

## THE DISTANCE MATRIX OF A GRAPH AND ITS TREE REALIZATION\*

BY

A. N. PATRINOS AND S. L. HAKIMI

*Northwestern University, Evanston, Illinois*

**Abstract.** The results of Hakimi and Yau and others in the realization of a distance matrix are generalized to graphs (digraphs) whose branches (arcs) may have negative weights. Conditions under which such matrices have a tree, hypertree or directed tree realization are given, uniqueness of these realizations is discussed and algorithms for their construction are indicated.

**1. Notation.** A number of definitions are given so that results will be presented in a self-contained manner. A *graph*  $G = (V, B)$  consists of a finite non-empty set  $V = \{v_1, v_2, \dots, v_n\}$  of *vertices* and a set  $B = \{b_1, b_2, \dots, b_m\}$  of unordered pairs of distinct vertices of  $V$ . Each such pair  $b_k = e(v_i, v_j)$  of vertices is a *branch* of  $G$  and is said to be incident at both  $v_i$  and  $v_j$ . A *directed graph* or *digraph*  $G = (V, A)$  consists of a finite non-empty set  $V = \{v_1, v_2, \dots, v_n\}$  of vertices and a set  $A = \{a_1, a_2, \dots, a_m\}$  of ordered pairs of distinct vertices of  $V$ . Each such pair  $a_k = e(v_i, v_j)$  of vertices is an *arc* of  $G$ , is directed from  $v_i$  to  $v_j$  and is incident at both  $v_i$  and  $v_j$ . A *subgraph* of a graph (digraph)  $G$  is a graph (digraph) which has all its vertices and branches (arcs) in  $G$ .

The *degree* of a vertex  $v_i$  in  $G$ , denoted  $\deg(v_i, G)$ , is the number of branches (arcs) incident at  $v_i$  in  $G$ . The *outdegree* of a vertex  $v_i$  in digraph  $G$ , denoted  $\text{outdeg}(v_i, G)$ , is equal to the number of arcs incident at  $v_i$  in  $G$  and directed away from  $v_i$ . The *indegree* of  $v_i$ , denoted  $\text{indeg}(v_i, G)$ , is equal to the number of arcs incident at  $v_i$  in  $G$  and directed towards  $v_i$ . A *weighted graph* (digraph) is a graph (digraph) together with a function which assigns a real number  $w_i$  to each branch  $b_i$  (arc  $a_i$ ). All graphs (digraphs) presented here are weighted.

An *edge-sequence* in a graph (digraph) between two vertices  $v_i$  and  $v_j$  is an alternating sequence of vertices and branches (arcs)  $v_i, b_1, v_2, b_2, \dots, b_r, v_j$  beginning and ending with  $v_i$  and  $v_j$ , in which each branch (arc) is incident at the vertex preceding and the vertex following it. A *path* from  $v_i$  to  $v_j$  is the set of all branches (arcs) in an edge-sequence between  $v_i$  and  $v_j$ . A *directed path* in a digraph is a path in which each arc is directed from the vertex preceding it to the vertex following it in the corresponding edge-sequence. A path or directed path is called *elementary* if all vertices in the edge-sequence are distinct. A path (directed path) is a *circuit* (*cycle*) if the first and last vertex in the edge-sequence are the same and all others distinct. The *length* of a path (directed path) is the sum of the weights of the branches (arcs) in it. A *connected graph* (digraph) is a graph (digraph) in which every pair of vertices is joined by a path.

---

\* Received July 13, 1971; revised version received October 9, 1971. This work was supported by the USAF Office of Scientific Research under Grant AFOSR-71-2103.

A *tree* (*directed tree*) is a connected graph (digraph) containing no circuits and thus any two vertices are joined by a unique elementary path. In directed trees not all vertices are joined by a directed path. In fact, if a directed path exists from  $v_i$  to  $v_j$ , there is no directed path from  $v_j$  to  $v_i$ . In a tree, if each branch is replaced by two oppositely directed arcs, the digraph so constructed is a *hypertree*. Two such arcs in the hypertree form an *elementary pair*, and the sum of the weights of the two arcs is the *weight of the elementary pair*. In a hypertree there is a unique elementary directed path between any pair of vertices.

The *distance*  $d(v_i, v_j)$  of vertex  $v_j$  from vertex  $v_i$  in a graph (digraph) is the length of a shortest (i.e. minimum sum of weights) elementary path (directed path) from  $v_i$  to  $v_j$ . Clearly  $d(v_i, v_i) = d(v_j, v_i)$  in graphs while in general  $d(v_i, v_j) \neq d(v_j, v_i)$  in digraphs. We also have  $d(v_i, v_i) = 0$  and  $d(v_i, v_j) = \infty$  if there is no (directed) path from  $v_i$  to  $v_j$ . A *distance matrix*  $D(V_1) = [d_{ij}]$  of a graph (digraph)  $G = (V, B)$ ,  $V_1 = \{v_{k_1}, v_{k_2}, \dots, v_{k_n}\} \subseteq V$ , is an  $n \times n$  matrix in which entry  $d_{ij}$  ( $i, j = 1, 2, \dots, n$ ) is the distance of vertex  $v_{k_i}$  from vertex  $v_{k_j}$  ( $v_{k_i}, v_{k_j} \in V_1$ ). If  $v \in V_1$  then  $v$  is an *external vertex*, otherwise an *internal vertex*. Any vertex of degree one in a graph (digraph) is a *terminal* vertex. By a *realization* of an  $n \times n$  matrix  $D = [d_{ij}]$  we mean a graph (digraph)  $G = (V, B)$ ,  $|V| \geq n$ , such that for some  $V_1 \subseteq V$ ,  $|V_1| = n$ , we shall have  $D(V_1) = D$ . All graphs (digraphs) have distance matrices, graphs having symmetric ones and digraphs, in general, asymmetric ones. All entries in the distance matrix of a connected graph are finite. Connected digraphs may have infinite entries. A branch (arc) in a graph (digraph) is *redundant* if its removal results in a graph (digraph) with the same distance matrix. An internal vertex  $v$  in a graph (digraph) is *redundant* if it has  $\deg(v, G) < 3$  ( $\text{indeg}(v, G) < 2$  or  $\text{outdeg}(v, G) < 2$ ). The *nullity* of a connected graph (digraph) is equal to  $|B| - |V| + 1$  ( $|A| - |V| + 1$ ) and thus is zero if the graph (digraph) is a (directed) tree.

**2. Introduction.** Given a weighted graph (digraph), algorithms are available for computing the distance matrix  $D$  of (a subset of) its vertices. Of these, the most efficient is due to Floyd [1]. The algorithms fail, in general, if the graph has a branch with a negative weight or if the digraph has a cycle whose length is a negative number.

A number of papers have also been published on the realizability of a given  $n \times n$  matrix  $D$  by a graph (digraph). Hakimi and Yau [2] gave necessary and sufficient conditions for an  $n \times n$  symmetric matrix  $D$  with non-negative entries to be the distance matrix of a graph. They defined as 'optimum' that realization which has a minimum total sum of weights and proved that a tree realization, if one exists, is the unique optimum realization. Goldman [3] and Murchland [4] extended some of these results to digraphs. Generalizing the above results, we have proved that any (symmetric) square matrix with zero diagonal elements is the distance matrix of some (graph) digraph.

Zaretskii [5, 6] gave necessary and sufficient conditions for the existence of a unique unweighted tree with  $n$  terminal vertices whose distance matrix equals a given matrix of order  $n$ . Simoes-Pereira [7] gave, without proof, a weaker statement of Theorem 2 presented in this work. Theorem 2 also provides a generalization of Zaretskii's results to the weighted case. Boesch [8], considering strictly non-negative weighted graphs, gave some properties of the distance matrix of a tree and suggested two algorithms for a tree realization. We indicate here that one of these algorithms (the one derived from theorem II of his work) can be successfully used in the general case. Shay [9] introduced the 'hypertree' and gave a necessary condition for its realization. We have completed his

work on the hypertree. Finally, we attacked the case of the distance matrix and its realization as a directed tree.

Almost all previous work restricted itself to non-negative entries in the given matrix  $D$  and non-negative weights in its realization. We have placed no such restriction in this work and admitted negative weights. Further, we introduce non-redundant internal vertices if they permit a realization which could not have been achieved otherwise. With these considerations in mind, the objectives of this paper are:

1. to find necessary and sufficient conditions for a given matrix to have a tree, hypertree or directed tree realization,
2. to find whether the above realizations are unique,
3. to indicate algorithms for construction of the tree, hypertree or directed tree if such realizations exist.

**3. The distance matrix and its realizability.** Given an  $n \times n$  symmetric matrix  $D = [d_{ij}]$ , necessary and sufficient conditions for the existence of an  $n$ -vertex graph  $G$  with non-negative weights having  $D$  as its distance matrix were given by Hakimi and Yau [2]. Specifically:

- 1)  $d_{ii} = 0$  for all  $i$ ,
- 2)  $d_{ij} + d_{ik} \geq d_{jk}$  for all  $i, j$  and  $k$ .

The  $n$ -vertex graph  $G$  realizing  $D$  can be constructed as follows: pick  $n$  vertices, labeling them  $v_1, v_2, \dots, v_n$ , and for every entry  $d_{ij}$  ( $i \neq j$ ) of  $D$  draw a branch  $e(v_i, v_j)$  assigning to it the weight  $d_{ij}$ .

Since  $d_{ij} + d_{jk} \geq d_{ik}$  and  $d_{ki} + d_{ij} \geq d_{kj}$  for any  $i, j$  and  $k$  we have  $d_{ij} \geq 0$ . This implies that  $D$  must contain strictly non-negative entries. If no restriction is placed on the type of weights in a realization of  $D$  then we can state:

**THEOREM 1:** Any  $n \times n$  symmetric matrix  $D = [d_{ij}]$  with zero diagonal elements is a distance matrix of some graph  $G$ .

*Proof:* Consider the  $n \times n$  matrix  $D' = [d'_{ij}]$ , where

$$d'_{ij} = \begin{cases} d_{ij} + a_i + a_j & \text{if } i \neq j \\ 0 & \text{if } i = j \end{cases}$$

and

$$a_k = \max \left\{ \max_{r, s = 1, 2, \dots, n} \frac{1}{2}(d_{rs} - (d_{rk} + d_{ks})), \quad 0 \right\}.$$

If  $d_{rk} + d_{ks} \geq d_{rs}$  for some  $r, k$  and  $s$  then clearly  $d'_{rk} + d'_{ks} \geq d'_{rs}$ . If  $d_{rk} + d_{ks} < d_{rs}$  for some  $r, k$  and  $s$  then  $0 < \frac{1}{2}(d_{rs} - (d_{rk} + d_{ks})) \leq a_k$ , or  $(d_{rk} + a_r + a_k) + (d_{ks} + a_k + a_s)$ , or  $d'_{rk} \leq d'_{rk} + d'_{ks}$ . Thus there exists an  $n$ -vertex graph  $G'$  which has  $D'$  as its distance matrix. Let  $G'$  be such a graph and  $v'_1, v'_2, \dots, v'_n$  be its vertices. To construct a graph  $G$  with distance matrix  $D$  add to  $G'$  vertex  $v_i$  and connect it with  $v'_i$  through a branch  $e(v_i, v'_i)$  of weight  $-a_i$  (for all  $i$ ). We have in  $G$ :  $d(v_i, v_j) = d(v_i, v'_i) + d(v'_i, v'_j) + d(v'_j, v_j) = -a_i + d'_{ij} - a_j = d_{ij}$  (for all  $i$  and  $j$ ). This proves the theorem.

**4. The distance matrix and the tree realization.** By virtue of Theorem 1 let us call a symmetric matrix  $D$  with zero diagonal elements a 'distance matrix'. We have proved that there exists at least one graph  $G$  realizing distance matrix  $D$ . If  $G[D]$  is the set of all

graphs realizing  $D$  we would like to know if there exists a tree  $t$  such that  $t \in G[D]$ . The following theorem provides an answer to this question. It was given by Simoes-Pereira [7] in a slightly weaker form. The interested reader should also consult Zaretskii [5, 6] for the unweighted case. Let us set  $\varphi(x, y, z) = 1$  if at least two of the numbers  $x, y$  and  $z$  are equal. Since any  $3 \times 3$  distance matrix is tree-realizable we shall assume that  $n \geq 4$ .

**THEOREM 2:** *Given an  $n \times n$  ( $n \geq 4$ ) distance matrix  $D = [d_{ij}]$ , a necessary and sufficient condition for  $D$  to be tree-realizable is that  $\varphi(d_{ik} + d_{il}, d_{il} + d_{jk}, d_{ij} + d_{kl}) = 1$  for all distinct  $i, j, k, l$ .*

*Proof: Necessity.* Assume matrix  $D$  is tree-realizable. Let  $t$  be some such tree. Pick any four external vertices  $v_i, v_j, v_k, v_l$  in  $t$  and consider the subtree connecting these four vertices. This subtree necessarily corresponds to at least one of the trees illustrated in Fig. 1. In all three cases we have

$$d_{m_1, m_2} + d_{m_3, m_4} = d_{m_1, m_3} + d_{m_2, m_4}$$

where  $m_1, m_2, m_3$  and  $m_4$  is some permutation of  $i, j, k$  and  $l$ . Thus  $\varphi(d_{ik} + d_{il}, d_{il} + d_{jk}, d_{ij} + d_{kl}) = 1$ ; hence the theorem's necessity conditions follow. *Sufficiency.* Let  $n = 4$  and  $d_{i_1, i_2} + d_{i_3, i_4} = d_{i_1, i_3} + d_{i_2, i_4}$ , where  $i_1, i_2, i_3$  and  $i_4$  is a permutation of 1, 2, 3 and 4. The tree realizing  $D$  is shown in Fig. 2.

Let us assume that the theorem is correct if the order of  $D$  is  $n - 1$ . Consider an  $n \times n$  distance matrix  $D$  satisfying the theorem's conditions. Let  $D'$  be the  $(n - 1) \times (n - 1)$  leading principal submatrix of  $D$ . According to the induction hypothesis the theorem is true for  $D'$  and let  $T_{n-1}$  be some such tree with  $(n - 1)$  external vertices realizing  $D'$ . In  $T_{n-1}$  let all external vertices be made terminal vertices by using branches of weight zero, and let all branches of weight zero joining two internal vertices be shorted (i.e. eliminate the branch by identifying the two vertices by a single internal vertex). Assume further that all internal vertices in  $T_{n-1}$  are non-redundant.

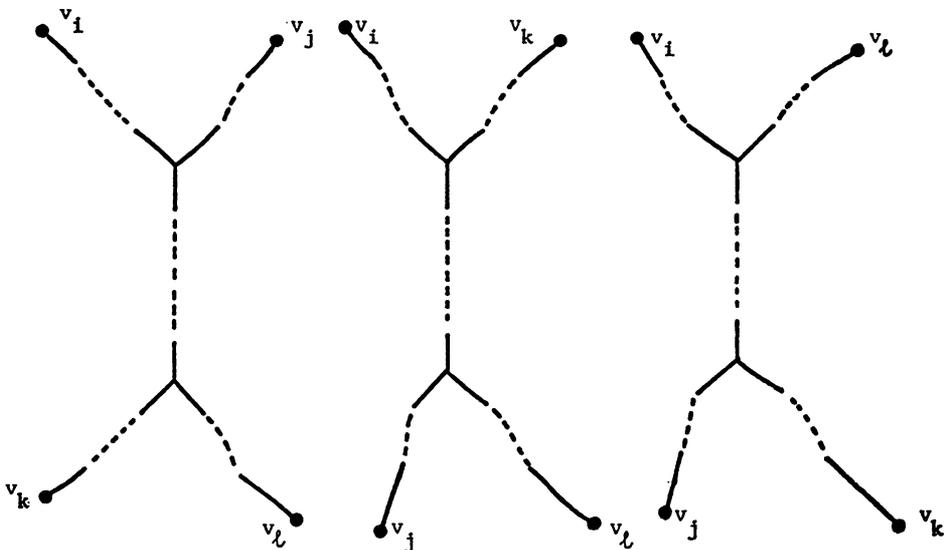


FIG. 1.

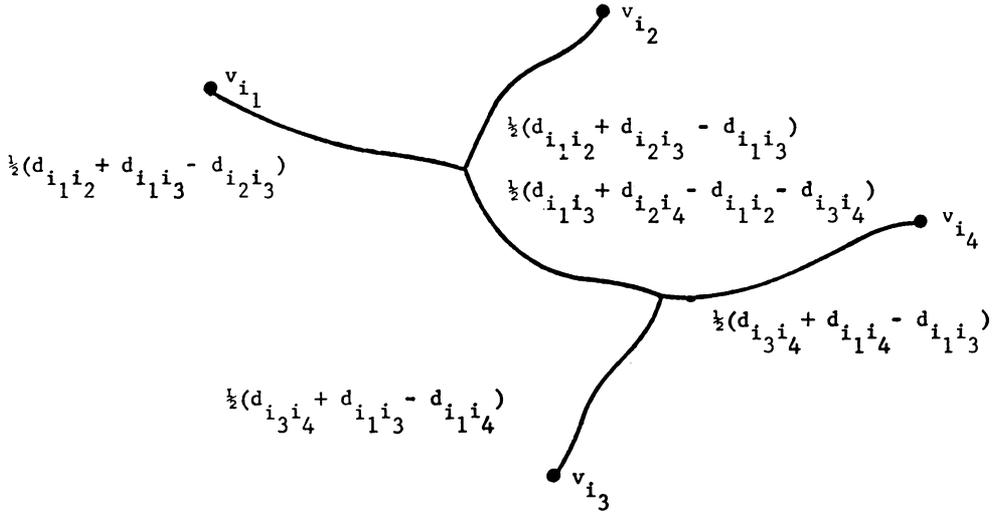


FIG. 2.

Let  $u$  be an internal vertex in  $T_{n-1}$  (clearly such a vertex always exists since  $n \geq 4$  and all external vertices are terminal vertices in  $T_{n-1}$ ). Since  $\text{deg}(u, T_{n-1}) \geq 3$ ,  $T_{n-1}$  can be divided as shown in Fig. 3, where external vertices  $v_1, v_2$  and  $v_3$  are assumed to be, without loss in generality, in the subtrees indicated. Define  $L(v_i, u)$  as the maximal subtree of  $T_{n-1}$  containing vertex  $v_i$  and having  $u$  as a terminal vertex.  $L(v_1, u)$  and  $L(v_2, u)$  are indicated in Fig. 3.

Case A:  $d_{1n} + d_{23} = d_{13} + d_{2n} \neq d_{12} + d_{3n}$ . Insert vertex  $v_n$  as shown in Fig. 4, with  $w = d_{n1} - d(u, v_1)$ . Clearly  $d(v_n, v_1) = d_{n1}$  and

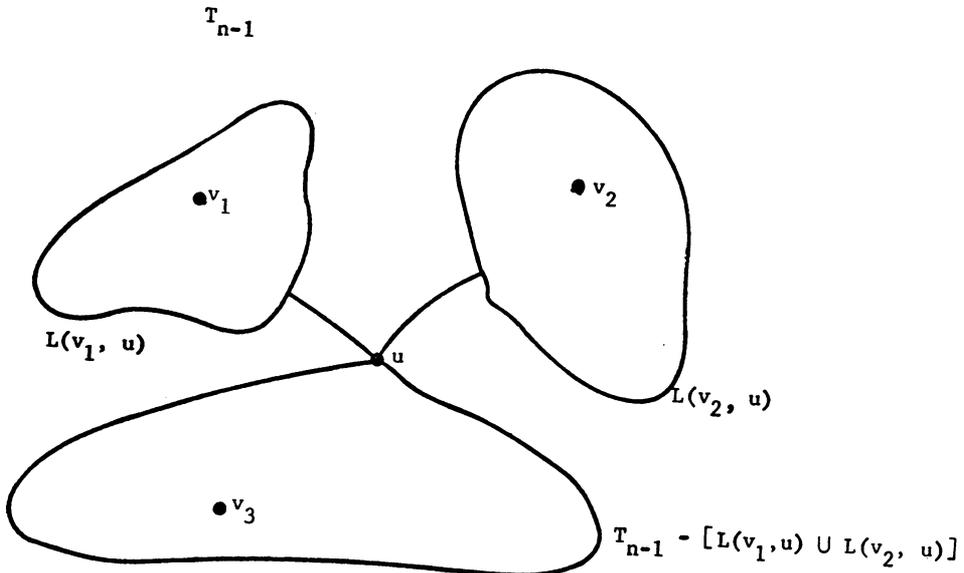


FIG. 3.

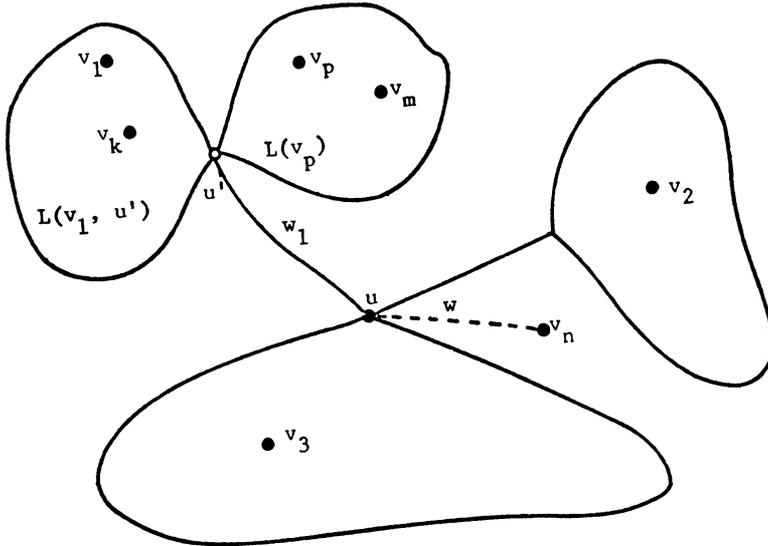


FIG. 4.

$$d(v_n, v_2) = w + d(u, v_2) = d_{n1} - d(u, v_1) + d(u, v_2) = d_{n1} - [d(u, v_1) + d(u, v_3)] + [d(u, v_2) + d(u, v_3)] = d_{n1} - d_{13} + d_{23} = d_{n2} .$$

If  $v_1$  is the only external vertex in  $L(v_1, u)$  then  $v_1 \equiv u'$  and we have shown that  $d(v_n, v_i) = d_{ni}$  for all  $v_i \in L(v_1, u)$ . Assume, therefore, that  $L(v_1, u)$  contains more than one external vertex. Then  $u'$  is an internal vertex and hence  $w_1 \neq 0$ . Since  $\text{deg}(u', T_{n-1}) \geq 3$  there exists at least one external vertex  $v_p$  in  $L(v_1, u) - [L(v_1, u') \cup e(u', u)] = L(v_p)$ . If  $v_m$  is an arbitrary external vertex in  $L(v_p)$  then, since  $w_1 \neq 0$ , we have

$$d_{12} + d_{3m} = d_{13} + d_{2m} \neq d_{1m} + d_{23} .$$

The above relation, together with

$$\begin{aligned} d_{13} + d_{2n} &= d_{1n} + d_{23} \neq d_{12} + d_{3n} , \\ \varphi(d_{12} + d_{mn}, d_{1n} + d_{2m}, d_{1m} + d_{2n}) &= 1, \\ \varphi(d_{13} + d_{mn}, d_{1n} + d_{3m}, d_{1m} + d_{3n}) &= 1, \\ \varphi(d_{23} + d_{mn}, d_{2n} + d_{3m}, d_{2m} + d_{3n}) &= 1, \end{aligned}$$

yield for all possible cases:  $d_{nm} = d_{1n} + d_{2m} - d_{12}$ . However, in the graph of Fig. 4,  $d(v_n, v_m) = w + d(u, v_m) = d_{n1} - d(u, v_1) + d(u, v_m) = d_{n1} - [d(u, v_1) + d(u, v_2)] + [d(u, v_m) + d(u, v_2)] = d_{n1} - d_{12} + d_{2m}$ . Thus  $d(v_n, v_m) = d_{nm}$  for all external vertices  $v_m \in L(v_p)$ . Since by assumption

$$d_{13} + d_{2n} = d_{1n} + d_{23} \neq d_{12} + d_{3n}$$

and by induction hypothesis

$$d_{12} + d_{3m} = d_{12} + d_{3m} \neq d_{1m} + d_{23} ,$$

then

$$d_{13} + d_{2n} \neq d_{13} + d_{2m} - d_{3m} + d_{3n}$$

or

$$d_{2n} + d_{3m} \neq d_{2m} + d_{3n}.$$

But in the graph of Fig. 4 we have

$$d(v_2, v_3) + d(v_m, v_n) = d(v_2, v_n) + d(v_3, v_m) \quad \text{or} \quad d_{23} + d_{mn} = d_{2n} + d_{3m}.$$

Hence

$$d_{23} + d_{mn} = d_{2n} + d_{3m} \neq d_{2m} + d_{3n}.$$

Then if  $v_k (\neq v_1)$  is an arbitrary external vertex in  $L(v_1, u')$ , the previous relation together with

$$d_{2k} + d_{3m} = d_{2m} + d_{3k} \neq d_{mk} + d_{23} \quad (\text{since } w_1 \neq 0),$$

$$\varphi(d_{2k} + d_{mn}, d_{kn} + d_{2m}, d_{km} + d_{2n}) = 1,$$

$$\varphi(d_{3k} + d_{mn}, d_{kn} + d_{3m}, d_{km} + d_{3n}) = 1,$$

$$\varphi(d_{23} + d_{kn}, d_{2n} + d_{3k}, d_{2k} + d_{3n}) = 1$$

yields, for all possible cases,  $d_{nk} = d_{nm} + d_{2k} - d_{2m}$ . But in the graph of Fig. 4,

$$d(v_n, v_k) = w + d(u, v_k) = d_{n1} - d(u, v_1) + d(u, v_k) = d_{n1} - [d(u, v_1) + d(u, v_2)]$$

$$+ [d(u, v_k) + d(u, v_2)] = d_{n1} - d_{12} + d_{2k} = d_{nm} + d_{2k} - d_{2m}.$$

This proves that for any external vertex  $v_i \in L(v_1, u)$  we have  $d(v_n, v_i) = d_{ni}$ . Similarly we can prove that for any external vertex  $v_j$  in  $L(v_2, u)$  we have  $d(v_n, v_j) = d_{nj}$ .

Let  $S = \{v_{i_1}, v_{i_2}, \dots, v_{i_s}\}$  be the set of all external vertices in  $T_{n-1} - [L(v_1, u) \cup L(v_2, u)]$ . Clearly  $\delta \leq n - 3$ . Construct an  $(\delta + 2) \times (\delta + 2)$  distance matrix  $D^* = [d_{ij}^*]$  as follows: set

$$d_{km}^* = d_{i_k i_m} \quad \text{for all } v_{i_k}, v_{i_m} \in S,$$

$$d_{k\delta+1}^* = d_{i_k n} \quad \text{for all } v_{i_k} \in S,$$

$$d_{\delta+1\delta+2}^* = w,$$

$$d_{k\delta+2}^* = d(v_{i_k}, u) \quad \text{for all } v_{i_k} \in S.$$

In  $D^*$  rows and columns  $\delta + 1$  and  $\delta + 2$  correspond to vertices  $v_n$  and  $u$  respectively. Clearly

$$\varphi(d_{qr}^* + d_{si}^*, d_{qs}^* + d_{ri}^*, d_{qi}^* + d_{rs}^*) = 1,$$

$$\varphi(d_{qr}^* + d_{s\delta+1}^*, d_{qs}^* + d_{r\delta+1}^*, d_{q\delta+1}^* + d_{rs}^*) = 1$$

for all  $v_{i_q}, v_{i_r}, v_{i_s}, v_{i_t}$  in  $S$  by the theorem's conditions. We also have  $\varphi(d_{qr}^* + d_{s\delta+2}^*, d_{qs}^* + d_{r\delta+2}^*, d_{q\delta+2}^* + d_{rs}^*) = 1$  for all  $v_{i_q}, v_{i_r}, v_{i_s}$  in  $S$  by the fact that these distances are distances in a subtree of  $T_{n-1}$ . Since

$$\varphi(d_{1i_q} + d_{ni_r}, d_{1i_r} + d_{ni_q}, d_{1n} + d_{i_q i_r}) = 1$$

for all  $v_{i_q}, v_r$ , in  $S$  we can write

$$\varphi(d_{1i_q} + d_{ni_r} - d(u, v_1), d_{1i_r} + d_{ni_q} - d(u, v_1), d_{1n} + d_{i_q i_r} - d(u, v_1)) = 1$$

or

$$\varphi(d(u, v_{i_q}) + d_{ni_r}, d(u, v_{i_r}) + d_{ni_q}, w + d_{i_q i_r}) = 1$$

or

$$\varphi(d_{q\delta+2}^* + d_{r\delta+1}^*, d_{r\delta+2}^* + d_{q\delta+1}^*, d_{\delta+1\delta+2}^* + d_{qr}^*) = 1.$$

Thus the distance matrix  $D^*$  satisfies the theorem's conditions and is of order at most  $n - 1$ . Hence  $D^*$  is tree-realizable and so is  $D$ .

The cases:

$$d_{12} + d_{3n} = d_{1n} + d_{23} \neq d_{13} + d_{2n},$$

$$d_{12} + d_{3n} = d_{13} + d_{2n} \neq d_{1n} + d_{23}$$

can be treated similarly by considering  $L(v_1, u)$ ,  $L(v_3, u)$  and  $L(v_2, u)$ ,  $L(v_3, u)$  respectively.

*Case B:*  $d_{1n} + d_{23} = d_{13} + d_{2n} = d_{12} + d_{3n}$ ,  $d_{i_1 n} + d_{i_2 i_3} = d_{i_1 i_2} + d_{i_3 n} = d_{i_1 i_3} + d_{i_2 n}$  for all  $v_{i_1} \in L(v_1, u)$ ,  $v_{i_2} \in L(v_2, u)$  and  $v_{i_3} \in T_{n-1} - [L(v_1, u) \cup L(v_2, u)]$  (otherwise reduce to case A). Then, setting  $v_n$  as before in  $T_{n-1}$  with  $w = d_{1n} - d(u, v_1)$ , we have

$$d(v_n, v_1) = d_{n1}, \quad d(v_n, v_2) = d_{n2}, \quad d(v_n, v_3) = d_{n3}$$

and for any external vertex  $v_k$  in  $T_{n-1}$  we can immediately show that  $d(v_n, v_k) = d_{nk}$ . Thus  $D$  has a tree realization. This completes the proof of the theorem.

Define an *elementary contraction* in a graph  $G$  by:

- a) shorting a branch of weight zero joining two internal vertices, or
- b) shorting a branch of weight zero joining an external vertex with an internal vertex.

The tree  $T_b$  in Fig. 5 is obtained from tree  $T_a$  by an elementary contraction of type a, while  $T_c$  is obtained from  $T_a$  by an elementary contraction of type b.

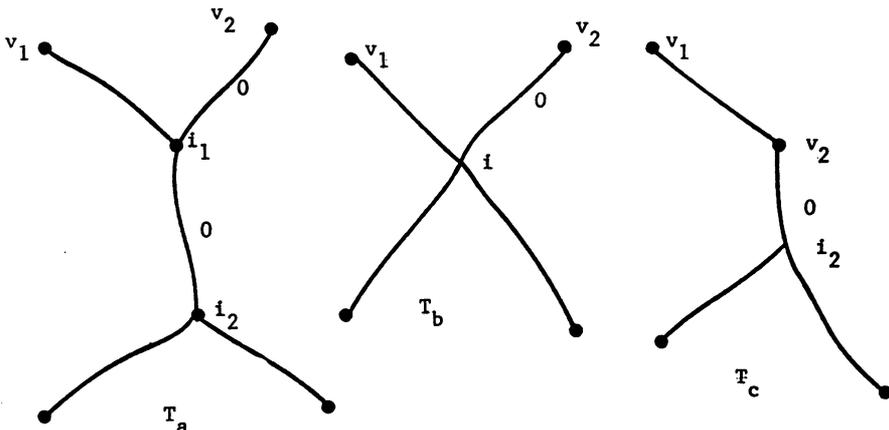


FIG. 5.

Define an *elementary expansion* in a graph  $G$  by:

- a) replacing an internal vertex by two internal vertices joined by a branch of weight zero, or
- b) replacing an external vertex of degree  $\geq 2$  in  $G$  by an internal vertex and joining it to the external vertex with a branch of weight zero.

Thus in Fig. 5 tree  $T_a$  is obtained from  $T_b$  by an elementary expansion of type *a*, while  $T_c$  is obtained from  $T_b$  by an elementary expansion of type *b*.

For the purpose of the following theorem, if graph  $G_2$  is obtained from  $G_1$  by a sequence of elementary contractions and expansions it will be considered 'identical' to  $G_1$ . In that sense trees  $T_a$ ,  $T_b$  and  $T_c$  of Fig. 5 are considered identical. For the proof, the reader is referred to [12].

**THEOREM 3:** *If distance matrix  $D = [d_{ij}]$  is realizable as a tree  $t$ , then  $t$  is the unique circuitless realization of  $D$  (without redundant internal vertices).*

There are some elegant methods of constructing the tree if the given matrix is tree-realizable. Although these were designed primarily for non-negative weight realizations, the following algorithm (derived from Boesch [8, theorem II]) equally applies for the general case.

Given an  $n \times n$  distance matrix  $D = [d_{ij}]$  which is tree-realizable, choose a reference vertex, say  $v_n$ , and construct the  $(n - 1) \times (n - 1)$  matrix  $Q = [q_{ij}]$  where:

$$q_{ij} = \frac{1}{2}(d_{in} + d_{jn} - d_{ij}), \quad i, j \neq n.$$

The reader can easily verify that:  $q_{ii}$  = length of the elementary path between vertex  $v_i$  and vertex  $v_n$  ( $= d_{in}$ ), and  $q_{ij}$  = length of the path that is common to the elementary paths from  $v_i$  to  $v_n$  and from  $v_j$  to  $v_n$ . Having  $Q$ , the reader can convince himself that it is simple to draw the tree [12].

**5. The distance matrix and the hypertree realization.** Given an  $n \times n$  matrix  $D = [d_{ij}]$ , necessary and sufficient conditions for the existence of an  $n$ -vertex digraph  $G$  having  $D$  as its distance matrix are:

- 1)  $d_{ii} = 0$
- 2)  $d_{ij} + d_{jk} \geq d_{ik}$  for all  $i, j$  and  $k$ .

The  $n$ -vertex digraph  $G$  realizing  $D$  can be constructed as follows: pick  $n$  vertices, labeling them  $v_1, v_2, \dots, v_n$ , and for every finite entry  $d_{ij}$  ( $i \neq j$ ) of  $D$  draw an arc  $e(v_i, v_j)$  assigning to it the weight  $d_{ij}$ . The construction shows that  $G$  has no cycles of negative length. The condition  $d_{ij} \geq 0$  for all  $i$  and  $j$  is necessary only if a non-negative weight realization is required. If no such restriction is placed on the type of weights in a realization of  $D$  then, we can state the following theorem [12].

**THEOREM 4:** *Any  $n \times n$  matrix  $D = [d_{ij}]$  with zero diagonal elements is the distance matrix of some digraph  $G$ .*

Let us define a strongly connected digraph  $G$  as a digraph in which for every two vertices  $v_i$  and  $v_j$ ,  $d(v_i, v_j)$  and  $d(v_j, v_i)$  are finite. Thus the distance matrix  $D = [d_{ij}]$  of a strongly connected digraph is asymmetric, has zero diagonal elements and finite entries. By Theorem 4 any square matrix  $D = [d_{ij}]$  with zero diagonal elements and finite entries is the distance matrix of some strongly connected digraph  $G$ . Hence such a matrix may be called 'distance matrix' (in this section not necessarily symmetric) and  $G$  its realization.

Let  $D = [d_{ij}]$  be the distance matrix of a hypertree  $h$ . Consider four external vertices  $v_i, v_k, v_j$  and  $v_l$  in  $h$ . A check will reveal that  $d_{i,i_2} + d_{i,i_4} = d_{i,i_4} + d_{i,i_2}$  and  $d_{i,i_1} + d_{i,i_3} = d_{i,i_3} + d_{i,i_1}$ , for some permutation  $i_1, i_2, i_3, i_4$  of  $i, j, k$  and  $l$ . By adding the two equations we obtain

$$(d_{i,i_2} + d_{i,i_1}) + (d_{i,i_4} + d_{i,i_3}) = (d_{i,i_4} + d_{i,i_1}) + (d_{i,i_2} + d_{i,i_3}).$$

By Theorem 2, this proves that  $D + D^t$  is the distance matrix of a tree, where  $D^t$  is the transpose of  $D$ . Let  $t_1$  be the tree realizing  $D + D^t$ . Construct a tree  $t_2$  from  $h$  by replacing each elementary pair in  $h$  by a branch of weight equal to the weight of the elementary pair. If  $v_i$  and  $v_j$  are two external vertices in  $t_2$ , then by construction  $d_{t_2}(v_i, v_j) = d_{ij} + d_{ji}$ . Hence the distance matrix of the external vertices of  $t_2$  is  $D + D^t$ . Since  $D + D^t$  must have a unique circuitless realization (Theorem 3)  $t_1$  and  $t_2$  are identical.

**THEOREM 5:** *Necessary and sufficient conditions for a distance matrix  $D = [d_{ij}]$  to be realizable as a hypertree are:*

- 1) that  $D + D^t$  be the distance matrix of a tree,
- 2) that  $d_{ij} + d_{jk} + d_{ki} = d_{ik} + d_{ki} + d_{ji}$  for all distinct  $i, j, k$ .

*Proof:* The necessity of condition 1 has been proven above. The necessity of condition 2 should be clear. The sufficiency will be proved by induction on the order of  $D$ . The case  $n = 3$  is illustrated in Fig. 6. Suppose the theorem is true if  $D$  is of order  $n - 1$ . Consider a distance matrix  $D$  of order  $n$  satisfying the theorem's conditions and let  $t$  be the tree realization of  $D^t = D + D^t$ . Make external vertex  $v_n$  a terminal vertex in  $t$  (if it is not one already) by an elementary expansion in  $t$ . Let  $u$  be the vertex adjacent to  $v_n$  in  $t$ . We shall assume for the moment that  $u$  is either an external vertex or an internal vertex with  $\text{deg}(u, t) \geq 4$ , with the weight of branch  $e(u, v_n)$  equal to  $w_n$ . Construct the hypertree  $h_{n-1}$  whose distance matrix is the  $(n - 1) \times (n - 1)$  leading principal submatrix of  $D$ . Add vertex  $v_n$  to  $h_{n-1}$ , connecting it with  $u$  through the elementary pair

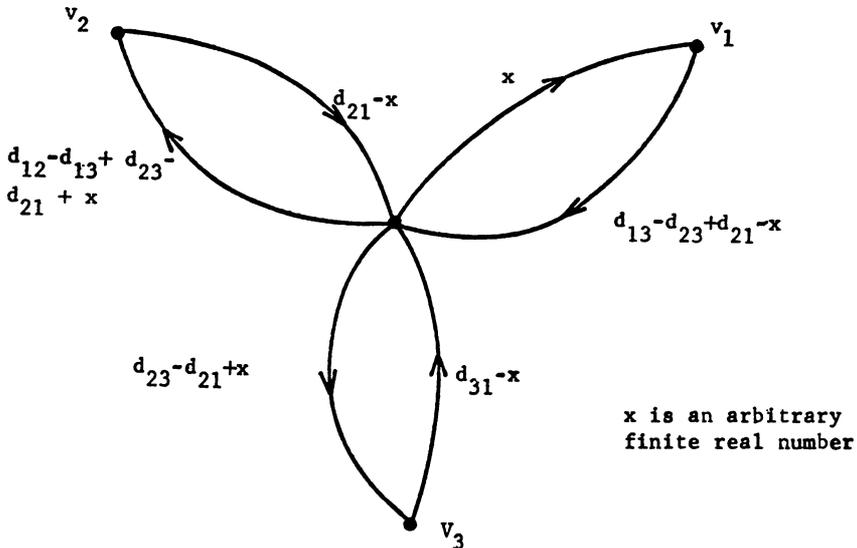


FIG. 6.

with weights as shown in Fig. 7. Call  $h_n$  the resulting hypertree. Let us denote by  $d'(v_i, v_j)$  the distance between vertices  $v_i, v_j$  in  $t$  and by  $d(v_i, v_j)$  their distance in  $h_n$ . By construction  $d(v_n, v_1) = d_{n1}$ . We also have  $d(v_1, v_n) = d(v_1, u) + w_n - d_{n1} + d(u, v_1) = [d(v_1, u) + w_n + d(u, v_1)] - d_{n1} = (d_{1n} + d_{n1}) - d_{n1} = d_{1n}$ . Let  $v_k$  be an arbitrary external vertex in  $h_n$ . We must show that  $d(v_k, v_n) = d_{kn}$  and  $d(v_n, v_k) = d_{nk}$ . By construction we have in  $t$

$$d'(v_n, v_1) + d'(u, v_k) = d'(v_n, v_k) + d'(u, v_1)$$

or, since  $D' = D + D^t$ ,

$$d_{n1} + d_{1n} + d(u, v_k) + d(v_k, u) = d_{nk} + d_{kn} + d(u, v_1) + d(v_1, u). \tag{1}$$

Since  $u, v_1, v_k$  are vertices of a hypertree  $h_n$ ,

$$d(u, v_1) + d(v_1, v_k) + d(v_k, u) = d(u, v_k) + d(v_k, v_1) + d(v_1, u)$$

and by the induction hypothesis this can be written

$$d(u, v_1) + d_{1k} + d(v_k, u) = d(u, v_k) + d_{k1} + d(v_1, u). \tag{2}$$

By condition 2 of the theorem,

$$d_{n1} + d_{1k} + d_{kn} = d_{nk} + d_{k1} + d_{1n}. \tag{3}$$

Eqs. (1), (2) and (3) give

$$d_{nk} = d_{n1} + d(u, v_k) - d(u, v_1).$$

But in  $h_n, d(v_n, v_k) = d(v_n, u) + d(u, v_k) = d_{n1} + d(u, v_k) - d(u, v_1) = d_{nk}$ . Similarly

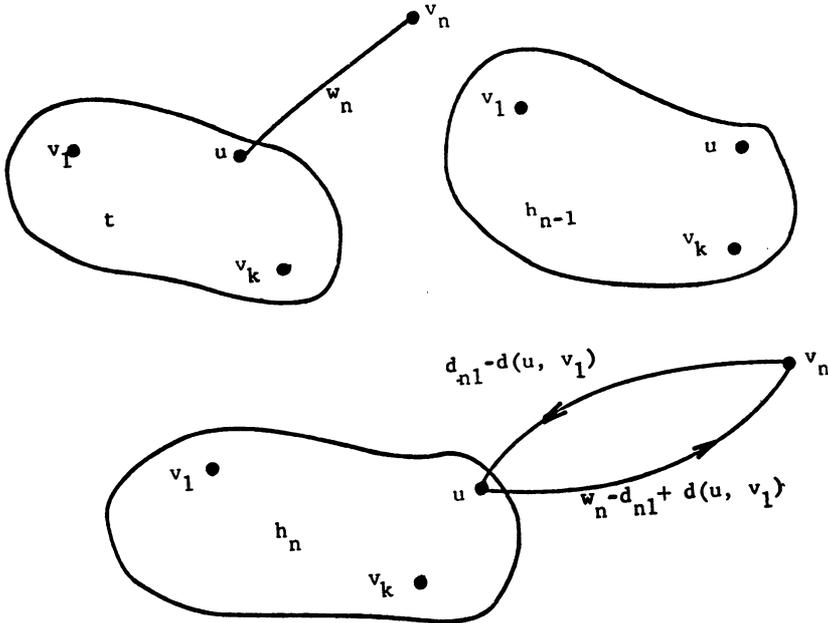


FIG. 7.

$d(v_k, v_n) = d_{kn}$ . Hence the theorem follows. The case where  $u$  is an internal vertex of degree 3 in  $t$  is treated similarly as illustrated in Fig. 8, where  $a$  is an arbitrary weight.

**COROLLARY:** All hypertrees realizing a given distance matrix  $D$  have the same total weight.

**THEOREM 6 [12]:** If  $D$  is realizable as a hypertree  $h$  which has no redundant internal vertices then  $h$  is geometrically unique (only the weights of the corresponding arcs may differ).

The proof of Theorem 5 suggests an algorithm for constructing the hypertree  $h_n$  given its distance matrix  $D$ . The algorithm is simple once the undirected tree with distance matrix  $D + D'$  has been drawn.

**6. The distance matrix and the directed tree realization.** Given a digraph  $G$  with  $m$  vertices, the reachability matrix  $R = [r_{ij}]$  between vertices  $v_{k_1}, v_{k_2}, \dots, v_{k_n}$  of  $G$  is an  $n \times n$  matrix ( $n \leq m$ ) defined as follows:

- $r_{ii} = 1$  for all  $i$ ,
- $r_{ij} = 1$  if there exists a directed path from vertex  $v_{k_i}$  to vertex  $v_{k_j}$ ,
- $r_{ij} = 0$  otherwise ( $i \neq j$ ).

Note that  $R$  does not depend on the weights of  $G$ . Define the *incidence set* of a vertex  $u$  in digraph  $G$  as the set of all vertices  $u_i$  in  $G$  such that  $e(u, u_i)$  is an arc of  $G$ , and the *external incidence set* of  $u$  in  $G$  as the set of all external vertices  $v_i$  in  $G$  such that a directed path exists from  $u$  to  $v_i$  in  $G$ . Define a *block* in  $G$  as a maximal subgraph  $B$  of  $G$  such that every two arcs in  $B$  lie on a common circuit.

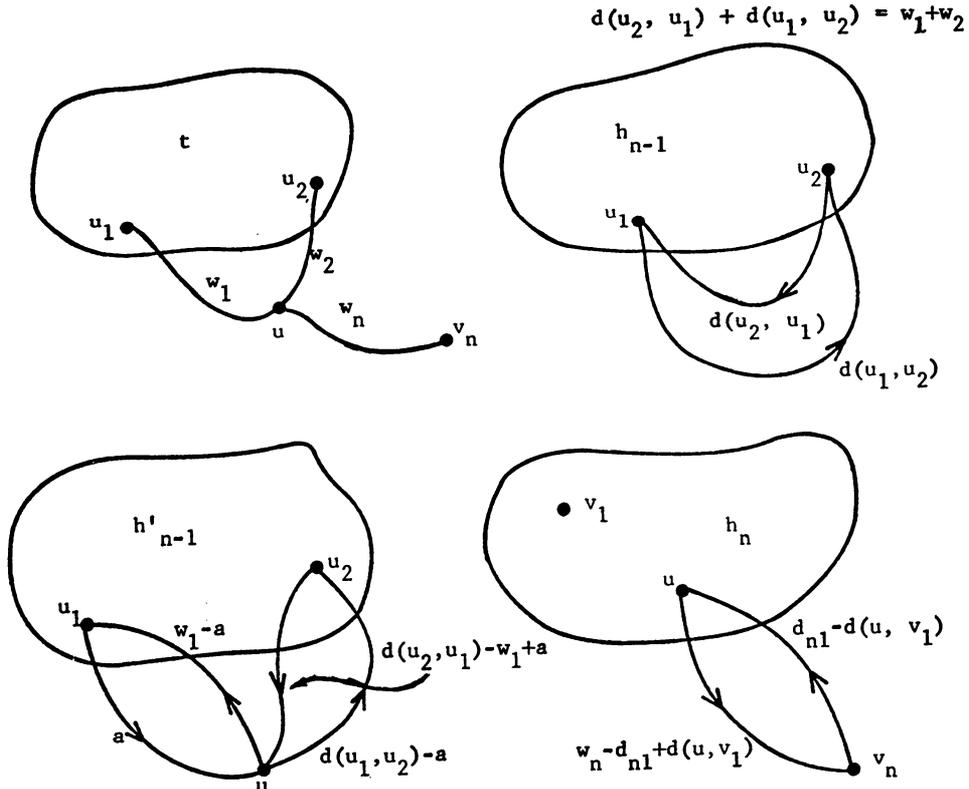


FIG. 8.

Given an  $n \times n$  binary matrix  $R = [r_{ij}]$  with unity diagonal elements, we would like to determine whether there exists a directed tree having  $R$  as its reachability matrix. A necessary and sufficient condition for  $R$  to be the reachability matrix of some digraph  $G$  is  $r_{ij} \cdot r_{jk} \leq r_{ik}$  for all  $i, j$  and  $k$ . Such a digraph can be constructed as follows: pick  $n$  vertices, labeling them  $v_1, v_2, \dots, v_n$ , and connect vertices  $v_i$  and  $v_j$  ( $i \neq j$ ) with an arc directed from  $v_i$  to  $v_j$  if and only if  $r_{ij} = 1$ . Clearly, the resulting digraph  $G_R$  is an  $n$ -vertex realization of the reachability matrix  $R$ . To avoid unnecessary complications we shall assume that  $G_R$  is a connected digraph.

Remove all redundant arcs from  $G_R$  (an arc being redundant if its removal results in a digraph with the same reachability matrix). If the resulting digraph  $G_R^0$  is circuitless, then  $R$  is the reachability matrix of the directed tree  $G_R^0$ . Otherwise let  $B_1$  be a block in  $G_R^0$  and  $v_1, v_2, \dots, v_r$  be vertices in  $B_1$  having the same incidence set in  $B_1$ , say  $\{v'_1, v'_2, \dots, v'_s\}$  ( $r, s \geq 2$ ). Apply the following nullity reduction cycle: remove all arcs  $e(v_k, v'_j)$ ,  $k = 1, 2, \dots, r, j = 1, 2, \dots, s$ , add an internal vertex  $i_1$  and draw arcs  $e(v_k, i_1), e(i_1, v'_j)$   $k = 1, 2, \dots, r, j = 1, 2, \dots, s$ . The resulting digraph  $G_R^1$  has the same reachability matrix as  $G_R^0$  between external vertices but the nullity has been reduced by  $r \cdot s - (r + s) + 1$ . If  $G_R^1$  is circuitless then we are finished. Otherwise a new block  $B_2$  can be located and the above cycle repeated. We can now state the main theorem.

**THEOREM 7:** *If  $R$  is an  $n \times n$  binary matrix with unity diagonal elements, then  $R$  is the reachability matrix of a directed tree if and only if:*

- 1)  $r_{ij} + r_{ji} \leq 1$   $i \neq j$
- 2)  $r_{ij} \cdot r_{jk} \leq r_{ik}$  for all  $i, j$  and  $k$
- 3) in a block  $B_{i+1}$  of  $G_R^i$  containing no redundant elements there exist at least two vertices with the same incidence set in  $B_{i+1}$ .

*Proof:* The necessity of conditions 1 and 2 should be clear. The necessity of condition 3 is proved as follows. Let  $B_{i+1}$  be a block in  $G_R^i$  such that there exist no two vertices in  $B_{i+1}$  with the same incidence set in  $B_{i+1}$ . Further, let  $T_R$  be a directed tree realization of the reachability matrix  $R$ . We first claim that if  $u$  is a vertex in  $B_{i+1}$  then we shall have either  $\text{indeg}(u, B_{i+1}) = 0$  or  $\text{outdeg}(u, B_{i+1}) = 0$ . We prove the claim as follows. If  $u$  is an external vertex, the claim is obvious since  $R$  has a directed tree realization  $T_R$  and  $G_R^i$  has no redundant elements. Let  $u$  be an internal vertex in  $B_{i+1}$  with  $\text{indeg}(u, B_{i+1}) \geq 1$  and  $u_1, u'_1$  be vertices in  $B_{i+1}$  such that  $e(u_1, u), e(u, u'_1) \in B_{i+1}$  (see Fig. 9). Since  $B_{i+1}$  is a block, there exists a circuit  $C_u$  in  $B_{i+1}$  containing the above two arcs. Also, the arc in  $C_u$  incident at  $u_1$  other than  $e(u_1, u)$  is directed towards  $u_1$ , since  $u$  has been introduced by a nullity reduction cycle. Thus  $\text{indeg}(u_1, B_{i+1}) \geq 1$ . Hence  $u_1$  must be an internal vertex. If the same reasoning is repeated with  $u_1$  using the same circuit  $C_u$ , etc., we conclude that  $C_u$  is a cycle. Thus in  $B_{i+1}$  we have a cycle  $C_u$  containing only internal vertices. Since originally, in  $G_R^0$ , no such cycle existed, it must have been introduced by a sequence of nullity reduction cycles, which is a contradiction. This proves the claim.

Let  $S_e = \{v_1, v_2, \dots, v_r\}, S_i = \{i_1, i_2, \dots, i_s\}$  be the sets of external and internal vertices in  $B_{i+1}$  respectively. Since, by hypothesis,  $G_R^i$  contains no redundant elements, the previous claim tells us that we can find for every  $i_k \in S_i$  an external vertex  $v_{i_k}$  not in  $B_{i+1}$  such that a directed path exists from  $v_{i_k}$  to  $i_k$ . Let  $S_{e,i} = \{v_{i_1}, v_{i_2}, \dots, v_{i_s}\}$  be the set of such external vertices corresponding to  $S_i$ . Consider now the set of external vertices  $S = \{v_1, v_2, \dots, v_r, v_{i_1}, v_{i_2}, \dots, v_{i_s}\}$  in  $T_R$  and let  $T_R^S$  be the subtree of  $T_R$

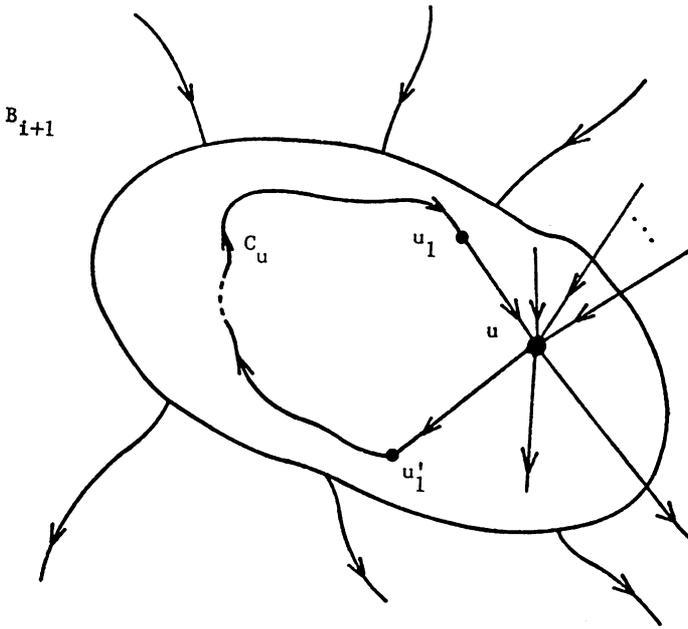


FIG. 9.

joining the vertices in  $S$ . If  $T_R^S$  contains no internal vertices, the block  $B_{i+1}$  in  $G_R^i$  necessarily contains redundant elements, which is a contradiction. Thus  $T_R^S$  contains at least one internal vertex and, therefore there exist at least two vertices in  $S$  of the same external incidence set in  $T_R^S$ . Without loss of generality, let us assume that  $v_1$  and  $v_i$  are two such vertices having  $v_2, v_3$  as their external incidence set in  $T_R^S$ . This implies that in  $G_R^i$  we have directed paths from  $v_1$  and  $v_i$  to  $v_2$  and  $v_3$ . Since by the previous claim the only directed path in  $B_{i+1}$  is an arc, the two vertices  $v_1$  and  $v_i$  in  $B_{i+1}$  have the same incidence set  $v_2, v_3$  in  $B_{i+1}$ , a contradiction. This proves the necessity of condition 3. *Sufficiency:* using condition 2 we can always construct an  $n$ -vertex digraph  $G_R^0$  without redundant arcs having reachability matrix  $R$ . Condition 3 gives a means of successively reducing the nullity of the digraph. Since the graph is finite and the nullity reduction cycle can always be applied as long as blocks exist in the digraph, the result will be a directed tree with reachability matrix  $R$ . This completes the proof of Theorem 7.

**COROLLARY:** *If  $R$  has a directed tree realization and  $u_k$  in  $B_{i+1}$  is the vertex of maximum incidence set in  $B_{i+1}$  then there exists at least one other vertex  $u_i$  in  $B_{i+1}$  of the same incidence set as  $u_k$  in  $B_{i+1}$ .*

Given an  $n \times n$  matrix  $D = [d_{ij}]$  construct an  $n \times n$  binary matrix  $R_D = [r_{ij}]$  as follows: set  $r_{ij} = 1$  if  $d_{ij}$  is finite and  $r_{ij} = 0$  otherwise.

**THEOREM 8:** *Given an  $n \times n$  matrix  $D = [d_{ij}]$ , necessary and sufficient conditions for  $D$  to be the distance matrix of a directed tree are that:*

- 1)  $R_D$  be the reachability matrix of a directed tree,
- 2)  $d_{ik} = d_{ij} + d_{jk}$  if  $d_{ik}, d_{ij}$  and  $d_{jk}$  are finite ( $i, j, k$  distinct),
- 3)  $d_{ik} + d_{il} = d_{il} + d_{lk}$  if  $d_{ik}, d_{il}, d_{il}$  and  $d_{lk}$  are finite ( $i, j, k, l$  distinct).

*Proof:* The necessity of the above conditions should be obvious. The sufficiency is proved as follows. If  $G_R^0$  is the  $n$ -vertex digraph without redundant arcs realizing  $R_D$ , set the weight of arc  $e(v_i, v_j)$  equal to  $d_{ij}$ , for all  $i$  and  $j$  ( $i \neq j$ ). By condition 2,  $D(G_R^0) = D$ . Then if  $u_1, u_2, \dots, u_r$  are the vertices of the same incidence set  $u'_1, u'_2, \dots, u'_s$  in  $B_k$  of  $G_R^{k-1}$  and  $i_k$  is the internal vertex introduced by the nullity reduction cycle, assign weights to the arcs incident at  $i_k$  as follows: set weight of arc  $e(u_1, i_k) = a$  (where  $a$  is an arbitrary finite number), weight of  $e(i_k, u'_j) = \text{weight of } e(u_1, u'_j) \text{ in } G_R^{k-1} - a$ , and weight of  $e(u_l, i_k) = \text{weight of } e(u_l, u'_1) \text{ in } G_R^{k-1} - \text{weight of } e(i_k, u'_1)$  for  $l = 2, 3, \dots, r$  and  $j = 1, 2, \dots, s$ . Clearly the distance matrix between the external vertices of  $G_R^{k-1}$  equals the distance matrix between the external vertices of  $G_R^k$  by construction and condition 3.

Theorem 7, its corollary, and Theorem 8 suggest an algorithm for constructing a directed tree realization of  $D$ , if one exists.

It can be shown [12] that if  $T$  is a directed tree realization of the  $n \times n$  matrix  $D$ , without redundant internal vertices, then  $T$  is geometrically unique.

**7. Conclusion.** Necessary and sufficient conditions for realizing a distance matrix as a tree, hypertree and directed tree were given and proved. Algorithms for their realization were suggested and the uniqueness of these realizations was discussed. It was found that the tree realization of a distance matrix is unique, the hypertree realization is geometrically unique with a constant total sum of weights and the directed tree realization is only geometrically unique.

The basic problem of finding the 'optimum' realization (i.e. the realization with the minimum total sum of weights) of a distance matrix as an undirected graph or digraph continues to be unsolved in the general case. The problem of finding distances in graphs with negative weight branches and in digraphs with negative weight cycles is also unsolved.

#### REFERENCES

- [1] R. W. Floyd, *Algorithm 97: shortest path*, Comm. ACM 5, 345 (1962)
- [2] S. L. Hakimi and S. S. Yau, *Distance matrix of a graph and its realizability*, Quart. Appl. Math., 12, 305-317 (1965)
- [3] A. J. Goldman, *Realizing the distance matrix of a graph*, J. Res. Nat. Bur. Standards B70, 153-54 (1966)
- [4] J. D. Murchland, *A fixed matrix method for all shortest distances in a directed graph and for the inverse problem*, Report LBS-TNT-91, Transport Network Theory Unit, London Business School, January 1969
- [5] K. A. Zaretskii, *Constructing a tree on the basis of a set of distances between hanging vertices*, Usp. Mat. Nauk, Akademiya Nauk SSSR, Moskovshoe Matematicheskoe Obshchestvo, 20, 90-92 (1965) (in Russian)
- [6] J. Turner and W. H. Kautz, *A survey of progress in graph theory in the Soviet Union*, SIAM Review, 12, 17 (1970)
- [7] J. Simoes-Pereira, *Some results on the tree realization of a distance matrix*, in *Theory of graphs* (P. Rosenstiehl, ed.), Gordon and Breach, N. Y., (1967), pp. 383-88
- [8] F. T. Boesch, *Properties of the distance matrix of a tree*, Quart. Appl. Math. 16, 607-09 (1969)
- [9] B. P. Shay, *Some considerations of distances in a linear directed graph*, M. S. thesis, E. E. Dept., Northwestern University, Evanston, Ill., 1965
- [10] Frank Harary, *Graph theory*, Addison-Wesley, Reading, Mass., 1969
- [11] Claude Berge, *The theory of graphs* (English translation), Methuen, 1962
- [12] A. N. Patrinos, *The distance matrix of a graph and its tree realization*, M. S. thesis, E. E. Dept., Northwestern Univ., Evanston, Ill., 1971