F\| = k+1$

I. $(n-1)(d_{j_1}) \cdots (d_{j_k})$ (see Fig. 6). If the edge e corresponding to the choice for the factor $(n-1)$ is removed from C^*, exactly two of the resulting components can be arbitrarily marked normally, i.e., using (ii). The directions of the arrows on all other edges are forced. Thus, for each choice of "source" vertices x and y in C^*, there are $d_T(x, y) = d(x, y)$ possible locations of the edge e. Therefore, there are exactly

$$\sum_{\{x,y\}\subseteq C^*} d(x, y)$$

ways of marking C^* in this case. Since the sign of the permutation π in (20) is $+1$, the total contribution to the determinant is

$$\begin{aligned} (-1)^{k+1} \Big\{ &| C_2 | \cdots | C_t | \sum_{\{x,y\}\subseteq C_1} d(x, y) \\ &+ | C_1 | | C_3 | \cdots | C_t | \sum_{\{x,y\}\subseteq C_2} d(x, y) + \cdots \\ &+ | C_1 | \cdots | C_{t-1} | \sum_{\{x,y\}\subseteq C_t} d(x, y) \Big\} \\ = (-1)^{k+1}\pi(F) &\sum_{i=1}^{t} \frac{1}{| C_i |} \sum_{\{x,y\}\subseteq C_i} d(x, y). \qquad (21) \end{aligned}$$

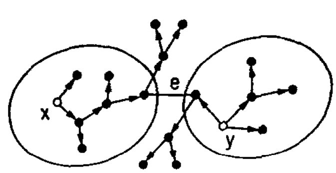


FIGURE 6

II(i). $d_{j_1} d_{j_1} (d_{j_2}) \cdots (d_{j_k})$ (see Fig. 7). Since $\|F\| = k+1$, there are no multiply marked edges. Using an argument similar to that in the preceding case (where an extra factor of 2 comes from the labeling of the edges with ① and ②), we obtain a total contribution in this case of

$$(-1)^k \pi(F) \sum_{i=1}^{t} \frac{1}{| C_i |} \sum_{\{x,y\}\subseteq C_i} (d(x, y) - 1). \qquad (22)$$

The summand here is $d(x, y) - 1$ instead of $d(x, y)$, since we have a choice of a *vertex* on the path between the sources x and y instead of the *edge* we had in the preceding case.

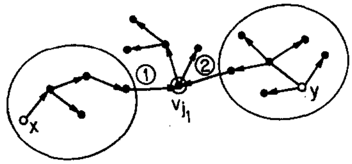


FIGURE 7

III(i). $(d_{j_1} - 1)\, a_{j_1, j_2} \cdots a_{j_{l-1}, j_l}(d_{j_l} - 1)(d_{j_{l+1}}) \cdots (d_{j_k})$ (see Fig. 8). Now on the path between the sources x and y we must choose the two points v_{j_1} and v_{j_2} as well as a direction. Thus, the total contribution in this case is

$$(-1)^k 2\pi(F) \sum_{i=1}^{t} \frac{1}{|C_i|} \sum_{\{x,y\} \subseteq C_i} \binom{d(x, y) - 1}{2}. \tag{23}$$

The remaining cases II(ii), (iii) and III(ii), (iii) cannot contribute for $\|F\| = k + 1$.

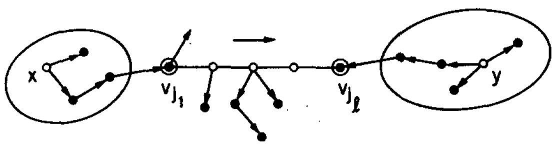


FIGURE 8

Hence, combining (21), (22), and (23) we obtain the following result by the use of (7), (13), and (14).

FACT 5. *If* $\|F\| = k + 1$,

$$A_F^{(k)} = \pi(F) \sum_{i=1}^{t} \frac{1}{|C_i|} \sum_{\{x,y\} \subseteq C_i} (2d(x, y) - d^2(x, y)). \tag{24}$$

Note that (24) *can be written alternatively as*

$$A_F^{(k)} = \pi(F) \left(\frac{1}{2} \|F\| - \sum_{i=1}^{t} \frac{1}{|C_i|} \sum_{\{x,y\} \subseteq C_i} (d(x, y) - 1)^2 \right). \tag{24'}$$

$\|F\| = k$

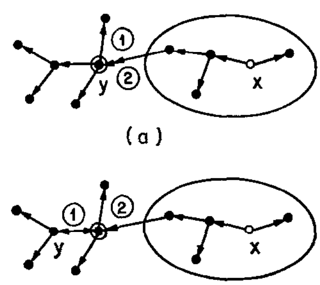
In some ways this is the most difficult case.

I. $(n-1)(d_{j_1}) \cdots (d_{j_k})$. The multiply marked edge must be an edge which has both an arrow and the symbol e. All components are initially marked normally (i.e., using (ii)). Then an arbitrary edge of F is selected for e. The total contribution is therefore

$$(-1)^{k+1}\,\pi(F)\,\|F\|. \tag{25}$$

II(i). $d_{j_1} d_{j_1} (d_{j_2}) \cdots (d_{j_k})$. There are two ways an edge can be lost. They are shown in Fig. 9. In Fig. 9a, for each choice of a pair of points of C^*, there are two ways of marking C^*, corresponding to the choices of the vertex to be called x and the vertex to be called y. The total contribution is

$$(-1)^k 2\pi(F) \sum_{i=1}^{t} \frac{1}{|C_i|} \sum_{\{x,y\}\subseteq C_i} 1 = (-1)^k \pi(F)\,\|F\|. \tag{26}$$



(a)

(b)

FIGURE 9

In Fig. 9b, the points x and y cannot be adjacent. Also, we have a factor of 4 corresponding to the choice of the pair selected from C^* to be called x and the assignment of ① and ②. Thus, the total contribution in this case is

$$\begin{aligned}
&(-1)^k 4\pi(F) \sum_{i=1}^{t} \frac{1}{|C_i|} \sum_{\substack{\{x,y\}\subseteq C_i \\ d(x,y)\geq 2}} 1 \\
&= (-1)^k 4\pi(F) \sum_{i=1}^{t} \frac{1}{|C_i|} \Bigg(\sum_{\{x,y\}\subseteq C_i} 1 - \sum_{\substack{\{x,y\}\subseteq C_i \\ d(x,y)=1}} 1 \Bigg) \\
&= (-1)^k 4\pi(F) \sum_{i=1}^{t} \frac{1}{|C_i|} \left\{ \binom{|C_i|}{2} - \|C_i\| \right\} \\
&= (-1)^k \pi(F) \left\{ 2\|F\| - 4 \sum_{i=1}^{t} \frac{\|C_i\|}{|C_i|} \right\}.
\end{aligned} \tag{27}$$

III(i). $(d_{j_1} - 1)\, a_{j_1,j_2} \cdots a_{j_{l-1},j_l}(d_{j_l} - 1)(d_{j_{l+1}}) \cdots (d_{j_k})$. The factor $(d_{j_1} - 1)$ is interpreted as choosing any edge of v_{j_1} *except* the one on the path between v_{j_1} and v_{j_l} (with $(d_{j_l} - 1)$ interpreted similarly). The only possibility for marking C^* is shown in Fig. 10. We must have $d(x, y) \geqslant 3$, since $l \geqslant 2$. Once x and y and the direction are chosen, there are $d(x, y) - 2$ choices for v_{j_l}. Thus, in this case the contribution is

$$(-1)^k 4\pi(F) \sum_{i=1}^{t} \frac{1}{|C_k|} \sum_{\substack{\{x,y\}\subseteq C_i \\ d(x,y)\geqslant 3}} (d(x, y) - 2)$$

$$= (-1)^k 4\pi(F) \left\{ \sum_{i=1}^{t} \frac{1}{|C_i|} \sum_{\{x,y\}\subseteq C_i} d(x, y) - \|F\| + \sum_{i=1}^{t} \frac{\|C_i\|}{|C_i|} \right\}. \quad (28)$$

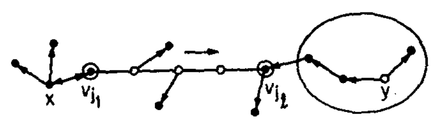


FIGURE 10

II(ii). $4d_{j_1}(d_{j_2}) \cdots (d_{j_k})$. This term has the property that it appears k times in the expansion of the determinant, once for each choice of the unparenthesized term d_{j_1}. Since there are just k factors in it and $\|F\| = k$ in this case, no edges are lost, all components are marked normally, and the total contribution is

$$(-1)^{k+1}\, 4k\pi(F). \quad (29)$$

III(ii). $-((d_{j_1} - 1) + (d_{j_l} - 1))\, a_{j_1,j_2} \cdots a_{j_{l-1},j_l}(d_{j_{l+1}}) \cdots (d_{j_k}^k)$. The markings corresponding to this case are shown in Fig. 11. As before, we must have $d(x, y) \geqslant 2$. The contribution is readily calculated to be

$$(-1)^{k+1} 4\pi(F) \sum_i \frac{1}{|C_i|} \sum_{\{x,y\}\subseteq C_i} (d(x, y) - 1)$$

$$= (-1)^{k+1}\pi(F) \left(4 \sum_{i=1}^{t} \frac{1}{|C_i|} \sum_{\{x,y\}\subseteq C_i} d(x, y) - 2\|F\| \right). \quad (30)$$

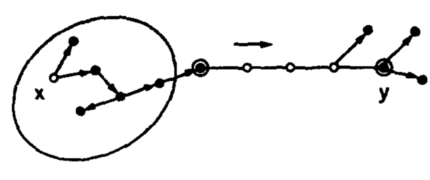


FIGURE 11

II(iii). $4(d_{j_2})\cdots(d_{j_k})$. At first sight it would appear that there are no contributions to $\|F\| = k$ from this case. However, it must be recognized that this term actually occurs $n - k + 1$ times in the expansion of $\delta'_{n-k}(T)$. We can write this as

$$4(n-k+1)(d_{j_2})\cdots(d_{j_k}) = 4(n-k)(d_{j_2})\cdots(d_{j_k}) + 4(d_{j_2})\cdots(d_{j_k}). \tag{31}$$

We interpret the term $4(n-k)(d_{j_2})\cdots(d_{j_k})$ as selecting distinct edges incident to $v_{j_2},\ldots, v_{j_k}$ (as usual) together with another distinct edge e^* (since T has $n-1$ edges altogether). The corresponding marking is shown in Fig. 12. This therefore contributes

$$(-1)^k 4\pi(F) \sum_{i=1}^{t} \frac{1}{|C_i|} \sum_{\{x,y\}\subseteq C_i} d(x,y). \tag{32}$$

There are no other contributions to $\|F\| = k$. We may now sum all the preceding expressions for the case $\|F\| = k$ to obtain the following result.

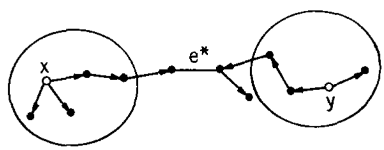


FIGURE 12

FACT 6. *If* $\|F\| = k$ *then*

$$A_F^{(k)} = 4\pi(F)\left\{\|F\| - \sum_{i=1}^{t} \frac{1}{|C_i|} \sum_{\{x,y\}\subseteq C_i} d(x,y)\right\}. \tag{33}$$

$\|F\| = k - 1$

There are no contributions here from I.

II(i). $d_{j_1}d_{j_1}(d_{j_2})\cdots(d_{j_k})$. There are two possibilities here. They are shown in Fig. 13. In Fig. 13a, an edge of C^* is chosen and one end is distinguished. Thus, this case contributes

$$(-1)^k 2\pi(F) \sum_{i=1}^{t} \frac{\|C_i\|}{|C_i|}. \tag{34}$$

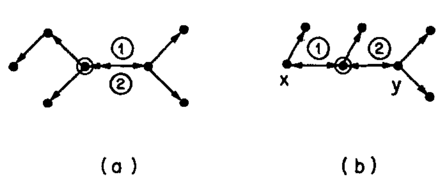


FIGURE 13

In Fig. 13b, a pair of points x, y with $d(x, y) = 2$ is chosen and the two edges between them are ordered. Hence, this case contributes

$$(-1)^k 2\pi(F) \sum_{i=1}^{t} \frac{1}{|C_i|} \sum_{\substack{\{x,y\}\subseteq C_i \\ d(x,y)=2}} 1. \tag{35}$$

II(ii). $4d_{j_1}(d_{j_2}) \cdots (d_{j_k})$. The marking shown in Fig. 14 can contribute to only *two* terms, namely, $4(d_{j_1}) \cdots d_i \cdots (d_j) \cdots (d_{j_k})$ and $4(d_{j_1}) \cdots (d_i) \cdots d_j \cdots (d_{j_k})$, since the parenthesized d_l's cannot place arrows on the same edge. Hence, to mark C^*, we simply choose a distinguished edge. The factor of 2 comes from the *two* terms to which this marking contributes. The resulting expression is therefore

$$(-1)^{k+1} 8\pi(F) \sum_{i=1}^{t} \frac{\| C_i \|}{|C_i|}. \tag{36}$$

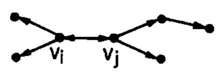


FIGURE 14

II(iii). $4(d_{j_2}) \cdots (d_{j_k})$. From the discussion of case II(iii) for $\|F\| = k$, we may use expansion (31). In particular, the second term $4(d_{j_2}) \cdots (d_{j_k})$, now summed just over $1 \leqslant j_2 < \cdots < j_k \leqslant n$, results in a contribution of

$$(-1)^k \, 4\pi(F). \tag{37}$$

III(i). $(d_{j_1} - 1)\, a_{j_1,j_2} \cdots a_{j_{l-1},j_l}(d_{j_l} - 1)(d_{j_{l+1}}) \cdots (d_{j_k})$. The markings of C^* contributing to $\|F\| = k - 1$ are shown in Fig. 15. Since $l \geqslant 2$, we must have $d(x, y) \geqslant 3$. The contribution here is

$$\begin{aligned}
&(-1)^k 2\pi(F) \sum_{i=1}^{t} \frac{1}{|C_i|} \sum_{\substack{\{x,y\}\subseteq C_i \\ d(x,y)\geqslant 3}} 1 \\
&= (-1)^k 2\pi(F) \left\{ \sum_{i=1}^{t} \frac{1}{|C_i|} \left(\sum_{\{x,y\}\subseteq C_i} 1 - \sum_{\substack{\{x,y\}\subseteq C_i \\ d(x,y)=1}} 1 - \sum_{\substack{\{x,y\}\subseteq C_i \\ d(x,y)=2}} 1 \right) \right\} \\
&= (-1)^k \pi(F) \left(\|F\| - 2 \sum_{i=1}^{t} \frac{\| C_i \|}{|C_i|} - 2 \sum \frac{1}{|C_i|} \sum_{\substack{\{x,y\}\subseteq C_i \\ d(x,y)=2}} 1 \right).
\end{aligned} \tag{38}$$

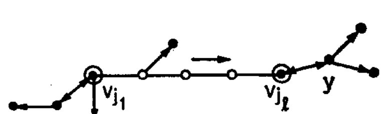


FIGURE 15

III(ii). $((d_{j_1}-1)+(d_{j_l}-1))\,a_{j_1,j_2}\cdots a_{j_{l-1},j_l}(d_{j_{l+1}})\cdots(d_{j_k})$. The corresponding marking is shown in Fig. 16. We must have $d(x,y)\geqslant 2$. Thus, we obtain a contribution of

$$(-1)^{k+1}4\pi(F)\sum_{i=1}^{t}\frac{1}{|C_i|}\sum_{\substack{\{x,y\}\subseteq C_i\\ d(x,y)\geqslant 2}}1$$

$$=(-1)^{k+1}4\pi(F)\sum_{i=1}^{t}\frac{1}{|C_i|}\Big(\sum_{\{x,y\}\subseteq C_i}1-\sum_{\substack{\{x,y\}\subseteq C_i\\ d(x,y)=1}}1\Big)$$

$$=(-1)^{k+1}\pi(F)\Big(2\,\|F\|-4\sum_{i=1}^{t}\frac{\|C_i\|}{|C_i|}\Big). \tag{39}$$

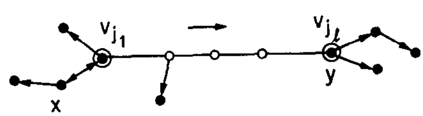


FIGURE 16

III(iii). $a_{j_1,j_2}\cdots a_{j_{l-1},j_l}(d_{j_{l+1}})\cdots(d_{j_k})$. We show the marking in Fig. 17. The contribution is

$$(-1)^{k}2\pi(F)\sum_{i=1}^{t}\frac{1}{|C_i|}\sum_{\{x,y\}\subseteq C_i}1$$

$$=(-1)^{k}\pi(F)\,\|F\|. \tag{40}$$

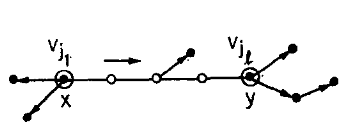


FIGURE 17

By combining the expressions in (34)–(40) we obtain the following result.

FACT 7. *If* $\|F\|=k-1$ *then*

$$A_F^{(k)}=4\pi(F)\Big(\sum_{i=1}^{t}\frac{\|C_i\|}{|C_i|}-1\Big). \tag{41}$$

In Table III (see Appendix) we list the values of $A_F^{(k)}$ for $F\in\mathscr{F}_m$, $m\leqslant 5$. We combine Facts 5, 6, and 7 into the main result of the paper.

THEOREM. *Let T be a tree with $n \geqslant 2$ vertices and distance matrix $D(T)$. If we write*

$$\Delta_T(\lambda) = \det(D(T) - \lambda I) = \sum_{k=0}^{n} \delta_k(T)\, \lambda^k$$

then

$$\delta_k(T) = (-1)^{n-1} 2^{n-k-2} \Bigg\{ \sum_{F \in \mathscr{F}_{k+1}} \pi(F) \left(\frac{1}{2} \|F\| - \sum_C \frac{1}{|C|} \sum_{\{x,y\} \subseteq C} (d(x,y) - 1)^2 \right) N_F(T)$$

$$+ 4 \sum_{F \in \mathscr{F}_k} \pi(F) \left(\|F\| - \sum_C \frac{1}{|C|} \sum_{\{x,y\} \subseteq C} d(x,y) \right) N_F(T)$$

$$+ 4 \sum_{F \in \mathscr{F}_{k-1}} \pi(F) \left(\sum_C \frac{\|C\|}{|C|} - 1 \right) N_F(T) \Bigg\}, \tag{42}$$

where C ranges over all components of F.

It follows from the theorem that we can rephrase (42) in the following appealing recursive form.

THEOREM′. *For all trees T with $n \geqslant 2$ vertices*

$$\delta_k(T) = (-1)^{n-1} 2^{n-k-2}$$
$$\times \Bigg\{ \sum_{F \in \mathscr{F}_{k+1}} a_F \pi(F)\, N_F(T) + \sum_{F \in \mathscr{F}_k} b_F \pi(F)\, N_F(T) + \sum_{F \in \mathscr{F}_{k-1}} c_F \pi(F)\, N_F(T) \Bigg\}$$

where a_F, b_F, c_F are defined as follows:

(i) *For the empty tree F^*, $a_{F^*} = b_{F^*} = 0$, $c_{F^*} = -4$.*

(ii) *If F is a* tree *T' with $n' \geqslant 2$ vertices and distance matrix (d'_{ij}) then*

$$a_{F'} = \frac{1}{n'} \sum_{i<j} d'_{ij}(2 - d'_{ij}),$$

$$b_{F'} = \frac{4}{n'} \sum_{i<j} (2 - d'_{ij}),$$

$$c_{F'} = -\frac{4}{n'}.$$

(iii) *If F is the disjoint union of forests F_1 and F_2 then*

$$a_F = a_{F_1} + a_{F_2},$$

$$b_F = b_{F_1} + b_{F_2},$$

$$c_F = c_{F_1} + c_{F_2} + 4.$$

Concluding Remarks

From Table II (see Appendix), it is apparent that in many cases $A_F^{(k)} = 4A_F^{(k-1)}$ when $\|F\| = k$. The following result explains this phenomenon. By a *star* S_m we mean a tree having m edges and at most one vertex of degree exceeding 1.

Fact 9. *If F is a union of stars and $\|F\| = k$ then*

$$A_F^{(k)} = 4A_F^{(k-1)}.$$

Proof. We first calculate

$$A_{S_m}^{(m-1)} = m, \qquad A_{S_m}^{(m)} = 4m, \qquad \pi(S_m) = m,$$

which implies

$$B_{S_m}^{(m-1)} = 1, \qquad B_{S_m}^{(m)} = 4.$$

The desired result now follows by Fact 8. ∎

Note that by Fact 7, for $\|F\| = k - 1$,

$$A_F^{(k)} < 0 \qquad \text{iff } F \text{ is a tree,}$$

$$A_F^{(k)} = 0 \qquad \text{iff } \ F = 2P_1 .$$

We remark that it is possible to have two different forests F and F' such that $A_F^{(k)} = A_{F'}^{(k)}$ for all k. An example of such a pair is

$$F = S_1 \cup 3S_3 \cup 3S_5 \cup S_{11},$$
$$F' = 4S_2 \cup 4S_7,$$

for which

$$\|F\| = \|F'\| = 36, \qquad \pi(F) = \pi(F') = 2^{12}3^4,$$
$$A_F^{(35)} = A_{F'}^{(35)} = 2^{11} \cdot 3^3 \cdot 37,$$
$$A_F^{(36)} = A_{F'}^{(36)} = 2^{13} \cdot 3^3 \cdot 37 \qquad \text{(by Fact 9)},$$
$$A_F^{(37)} = A_{F'}^{(37)} = 2^{11} \cdot 3^3 \cdot 31.$$

Since we know $\delta_{n-1}(T) = 0$ for all T, then for $k = n - 1$ the theorem reduces to the simple identity

$$\sum_{\{x,y\}\subseteq T} d(x, y) - \sum_{\{e_1,e_2\}\subseteq \text{edges of } T} d(e_1, e_2) = \| T \|^2, \tag{43}$$

where $d(e_1, e_2)$ is defined to be the length of the path joining the edges e_1 and e_2.

In [5], it is shown that for all trees T,

$$(-1)^{n-1}\,\delta_k(T) > 0 \qquad \text{for} \quad 0 \leqslant k < n-1, \tag{44}$$

where $n = |T|$. It appears that in fact for each tree T, the quantities $(-1)^{n-1}\,\delta_k(T)/2^{n-k-2}$ are *unimodal* with the maximum value occurring for $k = [n/2]$. We see no way to prove this, however.

As we noted in the Introduction, it has recently been shown that nonisomorphic trees can have the same distance matrix polynomial. The smallest such pair, due to McKay [15], is shown in Fig. 18. In fact, just as in the cor-

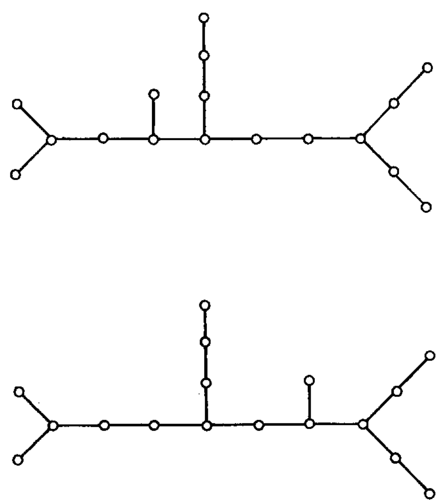


FIG. 18. The smallest pair of nonisomorphic trees T and T', $\Delta_T(\lambda) = \Delta_{T'}(\lambda)$.

responding case for adjacency matrix polynomials of trees [13], McKay has shown that almost all large trees T have (exponentially) may distance cospectral mates, i.e., nonisomorphic trees which share the same distance matrix polynomial $\Delta_T(\lambda)$.

More generally, one might consider $\Delta_G(\lambda)$ for general graphs G. It would be interesting to know the result which corresponds to the theorem in this case. It has been shown [14] that in general the determinant $\det(D(G))$ of the distance matrix of a connected graph G depends only on the *blocks* (= maximal 2-connected subgraphs) of G and *not* on how they are interconnected. In particular, this gives a particularly lucid explanation of why $\det(D(T))$ depends only on the size of T and not on its structure.

Appendix

In Tables I, II, and III we give short lists of the trees on at most eight vertices (Table I) (see [10]), the $\Delta_T(\lambda)$ for these trees (Table II), and the coefficients $A_F^{(k)}$ for $k \leqslant 5$ (Table III).

TABLE I

Trees on n Vertices, $n \leqslant 8$

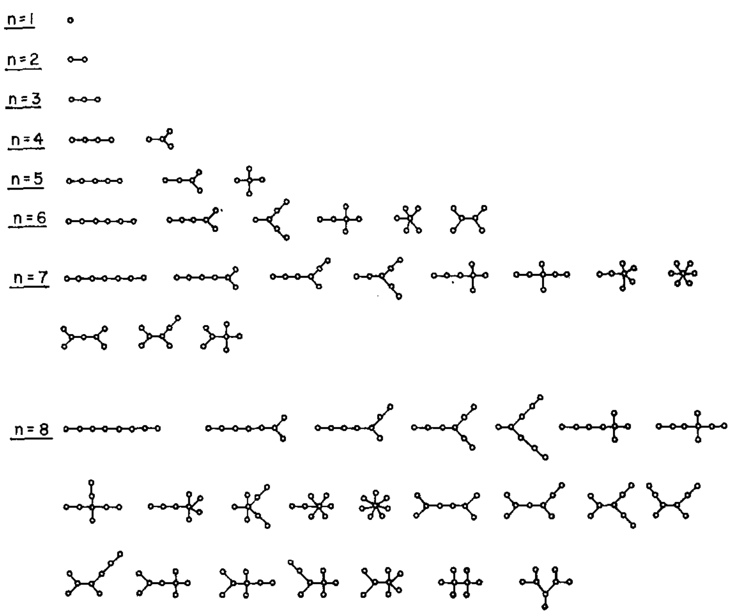


TABLE II

Coefficients of $\Delta_{t_i}(\lambda)$ for All Trees on at Most Eight Vertices[a]

t_i	k: 0	1	2	3	4	5	6	7	8
$n = 1$									
t_1	0	−1							
$n = 2$									
t_1	1	0	−1						
$n = 3$									
t_1	2	6	0	−1					
$n = 4$									
t_1	3	16	20	0	−1				
t_2	3	14	15	0	−1				
$n = 5$									
t_1	4	30	70	50	0	−1			
t_2	4	28	58	38	0	−1			
t_3	4	24	44	28	0	−1			
$n = 6$									
t_1	5	48	162	224	105	0	−1		
t_2	5	46	145	184	84	0	−1		
t_3	5	46	143	178	77	0	−1		
t_4	5	44	126	148	65	0	−1		
t_5	5	42	117	136	60	0	−1		
t_6	5	36	90	100	45	0	−1		
$n = 7$									
t_1	6	70	308	630	588	196	0	−1	
t_2	6	68	286	552	488	164	0	−1	
t_3	6	68	284	540	464	148	0	−1	
t_4	6	68	282	528	438	132	0	−1	

Table continued

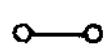
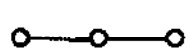
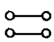
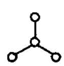
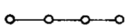
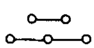
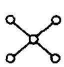
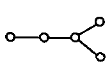
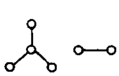
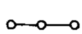
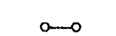
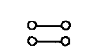
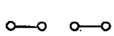
TABLE II (*continued*)

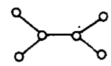
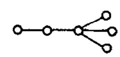
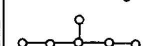
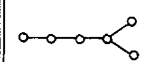
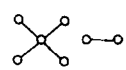
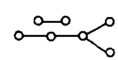
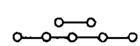
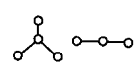
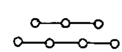
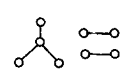
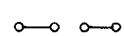
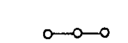
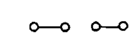
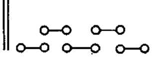
t_i	k = 0	1	2	3	4	5	6	7	8
t_5	6	64	248	438	366	122	0	−1	
t_6	6	64	244	442	340	108	0	−1	
t_7	6	58	200	324	208	86	0	−1	
t_8	6	50	156	240	190	66	0	−1	
t_9	6	66	264	476	402	134	0	−1	
t_{10}	6	66	260	460	376	120	0	−1	
t_{11}	6	62	222	366	292	96	0	−1	
$n = 8$									
t_1	7	96	530	1408	1980	1344	336	0	−1
t_2	7	94	493	1280	1721	1134	291	0	−1
t_3	7	94	491	1264	1672	1068	264	0	−1
t_4	7	94	489	1246	1611	984	228	0	−1
t_5	7	94	491	1262	1662	1056	255	0	−1
t_6	7	90	445	1080	1365	875	227	0	−1
t_7	7	90	441	1052	1293	790	195	0	−1
t_8	7	90	437	1026	1227	720	172	0	−1
t_9	7	84	426	852	1011	624	164	0	−1
t_{10}	7	84	376	822	948	564	143	0	−1
t_{11}	7	76	310	636	715	432	116	0	−1
t_{12}	7	66	245	476	525	322	91	0	−1
t_{13}	7	92	466	1156	1483	952	248	0	−1
t_{14}	7	92	464	1138	1432	892	223	0	−1
t_{15}	7	92	460	1110	1356	808	191	0	−1
t_{16}	7	92	460	1112	1368	824	200	0	−1
t_{17}	7	92	462	1128	1411	872	216	0	−1
t_{18}	7	88	418	960	1159	720	188	0	−1
t_{19}	7	88	410	920	1075	640	160	0	−1
t_{20}	7	88	412	930	1096	660	167	0	−1
t_{21}	7	82	349	732	829	498	131	0	−1
t_{22}	7	84	362	764	867	520	136	0	−1
t_{23}	7	90	433	996	1185	714	179	0	−1

[a] The t_i refer to trees listed in the order of Table I. The automatic factor of 2^{n-k-2} has been removed from all coefficients except $\delta_n(t_i)$, which is always -1.

TABLE III

Coefficients $A_F^{(k)}$ in the Expansion $\delta_k(T) = \Sigma_F A_F^{(k)} N_F(T)$ for $\|F\| \leqslant 5$ [a]

F	$A_F^{(0)}$	$A_F^{(1)}$	$A_F^{(2)}$	$A_F^{(3)}$	$A_F^{(4)}$	$A_F^{(5)}$
(empty)	0	-4				
	1	4	-4			
		2	8	-4		
		4	16	0		
			3	12	-4	
			0	8	-4	
			7	28	4	
			12	48	16	
				4	16	-4
				-2	8	-4
				-10	0	-4
				10	40	8
				12	48	12
				4	32	8
				20	80	32
				32	128	64

F	$A_F^{(4)}$	$A_F^{(5)}$	$A_F^{(6)}$
	5	20	-4
	-7	4	-4
	-4	8	-4
	-15	-4	-4
	-20	-8	-4
	-35	-20	-4
	13	52	12
	1	36	12
	-15	10	12
	17	68	20
	8	56	20
	28	112	48
	16	96	48
	33	132	60
	52	208	112
	80	320	192

[a] The shaded regions are 0.

REFERENCES

1. L. H. BRANDENBURG, B. GOPINATH, AND R. P. KURSHAN, On the addressing problem of loop switching, *Bell System Tech. J.* 51 (1972), 1445–1469.

2. K. L. Chew and K. Y. Yap, On the addressing problem for graphs, *Nanta Math.* **6** (1973), 77–84.
3. L. Collatz and U. Singowitz, Spektren endlicher Grafen, *Abh. Math. Sem. Univ. Hamburg* **21** (1957), 63–67.
4. D. M. Cvetković, "Graphs and Their Spectra," Pub. de la Faculté d'Électrotechnique de l'Université à Belgrade, Serie: Mathématiques et Physique, No. 354 (1971).
5. M. Edelberg, M. R. Garey, and R. L. Graham, On the distance matrix of a tree, *Discrete Math.* **14** (1976), 23–29.
6. R. L. Graham and H. O. Pollak, On the addressing problem for loop switching, *Bell. System Tech. J.* **50** (1971), 2495–2519.
7. R. L. Graham and H. O. Pollak, "Embedding Graphs in Squashed Cubes," pp. 99–110, Springer Lecture Notes in Mathematics, Vol. 303, Springer-Verlag, Berlin/New York, 1973.
8. R. L. Graham and E. Szemerédi, On subforest number independence in trees, *J. Combinatorial Theory* (B) **24** (1978), 213–222.
9. F. Harary, "Graph Theory," Addison–Wesley, New York, 1969.
10. A. J. Hoffman, The eigenvalues of the adjacency matrix of a graph, *in* "Combinatorial Mathematics and Its Applications," pp. 578–584, Univ. of North Carolina Press, Chapel Hill, 1969.
11. L. Lovász and J. Pelikán, *Per. Math. Hungar.* **3** (1973), 175–182.
12. A. Moshowitz, The characteristic polynomial of a graph, *J. Combinatorial Theory, Ser. B* **12** (1972), 177–193.
13. A. J. Schwenk, Almost all trees are cospectral, *in* "New Directions in the Theory of Graphs" (F. Harary, Ed.), pp. 275–307, Academic Press, New York, 1973.
14. R. L. Graham, A. J. Hoffman and H. Hosoya, On the distance matrix of a directed graph, *J. Graph Theory* 1 (1977), 85–88.
15. B. D. McKay, On the spectral characterization of trees, *Ars Combinatoria III* (1977), 219–232.