Published in *Journal of classification*, 2008, vol. 25, no. 2, p. 159-175
which should be cited to refer to this work

Journal of Classification manuscript No.
(will be inserted by the editor)

The Metric Cutpoint Partition Problem

Alain Hertz[1], **Sacha Varone**[2⋆]

[1] Département de mathématiques et de génie industriel, École Polytechnique, Montréal, Canada, e-mail: alain.hertz@gerad.ca

[2] Haute École de Gestion, Geneva, Switzerland, e-mail: sacha.varone@hesge.ch

The date of receipt and acceptance will be inserted by the editor

Abstract Let $G = (V, E, w)$ be a graph with vertex and edge sets V and E, respectively, and $w : E \rightarrow I\!R^+$ a function which assigns a positive weight or length to each edge of G. G is called a realization of a finite metric space (M, d), with $M = \{1, ..., n\}$ if and only if $\{1, ..., n\} \subseteq V$ and $d(i, j)$ is equal to the length of the shortest chain linking i and j in G $\forall i, j = 1, ..., n$. A realization G of (M, d), is said optimal if the sum of its weights is minimal among all the realizations of (M, d). A cutpoint in a graph G is a vertex whose removal strictly increases the number of connected components of G. The Metric Cutpoint Partition Problem is to determine if a finite metric space (M, d) has an optimal realization containing a cutpoint. We prove in this paper that this problem is polynomially solvable. We also describe

⋆ Supported by grant PA002-104974/2 from the Swiss National Science Foundation.

an algorithm that constructs an optimal realization of (M, d) from optimal realizations of subspaces that do not contain any cutpoint.

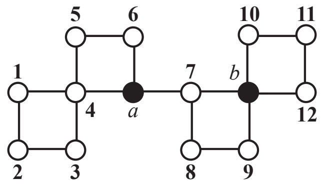
1 Introduction

A *metric space* is a couple (M, d) such that M is a set and d is a function defined on $M \times M$ such that $d(x, y) = d(y, x)$ and is a strictly positive finite number $\forall x \neq y$, $d(x, x) = 0\ \forall x$, and $d(x, z) \leq d(x, y)+d(y, z)\ \forall x, y, z$. Moreover, (M, d) is a finite metric space if M has a finite number of elements.

Let $G = (V, E, w)$ be a graph, with vertex and edge sets V and E, respectively, and $w : E \rightarrow I\!R^+$ a function which assigns a strictly positive weight or length to each edge of G. Furthermore, let $d^G(i, j)$ denote the length of a shortest chain in G linking vertices i and j. We say that G is a realization of a finite metric space (M, d), with $M = \{1, ..., n\}$ if and only if $\{1, ..., n\} \subseteq V$ and $d^G(i, j) = d(i, j)\ \forall i, j = 1, ..., n$. The elements in $V \backslash M$ are called *auxiliary vertices*. Without loss of generality, we can assume that every auxiliary vertex has at least three adjacent vertices. 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 weights of G is minimal among all realizations of (M, d). Clearly, every optimal realization is minimal. For illustration, a metric space together with an optimal realization G are shown in Figure 1. All edges of the graph have length one, and the black points a, b are two auxiliary vertices while the white ones are the elements of M.

	1	2	3	4	5	6	7	8	9	10	11	12
1	0	1	2	1	2	3	3	4	5	5	6	5
2		0	1	2	3	4	4	5	6	6	7	6
3			0	1	2	3	3	4	5	5	6	5
4				0	1	2	2	3	4	4	5	4
5					0	1	3	4	5	5	6	5
6						0	2	3	4	4	5	4
7							0	1	2	2	3	2
8								0	1	3	4	3
9									0	2	3	2
10										0	1	2
11											0	1
12												0

Upper diagonal of the distance matrix of a metric space (M,d) with $M=\{1,...,12\}$



Optimal realization of (M,d)

Figure 1. A metric space with an optimal realization

The embedding of finite metric spaces in graphs has applications in varied fields such as computational biology [13, 15] (e.g., constructing phylogenetic trees from genetic distances among living species), electrical networks [9], coding techniques [8], psychology [5], internet tomography [4], and compression softwares [14].

The problem of finding optimal realizations of metric spaces was first proposed by Hakimi and Yau [9] in 1964 who also gave a polynomial algorithm for the special case where the metric space has a realization as a tree. While every finite metric space has an optimal realization [12, 11], finding such realizations is an NP-hard problem [16]. Approximation algorithms for the embedding of metric spaces in graphs have also been a subject of extensive mathematical studies. Recent developments and references to earlier works on this subject can be found in [1, 3].

Optimal realizations can be constructed using building blocks. More precisely, for a graph G, we recall that a *cutpoint*, respectively a *bridge*, is a vertex, respectively an *edge*, whose removal strictly increases the number of connected component of G; a *block* is a maximal two-connected subgraph or a bridge in G. Imrich *et al.* [11] have proved the following theorem.

Theorem 1 *[11] Let G be a minimal realization of a finite metric space (M, d), let $G_1, \cdots, G_k$ be the blocks of G, and let M_r be the union of the points of M in G_r and the cutpoints of G in G_r. If every G_r is an optimal realization of the metric space induced by G on M_r, then G is also optimal.*

For example an optimal realization of the metric space of Figure 1 can be obtained by putting together optimal realizations of the metric spaces induced on $\{1, 2, 3, 4\}$, $\{4, 5, 6, a\}$, $\{a, 7\}$, $\{7, 8, 9, b\}$, and $\{10, 11, 12, b\}$.

We call *Metric Cutpoint Partition Problem* (MCPP for short) the problem of determining whether a given finite metric space (M, d) has an optimal realization containing a cutpoint. For example, on the basis of the distance matrix of Figure 1 (and without any knowledge of the optimal realization), we would like to be able to state that there is an optimal realization containing the cutpoint $4, 7, a$ or b. Similarly, the *Metric Bridge Partition Problem* (MBPP for short) is to recognize metric spaces (M, d) to which there exists an optimal realization containing a bridge.

If M contains only two elements, then the unique optimal realization G of (M, d) is a graph with two vertices linked by an edge. Obviously, such a graph G has a bridge and no cutpoint. If M has more than two elements, then at least one endpoint of every bridge is a cutpoint. Hence, the MCPP is more general than the MBPP.

We have shown in [10] that the MBPP can be solved in polynomial time. More precisely, we have presented an algorithm with running time $O(|M|^6)$ that decides whether a given metric space (M, d) has an optimal realization containing a bridge. We prove in this paper that the MCPP is also polynomially solvable.

2 Definitions and Known Results

It is well-known that the unique optimal realization of a metric space on three points i, j, k is a tree T. The *hub* of i, j, k, denoted h_{ijk}, is the point in T such that:

$$d^T(h_{ijk}, i) = \tfrac{1}{2}(d(i,j) + d(i,k) - d(j,k)),$$
$$d^T(h_{ijk}, j) = \tfrac{1}{2}(d(j,i) + d(j,k) - d(i,k)),$$
$$d^T(h_{ijk}, k) = \tfrac{1}{2}(d(k,i) + d(k,j) - d(i,j)).$$

Assume that the distance $d(i,j)$ is larger than or equal to $d(i,k)$ and $d(j,k)$. If $d(i,j) < d(i,k) + d(j,k)$, then T has three leaves i, j and k, and one auxiliary vertex corresponding to the hub h_{ijk}, else T is a chain linking i and j that traverses $k = h_{ijk}$ (see Figure 2).

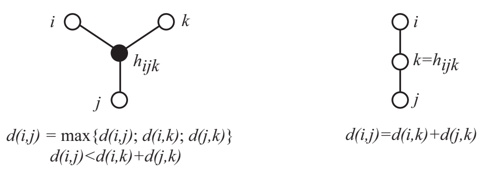


Figure 2. Optimal realizations of three points

Let $s_{ijk\ell}$ denote the sum $d(i,j) + d(k,\ell)$. It is also well-known that the optimal realization of a metric space on four points i, j, k, ℓ is unique and is a tree if and only if two of the sums $s_{ijk\ell}, s_{ikj\ell}, s_{i\ell jk}$ are equal and not smaller than the third [2]. The five possible configurations with $s_{ijk\ell} \leq s_{ikj\ell} = s_{i\ell jk}$ are represented in Figure 3.

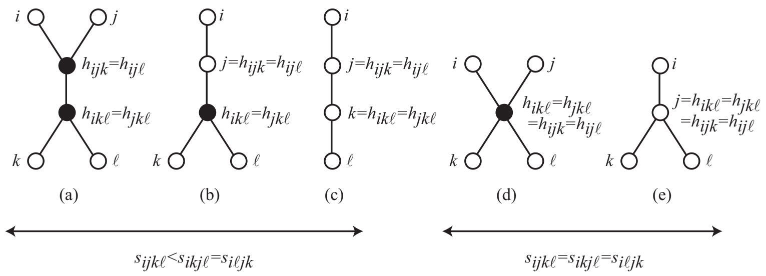


Figure 3. Optimal realizations of four points

Definition 1 *Consider a finite metric space* (M, d), *a partition of* M *into two non-empty subsets* K, L *and a mapping* $f : M \to I\!R^+$. *The triplet* (K, L, f) *is said nice if*

- $d(x, y) \leq f(x) + f(y)$ *for all* x, y *in* M, *equality holding whenever* $x \in K$ *and* $y \in L$, *and*
- $f(x) > 0$ *at least once in* K *and once in* L.

The above definition is motivated by the following result proved in [12] and [11]

Theorem 2 *[12, 11] Let* (M, d) *be a finite metric space to which there exists a nice triplet* (K, L, f). *Then every optimal realization* G *of* (M, d) *has a cutpoint* c *or a bridge with a point* c *on it such that all chains linking* K *with* L *go through* c, *and* $d^G(x, c) = f(x)$ $\forall x \in M$.

We have proved in [10] that the MBPP is polynomially solvable. In particular, we have proved the following theorem that provides a sufficient condition for the existence of a bridge in optimal realizations of a metric space.

Theorem 3 *Let (M, d) be a finite metric space to which there exists a partition of M into two non-empty subsets K, L with $|K| > 1$ and $|L| > 1$. If $s_{ijk\ell} < s_{ikj\ell} = s_{i\ell jk}$ $\forall i, j \in K$ and $k, \ell \in L$, then every optimal realization of (M, d) has a bridge.*

Also, we have designed in [10] a polynomial algorithm that produces one of the two following outputs for every given metric space (M, d):

- the first possible output is a message indicating that no optimal realization of (M, d) has a bridge,
- the second possible output is of the form $(K, d_K), (L, d_L), u \in K, v \in L, \ell$ with the following meaning : an optimal realization of (M, d) can be obtained by constructing optimal realizations of (K, d_K) and (L, d_L), and by linking u and v with an edge of length ℓ.

To show that the MCPP is also polynomially solvable, we can therefore restrict our attention to metric spaces (M, d) that have no optimal realization containing a bridge. Such metric spaces are said *bridgeless*.

The following definition associates a partition of M with each cutpoint in an optimal realization of a finite metric space (M, d).

Definition 2 *Let G be an optimal realization of a finite metric space (M, d) with a cutpoint u, and let H be the graph obtained from G by removing all edges incident to u (while keeping vertex u in H). Let $G_1, \cdots, G_k$ denote the connected components of H that contain at least one element of M, and let M_r be the union of the elements of M in G_r. We say that $\{M_1, \cdots, M_k\}$ is a u-partition of M.*

For example, the u-partition associated with $u = 4$ in Figure 1 is $\{\{1, 2, 3\}$, $\{4\}$, $\{5, \cdots, 12\}\}$ while it is equal to $\{\{1, \cdots, 6\}$, $\{7, \cdots, 12\}\}$ for $u = a$.

3 New Results

Lemma 1 *Let e be any edge in a minimal realization G of a finite metric space (M, d). Then there are two vertices a and b in M such that all shortest chains linking a and b traverse e.*

Proof Assume that for every two vertices a and b in M there exists a shortest chain linking a and b that does not traverse e. Then the graph obtained from G by removing e is still a realization of (M, d), which contradicts the minimality of G.

Lemma 2 *Let (M, d) be a bridgeless finite metric space to which there exists an optimal realization G with a cutpoint u, and let e be any edge in G that does not contain u as endpoint Then there is a chain linking two vertices a and b of M that traverses e, has a total length strictly smaller than $d^G(a, u)+ d^G(b, u)$, and has no intermediate vertex in M.*

Proof Since optimal realizations are minimal, we know from Lemma 1 that there are two vertices a and b in M such that all shortest chains linking a and b traverse e. Among all such chains, let us choose one that minimizes $d(a,b)$. It follows that no shortest chain linking a and b contains an intermediate vertex $c \in M$, else the pair (a,c) or (c,b) contradicts the minimality of (a,b).

If $d^G(a,b) < d^G(a,u)+d^G(b,u)$, then we are done. So let us assume that $d^G(a,b) = d^G(a,u) + d^G(b,u)$. Without loss of generality, we can assume that e belongs to all shortest chains linking a and u. So, consider such a shortest chain $(a = v_0, v_1, \cdots, v_k = u)$ with $e = (v_t, v_{t+1})$ for some $t < k-1$.

Since v_{k-1} is an auxiliary vertex, there is a vertex w adjacent to v_{k-1} with $w \neq v_{k-2}, u$. Now, since (v_{k-1}, w) is an edge in E that does not contain u as endpoint, we know from Lemma 1 that there are two vertices c and d in M such that all shortest chains linking c and d traverse (v_{k-1}, w) and have no intermediate vertex in M. Consider any such chain $(c = w_0, \cdots, w_r = w, w_{r+1} = v_{k-1}, w_{r+2}, \cdots, w_s = d)$. Since $d(v_{k-1}, u) > 0$, we have

$$2d^G(a, v_{k-1}) + \sum_{i=0}^{s-1} d^G(w_i, w_{i+1}) < 2d^G(a,u) + d^G(c,u) + d^G(d,u).$$

Hence, we are in at least one of the following two cases :

- $d^G(a, v_{k-1}) + \sum_{i=0}^{r} d^G(w_i, w_{i+1}) < d^G(a,u) + d^G(c,u)$: this means that the chain $(a = v_0, \cdots, v_{k-1} = w_{r+1}, w_r, \cdots, w_0 = c)$ traverses e, has no intermediate vertex in M, and its total length is strictly smaller than $d^G(a,u) + d^G(c,u)$,

– $d^G(a, v_{k-1}) + \sum_{i=r+1}^{s-1} d^G(w_i, w_{i+1}) < d^G(a, u) + d^G(d, u)$: this means that the chain $(a = v_0, \cdots, v_{k-1} = w_{r+1}, w_{r+2}, \cdots, w_s = d)$ traverses e, has no intermediate vertex in M, and its total length is strictly smaller than $< d^G(a, u) + d^G(d, u)$. □

Before proving the next theorem, we need to define two additional concepts.

Definition 3 *Let (x, y) and (z, t) be two pairs of distinct elements in M such that $s_{xyzt} = s_{xzyt} = s_{xtyz}$. The function $f_{(x,y)(z,t)} : M \rightarrow I\!R^+$ and the graph $H_{(x,y)(z,t)}$ are defined as follows :*

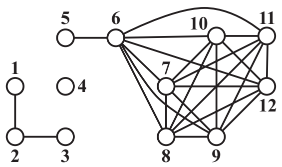
- *$f_{(x,y)(z,t)}(v)$ is the maximum between the distance from v to the hub h_{xyv} and the distance from v to the hub h_{ztv}. Formally, $f_{(x,y)(z,t)}(v)$= $max\{\frac{1}{2}(d(x, v) + d(y, v) - d(x, y)), \frac{1}{2}(d(z, v) + d(t, v) - d(z, t))\}$.*
- *The vertex set of $H_{(x,y)(z,t)}$ is M, and two vertices v and w are linked by an edge in $H_{(x,y)(z,t)}$ if and only if $f_{(x,y)(z,t)}(v) + f_{(x,y)(z,t)}(w) > d(v, w)$.*

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
1	0	1	2	1	2	3	3	4	5	5	6	5
2		0	1	2	3	4	4	5	6	6	7	6
3			0	1	2	3	3	4	5	5	6	5
4				0	1	2	2	3	4	4	5	4
5					0	1	3	4	5	5	6	5
6						0	2	3	4	4	5	4
7							0	1	2	2	3	2
8								0	1	3	4	3
9									0	2	3	2
10										0	1	2
11											0	1
12												0

Metric space of Figure 1 in which $s_{1357}=s_{1537}=s_{1735}=5$

v	1	2	3	4	5	6	7	8	9	10	11	12
$1/2(d(1,v)+d(3,v)-d(1,3))$	0	0	0	0	1	2	2	3	4	4	5	4
$1/2(d(5,v)+d(7,v)-d(5,7))$	1	2	1	0	0	0	0	1	2	2	3	2
$f_{(1,3)(5,7)}(v)$	1	2	1	0	1	2	2	3	4	4	5	4

Computation of $f_{(1,3)(5,7)}(v)$



Graph $H_{(1,3)(5,7)}$

Figure 4. Illustration of $f_{(x,y)(z,t)}$ and $H_{(x,y)(z,t)}$

The above concepts are illustrated in Figure 4 for the two pairs (1,3) and (5,7) of elements chosen in the metric space of Figure 1.

Theorem 4 *Let (M, d) be a bridgeless finite metric space to which there exists an optimal realization G with a cutpoint u, and let (x, y) and (z, t) be two pairs of vertices such that $f_{(x,y)(z,t)}(v) = d^G(v, u)\ \forall v \in M$. Then the blocks of the u-partition of M are the vertex sets of the connected components of $H_{(x,y)(z,t)}$.*

Proof Consider any two elements a and b in M. If a and b belong to two different subsets of the u-partition, then all chains linking a and b go through u, which means that a and b are not adjacent in $H_{(x,y)(z,t)}$ since

$$d(a, b) = d^G(a, u) + d^G(b, u) = f_{(x,y)(z,t)}(a) + f_{(x,y)(z,t)}(b).$$

We now prove that if a and b belong to the same subset of the u-partition, then a and b belong to the same connected component of $H_{(x,y)(z,t)}$. So consider any chain $C = (a = v_0, v_1, \cdots, v_k = b)$ linking a and b in G that does not go through u. By Lemma 2, we can associate to each edge (v_i, v_{i+1}) $(i = 0, \ldots, k-1)$ on C two vertices c_i and d_i in M and a chain C_i that traverses (v_i, v_{i+1}), has total length strictly smaller than $d^G(c_i, u) + d^G(d_i, u)$, and has no intermediate vertex in M. Notice that $c_0 = v_0 = a$ and $d_{k-1} = v_k = b$. Notice also that the graph $H_{(x,y)(z,t)}$ contains all edges (c_i, d_i) since $d^G(c_i, d_i)$ is at most equal to the total length of C_i, which is strictly smaller than $d^G(c_i, u) + d^G(d_i, u) = f_{(x,y)(z,t)}(c_i) + f_{(x,y)(z,t)}(d_i)$.

If $k = 1$ then a and b are adjacent in $H_{(x,y)(z,t)}$. Else, let us denote $C_i = (c_i = w_1^i, \cdots, w_{r_i}^i = d_i)$ for all $i = 1, \cdots, k-2$, and let $p(i)$ be the index such that $(w^i_{p(i)}, w^i_{p(i)+1}) = (v_i, v_{i+1})$. We have

$$\begin{aligned} &d^G(c_i,u) + d^G(d_i,u) + d^G(c_{i+1},u) + d^G(d_{i+1},u) \\ &> \textstyle\sum_{j=1}^{r_i-1} d^G(w^i_j, w^i_{j+1}) + \sum_{j=1}^{r_{i+1}-1} d^G(w^{i+1}_j, w^{i+1}_{j+1}). \end{aligned}$$

Hence,

$$\begin{aligned} &d^G(c_i,u) + d^G(d_{i+1},u) \\ &> \textstyle\sum_{j=1}^{p(i)} d^G(w^i_j, w^i_{j+1}) + \sum_{j=p(i+1)}^{r_{i+1}-1} d^G(w^{i+1}_j, w^{i+1}_{j+1}) \\ &\geq d^G(c_i, v_{i+1}) + d^G(v_{i+1}, d_{i+1}) \geq d^G(c_i, d_{i+1}) \end{aligned}$$

or/and

$$\begin{aligned} &d^G(d_i,u) + d^G(c_{i+1},u) \\ &> \textstyle\sum_{j=p(i)+1}^{r_i-1} d^G(w^i_j, w^i_{j+1}) + \sum_{j=1}^{p(i+1)-1} d^G(w^{i+1}_j, w^{i+1}_{j+1}) \\ &\geq d^G(v_{i+1}, d_i) + d^G(c_{i+1}, v_{i+1}) \geq d^G(c_{i+1}, d_i). \end{aligned}$$

In other words, the graph $H_{(x,y)(z,t)}$ contains the edge (c_i, d_{i+1}) or/and (c_{i+1}, d_i). It follows that all c_i's and d_i's belong to the same connected component of $H_{(x,y)(z,t)}$. This is in particular true for $a = c_0$ and $b = d_{k-1}$.

□

The next theorem gives necessary conditions for the existence of a cutpoint in at least one optimal realization of a bridgeless finite metric space.

Theorem 5 *Let (M,d) be a bridgeless finite metric space to which there exists an optimal realization G with a cutpoint u, and let $\{M_1, \cdots, M_k\}$ be a u-partition of M. Then*

(1) $s_{abcd} \leq s_{acbd} = s_{adbc}$ $\forall a, b \in M_r,\ c, d \notin M_r,\ r = 1, \cdots, k$

(2) there are four elements $x, y \in M_r$ and $z, t \in M_s$ $(r \neq s)$ such that

– $s_{xyzt} = s_{xzyt} = s_{xtyz}$

– $M_1, \cdots, M_k$ *are the vertex sets of the connected components of* $H_{(x,y)(z,t)}$

Proof Observe first that each M_r different from $\{u\}$ contains at least two elements. Indeed, if $M_r = \{a\}$ with $a \neq u$, then (a,u) is a bridge in G, a contradiction. So, consider any M_r with at least two elements, and define $K = M_r$ and $L = \cup_{k\neq r} M_k$. We have $|K| > 1$ and $|L| > 1$. Now choose any four elements $a, b \in K$ and $c, d \in L$. Since u is a cutpoint, we have

$$\begin{aligned} s_{acbd} = s_{adbc} &= d^G(a,u) + d^G(b,u) + d^G(c,u) + d^G(d,u) \\ &\geq d(a,b) + d(c,d) = s_{abcd} \end{aligned}$$

Since (M,d) is bridgeless, we know from Theorem 3 that there are four elements $x, y \in K$ and $x', y' \in L$ such that $s_{xyx'y'} = s_{xx'yy'} = s_{xy'yx'}$, which means that $d^G(x,y) = d^G(x,u) + d^G(y,u)$. Similarly, for every $M_s \subseteq L$ with at least two elements, there exist $z, t \in M_s$ such that $d^G(z,t) = d^G(z,u) + d^G(t,u)$. Hence,

$$s_{xzyt} = s_{xtyz} = d^G(x,u) + d^G(y,u) + d^G(z,u) + d^G(t,u) = s_{xyzt}.$$

Consider now any element $v \notin M_r$. Since the chain linking v and x goes through u, we have $d(x,v) = d^G(x,u) + d^G(v,u)$. By permuting the roles of x and y, we also have $d(y,v) = d^G(y,u) + d^G(v,u)$. Now, since $d(x,y) = d^G(x,u) + d^G(y,u)$ and $d(z,t) = d^G(z,u) + d^G(t,u)$, we have

$$\begin{aligned} &\tfrac{1}{2}(d(x,v) + d(y,v) - d(x,y)) \\ &= \tfrac{1}{2}(d^G(x,u) + d^G(y,u) + 2d^G(u,v) - d^G(x,u) - d^G(y,u)) \\ &= d^G(v,u) = \tfrac{1}{2}(d^G(z,u) + d^G(t,u) + 2d^G(u,v) - d^G(z,u) - d^G(t,u)) \\ &\geq \tfrac{1}{2}(d(z,v) + d(t,v) - d(z,t)). \end{aligned}$$

This means that $f_{(x,y,z,t)}(v) = d^G(v,u)\ \forall v \notin M_r$. By symmetry, the same holds for all $v \notin M_s$, which proves that $f_{(x,y,z,t)}(v) = d^G(v,u)\ \forall v \in M$. We therefore conclude from Theorem 4 that $M_1, \cdots, M_k$ are the vertex sets of the connected components of $H_{(x,y)(z,t)}$. $\square$

We finally give a sufficient condition for the existence of a cutpoint in at least one optimal realization of a metric space.

Theorem 6 *Let (M,d) be a bridgeless finite metric space, and let $M_1, \cdots, M_k$ be a partition of M into k non-empty subsets. Assume the existence of four distinct elements $x, y \in M_r$ and $z, t \in M_s$ $(r \neq s)$ such that*

(a) $s_{xyzt} = s_{xzyt} = s_{xtyz}$,

(b) $s_{abcd} \leq s_{acbd} = s_{adbc}$ for all a, b, c, d such that $a, b \in M_q$ and $c, d \notin M_q$ for some $q \in \{1, \cdots, k\}$, and $|\{a, b, c, d\} \cap \{x, y, z, t\}| \geq 2$.

Then every optimal realization G of (M,d) has a cutpoint u with $d^G(v,u) = f_{(x,y)(z,t)}(v)$ $\forall v \in M$.

Proof It is sufficient to prove that $(M_r, \cup_{j \neq r} M_j, f_{(x,y)(z,t)})$ is a nice triplet. Indeed, since (M,d) is bridgeless, we know from Theorem 2 that this will prove that each realization G of (M,d) has a cutpoint u such that all chains linking M_r with $\cup_{j \neq r} M_j$ traverse u, and $d^G(v,u) = f_{(x,y)(z,t)}(v)$ $\forall v \in M$.

So let T be an optimal realization of the metric space induced by x, y, z, t. We know from (a) that T is a tree in which all hubs $h_{xyz}, h_{xyt}, h_{xzt}, h_{yzt}$ coincide at one point which we call h.

Consider any element $v \notin M_r$. If $v \neq z$ then let U denote the optimal realization of the metric space induced on x, y, z and v. We know from (b) that U is a tree with hubs $h_{xyz} = h_{xyv} = h$ and $h_{xzv} = h_{yzv}$ (which are possibly all equal). We have

$$\begin{aligned}
& d(z,v) - d^T(z,h) \\
&= d^U(z,h_{xzv}) + d^U(h_{xzv},v) - d^U(z,h_{xzv}) - d^U(h_{xzv},h) \\
&= d^U(h,v) - 2d^U(h,h_{xzv}) \\
&\leq d^U(h,v) = d(x,v) - d^T(x,h) = \tfrac{1}{2}(d(x,v) + d(y,v) - d(x,y)).
\end{aligned}$$

If $v = z$, then $d(z,z) - d^T(z,h) \leq d(x,z) - d^T(x,h) = \frac{1}{2}(d(x,z) + d(y,z) - d(x,y))$. Hence, we have

$$d(z,v) - d^T(z,h) \leq d(x,v) - d^T(x,h) = \tfrac{1}{2}(d(x,v) + d(y,v) - d(x,y)) \forall v \notin M_r,$$

and by permuting the roles of z and t, we also have

$$d(t,v) - d^T(t,h) \leq d(x,v) - d^T(x,h) = \tfrac{1}{2}(d(x,v) + d(y,v) - d(x,y)) \forall v \notin M_r.$$

Since $d(z,t) = d^T(z,h) + d^T(t,h)$, we therefore have

$$\begin{aligned}
& \tfrac{1}{2}(d(z,v) + d(t,v) - d(z,t)) \\
&= \tfrac{1}{2}(d(z,v) - d^T(z,h) + d(t,v) - d^T(t,h)) \\
&\leq d(x,v) - d^T(x,h) = \tfrac{1}{2}(d(x,v) + d(y,v) - d(x,y)),
\end{aligned}$$

which means that $f_{(x,y)(z,t)}(v) = d(x,v) - d^T(x,h)$ for all $v \notin M_r$. By permuting the roles of x, y with those of z, t, we also have $f_{(x,y)(z,t)}(v) = d(z,v) - d^T(z,h)$ for all $v \notin M_s$. So, consider any two elements $v \notin M_r$ and $w \notin M_s$. We have

$$\begin{aligned}
f_{(x,y)(z,t)}(v) + f_{(x,y)(z,t)}(w) &= d(x,v) - d^T(x,h) + d(z,w) - d^T(z,h) \\
&= d(x,v) + d(z,w) - d(x,z).
\end{aligned}$$

It follows that if $v = z$ or/and $w = x$ then $f_{(x,y)(z,t)}(v) + f_{(x,y)(z,t)}(w) = d(v,w)$. Otherwise, let U denote the optimal realization of the metric space induced by x, z, v and w. We know from (b) that U is a tree with hubs $h_{xwz} = h_{xwv}$ and $h_{xzv} = h_{wzv}$ (which are possibly all equal). Since $d(x, v) + d(z, w) - d(x, z) = d^U(v, w) = d(v, w)$. We conclude that $f_{(x,y)(z,t)}(v) + f_{(x,y)(z,t)}(w) = d(v, w)$ for all $v \notin M_r$ and $w \in M_s$.

Consider now two elements $v, w \notin M_r$, and let U denote the optimal realization of the metric space induced by x, y, v and w. Again, we know from (b) that U is a tree with hubs $h_{xyv} = h_{xyw} = h$ and $h_{xvw} = h_{yvw}$ (which are possibly all equal), and we have

$$\begin{aligned} f_{(x,y)(z,t)}(v) + f_{(x,y)(z,t)}(w) &= d(x, v) + d(x, w) - 2d^T(x, h) \\ &= d^U(x, v) + d^U(x, w) - 2d^U(x, h_{xyv}) \\ &= d^U(v, w) + 2d^U(h_{xyv}, h_{xvw}) \\ &\geq d^U(v, w) = d(v, w). \end{aligned}$$

By symmetry, we also know that $f_{(x,y)(z,t)}(v) + f_{(x,y)(z,t)}(w) \geq d(v, w)$ for all $v, w \notin M_s$. Hence this is true for all $v, w \in M_r$.

Since $0 < d(x, y) \leq f_{(x,y)(z,t)}(x) + f_{(x,y)(z,t)}(y)$ we know that $f_{(x,y)(z,t)}(x)$ or/and $f_{(x,y,z,t)}(y)$ is strictly positive. Similarly, $f_{(x,y)(z,t)}(z)$ or/and $f_{(x,y)(z,t)}(t)$ is strictly positive. We can therefore conclude that $(M_r, \cup_{j \neq r} M_j, f_{(x,y,z,t)})$ is a nice triplet. $\square$

4 Algorithms

The following algorithm determines if a given finite bridgeless metric space (M, d) has an optimal realization containing a cutpoint u. Moreover, if such a realization G exists, then the algorithm also provides a u-partition of

M as well as two pairs (x,y) and (z,t) of elements such that $d^G(v,u) = f_{(x,y)(z,t)}(v)\ \forall v \in M$.

Algorithm 1 *MCPP*

Require: A finite bridgeless metric space (M,d);

Ensure: Either a message indicating that no optimal realization of (M,d) has a cutpoint, or two pairs (x,y) and (z,t) of elements and a u-partition $\{M_1,\cdots,M_k\}$ of M ;

for all couples of pairs (x,y) and (z,t) such that $s_{xyzt} = s_{xzyt} = s_{xtyz}$ **do**

set $M_1,\cdots,M_k$ equal to the vertex sets of the connected components of the graph $H_{(x,y)(z,t)}$

if there exist $r \neq s$ with $x,y \in M_r$ and $z,t \in M_s$ **then**

if $s_{abcd} \leq s_{acbd} = s_{adbc}$ for all a,b,c,d such that $a,b \in M_q$ and $c,d \notin M_q$ for some $q \in \{1,\cdots,k\}$, and $|\{a,b,c,d\} \cap \{x,y,z,t\}| \geq 2$ **then**

STOP: return $(x,y),(z,t)$ and $\{M_1,\cdots,M_k\}$.

end if

end if

end for

STOP : return a message indicating that no optimal realization of (M,d) has a cutpoint.

Theorem 7 *The MCPP algorithm works correctly and is polynomial.*

Proof Correctness of the algorithm follows from the results of the previous section. More precisely, if the algorithm stops with two pairs (x,y),(z,t) of elements and a partition $\{M_1,\cdots,M_k\}$ of M, then properties (a) and (b) of Theorem 6 are satisfied, and we conclude that every optimal realization G

of (M,d) has a cutpoint u with $d^G(v,u) = f_{(x,y)(z,t)}(v)\ \forall v \in M$. Moreover, we know from Theorem 4 that $\{M_1,\cdots,M_k\}$ is a u-partition of M since the M_i's correspond to the vertex sets of the connected components of $H_{(x,y)(z,t)}$.

Now, if (M,d) has an optimal realization G containing a cutpoint u, then we know from Theorem 5 that such a situation is detected. Indeed, we enumerate all couples of pairs $(x,y),(z,t)$ such that $s_{xyzt} = s_{xzyt} = s_{xtyz}$, and for each such couple, we build the partition $\{M_1,\cdots,M_k\}$ corresponding to the vertex sets of the connected components of $H_{(x,y)(z,t)}$. Moreover, we ask for the existence of two indices r and s such that $x,y \in M_r$ and $z,t \in M_s$, and we require that $s_{abcd} \leq s_{acbd} = s_{adbc}$ for all a,b,c,d such that $a,b \in M_q$ and $c,d \notin M_q$ for some $q \in \{1,\cdots,k\}$, and $|\{a,b,c,d\} \cap \{x,y,z,t\}| \geq 2$. This is less restrictive than the necessary conditions of Theorem 5.

Finally, the algorithm is polynomial since it can easily be implemented with a time complexity in $O(|M|^6)$ $\square$.

According to Theorem 1, one can build an optimal realization of (M,d) from an output (x,y), (z,t), and $\{M_1,\cdots,M_k\}$ of the $MCPP$ algorithm as follows:

- If the cutpoint u belongs to M (i.e, one of the blocks of the partition is a singleton), then consider the index r such that $M_r = \{u\}$, and construct for each $q \neq r$ an optimal realization G_q of the metric space $(M_q \cup \{u\}, d|_{M_q\cup\{u\}})$.

– If the cutpoint is an auxiliary vertex, then construct for each $q = 1, \cdots, k$ an optimal realization G_q of the metric space $(M'_q, d_{M'_q})$, where $M'_q = M_q \cup \{u\}$, $d_{M'_q}(v, w) = d(v, w)$ for all $v, w \in M_q$ and $d_{M'_q}(v, u) = f_{(x,y)(z,t)}(v)$ for all $v \in M_q$.

An optimal realization of (M, d) can then simply be obtained by gluing all G_i's at their unique common vertex u.

Assume the existence of an algorithm, called *Bridge*, which either indicates that the given metric space (M, d) is bridgeless, or provides an output of the form $(K, d_K), (L, d_L), a \in K, b \in L, \ell$ with the following meaning : an optimal realization of (M, d) can be obtained by constructing optimal realizations of (K, d_K) and (L, d_L), and by linking a and b with an edge of length ℓ. Algorithm *Bridge* can be implemented in polynomial time, as shown in [10]. Assume also the existence of an algorithm, called *NoCutpoint* that constructs an optimal realization of a bridgeless finite metric space if such a realization has no cutpoint. No polynomial algorithm is known for solving this problem.

The following algorithm, called *OptimalRealization*, uses the *MCPP* algorithm recursively, as well as *Bridge* and *NoCutpoint*, to build an optimal realization of any given finite metric space (M, d). The use of *NoCutpoint* makes it non polynomial.

Algorithm 2 *OptimalRealization*

Require: A finite metric space (M, d);

Ensure: An optimal realization G of (M, d);

Apply *Bridge* on (M, d);

if the output is of the form $(K, d_K), (L, d_L), a, b, \ell$ **then**

Construct optimal realizations G_K and G_L of (K, d_K) and (L, d_L) by applying *OptimalRealization*;

Add an edge of length ℓ linking a in G_K and b in G_L;

else

Apply *MCPP* on (M, d);

if the output indicates that no optimal realization of (M, d) has a cutpoint **then**

Apply *NoCutpoint* on (M, d) to build an optimal realization G of (M, d);

else

Let $(x, y)(z, t), \{M_1, \cdots, M_k\}$ be the output of *MCPP*;

if one of the sets M_r is a singleton {u} **then**

for all $q \neq r$ **do**

Apply *OptimalRealization* to construct an optimal realization G_q of $(M_q \cup \{u\}, d|_{M_q \cup \{u\}})$;

end for

else

for all $q = 1 \cdots k$ **do**

build M'_q by adding an auxiliary element u to M_q, and define $d_{M'_q}(v, q) = d(v, w)$ for all $v, w \in M_q$ and $d_{M'_q}(v, u) = f_{(x,y)(z,t)}(v)$ for all $v \in M_q$;

Apply *OptimalRealization* to construct an optimal realization G_q of $(M'_q, d_{M'_q})$;

end for

an optimal realization G of (M, d) is obtained by gluing all G_i's at their unique common vertex u;

end if

end if

end if

Figure 4 illustrates its use for the example of Figure 1. Since the given metric space (M, d) has an optimal realization that contains a bridge, algorithm *Bridge* determines two metric spaces $\mathbf{M}_1$ on $K = \{1, 2, 3, 4, 5, 6, a\}$ and $\mathbf{M}_2$ on $L = \{b = 7, 8, 9, 10, 11, 12\}$, these two metric spaces being linked by a bridge $(a, b = 7)$ of length 1. The Metric $\mathbf{M}_1$ is bridgeless but contains a cutpoint. A possible output of the $MCPP$ algorithm is then $(x = 1, y = 3), (z = 5, t = a)$ and $M_1 = \{1, 2, 3\}, M_2 = \{4\}, M_3 = \{5, 6, a\}$. For illustration, we represent the function $f_{(1,3)(5,a)}$ as well as the graph $H_{(1,3)(5,a)}$. We therefore create two metric spaces $\mathbf{M}_3$ and $\mathbf{M}_4$ on $\{1, 2, 3, 4\}$ and $\{4, 5, 6, a\}$. Since $\mathbf{M}_3$ and $\mathbf{M}_4$ have no cutpoint (which is detected by applying $MCPP$, an optimal realization of $\mathbf{M}_1$ is obtained by making the union of optimal realizations G_3 and G_4 of $\mathbf{M}_3$ and $\mathbf{M}_4$, these being obtained by applying $NoCutpoint$.

Similarly, the Metric $\mathbf{M}_2$ is bridgeless but contains a cutpoint. A possible output of $MCPP$ is $(x = 7, y = 9), (z = 10, t = 12)$, and $M_1 = \{7, 8, 9\}, M_2 = \{10, 11, 12\}$. Again, we represent the function $f_{(7,9)(10,12)}$ and the graph $H_{(7,9)(10,12)}$. We then create two metric spaces $\mathbf{M}_5$ and $\mathbf{M}_6$ on $\{7, 8, 9, u\}$ and $\{10, 11, 12, u\}$, where u is an auxiliary element at distance 1 from $7, 9, 10, 12$, and at distance 2 from 8 and 11. Since $\mathbf{M}_5$ and $\mathbf{M}_6$ have no cutpoint (which is detected by using $MCPP$), an optimal realization of $\mathbf{M}_2$ is obtained by making the union of optimal realizations G_5 and G_6 of $\mathbf{M}_5$ and $\mathbf{M}_6$, these being obtained by applying $NoCutpoint$.

Finally, an optimal realization G of (M, d) is obtained by linking a in G_1 with $b = 7$ in G_2 with a bridge of length 1.

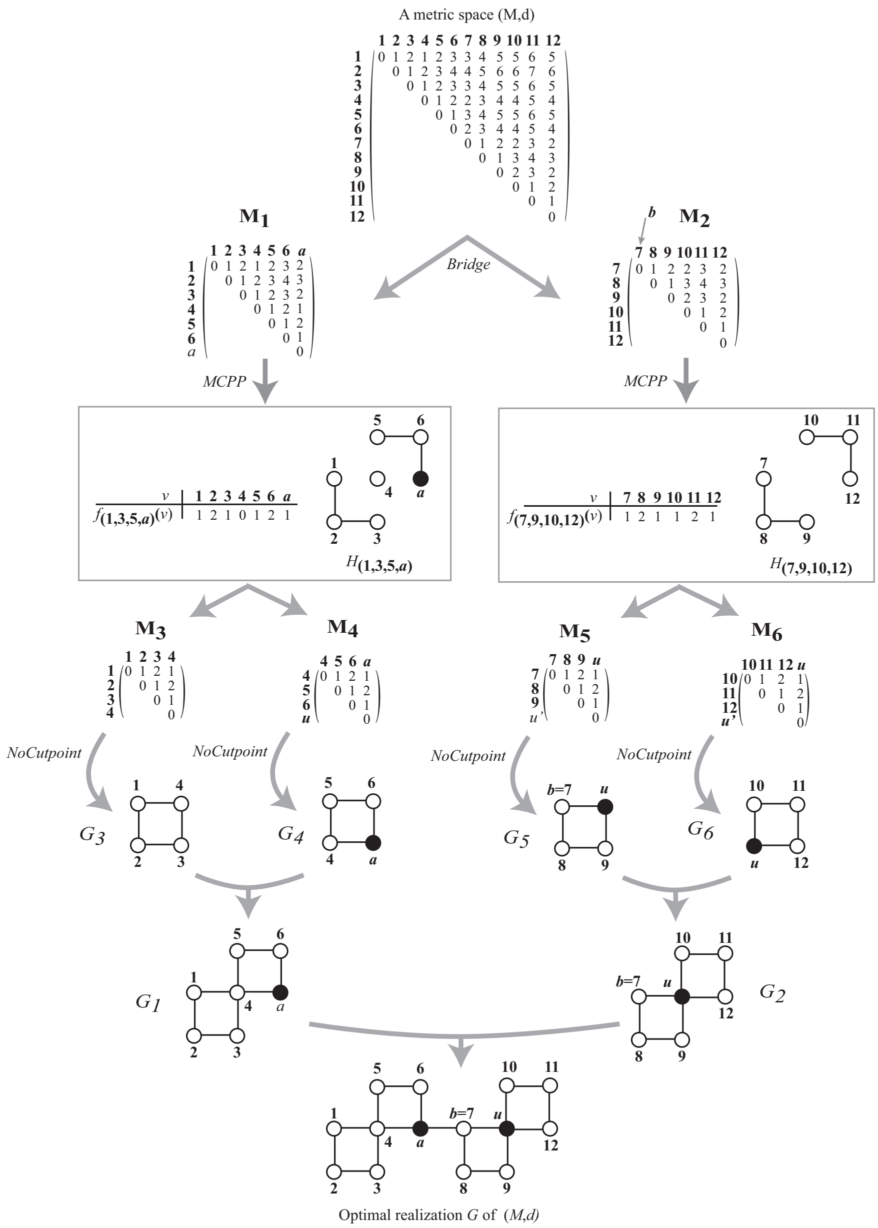


Figure 4. Construction of an optimal realization

5 Conclusion

We have proved that the Metric Cutpoint Partition Problem is polynomially solvable. The proposed algorithm can be used to construct an optimal realization of a metric space (M, d) using building blocks. More precisely, let G be a minimal realization of a finite metric space (M, d), let $G_1, \cdots, G_k$ be the blocks of G, and let M_r be the union of the points of M in G_r together with the cutpoints of G in G_r, $r = 1, \ldots, k$. Imrich *et al.* [11] have proved that if every G_r is an optimal realization of the metric space induced by G on M_r, then G is also optimal. We have shown in this paper that the sets M_r can be constructed in $O(|M|^6)$ time. Dress et al. [7] have recently shown that, using the algorithm described in [6] for the computation of so-called virtual cutpoints in finite metric spaces, it is possible to construct the above sets M_r in $O(|M|^3)$ time.

References

1. I. Abraham, Y. Bartal, and O. Neiman. Embedding metrics into ultrametrics and graphs into spanning trees with constant average distortion. In *Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 502–511, New Orleans, Louisiana, 2007.
2. J.P. Barthélemy and A. Guenoche. *Les arbres et les représentations des proximités.* Masson, Paris, 1988.
3. M. Bădoiu, P. Indyk, and A. Sidiropoulos. Approximation algorithms for embedding general metrics into trees. In *Proceedings of the Eighteenth Annual*

ACM-SIAM Symposium on Discrete Algorithms, pages 512–521, New Orleans, Louisiana, 2007.

4. F. Chung, M. Garrett, R. Graham, and D. Shallcross. Distance realization problems with applications to internet tomography. *Journal of Computer and System Sciences*, 63:432–448, 2001.
5. J. A. Cunningham. Free trees and bidirectional trees as representations of psychological distance. *J. Math. Psychol*, 17:165–188, 1978.
6. A. Dress, K. Huber, J. Koolen, and V. Moulton. An algorithm for computing virtual cut points in finite metric spaces. In *International Conference on Combinatorial Optimization and Applications (COCOA)*. Lecture Notes in Computer Science. Springer. in press.
7. A. Dress, K. Huber, J. Koolen, V. Moulton, and A. Spillner. A note on metric cut points and bridges. Technical report, School of Computing Sciences, University of East Anglia, Norwich, NR4 7TJ, UK, 2007.
8. 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.*, 53:321–402, 1984.
9. S. L. Hakimi and S. S. Yau. Distance matrix of a graph and its realizability. *Quart. Appl. Math.*, 22:305–317, 1964.
10. A. Hertz and S. Varone. The metric bridge partition problem : partitioning of a metric space into two subspaces linked by an edge in any optimal realization. *Journal of Classification*, to appear.
11. W. Imrich, J. M. S. Simões-Pereira, and C. M. Zamfirescu. On optimal embeddings of metrics in graphs. *J. Combin. theory*, 36B:1–15, 1984.
12. W. Imrich and E. Stockiĭ. On optimal embeddings of metrics in graphs. *Sibirsk. Mat. Ž.*, 13:558–565, 1972.

13. P. A. Landry, F. J. Lapointe, and J. A. W. Kirsch. Estimating phylogenies from distance matrices: additive is superior to ultrametric estimation. *Molecular Biology and Evolution*, 13:818–823, 1996.

14. M. Li, X. Chen, B. Ma, and P.M.B. Vitányi. The similarity metric. *IEEE Transactions on Information Theory*, 50(12):3250–3264, 2004.

15. V. Makarenkov. Comparison of four methods for inferring phylogenetic trees from incomplete dissimilarity matrices. In *Classification, Clustering, and Data Analysis*, pages 371–378. Springer, Cracow, Poland, 2002.

16. P. Winkler. The complexity of metric realisation. *SIAM J. Disc. math.*, 1(4):552–559, 1988.