JOURNAL OF COMBINATORIAL THEORY, Series B **36**, 1–15 (1984)

On Optimal Embeddings of Metrics in Graphs

WILFRIED IMRICH

Institut fuer Mathematik, Montanuniversitaet Leoben, Austria

J. M. S. SIMÕES-PEREIRA AND CHRISTINA M. ZAMFIRESCU

Department of Computer Science, Hunter College and The Graduate School of The City University of New York, 965 Park Avenue, New York, New York 10021

Communicated by the Editors

Received March 11, 1983

This paper extends previous results of the authors. In particular, non-tree-realizable metrics are investigated and it is shown that every finite metric has an optimal realization by a graph.

1. INTRODUCTION

To embed metrics in graphs, or, in other words, to find graph realizations of distance matrices, is an area of research which has been given much attention (see [1–14, 16–19]). The subject has many applications in such varied fields as operations research [18], the study of electrical networks [6], the design of coding techniques [4]; most interesting is its use in a biological model to reconstruct phylogenetic trees from matrices whose entries represent certain genetic distances among contemporary biological species, in an attempt to verify whether these species might have evolved from a single one, possibly extinct [5].

Tree-realizable metrics are the best known; tree realizations are optimal in a sense to be made precise below. In this paper we present several new results on optimal and unique optimal realizations of non-tree-realizable metrics.

Let $G(V, E, w)$ be a graph with vertex and edge sets V and E, respectively, and w: $E \to \mathbb{R}^+$ a function which assigns a positive weight or length to each edge of G. Furthermore, let $d_G(x, y)$ denote the length of a shortest path in G joining the vertices x and y. We say G *realizes* a metric (M, d) if M is a subset of V and if

$$d(a, b) = d_G(a, b)$$

0095-8956/84 $3.00

for all elements a, b of M. Unless otherwise specified we assume M to be finite.

The elements in $V\backslash M$ are called *auxiliary vertices* of the realization. We shall always suppress auxiliary vertices of degree 2. We also note that we can consider the edges of G as unordered pairs of vertices since there will be no need for parallel edges.

A realization G of (M, d) is called *minimal* if the removal of an arbitrary edge of G yields a graph which does not realize (M, d). A realization G of (M, d) is called *optimal* if the sum of all edge lengths of G, denoted $w(G)$, is minimal among all realizations of (M, d).

The notation in this paper is standard. Moreover, we shall use p_{xy} to denote a shortest path between vertices x and y and $|p_{xy}|$ to denote its length. To denote the length of an edge $[a, b]$ we simply write $w(a, b)$.

2. Existence of Optimal Realizations

LEMMA 2.1. *Let (M, d) be a metric on n points and G a realization of (M, d) on more than $2^{\binom{n}{2}+1}$ vertices. Then (M, d) is realized by a proper subgraph of G with at most $2^{\binom{n}{2}+1}$ vertices.*

Proof. For every pair x, y of distinct elements of M we choose a shortest path p_{xy} from x to y in G. Let P be the set of these paths. Clearly P has $\binom{n}{2}$ elements and we can suppose that G is the union of these paths.

As at least two paths of P intersect in every auxiliary vertex, we can label every auxiliary vertex v of G by the set P_v of paths in P containing v

$$P_v = \{p_{xy} \mid v \in p_{xy}, p_{xy} \in P\}.$$

As G contains more than $2^{\binom{n}{2}+1}$ vertices, there must be at least three auxiliary ones, say u, v, w, labelled with the same subsets of P, i.e.,

$$P_u = P_v = P_w.$$

Let the notation be chosen such that

$$d_G(u, v) + d_G(v, w) = d_G(u, w).$$

Choosing a path p_{xy} in P_v and deleting all edges of G incident with v which are not on p_{xy} we obtain a new graph which still realizes (M, d) but has fewer auxiliary vertices (of degree $\geqslant 3$) than G.

Thus there can be at most one v to every u such that $P_u = P_v$. This proves the lemma.

We note that the bound $2^{\binom{n}{2}+1}$ is by no means best possible.

THEOREM 2.2. *Every finite metric (M, d) has an optimal realization.*

Proof. Let a be the infimum of the total lengths of realizations of (M, d) and let

$$G_1, G_2, G_3, \ldots,$$

be a sequence of realizations of (M, d) with $G_i = G_i(V_i, E_i, w_i)$ and

$$\lim_{i \to \infty} w_i(G_i) = a.$$

By Lemma 2.1 we can assume that every G_i has at most $2^{\binom{n}{2}+1}$ vertices, where n is the number of elements of M. Since there exist only finitely many (unweighted) graphs on a given finite number of vertices there must be infinitely many G_i which differ only in the w_i. By choosing a subsequence of the G_i we can in fact suppose that all G_i differ only in the lengths of their edges from a given unweighted graph $G(V, E)$, i.e., $G_i = G_i(V, E, w_i)$.

The lengths of the edges of the G_i are bounded from below by 0 and from above by the diameter of (M, d). Furthermore, the total length of a finite weighted graph is a continuous function of the lengths of its edges. Since a continuous function on a compact set attains its minimum, there is a realization $H(V, E, w)$ of (M, d) such that $w(H) = a$.

We should like to add that the weights of some of the edges of H may be zero. We contract these edges to single points and the new graph obtained still realizes (M, d) and has the same total weight as H. This proves the theorem.

Let (M, d) be a metric and N a finite subset of the set of unordered pairs $[a, b]$ of distinct elements of M. We call a graph $G(V, E, w)$ a *partial realization* or *N-realization* if the following conditions hold:

(i) $d(x, y) \leqslant d_G(x, y)$ for all $x, y \in M \cap V$.

(ii) $a, b \in V$ and $d(a, b) = d_G(a, b)$ for every pair $[a, b]$ in N.

It is obvious what is meant by *minimal* and *optimal partial realizations.*

COROLLARY 2.3. *To every finite subset N of the set of unordered pairs of distinct elements of a metric (M, d) there exists an optimal N-realization.*

Proof. As in Lemma 2.1, one can show that one can restrict attention to partial realizations with at most $2^{\binom{n}{2}+1}$ vertices. Then one can proceed as in the proof of Theorem 2.2 to show the existence of optimal partial realizations.

If one drops the condition that a metric be symmetric one obtains so-called quasimetrics. They can be realized by oriented weighted graphs. In [15] it was shown that optimal realizations without zero weights need not exist for quasimetrics.

3. Unique Optimal Realizations

Having established the existence of optimal realizations for finite metrics we shall investigate cases in which they are unique. A simple, but useful tool will be

LEMMA 3.1. *Let x, y, z, t be four distinct points of a metric (M, d) and suppose that one of the following conditions is satisfied*:

(i) $d(x, y) + d(y, z) = d(x, z)$.

(ii) $d(x, y) + d(z, t) < \max\{d(x, z) + d(y, t), d(x, t) + d(y, z)\}$.

If the first condition holds y is the only common point of any shortest path between x and y and any shortest path between y and z in any realization of (M, d). If the second condition holds every shortest path between x and y is disjoint from any shortest path between z and t in any realization of (M, d).

Proof. Suppose there is a point v on a shortest path between x and y and on a shortest path between y and z in a realization of (M, d). Then

$$d(x, z) \leqslant d(x, v) + d(v, z) < d(x, y) + d(y, z),$$

in violation of (i). This proves the first assertion.

To prove the second we assume the existence of a point v on a shortest path between x and y and on a shortest path between z and t in some realization of (M, d). Then

$$\begin{aligned} d(x, z) + d(y, t) &\leqslant d(x, v) + d(v, z) + d(y, v) + d(v, t) \\ &= d(x, y) + d(z, t) \end{aligned}$$

and similarly

$$d(x, t) + d(y, z) \leqslant d(x, y) + d(z, t),$$

which proves the lemma.

Let x, y, z be arbitrary points of a metric (M, d). We introduce the notation

$$c(x, y, z) = (d(x, y) + d(y, z) - d(x, z))/2$$

and observe that $c(x, y, z) \geqslant 0$ because of the triangle inequality. Also, it is readily verified that

$$d(x, y) = c(x, y, z) + c(z, x, y)$$

and that similar expressions hold for $d(y, z)$ and $d(z, x)$. Thus every metric

(M, d) on three points x, y, z can be realized by a star with center v and edges $[v, x]$, $[v, y]$, $[v, z]$ of lengths $c(z, x, y)$, $c(x, y, z)$, and $c(y, z, x)$, respectively. Of course one of these edges can degenerate to a single point. (This realization is the unique optimal realization of (M, d) as the reader can verify by invoking Corollary 5.4.)

THEOREM 3.2. *Let $G(V, E, w)$ be a realization of (M, d) with $V = M$. Then G is the unique optimal realization of (M, d) if and only if the following two conditions are satisfied*:

(i) $d(x, z) = d(x, y) + d(y, z)$ *for all edges* $[x, y]$ *and* $[y, z]$ *of* G.

(ii) $d(x, y) + d(t, z) < \max\{d(x, t) + d(y, z), d(x, z) + d(y, t)\}$ *for all edges* $[x, y]$, $[t, z]$ *with no common endpoints.*

Proof. Let $G(V, E, w)$ be a realization of (M, d) satisfying (i) and (ii). Since every realization $G'(V', E', w')$ of (M, d) must contain shortest paths between any two vertices a, b with $[a, b]$ in E and since no two such paths can have interior points in common by Lemma 3.1 it is clear that G' contains G as a proper subgraph. This proves the sufficiency of the conditions.

On the other hand, let $G(V, E, w)$ be the unique optimal realization of (M, d), where $V = M$, and suppose first that there are two edges $[x, y]$, $[y, z]$, such that

$$d(x, z) < d(x, y) + d(y, z).$$

Then replacement of the edges $[x, y]$, $[y, z]$ by a star with center v and edges $[v, x]$, $[v, y]$, $[v, z]$ of lengths $c(z, x, y)$, $c(x, y, z)$, and $c(y, z, x)$ contradicts the optimality of G.

It remains to be shown that G is not optimal or not unique if (ii) does not hold. Setting

$$s_1 = d(x, y) + d(t, z),$$
$$s_2 = d(x, t) + d(y, z),$$
$$s_3 = d(x, z) + d(y, t),$$

we thus have

$$s_1 \geqslant s_2 \qquad \text{and} \qquad s_1 \geqslant s_3.$$

From this follows that

$$d(x, y) - c(z, x, y) - c(z, y, t) = (s_1 - s_3)/2 \geqslant 0,$$

the expression being zero if and only if $s_1 = s_3$. We now introduce two new vertices u, v and replace the edges $[x, y]$, $[t, z]$ by the edges $[x, u]$, $[z, u]$, $[u, v]$, $[v, y]$, $[v, t]$ of lengths $c(z, x, y)$, $c(x, z, y)$, $(s_1 - s_3)/2$, $c(z, y, t)$, and $c(y, t, z)$, respectively. (If $(s_1 - s_3)/2 = 0$ we identify u with v and omit the edge $[u, v]$.) It is easily seen that the graph still realizes (M, d), thereby contradicting either the optimality or the uniqueness assumption.

COROLLARY 3.3. *Let $G(V, E, w)$ be an N-realization of (M, d) with $E \subseteq N$. Then G is the unique and optimal N-realization of (M, d) if and only if the following conditions are satisfied*:

(i) $d(x, z) = d(x, y) + d(y, z)$ *for all edges* $[x, y]$, $[y, z]$.

(ii) $d(x, y) + d(t, z) < \max\{d(x, t) + d(y, z), d(x, z) + d(y, t)\}$ *for all edges* $[x, y]$, $[t, z]$ *with no common endpoints.*

Proof. Clear.

THEOREM 3.4. *Let $G_i(V_i, E_i, w_i)$, $1 \leqslant i \leqslant k$, be optimal N_i-realizations of a metric (M, d) with mutually disjoint V_i, except for points in M. Let $G(V, E, w)$ be defined by*

$$V = \bigcup_{1 \leqslant i \leqslant k} V_i, \qquad E = \bigcup_{1 \leqslant i \leqslant k} E_i, \qquad w \mid E_i = w_i$$

and let N be the union of the N_i. Then G is an optimal N-realization of (M, d) if:

(i) $d(x, z) = d(x, y) + d(y, z)$ *whenever* $[x, y]$ *and* $[y, z]$ *are in different* N_i.

(ii) $d(x, y) + d(t, z) < \max\{d(x, t) + d(y, z), d(x, z) + d(y, t)\}$ *for all disjoint pairs* $[x, y]$, $[t, z]$ *in different* N_i.

Let n_i be the number of optimal N_i-realizations of (M, d) and suppose (i) *and* (ii) *hold. Then the number of optimal N-realizations of (M, d) is*

$$n = \prod_{1 \leqslant i \leqslant k} n_i.$$

In particular, G is the unique optimal N-realization of (M, d) if all the G_i are unique and optimal.

Proof. Clear.

4. Examples of Optimal Realizations

4.1. *Consider the complete bipartite graph $K_{2,3}$ (Fig. 1) with every edge having length 1. By Theorem 3.2 this graph is the unique optimal realization of the metric induced on its vertex set. (It was also shown to be optimal in* [11].)

4.2. *Let (M, d) be the metric on $\{x, y, z, t\}$ where all distances between distinct vertices are 2. Let N consist of the pairs $[x, y]$ and $[z, t]$. It is easy to see that (M, d) has exactly two optimal N-realizations: One consists of two disjoint edges $[x, y]$ and $[z, t]$ of length two and the other of a star $K_{1,4}$, every edge having length* 1.

4.3. *Let (M, d) be the metric induced by the graph G of Fig. 2 on its vertex set. Let N_1 consist of the pairs $[x, y]$, $[z, t]$ and N_2 of all the other pairs of distinct elements of M. Then (M, d) has exactly two optimal realizations by Theorem 3.4: One is the graph G itself and the other one is obtained by identifying the midpoints of the edges $[x, y]$, $[z, t]$ of G in an auxiliary vertex, as seen in Fig. 3.*

Theorem 4.4. *Suppose a metric (M, d) has a realization by a circuit C on at least four vertices with vertex set $M = \{v_i \mid 1 \leqslant i \leqslant n\}$ and edge set $\{[v_i, v_{i+1}] \mid 1 \leqslant i \leqslant n\}$, indices modulo n. This realization is unique and optimal if and only if*

$$d(v_i, v_{i+1}) + d(v_{i+1}, v_{i+2}) = d(v_i, v_{i+2})$$

for all i.

Proof. By Theorem 3.2 the condition is necessary. It is also sufficient if it implies condition (ii) of Theorem 3.2. It therefore suffices to show

$$d(v_1, v_2) + d(v_i, v_{i+1}) < \max\{d(v_1, v_i) + d(v_2, v_{i+1}), d(v_1, v_{i+1}) + d(v_2, v_i)\}.$$

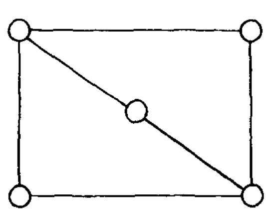

Figure 1

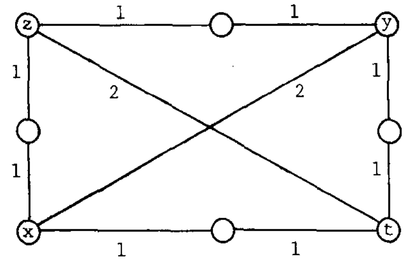

FIGURE 2

Since d is induced by C this inequality must also hold in C. We observe that it remains invariant if v_1 is interchanged with v_i and v_2 with v_{i+1}. As every shortest path in C from v_1 to v_i contains v_2 or v_{i+1} we can assume without loss of generality that there is one containing v_2. Similarly, every shortest path from v_2 to v_{i+1} contains v_1 or v_i. If there is one containing v_i we have

$$d(v_1, v_2) + d(v_i, v_{i+1}) < d(v_1, v_i) + d(v_2, v_{i+1}).$$

We can thus assume that there is one containing v_1. Then

$$d(v_2, v_{i+1}) = d(v_1, v_2) + d(v_1, v_{i+1}),$$
$$d(v_1, v_{i+2}) < d(v_1, v_{i+1})$$

and

$$\begin{aligned} d(v_i, v_{i+1}) &< d(v_i, v_{i+1}) + d(v_{i+1}, v_{i+2}) = d(v_i, v_{i+2}) \\ &\leqslant d(v_1, v_i) + d(v_1, v_{i+2}) < d(v_1, v_i) + d(v_1, v_{i+1}) \\ &= d(v_1, v_i) + d(v_2, v_{i+1}) - d(v_1, v_2). \end{aligned}$$

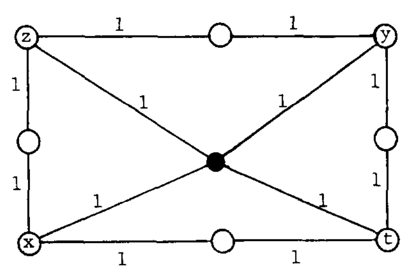

FIGURE 3

This yields the desired result

$$d(v_1, v_2) + d(v_i, v_{i+1}) < d(v_1, v_i) + d(v_2, v_{i+1}).$$

COROLLARY 4.5. *The cycle of length 4 is the unique optimal realization of the metric it defines on its vertices v_1, v_2, v_3, v_4 if and only if*

$$d(v_1, v_2) = d(v_3, v_4), \qquad d(v_1, v_4) = d(v_2, v_3).$$

Proof. By Theorem 4.4 we have

$$d(v_1, v_3) = d(v_1, v_2) + d(v_2, v_3) = d(v_3, v_4) + d(v_1, v_4),$$
$$d(v_2, v_4) = d(v_2, v_3) + d(v_3, v_4) = d(v_1, v_4) + d(v_1, v_2).$$

This immediately implies the assertion of the corollary.

4.6. *Let $G(V, E, w)$ be the finite lattice graph with vertex set*

$$V = \{(x_1, ..., x_m) \mid x_i \in N,\ 0 \leqslant x_i < n_i\},$$

where N denotes now the nonnegative integers, and edge set

$$E = \left\{ [(x_1, ..., x_m), (y_1, ..., y_m)] \;\middle|\; \sum_{1 \leqslant i \leqslant m} (x_i - y_i)^2 = 1 \right\}.$$

It is straightforward to verify that G is the unique optimal realization of the metric it defines on V if and only if there exist numbers $r_i = r_i(x_i)$, $1 \leqslant i \leqslant m$, such that the relation

$$w((x_1, ..., x_m), (y_1, ..., y_m)) = r_i \qquad \text{for} \quad x_i \neq y_i$$

holds for all edges $[(x_1, ..., x_m), (y_1, ..., y_m)]$. Furthermore, an induced connected subgraph H of G is an optimal realization of the metric it defines on its vertices if and only if the above relation holds for the edges of H.

5. CUTPOINTS IN OPTIMAL REALIZATIONS

Our next theorem is another valuable tool for the investigation of the optimality of large classes of graph realizations. It was first published in [10]. We give a new and shorter proof. To state this result it is convenient to look at edges of G as geometric lines. Any point c on such a line is a potential new vertex of G. A point c on the edge $[a, b]$ of length w becomes a vertex if $[a, b]$ is replaced by edges $[a, c]$ and $[c, b]$ of lengths w_1 and w_2, respectively, where $w_1 + w_2 = w$.

THEOREM 5.1 [10]. *Suppose (M, d) is a metric to which there exists a partition of M into two nonempty subsets K, L and a mapping $f\colon M \to \mathbb{R}^+$ with the property*

$$d(x, y) \leqslant f(x) + f(y),$$

equality holding whenever $x \in K$ and $y \in L$, and where $f(x) > 0$ at least once in K and once in L.

Then every optimal realization G of M has a cutpoint c or a bridge with a point c on it such that $d_G(x, c) = f(x)$ for all x in M.

Proof. Let H be a realization of M and let S be the set of all shortest paths between points in K and points in L. Let the origin $o(s)$ of $s \in S$ be in K and the terminus $t(s)$ be in L. Since $d_H(o(s), t(s)) = d(o(s), t(s)) = f(o(s)) + f(t(s))$ there is exactly one point c_s on s with

$$d_H(o(s), c_s) = f(o(s)) \qquad \text{and} \qquad d_H(c_s, t(s)) = f(t(s)).$$

We define $s_x = d_H(x, c_s)$ for every x in M. For every u in K we then have

$$\begin{aligned} f(u) + f(t(s)) = d(u, t(s)) &\leqslant d_H(u, c_s) + d_H(c_s, t(s)) \\ &= s_u + f(t(s)) \end{aligned}$$

and $f(u) \leqslant s_u$. Analogously $f(v) \leqslant s_v$ for $v \in L$. Hence

$$f(x) \leqslant s_x \qquad \text{for all} \quad x \in M.$$

If all points c_s of H coincide in a point c, then this point is the point c of the statement of the theorem. If they do not, then form H^* from H by identifying some or all points c_s in a single point c. We claim that H^* still realizes (M, d). For, suppose there are points $u, v \in M$ with

$$d_{H^*}(u, v) < d(u, v) = d_H(u, v).$$

Then we must have a shortest path in H^* from u to v via c, i.e., there must exist paths s, r in S with

$$s_u + r_v < d_H(u, v) \leqslant f(u) + f(v) \leqslant s_u + r_v,$$

which is not possible. A similar argument shows that

$$a_u = f(u) \qquad \text{and} \qquad a_v = f(v)$$

for any path a in S and every pair of points u, v in K for which there exists a shortest path from u to v via c_a. In this case there are also shortest paths from u and v to $t(a)$ via c_a.

If not all points c_s of H coincide, choose two distinct points c_a and c_b. Let A be the set of shortest paths containing c_a and B the set of shortest paths containing c_b. Let $o(A)=\{o(s) \mid s \in A\}$ and $t(A)$, $o(B)$, $t(B)$ be defined analogously. Either $o(A) \subseteq o(B)$ or not.

Suppose $o(A) \subseteq o(B)$. Then

$$d_H(u, c_a) = a_u = f(u) = b_u = d_H(u, c_b)$$

for all $u \in o(A)$. For any shortest path between a point u of $o(A)$ to c_a there thus exists a path of the same length from u to c_b, and of course this path cannot meet c_a. Furthermore, if there is a shortest path between points u, v of K via c_a there are shortest paths from u and v to $t(a)$ via c_a, as we have seen above, and thus $u, v \in o(A)$. But then there is also a shortest path from u to v which contains c_b but not c_a. The metric (M, d) will thus still be realized if we delete all edges of H incident with c_a which are on a shortest path from $o(A)$ to c_a, provided that we identify c_a with c_b at the same time.

Suppose now that $o(A) \backslash o(B)$ is nonempty. Then, $t(B) \backslash t(A)$ must be empty, for there are no shortest paths from $o(A) \backslash o(B)$ to $t(B) \backslash t(A)$. This implies $t(B) \subseteq t(A)$, which is equivalent to the previous case.

In all cases the assumption that H is optimal is violated. This completes the proof.

Corollaries 5.2 and 5.3 have obvious proofs.

COROLLARY 5.2. *Let (M, d) satisfy the assumptions of Theorem* 5.1 *and let (M', d') be a metric on $M \cup \{u\}$, where $d' \mid M \times M = d$ and, for any $x \in M$, $d'(u, x) = f(x)$. Then (M, d) and (M', d') have the same optimal realizations, except for the possibility that u is not a vertex but simply a point on a bridge in the optimal realization of (M, d).*

The graphs in Figs. 4 and 5 illustrate Corollary 5.2 with $M = \{1, 2\}$, $K = \{1\}$, $L = \{2\}$, $f(1) = 1$, $f(2) = 3$, $d(1, 2) = d'(1, 2) = 4$, $d'(1, u) = 1$, and $d'(2, u) = 3$.

COROLLARY 5.3. *Let (M, d) and (M', d') be defined as in Corollary* 5.2. *If the submetrics of (M', d') on $K \cup \{u\}$ and $L \cup \{u\}$ have unique optimal realizations, then this is also true of (M', d') and (M, d).*

Before we state and prove Corollary 5.4, let us call a point z of a metric

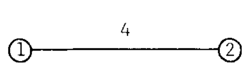

FIGURE 4

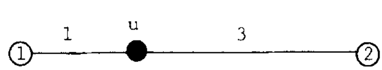

FIGURE 5

(M, d) an *endpoint* when $c(z) = \min\{(d(x, z) + d(z, y) - d(x, y))/2\} > 0$, the minimum being taken over all x, y in M different from z.

COROLLARY 5.4 [6]. *If a is an endpoint of (M, d), then a is the endpoint of a pendant edge of length $c(a)$ in every optimal realization of the metric.*

Proof. Set $K = \{a\}$, $L = M - \{a\}$, $f(a) = c(a)$ and $f(x) = d(a, x) - f(a)$ for all $x \in L$. If L has at least two elements, then the assumptions of the theorem are satisfied.

To every endpoint a of a metric (M, d) there thus exists a metric (M', d') on $(M\backslash\{a\}) \cup \{u\}$, where $d'(x, y) = d(x, y)$ for all x, y in $M\backslash\{a\}$ and $d'(x, a) = d(x, a) - c(a)$, such that every optimal realization of (M, d) can be obtained from one of (M', d') by attaching an edge of length $c(a)$ to u. We can thus reduce the problem of optimally realizing (M, d) to the problem of optimally realizing (M', d'). We call this process *reduction of an endpoint* and observe that M' has fewer endpoints or fewer points than M.

COROLLARY 5.5. *If the tree T is an optimal realization of the metric (M, d), then T is the unique optimal realization of (M, d).*

Proof. By reduction of endpoints.

COROLLARY 5.6 [16]. *Every metric on four points has a unique optimal realization by a (possibly degenerate) rectangle with (possibly degenerate) pendant edges from the corners of the rectangle.*

Proof. By reduction of endpoints and application of Corollary 5.5 or 4.5.

Let (M, d) be the metric on four points x, y, z, t and consider again the sums s_1, s_2, and s_3 introduced in the proof of Theorem 3.2. Using Corollary 5.6 and straightforward calculations we obtain

COROLLARY 5.7 [2]. *A metric on four points is tree-realizable if and only if two of the sums s_1, s_2, s_3 are equal and not smaller than the third.*

Following Buneman [2], let us say a metric satisfies the *four-point condition* if, for every submetric on four points, two of the respective sums s_1, s_2, s_3, are equal and not smaller than the third.

THEOREM 5.8 [14]. *A finite metric is tree-realizable if and only if it satisfies the four-point condition, i.e., if and only if every submetric on four points is tree-realizable.*

Proof. The condition is obviously necessary. Suppose therefore that (M, d) satisfies the four-point condition. Then (M, d) has at least two endpoints. For, let a, b be a pair of points of maximal distance in (M, d). If b

is not an endpoint, it must be on a shortest path between points p and q of M, both different from b. Since the submetric on $\{a, b, p, q\}$ is tree-realizable it is easily seen that either $d(a, p)$ or $d(a, q)$ is larger than $d(a, b)$. The proof is completed by reduction of endpoints, each step reducing the number of points or the number of endpoints.

For integer valued metrics, the result also holds when M is infinite. If we generalize the concept of a tree to mean a nonempty space T which has no subspace homeomorphic to a circle and in which any two points are the endpoints of an interval (i.e., for any two points, x, y there exists an isometry g of an interval $[0, d]$ into T with $g(0) = x$ and $g(d) = y$), then Theorem 5.8 remains true also for real-valued infinite metrics. For a proof see [8] and for applications to groups with length functions see [9].

Perhaps the most important consequence of Theorem 5.1 is the possibility of constructing certain optimal realizations using known optimal realizations as building blocks in a sense which is made precise in the next theorem. Before stating it, we recall that a *block* of a graph is a maximal two-connected subgraph or a bridge.

THEOREM 5.9. *Let G be a minimal realization of a metric (M, d), let $G_1, \ldots, G_k$ be the blocks of G and let M_i be the union of the points of M in G_i together with the cutpoints of G in G_i. If every G_i is an optimal realization of the metric induced by G on M_i, then G is also optimal. If every G_i, besides being optimal, is also unique, then G is optimal and unique too.*

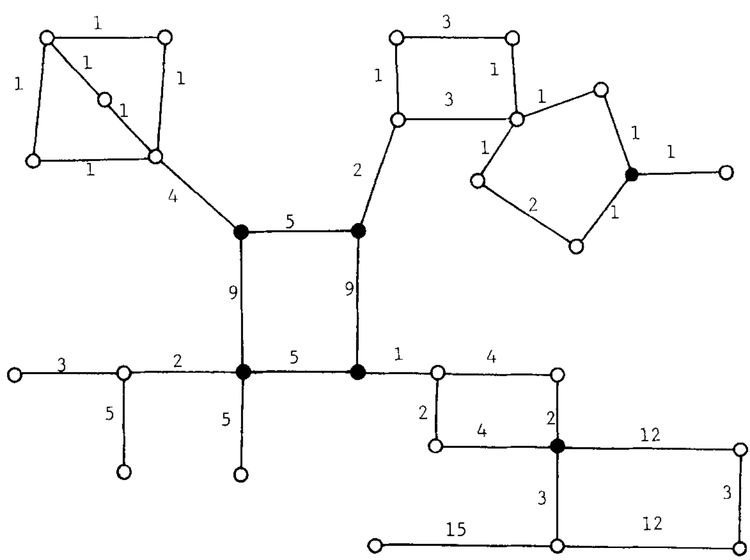

FIGURE 6

Proof. By induction with respect to k. The case $k = 1$ being trivial, assume that the assertion is true for all graphs with less than k blocks. Since every graph which is not a block itself contains a block having exactly one cutpoint of the graph, we can assume that G_k is such a block. Let c be the cutpoint separating it from $\bigcup_{i<k} G_i$. We set $K = M \cap \bigcup_{i<k} G_i$, $L = M \cap G_k$ and $f(x) = d(x, c)$ for all $x \in M$. Since G is minimal K, L and f satisfy the conditions of Theorem 5.1.

If $\bigcup_{i<k} G_i$ is an optimal realization of the metric on $K \cup \{c\}$ induced by G, then the assertion of the theorem follows by an application of Corollary 5.2. If $\bigcup_{i<k} G_i$ is unique and optimal, then we apply Corollary 5.3. This completes the proof.

In this theorem it is possible that a block G_i contains no point of M. This is already exemplified by metrics on four points and also by the more complicated graph of Fig. 6 which is the unique and optimal realization of the metric it induces on its white vertices. (The black ones are auxiliary.)

Acknowledgments

Research by Professors Simões-Pereira and Zamfirescu was done, for the most part, during a visit to the University of Coimbra in the Summer of 1982. Support was given by the Scientific Affairs Division of the North Atlantic Treaty Organization under collaborative research Grant RG 167.81 and by grants from The City University of New York PSC/CUNY Research Award Program.

Note added in proof. Theorem 2.2 has also been proved independently of us by Andreas W. M. Dress, "Trees, Tight Extensions of Metric Spaces and the Cohomological Dimension of Certain Groups—A Note on Combinatorial Properties of Metric Spaces," unpublished manuscript.

References

1. F. T. Boesch, Properties of the distance matrix of a tree, *Quart. Appl. Math.* **26** (1968), 607–609.
2. P. Buneman, A note on the metric properties of trees, *J. Combin. Theory Ser. B* **17** (1974), 48–50.
3. S. Chaiken, A. K. Dewdney, and P. S. Slater, An optimal diagonal tree code, *SIAM J. Algebraic Discrete Methods* **4** (1983), 42–49.
4. A. K. Dewdney, Diagonal tree codes, *Inform. and Control* **40** (1979), 234–239.
5. M. Eigen, W. Gardiner, P. Shuster, and R. Winkler-Oswatitsch, The origin of genetic information, *Sci. Amer.*, April 1981, 88–118.
6. S. L. Hakimi and S. S. Yau, Distance matrix of a graph and its realizability, *Quart. Appl. Math.* **22** (1964), 305–317.
7. W. Imrich, Realisierung von Metriken in Graphen, *Österreich. Akad. Wiss. Math.-Natur. Kl. Sitzungsber. II* **178** (1969), 19–24.

8. W. IMRICH, On metric properties of tree like spaces, *in* "Contributions to Graph Theory and Its Applications," pp. 129–156, Intern. Colloq. Oberhof, 1977, Technische Hochschule Ilmenau, Ilmenau, 1977.
9. W. IMRICH AND G. SCHWARZ, Trees and length functions in groups, *Ann. Discrete Math.* **17** (1983), 347–359.
10. W. IMRICH AND E. STOCKIĬ, On optimal embeddings of metrics in graphs, *Sibirsk. Mat. Ž.* **13** (1972), 558–565.
11. J. LAWRENCE, C. R. JOHNSON, AND E. HOWE, The structure of distances in networks, to appear.
12. A. N. PATRINOS AND S. L. HAKIMI, The distance matrix of a graph and its tree realization, *Quart. Appl. Math.* **30** (1972), 255–269.
13. P. H. SELLERS, The theory and computation of evolutionary distances: pattern recognition, *J. Algorithms* **1** (1980), 359–373.
14. J. M. S. SIMÕES-PEREIRA, A note on the tree realizability of a distance matrix, *J. Combin. Theory* **6** (1969), 303–310.
15. J. M. S. SIMÕES-PEREIRA, A note on optimal and suboptimal digraph realizations of quasidistance matrices, *SIAM J. Algebraic Discrete Methods* **5** (1984), 117–132.
16. J. M. S. SIMÕES-PEREIRA AND C. M. ZAMFIRESCU, Submatrices of non-tree-realizable distance matrices, *Linear Algebra Appl.* **44** (1982), 1–17.
17. E. A. SMOLENSKIĬ, A method for the linear recording of graphs, *Ž. Vyčisl. Mat. i Mat. Fiz.* **2** (1962), 371–372. [Russian]
18. B. C. TANSEL, R. L. FRANCIS, T. J. LOWE, AND M. L. CHEN, Duality and distance constraints for the non-linear *p*-center problem and covering problem on a tree network, *Oper. Res.* **30** (1982), 725–744.
19. K. A. ZARECKIĬ, Constructing a tree on the basis of a set of distances between the hanging vertices, *Uspehi Mat. Nauk* **20** (1965), 90–92. [Russian]

582b/36/1-2