OPTIMAL REALIZATIONS AND THE BLOCK DECOMPOSITION OF A FINITE METRIC SPACE

KATHARINA T. HUBER,, VINCENT MOULTON, AND ANDREAS SPILLNER

Abstract. Finite metric spaces are an essential tool in discrete mathematics and have applications in several areas including computational biology, image analysis, speech recognition, and information retrieval. Given any such metric D on a finite set X, an important problem is to find appropriate ways to *realize* D by weighting the edges in some graph G containing X in its vertex set such that $D(x, y)$ equals the length of a shortest path from x to y in G for all $x, y \in X$. Here we focus on realizations with minimum total edge weight, called *optimal* realizations. By considering the 2-connected components and bridges in any optimal realization G of D we obtain an additive decomposition of D into simpler metrics. We show that this decomposition, called the *block decomposition*, is canonical in that it only depends on D and *not* on G, and that the decomposition can be computed in $O(|X|^3)$ time. As well as providing a fundamental new way to decompose any finite metric space, we expect that the block decomposition will provide a useful preprocessing tool for deriving metric realizations.

Keywords finite metric space; optimal realization; block decomposition; block realization; cut points

1. Introduction

Let X denote a finite non-empty set, and D a *metric* on X, i.e., a symmetric map $D : X \times X \to \mathbb{R}_{\geq 0}$ with $D(x, x) = 0$ and $D(x, z) \leq D(x, y) + D(y, z)$, for all $x, y, z \in X$. Note that in this paper we allow $D(x, y) = 0$ even when $x \neq y$. In case there is no such pair in X, a *realization* of D is a graph $G = (V, E)$ with non-negative edge weights such that $X \subseteq V$ and $D(x, y)$ equals the length of a shortest path from x to y in G, for all $x, y \in X$ (see Fig. 1 for an example). A realization of D is *optimal* if it has minimum total edge weight among all realizations of D. It is known that every metric has at least one optimal realization [22, Theorem 2.2], but computing optimal realizations is an NP-hard problem [1, 24].

Applications of optimal realizations include internet traffic-flow analysis [4], electrical circuit design [15], the minimum Manhattan network problem (see e.g. [5] and the references therein), and phylogenetics. The latter application is motivated by the fact that in case a metric D on X can be realized by a tree (i.e. a connected graph with no cycles) with leaf set X, then this realization is necessarily optimal [15, Theorem 6]. Since metrics are now routinely generated from molecular data (essentially by computing the Hamming distance between aligned DNA sequences), optimal realizations and related graphs can be used to

Date: June 21, 2021.

D	a	b	c	d	e	f	g
a	0	2	3	5	7	7	5
b	2	0	3	5	7	7	5
c	3	3	0	2	4	4	2
d	5	5	2	0	2	4	4
e	7	7	4	2	0	2	4
f	7	7	4	4	2	0	2
g	5	5	2	4	4	2	0

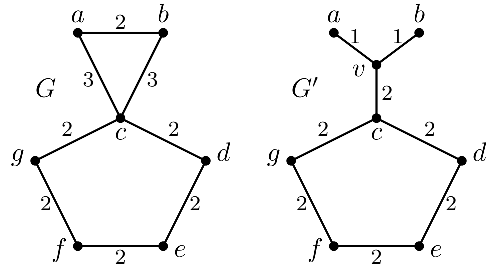


FIGURE 1. A metric D on $X = \{a, b, c, d, e, f, g\}$. G is a realization of D with total edge weight $|G| = 18$ and G' is an optimal realization of D. We have $Cut(G') = \{c, v\}$ and $\mathfrak{P}_v = \{\{a\}, \{b\}, \{c, d, e, f, g\}\}$ is the partition of X associated to the cut vertex v of G'. $B = \{a, v\}$ is one of the four blocks of G' and yields the metric D_B with $D_B(a, x) = 1$ for all $x \in X - \{a\}$ and $D_B(x, y) = 0$ for all other $x, y \in X$.

elucidate the evolutionary relationships between species in the form of trees or more general graphs called phylogenetic networks (see e.g. [21]). Such networks are commonly used to analyze the evolution of species that evolve in a non tree-like fashion (e.g. viruses which can recombine to form hybrid forms). Early work in [6] on understanding optimal realizations and their relationship with a structure called the tight-span (defined in Section 2.2 below) led to the concept of so-called split-networks [3], a type of phylogenetic network which is now routinely used in evolutionary analysis. For example, in Fig. 2 we present a split-network derived from DNA sequences of five coronaviruses, among them the one which causes the coronavirus disease 2019 (SARS-CoV-2) [25]. Developing approaches to generate phylogenetic networks remains an active area of research in the field of phylogenetics [14].

1.1. **Our main result and a sketch of its proof.** We shall focus on two features of optimal realizations G of a metric D on X: The set $Cut(G)$ of *cut vertices* of G and the collection $\mathcal{D}(G)$ consisting of the metrics D_B associated with each *block* B (i.e. 2-connected component or bridge) of G by putting $D_B(x, y)$, for any $x, y \in X$, to be the length of that part of any shortest path from x to y that is contained in B (see Fig. 1 again for an example). Clearly $\mathcal{D}(G)$ is an additive *decomposition* of D, that is, $D = \sum_{D' \in \mathcal{D}(G)} D'$. In the main result of this paper we show that $Cut(G)$ is completely determined by D and hence the decomposition $\mathcal{D}(G)$, or *block decomposition of* D is also determined by D. In particular, this provides a fundamental new way to decompose any finite metric into simpler metrics.

A key ingredient to proving our main results is the theory of the *tight span* $T(D)$ of a metric D [6] which, since X is finite, forms a connected polytopal complex which is a subset of $\mathbb{R}^X$. To see how the tight span arises, note that we can associate a map $f_v \in \mathbb{R}^X$ to any vertex v in an optimal realization G of a metric D on X which is necessarily in $T(D)$: Set $f_v(x)$ to be the length of a shortest path from v to x in G, for all $x \in X$. In particular, there exists a *cut point* $f \in T(D)$, that is, some f so that the set $T(D) - \{f\}$ is disconnected, if and only if either every optimal realization G of D contains a cut vertex v

D	a	b	c	d	e
a	0.000	0.514	0.514	0.519	0.970
b	0.514	0.000	0.341	0.373	0.920
c	0.514	0.341	0.000	0.139	0.766
d	0.519	0.373	0.139	0.000	0.742
e	0.970	0.920	0.766	0.742	0.000

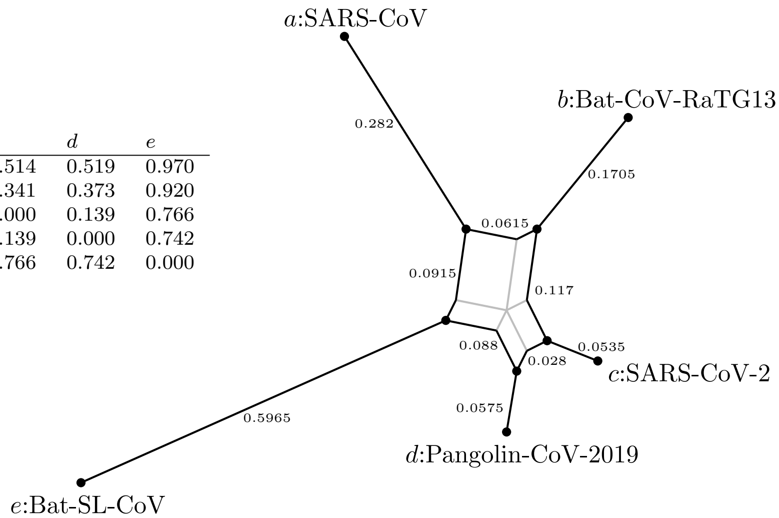


FIGURE 2. A split-network for estimated distances D between five DNA sequences related to the SARS-CoV-2 S protein given in [25, Tab. 2]. The lengths of the edges are proportional to their weights, and the dark part of the network is in fact an optimal realization of D. The network was produced with the help of the program SplitsTree [21] and it is a graphical representation of the so-called split decomposition of D (see also Section 5).

with $f = f_v$ or every optimal realization G of D contains a bridge into which we can insert a new vertex v of degree two with $f = f_v$. It follows that the set $Virt(D)$ of cut points of $T(D)$ – or "virtual cut points" of D [10] – corresponds, in the sense described above, precisely to the set of cut vertices or new degree 2 vertices within bridges that must occur in every optimal realization of D.

In the light of these facts, a full knowledge of the set $Virt(D)$ gives access to the cut points and bridges that appear in *every* optimal realization G of D and hence to the decomposition $\mathcal{D}(G)$. The problem of computing an optimal realization of D can therefore be reduced to the problem of computing optimal realizations for the metrics in $\mathcal{D}(G)$. To complete our proof of the uniqueness of the block decomposition, we show that it is a *compatible decomposition* [9]. This allows us to exploit a correspondence given in [8] between certain finite subsets of $Virt(D)$ and *block realizations* of D (essentially, these are realizations in which every 2-connected component is a clique). In particular, block realizations give us a tool for controlling the way in which we reassemble an optimal realization for D from optimal realizations for the metrics in $\mathcal{D}(G)$.

1.2. **Related work and the structure of the rest of this paper.** In [22] it is shown that cut vertices v in optimal realizations of a metric D on X correspond to certain maps $f_v \in \mathbb{R}^X$. These maps were introduced as a useful technical tool for handling optimal realizations but without further investigating the space of all maps that arise from a given metric in this way. A rather technical outline how to possibly capture this space of maps was presented in [11, Section 4], but without giving any details or proof. In this paper, we first provide in Section 2 formal definitions of the concepts informally introduced above. Then, in Section 3, we give a proof that the space of maps f_v mentioned above can be concisely characterized by a certain finite subset $Virt^*(D) \subseteq Virt(D)$ (Theorem 5). After that, in Section 4, we prove that every metric has a unique block decomposition (Theorem 8). A first short outline of how such a decomposition might be obtained for any given metric can be found in [13, Section 6].

Computing the block decomposition is closely related to computing optimal realizations. In [18, 19] algorithms are presented to check whether a metric can be decomposed by a cut vertex or a bridge in polynomial time. Iterative application of these algorithms results in a decomposition $\mathcal{D}$ of the input metric D that cannot be further refined, and, assuming that optimal realizations of the metrics in $\mathcal{D}$ are provided, an optimal realization of D is then assembled in polynomial time. However, it is not shown in [18, 19] that the decomposition $\mathcal{D}$ obtained by this approach, which must be the block decomposition of D, is uniquely determined by D. Applying the algorithms from [18, 19] to a metric D on X, the block decomposition of D can be computed in $O(|X|^6)$ time. In Section 4 we show that the block decomposition of D can be computed in $O(|X|^3)$ time (Theorem 10). A key ingredient for achieving this run time is an algorithm presented in [12] for computing the set $Virt^*(D)$ in $O(|X|^3)$ time. As a corollary, we also give a bound on the time required to compute an optimal realization of a metric in terms of the components in its block decomposition (Theorem 11).

In Section 5 we first prove that the block decomposition of a metric always corresponds to a block realization with minimum total edge weight (Theorem 13). As a consequence, every metric has a unique block realization with minimum total edge weight, a fact that was stated without proof in [9, p.1619]. We then conclude by describing a relationship between the well-known split decomposition of a metric D [2] and the block decomposition of D (Theorem 14), as well as some related open problems.

2. Preliminaries

2.1. **Realizations.** A *realization* $G = ((V, E), \omega, \varphi)$ of a metric D on X consists of a graph (V, E), an edge-weighting $\omega : E \to \mathbb{R}_{>0}$ and a labeling map $\varphi : X \to V$ such that $D(x, y) = D_G(\varphi(x), \varphi(y))$ for all $x, y \in X$, where $D_G : V \times V \to \mathbb{R}_{\geq 0}$ is the shortest-path metric on V induced by (V, E) and the edge weighting ω. In addition, letting $|G| = \sum_{e \in E} \omega(e)$ denote the total length of a realization $G = ((V, E), \omega, \varphi)$ of a metric D, a realization $G^* = ((V^*, E^*), \omega^*, \varphi^*)$ of a metric D is *optimal* if $|G^*| \leq |G'|$ holds for all realizations G' of D and all vertices in $V^* - \varphi^*(X)$ have degree at least 3.

Now, let $G = ((V, E), \omega, \varphi)$ be a realization of a metric D on X. For any vertex $v \in V$ we denote by $G - v$ the graph $(V - \{v\}, E \cap \binom{V - \{v\}}{2})$, where $\binom{M}{2}$ denotes the set of all 2-element subsets of a finite set M. Similarly, for any edge $e \in E$ we denote by $G - e$ the graph $(V, E - \{e\})$. A vertex $v \in V$ is called a *cut vertex* of G if the graph $G - v$ has at least two connected components. We denote by $Cut(G)$ the set of all cut vertices of G. Similarly, an edge $e \in E$ is called a *bridge* of G if the graph $G - e$ has two connected components. In addition, let $\mathcal{B}(G)$ denote the collection of those subsets of V that are vertex sets of

D	a	b	c	d
a	0	3	4	3
b	3	0	3	4
c	4	3	0	3
d	3	4	3	0

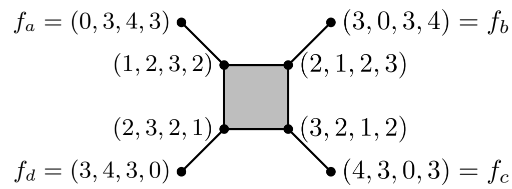


FIGURE 3. A metric D on the set $X = \{a, b, c, d\}$ and a projection of its tight span $T(D)$ into the plane with the gray square representing a 2-dimensional face. For every 0-dimensional face $\{f\}$ of $T(D)$ the 4-dimensional vector $(f(a), f(b), f(c), f(d))$ is given. The 1-dimensional face containing $(1, 2, 3, 2)$ and $(2, 1, 2, 3)$ corresponds to those $g \in P(D)$ that satisfy the equations $g(a) + g(b) = D(a, b)$, $g(a) + g(c) = D(a, c)$ and $g(b) + g(d) = D(b, d)$. Each face of $T(D)$ corresponds to such a set of equations. The structure of D determines which of the inequalities in the definition of $P(D)$ can simultaneously become equalities and, thus, yield faces of $T(D)$.

maximal 2-connected subgraphs or bridges of (V, E). We refer to the elements of $\mathcal{B}(G)$ as the *blocks* of G. Then, associating with each $B \in \mathcal{B}(G)$ the metric D_B on X defined by putting, for all $x, y \in X$, $D_B(x, y) = D_G(u, u')$, where u is the first vertex in B and u' is the last vertex in B along a shortest path from $\varphi(x)$ to $\varphi(y)$ in G, if such vertices exist, and $D_B(x, y) = 0$ otherwise, we have $D = \sum_{B\in\mathcal{B}(G)} D_B$.

In the following, we call a collection $\mathfrak{P}$ of non-empty subsets of X a *partition* of X if $\bigcup_{A\in\mathfrak{P}} A = X$ and $A \cap B = \emptyset$ for all $A, B \in \mathfrak{P}$ with $A \neq B$. A *split* of X is a partition $\mathfrak{P}$ of X with $|\mathfrak{P}| = 2$. To every cut vertex v of a realization $G = ((V, E), \omega, \varphi)$ of a metric D on X we associate the collection $\mathfrak{P}_v = \{\varphi^{-1}(C) : C \in \mathcal{C}_v\}$ of subsets of $X - \varphi^{-1}(v)$ where $\mathcal{C}_v$ is the collection of the vertex sets of the connected components of $G - v$. Similarly, to every bridge $e = \{u, v\}$ of G we associate the collection $S_e = \{\varphi^{-1}(C_u), \varphi^{-1}(C_v)\}$ of subsets of X where C_u and C_v are the vertex sets of the two connected components of $G - e$ containing u and v, respectively. For all realizations considered in this paper we assume that the collections $\mathfrak{P}_v$ and S_e are partitions of $X - \varphi^{-1}(v)$ and X, respectively (which in particular is necessarily the case for optimal realizations and also for weak block realizations which we define in Section 4).

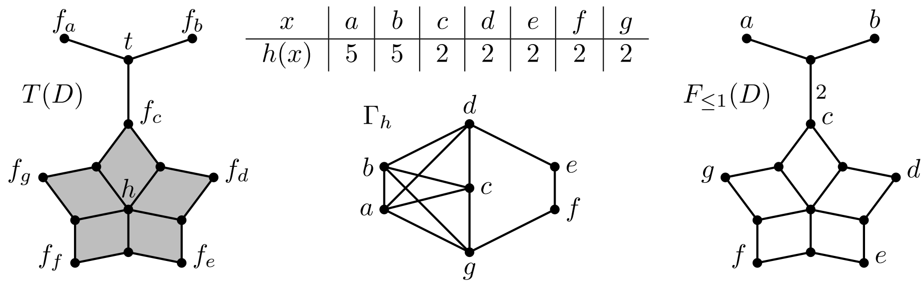
2.2. **Tight spans.** We now recall some concepts and basic facts concerning tight spans [6]. The *tight span* of a metric D on X is the set

$$T(D) = \{f \in \mathbb{R}^X : f(x) = \max(D(x, y) - f(y) : y \in X) \text{ for all } x \in X\}.$$

Note that $T(D)$ can be viewed as the polytopal complex obtained by taking the set of bounded faces of the polyhedron

$$P(D) = \{f \in \mathbb{R}^X : f(x) + f(y) \geq D(x, y) \text{ for all } x, y \in X\}.$$

In Fig. 3 we give an example. For every $k \in \mathbb{N}$ we denote by $F_k(D)$ the collection of k-dimensional faces of the polytopal complex $T(D)$. Later we will make use of the fact that $f \in P(D)$ is an element of $T(D)$ if and only if for all $x \in X$ there exists some $y \in X$



x	a	b	c	d	e	f	g
$h(x)$	5	5	2	2	2	2	2

FIGURE 4. A projection of the tight span $T(D)$ of the metric D in Fig. 1 into the plane. This tight span has only faces of dimension at most two and the gray parallelograms indicate the 2-dimensional faces. The graph Γ_h for the map h that forms a 0-dimensional face of $T(D)$ is connected and, thus, $h \not\in Virt(D)$, which is in agreement with the fact that $T(D) - \{h\}$ is connected. $Virt(D)$ corresponds to the set of points on the three 1-dimensional faces with one endpoint in t (except the points f_a and f_b) and $Virt^*(D) = \{t, f_c\}$. All edges in the realization $F_{\leq 1}(D)$ of D obtained from $T(D)$ have weight 1 (these weights are omitted in the figure), except for one edge that has weight 2.

with $f(x) + f(y) = D(x, y)$. In particular, for all $x \in X$, the map $f_x : X \to \mathbb{R}_{\geq 0}$ with $f_x(y) = D(x, y)$ for all $y \in X$ is contained in $T(D)$ and, more precisely, forms an element of $F_0(D)$. Note that $f \in T(D)$ with $f(x) = 0$ for some $x \in X$ implies $f = f_x$.

A map $f \in T(D)$ is a *virtual cut point* of D if there exists a split of the *support* $\text{supp}(f) = \{x \in X : f(x) > 0\}$ of f into two subsets A and B such that $D(a, b) = f(a)+f(b)$ holds for all $a \in A$ and $b \in B$. In addition, for every map $f \in T(D)$, let $\Gamma_f = (\text{supp}(f), E_f)$ denote the graph with vertex set $\text{supp}(f)$ and edge set

$$E_f = \{\{x, y\} \in \binom{\text{supp}(f)}{2} : f(x) + f(y) > D(x, y)\},$$

and let $\mathfrak{P}_f$ denote the partition of $\text{supp}(f)$ formed by the vertex sets of the connected components of Γ_f. We let $Virt(D) \subseteq T(D)$ denote the set of virtual cut points of D and we let $Virt^*(D)$ denote the subset of those $f \in Virt(D)$ with $f = f_x$ for some $x \in X$ or with Γ_f not a disjoint union of two cliques.

As mentioned in the introduction, $T(D)$ is a connected subset of $\mathbb{R}^X$. It is shown in [10] that the set $Virt(D)$ consists of precisely those $f \in T(D)$ for which $T(D) - \{f\}$ is disconnected and it follows from the characterization of $F_0(D)$ given in [7] that $Virt^*(D) = Virt(D) \cap F_0(D)$. Moreover, we put $F_{\leq 1}(D) = ((V, E), \omega, \varphi)$ with $V = F_0(D)$, E consisting of those $\{f, g\} \in \binom{F_0(D)}{2}$ for which there exists some $C \in F_1(D)$ with $\{f, g\} \subseteq C$, ω the map that assigns to every $\{f, g\} \in E$ the value $\max(|f(x) - g(x)| : x \in X)$ and φ mapping every $x \in X$ to f_x. It is shown in [6] that $F_{\leq 1}(D)$ is a realization of D and that $Cut(F_{\leq 1}(D)) = Virt^*(D)$. In Fig. 4 some of the concepts introduced above are illustrated.

3. Virtual cut points and optimal realizations

In this section, we describe in more detail how the maps in $Virt(D)$ are related to the cut vertices and bridges that occur in optimal realizations of a metric D. In particular, we will make precise in Theorem 5 below how the subset $Virt^*(D)$ of $Virt(D)$ defined in the previous section can be thought of as a representation of the set of cut vertices that are present in *every* optimal realization of D. To establish this, we shall need several observations beginning with the following (cf. [22, Theorem 5.1] and reference [23] therein):

Theorem 1. *Suppose that D is a metric on X for which there exists a partition of X into two nonempty subsets K, L and a map $f : X \to \mathbb{R}_{\geq 0}$ with $D(x, y) \leq f(x) + f(y)$ for all $x, y \in X$, with equality holding whenever $x \in K$ and $y \in L$ and with $f(x) > 0$ for at least one $x \in K$ and at least one $x \in L$. Then precisely one of the following statements holds:*

- (i) *For all optimal realizations $G = ((V, E), \omega, \varphi)$ of D there exists a cut vertex $u \in V$ with $D_G(\varphi(x), u) = f(x)$ for all $x \in X$.*
- (ii) *For all optimal realizations $G = ((V, E), \omega, \varphi)$ of D there exists a bridge $e = \{u, v\} \in E$ and a pair of strictly positive real numbers α and β with $\alpha + \beta = \omega(e)$ such that $D_G(\varphi(x), u) + \alpha = f(x)$ holds for all $x \in X$ with $D_G(\varphi(x), u) < D_G(\varphi(x), v)$, and $D_G(\varphi(x), v) + \beta = f(x)$ holds for all $x \in X$ with $D_G(\varphi(x), v) < D_G(\varphi(x), u)$.*

Using Theorem 1 we now relate the elements of $Virt(D)$ to cut vertices and bridges in optimal realizations of D.

Lemma 2. *Let D be a metric on X and $f \in Virt(D)$. Then precisely one of the following holds:*

- (i) *For all optimal realizations $G = ((V, E), \omega, \varphi)$ of D there exists a unique $u \in Cut(G)$ with $\mathfrak{P}_u = \mathfrak{P}_f$ and $D_G(\varphi(x), u) = f(x)$ for all $x \in X$.*
- (ii) *For all optimal realizations $G = ((V, E), \omega, \varphi)$ of D there exists a unique bridge $e = \{u, v\}$ of G and a pair of strictly positive real numbers α and β with $\alpha + \beta = \omega(e)$ such that $S_e = \mathfrak{P}_f$, $D_G(\varphi(x), u) + \alpha = f(x)$ holds for all $x \in X$ with $D_G(\varphi(x), u) < D_G(\varphi(x), v)$ and $D_G(\varphi(x), v) + \beta = f(x)$ holds for all $x \in X$ with $D_G(\varphi(x), v) < D_G(\varphi(x), u)$.*

Proof: Let $f \in Virt(D)$. Then, by definition, there exists a split of $\text{supp}(f)$ into subsets A and B such that $D(a, b) = f(a) + f(b)$ for all $a \in A$, $b \in B$. Define $K = A$ and $L = X - A$. Then $\{K, L\}$ is a split of X. Moreover, since $f \in T(D)$, $D(x, y) \leq f(x) + f(y)$ for all $x, y \in X$ and, since $\{A, B\}$ is a split of $\text{supp}(f)$, $f(x) > 0$ for at least one $x \in K$ and for at least one $x \in L$. Also note that if there exist any $z \in L - B$ we have $f(z) = 0$ and, thus, $f = f_z$ (cf. Section 2.2) implying that $f(z) + f(x) = D(z, x)$ for all $x \in K$. Therefore, in view of $D(a, b) = f(a) + f(b)$ for all $a \in A$, $b \in B$, we have $f(x) + f(y) = D(x, y)$ for all $x \in K$, $y \in L$. Hence, either statement (i) or (ii) in Theorem 1 must hold for all optimal realizations of D.

Case 1: Statement (i) in Theorem 1 holds for all optimal realizations of D. Let $G = ((V, E), \omega, \varphi)$ be an optimal realization of D and let $u \in V$ be a cut vertex with

$D_G(\varphi(x),u)=f(x)$ for all $x \in X$. First we show that u is unique. Assume for a contradiction that there exists a cut vertex $u' \in V$ with $u' \neq u$ and $D_G(\varphi(x),u')=f(x)$ for all $x \in X$. In view of $u \neq u'$ there must exist a connected component C of $G-u$ that does not contain u' and, similarly, a connected component C' of $G-u'$ that does not contain u. Moreover, since G is optimal, there must exist some $x \in X$ with $\varphi(x) \in C$ and some $x' \in X$ with $\varphi(x') \in C'$. Then, in view of $D_G(u,u')>0$, we have

$$f(x')=D_G(u,\varphi(x'))=D_G(u,u')+D_G(u',\varphi(x'))=D_G(u,u')+f(x')>f(x'),$$

a contradiction.

It remains to show that $\mathfrak{P}_u=\mathfrak{P}_f$. Note that both $\mathfrak{P}_u$ and $\mathfrak{P}_f$ are partitions of $X-\varphi^{-1}(u)=\text{supp}(f)$. We first establish that for all $A \in \mathfrak{P}_f$ there exists some $B \in \mathfrak{P}_u$ with $A \subseteq B$. This follows immediately if $|A|=1$. If $|A| \geq 2$, let x and y be arbitrary distinct elements of A. Since x and y belong to the same connected component of Γ_f there exists a sequence $x=z_1,z_2,\dots,z_l=y$ of elements of X with the property that $f(z_i)+f(z_{i+1})>D(z_i,z_{i+1})$ for all $i \in \{1,\dots,l-1\}$. Therefore, since we have $f(x')+f(y')=D(x',y')$ for any $x',y' \in X-\varphi^{-1}(u)$ that are contained in different sets in $\mathfrak{P}_u$, z_i and z_{i+1} must be contained in the same set of $\mathfrak{P}_u$ for all $i \in \{1,\dots,l-1\}$. Hence, x and y belong to the same set in $\mathfrak{P}_u$.

Now, in order to show $\mathfrak{P}_u=\mathfrak{P}_f$, assume for a contradiction that there exist $A \in \mathfrak{P}_f$ and $B \in \mathfrak{P}_u$ with $A \subsetneq B$. Put $X'=B \cup \{u\}$, $K=A$, $L=X'-A$ and consider the metric D' on X' defined by putting $D'(x,y)=D(x,y)$ and $D'(u,x)=D_G(u,\varphi(x))$ for all $x,y \in B$. Let C' be the vertex set of the connected component of $G-u$ that contains $\varphi(B)$ and put $V'=C' \cup \{u\}$. Note that since G is an optimal realization of D it follows that $G'=((V',E'),\omega',\varphi')$ with $E'=E \cap \binom{V'}{2}$, ω' the restriction of ω to E' and $\varphi': X' \to V'$ defined by $\varphi'(x)=\varphi(x)$ for all $x \in B$ and $\varphi'(u)=u$ is an optimal realization of D'. Moreover, for the map $f': X' \to \mathbb{R}_{\geq 0}$ with $f'(x)=f(x)$ for all $x \in B$ and $f'(u)=0$, we have $D(x',y') \leq f'(x')+f'(y')$ for all $x',y' \in X'$, $D(x',y')=f'(x')+f'(y')$ for all $x' \in K$ and all $y' \in L$, $f'(x')>0$ for some $x' \in K$ and $f'(y')>0$ for some $y' \in L$. Thus, in view of Theorem 1 and the fact that we have $f'(x')=D_{G'}(u,\varphi'(x'))$ for all $x' \in X'$, it follows that u is a cut vertex in G'. But then the vertices in $\varphi(B)$ cannot be contained in a single connected component of $G-u$, a contradiction.

Case 2: Statement (ii) in Theorem 1 holds for all optimal realizations of D. Let $G=((V,E),\omega,\varphi)$ be an arbitrary optimal realization of D and let $e=\{u,v\} \in E$ be a bridge in G such that, for a suitable pair of positive real numbers α and β with $\alpha+\beta=\omega(e)$, we have that $D_G(\varphi(x),u)+\alpha=f(x)$ holds for all $x \in X$ with $D_G(\varphi(x),u)<D_G(\varphi(x),v)$, and that $D_G(\varphi(x),v)+\beta=f(x)$ holds for all $x \in X$ with $D_G(\varphi(x),v)<D_G(\varphi(x),u)$. Using a similar argument as in Case 1 it follows that the bridge e is unique. In order to show that $S_e=\mathfrak{P}_f$, let C_u and C_v denote the vertex sets of the connected components of $G-e$ that contain u and v, respectively. Define $K=\varphi^{-1}(C_u)$ and $L=\varphi^{-1}(C_v)$. Then $S_e=\{K,L\}$ and, for all $x,y \in K$ with $x \neq y$, we have

$$f(x)+f(y)=D_G(\varphi(x),u)+D_G(\varphi(y),u)+2\alpha>D_G(\varphi(x),\varphi(y))=D(x,y).$$

Similarly, it follows that $f(x) + f(y) > D(x, y)$ for all $x, y \in L$ with $x \neq y$. Since, in addition, we have

$$f(x) + f(y) = D_G(\varphi(x), u) + \alpha + D_G(\varphi(y), v) + \beta = D_G(\varphi(x), \varphi(y)) = D(x, y)$$

for all $x \in K$ and $y \in L$, it follows that $\mathfrak{P}_f = \{K, L\} = S_e$, as required. □

Remark 3. *The last step in the argument given in Case 1 is related to a result in* [22, Corollary 5.2]*, which is given there without proof.*

Next we observe that the converse of Lemma 2 also holds. To state this result, we define, for every vertex v in a realization $G = ((V, E), \omega, \varphi)$ of a metric on X, the map $f_v : X \to \mathbb{R}_{\geq 0}$ with $f_v(x) = D_G(v, \varphi(x))$ for all $x \in X$.

Lemma 4. *Let $G = ((V, E), \omega, \varphi)$ be an optimal realization of a metric D on X.*

- (i) *For every $u \in Cut(G)$ we have $f_u \in Virt(D)$ and $\mathfrak{P}_u = \mathfrak{P}_{f_u}$.*
- (ii) *For every bridge $e = \{u, v\}$ in G and every pair of strictly positive real numbers α and β with $\alpha + \beta = w(e)$ we have $f_{(e,\alpha,\beta)} \in Virt(D)$ for the map $f_{(e,\alpha,\beta)} : X \to \mathbb{R}_{\geq 0}$ defined by putting $f_{(e,\alpha,\beta)}(x) = D_G(\varphi(x), u) + \alpha$ for all $x \in X$ with $D_G(\varphi(x), u) < D_G(\varphi(x), v)$, and $f_{(e,\alpha,\beta)}(x) = D_G(\varphi(x), v) + \beta$ for all $x \in X$ with $D_G(\varphi(x), v) < D_G(\varphi(x), u)$ and $S_e = \mathfrak{P}_{f_{(e,\alpha,\beta)}}$.*

Proof: To show (i), consider $u \in Cut(G)$. It suffices to show that $f_u \in Virt(D)$. From this it follows by Lemma 2 that $\mathfrak{P}_u = \mathfrak{P}_{f_u}$. First note that, by definition of the map f_u, we have

$$f_u(x) + f_u(y) = D_G(u, \varphi(x)) + D_G(u, \varphi(y)) \geq D_G(\varphi(x), \varphi(y)) = D(x, y)$$

for all $x, y \in X$, implying $f_u \in P(D)$. To show that $f_u \in T(D)$, consider the vertex set C of an arbitrary connected component of $G - u$ and define $K = \varphi^{-1}(C)$ and $L = \varphi^{-1}(V - C)$. Then we have

$$f_u(x) + f_u(y) = D_G(u, \varphi(x)) + D_G(u, \varphi(y)) = D_G(\varphi(x), \varphi(y)) = D(x, y)$$

for all $x \in K$ and all $y \in L$, implying that, for all $x \in X$, there exists some $y \in X$ with $f(x) + f(y) = D(x, y)$. Hence, $f_u \in T(D)$. Moreover, $\{A, B\}$ with $A = K \cap \text{supp}(f_u)$ and $B = L \cap \text{supp}(f_u)$ is a split of $\text{supp}(f_u)$ into subsets A and B with $f(a) + f(b) = D(a, b)$ for all $a \in A$, $b \in B$. Thus, $f_u \in Virt(D)$, as required.

Next, to show (ii), consider a bridge $e = \{u, v\}$ in G together with a pair of strictly positive real numbers α and β such that $\alpha + \beta = \omega(e)$. Again, it suffices to show that $f_{(e,\alpha,\beta)} \in Virt(D)$ which then, in view of Lemma 2, implies $S_e = \mathfrak{P}_{f_{(e,\alpha,\beta)}}$. Let C_u and C_v denote the vertex sets of the connected components of $G - e$ that contain u and v, respectively. Then we have

$$f_{(e,\alpha,\beta)}(x) + f_{(e,\alpha,\beta)}(y) \geq D_G(\varphi(x), \varphi(y)) = D(x, y)$$

for all $x, y \in X$. Furthermore, $\{A, B\}$ with $A = \varphi^{-1}(C_u)$ and $B = \varphi^{-1}(C_v)$ is a split of $X = \text{supp}(f_{(e,\alpha,\beta)})$ such that

$$f_{(e,\alpha,\beta)}(a) + f_{(e,\alpha,\beta)}(b) = D_G(\varphi(a), \varphi(b)) = D(a, b)$$

for all $a \in A$ and all $b \in B$. Thus, $f_{(e,\alpha,\beta)} \in Virt(D)$, as required. □

Using Lemmas 2 and 4, we next prove the main result of this section.

Theorem 5. *For every optimal realization $G = ((V, E), \omega, \varphi)$ of a metric D on X, the map*

$$\tau : Cut(G) \to Virt(D) \text{ with } \tau(u) = f_u$$

is an injection with $\tau(Cut(G)) = Virt^*(D)$. In particular, if G and G' are optimal realizations of the same metric D then $|Cut(G)| = |Cut(G')| = |Virt^*(D)|$.

Proof: As an immediate consequence of Lemmas 2 and 4 we obtain that τ is an injective map. It remains to show that $\tau(Cut(G)) = Virt^*(D)$. Consider $u \in Cut(G)$. First note that if $u = \varphi(x)$ for some $x \in X$ we immediately have $f_u = f_x \in Virt^*(D)$. Also note that if $|\mathfrak{P}_u| \geq 3$ then $|\mathfrak{P}_{f_u}| = |\mathfrak{P}_u| \geq 3$ and, thus, $f_u \in Virt^*(D)$ since Γ_{f_u} is not a disjoint union of two cliques.

Therefore, the only case left to consider is $u \in V - \varphi(X)$ and $|\mathfrak{P}_u| = 2$. Then, by the definition of an optimal realization, u has degree at least 3 in G. So, there must exist a connected component C of $G - u$ such that we have $|E_C| \geq 2$ for the set $E_C \subseteq E$ of edges connecting u with a vertex in C. Note that it suffices to show that there exist $x, y \in \varphi^{-1}(C)$ such that there is a shortest path from $\varphi(x)$ to $\varphi(y)$ that contains u, since then

$$f_u(x) + f_u(y) = D_G(u, \varphi(x)) + D_G(u, \varphi(y)) = D_G(\varphi(x), \varphi(y)) = D(x, y),$$

implying that Γ_{f_u} is not a disjoint union of two cliques and, thus, $f_u \in Virt^*(D)$. So, assume for a contradiction that there is no shortest path between two vertices in $\varphi^{-1}(C)$ that contains u and, therefore, no such shortest path contains an edge in E_C. Put $\varepsilon_1 = \frac{1}{2}\cdot\min(\omega(e) : e \in E_C)$ and $\varepsilon_2 = \frac{1}{2}\cdot\min(D_G(u, \varphi(x)) + D_G(u, \varphi(y)) - D(x, y) : x, y \in \varphi^{-1}(C))$. Note that in view of our assumption we have $\varepsilon = \min(\varepsilon_1, \varepsilon_2) > 0$. Let u' be a new vertex not contained in V. We construct a new realization $G' = ((V', E'), \omega', \varphi')$ of D by putting $V' = V \cup \{u'\}$,

$$E' = (E - E_C) \cup (\{\{u, u'\}\} \cup \{\{u', v\} : \{u, v\} \in E_C\}),$$

$\omega' : E' \to \mathbb{R}_{>0}$ with $\omega'(e) = \omega(e)$ for all $e \in E \cap E'$, $\omega'(\{u, u'\}) = \varepsilon$ and $\omega'(\{u', v\}) = \omega(\{u, v\}) - \varepsilon$ for all $\{u, v\} \in E_C$, and $\varphi' : X \to V'$ with $\varphi'(x) = \varphi(x)$ for all $x \in X$. Then, by construction and in view of $|E_C| \geq 2$, we have

$$|G'| = |G| + \varepsilon - |E_C| \cdot \varepsilon < |G|,$$

contradicting the assumption that G is an optimal realization of D. □

4. The block decomposition

In this section, we use the relationship between $Cut(G)$ and $Virt^*(D)$ for an optimal realization G of a metric D given in Theorem 8 to prove our main result, namely that every metric has a unique block decomposition (see Theorem 8 below).

We begin by introducing the necessary definitions. A pair of partitions $\mathfrak{P}, \mathfrak{Q}$ of X is called *strongly compatible* if there exist $A \in \mathfrak{P}$ and $B \in \mathfrak{Q}$ such that $A \cup B = X$. In addition, for any metric D on X, we let $\mathfrak{P}_D$ denote the partition of X associated to the equivalence

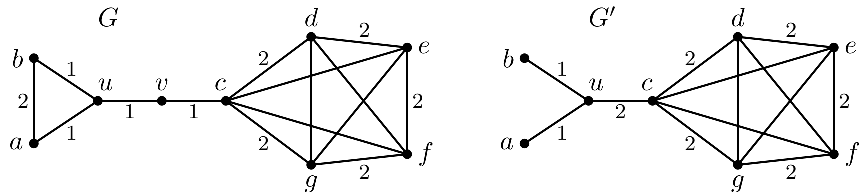


FIGURE 5. G is a weak block realization of the metric D in Fig. 1. All unlabeled edges in this figure have weight 4. For the blocks $B_1 = \{a, b, u\}$ and $B_2 = \{c, d, e, f, g\}$ of G and its associated metrics $D_1 = D_{B_1}$ and $D_2 = D_{B_2}$ we obtain the partitions $\mathfrak{P}_{D_1} = \{\{a\}, \{b\}, \{c, d, e, f, g\}\}$ and $\mathfrak{P}_{D_2} = \{\{a, b, c\}, \{d\}, \{e\}, \{f\}, \{g\}\}$, which are strongly compatible. G' is a block realization of D. In view of $Cut(G') = \{u, c\}$ and $Virt^*(D) = \{f_u, f_c\}$, it follows that $\{D_{B'} : B' \in \mathcal{B}(G')\}$ is the block decomposition of D.

relation on X defined by setting x and y in X to be equivalent if and only if $D(x, y) = 0$. Then a *compatible decomposition* of a metric D on X is a set $\mathcal{D}$ of metrics on X such that $D = \sum_{D' \in \mathcal{D}} D'$ and, for all $D_1, D_2 \in \mathcal{D}$ with $D_1 \neq D_2$, the partitions $\mathfrak{P}_{D_1}, \mathfrak{P}_{D_2}$ are strongly compatible and D_1 and D_2, considered as vectors in $\mathbb{R}^{X \times X}$, are linearly independent. A metric D on X is called a *block metric* if $\{D\}$ is the only compatible decomposition of D and a compatible decomposition $\mathcal{D}$ of a metric D is called a *block decomposition* of D if all metrics in $\mathcal{D}$ are block metrics. Note that $\mathfrak{P}_D = \{X\}$ for a metric D on X if and only if $D(x, y) = 0$ for all $x, y \in X$. Moreover, the metric D on X with $D(x, y) = 0$ for all $x, y \in X$ has $\{D\}$ as its unique compatible decomposition and, for all metrics $D' \neq D$ on X, no compatible decomposition of D' contains D.

As mentioned in the introduction, to prove the main result in this section, we use some facts concerning another type of realization. A realization $G = ((V, E), \omega, \varphi)$ of a metric D is a *weak block realization* of D if (V, E) is a block graph (i.e., a graph in which every maximal 2-connected subgraph is a clique), every vertex $v \in V - \varphi(X)$ is a cut vertex of G and, for all $x, y \in X$, the unique path in (V, E) from $\varphi(x)$ to $\varphi(y)$ with the smallest number of edges is a shortest path in G of length $D_G(\varphi(x), \varphi(y)) = D(x, y)$. A *block realization* is a weak block realization $G = ((V, E), \omega, \varphi)$ in which every vertex $v \in V - \varphi(X)$ has degree at least 3. See Fig. 5 for an example.

It is shown in [8, Theorem 1] that, for every metric D on X, there is a bijective correspondence between (isomorphism classes) of weak block realizations of D and finite subsets of $Virt(D)$. This correspondence works by mapping each cut vertex v in a weak block realization to f_v and, for every $x \in X$, the vertex $\varphi(x)$ to f_x. Therefore, in [8] the correspondence is described using the larger set $Virt(D) \cup \{f_x : x \in X\}$, but it is the subset of $Virt(D)$ that actually determines the isomorphism class of the corresponding weak block realization. Also note that it is shown in [9, Theorem 1] that, for every metric D on X, there is a bijective correspondence between (isomorphism classes) of block realizations of D and compatible decompositions of D. This correspondence is based on associating to each block $B \in \mathcal{B}(G)$ in a block realization $G = ((V, E), \omega, \varphi)$ of D the metric D_B on X

(cf. Section 2.1). The following two lemmas summarize some consequences of these two correspondences.

Lemma 6. *Let D be a metric on X with $|\mathfrak{P}_D| \geq 3$. Then $Virt^*(D) = \emptyset$ if and only if $Virt(D) = \emptyset$.*

Proof: In view of $Virt^*(D) \subseteq Virt(D)$ it suffices to show that $Virt^*(D) = \emptyset$ implies $Virt(D) = \emptyset$. Assume for a contradiction that there exists a metric D on X with $|\mathfrak{P}_D| \geq 3$ and $Virt^*(D) = \emptyset$ but there exists some $f \in Virt(D)$. We consider the finite subset $\{f\}$ of $Virt(D)$. By [8, Theorem 1], $\{f\}$ corresponds to an, up to isomorphism, unique weak block realization $G' = ((V', E'), \omega', \varphi')$ of D. Note that G' has precisely one cut vertex v and this cut vertex satisfies $f_v = f$. For all other vertices u of G' there exists at least one $x \in X$ with $\varphi'(x) = u$ and, therefore, $f_u = f_x$. Since $f \notin Virt^*(D)$ the graph Γ_f is a disjoint union of two cliques and $f_v \neq f_x$ for all $x \in X$, implying that G' consists of precisely two blocks with vertex sets A and B such that $A \cap B = \{v\}$.

Now, in view of $|\mathfrak{P}_D| \geq 3$ we may assume without loss of generality that there exist $a_1, a_2 \in \varphi'^{-1}(A)$ with $D(a_1, a_2) > 0$. Put $\varepsilon = \frac{1}{2} \cdot \min(f(x) + f(y) - D(x, y) : x, y \in \varphi'^{-1}(A))$ and consider the map $g : X \to \mathbb{R}_{\geq 0}$ defined by putting $g(x) = f(x) - \varepsilon$ for all $x \in \varphi'^{-1}(A)$ and $g(x) = f(x) + \varepsilon$ for all $x \in \varphi'^{-1}(B)$. To finish the proof, we show that $g \in Virt^*(D)$, which yields a contradiction to our assumption above that $Virt^*(D) = \emptyset$. First note that $\varepsilon \geq 0$ and consider any $x, y \in X$. Then, if $x, y \in \varphi'^{-1}(A)$, we have

$$g(x) + g(y) = f(x) + f(y) - 2\varepsilon \geq D(x, y).$$

Similarly, if $x, y \in \varphi'^{-1}(B)$, we have

$$g(x) + g(y) = f(x) + f(y) + 2\varepsilon \geq D(x, y)$$

and, if $x \in \varphi'^{-1}(A)$ and $y \in \varphi'^{-1}(B)$, we have

$$g(x) + g(y) = f(x) - \varepsilon + f(y) + \varepsilon = D(x, y).$$

Since $X = \varphi'^{-1}(A) \cup \varphi'^{-1}(B)$, this establishes that $g \in T(D)$. Moreover, in view of $\mathrm{supp}(f) = X$ and the assumption above that there exist $a_1, a_2 \in \varphi'^{-1}(A)$ with $D(a_1, a_2) > 0$ we have $\mathrm{supp}(g) \cap \varphi'^{-1}(A) \neq \emptyset$ and $\mathrm{supp}(g) \cap \varphi'^{-1}(B) \neq \emptyset$. Hence, Γ_g is disconnected. Moreover, in view of the choice of ε, we must have $g = f_x$ for some $x \in \varphi'^{-1}(A)$ or Γ_g is not a disjoint union of two cliques. But this implies $g \in Virt^*(D)$. □

Lemma 7. *Let D be a metric on X. The following are equivalent:*

- (i) *D is a block metric.*
- (ii) *Every block realization of D consists of a single block.*
- (iii) *$Virt^*(D) = \emptyset$.*
- (iv) *For every optimal realization G of D we have $Cut(G) = \emptyset$.*

Proof: First note that (i) ⇔ (ii) is a direct consequence of [9, Theorem 1].

Next, to establish (ii) ⇔ (iii), note that the equivalence clearly holds if $|\mathfrak{P}_D| \in \{1, 2\}$. So suppose $|\mathfrak{P}_D| \geq 3$. Then, to establish (iii) ⇒ (ii), it suffices to note that, by Lemma 6, $Virt^*(D) = \emptyset$ implies $Virt(D) = \emptyset$ and, therefore, by [8, Theorem 1], every weak block realization, and thus every block realization, consists of a single block. Conversely, suppose (ii) holds and assume for a contradiction that there exists some $f \in Virt^*(D)$. Then, by [8, Theorem 1], there exists an, up to isomorphism, unique weak block realization $G' = ((V', E'), \omega', \varphi')$ of D corresponding to the subset $\{f\} \subseteq Virt(D)$, that is, $Cut(G') = \{v\}$ consists of a single cut vertex v with $f_v = f$. In view of (ii) G' cannot be a block realization of D and, thus, v has degree 2, implying that Γ_f is a disjoint union of two cliques and $f_v \neq f_x$ for all $x \in X$. But this contradicts $f \in Virt^*(D)$.

(iii) ⇔ (iv) This is an immediate consequence of Theorem 5. □

We now prove the main result of this section.

Theorem 8. *Let D be a metric on X. Then D has a unique block decomposition $\mathcal{D}$. In particular, $\mathcal{D} = \{D_B : B \in \mathcal{B}(G)\}$ for the (up to isomorphism) unique block realization G of D whose cut vertices are in one-to-one correspondence with $Virt^*(D)$.*

Proof: We first show that D has a block decomposition. To see this, note that, by [22, Theorem 2.2], D has an optimal realization $G = ((V, E), \omega, \varphi)$. From this we construct a block realization $G' = ((V', E'), \omega', \varphi')$ of D by putting $V' = \varphi(X) \cup Cut(G)$, E' consisting of all $\{u, u'\} \in \binom{V'}{2}$ with u and u' contained in the same maximal 2-connected subgraph of (V, E) or $\{u, u'\} \in E$, $\omega'(\{u, u'\}) = D_G(u, u')$ for all $\{u, u'\} \in E'$ and $\varphi'(x) = \varphi(x)$ for all $x \in X$. Then, by construction, $Cut(G') = Cut(G)$ and, thus, by Theorem 5, $Cut(G')$ is in bijective correspondence with $Virt^*(D)$. Since, by [9, Theorem 1], $\{D_B : B \in \mathcal{B}(G')\}$ is a compatible decomposition of D, it suffices to show that D_B is a block metric for every $B \in \mathcal{B}(G')$.

So, consider a block $B \in \mathcal{B}(G')$ and assume for a contradiction that the metric D_B is not a block metric. Note that this implies that B consists of at least three vertices and, thus, B arises from some maximal 2-connected subgraph H of (V, E). Moreover, by Lemma 7(iv), D_B has an optimal realization $G_B = ((V_B, E_B), \omega_B, \varphi_B)$ with $Cut(G_B) \neq \emptyset$. We associate to every vertex v of B the set X_v comprising those $x \in X$ with $D_{G'}(\varphi(x), v) < D_{G'}(\varphi(x), u)$ for all vertices $u \neq v$ of B. Note that it follows from the definition of a block realization that $X_v \neq \emptyset$ for all $v \in B$. Moreover, for every vertex v of B, there must exist a unique $a_v \in V_B$ with $\varphi_B^{-1}(a_v) = X_v$. Thus, we obtain an optimal realization G'' of D by replacing the subgraph H in G with (V_B, E_B) in such a way that, for every vertex v of B, a_v takes the place of v. By Theorem 5, we have $|Cut(G)| = |Cut(G'')| = |Virt^*(D)|$, implying that there exists a vertex v of B with $v \in Cut(G)$ and $a_v \in Cut(G_B)$. Note that the maps f_v obtained from G and f_{a_v} obtained from G'' coincide. By Lemma 2, it follows that we must have $\mathfrak{P}_v = \mathfrak{P}_{f_v} = \mathfrak{P}_{f_{a_v}} = \mathfrak{P}_{a_v}$, where the first and the last of these partitions of X are derived from G and G'', respectively. But, since a_v is a cut vertex of (V_B, E_B) while v is not a cut vertex of H, we have $\mathfrak{P}_v \neq \mathfrak{P}_{a_v}$, a contradiction. Thus, D_B is a block metric.

It remains to show that the block decomposition of D is unique. So, let $\mathcal{D}_1$ and $\mathcal{D}_2$ be two block decompositions of D. By [9, Theorem 1] there exists a (up to isomorphism) unique block realization $G_i = ((V_i, E_i), \omega_i, \varphi_i)$ of D, $i \in \{1, 2\}$, such that $\mathcal{D}_i = \{D_B : B \in \mathcal{B}_i\}$ where $\mathcal{B}_i$ is the set of blocks of G_i. By [22, Theorem 2.2], for every $B \in \mathcal{B}_i$, the metric D_B has an optimal realization G_B. Similarly to the construction of G'' above, we replace every block B in G_i by G_B to obtain a realization G_i'' of D. Since our construction of G_i''

respects the cut vertices already present in the realization G_i of D, it follows from [22, Theorem 5.9] that G_i'' is an optimal realization of D. Moreover, since D_B is a block metric for every $B \in \mathcal{B}_i$, we have $Cut(G_B) = \emptyset$ in view of Lemma 7(iv). But this implies that the sets $Cut(G_i)$ and $Cut(G_i'')$ are in bijective correspondence. Since, by Theorem 5, the sets $Cut(G_1'')$ and $Cut(G_2'')$ are in bijective correspondence with $Virt^*(D)$, we obtain that both $Cut(G_1)$ and $Cut(G_2)$ are in bijective correspondence with $Virt^*(D)$. Therefore, by [8, Theorem 1], G_1 and G_2 are isomorphic weak block realizations of D, implying that $\mathcal{D}_1 = \{D_B : B \in \mathcal{B}_1\} = \{D_B : B \in \mathcal{B}_2\} = \mathcal{D}_2$. □

We conclude this section by pointing out some additional facts related to the block decomposition of a metric. First note that the optimal realization G used in the proof of Theorem 8 was chosen arbitrarily and, thus, we obtain:

Corollary 9. *Let D be a metric and $\mathcal{D}$ its block decomposition. Then, for every optimal realization G of D, we have $\mathcal{D} = \{D_B : B \in \mathcal{B}(G)\}$.*

Next we briefly describe how the block decomposition of a metric D can be computed efficiently. As noted in the introduction, the key is to efficiently compute the set $Virt^*(D)$.

Theorem 10. *Let D be a metric on a set X with $|X| = n$. Then the block decomposition of D can be computed in $O(n^3)$ time.*

Proof: By [12, Theorem 4.1] the set $M = Virt^*(D) \cup \{f_x : x \in X\}$ can be computed in $O(n^3)$ time and, by [12, Lemma 3.1], $|M| \in O(n)$. Moreover, in view of Theorem 8, M is in one-to-one correspondence with the vertex set of a block realization G of D with $Cut(G)$ corresponding to $Virt^*(D)$. To compute the blocks of G, consider the metric D^* on M defined by putting $D^*(f, f') = \max(|f(x) - f'(x)| : x \in X)$ for all $f, f' \in M$. Note that $D^*(f, f')$ coincides with $D_G(u, u')$ for the vertices u and u' in G with $f = f_u$ and $f' = f_{u'}$. Clearly, D^* can be computed in $O(n^3)$ time. Next, we compute for all $g \in M$ the graph $H_g = (V_g, E_g)$ with $V_g = M - \{g\}$ and

$$E_g = \{\{f, f'\} \in \binom{V_g}{2} : D^*(f, f') < D^*(f, g) + D^*(f', g)\}.$$

Note that $h, h' \in M$ correspond to vertices contained in the same block of G if and only if h and h' are contained in the same connected component of H_g for all $g \in M - \{h, h'\}$. Thus, first computing the collection of connected components of H_g for all $g \in M$, each collection in $O(n^2)$ time, the set $\mathcal{B}$ of blocks of G can be computed in $O(n^3)$ time. By Theorem 8, $\mathcal{D} = \{D_B : B \in \mathcal{B}\}$ is the block decomposition of D. It remains to compute the block metrics D_B, which can be done in $O(n^2)$ time for each $B \in \mathcal{B}$. Hence, noting that, in view of $|M| \in O(n)$, we must have $|\mathcal{B}| \in O(n)$, it follows that $\mathcal{D}$ can be computed in $O(n^3)$ time. □

Now, by Corollary 9, for any metric D with block decomposition $\mathcal{D} = \{D_1, D_2, \ldots, D_k\}$, the total length $|G|$ of any optimal realization G of D equals $\sum_{i=1}^{k} |G_i|$, where G_i is an optimal realization of D_i. Thus, in view of Lemma 10, we obtain:

Corollary 11. *If the block decomposition $\mathcal{D} = \{D_1, D_2, \dots, D_k\}$ of a metric D on X with $|X| = n$ is such that an optimal realization of D_i can be computed in $O(t(n))$ time for every $i \in \{1, 2, \dots, k\}$, then an optimal realization of D can be computed in $O(n^3 + k \cdot t(n))$ time.*

5. Some properties of the block decomposition

In this section, we present some further properties of the block decomposition of a metric. We first characterize the block decomposition of a metric in terms of its associated block realization. The proof of this characterization will make use of the following fact.

Lemma 12. *Suppose that $G = ((V, E), \omega, \varphi)$ is a block realization of a metric D on X with $|Cut(G)| \geq 1$. Then $|G| < \sum_{\{u,u'\} \in \binom{\varphi(X)}{2}} D_G(u, u')$.*

Proof: In view of the definition of a block realization, for all $\{u, u'\} \in \binom{\varphi(X)}{2}$, the path $\pi_{\{u,u'\}}$ with the smallest number of edges from u to u' has length $D_G(u, u')$. Let $E_{\{u,u'\}}$ be the set of edges contained in $\pi_{\{u,u'\}}$. It follows immediately from the definition of a weak block realization that every edge $e \in E$ is contained in at least one path $\pi_{\{u,u'\}}$ for some $\{u, u'\} \in \binom{\varphi(X)}{2}$, that is, $\bigcup_{\{u,u'\} \in \binom{\varphi(X)}{2}} E_{\{u,u'\}} = E$. Moreover, in view of $|Cut(G)| \geq 1$ and the fact that in a block realization every vertex $v \in V - \varphi(X)$ has degree at least 3, the union cannot be a disjoint union, and, therefore, $|G| < \sum_{\{u,u'\} \in \binom{\varphi(X)}{2}} D_G(u, u')$, as required. □

We now prove the aforementioned characterization.

Theorem 13. *Let D be a metric on X. Then the block decomposition of D coincides with $\{D_B : B \in \mathcal{B}(G)\}$ for the (up to isomorphism) unique block realization $G = ((V, E), \omega, \varphi)$ of D with $|G|$ minimum.*

Proof: Let $G = ((V, E), \omega, \varphi)$ be a block realization of D such that $|G|$ minimum. It suffices to show that $\mathcal{D} = \{D_B : B \in \mathcal{B}(G)\}$ is the block decomposition of D. Assume for a contradiction that $\mathcal{D}$ is not the block decomposition of D. Then there exists some $B \in \mathcal{B}(G)$ such that D_B is not a block metric and, by Lemma 7, D_B has a block realization G_B with $|Cut(G_B)| \geq 1$. Therefore, by Lemma 12, we have

$$|G_B| < \sum_{\{u,u'\} \in \binom{\varphi(X)}{2}} D_{G_B}(u, u') = \sum_{\{v,v'\} \in \binom{B}{2}} \omega(\{v, v'\}).$$

Thus, replacing the block B in G by the block realization G_B of D_B and suppressing any resulting vertices of degree 2 not labeled by an element in X, we obtain a block realization G' of D with $|G'| < |G|$, a contradiction. □

In view of Corollary 11, it is of interest to better understand the structure of the metrics that appear in the block decomposition of a metric D, ultimately with the aim of exploiting this structure to compute an optimal realization of D efficiently. We now present a description of the metrics occurring in the block decomposition of a so-called totally split decomposable metric D.

We first provide formal definitions of the concepts involved. For every split $S = \{A, B\}$ of X, we denote by D_S the metric on X with $D_S(x, y) = 0$ if $\{x, y\} \subseteq A$ or $\{x, y\} \subseteq B$ and

$S \in \mathcal{S}$	$\alpha(S)$
$S_1 = \{\{a\}, \{b, c, d, e, f, g\}\}$	1
$S_2 = \{\{b\}, \{a, c, d, e, f, g\}\}$	1
$S_3 = \{\{a, b\}, \{c, d, e, f, g\}\}$	2
$S_4 = \{\{a, b, c, d\}, \{e, f, g\}\}$	1
$S_5 = \{\{a, b, c, f, g\}, \{d, e\}\}$	1
$S_6 = \{\{a, b, c, d, g\}, \{e, f\}\}$	1
$S_7 = \{\{a, b, c, d, e\}, \{f, g\}\}$	1
$S_8 = \{\{a, b, c, g\}, \{d, e, f\}\}$	1

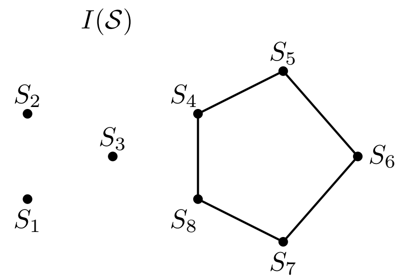


FIGURE 6. The collection $\mathcal{S} = \{S_1, S_2, \dots, S_8\}$ of splits of $X = \{a, b, c, \dots, g\}$ is weakly compatible. Using the weighting α, $D_{(\mathcal{S},\alpha)}$ yields the metric D on X given in Fig. 1, which is, therefore, totally split decomposable. The graph $I(\mathcal{S})$ has four connected components, each corresponding to a metric in the block decomposition of D.

$D_S(x, y) = 1$ otherwise. For a collection $\mathcal{S}$ of splits of X and a weighting map $\alpha : \mathcal{S} \to \mathbb{R}_{>0}$ we define the metric

$$D_{(\mathcal{S},\alpha)} = \sum_{S \in \mathcal{S}} \alpha(S) \cdot D_S.$$

Moreover, two splits $S = \{A, B\}$ and $S' = \{A', B'\}$ of X are called *compatible* if at least one of the intersections

$$A \cap A', \quad A \cap B', \quad B \cap A', \quad B \cap B'$$

is empty, otherwise S and S' are called *incompatible*. Similarly, three splits $\{A_1, B_1\}$, $\{A_2, B_2\}$, $\{A_3, B_3\}$ of X are called *weakly compatible* if at least one of the intersections

$$A_1 \cap A_2 \cap A_3, \quad A_1 \cap B_2 \cap B_3, \quad B_1 \cap A_2 \cap B_3, \quad B_1 \cap B_2 \cap A_3$$

is empty. We will also call a collection $\mathcal{S}$ of splits of X compatible if any two splits in $\mathcal{S}$ are compatible and, similarly, we call $\mathcal{S}$ weakly compatible if any three splits in $\mathcal{S}$ are weakly compatible. A metric D on X is *totally split decomposable* (cf. [2]) if there exists a weakly compatible collection $\mathcal{S}$ of splits of X together with a weighting map α such that $D = D_{(\mathcal{S},\alpha)}$. Note that if such an ordered pair $(\mathcal{S}, \alpha)$ exists for D it is unique and referred to as the *split decomposition* of D. For example, the split decomposition of the metric D in Fig. 2 can be read off from the split-network in that figure by taking the set of splits of the label set induced by the classes of parallel edges, with the weight of the split equal to the length of the edges in the class.

Denoting, for any collection $\mathcal{S}$ of splits of X, by $I(\mathcal{S}) = (V(\mathcal{S}), E(\mathcal{S}))$ the graph with $V(\mathcal{S}) = \mathcal{S}$ and

$$E(\mathcal{S}) = \{\{S, S'\} \in \binom{\mathcal{S}}{2} : S \text{ and } S' \text{ are incompatible}\}$$

and by $\mathcal{C}(\mathcal{S})$ the collection of those subsets of $\mathcal{S}$ that are vertex sets of connected components of $I(\mathcal{S})$ (see e.g. Fig. 6), we now present a description of the block decomposition of a totally split decomposable metric.

Theorem 14. *Let D be a totally split decomposable metric on X and let $\mathcal{S}$ be the associated weakly compatible split system together with the weighting map $\alpha : \mathcal{S} \to \mathbb{R}_{>0}$ such that $D = D_{(\mathcal{S},\alpha)}$. Then the block decomposition of D is*

$$\mathcal{D} = \{\sum_{S \in \mathcal{S}'} \alpha(S) \cdot D_S : \mathcal{S}' \in \mathcal{C}(\mathcal{S})\}.$$

Proof: It is shown in [20] that the collection $\mathcal{C}(\mathcal{S})$ is in bijective correspondence with the collection $\mathcal{B}(F_{\leq 1}(D))$ of vertex sets of maximal 2-connected components and bridges of the realization $F_{\leq 1}(D)$ of D obtained from $T(D)$. In particular, it is shown that, for all $B \in \mathcal{B}(F_{\leq 1}(D))$, there exists a unique $\mathcal{S}' \in \mathcal{C}(\mathcal{S})$ with $D_B = \sum_{S \in \mathcal{S}'} \alpha(S) \cdot D_S$, which, by Theorem 8 and in view of the fact that $Cut(F_{\leq 1}(D)) = Virt^*(D)$, must be a block metric. □

It would be interesting to explore how Theorem 14 can be used to efficiently compute optimal realizations of totally split decomposable metrics. More generally, one could ask whether there are other special types of metrics D that allow for a useful description of the block metrics arising in the block decomposition of D.

In another direction, it could be of interest to understand more deeply how the structure of the tight span of a metric is related to the structure of optimal realizations of the metric. For example, it is shown in [16] that optimal realizations of a totally split decomposable metric D whose tight span is 2-dimensional can always be found in $F_{\leq 1}(D)$, a fact which is used in [17] to help compute optimal realizations of such metrics D. In general, clarifying the precise relationship between optimal realizations of D and $F_{\leq 1}(D)$ remains an important open problem.

Acknowledgment

We would like to thank the two anonymous reviewers and the editor for their helpful comments on an earlier version of this paper.

References

[1] I. Althöfer. On optimal realizations of finite metric spaces by graphs. *Discrete and Computational Geometry*, 3(2):103–122, 1988.
[2] H. J. Bandelt and A. Dress. A canonical decomposition theory for metrics on a finite set. *Advances in Mathematics*, 92(1):47–105, 1992.
[3] H. J. Bandelt and A. Dress. Split decomposition: a new and useful approach to phylogenetic analysis of distance data. *Molecular Phylogenetics and Evolution*, 1(3):242–252, 1992.
[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(3):432–448, 2001.
[5] A. Das, E. Gansner, M. Kaufmann, S. Kobourov, J. Spoerhase, and A. Wolff. Approximating minimum Manhattan networks in higher dimensions. *Algorithmica*, 71(1):36–52, 2015.
[6] A. Dress. Trees, tight extensions of metric spaces, and the cohomological dimension of certain groups: A note on combinatorial properties of metric spaces. *Advances in Mathematics*, 53(3):321–402, 1984.
[7] A. Dress. Towards a classification of transitive group actions on finite metric spaces. *Advances in Mathematics*, 74(2):163–189, 1989.

[8] A. Dress, K. T. Huber, J. Koolen, and V. Moulton. Block realizations of finite metrics and the tight-span construction I: The embedding theorem. *Applied Mathematics Letters*, 21(12):1306–1309, 2008.

[9] A. Dress, K. T. Huber, J. Koolen, and V. Moulton. Compatible decompositions and block realizations of finite metrics. *European Journal of Combinatorics*, 29(7):1617–1633, 2008.

[10] A. Dress, K. T. Huber, J. Koolen, and V. Moulton. Cut points in metric spaces. *Applied Mathematics Letters*, 21(6):545–548, 2008.

[11] A. Dress, K. T. Huber, J. Koolen, V. Moulton, and A. Spillner. A note on metric cut points and bridges, 2007. Available from: `www.researchgate.net/publication/317741392_2007-07-21-cutpoints`.

[12] A. Dress, K. T. Huber, J. Koolen, V. Moulton, and A. Spillner. An algorithm for computing cutpoints in finite metric spaces. *Journal of Classification*, 27(2):158–172, 2010.

[13] A. Dress, V. Moulton, and W. Terhalle. T-theory: An overview. *European Journal of Combinatorics*, 17(2-3):161–175, 1996.

[14] R. Elworth, H. Ogilvie, J. Zhu, and L. Nakhleh. Advances in computational methods for phylogenetic networks in the presence of hybridization. In T. Warnow, editor, *Bioinformatics and Phylogenetics*, pages 317–360. Springer, 2019.

[15] S. Hakimi and S. Yau. Distance matrix of a graph and its realizability. *Quarterly of Applied Mathematics*, 22(4):305–317, 1964.

[16] S. Herrmann, J. Koolen, A. Lesser, V. Moulton, and T. Wu. Optimal realisations of two-dimensional, totally-decomposable metrics. 2011. Available from: `arXiv:1108.0290`.

[17] S. Herrmann, V. Moulton, and A. Spillner. Searching for realizations of finite metric spaces in tight spans. *Discrete Optimization*, 10(4):310–319, 2013.

[18] 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*, 24(2):235–249, 2007.

[19] A. Hertz and S. Varone. The metric cutpoint partition problem. *Journal of Classification*, 25(2):159–175, 2008.

[20] K. T. Huber, J. Koolen, and V. Moulton. The polytopal structure of the tight-span of a totally split-decomposable metric. *Discrete Mathematics*, 342(3):868–878, 2019.

[21] D. Huson and D. Bryant. Application of phylogenetic networks in evolutionary studies. *Molecular biology and evolution*, 23(2):254–267, 2006.

[22] W. Imrich, J. Simoes-Pereira, and C. Zamfirescu. On optimal embeddings of metrics in graphs. *Journal of Combinatorial Theory, Series B*, 36(1):1–15, 1984.

[23] W. Imrich and E. Stockii. The optimal embeddings of metrics into graphs. *Siberian Mathematical Journal (Sibirskiy Matematiki Zhurnal)*, 13:558–565, 1972.

[24] P. Winkler. The complexity of metric realisation. *SIAM Journal of Discrete Mathematics*, 1(4):552–559, 1988.

[25] Z. Zhu, K. Meng, and G. Meng. Genomic recombination events may reveal the evolution of coronavirus and the origin of SARS-CoV-2. *Scientific Reports*, 10(1):1–10, 2020.

School of Computing Sciences, University of East Anglia,, Norwich, NR4 7TJ, UK, E-mail addresses: {k.huber,v.moulton,taoyang.wu}@uea.ac.uk

Merseburg University of Applied Sciences,, 06217 Merseburg, Germany, E-mail address: andreas.spillner@hs-merseburg.de