MATHEMATICS OF OPERATIONS RESEARCH
Vol. 29, No. 2, May 2004, pp. 383–393
ISSN 0364-765X | EISSN 1526-5471 | 04 | 2902 | 0383
DOI 10.1287/moor.1030.0070

On Metric Generators of Graphs

András Sebő and Eric Tannier
Directeur de Recherche, C.N.R.S., Laboratoire Leibniz-IMAG, Département de Mathématiques Discrètes,
46 av. Felix Viallet, 38031 Grenoble Cedex, France {andras.sebo@imag.fr, eric.tannier@imag.fr}

We study generators of metric spaces—sets of points with the property that every point of the space is uniquely determined by the distances from their elements. Such generators put a light on seemingly different kinds of problems in combinatorics that are not directly related to metric spaces. The two applications we present concern combinatorial search: problems on false coins known from the borderline of extremal combinatorics and information theory; and a problem known from combinatorial optimization—connected joins in graphs.

We use results on the detection of false coins to approximate the metric dimension (minimum size of a generator for the metric space defined by the distances) of some particular graphs for which the problem was known and open. In the opposite direction, using metric generators, we show that the existence of connected joins in graphs can be solved in polynomial time, a problem asked in a survey paper of Frank. On the negative side we prove that the minimization of the number of components of a join is NP-hard.

We further explore the metric dimension with some problems. The main problem we are led to is how to extend an isometry given on a metric generator of a metric space.

Key words: metric space; isometric embedding of graphs; join; T-join; combinatorial search; false coins
MSC2000 subject classification: Primary: 05C12, 90C27, 05C70, 05C85, 05C38
OR/MS subject classification: Primary: Networks/graphs
History: Received January 27, 2003; revised March 8, 2003.

Introduction. For any set X, a function $\mu: X \times X \to \mathbb{R}^+$ is called a *metric* if it satisfies three conditions: $\mu(x,x)=0$ for all $x \in X$, $\mu(x,y)=\mu(y,x)$ for all $x, y \in X$, and $\mu(x,y) \leq \mu(x,z)+\mu(z,y)$ for all $x, y, z \in X$. The pair (X, μ) is called a *metric space*.

In this paper graphs are assumed to be connected and undirected. If G is a graph, $V(G)$ denotes the set of its vertices, and $E(G)$ the set of its edges. The *length* of a path P is the number of its edges, and will be denoted by $|P|$. For a pair of vertices x, y, $\mu_G(x,y)$ denotes the length of a shortest (that is, minimum length) path between x and y. $(V(G), \mu_G)$ is clearly a metric space, which we will also simply denote by G.

A vertex $t \in V(G)$ is said to *distinguish* two vertices x and y if $\mu_G(t,x) \neq \mu_G(t,y)$. A set $T \in V(G)$ is said to be a *metric generator* for G if any pair of vertices of G is distinguished by some element of T. A minimum generator is called a (metric) *basis*, and its cardinality the (metric) *dimension* of G, denoted by $\dim(G)$. For a generator T and a vertex $x \in V(G)$, $(\mu_G(t,x))_{t\in T}$ is the vector of *metric coordinates* of x. All these notions can, of course, be defined for any metric space.

For a detailed review about the metric dimension of graphs see Chartrand et al. (2000). Here we only mention that its computation is NP-hard and that it can be computed for trees in polynomial time (see §2), but the algorithmic behavior has not really been explored more generally.

The following §1 is devoted to some simple examples of metric bases: trees, affine spaces, and the hypercube. The metric dimension of the hypercube came up in Chartrand et al. (2000) as an open problem. We will see below that merely recognizing the connection with combinatorial search (of false coins) provides a solution to the problem. We also state a combinatorial optimization problem that refines this well-known search problem.

In §2 we are led to some notions that are more restrictive than metric bases, with the aim of extending isometric embeddings given on subsets of vertices of graphs. The results we prove about these lead to a characterization of the existence and a polynomial algorithm for finding "connected joins" asked by Frank (1996). The consequences of the results for joins are worked out in §3 along with the NP-hardness of the minimization version of the same problem.

1. Three examples. Let us first show here how to find a metric basis for trees, because it will be used for our purposes later. In a tree A, the *leaves* are the vertices of degree 1, and the *inner vertices* are the vertices of degree at least 3. We say that an inner vertex v is *close* to a leaf l if there is no other inner vertex in the unique path between v and l in A. The number of leaves which an inner vertex v is close to is denoted by n_v. The following has been observed in several papers (see for example Chartrand et al. 2000).

THEOREM 1. *If A is a tree but not a path, then*

$$\dim(A) = \sum_{v\in V(G),\, n_v\geq 1} n_v - 1.$$

The metric bases are all constructed in the following way: for any inner vertex v for which $n_v \geq 1$, let $l_1, \ldots, l_k$ be the leaves close to v; for every $i = 1, \ldots, k$ except one, take a vertex distinct from v on the path from v to l_i.

A second simple example will not be directly used in the sequel, but the intuition it provides will be helpful. An affine space is a translated linear space.

The metric bases of an affine space endowed with the Euclidian metric are exactly the maximum affinely independent subsets of the space.

Indeed, we can suppose that the mentioned affine space is $\mathbb{R}^n$. Let $T \subseteq \mathbb{R}^n$. The affine rank is invariant with respect to translations, and the property of being a metric generator also; therefore we can suppose without loss of generality that one of the vectors is 0. We prove then that T is a metric generator if and only if its linear rank is at least n.

If it is less than n, two points symmetric to the subspace generated by T have the same metric coordinates. Conversely, if $x, y \in \mathbb{R}^n$ have the same metric coordinates, then T is a subset of the hyperplane of points that are equidistant from x and y. The statement is proved.

Let us state now the main facts concerning the metric dimension of hypercubes by translating some results on combinatorial search problems. If x and y are elements of $\mathbb{Z}^n$, we call the *Hamming distance* between x and y, $\mu_H(x, y) = \sum_{i=1,\ldots,n} |x_i - y_i|$. We define the hypercube H_n as the graph on $\{0, 1\}^n$ as a vertex set, where two vertices are joined by an edge if and only if their Hamming distance is 1. So the distance in any hypercube H_n is the Hamming distance μ_H.

In Chartrand et al. (2000), the problem of finding the metric dimension of hypercubes is mentioned with no better upper bound than n. As shown below, the *asymptotically exact* length of the best strategy is in fact $2n/\log_2 n$ and this is just an equivalent translation of a problem on false coins. We state here the main results and include some references.

Given n coins some of which weigh 10 grams (the genuine ones) and the others 9 grams (the counterfeit ones), how many tests for the weight of a subset of coins are necessary to detect all the counterfeit ones? In fact, each test of a subset of coins gives the number of counterfeit ones among them, that is, the cardinality of the common part of the tested subset and the set of counterfeit coins. This number can also be interpreted as the scalar product of the 0–1 characteristic vectors of the tested subset and the subset of all counterfeit coins.

We consider the variant of the problem where all the family of tested subsets has to be given in advance, and therefore none of them can depend on the results of the previous tests. Let $f(n)$ be the minimum number of necessary tests. Lindström (1964) has determined the asymptotic value of $f(n)$ to be $2n/\log_2 n$.

More precisely, the problem is to find a *strategy*, that is, a family $s_1, \ldots, s_m \in \{0, 1\}^n$ minimizing m, such that each vector $x \in \{0, 1\}^n$ is uniquely determined by the scalar products $s_i{}^T x$, $i = 1, \ldots, m$. "Uniquely determined" means that for $x \neq x'$ there exists $i = 1, \ldots, m$ with $s_i{}^T x \neq s_i^T x'$. Then $f(n)$ is the minimum of m for a given n.

In order to translate this into the language of metrics, note that for $u, x \in \{0, 1\}^n$, $\mu_H(u, x) = \sum_{i=1}^n (u_i + x_i) - 2u^T x$. Adding the all 1 vector $\underline{1} \in \mathbb{R}^n$ to the tests changes the length of the strategy only by 1; on the other hand, $\mu_H(\underline{1}, x) = n - \sum_{i=1}^{n} x_i$. Knowing this value, the function $\mu_H(u, x)$ and $u^T x$ can be computed from one another, and it follows that the difference of $f(n)$ and the metric dimension of H_n is at most 1. Therefore:

THEOREM 2.
$$\lim_{n\to\infty} \frac{\dim(H_n)\log_2(n)}{n} = 2.$$

This translation works also for deriving bounds for the metric dimension of grid graphs, and of many other graphs, but the translation does not always work for asymptotically close lower and upper bounds. A *grid graph* $H_{n,k}$ is a graph with vertex set $\{0, \ldots, k-3\}^n$ where two vertices are joined if and only if their Hamming distance is 0. With a slight generalization of the above construction, using a result of Lindström (1965) one can show that $\limsup_{n\to\infty}(\dim(H_{n,k})\log_k(n))/n \le 2$, and this can be realized with metric generators whose coordinates take only the values $k-1$ or 0. We think it does not help to be allowed to take arbitrary points of $H_{n,k}$; this upper bound is probably the asymptotically correct value. However, the lower bounds for the problems of false coins do not obviously go through since it is no more true that distances can be computed from scalar products. Yet, since $(nk)^m \ge k^n$ we have $\liminf_{n\to\infty}(\dim(H_{n,k})\log_k(n))/n \ge 1$.

Distances can be computed in this case from "products" of the form $\sum_{i=1}^{n}\min\{x_i, y_i\}$, $(x, y \in \mathbb{R}^n)$; for 0–1 vectors this is just the scalar product, but not in general.

What is the asymptotic length of a strategy with this test-evaluation function?

It is not our goal to survey problems on false coins with modified evaluations; rather, we point at the aforementioned results (Lindström 1964) and two follow-up papers (Lindström 1965, 1966) of the same author; a wider range of problems with similar interpretation is provided by Katona's survey (1971) on combinatorial search. Yet we are interested in the combinatorial optimization aspects of the problem:

Given a family $\mathcal{F}$ of subsets of a finite set as input, is it possible to determine the minimum of $|\mathcal{Q}|$, $\mathcal{Q} \subseteq \mathcal{F}$ so that the Hamming-distances (or the cardinalities of intersections or ...) of $F \in \mathcal{F}$ from each element of $\mathcal{Q}$ uniquely determine F among the elements of $\mathcal{F}$? Equivalently, we are asking for the dimension of the metric space generated by the Hamming distances induced by $\mathcal{F}$.

The metric dimension of the hypercube is a special case of this problem, but as such the size of the problem is measured by 2^n. Lindström's solution provides a randomized algorithm that finds the optimal strategy for the hypercube.

2. Extending isometric embeddings. An *isometry* from a metric space (X, μ_X) into another (Y, μ_Y) is a mapping $f: X \to Y$ such that for all $x, y \in X$, $\mu_X(x, y) = \mu_Y(f(x), f(y))$.

An *extension* of a function defined on a set T is a function $\bar{f}$ defined on a set including T, such that for all $x \in T$, $\bar{f}(x) = f(x)$. If g is an arbitrary function defined on X, then as usual, $g_{|T}$ $(T \subseteq X)$ denotes the *restriction* of g to T, that is, the function defined on T that has the same values as g. If μ is a metric we also note $\mu_{|T}$ the metric μ *restricted to* T, that is, the metric on $T \times T$, with the same values as μ on $T \times T$.

In the following we would like to study how to extend an isometry between a subset T of vertices of a graph H and the vertices of another graph G, to the set of all vertices of H. This formulation is still too poor: If $T = \varnothing$, we get to the problem of isometrically embedding H to G, an NP-hard problem as shown by choosing H to be a clique. (Then the problem specializes to deciding whether G has a clique of size $|V(H)|$.)

We are interested in making clear what properties of T make the problem interesting on one hand and solvable on the other, and solve it for a particular choice of H motivated by an application.

This setting appears to be different from the usual and sounds even exotic to those who are used to embedding metrics into particular metric spaces (defined say by the l_1, l_2

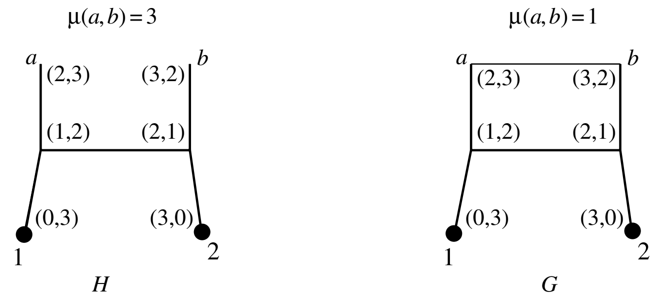

Figure 1. The metric basis $\{1, 2\}$ is not a strong metric generator.

norm, or the hypercube). The vast and interesting subject of metric embeddings is widely overviewed in Deza and Laurent (1991), but we did not find any reference to the problem when the choice of H and that of T are restricted. This may also lead to an interesting theory, and one of the goals of the present work is to exhibit this problem and solve it in a useful case.

The NP-hardness of a general metric embedding shows that our goal cannot be to capture all aspects of this problem. We are motivated by solving the problem when the generator T of H has the following strong property: whenever H is a subgraph of some graph G, and the metric coordinates are the same in H and in G (for all $t \in T$, $x \in V(H)$, $\mu_H(t, x) = \mu_G(t, x)$), then we also have that H is isometrically embedded in G ($\mu_{G|V(H)} = \mu_H$). Although a metric basis distinguishes all the vertices of the graph, it does not determine all the distances. Indeed, in Figure 1, we see two different graphs on the same vertex set with the same metric coordinates. All the metric vectors on different vertices of H (left) are different, so $T = \{1, 2\}$ is indeed a metric basis. But on the right figure the distance of a and b in G is not equal to the distance of a and b in the subgraph H, even though the distances to the bases elements are equal in the two!

We do not know whether isometries given on metric bases of H can be easily extended, but it is clear that our methods—and fortunately at the same time the immediate goal of handling connected joins as well—involve only the more restrictive property that the metric coordinates determine all the distances in G between vertices of H. The following definition will be sufficient for ensuring this.

A *strong metric generator* of a connected graph $G = (V, E)$ is a subset $S \subseteq V$ of its vertices, such that for any pair $x, y \in V$ there is an element $s \in S$, such that there exists a shortest path from s to x that contains y, or a shortest path from s to y that contains x. Equivalently, for all $x, y \in V$ some shortest path between x and y is contained in some shortest path from a vertex of S.

One can immediately see that a strong metric generator is always a metric generator, and Figure 1 shows that the converse is not true. Strong metric generators have a built-in reason that inherently makes sure (and obvious), independently of the supergraph G, that we only have to care about the distance from the basis elements: if these are isometrically embedded in G, then all the distances of H are.

Theorem 3. *In trees, a set containing all the leaves except one is a minimum strong metric generator.*

Indeed, the path between two leaves is not contained in any other shortest path, and therefore every strong metric generator must contain at least one of them. Conversely, for any two vertices of a tree, the unique path joining them is a subpath of a path between two leaves. If at most one of these is missing from a set (for any pair of vertices), then the set is a strong metric generator.

PROBLEM. ISOMETRIC EMBEDDING WITH A STRONG METRIC GENERATOR.

Instance. Two graphs H and G, a strong generator T of H, and an isometry $f: T \to V(G)$.

Question. Is there an isometry from H to G, which is an extension of f?

Is this problem polynomially solvable?

We provide a polynomial algorithm if, in addition, $H - T$ is a forest. This algorithm will be the key step towards the results of the next section.

Proceed as follows: for each vertex u of H, define a set of "candidates" for being the image of u in G: $C_u^{(0)} := \{v \in V(G),\ \mu_G(f(t), v) = \mu_H(t, u) \text{ for all } t \in T\}$. As T is a metric generator of H, the $C_u^{(0)}$ $(u \in V(H))$ are pairwise disjoint. Moreover, using also that $f: T \to V(G)$ is an isometry from H to G, we get that $C_t^{(0)} = \{f(t)\}$ for all $t \in T$.

Denote $C^{(0)} := \bigcup_{u \in V(H)} C_u^{(0)}$.

For $v \in V(G)$ to be the image of $u \in V(H)$, the following two conditions are obviously necessary:

- $v \in C_u^{(0)}$;
- *For all neighbors x of u in H, there exists $y \in C_x^{(0)}$, such that $vy \in E(G)$.*

Indeed, in an appropriate extension of f, the image of u must be in $C_u^{(0)}$, the image of x must be in $C_x^{(0)}$, and it must also be adjacent to $f(u) = v$.

The key observation for the main result of this section is that these two necessary conditions are already sufficient!

Deleting from $C^{(0)}$ those vertices that do not satisfy this condition, the condition may become invalid for some other vertex, and therefore it has to be tested again. We have informally described a recursive procedure that leads to sets that satisfy the condition, or the emptiness of one of them certifies that f cannot be extended. Here is a more formal description:

$$C^{(i+1)} := \{x \in C^{(i)} : \text{if } x \in C_u^{(i)},\ uv \in E(H), \text{ then there exists } y \in C_v^{(i)} \text{ such that } xy \in E(G)\}$$

$$C_u^{(i+1)} := C_u^{(i)} \cap C^{(i+1)}, \qquad (u \in V(H)).$$

We get a decreasing sequence $C^{(0)} \supseteq C^{(1)} \supseteq \cdots$. As $C^{(i)} = C^{(i+1)}$ clearly implies $C^{(i)} = C^{(i+1)} = C^{(i+2)} = \cdots$, and the cardinality of $C^{(i)}$ is not increasing with i, this procedure is stabilized; that is, $C^{(i)}$ reaches its final value after at most $n := |V(G)|$ steps.

Let $C^* := C^{(n)}$, $C_u^* := C_u^{(n)}$.

We can now state the following good ($\text{NP} \cap \text{co NP}$) characterization of the existence of an isometry that extends f, and the proof also implies a polynomial algorithm for testing the isometry. We only need one more definition.

Given two graphs H and G, a strong metric generator T of H and an isometry $f: (T, \mu_{H|T}) \to G$, we say that the sets $\varnothing \neq C_u \subseteq V(G)$ $(u \in V(H))$ *represent* H in G with respect to $T \subseteq V(H)$ if the following conditions are satisfied:

(i) If $t \in T$, $u \in V(H)$, and $x \in C_u$, $\mu_G(f(t), x) = \mu_H(t, u)$;

(ii) If $uv \in E(H)$ and $x \in C_u \subseteq V(G)$, there exists $y \in C_v$, such that $xy \in E(G)$.

Since T is a metric generator in H, condition (i) implies, as before, $C_t = \{f(t)\}$ for all $t \in T$ (using also that f is an isometry from T), and $C_u \cap C_v = \varnothing$ if $u \neq v \in V(H)$.

It will also be helpful to define the distances of sets of vertices: For $X, Y \subseteq V(G)$ we define $\mu_G(X, Y) := \min\{\mu_G(x, y): x \in X, y \in Y\}$; in particular $\mu_G(X, Y) = 0$ if $X \cap Y \neq \varnothing$.

In fact, an isometry is nothing else but a representation where the sets $C_u \subseteq V(G)$ $(u \in V(H))$ have exactly one element. We sketched above how to find a representation or show that there is none. We will see that in the case where $H - T$ is a tree, a representation is sufficient for having an isometry. The following lemma provides an intermediate stage between a representation and an isometry.

LEMMA 1. *If there exist $C_u \neq \varnothing$ $(u \in V(H))$ that represent H in G with respect to T, then for all $u, v \in V(H)$ and $x \in C_u$, $\mu_H(u, v) = \mu_G(x, C_v)$.*

Indeed, let $k := \mu_H(u, v)$. We first prove

$$\mu_G(x, C_v) \le \mu_H(u, v) = k$$

by induction on k. Let $u_0 := u, u_1, \ldots, u_k := v$ be the vertices of a shortest path between u and v in H, $(u_{i-1}u_i \in E(H),\ i = 1, \ldots, k)$.

If $k = 0$, then $\mu_G(x, C_v) = 0 = \mu_H(u, v)$. Otherwise by (ii) there exists $x_1 \in C_{u_1}$ such that $xx_1 \in E(G)$. Since $\mu_H(u_1, v) = k - 1$, we have $\mu_G(x_1, C_v) \le \mu_H(u_1, v) = k - 1$ by the induction hypothesis, and

$$\mu_G(x, C_v) \le \mu_G(x, x_1) + \mu_G(x_1, C_v) \le 1 + k - 1 = k,$$

as claimed.

Now to finish the proof of the lemma we show for all $y \in C_v$, $\mu_G(x, y) \ge \mu_H(u, v)$. Otherwise, the opposite inequality $\mu_G(x, y) < \mu_H(u, v)$ together with $t \in T$ such that a shortest path in H between t and v contains u (this is the definition of strong metric generators after possibly interchanging x and y), and (i) would imply:

$$\mu_G(f(t), y) = \mu_H(t, v) = \mu_H(t, u) + \mu_H(u, v) > \mu_G(f(t), x) + \mu_G(x, y) \ge \mu_G(f(t), y),$$

and this contradiction proves the lemma.

Theorem 4. *Suppose we are given two graphs H and G, a strong metric generator T of H where $H - T$ is a forest, and an isometry $f: (T, \mu_{H|T}) \to G$. Then there exists an isometry from H to G extending f if and only if there exist C_u $(u \in V(H))$ that represent H in G with respect to T.*

Moreover, if the isometry exists, the sets C_u^* $(u \in V(H))$ are the unique inclusionwise maximal sets that represent H in G with respect to T.

Proof. The necessity is obvious, since if the extension $\bar{f}$ exists, then $C_u := \{\bar{f}(u)\}$ $(u \in V(H))$ does represent H with respect to G. Moreover, if C_u $(u \in V(H))$ represent H in G with respect to T, then according to (i) we have $C_u \subseteq C_u^{(0)}$, and clearly, by the recursive definition of $C^{(i)}$ we see that $C_u \subseteq C_u^{(i)}$ implies $C_u \subseteq C_u^{(i+1)}$ $(i = 0, 1, \ldots)$. Therefore, by induction, $C_u \subseteq C_u^*$, as claimed.

To prove the sufficiency, suppose that C_u $(u \in V(H))$ represent H in G with respect to T, and construct the extension $\bar{f}$.

Orient each component of $H - T$ so that each of its vertices is reachable from an arbitrarily chosen root, and denote the obtained directed graph by $\overrightarrow{H - T}$. For the roots r of the obtained arborescences we define $\bar{f}(r) := c_r$, where $c_r \in C_r$ is arbitrary. Now the indegree of all the remaining vertices is 1, and clearly, it is sufficient to define $\bar{f}(v)$ supposing that $\bar{f}(u)$ has already been defined, where u is the tail of the unique arc (u, v) entering v. According to (ii) for $x = \bar{f}(u)$, there exists $y \in C_v$ such that $xy \in E(G)$. Let $\bar{f}(v) := y$.

Now $\bar{f}$ is defined on $V(H)$, and we have to check that $\bar{f}$ is an isometry from H to G. Let $u, v \in V(H)$, and $x = \bar{f}(u)$, $y = \bar{f}(v)$. According to the lemma, $\mu_H(u, v) = \mu_G(x, C_v) \le \mu_G(x, y)$, so it remains only to exhibit a path of length at most $\mu_H(u, v)$ in G between x and y.

Let P be a shortest path in H between u and v, $|P| = \mu_H(u, v)$. Since H is connected, $\mu_H(u, v)$ is finite and P exists.

Case 1. P is entirely in a component of $H - T$.

In this case u and v have a "least common root" in their arborescence, that is, a vertex w so that there exist arc-disjoint directed paths P_u and P_v from w to u and from w to v in $\overrightarrow{H - T}$, and $P = P_u \cup P_v$. Therefore both P_u and P_v are also shortest paths between their endpoints.

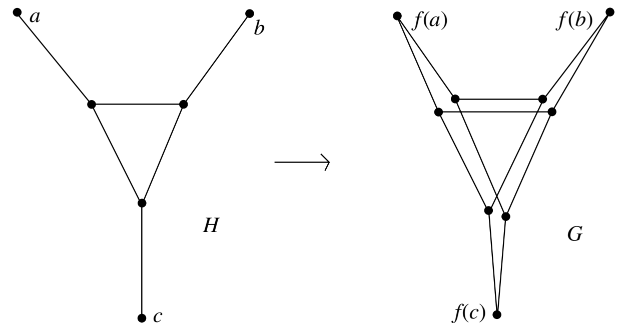

FIGURE 2. Despite a representation of H in G with respect to $\{a, b, c\}$, the given isometric embedding cannot be extended!

By the above construction of $\bar{f}$ we see that the corresponding paths also exist in G: a path Q_u from $\bar{f}(w)$ to $\bar{f}(u) = x$, and a path Q_v from $\bar{f}(w)$ to $\bar{f}(v) = y$ of the same sizes, whence:

$$\mu_H(u, v) = |P| = |P_u| + |P_v| = |Q_u| + |Q_v| \geq \mu_G(x, y).$$

Case 2. There exists $t \in V(P) \cap T$. In this case by (i),

$$\mu_H(u, v) = \mu_H(u, t) + \mu_H(t, v) = \mu_G(x, f(t)) + \mu_G(f(t), y) \geq \mu_G(x, y).$$

Clearly, one of the two cases holds and in either cases we are done. □

COROLLARY 1. *Under the condition of the theorem it can be decided in polynomial time whether there exists an isometry from H to G extending f, and if one exists, it can also be found in polynomial time.*

Indeed, it follows from the theorem that there exists an isometry from H to G extending f if and only if $C_u^* \neq \varnothing$ $(u \in V(H))$. However, it is straightforward to check that the construction of the sets C_u^* requires at most $|V(G)|$ iterations of linear complexity each.

Figure 2 shows that if $H - T$ contains a triangle, the methods do not work anymore.

3. The connected components of a minimum T-join. For basic facts and context related to T-joins, see Frank (1996). If T is an even cardinality subset of vertices of a connected graph $G, F \subseteq E(G)$ is called a *T-join* if for all $v \in V$, $\deg_F(v)$ is odd exactly when v is in T. A cut $\delta(X)$, where $X \subseteq V(G)$ is called a T-cut if $|X \cap T|$ is odd.

It is useful to know and easy to check that *a minimal T-join is the disjoint union of $|T|/2$ paths pairing the vertices of T*. (Indeed, there exists a path between two vertices of T; deleting this path, $|T|$ decreases by 2, and if $|T| > 2$ we can continue, etc.)

In this section we develop our most important application for the isometric embedding problem studied in the previous section: the problem of finding minimum T-joins with a smallest number of connected components, and as a result of finding a relatively large integral packing of T-cuts.

We call $\nu(G, T)$ the maximum cardinality of a family of pairwise disjoint T-cuts, and $\tau(G, T)$ the minimum cardinality of a T-join. It is easy to see that for all G and T,

$$\nu(G, T) \leq \tau(G, T)$$

(a T-cut and a T-join always have an edge in common). The characterization of the equality is the subject of extensive studies in the literature because of its links to integral multiflows. For example, in bipartite graphs, the equality holds for all T (a result of Seymour 1981; see also Frank 1996). Middendorf and Pfeiffer (1998) proved that deciding if the equality holds is an NP-complete problem. However, Korach and Penn (1992) proved that for a minimum cardinality T-join F with k connected components,

$$\tau(G,T) - k + 1 \le \nu(G,T) \le |F|,$$

and therefore, if a minimum T-join is connected, then $\nu(G,T) = \tau(G,T)$.

Minimizing the number of components of a minimum T-join is then a way to guarantee the existence of a packing of T-cuts of a certain size. Formally:

PROBLEM. MINIMUM T-JOIN WITH THE SMALLEST POSSIBLE NUMBER OF COMPONENTS (MTSC).

Instance. A connected graph G, $T \subseteq V(G)$ of even cardinality, and an integer k.

Question. Is there a T-join of minimum cardinality with at most k connected components?

In the following two subsections we prove that:

- this problem is NP-hard in general;
- if we restrict it to instances with $k = 1$, then it is polynomially solvable.

3.1. Minimizing the number of connected components.

THEOREM 5. *MTSC is NP-complete.*

PROOF. MTSC is obviously in NP. We reduce to it the three-dimensional matching problem.

PROBLEM. THREE-DIMENSIONAL MATCHING (3DM).

Instance. Three finite and disjoint sets W, X, and Y of cardinality q, and $H \subseteq W \times X \times Y$.

Question. Is there a *matching*, that is, $M \subseteq H$, so that every $w \in W \cup X \cup Y$ occurs in exactly one element of M?

This problem is known to be NP-complete (Garey and Johnson 1979).

Let W, X, Y, and $H \subseteq W \times X \times Y$ be an instance of 3DM, $q := |W| = |X| = |Y|$. Construct the graph G (see Figure 3).

- Let $V'(G) := W \cup X \cup Y \cup Z \cup H$, where Z is a set disjoint from the others, $|Z| = q$. Let $T := W \cup X \cup Y \cup Z$ ($|T| = 4q$).
- Join every pair of vertices in T by a path of length 2 (the vertex inside the path is a new vertex of the graph). This achieves the definition of $V(G)$.
- Join each vertex $h = (w, x, y) \in H$, to w, x, y and all the points of Z.

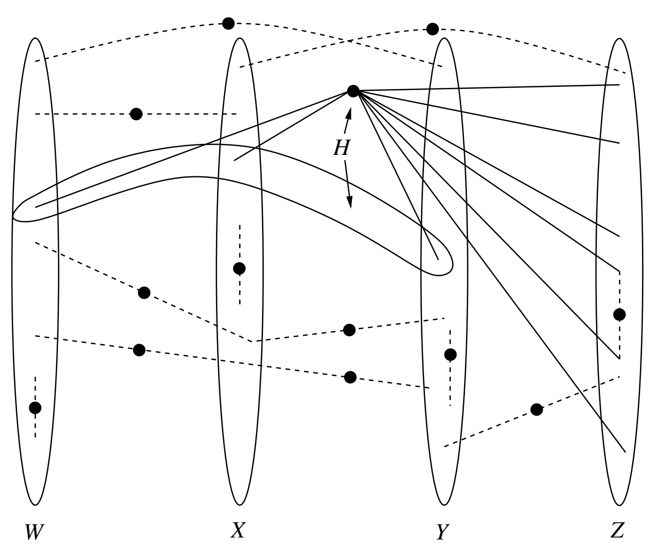

FIGURE 3. The MTSC instance constructed from an instance of 3DM.

Claim. *There exists a minimum T-join with at most q connected components in G if and only if there exists a matching $M \subseteq H$.*

Since T does not induce any edge, every T-join F must contain at least $|T| = 4q$ edges, and the equality holds if and only if F contains exactly one edge incident to each $t \in T$.

If M is a matching, then Z can be indexed with M: $Z = \{z_m : m \in M\}$. Now

$$\bigcup \{\{mw, mx, my, mz_m\} : m = (w, x, y) \in M\}$$

is a minimum T-join with q components.

Conversely, supposing that F is a minimum T-join, and F has at most q components, we will construct a matching.

Partitioning T in $2q$ pairs in an arbitrary way, and taking the union of paths of length 2 between all pairs, we get a T-join of size $4q$; consequently $|F| \leq 4q$. We now get $|F| = 4q$ from the preliminary remark, and also that every $t \in T$ is then incident to exactly one edge th of F. Clearly, $h \notin T$; whence h is either a vertex of degree 2, or $h \in H$. It is true in both cases that h is adjacent to at most one vertex in each of W, X, Y, and together with its neighbors it forms a component of F. It follows that F has at least q components and the equality holds if and only if each component of F contains a vertex in H and exactly one vertex from each of W, X, and Y. □

3.2. Connected minimum T-joins. Here we give a polynomial algorithm that finds a minimum T-join that is also connected, or concludes that such a T-join does not exist, that is, to solve MTSC for $k = 1$. A first proof of these results appeared in Sebő and Tannier (2001) together with some variants and ramifications. Here we concentrate only on the main result which follows almost immediately from the connection to metric bases.

First, we introduce tree metrics. A metric μ on a set X is said to be a *tree metric* if there exists a tree A, and an isometry from (X, μ) to A. The tree A will be called a *realization* of (X, μ). Tree metrics have been characterized, and their polynomial recognition as well as the construction of the tree have been solved in more than one way (e.g., Buneman 1974). We include a construction for the sake of self-containedness, and to provide at the same time a proof of the following claim that will be needed in the sequel:

If μ is a tree metric on a set X, there is a unique (up to isomorphism) inclusionwise minimal tree A which is a realization of μ.

Indeed, if $|X| = 2$ then obviously the unique minimal realization is a path of length $\mu(a, b)$ between the two elements a, b of X. So suppose that μ is a tree metric, $|X| > 2$, and let $a \in X$.

Now supposing that there is a unique minimal tree A_a realizing $\mu_{|X\setminus\{a\}}$, note first that for any tree A realizing μ and for any $x, y, z \in X$, the length of the path joining x and the path between y and z in A is determined by

$$\mu(x, y, z) = \frac{1}{2}(\mu(x, y) + \mu(x, z) - \mu(y, z)),$$

independently of the realization. Since μ is a tree metric, the right-hand side is integer, and we can define $\mu(x, y, z)$ with this formula. Now choose $b, c \in X\setminus\{a\}$ such that $\mu(a, b, c)$ is minimum, and let d be the vertex on the path joining b and c in A_a for which $\mu(a, b, c) = \mu(a, d)$. (Again, if μ is a tree metric, d exists and is unique.) We add now a new path of length $\mu(a, d)$ from a to d. (If $\mu(a, d) = 0$, then $a = d$ is already present in A_a; otherwise it is a new vertex.) The constructed tree is clearly uniquely determined.

This proof is algorithmic, the underlying algorithm is short and can be easily completed. As for any input μ, it either finds the tree or concludes that μ is not a tree metric. (The latter holds if $\mu(x, y, z)$ is not integer for some x, y, z; if d does not exist or is not unique; if the distance function of the uniquely constructed tree is not μ; etc.)

Given two sets X and Y we note their symmetric difference $X \triangle Y := (X \backslash Y) \cup (Y \backslash X)$. The link between T-joins and tree metrics is provided by the following:

LEMMA 2. *A connected T-join F in a graph G is minimum if and only if it is a tree, and for all $x, y \in V(F)$: $\mu_F(x, y) = \mu_G(x, y)$.*

Indeed, let F be a connected T-join.

In order to prove the necessity of the condition, suppose F is a minimum T-join. Then F is a tree, so we only have to show that the path between any two vertices of F is a minimum path in G. If this did not hold, let $a, b \in V(F)$, and let P_1 be an a, b path in F, and P_2 a strictly smaller a, b path in G. Then $C = P_1 \triangle P_2$ is the disjoint union of circuits, and $|C \cap F| \geq |C \cap P_1| > |C \cap P_2| \geq |C \backslash F|$; whence $F \triangle C$ is a smaller T-join, a contradiction.

Conversely, to prove the sufficiency, suppose that F is a tree, and $\mu_F(x, y) = \mu_G(x, y)$ for all $x, y \in V(F)$. Let J be a minimum T-join in G. As noticed before, J is the edge-disjoint union of $m := |T|/2$ paths $P_1, \ldots, P_m$ between pairs of vertices $\{s_1, t_1\}, \ldots, \{s_m, t_m\}$ partitioning T. Since F is a connected T-join, it also contains shortest paths Q_i between s_i and t_i $(i = 1, \ldots, m)$, and therefore

$$|P_1| \geq |Q_1|, |P_2| \geq |Q_2|, \ldots, |P_m| \geq |Q_m|.$$

On the other hand, J is a minimum T-join; whence

$$|P_1| + \cdots + |P_m| = |J| \leq |F| \leq |Q_1| + \cdots + |Q_m|.$$

The last inequality follows from $F = Q_1 \cup \cdots \cup Q_m$. This is true, because $Q_1 \cup \cdots \cup Q_m \subseteq F$ contains a T-join F'; therefore $F \backslash F'$ has only even degree vertices. Since F is a tree, $F' = Q_1 \cup \cdots \cup Q_m = F$ follows.

From the two proven inequalities we see that there is equality throughout, and F is a minimum T-join, as claimed. □

THEOREM 6. *Let G be a connected graph, and let T be a subset of vertices of even cardinality. There exists a connected minimum T-join in G if and only if all of the following conditions hold*:

- *$\mu_{G|T}$ is a tree metric, A denotes its unique minimal realization, and g is the isometry from $(T, \mu_{G|T})$ to A, $T' := g(T)$; $f : T' \to T$ is the inverse of g*;
- *The tree A is a T'-join*;
- *There exists an isometry from A to G extending f.*

PROOF. The necessity of each of the conditions is straightforward to check: the first and third conditions follow from Lemma 2 and the second from the uniqueness of A.

To prove the sufficiency of the conditions, let A be the tree provided by the first condition. Let F be the image of A in G provided by the isometry of the third condition. Then by the second condition, F is a T-join, and since it is a tree, it is connected. Now we see from this construction that the condition of Lemma 2 is satisfied: μ_F arises from $\mu_{G|T}$ by two isometries. Then according to this lemma F is a minimum T-join, as claimed. □

Note the miraculous role of the uniqueness of A in this proof. If the tree A exists but is not a T'-join, then there is no connected T-join!

It is an immediate consequence of this theorem that the existence of connected minimum T-joins can be tested in polynomial time: as it has been said above, it is possible to recognize a tree metric and to construct the tree A of the first condition in polynomial time; the second condition is trivial to check; the third condition is exactly the solution of the isometric embedding problem studied in the previous section: T' is a strong metric generator of A since it contains at least all of its leaves.

References

Buneman, Peter. 1974. A note on the metric properties of trees. *J. Combin. Theory (B)* **17** 48–50.

Chartrand, Gary, Linda Eroh, Mark A. Johnson, Ortrud R. Oellermann. 2000. Resolvability in graphs and the metric dimension of a graph. *Discrete Appl. Math.* **105** 99–113.

Deza, Michel Marie, Monique Laurent. 1991. *Geometry of Cuts and Metrics*. Springer, Berlin, Germany.

Frank, András. 1996. A survey on T-joins, T-cuts, and conservative weightings. *Combinatorics, Paul Erdős is Eighty* **2** 213–252.

Garey, M., D. Johnson. 1979. *Computers and Intractability.* W. H. Freeman, San Francisco, CA.

Katona, G. O. H. 1971. Combinatorial search problems. *Survey Comb. Theory Symp.* Colorado State University, Fort Collins, CO, 1971, 285–308.

Korach, E., M. Penn. 1992. Tight integral duality gap in the Chinese postman problem. *Math. Programming* **55** 183–191.

Lindström, Bernt. 1964. On a combinatory detection problem, I. *A Magyar Tudományos Akadémia Matematikai Kutató Intézetének Közleményei* **9** 195–206.

Lindström, Bernt. 1965. On a combinatorial problem in number theory. *Canadian Math. Bull* **8** 477–490.

Lindström, Bernt. 1966. On a combinatory detection problem, II. *Stud. Sci. Math. Hungaria* **1** 353–361.

Middendorf, M., F. Pfeiffer. 1998. On the complexity of the edge-disjoint path problem. *Combinatorica* **8** 103–116.

Sebő, András, Eric Tannier. 2001. Connected joins in graphs. Aardal, Gerards, eds. *Integer Programming and Combinatorial Optimization, Lecture Notes in Computer Sci.*, No. 2081, 383–395.

Seymour, Paul. 1981. On odd cuts and plane multicommodity flows. *Proc. London Math. Soc.* **42** 178–192.