On a Conjecture of Graham and Lovász about Distance Matrices

Karen L. Collins
Wesleyan University, Dept. of Mathematics
Middletown, CT 06457

Abstract

In their 1978 paper "Distance Matrix Polynomials of Trees", [4], Graham and Lovász proved that the coefficients of the characteristic polynomial of the distance matrix of a tree ($CPD(T)$) *can be expressed in terms of the numbers of certain subforests of the tree. This result was generalized to trees with weighted edges by Collins, [1], in 1986. Graham and Lovász computed these coefficients for all trees on less than 8 vertices, noticed that the sequence of coefficients was unimodal with peak at the center, and conjectured that this was always true. In this paper, we disprove the conjecture. The coefficients for a star on* n *vertices are indeed unimodal with peak at* $\left[\frac{n}{2}\right]$*, but the coefficients for a path on* n *vertices are unimodal with peak at* $n(1 - 1/\sqrt{5})$.

Adjacency matrices have received a lot of attention, particularly with respect to their characteristic polynomials (CP's). It has only been in the last few years that very much work has been done with distance matrices. This is partly because adjacency matrices have many more zeroes than distance matrices, and therefore it is easier to compute their CP's, and partly because even adjacency matrices themselves are not well understood. When the first pair of trees (surely the simplest kind of graphs) were found that had the same CP of adjacency matrices, it was shown that we could not so easily find a mapping from the set of all trees with n vertices to the set of polynomials of degree n. Later this desire lead to the discovery of two trees with the same CP of distance matrices [6]. If only each tree could have associated with it some unique polynomial; for then the set of trees would somehow be like an algebra, and might yield information from algebraic techniques. Unfortunately, at the moment there are no good candidates for such an assignment of polynomials (see [6]). Meanwhile, we seek to understand what such classifications as we have discovered may tell us about trees.

Distance matrices also have arisen independently from a data communication problem studied by Graham and Pollack, [5], in 1971. See also Graham and Lovász, [4], for a description of the problem.

In this paper we investigate some of the properties of the characteristic polynomials of the distance matrices of trees. Wherever possible, we compare these results with the results for adjacency matrices, in order to think about the question of whether distance matrices do give us more information about trees than adjacency matrices, or whether the information we get from distance matrices is contained in the information from adjacency matrices. For instance, we prove that coefficients of the CP of the distance matrix of a path on n vertices are unimodal with peak at $n(1-1/\sqrt{5})$. Thus, the ratio of the position of the peak in the sequence, to the number of terms in the sequence is $(1-1/\sqrt{5})$. If we consider only the absolute values of the nonzero terms of the sequence of coefficients of the CP of the adjacency matrix of an n vertex path, its ratio of position of the peak to the number of terms in the sequence is also $(1-1/\sqrt{5})$. This also holds for stars. We conjecture that this is always true (see Conjecture 2 at the end).

The distance matrix of a graph with n vertices is defined to be an n x n matrix with ij entry the distance in the graph between vertex i and vertex j. The adjacency matrix is an n x n matrix with ij entry 1 if vertex i and vertex j are adjacent and 0 otherwise. It is well-known (see [2]) that we can describe the characteristic polynomial of the adjacency matrix of a tree in terms of the number of different size matchings in a tree:

$$a_{n-k} = \begin{cases} (-1)^{n+k/2} \textit{ times the number of } k/2 \textit{ matchings in } T \textit{ when } k \textit{ is even} \\ 0 \textit{ otherwise} \end{cases}$$

In their 1978 paper, Graham and Lovász show a similar (but complicated) result for distance matrices. In her 1986 MIT thesis, Collins generalizes their result to trees with weighted edges. For example, the determinant of the distance matrix of a tree T with n vertices is always $(-1)^{n-1}(n-1)2^{n-2}$. The determinant of the distance matrix depends only on the subforests of the tree that have one edge; since every tree has the same number of edges, the determinants are all the same. The determinant of the distance matrix of a tree with edges weighted $x_1, x_2, \ldots, x_{n-1}$ is

$$(-1)^{n-1}(\sum_{i=1}^{n-1} x_i) \prod_{j=1}^{n-1} x_j$$

Compare this with the case for adjacency matrices: $a_0 = 0$ in both the weighted and unweighted versions, unless T has an $n/2$ matching. But if a tree has one $n/2$ matching, it can have only one; in the unweighted case, n must be even and $a_0 = (-1)^{3n/2}$. In the general weighted case,

$$a_{n-k} = \{ {(-1)^{n+k/2} \sum_M w(M), for\ k\ even \atop 0\ otherwise}$$

where $M = a weighted matching$ and $w(M) = \prod x_i^2$ over every weighted edge x_i in M. Therefore, when T has an $n/2$ matching, M_0,

$$a_0 = (-1)^{3n/2} \prod_{x_i\ is\ in\ M_0} x_i^2$$

Let δ_k be the coefficient of λ^k in CPD(T) for tree T. In [4] it is shown that

$$\delta_k = (-1)^{n-1} 2^{n-k-2} \sum_F A_{F,k} N_F$$

where F is a forest with $k-1, k$ or $k+1$ vertices, where $A_{F,k}$ is a number depending only on F and k, and where N_F is the number of copies of F contained as a subgraph in T. Let

$$d_k = (-1)^{n-1} \delta_k / 2^{n-k-2}$$

Then Graham and Lovász conjectured that the sequence $d_1, d_2, \ldots, d_{n-2}$ is unimodal, with peak at $k = n/2$. This is true for stars, but not for paths. It is not known whether the sequence is unimodal (with varying peaks) for general trees. (Note that $d_{n-1} = 0$ and $d_n = -1$ always.)

Theorem 1 *The coefficients of CPD(S) where S is a star on n vertices are unimodal with peak at* $[n/2]$.

Proof:

$$CPD(S) = det \begin{pmatrix} -\lambda & 1 & 1 & 1 & \ldots & 1 \\ 1 & -\lambda & 2 & 2 & \ldots & 2 \\ 1 & 2 & -\lambda & 2 & \ldots & 2 \\ 1 & 2 & 2 & -\lambda & \ldots & 2 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 2 & 2 & 2 & \ldots & -\lambda \end{pmatrix}$$

The coefficient of $(-\lambda)^k$ in CPD(S) (for a star with n vertices) is

$$\binom{n-1}{k} det \underbrace{\begin{pmatrix} 0 & 1 & 1 & 1 & \dots & 1 \\ 1 & 0 & 2 & 2 & \dots & 2 \\ 1 & 2 & 0 & 2 & \dots & 2 \\ 1 & 2 & 2 & 0 & \dots & 2 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 2 & 2 & 2 & \dots & 0 \end{pmatrix}}_{n-k} + \binom{n-1}{k-1} det \underbrace{\begin{pmatrix} 0 & 2 & 2 & 2 & \dots & 2 \\ 2 & 0 & 2 & 2 & \dots & 2 \\ 2 & 2 & 0 & 2 & \dots & 2 \\ 2 & 2 & 2 & 0 & \dots & 2 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 2 & 2 & 2 & 2 & \dots & 0 \end{pmatrix}}_{n-k}$$

For an m x m matrix,

$$det \begin{pmatrix} 0 & 1 & 1 & 1 & \dots & 1 \\ 1 & 0 & 2 & 2 & \dots & 2 \\ 1 & 2 & 0 & 2 & \dots & 2 \\ 1 & 2 & 2 & 0 & \dots & 2 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 2 & 2 & 2 & \dots & 0 \end{pmatrix} = (-1)^{m-1} 2^{m-2} (m-1)$$

and

$$det \begin{pmatrix} 0 & 2 & 2 & 2 & \dots & 2 \\ 2 & 0 & 2 & 2 & \dots & 2 \\ 2 & 2 & 0 & 2 & \dots & 2 \\ 2 & 2 & 2 & 0 & \dots & 2 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 2 & 2 & 2 & 2 & \dots & 0 \end{pmatrix} = (-1)^{m-1} 2^{m} (m-1)$$

Therefore, the coefficient of λ^k is

$$(-1)^n (n-k-1) 2^{n-k-2} (\binom{n-1}{k} + 4\binom{n-1}{k-1})$$

Hence $d_k = (n-k-1)(\binom{n-1}{k} + 4\binom{n-1}{k-1})$. Therefore

$$d_k - d_{k-1} = \binom{n-1}{k}(\frac{n^3 + n(-7k^2 + 3k - 1) + 6k^3 - 3k^2 - 3k}{n^2 + n(1-2k) + k^2 - k})$$

The denominator is positive for $1 \leq k \leq n-1$. Fix n as a positive integer greater than 1. Let $P(k) = n^3 + n(-7k^2 + 3k - 1) + 6k^3 - 3k^2 - 3k$. The sign of

$d_k - d_{k-1}$ depends on the sign of $P(k)$. We show $P(k)$ has one negative root, one root near $n/2$ and one root that is greater than n. If k is much less than 0, the term $6k^3$ dominates $P(k)$ and $P(k) < 0$. If k is much greater than 0, again, $6k^3$ dominates and $P(k) > 0$. At $k = 0$, $P(k) = n^3 - n > 0$. Therefore $P(k)$ has a negative root. At $k = n$, $P(k) = -2n < 0$ and therefore $P(k)$ has a root greater than n. The remaining root is between $0 \leq k \leq n$.

Suppose n is even, say $n = 2m$. Then

$$P(k) = 8m^3 + 2m(-7k^2 + 3k - 1) + 6k^3 - 3k^2 - k$$

At $k = m$, $P(k) = 3m^2 - 3m > 0$. At $k = m + 1$, $P(k) = -7m^2 + m + 2 < 0$. So the peak of unimodality occurs at $k = n/2$.

Suppose n is odd, say $n = 2m + 1$. Then

$$P(k) = 8m^3 + 12m^2 + m(-14k^2 + 6k + 4) + 6k^3 - 10k^2 + 2k$$

At $k = m$, $P(k) = 8m^2 + 6m > 0$. At $k = m + 1$, $P(k) = -2m^2 - 4m - 2 < 0$. Thus the peak of unimodality occurs at $(n - 1)/2$.

Remark 1 *$CPA(S) = (-1)^n(\lambda^n - \lambda^{n-2})$, where S is a star on n vertices.*

Theorem 2 *The coefficients of $CPD(P)$ where P is a path on n vertices are unimodal with peak at $n(1 - 1/\sqrt{5})$.*

Proof: We take a description of the coefficients of $CPD(P)$ from [1]. Let $D(P)$ be the distance matrix of the path P. Let $D[v_1, v_2, \ldots v_k]$ be the k x k submatrix of $D(T)$ for any tree T whose rows and columns are indexed by $v_1, v_2, \ldots v_k$. Then the coefficient of λ^{n-k} in $CPD(P)$ is (for $k \geq 2$)

$$(-1)^{n-k} \sum_{1 \leq i_1 < i_2 < \ldots < i_k \leq n} det(D[v_{i_1}, v_{i_2}, \ldots v_{i_k}])$$

For path P, we may interpret $D[v_{i_1}, v_{i_2}, \ldots v_{i_k}]$ as the distance matrix of the path $P[v_{i_1}, v_{i_2}, \ldots v_{i_k}]$ whose vertices are $v_{i_1}, v_{i_2}, \ldots v_{i_k}$ and whose edges are the $k - 1$ shortest distances from the set of distances between $v_{i_1}, v_{i_2}, \ldots v_{i_k}$. Since we started with a path, $P[v_{i_1}, v_{i_2}, \ldots v_{i_k}]$ is still a path and we can apply the following theorem from [3].

Theorem 3 (Graham, Hoffman and Hosoya) *Let G be a finite graph in which each edge e has associated with it an arbitrary non-negative length $w(e)$. (The usual weight chosen for edges is $w(e) = 1$.) Let*

$$d_{i,j} = \min_{P(v_i,\, v_j)} w(P(v_i, v_j))$$

where $P(v_i, v_j)$ ranges over all paths from v_i to v_j and $w(P(v_i, v_j))$ denotes the sum of all edge-lengths in $P(v_i, v_j)$. Let the 2-connected pieces of G be $G_1, G_2, \ldots, G_k$. Let $D(G)^{-1} = (d_{i,j}^{-1})$ and $cof(D(G)) = det(D(G)) \sum_{i,j} d_{i,j}^{-1}$. Then

$$det(D(G)) = \sum_{i=1}^{k} det(D(G_i)) \prod_{j \neq i} cof(D(G_j))$$

Lemma 1 $\delta_{n-k} =$

$$(-1)^{n-1} 2^{k-2} \sum_{1 \leq s_1 < s_2 < \ldots < s_k \leq n} (s_2 - s_1)(s_3 - s_2) \ldots (s_k - s_{k-1})(s_k - s_1)$$

Proof of Lemma: Since $P[v_{i_1}, v_{i_2}, \ldots v_{i_k}]$ is a tree, its 2-connected pieces are its edges. Therefore, by the above theorem, its determinant is $-(-2)^{k-2}$ times

$$(s_2 - s_1)(s_3 - s_2) \ldots (s_k - s_{k-1})(s_k - s_1)((s_2 - s_1) + (s_3 - s_2) + \ldots + (s_k - s_{k-1}))$$

since the distance between s_{i+1} and s_i is $(s_{i+1} - s_i)$. When we multiply this by $(-1)^{n-k}$ we get the above.

Let $d_{n-k} = f(n,k) = \sum_{1 \leq s_1 < s_2 < \ldots < s_k \leq n} (s_2 - s_1)(s_3 - s_2) \ldots (s_k - s_{k-1})(s_k - s_1)$

We show that $d_{n-k} = f(n,k) = \binom{n+k-1}{2k-1}(k-1)\frac{n}{k}$.

Lemma 2

$$\textit{Let} \quad g(m,k) = \sum_{1 = s_1 < s_2 < \ldots < s_k = m} (s_2 - s_1)(s_3 - s_2) \ldots (s_k - s_{k-1})$$

Then $g(m,k) = \binom{m+k-3}{2k-3}$.

Proof of Lemma: We can choose s_{k-1} to be $n-1$, $n-2$, etc. Therefore:

$$g(m,k) = \sum_{i=1}^{m-k+1} i\, g(m-i, k-1)$$

However, $\binom{m+k-3}{2k-3}$ also satisfies this recurrence. To prove this, we use the following formulae:

$$\sum_{m=k}^{n} \binom{m}{k} = \binom{n+1}{k+1} \tag{1}$$

$$\sum_{m=k}^{n} m\binom{m}{k} = (k+1)\binom{n+1}{k+2} + k\binom{n+1}{k+1} \tag{2}$$

Now the recurrence for $\binom{m+k-3}{2k-3}$ gives:

$$\sum_{i=1}^{m-k+1} i \binom{m+k-4-i}{2k-5} = \sum_{j=1}^{m-k+1} \sum_{i=j}^{m-k+1} \binom{m+k-i-4}{2k-5}$$

by formula (1)

$$= \sum_{j=1}^{m-k+1} \binom{m+k-j-3}{2k-4}$$

by formula (1)

$$= \binom{m+k-3}{2k-3}$$

It is easy to check that the initial values of $g(m,k)$ and $\binom{m+k-3}{2k-3}$ are the same.

The proof for $f(n,k)$ is by induction on n. The values $f(2,1) = 0$ and $f(2,2) = 1$ and $f(n,n) = n-1$ agree with our hypothesis. Now $f(n,k)$ also satisfies a recurrence:

$$f(n,k) = f(n-1,k) \; + \; \mathit{all\ terms\ for\ sequences\ with}\ s_n = n$$

that is,

$$f(n,k) = f(n-1,k) + (n-1)g(n,k) + (n-2)g(n-1,k) + \ldots + (k-1)g(k,k)$$

By induction, we have $f(n,k) =$

$$\binom{n+k-2}{2k-1}(k-1)(n-1)/k + (n-1)\binom{n+k-3}{2k-3} + (n-2)\binom{n+k-4}{2k-3} + \ldots + (k-1)\binom{2k-3}{2k-3}$$

We can manipulate the terms to get

$$\sum_{m=k}^{n}(m-1)\binom{m+k-3}{2k-3} = \sum_{l=2k-3}^{n+k-3}(l-k+2)\binom{l}{2k-3}$$

$$= \sum_{l=2k-3}^{n+k-3} l\binom{l}{2k-3} \; + \; \sum_{l=2k-3}^{n+k-3}(2-k)\binom{l}{2k-3}$$

Using formula (2) and (1), we can get

$$kf(n,k) = (k-1)(k\binom{n+k-2}{2k-2} \; + \; (n+2k-1)\binom{n+k-2}{2k-1})$$

and this reduces to the desired form.

Next we show that $\binom{n+k-1}{2k-1}(k-1)\frac{n}{k}$ is unimodal at $n/\sqrt{5}$. Since $f(n,k) = d_{n-k}$, this proves our theorem. Now

$$f(n,k+1) - f(n,k) =$$

$$\binom{n+k}{2k+1}k\frac{n}{(k+1)} \; - \binom{n+k-1}{2k-1}(k-1)\frac{n}{k} =$$

$$n(\frac{\binom{n+k}{2k+1}k^2 - \binom{n+k-1}{2k-1}(k-1)(k+1)}{k(k+1)}) =$$

$$\frac{n}{k(k+1)}(k^2\binom{n+k}{2k+1} - (k^2-1)\binom{n+k-1}{2k-1}) =$$

$$\frac{n(n+k-1)!}{k(k+1)(2k-1)!(n-k-1)!}(\frac{k^2(n+k)}{(2k+1)(2k)} - \frac{k^2-1}{n-k})$$

Thus, the sign of $f(n,k+1) - f(n,k)$ depends on

$$\frac{kn^2 - 5k^3 - 2k^2 + 4k + 2}{(4k+2)n - 4k^2 - 2k}$$

The denominator is always positive for $1 \leq k+1 \leq n$. The numerator is zero when

$$n = \pm\sqrt{5k^2 + 2k - \frac{2}{k} - 4}$$

We ignore the negative term. Where is the one remaining root? When $k+1 = \frac{n}{\sqrt{5}} + 1$, the numerator is $-\frac{2n}{\sqrt{5}} + \frac{2\sqrt{5}}{n} + 4$ and when $k+1 = \frac{n}{\sqrt{5}}$, the numerator is $\frac{8n^2 - 7\sqrt{5}n + 5}{\sqrt{5}n - 5}$. So $kn^2 - 5k^3 - 2k^2 + 4k + 2$ is zero when $\frac{n}{\sqrt{5}} < k+1 < \frac{n}{\sqrt{5}} + 1$. There must be some integer in this range, say $\frac{n}{\sqrt{5}} < M < \frac{n}{\sqrt{5}} + 1$, and M will be the peak of the unimodal sequence unless M is too close to $\frac{n}{\sqrt{5}} + 1$. In that case, the peak of the sequence will be $M - 1$.

Remark 2 *In [4], the values of the d_k's for $0 \le k \le n \le 8$ are given for any tree T. They are all unimodal with peak at $\left[\frac{n}{2}\right]$. However, we can see the effect of the $\frac{1}{\sqrt{5}}$ on the coefficients of a path on 9 vertices demonstrated in the following table of d_k's.*

	$k=0$	$k=1$	$k=2$	$k=3$	$k=4$	$k=5$	$k=6$	$k=7$
$n=2$	1	0	-1					
$n=3$	2	**6**	0	-1				
$n=4$	3	16	**20**	0	-1			
$n=5$	4	30	**70**	50	0	-1		
$n=6$	5	48	162	**224**	105	0	-1	
$n=7$	6	70	308	**630**	588	196	0	-1
$n=8$	7	96	520	1408	**1980**	1344	336	0
$n=9$	8	126	810	2730	5148	**5346**	2772	540

Theorem 4 *The coefficients of CPA(P) where P is a path on n vertices are "unimodal" at $n(1 - 1/\sqrt{5})$.*

Proof of Theorem: Recall that

$$a_{n-k} = \begin{cases} (-1)^{n+k/2} \textit{ times the number of } k/2 \textit{ matchings in } T \textit{ when } k \textit{ is even} \\ 0 \textit{ otherwise} \end{cases}$$

The number of i-matchings in a path on n vertices is $\binom{n-i}{i}$. Therefore

$$a_{n-k} = \begin{cases} (-1)^{n+k/2}\binom{n-i}{i} \\ 0 \textit{ otherwise} \end{cases}$$

Since every other term is 0 in the sequence of a's, and since the non-zero terms alternate in sign, we cannot really call the sequence unimodal.

However, if we take only the non-zero terms, and take the absolute value of the terms, the remaining sequence:

$$\binom{n-[\frac{n}{2}]}{[\frac{n}{2}]} \ldots \binom{n-2}{2}, (n-1), 1$$

is unimodal, with peak at $n/2(1-1/\sqrt{5})$. Now

$$\binom{n-i+1}{i-1} - \binom{n-i}{i} = \frac{(n-i)!}{(i-1)!(n-2i)!}\left[\frac{n-i+1}{(n-2i+1)(n-2i+2)} - \frac{1}{i}\right]$$

$$= -\frac{n^2+n(3-5i)+5i^2-7i+2}{n^2i+n(3i-4i^2)+4i^3-6i^2+2i}$$

The denominator is positive for $0 \le i \le n/2$. The numerator has roots

$$i = \frac{5n+7 \pm \sqrt{5n^2+10n+9}}{10}$$

Since $i \le n/2$, we take the root with the negative sign. When $i = \frac{n(5-\sqrt{5})}{10}$, the numerator is $-\frac{n(7\sqrt{5}-5)+20}{10}$ and hence negative; when $i = \frac{n(5-\sqrt{5})}{10}+1$, the numerator is $\frac{n(3\sqrt{5}+5)}{10}$ and hence positive. Therefore the peak of unimodality is the integer M such that $\frac{n(5-\sqrt{5})}{10} \le M \le \frac{n(5-\sqrt{5})}{10}+1$, unless the value of the numerator at M is negative. In that case, the peak of unimodality is at $M+1$. [1]

Conjecture 1 (Peter Shor) *The coefficients of $CPD(T)$ for any tree T with n vertices are unimodal with peak between $n/2$ and $n(1-1/\sqrt{5})$.*

Recall that the coefficients of the CP of the adjacency matrix are given by:

$$a_{n-k} = \begin{cases} (-1)^{n+k/2} \textit{ times the number of } k/2 \textit{ matchings in } T \textit{ when } k \textit{ is even} \\ 0 \textit{ otherwise} \end{cases}$$

Thus, every other term in the sequence is zero, and the non-zero terms alternate in sign. Hence the sequence of coefficients cannot be unimodal. However, if we take the absolute value of the non-zero terms, we have a sequence that could be unimodal.

[1]The author would like to thank Mark Hovey for many useful discussions.

Conjecture 2 *The sequence $\tilde{A}$ of the absolute values of the non-zero coefficients of $CPA(T)$ are unimodal with peak at the "same" place as that of the sequence $\tilde{D}$ of the (conjectured unimodal) coefficients of $CPD(T)$. That is, let m be the number of non-zero terms in the sequence of coefficients of the CP of the adjacency matrix of an n vertex tree. For any sequence S, let $p(S)$ be the place where the peak comes in the sequence. Then the integer closest to $\frac{p(\tilde{A})}{m}$ is the same as the integer closest to $\frac{p(\tilde{D})}{n}$.*

[2]

References

[1] Karen L. Collins, MIT Thesis, "Distance Matrices of Trees", 1986.

[2] Dragoš M. Cvetković, Michael Doob, Horst Sachs, Spectra of Graphs, Academic Press, New York, 1980, vol 87 in the series "Pure and Applied Math".

[3] Ron L. Graham, Alan J. Hoffman and H. Hosoya, "On the Distance Matrix of a Directed Graph", **Journal of Graph Theory**, Vol I, 85-88, 1977.

[4] Ron L. Graham and Laszlo Lovász, "Distance Matrix Polynomials of Trees", **Advances in Math.,** Vol 29, 60-88, July 1978.

[5] Ron L. Graham and H. O. Pollack, "Embedding Graphs in Squashed Cubes", **Springer Lecture Notes in Mathematics**, Vol 303, Springer-Verlag, Berlin/New York, 99-110, 1973.

[6] B. D. McKay, "On the Spectral Characterization of Trees", **Ars Combinatoria III**, 219-232, 1977.

[2]The author would like to thank the referee for suggesting that such a conjecture as this be made.