STUDIA UNIV. BABEŞ–BOLYAI, INFORMATICA, Volume **XLIX**, Number 2, 2004

A NOTE ON THE COMPLEXITY OF THE GENERALIZED MINIMUM SPANNING TREE PROBLEM

PETRICĂ CLAUDIU POP AND CORINA POP SITAR

Abstract. We consider the Generalized Minimum Spanning Tree Problem denoted by GMST. It is known that the GMST problem is $\mathcal{NP}$-*hard*. We present a stronger result regarding the complexity of the problem, namely, the GMST problem even on trees is $\mathcal{NP}$-*hard*. As well as we present three cases when the GMST problem is solvable in polynomial time.
Keywords: combinatorial optimization, complexity theory, $\mathcal{NP}$-*hard*, generalized minimum spanning tree problem, dynamic programming.

1. Introduction

We are concerned with the generalized version of the minimum spanning tree problem (MST) called the generalized minimum spanning tree problem (GMST). Given an undirected graph whose nodes are partitioned into a number of subsets (clusters), the GMST problem is then to find a minimum-cost tree which includes *exactly* one node from each cluster. Therefore, the MST is a special case of the GMST problem where each cluster consists of exactly one node.

The model fits various problems of determining the location of regional service centers (e.g. public facilities, branches, distribution centers) which should be connected by building links (e.g. highways, communication links). For example, when a company tries to establish marketing centers, one for each market segment, and construct a communication network which interconnects the established centers, the company faces a GMST problem. For another example, when designing metropolitan area networks [7] and regional area networks [15], we are to interconnect a number of local area networks. For this model, we must select a node in each local network as a hub (or a gateway) and connect the hub nodes via transmission links such as optical fibers. Then, such a network design problem reduces to a GMST problem.

The GMST problem has been introduced by Myung, Lee and Tcha in [10]. Since then it appeared several papers studying different aspects of the GMST problems: Feremans, Labbé and Laporte in [5] and Pop in [11] present several integer

Received by the editors: October 5, 2004.
2000 *Mathematics Subject Classification.* 90C05,90C25, 90C34.

formulations of the GMST problem, in [5] Feremans *et al.* study the polytope associated with the GMST problem, in [12] Pop presents approximation results of the problem, etc. The GMST problem was solved to optimality for graphs with up to 200 nodes by Feremans [4] using a branch-and-cut algorithm and by Pop [13] for graphs with up to 240 nodes using a so-called rooting procedure.

A variant of the GMST problem is the problem of finding a minimum cost tree including *at least* one vertex from each cluster. This problem was introduced by Dror *et al.* in [3]. These authors provide also five heuristics including a genetic algorithm [8]. In the present paper we confine ourselves to the problem of choosing exactly one vertex per cluster.

2. Definition of the problem

We begin this section with the definition of the well-known minimum spanning tree problem.

Let $G = (V, E)$ be a connected graph. A spanning tree is a graph where all the nodes in the graph are connected in some way with the requirement that there are no cycles in the graph. If each edge has a weight or cost connected to it, denoting how much you have to pay in order to use the edge, the total sum of the costs of the edges can vary from one edge to another, in the same graph. The one of these possibilities which has the lowest sum is called the minimum spanning tree.

The minimum spanning tree problem can be solved by a polynomial time algorithm, for instance the algorithms of Kruskal [9] or Prim [14]. However as we will show the GMST problem is $\mathcal{NP}$-*hard* [1].

The GMST problem is defined on an undirected graph $G = (V, E)$ with nodes partitioned into m clusters. Let $|V| = n$ and $K = \{1, 2, \dots, m\}$ be the index set of the node sets (clusters). Then, $V = V_1 \cup V_2 \cup \ldots \cup V_m$ and $V_l \cap V_k = \emptyset$ for all $l, k \in K$ such that $l \neq k$. We assume that edges are defined only between nodes belonging to different clusters and each edge $e = \{i, j\} \in E$ has a nonnegative cost denoted by c_{ij} or by $c(i, j)$.

The GMST problem is the problem of finding a minimum-cost tree spanning a subset of nodes which includes exactly one node from each cluster. We will call a tree containing one node from each cluster a generalized spanning tree.

3. Complexity of the GMST problem

Garey and Johnson [6] have shown that for certain combinatorial optimization problems, the simple structure of trees can offer algorithmic advantages for efficiently solving them. Indeed, a number of problems that are $\mathcal{NP}$-*complete*, when are formulated on a general graph, become polynomially solvable when the graph is a tree. Unfortunately, this is not the case for the GMST problem. We will show that on trees the GMST problem is $\mathcal{NP}$-*hard*.

Let us consider the case when the GMST problem is defined on trees, i.e. the graph $G = (V, E)$ is a tree.

To show that the GMST problem on trees is $\mathcal{NP}$-*hard* we introduce the so-called *set cover problem* which is known to be $\mathcal{NP}$-*complete* (see [6]).

Given a finite set $X = \{x_1, ..., x_a\}$, a collection of subsets, $S_1, ..., S_b \subseteq X$ and an integer $k < |X|$, the *set cover problem* is to determine whether there exists a subset $Y \subseteq X$ such that $|Y| \leq k$ and

$$S_c \cap Y \neq \emptyset, \ \forall\, c \text{ with } 1 \leq c \leq b.$$

We call such a set Y a *set cover* for X.

Theorem 1. *The Generalized Minimum Spanning Tree problem on trees is $\mathcal{NP}$-hard.*

Proof: In order to prove that the GMST problem on trees is $\mathcal{NP}$-*hard* it is enough to show that there exists an $\mathcal{NP}$-*complete* problem that can be polynomially reduced to GMST problem.

We consider the set cover problem for a given finite set $X = \{x_1, ..., x_a\}$, a collection of subsets of X, $S_1, ..., S_b \subseteq X$ and an integer $k < |X|$.

We show that we can construct a graph $G = (V, E)$ having a tree structure such that there exists a set cover $Y \subseteq X$, $|Y| \leq k$ if and only if there exists a generalized spanning tree in G of cost at most k.

The constructed graph G contains the following $m = a + b + 1$ clusters $V_1, ..., V_m$:

- V_1 consists of a single node denoted by r
- $V_2, ..., V_{a+1}$ node sets (corresponding to $x_1, x_2, ..., x_a \in X$) each of which has two nodes: one 'expensive' (see the construction of the edges) say $\overline{x_i}$ and one 'non-expensive' say $\hat{x_i}$, for $i = 2, ..., a$, and
- b node sets, $V_{a+2}, ..., V_m$ with $V_\nu = S_{\nu-(a+1)}$, for $\nu = a + 2, ..., m$.

Edges in G are constructed as follows:

(i) Each 'expensive node', say $\overline{x_t}$ of V_t for all $t = 2, ..., a + 1$, is connected with r by an edge of cost 1 and each 'non-expensive' node, say $\hat{x}_t$ of V_t for all $t = 2, ..., a + 1$, is connected with r by an edge of cost 0.

(ii) Choose any node $j \in V_t$ for any $t \in \{a + 2, ..., m\}$. Since $V_t \subset X$, then j coincides with a node in X, say $j = x_l$. We construct an edge between j and (the expensive node) $\overline{x_l} \in V_l$ with $l \in \{2, ..., a\}$. The cost of the edges constructed in this way is 0.

By construction the graph $G = (V, E)$ has a tree structure.

Suppose now that there exists a generalized spanning tree in G of cost at most k then by choosing

$$Y := \{x_l \in X \mid \text{the expensive vertex } \overline{x_l} \in V_{l+1} \text{ corresponding to } x_l \text{ is a vertex of the generalized spanning tree in } G\}$$

we see that Y is a set cover of X.

On the other hand, if there exists a set cover $Y \subseteq X$, $|Y| \leq k$ then according to the construction of G there exists a generalized spanning tree in G of cost at most k. ■

The following theorem due originally to Myung *et al.* [10] is an easy consequence of Theorem 1.

Theorem 2. *The Generalized Minimum Spanning Tree problem is $\mathcal{NP}$-hard.*

Remark 3. To show that the GMST problem is $\mathcal{NP}$-*hard*, Myung, Lee and Tcha [10] used the so-called *node cover problem* which is known that is $\mathcal{NP}$-*complete* (see [6]) and showed that it can be polynomially reduced to GMST problem. Recall that given a graph $G = (V, E)$ and an integer $k < |V|$, the *node cover problem* is to determine whether a graph has a set C of at most k nodes such that all the edges of G are adjacent to at least one node of C. We call such a set C a *node cover* of G.

4. Polynomially solvable cases of the GMST problem

As we have seen the GMST problem is $\mathcal{NP}$-*hard*. In this section we present some cases when the GMST problem can be solved in polynomial time.

A special case in which the GMST problem can be solved in polynomial time is the following:

Remark 4. If $|V_k| = 1$, for all $k = 1, ..., m$ then the GMST problem trivially reduces to the classical Minimum Spanning Tree problem which can be solved in polynomial time, by using for instance the algorithm of Kruskal or the algorithm of Prim.

Another case in which the GMST problem can be solved in polynomial time is given in the following proposition:

Proposition 5. *If the number of clusters m is fixed then the GMST problem can be solved in polynomial time (in the number of nodes n).*

Proof: We present a polynomial time procedure based on dynamic programming which solves the GMST problem in this case.

Let G' be the graph obtained from G after replacing all nodes of a cluster V_i with a supernode representing V_i, that we will call the *global graph*. For convenience, we identify V_i with the supernode representing it. We assume that G' with vertex set $\{V_1, ..., V_m\}$ is complete.

Given a global spanning tree of G', which we shall refer to as the *global spanning tree*, we use dynamic programming in order to find the best (w.r.t. cost minimization) generalized spanning tree.

Fix an arbitrary cluster V_{root} as the root of the global spanning tree and orient all the edges away from vertices of V_{root} according to the global spanning tree. A directed edge $\langle V_k, V_l \rangle$ resulting from the orientation of edges of the global spanning tree defines naturally an orientation $\langle i, j \rangle$ of an edge $(i, j) \in E$ where $i \in V_k$ and $j \in V_l$. Let v be a vertex of cluster V_k for some $1 \leq k \leq m$. All such nodes v are potential candidates to be incident to an edge of the global spanning tree.

The "subtree" rooted at a vertex v, $v \in V_k$ with $k \leq m$, denoted by $T(v)$ includes all the vertices reachable from v under the above orientation of the edges of G, based on the orientation of the edges of the global spanning tree of G'. The *children* of v denoted by $C(v)$ are all those vertices u with a directed edge (v, u). Leaves of the tree have no children.

Let $W(T(v))$ denote the minimum weight of a generalized "subtree" rooted at v. We want to compute:

$$\min_{r \in V_{root}} W(T(r)).$$

We give now the dynamic programming recursion to solve the subproblem $W(T(v))$. The initialization is:

$$W(T(v)) = 0 \text{ if } v \in V_k \text{ and } V_k \text{ is a leaf of the global spanning tree.}$$

The recursion for $v \in V$ an interior vertex is then as follows:

$$W(T(v)) = \sum_{l, C(v) \cap V_l \neq \emptyset} \min_{u \in V_l} \{c(v, u) + W(T(u))\},$$

where by $c(v, u)$ we denoted the cost of the edge (v, u).

For computing $W(T(v))$, i.e. find the optimal solution of the subproblem $W(T(v))$, we need to look at all the vertices from the clusters V_l such that $C(v) \cap V_l \neq \emptyset$. Therefore for fixed v we have to check at most n vertices. So the overall complexity of this dynamic programming algorithm is $O(n^2)$, where $n = |V|$.

Notice that the above procedure leads to an $O(m^{m-2} n^2)$ time exact algorithm for GMST problem, obtained by trying all the global spanning trees, i.e. the possible trees spanning the clusters, where m^{m-2} represents the number of distinct spanning trees of a completely connected undirected graph of m vertices given by Cayley's formula [2]. Therefore when the number of clusters m is fixed the above procedure leads to a polynomial time algorithm for solving the GMST problem. ■

The last case when the GMST problem can be solved in polynomial time is given in the following proposition:

Proposition 6. *Consider the GMST problem on trees. If the number of leaves is bounded then the problem can be solved in polynomial time.*

Proof: Because the number of leaves of the graph $G = (V, E)$ (having a tree structure) is bounded, the number of possible generalized spanning trees, i.e. trees containing exactly one node from each cluster is finite. Therefore the GMST problem in this case is solvable in polynomial time. ■

References

[1] G. Ausiello, Crescenzi, P., Gambosi, G., Kann,V., Marchetti-Spaccamela, A., Protasi, M., *Complexity and Approximation Combinatorial optimization problems and their approximability properties*, Springer Verlag, 1999.

[2] A. Cayley , *On the mathematical theory of isomers*, Philosophical Magazine, 67, 1874, 444.

[3] M. Dror, Haouari, M., Chaouachi, J., *Generalized Spanning Trees*, European Journal of Operational Research, 120, 2000, 583-592.

[4] C. Feremans, *Generalized Spanning Trees and extensions*, Ph.D. thesis, Universite Libré de Bruxelles, Belgium, 2002.

[5] C. Feremans, Labbé, M., Laporte, G., A *Comparative Analysis of Several Formulations for the Generalized Minimum Spanning Tree Problem*, Networks 39(1), 29-34, 2002.

[6] M.R. Garey, M.R., Johnson, D.S., *Computers and Intractability: A Guide to the Theory of NP-Completeness*, Freeman, San Francisco, California, 1979.

[7] M. Gerla, Frata, L., *Tree structured fiber optics MAN's*, IEEE J.Select. Areas Comm. SAC-6, 1988, 934-943.

[8] J.H. Holland, *Adaptation in Natural and Artificial Systems*, University of Michigan Press, Ann Arbor, 1975.

[9] J.H. Kruskal, *On the shortest spanning subtree of a graph and the traveling salesman problem*, Proceedings of the American Mathematical Society 7, 48-50, 1956.

[10] Y.S. Myung, Lee, C.H., Tcha, D.w., *On the Generalized Minimum Spanning Tree Problem. Networks*, Vol 26, 1995 231-241.

[11] P.C. Pop, *The Generalized Minimum Spanning Tree Problem*, Ph.D. thesis, University of Twente, The Netherlands, 2002.

[12] P.C. Pop, Kern, W., Still, G., *An Approximation Algorithm for the Generalized Minimum Spanning Tree Problem with bounded cluster size*, EIDMA 2001 Symposium, Oostende, Belgium, 25-26 October, 2001.

[13] P.C. Pop, Kern, W., Still, G., *A New Relaxation method for the Generalized Minimum Spanning Tree Problem*, to appear in European Journal of Operational Research.

[14] R.C. Prim, *Shortest connection networks and some generalizations*, Bell Systems Technical Journal, 36, 1389-1401, 1957.

[15] J.J. Prisco, *Fiber optic regional area networks in New York and Dallas*, IEEE J. Select. Areas Comm. SAC-4, 1986, 750-757.

Department of Mathematics and Computer Science, Faculty of Sciences, North University of Baia Mare, Romania

E-mail address: pop_petrica@yahoo.com