ELSEVIER

Discrete Mathematics 192 (1998) 337–346

DISCRETE MATHEMATICS

Trees related to realizations of distance matrices

Sacha C. Varone*

Dept. de Mathematiques, Ecole Polytechnique Federal de Lausanne, MA-Ecublens, CH-1015 Lausanne, Switzerland

Received 16 December 1996; revised 27 May 1997; accepted 10 June 1997

Abstract

Let D be a distance matrix. We use heuristics for finding the largest tree that realizes a submatrix of D, or for decomposing D in a minimum number of tree-realizable submatrices.

0. Introduction

Definition 1. A *metric space* is a couple (M,d) such that M is a set and d is a function defined on $M \times M$ satisfying

- $\forall x, y\ d(x,y) = d(y,x)$.
- $\forall x, y\ d(x,y) \geqslant 0$ and $d(x,y) = 0 \Leftrightarrow x = y$.
- $\forall x, y, z\ d(x,z) \leqslant d(x,y) + d(y,z)$.

Moreover (M,d) is a *finite metric space* if $|M| < \infty$.

The finite metric spaces are represented by symmetrical distance matrices.

Definition 2. A matrix $D = (d_{ij})$ with non-negative values $d_{ij} \in \mathbb{R}^+$ is called a *distance matrix* iff $d_{ii} = 0$, $d_{ij} > 0$ and $d_{ij} \leqslant d_{ik} + d_{kj}\ \forall i,j,k$.

Let $G = \langle V, E, w \rangle$ be a weighted (non directed) graph, where V is the set of vertices, E is the set of edges and $w : E \to \mathbb{R}^+$ is the function assigning a length (weight) to each edge. The shortest path in G between i and j will be denoted by d_{ij}^G.

Definition 3. A weighted graph $G = \langle V, E, w \rangle$ *realizes* the distance matrix D of order n if and only if $\{1, \ldots, n\} \subseteq V$ and $d_{ij}^G = d_{ij}\ \forall i, j = 1, \ldots, n$. G is then called a *realization* of D.

* E-mail: sacha.varone@epfl.ch.

0012-365X/98/$19.00
PII S0012-365X(98)00081-8

The embedding of finite metric spaces in graphs, or in other words the realization of distance matrices by graphs, is a problem that occurs in various fields such as the study of electrical networks [7], coding techniques [4], psychology [3] or genetics [6].

We restrict our discussion to symmetrical distance matrices. Those realizable by trees are the best known. This paper presents a way to find tree-realizable distance matrices that are submatrices of non tree-realizable distance matrices. The first part contains the basic definitions and theorems. The second part deals with the special case of distance matrices with infinite values. The next part explains how to construct a graph related to non tree-realizable submatrices and gives some statistical results on this graph. In the last part the search for a tree of maximal order and a minimal forest is solved by heuristics and algorithms.

1. Preliminaries

Definition 4. A realization $G=\langle V,E,w\rangle$ of D is called *optimal* if the sum $\sum_{e\in E} w(e)$ is minimal among all realizations of D.

Theorem 5 (Hakimi and Yau [7]). *No optimal realization contains a triangle.*

We will consider only optimal realizations that have no internal vertices of degree two, since they can always be removed.

Definition 6. A graph G is called a *sub-realization* of a distance matrix D if it is a realization of a principal submatrix [1] of D.

Definition 7. A distance matrix D is called *tree-realizable* if there exists a weighted tree $T=\langle V,E,w\rangle$ that realizes D.

Theorem 8 (Dress [5], Simões-Pereira et al. [11]). *Every finite metric space (M,d) has an optimal realization.*

Lemma 9. *Every distance matrix of order $n\leqslant 3$ is tree-realizable.*

Proof. We only consider the non-trivial case $n=3$.

Suppose the distance matrix D of order 3 is not tree-realizable. Consider G a realization of D. Since all values are finite, G is connected. So there must be a cycle.

[1] A $m\times m$ matrix restriction of the $n\times n$ matrix D $(n\geqslant m)$ in which a same permutation of the rows and the columns has been done.

Since the elements on the diagonal are all zero there is no cycle of order 1. If there is a cycle of order 2, we could remove the heaviest edge (or any of them if both have the same weight) without losing realizability. So the cycle has to be of order 3. But the realization is not optimal by Theorem 5. Since an optimal realization exists by Theorem 8, D has to be tree-realizable. □

Note that for D of order 3 the length of the edges of a tree-realization are the numbers l_i given by $2l_i = d_{ij} + d_{ik} + d_{jk}$, where $i \neq j \neq k$.

Theorem 10 (Hakimi and Yau [7]). *If D has a tree-realization, then this realization is optimal and unique.*

Theorem 11 (Simoes-Pereira [9]). *The matrix D has a tree-realization if and only if all its principal submatrices of order 4 have a tree-realization.*

Theorem 12 (Boneman [1], Imrich [8]). *A matrix D of order 4 is tree-realizable if and only if, among the three sums $s_1 = d_{12} + d_{34}$, $s_2 = d_{13} + d_{24}$, $s_3 = d_{14} + d_{23}$, two are equal and not smaller than the third one.*

Remark 1. The condition in Theorem 12 is known as the 4-*point condition*.

Culberson and Rudnicki [2] gave an $O(n^2)$ running time algorithm which constructs a tree that realizes a given tree-realizable distance matrix. Their algorithm is based on the simple observation: for any three vertices in a tree, there is a unique vertex which lies on all three simple paths between pairs of vertices.

2. Disconnected realizations

We extend the definition of a distance matrix D in the case of positive infinite values and we want to partition D in the smallest number of distance matrices whose realizations are connected.

The meaning of an infinite distance $d_{ij} = \infty$ between two points i and j is that these points are neither connected together nor there exists a path that joins them in any realization of D. We give the mathematical formulation.

Proposition 13. *Let D be a distance matrix and G any realization of D. $d_{ij} < \infty \Rightarrow \exists Path_{ij} = (k_0, k_1, \ldots, k_r)$ such that $r < \infty$, $k_0 = i$, $k_r = j$, $d^G_{k_u k_{u+1}} < \infty$ $\forall u = 0, \ldots, r-1$.*

Proof.

$\Rightarrow$

Take $r=1$. So $d^G_{k_0k_1}=d_{ij}<\infty$.

$\Leftarrow$

$d_{ij}=d^G_{ij}\leqslant d^G_{k_0k_1}+d^G_{k_1k_2}+\cdots+d^G_{k_{r-1}k_r}<\infty$. □

Corollary 14. *If the distance matrix D contains an entry of infinite value then all realizations of D are disconnected.*

Proof. Immediate from Proposition 13. □

Consider the relation $\sim$ defined as $x\sim y\Rightarrow d_{xy}<\infty$. It is easy to verify that $\sim$ is an equivalence relation. It defines a partition of the entries of D into equivalence classes.

Corollary 15. *Let D^i be the submatrix of D induced by the ith equivalence class, G^i one of its realization and k the number of equivalence classes. Then each G^i is connected and $\bigcup_{i=1}^{k} G^i$ is a realization of D.*

Proof. Immediate from Proposition 13. □

We give an algorithm based on Proposition 13 for finding the partition defined above. For each class, the algorithm checks a line of the matrix D to find all the vertices that belong to this class.

Algorithm.

```
Input: D a n × n distance matrix
Output: V the n-vector of connected components
NbSelected = 0
Num = 0
Next = 1
V[i] = 0   i = 1,...,n
While (Nbselected < n)
    Current = Next
    Num = Num + 1
      For j = 1 to n
          if V[j] = 0
            if d_{Current,j} ≠ ∞
                NbSelected = Nbselected + 1
                V[j] = Num
             else Next = j
```

This algorithm runs in $O(n^2)$. Now we will only consider distance matrices with finite values.

3. Tree and non-tree realizable submatrices

Let D be a distance matrix of order n and consider the graph G_{ns} on n vertices constructed as follows:

For any four different entries i,j,k and l that do not satisfy the 4-point condition 12, we construct a clique between these 4 vertices. G_{ns} is the graph obtained in which all 4-uplets of D have been tested.

Proposition 16. *Every induced subgraph of G_{ns} that does not contain a clique on 4 points corresponds to a tree-realizable principal submatrix of D.*

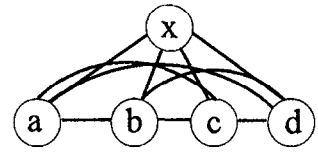
Proof. Immediately follows. □

Remark 2. G_{ns} can contain cliques on 4 vertices that correspond to a tree-realizable principal submatrix of D.

Example 17.

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

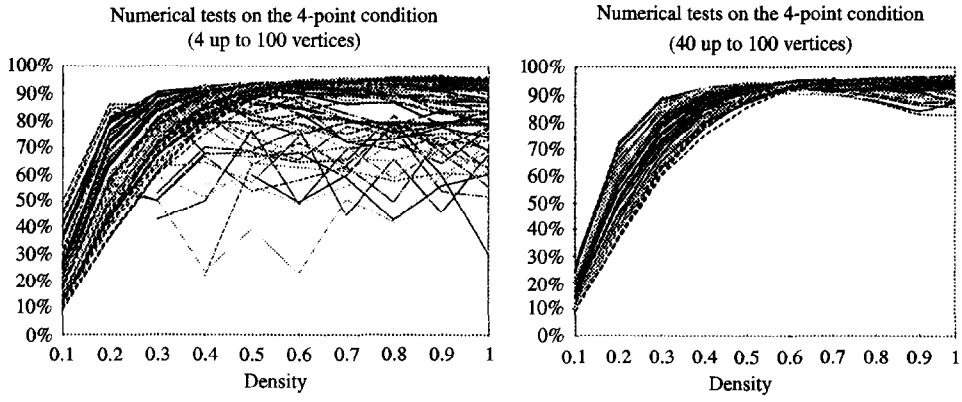
with a,b,c,d,x its associated vertices. The graph G_{ns} is



The principal submatrices related to the vertices a,b,c,x; a,b,d,x; a,c,d,x are all non-tree-realizable. The principal submatrix related to the vertices a,b,c,d is tree-realizable but there is a clique on these four vertices in G_{ns}.

Define F as the set of submatrices of size 4 of D that do not satisfy the 4-point condition. We give some numerical tests on $|F|$ as follows: we have created 10 series of random graphs $G=(V,E)$ with 4 up to 100 vertices and density 0.1 up to 1.0 (step 0.1). Then we have calculated the associated distance matrices and the average number of $|F|$. The figures below shows the graphical results. The first one is for random graphs with 4 up to 100 vertices and the second figures contains only the results for random

graphs with 40 up to 100 vertices.



The x-axis represents the density of the random graph $G=(V,E)$ and the y-axis $100 \times |F|/C_4^{|V|}$. This graphic leads to the assumption that there exists an asymptotical convergence in the number of vertices.

4. Maximal tree and minimal forest

Theorem 18 (Yannakakis [12]). *The connected maximum subgraph problem for graph properties π that are hereditary (if G has π then all subgraphs of G have π) on induced subgraphs, nontrivial (some graphs do have π, some do not) and interesting (graph of arbitrarily large order may have π) on connected graphs, is NP-hard.*

4.1. Maximal tree

We are looking for the largest tree that is a sub-realization of a distance matrix. It follows from Theorem 18 that:

Lemma 19 (Simões-Pereira and Zamfirescu [10]). *The problem of finding a tree-realizable submatrix of maximum order of a non-tree-realizable distance matrix is NP-hard.*

4.1.1. Graph method

Determination of a tree-realizable principal submatrix. We want to determine the largest tree that is a sub-realization of the distance matrix D. First we construct the graph G_{ns} as in Section 3, then we solve the graph coloring problem (GCP) on G_{ns}. The minimum number of colors needed to color G_{ns} will be denoted $\chi(G_{ns})$. As this problem is known to be *NP*-complete, we use an heuristic solution method. Then we

select the vertices whose colors are one of the three most frequent colors. Define S as the set of these vertices. By Proposition 16, S corresponds to a tree-realizable principal submatrix of D.

Search for a maximal tree-realizable principal submatrix. We want to increase the cardinality of the set S defined above. According to Remark 2, we have

Proposition 20. *Consider a clique K on 4 vertices in G_{ns} and assume that K corresponds to a tree-realizable principal submatrix of D. Then the degree of all these 4 vertices in G_{ns} is greater or equal to 4.*

Proof. The degree of these 4 vertices a,b,c,d cannot be less than 3 since they form a clique.

Suppose one of them, say a, is of degree 3. Each edge introduced in G_{ns} by our construction results from a set of 4 points corresponding to a non-tree-realizable submatrix. Then a and three other vertices $\{b',c',d'\}$ correspond to a non-tree-realizable submatrix. Since the degree of a is 3, a cannot be involved in another clique and $\{b,c,d\}=\{b',c',d'\}$. This contradicts the fact that a,b,c,d correspond to a tree-realizable principal submatrix. □

We propose a basic heuristic (similar to the descent method) to increase the cardinality of the set S.

1. If all vertices in $G_{ns}\backslash S$ are of degree less or equal to 3 then STOP (S is optimal)
2. Sort the vertices of $G_{ns}\backslash S$ by non increasing degree in a list L
3. Set the current vertex v_c to the first item in the list L
4. If the degree of v_c is strictly less than 4, STOP (S is a local maximal tree-realizable principal submatrix)
5. If the set $S\cup\{v_c\}$ satisfies the 4-point condition then $S=S\cup\{v_c\}$
6. Else set the current verted v_c to the next vertex in the list L (STOP if there is no more vertices) and go to 4

Notice that about the part 2 of this heuristic, one may also list the vertices of G_{ns} by nondecreasing degrees (and then skip the part 4). As pointed out by one of the referees the quality of the output strongly depends on the initial coloring of G_{ns}. Maximizing the size of S can be obtained more efficiently by means of different local search algorithms (e.g. Taboo Search, Genetic Algorithms).

4.1.2. Hypergraph method

Let $H=(V,F)$ be a 4-uniform[2] hypergraph such that each edge is a set of four vertices that correspond to entries in the distance matrix D that do not satisfy the 4-point condition 12. Let S be a stable set in H. S is a set of points that corresponds to a tree-realizable principal submatrix of D. The problem is therefore to maximize the cardinality of S.

[2] A 4-uniform hypergraph is a hypergraph in which each edge is of cardinality 4.

Mathematical formulation. Let A be the incident $n \times m$ vertex-edge matrix of H and let y be a n-vector, $y \in \{0,1\}^n$.

$$\text{Max} \quad \vec{1} \cdot y$$
$$\text{with} \quad A^t y \leqslant \vec{3}.$$

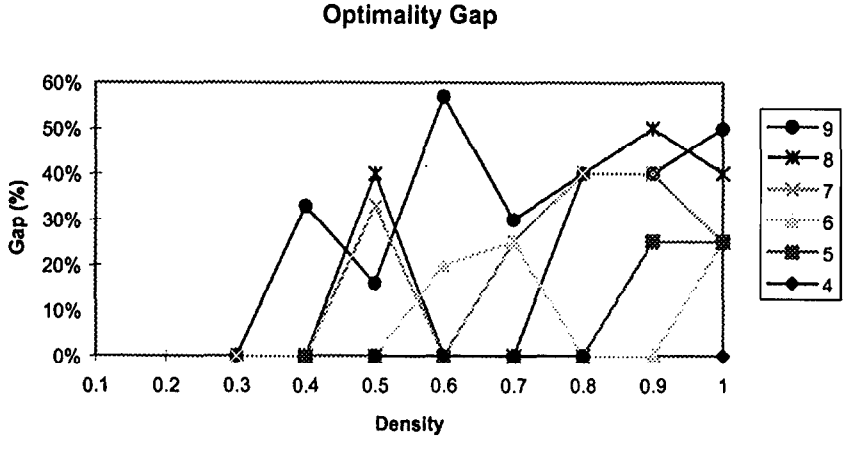
Lemma 21. *Let H be a hypergraph defined as above and S be a maximal stable set of H. Let T be a maximal tree-realizable submatrix of D and let $|T|$ denote the order of T. Then $|S| = |T|$.*

Proof. Each stable set of H corresponds to a tree-realizable submatrix of D. Therefore $|S| \leqslant |T|$.

Suppose $|S| < |T|$. Then there must exist four vertices s_1, s_2, s_3, s_4 in T such that $\{s_1, s_2, s_3, s_4\}$ is an edge of H. This edge corresponds to a violation of the four-point condition. So T does not define a tree-realizable distance matrix, a contradiction. □

4.1.3. Optimality gap

If the Graph method is used and the associated GCP can be solved to optimality then we can compare the results with those of the Hypergraph method. The tests have been done on graphs from 4 up to 9 vertices and the problems were solved using a Branch and Bound method. We call *optimality gap* the difference between the results of these two methods. This optimality gap justifies the postoptimisation proposed in the graph method.



4.2. Minimal forest

We are looking for a partition of a distance matrix such that all elements of this partition is tree-realizable and the cardinality of this partition is minimal.

4.2.1. Graph method

We construct the graph G_{ns} as explained in Section 3 and solve the associated GCP. By Proposition 16 each set of vertices for which no more than three colors is used induce a tree-realizable submatrix of D. Therefore $\lceil \chi(G_{ns})/3 \rceil$ represents an upper bound on the minimal number of elements of the partition.

4.2.2. Hypergraph method

The problem can be formulated as a coloring problem on a 4-uniform hypergraph $H=(V,F)$ (H as in 4.1.2).

Mathematical formulation. Let A be the incident $n \times m$ vertex-edge matrix of H and let x be a $n-2$-vector and y^i be a n-vector $i=1 \ldots n-2$.

$$
\begin{array}{lll}
\text{Min } \vec{1} \cdot x & & \\
\text{with } x_i \geqslant y_j^i & \forall j=1 \ldots n & \text{All used colors are counted,} \\
A^t y^i \leqslant \vec{3} & \forall i=1 \ldots n-2 & \text{No monochromatic edge,} \\
y^1 + \cdots + y^{n-2} = \vec{1} & & \text{All vertices have to be colored,} \\
x \in \{0,1\}^{n-2} & y^i \in \{0,1\}^n. &
\end{array}
$$

We can understand this formulation as follows:

x is a counter of the number of used colors.

$$
x_i = \begin{cases} 1 & \text{if color } i \text{ is used } (\exists j \text{ such that } y_j^i = 1), \\ 0 & \text{else.} \end{cases}
$$

i is a specific color. As G is 4-uniform, we need at most $n-2$ colors.

y^i represents the vertices colored with color i.

$$
y_j^i = \begin{cases} 1 & \text{if vertice } j \text{ is colored with color } i, \\ 0 & \text{else.} \end{cases}
$$

Lemma 22. *Let P be a minimal partition of a distance matrix D into tree-realizable submatrices and let H be a hypergraph defined as above. Then $\chi(H)=|P|$.*

Proof. Let C_i be the set of vertices colored by color i in a coloring of H with $\chi(H)$ colors. Each set C_i represents a tree-realizable submatrix of D and, as one vertex of H receives one and exactly one color, $C_i \cap C_j = \emptyset$, $i \neq j$, $i,j=1 \ldots \chi(H)$. So $\{C_1, \ldots, C_{\chi(H)}\}$ is a partition of D into tree-realizable submatrices and $\chi(H) \geqslant |P|$.

Suppose $\chi(H) > |P|$. Then there must exist four vertices s_1, s_2, s_3, s_4 that belong to a class P_k of the partition P and such that $\{s_1, s_2, s_3, s_4\}$ is an edge of H. This edge corresponds to a violation of the four-point condition. So the class P_k does not define a tree-realizable distance matrix, a contradiction. □

Acknowledgements

The research presented in this paper results from fruitful discussions with Professor A. Hertz in Winter 95.

References

[1] P. Buneman, A note on metric properties of trees, J. Combin. Theory Ser. B 17 (1974) 48–50.
[2] J.C. Culberson, P. Rudnicki, A fast algorithm for constructing tree from distance matrices, Inform. Process. Lett. 30 (1989) 215–220.
[3] J.A. Cunningham, Free trees and bidirectional trees as representations of psychological distance, J. Math. Psychol. 17 (1978) 165–188.
[4] A.K. Dewdney, Diagonal tree codes, Inform. and Control 40 (1979) 234–239.
[5] A.W.M. Dress, Trees, tight extensions of metric spaces, and the cohomological dimension of certain groups: a note on combinatorial properties of metric spaces, Adv. Math. 53 (1984) 321–402.
[6] E. Eigen, W. Gardiner, P. Shuster, R. Winkler-Oswatitsch, The origin of genetic information, Sci. Amer. (1981) 88–118.
[7] S.L. Hakimi, S.S. Yau, Distance matrix of a graph and its realizability, Quart. J. Appl. Math. 22 (1964) 305–317.
[8] W. Imrich, On metric properties of tree like spaces, Contrib. Graph Theory Appl. (1977) 129–156.
[9] J.M.S. Simões-Pereira, A note on the tree realizability of a distance matrix, J. Combin. Theory 6 (1969) 303–310.
[10] J.M.S. Simões-Pereira, C.M. Zamfirescu, Submatrices of non-tree realizable distance matrices, Linear Algebra Appl. 44 (1982) 1–17.
[11] J.M.S. Simões-Pereira, C.M. Zamfirescu, W. Imrich, On optimal embeddings of metrics in graphs, J. Combin. Theory 36B (1984) 1–15.
[12] M. Yannakakis, The effect of a connectivity requirement on the complexity of maximum subgraph problems, J. Assoc. Comput. Mach. 26 (1979) 618–630.