DETERMINANTS OF DISTANCE MATRICES OF TREES

LEQI ZHOU

1. Introduction

The object of this expository paper is to present a proof of Graham and Pollak's theorem. Some parts of the proof come from [1], while the rest are my own (see Remark 3.4).

Graham and Pollak's theorem finds the determinants of the distance matrices of trees, that for a tree on n vertices, the determinant is $-(n-1)(-2)^{n-2}$. It is simple and intriguing in that it shows the determinant depending only on the size of a given tree, independent of its structure or shape.

To discuss the theorem, let's look at the definitions first.

Definition 1.1 (distance)**.** The distance from one vertex to another is the length of the shortest path between them.

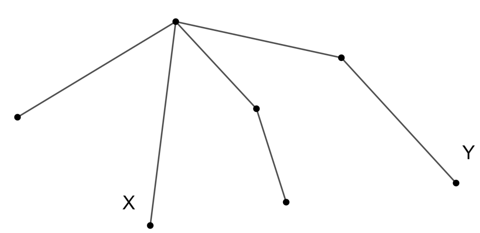


Figure 1. The distance between X and Y is 3.

Definition 1.2 (distance matrix)**.** For a graph G on n vertices, the distance matrix, denoted $D(G)$, of graph G is an n by n matrix whose (i, j)-th entry is the distance between vertices v_i and v_j.

Naturally, the distance matrix is a symmetric matrix whose diagonal consists entirely of zeros.

Date: December 5, 2020.

Before directly approaching the problem, we need to know some matrix properties that will be useful. The proof will be done by induction on the number of vertices n. Assuming the $(n-1)$-case, we want to know how the determinant of an $n \times n$ matrix (n-vertex tree) is related to those of $(n-1) \times (n-1)$ matrices ($(n-1)$-vertex trees). Actually, there is a method of expressing the determinant of a matrix by those of smaller ones inside it: cofactor expansion.

2. Cofactor Expansion

Cofactor expansion is a way of calculating the determinant of an $n \times n$ matrix M by taking the determinants of $(n-1) \times (n-1)$ matrices inside M.

Example 2.1. $n = 3$:

$$
\begin{aligned}
|A| &= \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} \\
&= a_{11}(+1) \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} + a_{12}(-1) \begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{13}(+1) \begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix} \\
&= a_{21}(-1) \begin{vmatrix} a_{12} & a_{13} \\ a_{32} & a_{33} \end{vmatrix} + a_{22}(+1) \begin{vmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{vmatrix} + a_{23}(-1) \begin{vmatrix} a_{11} & a_{12} \\ a_{31} & a_{32} \end{vmatrix} \\
&= a_{31}(+1) \begin{vmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{vmatrix} + a_{32}(-1) \begin{vmatrix} a_{11} & a_{13} \\ a_{21} & a_{23} \end{vmatrix} + a_{33}(+1) \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}
\end{aligned}
$$

To do the expansion, first pick out a row (or column) k, pick each entry $m_{k,j}$ in row k, cross out the row and column it sits in, and calculate $(-1)^{k+j} m_{k,j} \det(M_k^j)$ (M_k^j denotes the remaining $(n-1) \times (n-1)$ matrix after crossing out row k and column j). Note that in the expression of $\det(M)$ as the sum of entry products, $(-1)^{k+j} m_{k,j} \det(M_k^j)$ is the sum of all products containing $m_{k,j}$, where $(-1)^{j+k}$ denotes the change in sign function of the corresponding permutation after adding $m_{k,j}$. So by summing up $(-1)^{k+j} m_{k,j} \det(M_k^j)$ of all $m_{k,j}$ in row k, we obtain $\det(M)$.

Denote $\operatorname{cof}(M,k,j) = (-1)^{k+j}\det(M_k^j)$. The formulas are

$$\det(M) = \sum_{j=1}^{n} m_{k,j} \operatorname{cof}(M,k,j), \quad \det(M) = \sum_{i=1}^{n} m_{i,\ell} \operatorname{cof}(M,i,\ell)$$

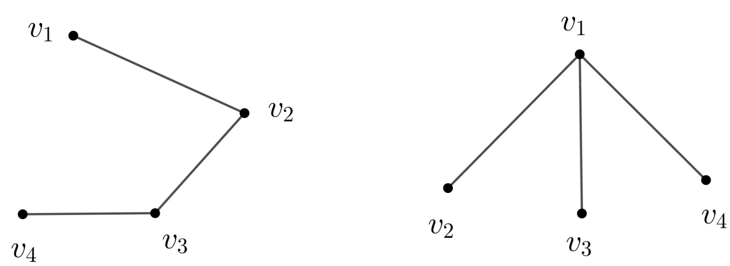
From the expression, we know it is better to use the formula when there are a lot of zeroes in a row or column.

3. The Proof

Theorem 3.1. *(Graham and Pollak, 1971) The determinant of the distance matrix D of a tree on n vertices is $-(n-1)(-2)^{n-2}$, or*

$$\det(D) = -(n-1)(-2)^{n-2}$$

Example 3.2. It's a little hard to believe, so we will look at examples with $n = 4$.



$$\begin{array}{c} \\ v_1 \\ v_2 \\ v_3 \\ v_4 \end{array} \begin{array}{c} \begin{array}{cccc} v_1 & v_2 & v_3 & v_4 \end{array} \\ \begin{bmatrix} 0 & 1 & 2 & 3 \\ 1 & 0 & 1 & 2 \\ 2 & 1 & 0 & 1 \\ 3 & 2 & 1 & 0 \end{bmatrix} \end{array} \qquad \begin{array}{c} \\ v_1 \\ v_2 \\ v_3 \\ v_4 \end{array} \begin{array}{c} \begin{array}{cccc} v_1 & v_2 & v_3 & v_4 \end{array} \\ \begin{bmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 2 & 2 \\ 1 & 2 & 0 & 2 \\ 1 & 2 & 2 & 0 \end{bmatrix} \end{array}$$

Figure 2. two trees and their respective distance matrices

The two trees have clearly different distance matrices, but the determinant of both matrices is -12, that is $-(4-1)(-2)^{4-2}$, as obtained by the theorem.

Proof. The proof is by induction, so base cases are shown that the theorem indeed works for a tree on 2 or 3 vertices. The 2-vertex tree has determinant -1, as obtained by $-(2-1)(-2)^{2-2}$.

For a tree on 3 vertices, the distance matrix is $D = \begin{bmatrix} 0 & 1 & 2 \\ 1 & 0 & 1 \\ 2 & 1 & 0 \end{bmatrix}$

$\det(D) = 4 = -(3-1)(-2)^{3-2}$.

Denote the distance matrix determinant if the n-vertex tree a_n. We now assume the theorem holds for trees on less than $n \geq 4$ vertices. It suffices to prove that $a_n = -(n-1)(-2)^{n-2}$.

Induction on n requires deleting some vertex from the tree to make it smaller. We know that there are at least two special vertices – leaves on a tree. More importantly, since a leaf has only one neighbor, deleting it won't change much of the graph: it is still a tree, and the distances between remaining vertices stay the same.

Changing the labeling doesn't change the determinant of the distance matrix. So for convenience, label the two leaves v_1 and v_n, and label vertices connected to them v_2 and v_{n-1} respectively.

$$D = \begin{bmatrix} 0 & d_{1,2} & \cdots & d_{1,n-1} & d_{1,n} \\ d_{2,1} & 0 & \cdots & d_{2,n-1} & d_{2,n} \\ \vdots & \vdots & & \vdots & \vdots \\ d_{n-1,1} & d_{n-1,2} & \cdots & 0 & d_{n-1,n} \\ d_{n,1} & d_{n,2} & \cdots & d_{n,n-1} & 0 \end{bmatrix}$$

where $d_{i,j}$ denotes the distance between v_i and v_j.

In order to use cofactor expansion, we want to do column (or row) operations that don't change the determinant while creating a lot of zeroes in one column (or row). The leaves enable us to do just that. Observe that for any $2 \leq i \leq n$, $d_{i,1} = d_{i,2}+1$, and for any $1 \leq i \leq n-1$, $d_{i,n} = d_{i,n-1} + 1$. Denote the i-th column vector $\mathbf{c}_i$. We have

$$\begin{aligned} (\mathbf{c}_1 - \mathbf{c}_2)^T + (\mathbf{c}_{n-1} - \mathbf{c}_n)^T &= \begin{bmatrix} -1 & 1 & \cdots & 1 & 1 \end{bmatrix} + \begin{bmatrix} -1 & -1 & \cdots & -1 & 1 \end{bmatrix} \\ &= \begin{bmatrix} -2 & 0 & \cdots & 0 & 2 \end{bmatrix} \end{aligned}$$

So we obtain $\det(D')$=$\det(D)$.

$$D' = \begin{bmatrix} -2 & d_{1,2} & \cdots & d_{1,n-1} & d_{1,n} \\ 0 & 0 & \cdots & d_{2,n-1} & d_{2,n} \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & d_{n-1,2} & \cdots & 0 & d_{n-1,n} \\ 2 & d_{n,2} & \cdots & d_{n,n-1} & 0 \end{bmatrix}$$

Cofactor expansion of the first column gives $\det(D) = \det(D') = -2\det(D_1^1) + 2(-1)^{n+1}\det(D_n^1)$. Remind that D_1^1 is the distance matrix of the $(n-1)$-vertex tree with leaf v_1 removed, so $\det(D_1^1)= a_{n-1}$. Thus

$$a_n = \det(D) = \det(D') = -2a_{n-1} + 2(-1)^{n+1}\det(D_n^1) \tag{3.3}$$

Remark 3.4. Yan and Yeh's proof [1] of the theorem then relies on the Desnanot-Jacobi identity

$$\det(D)\det\left(D_{1,n}^{1,n}\right) = \det\left(D_1^1\right)\det\left(D_n^n\right) - \det\left(D_1^n\right)\det\left(D_n^1\right)$$

in which $\det\left(D_{1,n}^{1,n}\right) = a_{n-2}$, $\det\left(D_1^1\right) = \det\left(D_n^n\right) = a_{n-1}$ and $\det\left(D_1^n\right) = \det\left(D_n^1\right)$. Substituting in (3.3) gives a quadratic equation in terms of $\det\left(D_1^n\right)$. But the quadratic equation has two solutions and I do not find it very obvious to eliminate one of them. The elimination is not included in [1]. Thus I tried to fix the problem and found that $\det\left(D_1^n\right)$ can be directly calculated. I obtained the same recurrence sequence (3.6) as in [2], with a probably shorter proof as the following.

Now we wish to obtain $\det(D_n^1)$, which is the original matrix with the n-th row and first column deleted. Note that the properties of leaf v_1 and its neighbor v_2 still exist, namely, $d_{1,i} = d_{2,i} + 1$ for every $2 \le i \le n$. Similarly $d_{n,i} = d_{n-1,i} + 1$, $1 \le i \le n-1$. Let $\mathbf{r}_i$ denote the i-th row vector of D_n^1, and we have $\mathbf{r}_1 - \mathbf{r}_2 = \mathbf{1}$. Thus, first swapping the first row to the last then adding $\mathbf{r}'_{n-1}$ to $\mathbf{r}'_n$, we obtain

$$\det(D_n^1) = \begin{vmatrix} 1 & 1 & \cdots & 1 & 1 \\ 0 & d_{2,3} & \cdots & d_{2,n-1} & d_{2,n} \\ \vdots & \vdots & & \vdots & \vdots \\ d_{n-1,2} & d_{n-1,3} & \cdots & 0 & d_{n-1,n} \end{vmatrix}$$

$$= (-1)^{n-2}\begin{vmatrix} 0 & d_{2,3} & \cdots & d_{2,n-1} & d_{2,n} \\ \vdots & \vdots & & \vdots & \vdots \\ d_{n-1,2} & d_{n-1,3} & \cdots & 0 & d_{n-1,n} \\ 1 & 1 & \cdots & 1 & 1 \end{vmatrix}$$

$$= (-1)^{n-2}\begin{vmatrix} 0 & d_{2,3} & \cdots & d_{2,n-1} & d_{2,n} \\ \vdots & \vdots & & \vdots & \vdots \\ d_{n-1,2} & d_{n-1,3} & \cdots & 0 & d_{n-1,n} \\ d_{n,2} & d_{n,3} & \cdots & d_{n,n-1} & 2 \end{vmatrix}$$

$$= (-1)^{n-2} \begin{vmatrix} 0 & d_{2,3} & \cdots & d_{2,n-1} & d_{2,n} \\ \vdots & \vdots & & \vdots & \vdots \\ d_{n-1,2} & d_{n-1,3} & \cdots & 0 & d_{n-1,n} \\ d_{n,2} & d_{n,3} & \cdots & d_{n,n-1} & 0 \end{vmatrix} \tag{3. 5}$$

$$+ (-1)^{n-2} \cdot 2 \begin{vmatrix} 0 & d_{2,3} & \cdots & d_{2,n-1} \\ \vdots & \vdots & & \vdots \\ d_{n-1,2} & d_{n-1,3} & \cdots & 0 \end{vmatrix}$$

Note that the two matrices in (3.5) are D_1^1 and $D_{1,n}^{1,n}$ respectively, so the determinants of both can be obtained by induction hypothesis.

Therefore,

$$\det(D_n^1) = (-1)^{n-2} \det(D_1^1) + (-1)^{n-2} \cdot 2 \det(D_{1,n}^{1,n})$$
$$= (-1)^{n-2} a_{n-1} + 2(-1)^{n-2} a_{n-2}$$

Remind that $a_n = -2a_{n-1} + 2(-1)^{n+1} \det(D_n^1)$. (3.3)

Now we have the recurrence sequence

$$a_n = -2a_{n-1} + 2(-1)^{n+1}((-1)^{n-2} a_{n-1} + 2(-1)^{n-2} a_{n-2})$$
$$= -4a_{n-1} - 4a_{n-2} \tag{3.6}$$

By the induction hypothesis, we easily obtain $a_n = -(n-1)(-2)^{n-2}$.

□

4. Questions and Remarks

Graham and Pollak's theorem leads to further inquiries, namely:

1. How does the determinant relate to the tree geometrically? (if at all)

2. Looking at the formula combinatorially, one might think that the formula is counting the number of ways to pick an edge then decide whether to include/exclude the rest of the edges. Is this the case?

3. Why doesn't the structure matter?

The result of the theorem has profound implications. First, as a direct connection to linear algebra, since the determinant is nonzero for all trees on $n \geq 2$ vertices, the distance matrices of all trees are nonsingular.

Second, of the n^{n-2} trees on n vertices, every single one of their distance matrices has the same determinant, while the result is not true for more general graphs. This suggests an underlying similarity between all trees, which has persisted in many other results regarding trees, and a great diversity of conjectures.

References

[1] W. Yan, Y.-N. Yeh, A simple proof of Graham and Pollak's theorem, J. Combin. Theory Ser. A 113 (2006) 892–893.

[2] Zhibin Du, Jean Yeh, Another simple proof of Graham and Pollak's theorem, Discrete Mathematics 343 (2020) 111994.

[3] R.L. Graham, H.O. Pollak, On the addressing problem for loop switching, Bell Syst. Tech. J. 50 (1971) 2495–2519.