

PROPERTIES OF THE DISTANCE MATRIX OF A TREE*

By F. T. BOESCH (*Bell Telephone Laboratories, Incorporated, Whippany, N. J.*)

In a recent paper, Hakimi and Yau [1] have given the necessary and sufficient conditions for a matrix D to be the distance matrix of a linear graph. They also give a computational technique for determining a tree realization if it exists. The purpose of this note is to exhibit certain interesting algebraic properties of the distance matrix of a tree. Two algorithms for tree realization emerge from these properties. However, these algorithms are not elaborated on as there are many computational techniques which are faster than the original method proposed by Hakimi and Yau.

Let G be a nonoriented weighted linear graph with $n + 1$ vertices and m edges, such that the weight assigned to edge i is a positive number r_i . The length of any path from node i to node j is the sum of the weights of the edges in that path; the distance d_{ij} between nodes i and j is the length of the shortest path from i to j , with d_{ii} taken as zero. The matrix whose elements are d_{ij} is the distance matrix D . This development is limited to graphs which are trees [2]. It is convenient to introduce the path matrix of a tree. If an arbitrary vertex r is taken as a reference vertex, then the path matrix P is defined as follows.

$$P = [p_{ij}]$$

where

$$p_{ij} = \begin{cases} +1 & \text{if the path from vertex } j \text{ to the reference vertex includes edge } i, \text{ and} \\ 0 & \text{if the path from vertex } j \text{ to the reference vertex does not include} \\ & \text{edge } i. \end{cases}$$

Notice that, due to the deletion of the reference vertex, there is an implicit renumbering of the vertices contained in the definition of the P matrix. Now although the path matrix is not a customary matrix in graph theory, it can be shown that P is actually the negative of the inverse of the reduced vertex-arc incidence matrix of the arborescence corresponding to reference vertex r [3], [4].

A standard numbering of the edges of a tree will also prove to be convenient. A numbering of the edges of a tree will be called an adjacent numbering if the first edge on the path from vertex i to the reference vertex is labeled edge i . Clearly, this defines a unique edge numbering.

If D is the distance matrix of a tree, then a number of properties of the tree are immediately obvious from D . Recall that a terminal vertex is a vertex with only one incident edge; it is easily shown that every tree has at least two terminal vertices. Furthermore vertex s is said to be the center of the tree if the maximum distance from s to any other vertex in the tree is not larger than this maximum distance for any other vertex in the tree. Certainly the center may be obtained by inspection of the distance

*Received August 12, 1966.

matrix. At least two terminal vertices may also be readily identified, as shown in Theorem I.

THEOREM I. *If D is the distance matrix of a tree and $d_{s,k}$ is a maximum of all $d_{i,j}$, then s and k are terminal vertices.*

Proof. If k is not a terminal vertex then there is a path from s through k to some terminal vertex t . However, this implies that $d_{i,s} > d_{k,s}$. The same reasoning is valid for vertex s .

Theorem I could be utilized as the basis of a new computational technique to determine the tree corresponding to a given distance matrix D . The method would proceed by identifying terminal vertices and then proceeding to the subtree obtained by deleting the corresponding terminal edges. This technique is very simple and proceeds much faster than the method given by Hakimi and Yau. Since the purpose of this note is to prove the algebraic properties, the details of this technique will be left to the reader.

Let R denote the m by m diagonal matrix of the weights of a tree, and let an arbitrary vertex r be chosen as a reference. The reference distance matrix Q is defined by the following congruence transformation:

$$Q = P'RP = [q_{ij}],$$

where P is the path matrix corresponding to reference r . It will be shown in the next theorem that Q has a simple topological interpretation.

THEOREM II.

q_{ii} = length of the path from vertex i to vertex r

q_{ij} = length of the path that is common to the paths from i to r and j to r .

Proof. If p_i denotes the i th column of P , and $'$ denotes transposition then

$$q_{ij} = p_i' R p_j.$$

Thus

$$q_{ij} = \sum_{k=1}^m r_k \alpha_k$$

where

$$\begin{aligned} \alpha_k &= 0 & \text{if } p_{ki} p_{kj} &= 0 \\ &= 1 & \text{otherwise.} \end{aligned}$$

Clearly α_k is nonzero if and only if the paths from vertices i and j to the reference both include edge k .

This theorem shows that the matrices Q and D are intimately related. In fact the following corollary emerges directly from Theorem II.

COROLLARY.

$$q_{ij} = \frac{1}{2}[d_{ir} + d_{jr} - d_{ij}].$$

Thus given a distance matrix D , the Q matrix is uniquely defined by a choice of a reference vertex. Now the congruence transformation which defines Q has an interesting

consequence. In the first place, Q is certainly positive definite since R is. Hence Q has a unique real triangular factorization as $\Delta'\Delta$ where Δ is lower triangular with non-negative diagonal elements [5]. Now clearly,

$$Q = (PR^{1/2})(R^{1/2}P),$$

where $R^{1/2}$ denotes the positive square root of R . Thus if P were known, a priori, to be triangular, then the uniqueness implies that $\Delta = R^{1/2}P$ or $P = R^{-1/2}\Delta$. The triangular property of P , however, may be given a simple interpretation.

THEOREM III. *If Q is the reference distance matrix of a tree with an adjacent numbering such that*

$$q_{11} \geq q_{22} \geq q_{33} \cdots \geq q_{nn},$$

then P must be lower triangular.

Proof. Suppose P is not lower triangular. Then there exists a nonzero element p_{sk} with $s < k$. Thus the path from vertex k to the reference vertex includes edge s . The adjacent numbering implies, therefore, that the length of the path from k to the reference is strictly larger than the length of the path from s to the reference, i.e., $q_{kk} > q_{ss}$. However, this contradicts the assumed ordering of the diagonal elements of Q . Hence P is lower triangular.

In summary, it may be observed that Theorem I and Theorem III provide for two distinct algorithms for realizing the distance matrix of a tree. The method which emerges from Theorem I is extremely fast and may be utilized for hand calculations with large order matrices.¹ The algorithm which emerges from Theorem III is certainly longer. However, the purpose of this note is merely to exhibit the algebraic properties of these matrices.

REFERENCES

- [1] S. L. Hakimi and S. S. Yau, *Distance matrix of a graph and its realizability*, Quart. Appl. Math. **12**, 305-317 (1965)
- [2] C. Berge, *The theory of graphs and its applications*, translated by Alison Doig, Wiley, New York, 1962
- [3] F. H. Branin, *The relation between Kron's method and the classical methods of network analysis*, IRE WESCON Conv. Rec. Part 2, Aug. 1959, pp. 3-28
- [4] F. T. Boesch, *On the synthesis of resistor N -ports*, Polytechnic Institute of Brooklyn, RADC-TDR-63-294, PIBMR1-1143-63, Apr. 1963
- [5] F. R. Gantmacher, *Matrix theory*, Vol. I, Chelsea, New York, 1959

¹The reviewer pointed out that there is another fast algorithm based on the concept of a minimal tree.