Connected Joins in Graphs

András Sebő and Eric Tannier

Laboratoire Leibniz-Imag, 46, Avenue Félix Viallet, 38000 Grenoble, France
{Andras.Sebo, Eric.Tannier}@imag.fr

Abstract. A join in a graph is a set F of edges such that for every circuit C, $|C \cap F| \leq |C \setminus F|$. We study the problem of finding a connected join covering a given subset of vertices of the graph, that is a Steiner tree which is a join at the same time. This turns out to contain the question of finding a T-join of minimum cardinality (or weight) which is, in addition, connected. This last problem is mentioned to be open in a survey of Frank [7], and is motivated by its link to integral packings of T-cuts: if a minimum T-join F is connected, then there exists an integral packing of T-cuts of cardinality $|F|$.
The problems we deal with are closely related to some known NP-complete problems: deciding the existence of a connected T-join; finding the minimum cardinality of a connected T-join; the Steiner tree problem; subgraph isomorphism. We also explore some of these connections.

1 Introduction

Graphs $G = (V, E)$ are always undirected, $V = V(G)$ denotes the vertex-set of G and $E = E(G)$ its edge-set. Paths, trees and circuits will be considered as edge-sets.

A set $F \subseteq E(G)$ is said to be a *join* if for every circuit C in $E(G)$, $|C \cap F| \leq |C \setminus F|$. We say that $F \subseteq E(G)$ *covers* $v \in V(G)$, if $\deg_F(v) > 0$ ($\deg_F(v)$ denotes the *degree* of v in F, that is, the number of edges of F incident to v), and covers $S \subseteq V(G)$ if it covers every $v \in S$. We will denote by $V(F)$ the set of vertices covered by F. We will say that F is *connected* if the graph $(V(F), F)$ is connected, and generally F will also denote the subgraph $(V(F), F)$. The main result is a polynomial algorithm an related theorems for the following problem.
Connected join
INSTANCE. A graph G, a subset $S \subseteq V(G)$.
QUESTION. Is there a connected join covering S?

The connected join problem has interesting connections to well-known objects of combinatorial optimization, one of them is T-joins, and a ‘dual’ one is integral packings of T-cuts. For basic facts and context related to T-joins, see Frank [7]. If T is an even cardinality subset of vertices of 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 T-join of minimum cardinality is a join according to a result of Guan [10], [6]. Conversely, a join F is a minimum cardinality T-join, where T is defined as the set of vertices incident to an odd number of edges of F. (Indeed, F is not a join if and only if there is

K. Aardal, B. Gerards (Eds.): IPCO 2001, LNCS 2081, pp. 383–395, 2001.

a circuit C in G such that $|C \cap F| > |C \setminus F|$, that is, if and only if $F \Delta C$ is a T-join of smaller cardinality). A *T-cut* is a set of edges connecting two subsets of $V(G)$ both intersecting T in an odd number of vertices. A consequence of the characterization of connected joins is the solution the problem of deciding the existence of a connected minimum weight T-join. This problem is an (the unique) open problem stated in Frank's survey paper [7].

Another interesting related problem is minimizing the cardinality (or weight) of a Steiner tree. A Steiner tree is a tree covering a given set of vertices; to require that it is a join looks like a natural substitute for minimizing its cardinality. The main result of this paper tells that *in spite of the NP-completeness of Steiner tree minimization, the 'substitute' is polynomially solvable.*

We need to introduce some basic notions about metric spaces, an important tool in our solution. Let X be any finite set. A function $d : X \times X \longrightarrow \mathbb{R}_+$ is a *metric* if it is symmetric ($d(x,y) = d(y,x)$ for all $x, y \in X$), $d(x,y) = 0$ implies $x = y$ for all $x, y \in X$, and it satisfies the triangle inequality : $d(x,y) \leq d(x,z) + d(z,y)$ for all $x, y, z \in X$. (X, d) is a *metric space*. The *restriction* of d to $S \subseteq X$, denoted by $d|_S$, is the metric on $S \times S$ such that $d|_S(x,y) = d(x,y)$ for all x, y in S. The restriction of a function f to a subset S of its domain will also be denoted by $f|_S$.

If X is the set of vertices of a connected graph G, let $d_G(x,y)$ be the minimum number of edges in a path between x and y, $x, y \in V(G)$. It is easy to see that $(V(G), d_G)$ is a metric space. With an abuse of notation, this metric space $(V(G), d_G)$ will simply be denoted G.

An *isometry* from a metric space (X, d_X) into another (Y, d_Y) is a mapping $f : X \to Y$, such that for all $u, v \in X$, $d_X(u,v) = d_Y(f(u), f(v))$. Given a metric space (X, d), we say that a graph H *realizes* d if there is an isometry from (X, d) into H. (An alternative terminology in the literature: (X, d) is 'embeddable' in H). Note that f is an injection and therefore $|X| = |f(X)| \leq |V(H)|$ and the strict inequality may hold. If there is an isometry between two graphs G and H with $|V(G)| = |V(H)|$, then this isometry is in fact a bijection and the two graphs are *isomorphic*.

A metric d is said to be a *tree metric* if there exists a tree realizing it. (This definition slightly differs from the one used in the literature: all the weights in the realization must be equal to 1. This corresponds to 'tree realizable' metrics in terms of [2]. For instance the metric m which is 1 on every pair of a three element set is not a tree metric, but $2m$ is a tree metric in our sense.) For tree metrics, $d(x,y) + d(x,z) - d(y,z)$ is even for all x, y, z in X : this is twice the length of the path joining x to the y, z-path in any tree that realizes d. We note this length $d^x_{y,z} = \frac{1}{2}(d(x,y) + d(x,z) - d(y,z))$.

Metric spaces and embeddability are treated in details in Deza and Laurent's book [5].

The following observation, is not yet a good characterizaton, just an equivalent reformulation of the connected join problem. However, it provides a key to the solution.

Proposition 1. *Let G be a graph, $S \subseteq V(G)$, and F a subset of edges of G covering S. Then F is a connected join if and only if it is a tree, and $d_F = d_G|_{V(F)}$.*

Proof. The necessity is straightforward: a connected join F covering S is indeed a tree, so we only have to show that the path between any two covered vertices is a minimum path in G. Indeed, if not, then 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 $C = P_1 \Delta P_2$ is the disjoint union of circuits, and $|C \cap F| = |C \cap P_1| > |C \cap P_2| = |C \setminus F|$, contradicting that F is a join.

To prove the sufficiency, let F be a tree in G such that $d_G|_{V(F)} = d_F$. We have to prove that F is a join, that is, F is a minimum T-join, where $T = \{v \in V(G), \deg_F(v) \text{ is odd}\}$. Let J be a minimum T-join in G, with $|F \cap J|$ maximum.

Claim: $J \subseteq F$.

Indeed, suppose not: then there exists vertices $a, b \in V(F)$ so that both F and J contain an a, b-path, P and R respectively, so that the only two common vertices of these paths are a and b. (If $V(J) \subseteq V(F)$, then any $R = \{ab\}$, $ab \in J \setminus F$ will do, otherwise consider a vertex $c \in V(J) \setminus V(F)$. Since $c \notin T$, its degree is even, and one can walk in both directions on even vertices of J until reaching vertices a, b of F. In this way R is determined, and P can be defined as the a, b-path of F.)

Now $C = P \cup R$ is a circuit. By the assumption $d_F = d_G|_{V(F)}$, we have $|C \cap F| = |R| \leq |P| = |C \setminus F|$. It follows that $J \Delta C$ is also a minimum T-join and has bigger intersection with F than J, contradicting the definition of F, and finishing the proof of the claim.

So $J \subseteq F$. As F and J are both T-joins, $F \Delta J = F \setminus J$ is the union of disjoint circuits, which is necessarily empty since F is a tree. This implies $J = F$. □

We will treat the condition of Proposition 1 in two parts. In order to test whether there exists a join in G covering $S \subseteq V(G)$, we will first test the weaker condition :

(1) *The restriction $d_G|_S$ of d_G to S is a tree metric.*

In other words, there exists a tree A, and an isometry f_S from $(S, d_G|_S)$ into A. In order to check the condition of Proposition 1, choose A to be *inclusionwise minimal*, that is, all paths in A are subpaths of paths between vertices of $f_S(S)$. It is easy to see that (1) is not sufficient for the existence of a connected join: C_6 with every second vertex in S satisfies (1), and $K_{1,3}$ is the realization but there is no connected join covering S in C_6. This shows that the following condition (2) is essential. It is easy to see that (1) and (2) together are already equivalent to the condition of Proposition 1 (see the proof of Theorem 1).

(2) *There is a subgraph F of G and an isometry f from A into F, with $f(f_S(x)) = x$ for all x in S.*

It is easy and well-known to decide the condition (1) in polynomial time (and, more generally to decide if any metric is a tree metric, see section 2). In section 3, it is shown that (2) can also be tested in polynomial time: we construct the subgraph F whose existence is stated in (2), or certify its nonexistence, supposing that a tree satisfying (1) is already known. Since the tree realizing $d_G|_S$ in (1) *turns out to be unique*, applying the proposition to $S := T$, the solution of Frank's connected T-join problem follows (section 4). Then, in section 5, we show a way of constructing Korach's maximum integral packing of T-cuts whenever a connected minimum T-join exists. Finally, in section 6 the relation of the results to some well-known problems of combinatorial optimization is shown.

2 Constructing the Shape of a Join

In this section we provide a polynomial algorithm that either finds a tree realizing a metric d given on a finite set X, or certifies that such a tree does not exist. This problem has many applications in various fields, such as phylogeny [16], data mining, or psychology [4], and has been solved long ago [1], [2], [11] in a sharper sense; since we make only a very simple use of tree metrics, and for the sake of completeness, we include the simple treatment we worked out for our restricted goals.

Start with a copy $\{f(x), x \in X\}$ of the set X, and construct a path of length $d(a,b)$ between $f(a)$ and $f(b)$ by adding new vertices and edges. This path is a tree realizing $d|_{\{a,b\}}$, and $f|_{\{a,b\}}$ is an isometry.

Suppose now that we have a tree A realizing the restriction of d to some proper subset Y of X, and $f|_Y$ is the corresponding isometry. Let $c \in X \setminus Y$. Choose $a, b \in Y$ so that $d^c_{a,b}$ is minimum, and let d be the vertex of $V(A)$ at distance $d^a_{b,c}$ from $f(a)$ on the $f(a), f(b)$-path. Construct a path P from $f(c)$ to d, of length $d^c_{a,b}$.

Then either $A \cup P$ realizes $d|_{Y\cup\{c\}}$, and then $f|_{Y\cup\{c\}}$ is an isometry or there is no tree realizing $d|_{Y\cup\{c\}}$.

Note that an inclusionwise minimal tree realizing d, whenever it exists, is *uniquely determined* by this procedure. The unicity up to isomorphism is actually straightforward to prove from scratch: vertices of degree 1 of the tree can be characterized with the help of distances, and after deleting them we can continue by induction. A variant of this statement will be important in the sequal:

Proposition 2. *Let G be a graph and $S \subseteq V(G)$. An inclusionwise minimal connected join covering S is a subtree of G that realizes $d_G|_S$ and is uniquely determined up to isomorphisms leaving S fixed.*

The proof is easy, along any of the lines explained above (by the unicity of the construction, or from scratch), we leave it to the reader.

Note that tree metrics have been characterized with a certain 'four-point-condition', that ensures the existence of a realizing weighted tree (Buneman [2]):

$$\text{for all } a,b,c,d \in X, d(a,b) + d(c,d) \leq max \begin{cases} d(a,c) + d(b,d) \\ d(a,d) + d(b,c) \end{cases}$$

However, the realizing tree may have noninteger weights. As mentioned before, we also need the following condition : for all a, b, c in X, $d(a,b) + d(a,c) - d(b,c)$ is an even integer. Together with the four-point-condition, this is sufficient for a metric to be a tree metric, as it can be shown with the simple proof above (this is irrelevant for the main results of this paper).

3 Finding a Connected Join

The previous section allows us to decide whether the condition (1) holds or not. In this section we show how to test condition (2) in polynomial time, that is, we construct a subgraph of G covering S, isomorphic to A.

Suppose that condition (1) holds, and let A be the inclusionwise minimal tree that realizes the restriction of d_G to S. Let f_S be the isometry from $(S, d_G|_S)$ into A.

Let us recursively construct a set P_u^i for every $u \in V(A)$ and $i = 0, \dots, n$, where $n = |V(G)|$. (P_u^i is the set of vertices of G that are not excluded to be in the image of u after step i).

For $u \in f_S(S)$, define $P_u^i := \{f_S^{-1}(u)\}$ for all $i = 0, \dots, n$. Let now $u \in V(A)$ be arbitrary:

- $P_u^0 := \{v \in V(G) : d_G(v, s) = d_A(u, f_S(s))$ for all $s \in S\}$, and

- for some $i \in 0 \dots n-1$, if P_u^i has already been defined for all $u \in V(A)$, then we check for all $v \in P_u^i$, and all $x \in V(A)$ such that $ux \in A$, whether there is an edge from v to some vertex of P_x^i. If not, we define $P_u^{i+1} := P_u^i \setminus \{v\}$, and $P_{u'}^{i+1} := P_{u'}^i$ for all $u' \neq u$.

Clearly, the sets P_u^i ($u \in V(A)$) are pairwise disjoint, and after $k \leq n$ steps, either some P_u^k is empty, or for all $u \in V(A)$ all the vertices of P_u^k are adjacent to some vertex of P_x^k for every $x \in V(A)$ such that $ux \in A$. We then define $P_u := P_u^k$ the *representing set* of $u \in V(A)$ in G. If one of the representing sets is empty, then all are empty, and there is no connected join. If none of the representing sets is empty, then a connected join can be constructed 'greedily', as it is shown in the proof of the following good characterisation theorem for connected joins covering S:

Theorem 1. *Let G be a graph, and $S \subseteq V(G)$. There exists a connected join F covering S if and only if the distances in G between pairs of vertices of S form a tree metric, and the representing sets of the vertices of the realizing tree are nonempty.*

Proof. The necessity is obvious. Conversely, let A be the tree realizing the restriction of d_G to S, P_u the representing set of $u \in A$, and let us construct a tree F which is a subgraph of G covering S isomorphic to A, so that $d_G|_{V(F)} = d_F$. Then by Proposition 1 it follows that F is a connected join.

Suppose that the representing sets are nonempty. We construct the isomorphism f from A into an appropriate subgraph of G by choosing one $f(u) \in P_u$ for all $u \in V(A)$ as follows:

Start with an arbitrary $u \in V(A)$, and let $f(u) \in P_u$ be arbitrary.

- Choose one vertex $f(v)$ from every P_v such that uv is an edge of A; add to F the edges $f(u)f(v)$ for all $uv \in E(A)$. (The property of the representing sets guarantees that these edges exist in G for every $uv \in E(A)$.)

Then we can repeat the same procedure for any vertex $u \in V(A)$ such that $f(u) \in P_u$ has already been defined, but $0 < \deg_F(v_u) < \deg_A(u)$:

- Choose a vertex $f(v) \in P_v$ for all $v \in V(A)$ such that $uv \in E(A)$, and $f(v)$ has not yet been defined; add to F all edges $f(u)f(v)$, $uv \in E(A)$.

Since A is connected, after at most $|V(A)|$ steps, one vertex v_u in every representing set P_u is chosen, and $\deg_F(f(u)) = \deg_A(u)$. (The procedure is similar to usual 'in width label and scan' procedures.)

The result is a tree F, and clearly, f is an isomorphism between A and F. Then for all $x, y \in S$,

$$d_G(x,y) = d_A(f_S(x), f_S(y)) = d_F(f(f_S(x)), f(f_S(y))) = d_F(x,y).$$

Every path in F is a subpath of a path between vertices in S, that is, a subpath of a minimum path in G, hence is a minimum path in G. As a consequence, $d_G(x,y) = d_F(x,y)$ for all $x, y \in V(F)$, as claimed. □

4 Finding a Connected Minimum *T*-Join

The following problem was mentioned to be open by Frank [7] :
Connected Minimum T-join
INSTANCE. A graph G, a subset $T \subseteq V(G)$ of even cardinality.
QUESTION. Is there a T-join of minimum cardinality, which in addition is connected?

As an immediate corollary of Theorem 1, this problem can be solved in polynomial time:

Corollary 1. *A graph $G = (V, E)$, with a subset T of vertices has no connected minimum T-join if and only if it has no connected join covering T, or it has an inclusionwise minimal connected join F which is not a T-join.*

Proof. The necessity of the condition is obvious. To prove the sufficiency, suppose that F is an inclusionwise minimal connected join, which is not a T-join. We have to prove that there is no connected T-join in G at all !

Indeed, by Proposition 1 $d_F = d_G|_{V(F)}$, moreover, since F is inclusionwise minimal, by Proposition 2, F is uniquely determined up to isomorphisms leaving T fixed. Therefore, in any connected inclusionwise minimal join F' (and note that connected T-joins are like this), the degrees in the vertices of T are the same as in F, and the number of vertices of odd degree in $V(F') \setminus T$ is also the same as the number of such vertices in $V(F) \setminus T$. So F' is a T-join if and only if F is a T-join. □

Note that this corollary can also be proved independently of the theorem: the connected minimum T-join problem can be reduced to finding a connected join, and this reduction does not have to rely on the solution of the problem, it is actually much easier than solving either of the two subproblems.

5 Connected T-Joins and Integral Packings of T-Cuts

In this section we develop an important motivation for the connected join problem, which is the problem of finding integral packings 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 commun). The characterization of the equality is the subject of extensive studies in the literature because of its links to integral multiflows. Middendorf and Pfeiffer [15] proved that deciding if equality holds is an NP-complete problem. However, in [12], Korach and Penn proved that for a minimum cardinality T-join F with k connected components, $\nu(G,T) - k + 1 \leq \nu(G,T) \leq |F|$, and therefore, if a minimum T-join is connected, then $\nu(G,T) = \tau(G,T)$.

We will present in this section a simple polynomial algorithm that computes a packing of disjoint T-cuts of cardinality $\tau(G,T)$, or provides a certificate that a connected minimum T-join does not exist.

Such algorithms can be derived from any proof of Korach and Penn's above mentioned theorem [12], [8], [18]. The one we present here in the spirit of the present work is short, elementary, and maybe shows a somewhat simpler way of dealing with vertices of degree 1. The skeleton of the proof is similar to the proof in [17] of $\nu(G,T) = \tau(G,T)$ for bipartite graphs.

First we need one more definition. For a graph G, and $T \subseteq V(G)$, the T-*contraction* of $xy \in E(G)$ is the operation of contracting xy in the graph, and redefining T with $T' := T \setminus \{x,y\} \cup \{v_{xy}\}$ if exactly one of x and y are in T, and $T' := T \setminus \{x,y\}$ otherwise. In this paper we will apply this operation uniquely to the T-contraction of (all edges of) the star of a vertex, that is to identifying a vertex with all its neighbors, (and redefining T depending on the number of vertices of T in the identified set).

Lemma 1 can be viewed as one iteration of a polynomial algorithm finding :
- either a packing of disjoint T-cuts of cardinality $\tau(G,T)$
- or a certificate that a minimum connected join does not exist (when one of the conditions in lemma 1 is violated).

Lemma 1. *Let G be a graph, T a subset of its vertices. Let a and b be elements of T such that $d_G(a,b)$ is the maximum over all distances between pairs of vertices of T. If there exists a connected minimum T-join, then:*

(i) a (and also b) has degree 1 in any connected minimum T-join in G; in particular a and b are in T.
(ii) there exists at least one neighbor a' of a, such that $d_G(a',x) = d_G(a,x) - 1$ for all $x \in T$, and at most one of such neighbors is in T.
(iii) $\delta(a)$ is a T-cut and, if G', T' are obtained by T-contracting $\delta(a)$, then G' has a connected minimum T'-join of cardinality $\tau(G,T) - 1$.

Proof. Let F be a connected minimum T-join, and let P be the a,b-path of F. By Proposition 1 P is also a shortest path in G, and by definition, the longest among all shortest paths between vertices of T.

Proof of (i): If indirectly, F contains an edge incident to a which is not in P, then starting on that edge from a until reaching a vertex c, $deg_F(c) = 1$, the obtained a, c-path $R \subseteq F$ is vertex-disjoint of P, because F is a tree. But then $P \cup R$ is also a path, and is also a shortest path of G, by Proposition 1. Because of $deg_F(c) = 1$ we have $c \in T$, and therefore $P \cup R$ is a path between two vertices of T, and longer than P, a contradiction with the choice of a, b.

Proof of (ii): Let a' be the neighbor of a on P. Again, by Proposition 1, $d_G(a', b) = d_F(a', b) = d_F(a, b) - 1$, and if a had another neighbor in G such that $a'' \in T$, then a'' is also a vertex of F, and $d_G(a, a'') = d_F(a, a'') = 1$, that is $aa'' \in F$, contradicting (i).

Proof of (iii): Let us prove that $F \setminus \{aa'\}$ is a connected minimum T'-join in G'. By (i), it is connected. By Proposition 1 it is sufficient to prove that any path $R \subseteq P$ between two vertices $x, y \in V(F)$, $deg_F(x) = deg_F(y) = 1$, is also a shortest path in G'. Since we know that in R is a shortest path in G, a shorter x, y-path R' of G' must contain the new vertex a^* of G'. But $d_{G'}(x, a^*) \geq d_G(x, a) - 1 = d_G(x, a')$ by (ii), and therefore

$$|R| = d_G(x, y) \leq d_G(x, a') + d_G(a', y) \leq d_{G'}(x, a^*) + d_{G'}(a^*, y) = |R'|,$$

a contradiction. □

Clearly, the sets $\delta(a)$ found by successive application of the lemma are disjoint T-cuts, and each intersects any minimum T-join in one element. Hence, if there exists a connected T-join, we arrive at a packing of T-cuts of cardinality $\tau(G, T)$, as desired. Note that this algorithm neither implies nor is implied by the algorithm developed in the preceding sections for deciding the existence of a connected minimum T-join; indeed, in case a packing of disjoint T-cuts of cardinality $\tau(G, T)$ is found, no connected T-join is exhibited. We do not see how to decide the existence of a connected T-join in this way.

6 Concluding Remarks

6.1 Complexity

The complexity of the algorithm based on our results finds a connected join covering a subset S of vertices of a graph on n vertices (or certifies its nonexistence), in time at most $O(n^3)$.

Indeed, in the first step (constructing A from $d_G|_S$), at each step of a recurrence, we have to compute $d^c_{a,b}$ and $d^a_{c,b}$ for all a, b in $Y \subseteq S$; it takes $O(|Y|^2)$ operations. Then we construct a path (in $O(|Y|)$ time) and check whether the tree realizes $d_G|_Y$, that is we compute $d(c, a)$ for all $a \in Y$. The entire procedure takes $O(\sum_{i=1\ldots n} i^2) = O(n^3)$ time. Note that it is possible to determine if a metric satisfies the four-point-condition (which is equivalent to our first step) in time $O(n^2 log n)$ with the help of ultrametrics as it is proved in [1], that could improve our computation time.

The second step (embedding A in the graph G) has also a $O(n^3)$ time complexity: constructing a P^0_u requires the computation of all the distances between

u and every $x \in S$, then for all $u \in V(A)$, it takes at most $O(n^2)$ operations. Then computing P_u^i from P_u^{i-1} may need $O(nm)$ checks for a single $u \in V(A)$, and this operation is achieved at most n times. In conclusion, the entire process takes at most a time which is a function $O(n^3)$.

6.2 Weighted Case

As we mentioned in the introduction, the results of this paper, including polynomial algorithms, work for the weighted case too. As far as the results are concerned they can be immediately derived by subdivivision of the edges of a weighted graph G, w (subdivide every edge $e \in E(G)$ $w(e)$ times, or contract it if $w(e) = 0$, and apply Theorem 1). But this provides only an algorithm that depends linearly on the weights (instead of being bounded by a polynomial of the logarithm, or not depending on the numbers at all). Anyway it is possible to achieve a strongly polynomial algorithm for the weighted case. The problem can be reformulated as follows (a join is then defined as a set F of edges such that for all circuit C, $w(C \cap F) \leq w(C \setminus F)$):
Weighted Connected Join
INSTANCE. A graph G, an nonnegative integer weight function $w : E(G) \longrightarrow \mathbb{N}$, a subset S of vertices, an integer k.
QUESTION. Is there a connected join F covering S such that $w(F) \leq k$?

In particular, if $|S|$ is even, the question could concern the existence of a S-join of minimum weight which is connected. We reformulate Theorem 1 (and Corollary 1) and the principles of the construction in the following way.

Let the distances between pairs of vertices in G be the minimum weights of paths between the vertices. Then it is possible to restate the two steps of the algorithm.

First, construct a weighted tree realizing a metric d on a set X.

Start with a, b in X. Construct the images of a and b by an isometry f, and an edge between $f(a)$ and $f(b)$ with a weight $w(ab) = d(a, b)$.

Then suppose a weighted tree A, w realizes the restriction of d to $Y \subset X$. Choose $c \in X \setminus Y$ such that $d_{b,c}^a$ is minimum among all a, b in Y. Subdivide the edge e incident to $f(a)$ (it is unique, because of the choice of a), in two edges e_1, e_2, of respective weight $d_{b,c}^a$ and $w(e) - d_{b,c}^a$. Then join $f(c)$ and the new vertex incident to both e_1 and e_2 by an edge of weight $d_{a,b}^c$. Check that the result is a tree realizing $d_{Y \cup \{c\}}$ and start again. Remark that in A, there is no vertex of degree two, and therefore a minimum connected join in G is homeomorphic to A (if it is incusionwise minimal).

Then, construct the representing sets P_u of every vertex $u \in V(A)$: first, $x \in V(G)$ is in P_u if $d_{G,w}(x, s) = d_A(x, s)$ for all $t \in S$. Check for every $x \in P_u$, and $v \in V(A)$ such that uv is an edge of $E(A)$, that there is a path of weight $w(uv)$ between x and some $y \in P_v$. If not, delete x from P_u, and start again. Finally Theorem 1 can be reformulated for the weighted case as follows.

Theorem 2. *Let G be a graph, $w : E(G) \longrightarrow \mathbb{N}$, $S \subseteq V(G)$. There exists a minimum weight connected join F covering S if and only if the distances in G*

with respect to w form a tree metric, and the representing sets of the vertices of the realizing tree are nonempty.

6.3 Related Problems

There are several problems which are closely related to the connected join problem, which have been proved to be NP-complete. We provide some simple NP-completeness proofs and show their relation to our problem.

Connected T-join. This is a NP-complete variant of the connected minimum T-join problem : we don't require the T-join to have minimum cardinality.
Connected T-join
INSTANCE. A graph G, a subset $T \subseteq V(G)$, $|T|$ even.
QUESTION. Is there a connected T-join?

This problem is mentioned by Frank [7], and NP-competeness is said to be solved by a proof of Pulleyblank, with a reduction of the Hamilonian circuit problem, for 3-regular graphs, probably similar to the following :

Proof. Let G be a 3-regular graph. Let us construct G' in the following way : let f be a bijection between $V(G)$ and a new set U of the same cardinality. Let $V(G') = V(G) \cup U$, and $E(G') = E(G) \cup \{xf(x)$, for all $x \in V(G)\}$. Let $T = V(G')$. There exists a Hamiltonian circuit in G if and only if there exists a connected T-join in G'. Indeed, let $F \in E(G')$ be a connected T-join, then all degrees of vertices of $V(G)$ in F are equal to 3 (they are odd, less than 4 because G is 3-regular, and greater than 2 because F is connected). Then $F \cap E(G)$ is a connected subgraph of G such that every vertex has degree 2, in consequence a Hamiltonian circuit. The 'only if' part of the proof follows in the opposite way.

Note that in consequence, it is NP-hard to find a connected T-join, with a minimum number of edges.

Steiner Tree. Let G be a graph and $w : E(G) \longrightarrow \mathbb{N}$ a weight function, and T a subset of $V(G)$. We call a *Steiner tree* for T a tree in G covering T. The *optimal Steiner tree* is a Steiner tree of minimum weight. The following NP-complete problem has been much studied.
Steiner Tree
INSTANCE. A graph G, and $T \subseteq V(G)$, a weight function $w : E(G) \longrightarrow \mathbb{N}$, an integer k.
QUESTION. Is there a tree F in G covering T, such that $w(F) \leq k$?

For NP-completeness proofs, exact algorithms, polynomial heuristics, see [13], [20]. The connection to joins is that a connected join turns out to be an optimal Steiner tree.

Proposition 3. *Let G be a graph, $T \subseteq E(G)$ and $w : E(G) \longrightarrow \mathbb{N}$. If F is a w-minimum connected join covering T, then F is an optimal Steiner tree for the terminal set T.*

Proof. If $|T| = 2$, then the statement is obvious. Suppose now $|T| > 2$. Let A be a Steiner tree covering T, and F a connected join covering T, such that $w(F) > w(A)$. Choose A and F so that $|T|$ is minimum.

For any $t \in T$, call A_t (resp. F_t) the subset of edges of A (resp. F) not belonging to any path of A (resp. F) between two vertices of $T \setminus \{t\}$. For all $t \in T$, $F \setminus F_t$ is a minimum weight connected join covering $T \setminus \{t\}$, and $A \setminus A_t$ is a Steiner tree for $T \setminus \{t\}$. Then $w(F \setminus F_t) \leq w(A \setminus A_t)$, by the minimality of the example. Consequently, $w(F_t) > w(A_t)$ for all $t \in T$.

Now it is easy to see that there exists two vertices t_1, t_2 such that the t_1, t_2-path in A is exactly $A_{t_1} \cup A_{t_2}$ (For example, fix $x \in V(A)$, and let t_1, t_2 maximize $d_A(x, t_1) + d_A(x, t_2) - d_A(t_1, t_2)$). Then

$$d_A(t_1, t_2) = w(A_{t_1}) + w(A_{t_2}) < w(F_{t_1}) + w(F_{t_2}) \leq d_F(t_1, t_2)$$

This contradicts (the trivial part of) Proposition 1. □

Approximations, λ-joins and tree λ-spanners. The links to the Steiner tree problem lead us to the study of a relaxation of the connected join problem. It is possible for example to relax the property of Proposition 1: a λ-*tree* covering S ($\lambda \in \mathbb{R}_+$) is a tree $F \subseteq E(G)$ covering S, with the property that $d_F(x, y) \leq \lambda \times d_G(x, y)$ for all x, y in $V(F)$.

In fact, a λ-tree F is a Steiner tree, which is a most λ times bigger than the optimal Steiner tree. (Indeed, if w is the weight function on the edges of G, define $w'(e) := 1/\lambda w(e)$ for all $e \in F$, and $w'(e) = w(e)$ otherwise. F is a connected join according to w', and therefore $1/\lambda w(F) = w'(F) \leq w'(S) \leq w(S)$, where S is an optimal Steiner tree.)

We could not generalize our arguments to λ-trees, and this is not surprising in the view of the work of Cai and Corneil's on tree λ-spanners [3]: a *tree* λ-*spanner* in a weighted graph G, is a spanning tree F (that means 'covering $V(G)$') such that $d_F(x, y) \leq \lambda \times d_G(x, y)$ for all $x, y \in V(G)$.

Note that a tree λ-spanner is a connected λ-tree covering $S = V(G)$, a special case of our problem. In consequence, as a corollary of our work, it is polynomially solved when $\lambda = 1$ (it was also solved in [3] in a simpler way); but when $\lambda > 1$, it is NP-complete to determine whether a graph contains a tree λ-spanner, and that proves NP-completeness of the λ-tree. Note that tree λ-spanners become tractable for $\lambda \leq 2$ if the graph is unweighted, according to [3].

Isometric Subgraphs. We put the algorithm of Section 3 to the more general context of subgraph isomorphisms, showing also the limits of our method. It is well-known, that it is NP-complete to decide whether a graph G given as input contains a subgraph isomorphic to another given graph H, even if H is a tree [9]. Isomorphisms are exactly the isometries of the distance functions of unweighted graphs: indeed, xy is an edge of a graph, if and only if the distance of x and y in the graph is 1. The problem we have been studying in this paper (and mainly in Section 3) has the specificity, that the distances in the isomorphic subgraph must be the same as in the whole graph.

Subgraph Isometry
INSTANCE. Two graphs G, and A.
QUESTION. Is there an isometry from A into G?

In other words, does G contain a subgraph H isomorphic to A so that the distances in H are the same as in G. We will call H an A-*isometric* subgraph of G.

If A is a clique of size k, then an A-isometric subgraph of G is still a clique of size k: the subgraph isometry problem is in consequence NP-complete. However, the problem is not the same as subgraph isomorphism: indeed, suppose for instance that A is a path of length $|V(G)|$. The problem of deciding whether G contains a subgraph isomorphic to A is the Hamiltonian path problem, whereas the A-isometric subgraph problem is trivial in this case: G has such a subgraph if and only if G *is* a path of length n.

Another specificity of the problems we had to solve here is that the images of some vertices of A are fixed in advance, moreover, that the distances from these prefixed vertices uniquely determine every vertex of the graph. Indeed, the polynomial running time of the algorithm in Section 3 is based on the pairwise disjointness of the representing sets.

We say that $R \subseteq V(A)$ *determines* A if for all $x \in V(A)$, there exists $u, v \in R$, such that $d_A(u,x) \neq d_A(v,x)$.

Determined Subgraph Isometry
INSTANCE: Two graphs G and A, a set $R \subseteq V(A)$ that determines A and an isometry f from $(R, d_A|_R)$ into G.
QUESTION: Is there an extension of f, which is an isometry from A into G?

In other words, is there an isomorphism $\bar{f}$ between A and a subgraph H of G so that $\bar{f}(r) = f(r)$ for all $r \in R$, and $d_H(x,y) = d_G(x,y)$ for all $x, y \in V(H)$.

The algorithm in Section 3 can now be copied to provide a polynomial algorithm for this more general problem. The only fact to notice is that since R determines A, the representing sets are disjoint again.

In other words, the main point of this work expressed in Section 3 is working because the vertices of degree 1 of a tree determine it (and at least all the vertices of degree 1 are fixed in advance). In fact all the vertices which degree is at least 2 and some of the vertices of degree 1 can be deleted from the determining set (one from among each vertex in an equivalence class of vertices having degree one and the same neighbor). It is easy to see that in this way we get all the minimum determining sets of a tree.

While the smallest set that determines a clique has $n-1$ elements, some other interesting classes of graphs are determined by a subset of constant size. For instance paths are determined by only one vertex, and circuits by two. For classes of graphs that can be determined by a constant number of vertices the Subgraph Isometry problem can also be solved in polynomial time, because one can then define R to be a set of constant size that determines A, and one can check the existence of a subgraph isometry for all possible choices for an image of R. (There is a constant number of them.)

This raises the problem of finding the smallest determining set of a graph. Is this problem polynomially solvable or NP-hard? For arbitrary k can the structure of graphs that can be determined by at most k vertices be described? For dense graphs this problem is close to subgraph isomorphism.

Another generalization of the problem studied in this paper is the minimization of the number of components of a minimum T-join (or in a join). We do not know the complexity status of this problem.

References

1. Bandelt Hans-Jürgen, "Recognition of tree metrics", *SIAM Journal of Discrete Math.*, **3** (1990), 1-6.
2. Buneman Peter, "A note on the metric properties of trees", *Journal of Combinatorial Theory* (B), **17** (1974), 48-50.
3. Cai Leizhen and Corneil Derek, "Tree Spanners", *SIAM journal of Discrete Mathematics*, **8** (1995), 359-378.
4. Cunningham James, "Free Trees and Bidirectional Trees as Representations of Psychological distance", *Journal of mathematical psychology*, **17** (1978), 165-188.
5. Deza Michel Marie and Laurent Monique *Geometry of cuts and metrics*, Springer, 1991.
6. Edmonds Jack and Johnson Ellis, "Matching, Euler Tours and the Chinese Postman", *Mathematical Programming*, **5** (1973), 88-124.
7. Frank András, "A Survey on T-joins, T-cuts, and Conservative Weightings", *Combinatorics, Paul Erdős is eighty*, **2** (1996), 213-252.
8. Frank András and Szigeti Zoltán, "On packing T-cuts", *Journal of Combinatorial Theory* (B), **61** (1994), 263-271.
9. Garey Michael and Johnson David, *Computers and intractability, a Guide to the Theory of NP-Completeness*, Freeman, 1979.
10. Guan Mei Gu, "Graphic programming using odd and even points", *Chinese Journal of Mathematics*, **1** (1962), 273-277.
11. Hakimi S.L. and Yau S.S., "Distance matrix of a graph and its realizability", *Quarterly Applied Mathematics*, **22** (1964), 305-317.
12. Korach E. and Penn M., "Tight integral duality gap in the Chinese postman problem", *Mathematical Programming*, **55** (1992), 183-191.
13. Korte Bernhard, Prömel Hans Jürgen and Steger Angelika, "Steiner trees and VLSI-layout", *Paths, Flows and VLSI-layouts*, Korte, Lovász, Prömel, Schrijver, eds, Springer-Verlag, 1980.
14. Lovász László and Plummer M. D., *Matching Theory*, North-Holland, Amsterdam, 1986.
15. Middendorf M. and Pfeiffer F., "On the complexity of the edge-disjoint path problem", *Combinatorica*, **8** (1998), 103-116.
16. Penny David, Foulds and Hendy, "Testing the theory of evolution by comparing phylognenetic trees", *Nature*, **297** (1982), 197-200.
17. Sebő András, "A quick proof of Seymour's Theorem on t-joins", *Discrete Mathematics*, **64** (1987), 101-103.
18. Sebő András, "Undirected distances and the postman-structure of graphs", *Journal of Combinatorial Theory/B*, **49** (1990), No 1.
19. Seymour Paul, "On odd cuts and planar multicommodity flows", *Proc. London Math. Soc.*, **42** (1981), 178-192.
20. Winter Pawel, "Steiner Problem in Networks : A Survey", *Networks*, **17** (1987), 129-167.