A summary of

Distance realization problems with applications to Internet tomography

by Fan Chung, Mark Garrett, Ronald Graham, and David Shallcross

In recent years, a variety of graph optimization problems have arisen in which the graphs involved are much too large for the usual algorithms to be effective. In these cases, even though we are not able to examine the entire graph (which may be changing dynamically), we would still like to deduce various properties of it, such as the size of a connected component, the set of neighbors of a subset of vertices, mapping usage patterns, assessing the performance of software/hardware and dealing with a variety of security and reliability issues. An area of recent interest in the Internet research community is to measure and monitor various aspects of the Internet, such as its performance, or its topological structure. We consider the problem of inferring network topology from very non-intrusive measurements, such as the delay of packets sent between pairs of network monitoring nodes.

Here, we focus on several fundamental problems in combinatorial algorithms, which we call *distance realization problems*. The measured delays between network monitors can be thought of as distances between the nodes in a graph. We investigate the problem of reconstructing the entire graph based on a certain subset of distances. Our algorithms are on a project, "Independent Monitoring for Network Survivability" at Telcordia Technologies. Similar projects have been actively pursued by various research groups [3, 10, 11, 20], especially on the statistical and visualization aspects. In addition to the applications to Internet tomography, the distance realization problems that we discuss here turn out to have a large number of applications in computational biology (e.g., constructing phylogenetic trees from genetic distances among living species), classification, etc. [1, 2, 13].

Here we will formulate several versions of this problem, mention the relevant results, both new and old, and discuss their algorithmic implications.

First, we give some definitions. For a matrix D with rows and columns indexed by S, we say D has a *realization* if there is a graph whose node set contains S, and $D(u, v)$ is the distance between u and v. It is easy to see that a matrix D has a realization if its entries are non-negative and satisfy the triangle inequality.

$$D(u, v) + D(v, w) \geq D(u, w). \tag{1}$$

It turns out [18] that this necessary condition is also sufficient (as indicated by the example above). We say that D is a *distance matrix* for S if $D(u, v)$, $u, v \in S$, is non-negative and satisfies (1).

Problem 1: For a given distance matrix D on a set S of terminal nodes, find a graph G which is a realization of D so that the total sum of all edge lengths of G is minimized.

The above problem was first proposed by Hakimi and Yau [18] in 1965 who also gave an algorithm which will lead to the solution for the special case that the realization of the distance matrix is a tree. Since then, an extensive literature has developed for this problem. Special attention has been given to the case of *tree realizations*, i.e., when the graph that realized the distance matrix is a tree. Necessary and sufficient conditions for a distance matrix realizable by a tree were given in several papers [2, 13, 23, 22, 24]. An $O(n^2)$ time algorithm for testing and constructing a tree realization from a distance matrix was described in [12].

For a (general) distance matrix, it is not too hard to show that an *optimal realization* (i.e., the realization having the minimum total length) exists [14]. It was shown in [14] that an optimal realization can have at most n^4 nodes if the number of terminal nodes is n. On the other hand, there are some examples of optimal realizations of a distance matrix on n terminals which have at least n^2 vertices. Therefore there is a finite (but exponential) algorithm to find an optimal realization for a given distance matrix. Various heuristics are discussed in many papers [18, 5, 21, 24, 25, 26]. However, solutions to this problem seem to be elusive, and, in fact, computing optimal realizations for distance matrices with a small number of terminal nodes is already quite complicated. Indeed, Althöfer [1] showed that the problem of finding optimal realizations of distance matrices with integral entries is NP-complete. We here will provide additional evidence pointing to the difficulties in approximating the optimal realization. In [7], it is shown that there are distance matrices D and D' on n terminal nodes satisfying $D(u,v) \geq D'(u,v)$, but where the optimal realization of D' has total edge length much larger than that of D. In fact, the ratio of the respective sums of edge lengths can be as large as a factor of n.

Because the realization problem seems hard even to approximate, we introduce a number of more robust variations and generalizations: For a given distance matrix D on a set S of terminal nodes, a graph G is said to be a *weak* realization of D if
(i) the node set of G contains S,
(ii) the distance between u and v in G is greater than or equal to $D(u,v)$ for all u,v in S.

Problem 2: For a given distance matrix D on a set S of terminal nodes, find a weak realization of D so that the total sum of all edge lengths of G is minimized.

It is not hard to show that if a distance matrix D dominates another matrix D' (i.e., $D(u,v) \geq D'(u,v)$), an optimal weak realization of D has total edge length no smaller than that of D'.

For a given distance matrix D on a set S of terminal nodes, a graph G is said to be a *rooted weak* realization of D if
(i) the node set of G contains S,
(ii) the distance between u and v in G is greater than or equal to $D(u,v)$ for all u,v in S,
(iii) there is a special node v^* in S, called the *root*, and the distance between u and v^* in G is equal to $D(u,v)$ for all u,v in S.

Problem 3: For a given distance matrix D on a set S of terminal nodes and a special node v^* of S, find a rooted weak realization of D so that the total sum of all edge lengths of G is *minimized.*

As it turns out, both Problems 2 and 3 are NP-complete [7]. Approximation algorithms are given for both Problems 2 and 3, which are within a small constant factor of the optimum.

We remark that the weak realization problems are similar to but different from the so-called *Steiner tree problem* which also has an extensive literature [19].

Euclidean Steiner tree problem: Given n points in the plane (or, in general, $\mathbb{R}^n$), find the shortest tree connecting the n points (where this tree may have additional points as vertices).

Graph Steiner tree problem: For a given graph and a subset S of nodes, find the shortest tree containing nodes in S.

Both of the above Steiner problems are NP-complete [16]. These Steiner problems are different from the distance realization problems since the host graph is unknown for the distance realization problems.

There is yet another related version of the realization problem which has more detailed inputs. It has also come up in the applications to Internet tomography:

Problem : Suppose we are given a set S of terminal nodes and a set E of edges. Each pair of terminal nodes, u and v are associated with a subset of edges which are in a shortest path joining u and v in the host graph. (In other words, an incidence matrix of edges and pairs of nodes is given.) The goal is to determine the host graph (i.e., to describe the adjacencies of its nodes and edges).

The above problem turns out to be a special case of the problem of determining the graphical construction of a matroid. This problem was solved by Tutte in his seminal papers [28, 29]. Indeed, there is a polynomial time algorithm of order $O(n^2m)$ for graphs with m edges and n terminal nodes [4, 15].

Some related problems have similar flavor as various topics in extremal graph theory. For example, for a family $\mathcal{F}$ of graphs, it is desirable to find a minimum *universal* graph which contains all graphs in $\mathcal{F}$ as subgraphs [8, 9]. Suppose the host graph G is unknown, but we are given a subset of vertices S, the distance matrix D_S restricted to S, and all rooted weak realization T_v for all nodes v in S. The problem of interest is to find a universal graph that contains all T_v and has the minimum total length.

References

[1] I. Althöfer, On optimal realizations of finite metric spaces by graphs, *Discrete Comput. Geom.* **3** (1988), 103-122.

[2] H. Bandelt, Recognition of tree metrics, *SIAM J. Discrete Math.* **3** (1990), no. 1, 1-6.

[3] K. Bharat and A. Broder, A technique for measuring the relative size and overlap of public web search engines, *Proceedings of the 7th International World Wide Web conference*, Brisbane, Australia, (1998), 379-388.

[4] R. E. Bixby and D. K. Wagner, An almost linear-time algorithm for graph realization, *Math. of Operations Research* **13** (1988), 99-123.

[5] F. T. Boesch, Properties of the distance matrix of a tree, *Quart. Appl. Math.* **26** (1968/69), 607-609.

[6] K. Calvert, M, Doar, and E. W. Zegura, Modeling Internet Topology, *IEEE Communications Magazine* (1997), **35** no. 6, 160-163.

[7] F. Chung, Garrett, R. Graham and D. Shallcross, Distance realization problems with applications to Internet tomography, preprint

[8] F. R. K. Chung and R. L. Graham, On universal graphs, *Annals of the New York Academy of Sciences*, No. 319 (1979), 136-140.

[9] F. R. K. Chung and R. L. Graham, On universal graphs for spanning trees, *Journal of London Math. Soc.* 27 (1983), 203-211.

[10] K. C. Claffy, T. E. Monk and D. McRobb, Internet tomography, *Nature, Web Matters* http:// helix.nature.com/webmatters/tomog.html, Jan. 1999.

[11] K. C. Claffy and T. Monk, What's next for Internet data analysis? *IEEE Special Issue on Communications in the 21st Century* **85** (1997), 1563-1571.

[12] J. C. Culberson and P. Rudnicki, A fast algorithm for constructing trees from distance matrices, *Inform. Process. Lett.* **30** (1989), no. 4, 215-220.

[13] E. Dahlhaus, Fast parallel recognition of ultrametrics and tree metrics, *SIAM J. Discrete Math.* **6** (1993), no. 4, 523-532.

[14] 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. in Math.* **53** (1984), no. 3, 321-402.

[15] S Fujishige, An efficient PQ-graph algorithm for solving the graph realization problem, *J. Comput. System Sci.* **21** (1980), 63-86.

[16] M. R. Garey and D. S. Johnson, *Computer and Intractability, A Guide to the Theory of NP-completeness*, W. H. Freeman and Co., San Francisco, 1979.

[17] M. Goemans and D. Williamson, Improved approximation algorithms for maximum cut and satisfiability problem using semidefinite programming, *J. Assoc. Mach.* **42** (1995), 1115-1145.

[18] S. L. Hakimi and S. S. Yau, Distance matrix of a graph and its realizability, *Quart. Appl. Math.* **22** (1965), 305-317.

[19] F. K. Hwang, D. S. Richard, and P. Winter, The Steiner problem, *Annals of Discrete Math.* **53** North Holland Pub. Co., Amsterdam, 1992.

[20] T. Munzner, Exploring large graphs in 3D hyperbolic space, *IEEE Computer Graphics and Applications* **18** (1998), 18-23.

[21] J. Nieminen, Realizing the distance matrix of a graph, *Elektron. Informationsverarbeit. Kybernetik* **12** (1976), no. 1-2, 29-31.

[22] A. N. Patrinos and S. L. Hakimi, The distance matrix of a graph and its realization, *Quart. Appl. Math.* **30** (1972/73), 255-269.

[23] J. M. S. Simoes Pereira, A note on the tree realizability of a distance matrix, *J. Combinatorial Theory* **6** (1969), 303-310.

[24] J. M. S. Simoes Pereira, An algorithm and its role in the study of optimal graph realizations of distance matrices, *Discrete Math.* **79** (1990), no. 3, 299-312.

[25] J. M. S. Simoes Pereira, An optimality criterion for graph embeddings of metrics, *SIAM J. Discrete Math.* **1** (1988), no. 2, 223-229.

[26] J. M. S. Simoes Pereira and C. M. Zamfirescu, Submatrices of non-tree-realizable distance matrices, *Linear Algebra Appl.* **44** (1982), 1-17.

[27] I. Peterson Diagnosing the Internet's ills, *Science News* , vol. 155, January 16, 1999, page 37.

[28] W. T. Tutte, A homotopy theorem for matroids, I,II, *Trans. Amer. Math. Soc.* **88** (1958), 144-174.

[29] W. T. Tutte, An algorithm for determining whether a given binary matroid is graphic, *Proc. Amer. Math. Soc.* **11** (1960), 905-917.