Discrete Optimization 10 (2013) 310–319

Contents lists available at ScienceDirect

Discrete Optimization

journal homepage: www.elsevier.com/locate/disopt

Searching for realizations of finite metric spaces in tight spans

Sven Herrmann [a], Vincent Moulton [a,*], Andreas Spillner [b]

[a] School of Computing Sciences, University of East Anglia, Norwich, NR4 7TJ, United Kingdom

[b] Department of Mathematics and Computer Science, University of Greifswald, 17487, Greifswald, Germany

ARTICLE INFO

Article history:
Received 1 May 2012
Received in revised form 3 August 2013
Accepted 3 August 2013
Available online 18 September 2013

Keywords:
Combinatorial optimization
Metric
Graph
Realization
Tight span

ABSTRACT

An important problem that commonly arises in areas such as internet traffic-flow analysis, phylogenetics and electrical circuit design, is to find a representation of any given metric D on a finite set by an edge-weighted graph, such that the total edge length of the graph is minimum over all such graphs. Such a graph is called an *optimal realization* and finding such realizations is known to be NP-hard. Recently Varone presented a heuristic greedy algorithm for computing optimal realizations. Here we present an alternative heuristic that exploits the relationship between realizations of the metric D and its so-called tight span T_D. The tight span T_D is a canonical polytopal complex that can be associated to D, and our approach explores parts of T_D for realizations in a way that is similar to the classical simplex algorithm. We also provide computational results illustrating the performance of our approach for different types of metrics, including l_1-distances and two-decomposable metrics for which it is provably possible to find optimal realizations in their tight spans.

1. Introduction

An important problem that commonly arises in areas such as internet traffic-flow analysis [1], phylogenetics [2] and electrical circuit design [3], is to realize any given metric D on some finite set X by an edge-weighted graph with X labeling its vertex set, often with the additional requirement that the total edge length of the graph is minimum. This can be useful, for example, for visualizing the metric, or for trying to better understand its structural properties. More formally this optimization problem can be stated as follows. A *realization* (G, ω, τ) of D is a connected graph $G = (V, E)$ with vertex set V and edge set E, together with an edge-weighting $\omega : E \to \mathbb{R}_{>0}$ and a labeling map $\tau : X \to V$ such that, for all $x, y \in X$, $D(x, y)$ equals $D_G(\tau(x), \tau(y))$, that is, the length of a shortest path from $\tau(x)$ to $\tau(y)$ in G (cf. Fig. 1(a) and (b)). The problem then is to find an *optimal* realization of D, that is, a realization of D that has minimum total edge length over all possible realizations of D.

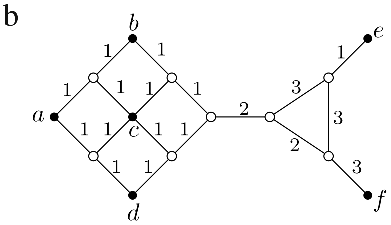
Early work on optimal realizations started with [3] (see also [4] for a comprehensive list of references), which focused mainly on special classes of metrics such as, for example, those that admit an optimal realization where the underlying graph is a tree (so-called *treelike* metrics). Subsequently it was found that *every* metric D on a finite set X has an optimal realization [5], although this need not be unique (cf. Fig. 1(c) and (d)). There even always exists an optimal realization of (X, D) with $O(|X|^4)$ vertices [6, p. 392], which implies that there is an exhaustive algorithm to search for an optimal realization. However, it was also shown that computing an optimal realization is NP-hard [7,8]. More recently, there has

* Corresponding author.
E-mail addresses: sherrmann@mathematik.tu-darmstadt.de (S. Herrmann), vincent.moulton@cmp.uea.ac.uk (V. Moulton), andreas.spillner@uni-greifswald.de (A. Spillner).

1572-5286/$ – see front matter Crown Copyright © 2013 Published by Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.disopt.2013.08.002

a

D	a	b	c	d	e	f
a	0	2	2	2	10	11
b	2	0	2	4	8	9
c	2	2	0	2	8	9
d	2	4	2	0	8	9
e	10	8	8	8	0	7
f	11	9	9	9	7	0



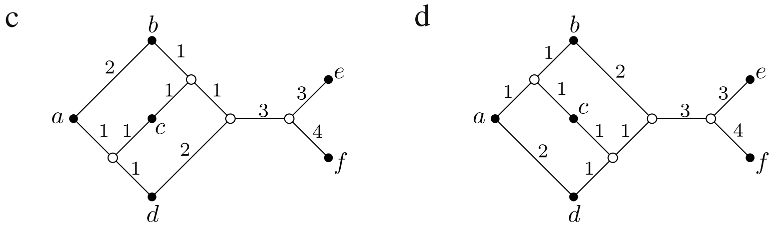


Fig. 1. (a) A metric D on $X = \{a, b, c, d, e, f\}$. (b) A realization of (X, D) that is not optimal. Vertices associated with an element of X are drawn as black dots, the remaining vertices are drawn as empty circles. (c), (d) Two optimal realizations of (X, D).

been renewed interest in computational aspects of this problem. For example, in [9,10] (see also [11]) a way to break up the problem of computing an optimal realization into subproblems using so-called *cut points* is presented, and in [4] a heuristic is presented for computing optimal realizations.

Here we present an alternative heuristic for systematically computing optimal realizations that exploits the relationship between optimal realizations of a metric D and its so-called *tight span* T_D [6,12]. In brief (see Section 2 for details), T_D is a polytopal complex (essentially a union of polytopes) that can be canonically associated to D which is itself a (non-finite) metric space and into which the metric D can be canonically embedded. Remarkably, in [6] it is shown that the 1-skeleton G_D of T_D (i.e., the edge-weighted graph formed essentially by taking all of the 0- and 1-dimensional faces of T_D) is always a realization of D. Moreover, Dress conjectured [6, (3.20)] that some optimal realization of D can always be obtained by removing some set of edges from G_D.

While Dress' conjecture is still open for metrics in general, recently it has been shown to hold for the class of so-called *two-decomposable* metrics [13, Theorem 1.2], a class which includes treelike metrics and l_1-distances between points in the plane (see Section 3 for more details). In particular, this and the aforementioned result in [6] suggest that it could be useful to consider G_D as a "search space" in which to look for some optimal realization of D (or at least some interesting realization of D which has relatively small total edge length).

Guided by this principle, given an arbitrary finite metric D, in Section 4 we propose a heuristic for computing a realization of D that is a subgraph of G_D. This heuristic explores parts of T_D in a way similar to the classical simplex algorithm [14]. Moreover, it does not explicitly compute G_D, whose vertex set can have cardinality that is exponential in $|X|$ (see e.g. [15] for some explicit bounds). We also show that the heuristic is guaranteed to find optimal realizations for some simple types of metrics.

Since, as mentioned above, the problem of finding optimal realizations is NP-hard, we assess the performance of our new heuristic using two strategies. First, we consider a special instance of the problem where we take metrics to be l_1-distances between points in the plane. In Section 5 we show that finding optimal realizations of such a metric D in G_D is equivalent to the so-called *minimum Manhattan network* problem (which was also recently shown to be NP-hard [16]). This allows us to compare the realizations computed by our heuristic with realizations computed using a mixed integer linear program (MIP) for the minimum Manhattan network problem presented in [17] (see also [18] for a comprehensive list of references on other approaches for solving this well-studied problem). Second, in Section 6 we describe a mixed integer program (MIP) for computing a *minimal subrealization* of a realization of some metric, that is, a subrealization with minimum total edge length. This allows us to obtain some impression of how close the realizations computed by our heuristic are to a minimal subrealization of G_D in case $|X|$ is not too large. Moreover, in case the metric is two-decomposable, a minimal subrealization of G_D is (by the aforementioned result in [13]) an optimal realization and so we can compare the realizations computed by our new heuristic with optimal ones for this special class of metrics.

Based on these considerations, in Section 7 we present simulations for l_1-distances, two-decomposable metrics and random metrics to assess the performance of our heuristic. An implementation of this heuristic is freely available for download at www.uea.ac.uk/cmp/research/cmpbio/CoMRiT/. This includes the algorithm for efficiently computing cut points as described in [11] and auxiliary programs that can be used to generate the MIP description for the minimum Manhattan network problem, as well as for the problem of computing a minimal subrealization so that they can be solved using existing MIP solvers (we used the solver that is part of the GNU linear programming kit (www.gnu.org/software/glpk/) in our experiments). We conclude the paper with a brief discussion of some possible future directions in Section 8.