

## DISTANCE MATRIX OF A GRAPH AND ITS REALIZABILITY\*

By

S. L. HAKIMI AND S. S. YAU

*Northwestern University, Evanston, Illinois*

**Abstract.** The distances in a linear graph are described by a distance matrix  $D$ . The realizability of a given  $D$  by a linear graph is discussed and conditions under which the realization of  $D$  is unique are established. The optimum realization of  $D$ , (i.e., the realization of  $D$  with "minimum total length"), is investigated. A procedure is given by which a tree realization of  $D$  can be found, if such a realization exists. Finally, it is shown that a tree realization, if it exists, is unique and is the optimum realization of  $D$ .

**1. Introduction.** A number of interesting papers on the distances in a linear graph have been published [1, 2, 3]. These papers deal with very specific problems concerning the distances in a graph. For example, Moore [1] described a method for finding the length of the shortest path (the distance) between any pair of vertices in a graph. Kruskal and Prim [2, 3] presented methods for finding a tree of a "weighted" graph whose total weight is minimum without examining all trees of that graph. This paper is devoted to a study of general properties of distances in a linear graph. More specifically, we will study distances in a graph by defining a distance matrix  $D$ , deriving necessary and sufficient conditions for the realizability of a given  $D$  by a graph, and describing conditions under which  $D$  has a unique realization. In general, it is shown that the realization is not unique, especially if we allow additional "internal vertices". Therefore, we define an optimum realization of  $D$ . The optimum realization of  $D$  is a realization of  $D$  with "minimum total weight". This optimum realization problem may be loosely described as a graph theoretic version of Steiner's problem [4]. Steiner's problem is: given a number of points in a two dimensional Euclidean space, find the "shortest highway network" that interconnects these points. Steiner's problem is reduced to the problem considered by Prim [3], if we do not allow the roads in the highway network to intersect at any points other than the given points. However, there is no known method of solving Steiner's problem in its general form. It should be noted that, in contrast with Steiner's problem, the distances in this paper, are not Euclidean.

An application of some of the results in this paper is discussed in Section V. For an introduction to the theory of linear graphs, the reader is referred to a book by Seshu and Reed [5].

**2. Distance matrix and its realizability.** Let  $G$  be a weighted linear graph with  $n$  vertices (nodes), i.e., a linear graph to each of whose elements (arcs) a real nonnegative number is attached.<sup>1</sup> The distance  $d_{ij} = d_{ji}$  between vertices  $v_i$  and  $v_j$  is the length of the shortest path in  $G$  between  $v_i$  and  $v_j$ , where the length of a path is defined as the sum of the weights of the arcs in that path. Clearly, between every pair of vertices in  $G$ ,

---

\*Received November 15, 1963; revised manuscript received February 28, 1964. This work was supported by the U. S. Army Research Office (Durham) and the Air Force Office of Scientific Research.

<sup>1</sup>Although the paper is concerned with nonoriented graphs, some of the results are readily generalized to oriented graphs.

there is a uniquely defined distance. The distance matrix of an  $n$ -vertex graph is an  $(n \times n)$  symmetric matrix  $D = [d_{ij}]$  whose typical entry  $d_{ij}$  is defined as

$$d_{ij} = \begin{cases} \text{distance between } v_i \text{ and } v_j, & \text{if } i \neq j, \\ 0, & \text{if } i = j. \end{cases} \tag{1}$$

For example the distance matrix of the graph of Fig. 1 is

$$D = \begin{pmatrix} 0 & 1 & 2 & 3 & 2 \\ 1 & 0 & 3 & 2 & 1 \\ 2 & 3 & 0 & 2 & 3 \\ 3 & 2 & 2 & 0 & 1 \\ 2 & 1 & 3 & 1 & 0 \end{pmatrix}. \tag{2}$$

One may immediately ask the following question: given a square symmetric matrix with real and nonnegative entries, under what circumstances is the given matrix the distance matrix of a linear graph? The following theorem answers this question.

*Theorem 1.* The necessary and sufficient conditions for an  $(n \times n)$  symmetric matrix  $D$  with nonnegative entries to be a distance matrix are as follows:

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

*Proof.* The necessity of condition (a) is obvious. To prove the necessity of condition (b), suppose that  $d_{ij} + d_{jk} < d_{ik}$  for some vertices  $i, j$ , and  $k$  in a graph  $G$ . ( $1 \leq i, j, k \leq n$ ). Denote the corresponding paths in  $G$  by  $p_{ij}$ ,  $p_{jk}$ , and  $p_{ik}$ . The paths  $p_{ij}$  and  $p_{jk}$  together form a path  $p'_{ik}$  between  $v_i$  and  $v_k$ . The length of  $p'_{ik}$  is equal to  $d_{ij} + d_{jk}$  which is shorter than the length of the path  $p_{ik}$ . Since  $p_{ik}$  was assumed to be the shortest path between  $v_i$  and  $v_k$ , this is a contradiction; condition (b) is therefore necessary.

To prove sufficiency, construct an  $n$ -vertex graph as follows: pick  $n$  points, labeling them  $v_1, v_2, \dots, v_n$ . For every entry  $d_{ij} = d_{ji}$ , ( $i \neq j$ ), of  $D$ , we draw an arc  $e'(v_i, v_j)$  joining  $v_i$  and  $v_j$ , and give it the weight  $d_{ij}$ . Let the resulting graph be  $G'$ . We would like to show that if  $D$  satisfies conditions (a) and (b), then  $G'$  will have the distance matrix  $D$ . To this end, denote the distance between  $v_i$  and  $v_j$  in  $G'$  by  $d'_{ij}$ , ( $1 \leq i, j \leq n$ ). Clearly  $d'_{ij} \leq d_{ij}$  for all  $i$  and  $j$ . If  $G'$  does not have the distance matrix  $D$ , then, for some  $i$  and  $j$ ,  $d'_{ij} < d_{ij}$ . If  $d'_{ij} < d_{ij}$ ,  $G'$  contains a path  $p'_{ij}$  between  $v_i$  and  $v_j$  whose length is less than  $d_{ij}$ . We shall prove by induction on the number of elements in  $p'_{ij}$  that this cannot happen. If  $p'_{ij}$  contains only one element, then  $d'_{ij} = d_{ij}$ . Assume  $d'_{ij} = d_{ij}$  if  $p'_{ij}$  contains  $m - 1$  elements, where  $m \leq n$ . Let  $p'_{ij}$  contain  $m$  elements. We may write  $p'_{ij} = p'_{ix} + e'(v_x, v_j)$ , from which we can conclude that  $d'_{ij} = d'_{ix} + d'_{xj}$ . Due to the induction hypothesis  $d'_{ix} = d_{ix}$  and  $d'_{xj} = d_{xj}$ ; therefore  $d'_{ij} = d_{ix} + d_{xj}$ . From the hypothesis  $d_{ix} + d_{xj} \geq d_{ij}$ ; hence  $d'_{ij} \geq d_{ij}$ , which is a contradiction. This proves Theorem 1.

Let  $w_{ij}$  be the weight of the element  $e(v_i, v_j)$  in a graph  $G$ . An element  $e(v_i, v_j)$  in  $G$  is *redundant* if there exists an integer  $k$ , ( $k \neq i, j$ ,  $1 \leq k \leq n$ ), such that  $w_{ij} \geq d_{ik} + d_{kj}$ . For example, elements  $e(v_1, v_5)$  and  $e(v_2, v_4)$  in Fig. 1 are redundant.

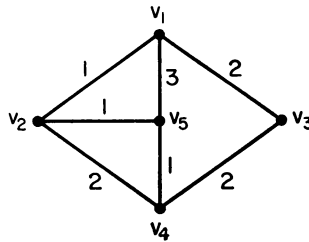


FIG. 1. A weighted graph and its distance matrix.

*Theorem 2.* If  $G$  is an  $n$ -vertex realization of  $D$  without redundant elements, then  $G$  is unique.

*Proof.* Let us assume that there are two "distinct"  $n$ -vertex graphs  $G$  and  $G'$  which realize the distance matrix  $D$ . If  $G$  and  $G'$  are not identical, then either (1) there is a corresponding pair of elements  $e(v_i, v_j)$  in  $G$  and  $e'(v_i, v_j)$  in  $G'$  which have different weights, or (2) there is an element  $e(v_x, v_y)$  in  $G$  (or  $e'(v_x, v_y)$  in  $G'$ ) but there is no corresponding element  $e'(v_x, v_y)$  in  $G'$  (or  $e(v_x, v_y)$  in  $G$ ). As before, let  $w_{ij}$  be the weight of element  $e(v_i, v_j)$  in  $G$  and similarly  $w'_{ij}$  be the weight of element  $e'(v_i, v_j)$  in  $G'$ . If element  $e(v_i, v_j)$  is present in  $G$ , then  $w_{ij} = d_{ij}$ ; for we know  $w_{ij} \geq d_{ij}$  and if  $w_{ij} > d_{ij}$ , then  $e(v_i, v_j)$  is a redundant element. Similarly we could prove that if  $e'(v_i, v_j)$  is present in  $G'$ , then  $w'_{ij} = d_{ij}$ , therefore case (1) cannot occur. If  $e(v_x, v_y)$  is not present in  $G$ , then there exists an integer  $t$ , ( $1 < t < n$ ), such that  $d_{xt} + d_{ty} = d_{xy}$  which proves that the corresponding element  $e'(v_x, v_y)$  in  $G'$  will also be redundant. Hence, case (2) cannot occur. This completes the proof of the theorem.

In contrast to Theorem 2, if we allow the number of vertices in  $G$  to be larger than  $n$  (the order of  $D$ ), then indeed there are many realizations of  $D$ . For example, consider the distance matrix

$$D = \begin{pmatrix} 0 & 2 & 3 \\ 2 & 0 & 2 \\ 3 & 2 & 0 \end{pmatrix}. \tag{3}$$

Two distinct realizations of  $D$ , none of which have any redundant elements, are shown in Fig. 2. Since the realization of  $D$ , in general, is not unique, we can define an optimum realization. Let  $S[G]$  be the sum of the weights of all elements in  $G$ , or the total weight of  $G$ . For example, for the graph of Fig. 2a,  $S[G] = 7$ , and for the graph of Fig. 2b,  $S[G] = 3.5$ . Graph  $G$  is said to be an *optimum realization* of  $D$ , if there exists no realization  $G'$  of  $D$  such that  $S[G'] < S[G]$ . In a realization of  $G$  of the given  $(n \times n)$  distance matrix  $D$ , the vertices  $v_1, v_2, \dots, v_n$  are called external vertices. All other vertices in  $G$  (if there are any) are called internal vertices of  $G$ . It must be noted that only the distances between pairs of external vertices are specified by  $D$ . For example, the graph of Fig. 2a has only external vertices because it realizes a  $3 \times 3$  distance matrix. In the graph of Fig. 2b, the vertices  $v_1, v_2, v_3$  are external vertices, but vertex  $v_4$  is an internal vertex.

Although we were unable to outline a procedure for arriving at the optimum realization, the following *elementary reduction cycle* can be used repeatedly to reduce the total

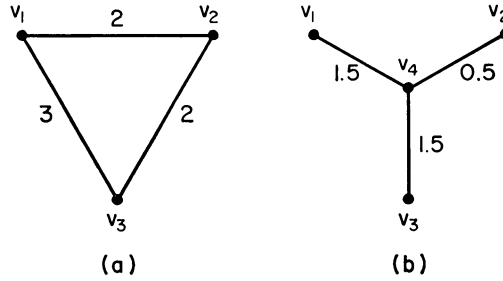


FIG. 2. Two graphs with identical distance matrices.

weight of the realization. In the next section we will show that a modification of this procedure leads to an optimum realization for a certain class of distance matrices.

Let  $G$  be a realization of  $D$ . Let in  $G$  there be three elements  $e(v_i, v_j)$ ,  $e(v_j, v_k)$ , and  $e(v_k, v_i)$ , such that  $w_{ij} + w_{jk} > w_{ik}$ ,  $w_{ij} + w_{ik} > w_{jk}$ , and  $w_{jk} + w_{ki} > w_{ij}$ , then we can find a graph  $G'$  which realizes  $D$  and  $S[G'] < S[G]$ .<sup>2</sup> To do this, we first delete elements  $e(v_i, v_j)$ ,  $e(v_j, v_k)$ , and  $e(v_k, v_i)$  from  $G$ . Let the resulting graph be  $G_1$ . We add a vertex  $v_{n+1}$  to  $G_1$  and we then add elements  $e(v_i, v_{n+1})$ ,  $e(v_j, v_{n+1})$ , and  $e(v_k, v_{n+1})$  to  $G_1$  where the weights of these elements are  $w_{i(n+1)} = \frac{1}{2}[w_{ij} + w_{jk} - w_{ki}]$ ,  $w_{j(n+1)} = \frac{1}{2}[w_{ij} + w_{ik} - w_{jk}]$ , and  $w_{k(n+1)} = \frac{1}{2}[w_{jk} + w_{ki} - w_{ij}]$ . The resulting graph  $G'$  realizes  $D$  and  $S[G'] < S[G]$ .

The following example shows that the elementary reduction cycle can be used repeatedly, and furthermore, the example is designed to illustrate a major difficulty in arriving at an optimum realization. Consider the distance matrix

$$D = \begin{pmatrix} 0 & 3 & 4 & 1 & 4 \\ 3 & 0 & 3 & 2 & 3 \\ 4 & 3 & 0 & 5 & 2 \\ 1 & 2 & 4 & 0 & 3 \\ 4 & 3 & 2 & 3 & 0 \end{pmatrix}. \tag{4}$$

A five-vertex realization of  $D$  is shown in Fig. 3a. Deleting all of the redundant elements, we obtain the optimum five-vertex realization of  $D$ , which is shown in Fig. 3b. An elementary reduction cycle can be carried out by considering the elements  $e(v_2, v_4)$ ,  $e(v_2, v_5)$ , and  $e(v_4, v_5)$ . After such a reduction, a graph of Fig. 3c is obtained. Another reduction cycle is possible, if we introduce a fictitious element  $e(v_2, v_5)$ . Then using elements  $e(v_2, v_5)$ ,  $e(v_2, v_3)$  and  $e(v_3, v_5)$  in a reduction cycle, we obtain the graph shown in Fig. 3d. At this stage it seems that no further reduction is possible. In the graph of Fig. 3b,  $S[G] = 18$ , in the graph of Fig. 3d  $S[G] = 13$ . However, the graph of Fig. 3d is not the optimum realization. The graph shown in Fig. 3e realizes the original distance matrix and,  $S[G] = 12$ .

We will postpone further discussion of optimum realization of the distance matrix to the next section. Here, we would like to examine under what conditions a distance matrix has a tree (i.e., a connected circuit-less) realization. The following theorem states

<sup>2</sup>Clearly, it is not necessary to check all three inequalities. If the sum of the smallest two numbers is larger than the third, then the other two inequalities will automatically be satisfied.

that such a realization (if it exists) is unique, and the proof of the theorem outlines the procedure for arriving at such a realization. The details of the procedure for constructing a tree realization of  $D$  are explained by an example following the proof of Theorem 3. The procedure always gives a tree realization when one exists and fails otherwise.

*Theorem 3.* If  $D$  is realizable as a tree  $t$ , then  $t$  is the only circuit-less realization of  $D$ .

*Remarks.* If  $t$  is a realization of  $D$ , we can find another realization  $t'$  of  $D$  by inserting an internal node  $v_p$  in the middle of an element  $e(v_i, v_j)$  in  $t$  and choosing the weights of the two resulting elements  $e(v_i, v_p)$  and  $e(v_p, v_j)$  in such a way that  $w_{i,p} + w_{p,j} = w_{i,j}$ . Clearly, if internal vertices of degree two, such as  $v_p$  in  $t'$ , are allowed, then every distance matrix will have infinitely many realizations. It should be understood that, in the statement of Theorem 3,  $t$  is a unique realization of  $D$ , if all internal vertices are of degree three or larger. It should be noted that any internal vertex of degree less than three is basically superfluous.

*Proof.* We prove this theorem by induction on the order of matrix  $D$ . If  $D$  is a  $(2 \times 2)$  matrix, then the theorem is clearly true. Let us assume the theorem is correct if the order of  $D$  was  $(n - 1 \times n - 1)$ . Consider an  $(n \times n)$  distance matrix  $D$ . Let  $D'$  be the  $(n - 1 \times n - 1)$  leading principal submatrix of  $D$ , (i.e.,  $D'$  be a matrix consisting of the first  $n - 1$  rows and  $n - 1$  columns of  $D$ .) Clearly, if  $D$  is realizable as a circuitless graph, so is  $D'$ . Let  $t'$  be the circuit-less realization of  $D'$ . According to the induction

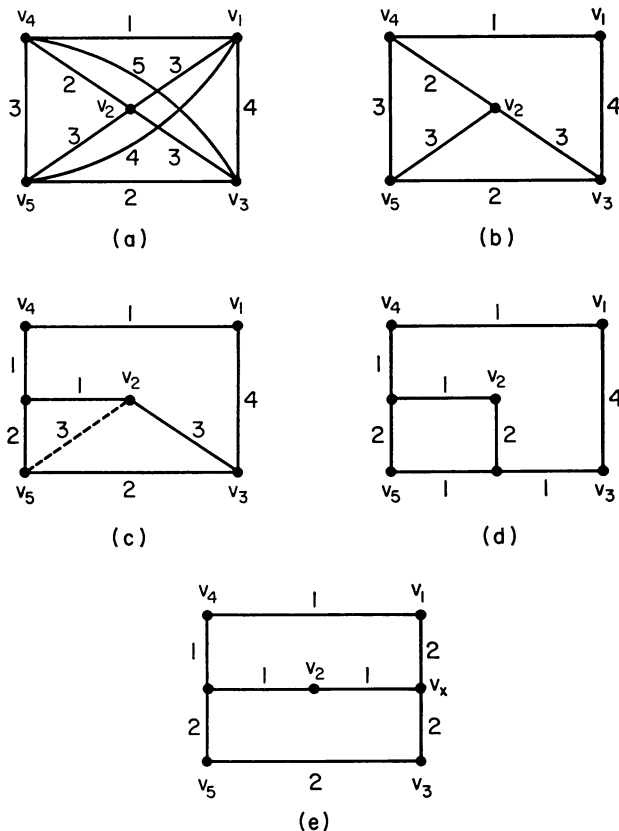


FIG. 3. Step by step reduction of a realization of a distance matrix.

hypothesis  $t'$  is unique. To construct tree  $t$  from  $t'$ , we add an isolated vertex  $v_n$  to  $t'$  and then connect an element  $e(v_n, x)$  of weight  $w$  between vertex  $v_n$  and some "point"  $x$  on an element of  $t'$ . We now prove that  $t$  obtained from  $t'$  is unique, by showing that weight  $w$  and point  $x$  are unique. To do this, assume that there are two points  $x_1$  and  $x_2$  on  $t$  and two elements  $e(v_n, x_1)$  and  $e(v_n, x_2)$  of weights  $w_1$  and  $w_2$  which would result in two acceptable realizations  $t_1$  and  $t_2$ . It can be seen that there exists an external vertex  $v_i$  in  $t_1$  and such that in  $t_1$

$$d_{ni} = w_1 + d(x_1, v_i), \tag{5}$$

and in  $t_2$

$$d_{ni} = w_2 + d(x_2, x_1) + d(x_1, v_i). \tag{6}$$

From the above two equalities, we obtain

$$w_1 = w_2 + d(x_2, x_1). \tag{7}$$

Using another terminal vertex  $v_i$  in  $t_1$  and  $t_2$ , we could arrive at the equation

$$w_2 = w_1 + d(x_2, x_1). \tag{8}$$

Equations (7) and (8) prove that  $d(x_1, x_2) = 0$  and  $w_2 = w_1$ . To prove the theorem it remains to show that if we start with another principal submatrix of  $D$ , we will not obtain a different realization of  $D$ . To do this, we assume that  $t^* \neq t$  is also a tree realization of  $D$ . It is clear that  $t^*$  will also realize  $D'$ , with the possibility of a superfluous element. We consider two cases: (1) degree of  $v_n$  in  $t^*$  is equal to one, and (2) degree of  $v_n$  in  $t^*$  is larger than one. In the first case, the element incident at  $v_n$  is deleted from  $t^*$  and resulting graph  $t_1^*$  realizes  $D'$ . However, due to induction hypothesis,  $t_1^* = t'$ . We have already shown a tree realization of  $D$  obtained from  $t'$  is unique, thus  $t^* = t$ . With similar reasoning, it can be shown that the second case will also lead to a contradiction, hence the theorem.

The proof of Theorem 3 suggests a procedure for arriving at a circuit-less realization of a distance matrix, if such a realization exists. The details of the procedure may be explained by means of an example. Consider

$$D = \begin{pmatrix} 0 & 3 & 5 & 4 & 5 \\ 3 & 0 & 6 & 5 & 6 \\ 5 & 6 & 0 & 5 & 4 \\ 4 & 5 & 5 & 0 & 3 \\ 5 & 6 & 4 & 3 & 0 \end{pmatrix}. \tag{9}$$

We start by choosing two vertices  $v_1$  and  $v_2$  and joining them by an element of weight 3. We next add a vertex  $v_3$  and we find a point  $x_0$  on element  $e(v_1, v_2)$  and insert an element  $e(v_3, x_0)$  of weight  $w$  between  $v_3$  and  $x_0$ . By a simple calculation, we can determine that point  $x_0$  is along the path from  $v_1$  to  $v_2$  and is of distance 1 from  $v_1$  and that weight  $w$  is equal to 4. These two steps are shown in Fig. 4a and b. The graph of Fig. 4b, which will be denoted by  $t_3$ , is the circuit-less realization of the  $(3 \times 3)$  leading principal submatrix of  $D$ . There is, however, a simpler method for arriving at  $t_3$ . To do this, we start, as in Theorem 1, by connecting three elements  $e(v_1, v_2)$ ,  $e(v_1, v_3)$ ,  $e(v_2, v_3)$  of weights



$v_2, v_3$ , and  $v_4$  of weights 5, 6, 4, and 3, as required by the given distance matrix. Such a graph is shown in Fig. 4f. To determine the location of  $x_2$ , we start by going through one elementary reduction cycle involving elements  $e(v_5, v_2), e(v_5, v_1)$ , and path  $p_{12}$ . The resulting graph is shown in Fig. 3g. The position of  $x_2$  is not entirely determined. Let us perform one more reduction cycle involving elements  $e(v_5, x_0), e(v_5, v_3)$ , and  $p_{3x_0}$ . The resulting graph is shown in Fig. 4h. At this stage the position of  $x_2$  is completely determined, and we notice that the distance from  $v_5$  to  $v_4$  within the tree is 5 but the required distance is 3. We may now conclude that the given matrix is not realizable as a circuit-less graph.

**3. Optimum reduction and realization.** Let  $G$  be a realization of  $D$  which has external vertices  $v_1, v_2, \dots, v_n$  and internal vertices  $v_x, v_y, v_z, \dots$ . Let  $d(v_p, v_q)$  represent the distance between vertices  $v_p$  and  $v_q$  in  $G$ . If  $v_p$  and  $v_q$  are both terminal vertices, then  $d(v_p, v_q) = d_{pq}$ . A vertex  $v_p$  in  $G$  is said to be a compact vertex of  $G$ , if there exists an entry  $d_{ij}$  in  $D$  such that

$$d_{ij} = d(v_i, v_p) + d(v_p, v_j), \text{ and } v_i, v_j \neq v_p.$$

*Theorem 4.* If  $G$  is an optimum realization of a distance matrix  $D$ , then every vertex in  $G$  of degree higher than one is a compact vertex.

*Proof.* Suppose  $G$  is an optimum realization of  $D$  and contains a vertex  $v_p$  which is not compact. Let two vertices which are adjacent to  $v_p$  in  $G$  be  $v_q$  and  $v_r$ .<sup>3</sup> If there is an element  $e(v_q, v_r)$  in  $G$  whose weight is any nonnegative number, then, either in  $G$  there is a redundant element, or by an elementary reduction cycle involving elements  $e(v_p, v_q), e(v_p, v_r), e(v_q, v_r)$ , we can obtain a graph  $G'$ , which realizes  $D$  and  $S[G'] < S[G]$ . This is a contradiction. Suppose element  $e(v_q, v_r)$  does not exist. Let  $G'$  be obtained from  $G$  by adding an element  $e(v_q, v_r)$  to  $G$  whose weight  $w_0$  is determined by the following equation:

$$w_0 = \max \{ [d(v_i, v_j) - d(v_i, v_q) - d(v_j, v_r)], 0 \}.$$
 (10)

over all possible pairs of external vertices  $v_i$  and  $v_j$  in  $G$ . To begin with, we would like to show that  $G'$  has the same distance matrix as  $G$ . Let the distance matrix of  $G'$  be  $D' = [d'_{ij}]$ . If  $G$  and  $G'$  do not have the same distance matrix, then, since the entries of  $D'$  cannot be larger than the corresponding entries of  $D$ , there must exist an entry  $d'_{k1} < d_{k1}$ . Since element  $e(v_q, v_r)$  in  $G'$  had the effect of reducing the entry  $d_{k1}$  in  $D$  to  $d'_{k1}$  in  $D'$ , we can conclude that

$$d'_{k1} = d(v_k, v_q) + d'(v_q, v_r) + d(v_r, v_1),$$
 (11)

where  $d(v_q, v_r) = w_0$ . Equation (10) implies that

$$d'(v_q, v_r) = w_0 \geq d(v_1, v_k) - d(v_k, v_q) - d(v_1, v_r).$$
 (12)

Substitution of (12) into (11) yields

$$d'_{k1} \geq d(v_1, v_k) = d_{k1}.$$
 (13)

Since (13) contradicts the original assumption,  $G$  and  $G'$  have the same distance matrix.

The next step in the proof is to show that

$$w_0 < d(v_q, v_p) + d(v_p, v_r).$$
 (14)

---

<sup>3</sup>Two vertices  $v_p$  and  $v_q$  in  $G$  are adjacent if there is an element  $e(v_p, v_q)$  in  $G$ .



Since  $v_p$  is not a compact vertex in  $G$ , we have

$$d(v_k, v_i) < d(v_k, v_p) + d(v_p, v_i), \tag{15}$$

where  $v_k$  and  $v_i$  are any two terminal vertices in  $G$ . We know that  $d(v_k, v_p) \leq d(v_k, v_a) + d(v_a, v_p)$  and  $d(v_p, v_i) \leq d(v_p, v_r) + d(v_r, v_i)$ . Making these substitutions in (15), we obtain

$$d(v_k, v_i) < d(v_k, v_a) + d(v_a, v_p) + d(v_p, v_r) + d(v_r, v_i), \tag{16}$$

which can be written as

$$d(v_k, v_i) - d(v_k, v_a) - d(v_i, v_r) < d(v_a, v_p) + d(v_p, v_r). \tag{17}$$

According to (10), if we choose  $v_k$  and  $v_i$  properly, the left hand side of the inequality (17) is equal to  $w_0$ ; therefore the desired inequality (14) is derived. Note that if  $w_0 = 0$ , then (14) is automatically satisfied.

Let us return to graphs  $G$  and  $G'$  which have the same distance matrices and assume that in graph  $G'$  we have an element  $e(v_a, v_r)$  whose weight  $w_0$  is less than the sum of the weights of elements  $e(v_a, v_p)$  and  $e(v_r, v_p)$ . It can now be seen that either  $e(v_a, v_p)$  or  $e(v_r, v_p)$  are redundant in  $G$ , or that, by an elementary reduction cycle involving the three elements mentioned above, we can obtain a new graph  $G''$  which has the same distance matrix as  $G$  such that  $S[G''] < S[G]$ . This is a contradiction to the assumption that  $G$  was an optimum realization.

Theorem 4 is useful in finding an optimum realization of a distance matrix. This can be illustrated by the example of distance matrix (4). The "best" realization we were able to obtain is shown in Fig. 3e. It can be shown that in that graph the terminal vertex  $v_3$  is not compact. By a simple calculation, we can find that if we add an element  $e(v_5, v_x)$  of weight 2 to the graph of Fig. 3e, the distance matrix of that graph will not be disturbed. The resulting graph is shown in Fig. 5a. Now, going through an elementary reduction cycle involving elements  $e(v_3, v_5)$ ,  $e(v_3, v_x)$ , and  $e(v_5, v_x)$ , we obtain the graph of Fig. 5b, which realizes the distance matrix (4), and in which the sum of all weights is equal to 11.

Let  $D$  be a given  $(n \times n)$  distance matrix. Let  $D_i(a)$  be a matrix obtained from  $D$  by subtracting a nonnegative number  $a$  from all entries in the  $i$ -th row and the  $i$ -th column of  $D$ , with the exception of the entry  $(i, i)$ . More precisely, if  $D = [d_{\alpha\beta}]$  and  $D_i(a) = [d'_{\alpha\beta}]$ , then

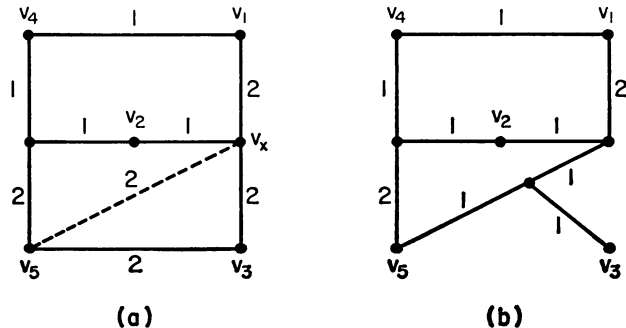


FIG. 5. A noncompact vertex and the resulting reduction.

$$d_{pq}^i = \begin{cases} d_{pq}, & \text{if } p \text{ and } q \neq i, \\ d_{pq} - a, & \text{if } p \text{ or } q = i, \quad p \neq q, \\ 0, & \text{if } p = q, \end{cases} \tag{18}$$

We would like to know, how could one choose the number  $a$  such that  $D_i(a)$  is a distance matrix?

*Lemma 1.* Matrix  $D_i(a)$  is a distance matrix if, and only if

$$a \leq \frac{1}{2}(d_{pi} + d_{ir} - d_{pr}), \text{ for } p, r = 1, 2, \dots, n; \text{ and } p, r \neq i. \tag{19}$$

*Proof.* If the above inequality is violated for some  $p$  and  $r$ , then we have

$$2a > (d_{pi} + d_{ir} - d_{pr}), \quad p, r \neq i, \tag{20}$$

which may be written as

$$d_{pr} > [(d_{pi} - a) + (d_{ir} - a)], \quad p, r \neq i \tag{21}$$

From the definition of  $D_i(a)$ , it follows from (21) that

$$d_{pr}^i > d_{pi}^i + d_{ir}^i, \tag{22}$$

which proves that  $D_i(a)$  is not a distance matrix.

Suppose the inequality of the hypothesis is satisfied, then we must show that  $D_i(a)$  is a distance matrix (if  $D$  is). To do this, we must prove that for every  $q$ , ( $1 \leq q \leq n$ ),

$$d_{pq}^i + d_{qr}^i \geq d_{pr}^i, \text{ for } p, r = 1, 2, \dots, n. \tag{23}$$

We will consider two cases.

*Case (a):*  $q \neq i$ . If  $p \neq i$  and  $r \neq i$ , then, from the definition of  $D_i(a)$  and the fact that  $D$  is a distance matrix, the inequality (23) is satisfied. If  $p = i$  (or  $r = i$ ), then for any value of  $a$ , the inequality (23) remains satisfied. If  $p = i$  and  $r = i$ , then (23) may be written in terms of the entries of  $D$  and the number  $a$  as follows:

$$(d_{iq} - a) + (d_{qi} - a) \geq 0, \text{ for } q = 1, 2, \dots, n, \quad q \neq i, \tag{24}$$

which implies that

$$a \leq d_{iq}, \text{ for } q = 1, 2, \dots, n \text{ and } q \neq i. \tag{25}$$

However, if we let  $p = r$  in the inequality of the hypothesis, then we see that (25) follows from the hypothesis.

*Case (b):*  $q = i$ . If  $p \neq i$  and  $r \neq i$ , then (23) may be written as

$$(d_{pi} - a) + (d_{ir} - a) \geq d_{pr}, \text{ for } p, r = 1, 2, \dots, n; \quad p, r \neq i. \tag{26}$$

The inequality (26), however, is identical with the inequality of the hypothesis. If  $p = i$  (or  $r = i$ ), then (23) is satisfied for all values of  $a$ , and finally if  $p = i$  and  $r = i$ , then again (23) is satisfied. This completes the proof of the Lemma.

Let  $a_0$  be a number defined as follows

$$a_0 = \min (d_{pi} + d_{ir} - d_{pr})/2, \quad (p, r = 1, 2, \dots, n; p, r \neq i). \tag{27}$$

It follows from Lemma 1 that  $D_i(a)$  is a distance matrix for all  $a$  satisfying  $0 \leq a \leq a_0$ .

Let  $G_i(a)$  be a realization of  $D_i(a)$ . Let the external vertices of  $G_i(a)$  be  $v'_1, v'_2, \dots, v'_n$ .

To construct a realization  $G$  of  $D$  from  $G_i(a)$ , we add a vertex  $v_i$  to  $G_i(a)$  and connect an element  $e(v_i, v'_i)$  between  $v_i$  and  $v'_i$  whose weight is equal to  $a$ .

*Theorem 5.* If  $0 \leq a \leq a_0$  and if  $G_i(a)$  is an optimum realization of  $D_i(a)$ , then  $G$  obtained from  $G_i(a)$  by the above process is an optimum realization of  $D$ .

*Proof.* If  $a = 0$ , then  $G_i(a)$  and  $G$  are identical.  $G$  is the optimum realization of  $D$ , otherwise  $G_i(a)$  could not be the optimum realization of  $D_i(a)$ . Suppose  $a > 0$  and the theorem is not correct. Then, there exists an optimum realization  $G^*$  of  $D$  such that

$$S[G^*] < S[G] = S[G_i(a)] + a. \tag{28}$$

Let the external vertices of  $G^*$  be  $v_1^*, v_2^*, \dots, v_n^*$ . We consider two cases: (a) the vertex  $v_i^*$  is of degree two or higher, and (b)  $v_i^*$  is of degree one.

*Case (a).* We would like to show that  $v_i^*$  is a noncompact vertex. From (27), we can conclude that

$$a_0 \leq \frac{d_{pi} + d_{ir} - d_{pr}}{2}, \text{ for } p, r = 1, 2, \dots, n; \quad p, r \neq i, \tag{29}$$

and we know that  $0 < a \leq a_0$ . Thus,

$$d_{pi} + d_{ir} - d_{pr} > 0, \text{ for } p, r = 1, 2, \dots, n, \text{ and } p, r \neq i, \tag{30}$$

which proves that  $v_i^*$  is not a compact vertex. Since  $v_i^*$  is also assumed to be of degree higher than one, according to Theorem 4,  $G^*$  is not an optimum realization, which is a contradiction.

*Case (b).* Vertex  $v_i^*$  is of degree one. Let the adjacent vertex to  $v_i^*$  in  $G^*$  be  $v_x^*$ . We first show that  $d(v_i^*, v_x^*) = a_1 < a$ . Suppose  $d(v_i^*, v_x^*) \geq a$ , then we can pick a point  $v_{i'}^*$  on element  $e(v_i^*, v_x^*)$  such that  $d(v_{i'}^*, v_i^*) = a$ . Let graph  $G_1^*$  be obtained by removing  $e(v_{i'}^*, v_i^*)$  from  $G^*$ . Then, it can be seen that  $G_1^*$  will realize  $D_i(a)$  and that

$$S[G_1^*] = S[G^*] - a < S[G] - a = S[G_i(a)] \tag{31}$$

which proves that  $G_i(a)$  is not an optimum realization of  $D_i(a)$ , and therefore,  $d(v_i^*, v_x^*) < a$ . If  $d(v_i^*, v_x^*) = a_1 < a$ , the  $v_x^*$  must be an internal vertex, for the distance of the closest external vertex from  $v_i^*$  is at least  $a_0$ . Since  $v_x^*$  is an internal vertex, it must be of at least degree three. Consider the graph that is obtained by removing element  $e(v_i^*, v_x^*)$  from  $G^*$ . If in the resulting graph  $v_x^*$  is assumed to be the  $i$ -th external vertex, then the resulting graph will realize the matrix  $D_i(a_1)$  which may therefore, be denoted by  $G_i^*(a_1)$ . With similar reasoning as in case (a), we can show that vertex  $v_x^*$  in  $G_i^*(a_1)$  is not compact, therefore  $G_i^*(a_1)$  is not an optimum realization of  $D_i(a_1)$ , which implies that  $G^*$  is not an optimum realization of  $D$ . This, however, is again a contradiction; hence, the theorem.

*Theorem 6.* If  $D$  has a tree realization  $t$ , then  $t$  is the optimum realization of  $D$ .

*Proof.* The theorem is true if  $D$  is of order 2. Suppose it is also true if  $D$  is of order  $n-1$ . Consider a distance matrix  $D$  of order  $n$ . Let the realization of the  $(n-1) \times (n-1)$  leading principal submatrix of  $D$  be  $t_{n-1}$ . From Theorem 3, there exists a point  $x$  on  $t_{n-1}$  such that if we connect element  $e(v_{n-1}, x)$  of weight  $w_0$  we obtain the realization  $t$  of  $D$ . Let point  $x$  in  $t_{n-1}$  be labeled  $v'_n$ . Let us assume that  $v'_n$  is also a terminal vertex of the modified graph  $t'_{n-1}$ . It can be seen that the  $(n \times n)$  distance matrix of  $t'_{n-1}$  is  $D_n(w_0)$ . Since  $t_{n-1}$  is the optimum realization of the  $(n-1) \times (n-1)$  leading principal submatrix of  $D$ ,  $t'_{n-1}$  must be the optimum realization of  $D_n(w_0)$ . According to Theorem 5,  $t$  con-

structed from  $t_{n-1}$  must be the optimum realization of  $D$ . This completes the proof of the theorem.

It is also possible to prove that  $t$  is the only optimum realization of  $D$ . The authors believe that in general the optimum realization of any distance matrix is unique. However, they were not able to prove it. The main problem in arriving at an optimum realization is that one does not know when to quit searching for "better" realizations. It would be of great help if there was an upper bound on the total weight of the optimum realization.

**4. Realizability of a distance matrix as a graph with uniform weights.** It might be interesting to study the distance matrices of graphs in which all elements have uniform weights. It is also clear that we do not lose any generality in such a study by assuming that the weights of all elements are equal to one. Therefore, we will concentrate on distance matrices of unweighted graphs. The length of a path in such a graph is the number of elements in that path, and the distance  $d_{ij}$  between two vertices  $v_i$  and  $v_j$  is the length of the shortest path between  $v_i$  and  $v_j$ . In this section by a graph we mean an unweighted graph.

It can easily be seen that every distance matrix whose entries are all integers is realizable as a (unweighted) graph with additional internal vertices. The interesting problem, therefore, seems to be to find the class of  $(n \times n)$  distance matrices which are realizable as an  $n$ -vertex graph. The statement of Theorem 2 by itself, although in an indirect manner, provides a solution to this problem. A more direct solution may be formulated as follows:

*Theorem 7.* A symmetric matrix  $D = [d_{ij}]$  of order  $n$  with non-negative integers as entries is realizable as a distance matrix of an  $n$ -vertex (unweighted) graph if, and only if

- (a)  $d_{ii} = 0$  for  $i = 1, 2, \dots, n$ ;
- (b) for every entry  $d_{ij}$  of  $D$

$$d_{ij} \leq d_{ip} + d_{pj} \quad \text{for } p = 1, 2, \dots, n;$$

- (c) for every entry  $d_{ij} > 1$ , there exists an integer  $k$ ,  $(1 \leq k \leq n)$ , such that

$$d_{ij} = d_{ik} + d_{kj}, \text{ and } d_{ik} = 1.$$

*Proof.* The necessity of (a) and (b) is already established. To prove the necessity of (c), consider an  $n$ -vertex graph  $G$  and a pair of vertices  $v_i$  and  $v_j$  in  $G$  such that  $d_{ij} > 1$ . Let  $p_{ij}$  be a shortest path between  $v_i$  and  $v_j$ . Since the length of this path is larger than or equal to two, there is a vertex  $v_k$  on  $p_{ij}$  such that  $p_{ij} = p_{ik} + e(v_k, v_j)$ . Path  $p_{ik}$  must be a shortest path between  $v_i$  and  $v_k$ , therefore  $d_{ij} = d_{ik} + 1$ .

To prove sufficiency, we construct a graph  $G$  as follows. Let  $n$  points be labeled  $v_1, v_2, \dots, v_n$ . An element  $e'(v_i, v_j)$  is placed between  $v_i$  and  $v_j$  if, and only if, the corresponding entry  $d_{ij}$  in  $D$  is one. This operation is continued for every entry of  $D$  which is equal to one. If for some  $i$  and  $j$ ,  $d_{ij} = 0$ , then points  $v_i$  and  $v_j$  are assumed to be identical. Let the resulting graph be  $G'$  and its distance matrix be  $D' = [d'_{ij}]$ . We would like to show that  $D' = D$ . Suppose not, then there exists at least one entry  $d'_{ij} \neq d_{ij}$ ,  $(i \neq j)$ , which implies that either  $d'_{ij} < d_{ij}$ , or  $d'_{ij} > d_{ij}$ .

(a)  $d'_{ij} < d_{ij}$ . Clearly if  $d'_{ij} \leq 1$ , then  $d'_{ij} = d_{ij}$ . Suppose that if  $d'_{ij} \leq k - 1$ , then  $d'_{ij} = d_{ij}$ . Let  $d'_{ij} = k$ , then in  $G'$  there exists a vertex  $v_p$ ,  $(v_p \neq v_i, v_j)$ , such that  $d'_{ij} = d'_{ip} + d'_{pj}$ . From condition (b) of the hypothesis, we have  $d_{ij} \leq d_{ip} + d_{pj}$ ; there-

fore,  $d'_{i_p} + d'_{p_i} < d_{i_p} + d_{p_i}$ . However, since  $d'_{i_p}$  and  $d'_{p_i} \leq k - 1$ ,  $d'_{i_p} = d_{i_p}$  and  $d'_{p_i} = d_{p_i}$  which is a contradiction, therefore  $d'_{i_i} \geq d_{i_i}$ .

(b)  $d'_{i_i} > d_{i_i}$ . If  $d_{i_i} \leq 1$ , then  $d'_{i_i} = d_{i_i}$ . Suppose that if  $d_{i_i} \leq k - 1$ , then  $d'_{i_i} = d_{i_i}$ . Let  $d_{i_i} = k$ , ( $k > 1$ ), according to condition (c) of the hypothesis there exists an integer  $p$  such that  $d_{i_i} = d_{i_p} + d_{p_i}$  and  $d_{i_p} = 1$ . Therefore, in  $G'$  there is an element  $e'(v_i, v_p)$  and  $d'_{i_i} \leq 1 + d'_{p_i}$ . This implies that  $d'_{p_i} > d_{p_i}$ . Since  $d_{p_i} = k - 1$ , this is a contradiction, hence the theorem.

**5. Conclusions and an application.** The properties of distances in a graph were studied by defining a distance matrix and investigating its realizability. It was shown that if we allowed nonterminal vertices in the realization of a distance matrix  $D$ , then  $D$  had more than one realization. Therefore, we defined the optimum realization and described methods for attempting to arrive at such a realization. It was shown that if  $D$  had a tree realization  $t$ , then  $t$  was unique and the optimum realization of  $D$ . We were not able to give a complete and finite procedure for obtaining the general optimum realization. The main difficulty was that we did not know when to stop searching for reduced realizations. Theorem 4 supplied a strong necessary condition and its proof gave a method to reduce a graph which had noncompact vertices. It is our feeling that the optimum realization is always unique, but we were not able to prove or disprove this. Some of the results of this paper can be easily generalized to directed graphs. For example, Theorem 1 supplies the necessary and sufficient conditions for realizability of a distance matrix  $D$  by a directed graph when the condition of symmetry of  $D$  is removed. The authors attempted to use the distance matrix to extract some results on centers and medians of a graph [6]. Some progress has already been made in this direction [7].

Some of the results of this paper can be used to attack the problem of synthesis of  $n$ -port resistive networks [8]. The open circuit impedance matrix of a grounded  $n$ -port resistive network  $Z_{oc}$  can be transformed to a driving-point impedance matrix,  $Z_d$ . It can then be shown that  $Z_d$  is a distance matrix, though not every distance matrix is a driving-point impedance matrix [8]. Since  $Z_d$  is a distance matrix, using the procedure described in the proof of Theorem 3, one may find a circuit-less realization of  $Z_d$ . In the resulting graph, if the weights of the elements are identified with the resistances of the same values, then the graph will also realize  $Z_d$  as a driving-point impedance matrix. Also the same graph will realize  $Z_{oc}$ , if the external terminals are properly identified. It is our feeling that further effort in this direction may lead to fruitful results.

#### REFERENCES

1. E. F. Moore, *Shortest path through a maze*, Proc. Internatl. Symposium on Switching Circuits, Harvard University (1957) 285-292
2. J. B. Kruskal, Jr., *On the shortest spanning subtree of a graph and the traveling salesman problem*, Proc. Amer. Math. Soc. 7 (1960) 48-50
3. R. C. Prim, *Shortest connection networks and some generalization*, Bell System Tech. Journal 36 (1957) 1389-1401
4. R. Courant and H. Robbins, *What is mathematics*, Oxford University Press, London, 1941
5. S. Seshu and M. B. Reed, *Linear graphs and electrical networks*, Addison-Wesley, Reading, Mass., 1961
6. O. Ore, *Theory of graphs*, Am. Math. Soc. Colloquium Publication 38 (1962)
7. S. L. Hakimi, *Optimum locations of switching centers and the absolute centers and medians of a graph*, J. Operations Res. 12 (1964) 450-459
8. S. L. Hakimi and S. S. Yau, *Distance matrix and the synthesis of  $N$ -port resistance networks*, Network Theory Group, Tech. Report No. 5, Northwestern University, Evanston, Ill., 1963