Faculty of Mathematical Sciences

University of Twente

University for Technical and Social Sciences

P.O. Box 217
7500 AE Enschede
The Netherlands
Phone: +31-53-4893400
Fax: +31-53-4893114
Email: memo@math.utwente.nl

Memorandum No. 1577

An approximation algorithm for the generalized minimum spanning tree problem with bounded cluster size

P.C. Pop, W. Kern and G.J. Still

March 2001

ISSN 0169-2690

An Approximation Algorithm for the Generalized Minimum Spanning Tree Problem with Bounded Cluster Size

P.C. Pop*, **W. Kern**, **G. Still**
Faculty of Mathematical Sciences, University of Twente
P.O. Box 217, 7500 AE Enschede, The Netherlands

March 7, 2001

Abstract

Given a complete undirected graph with the nodes partitioned into m node sets called clusters, the Generalized Minimum Spanning Tree problem denoted by GMST is to find a minimum-cost tree which includes exactly one node from each cluster. It is known that the GMST problem is NP-hard and even finding a near optimal solution is NP-hard. We give an approximation algorithm for the Generalized Minimum Spanning Tree problem in the case when the cluster size is bounded by ρ. In this case, the GMST problem can be approximated to within 2ρ.
Keywords: approximation algorithms, combinatorial optimization, generalized minimum spanning trees, LP relaxation.
Mathematical Subject Classification: 90C11, 90C27, 05C05, 90B10.

1 Introduction

The GMST problem was introduced by Myung et al. [7]. The problem arises in simultaneus selection and sequencing decision, e.g. as in location problems, telecommunications where metropolitan and regional area networks must be interconnected by a tree containing a gateway from each cluster, settings involving agricultural irrigation systems, etc.

The GMST problem is defined on an undirected graph $G = (V, E)$ with nodes partitioned into m clusters. Let $K = \{1, 2, \dots, m\}$ 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 between any two nodes

*Corresponding author. Tel: (00) 31 53 489 3385; fax: (00) 31 53 489 4858; e-mail: p.c.pop@math.utwente.nl

and each edge $\{i, j\} \in E$ has a nonnegative cost c_{ij}. 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.

The following two theorems were proved in [7].

Theorem 1 *The GMSTP is NP-hard.* ■

Even an approximation algorithm for the GMST problem can not be polynomial.

Theorem 2 *Let H be a polynomial-time heuristic for the GMST problem. Assume $P \neq NP$. Then no value $L \lneqq \infty$ can exist such that*

$$\frac{Z_H(I)}{Z(I)} \leq L$$

for every instance I where $Z(I)$ and $Z_H(I)$ are the values of an optimal solution and of the solution found by H, respectively. ■

However if the size of the clusters is bounded,

$$|V_k| \leq \rho \text{ ,} \qquad \text{for all k=1,...m} \qquad \text{(a)}$$

then a polynomial approximation algorithm is possible.

In this paper under condition (a) and the assumption that the costs c_{ij} are nonnegative and satisfy the triangle inequality an approximation algorithm for the GMST problem with performance ratio 2ρ is given. The approximation algorithm is constructed following the ideas in [9] where the Generalized Traveling Salesman Problem and Group Steiner problem have been treated.

2 Integer program and LP relaxation of GMST problem

We formulate the GMST problem as an integer programming problem. We define for each edge $\{i, j\}$ and each node i the binary variables:

$$x_{ij} = \begin{cases} 1 & \text{if edge } \{i,j\} \text{ is included in the selected subgraph} \\ 0 & \text{otherwise} \end{cases}$$

$$y_i = \begin{cases} 1 & \text{if node } i \text{ is included in the selected subgraph} \\ 0 & \text{otherwise} \end{cases}$$

The GMST problem can be formulated as the following integer programming problem:

Problem IP1:

$$\begin{array}{llll} Z_1 = \text{minimize} & \sum_{e \in E} c_e x_e & & \\ \text{subject to} & y(V_k) = 1, & \text{for all } \ k \in K & (1) \\ & x(\delta(S)) \geq y_i, & \text{for all } i \in S \subset V & (2) \\ & x(E) = m - 1 & & (3) \\ & x_e \in \{0, 1\}, & \text{for all } e \in E & (4) \\ & y_i \in \{0, 1\}, & \text{for all } i \in V & (5) \end{array}$$

We use here the standard shorthand notations: for every subset S of V $E(S) = \{(i, j) \in E | i, j \in S\}$, $x(E(S)) = \sum_{e \in E(S)} x_e$, $y(S) = \sum_{j \in S} y_j$ and as usual the cutset $\delta(S)$ is defined by

$$\delta(S) = \{\{i, j\} \in E \mid i \in S \text{ and } j \notin S\}.$$

Condition (1) guarantees that a feasible solution contains exactly one vertex from every cluster. Condition (2) guarantees that any feasible solution is a connected subgraph. Condition (3) simply assures that any feasible solution has m-1 edges and due to the fact that the cost function is non-negative this constraint is redundant.

Consider now the LP relaxation of the integer programming formulation of the GMST problem. In order to do that, we simply replace conditions (4) and (5) in IP1 by new conditions $(4^{'})$ and $(5^{'})$, where

$$\begin{array}{lll} (4^{'}) & 0 \leq x_e \leq 1, & \text{for all } e \in E \\ (5^{'}) & 0 \leq y_i \leq 1, & \text{for all } i \in V \end{array}$$

3 Approximation algorithm

First we assume that the graph has bounded cluster size, i.e. $|V_k| \leq \rho$ for all j=1,...,m and the cost function $c_e = c_{\{i,j\}}$ is symmetric and satisfies the triangle inequality, i.e. $c_{ij} + c_{jk} \geq c_{ik}$ for all $i, j, k \in V$.

The algorithm for approximating the optimal solution of the GMST problem is as follows:

Algorithm "Approximate the GMST problem"

Input: A complete graph G=(V,E) with non-negative symmetric cost function on the edges satisfying the triangle inequality, and clusters $V_1, ..., V_m$ such that $V = V_1 \cup V_2 \cup \ldots \cup V_m$ and with bounded size.

Output: A tree $T \subset G$ spanning some vertices $W^{'} \subset V$ which includes exactly one vertex from every cluster, that approximates the optimal solution to the GMST problem.

1. Solve the LP relaxation of the problem IP1 and let $(y^*, x^*, Z_1^*) = ((y_i^*)_{i=1}^n, (x_e^*)_{e \in E}, Z_1^*)$ be the optimal solution.

2. Set $W^* = \left\{ i \in V | y_i^* \geq \frac{1}{\rho} \right\}$ and consider $W^{'} \subset W^*$ with the property that $W^{'}$ has exactly one vertex from each cluster, and find a minimum spanning tree $T \subset G$ on the subgraph $G^{'}$ generated by $W^{'}$.

3. Output APP = length(T) and the tree T.

Even though the LP relaxation of the problem IP1 has exponentially many constraints, it can still be solved in polynomial time either using ellipsoid method with a min-cut max-flow oracle [4] or using Karmakar's algorithm [5] since the LP relaxation can be formulated compactly using flow variables [7, 8].

4 Auxiliary results

In order to stablish upper bounds on the performance ratio of the above algorithm, we now present some auxiliary results.

The parsimonius property proven by Goemans and Bertsimas [3] can be stated as follows: considering a complete undirected graph G=(V,E), for any pair (i,j) of vertices, let r_{ij} be the connectivity requirement between i and j (r_{ij} is assumed to be symmetric,i.e. $r_{ij} = r_{ji}$). Consider now the following two integer programs:

$$\begin{array}{llll} IZ_\phi(r) = \min & \sum_{e \in E} c_e x_e & & \\ & \text{s.t.} & x(\delta(S)) \geq max_{(i,j) \in \delta(S)} r_{ij}, & \text{for all } S \subset V,\ S \neq \phi \\ & & 0 \leq x_e, & \text{for all } e \in E \\ & & x_e \text{ integral}, & \text{for all } e \in E \end{array}$$

Let denote $Z_\phi(r)$ the optimal value of the LP relaxation (obtained by dropping the integrality restrictions). Clearly $Z_\phi(r)$ is a lower bound on $IZ_\phi(r)$.

$$
\begin{array}{rll}
IZ_D(r) = \min & \sum_{e\in E} c_e x_e & \\
\text{s.t.} & x(\delta(S)) \geq max_{(i,j)\in\delta(S)} r_{ij}, & \text{for all } S \subset V,\ S \neq \phi \\
& x(\delta(i)) = max_{j\in V\setminus\{i\}} r_{ij}, & i \in D,\ D \subseteq V \\
& 0 \leq x_e, & \text{for all } e \in E \\
& x_e \text{ integral}, & \text{for all } e \in E
\end{array}
$$

Let $Z_D(r)$ the optimal value of the LP relaxation.

Theorem 3 *(parsimonious property) If the costs c_{ij} satisfy the triangle inequality, then*

$$Z_\phi(r) = Z_D(r)$$

for all subsets $D \subseteq V$. ■

The proof of this theorem is based on a result due to Lovasz [6] on connectivity properties of Eulerian multigraphs.

Let $W \subset V$. Consider the following linear program.

Problem LP2:

$$
\begin{array}{rlll}
Z_2^*(W) = \min & \sum_{e\in E} c_e x_e & & \\
\text{s.t.} & x(\delta(S)) \geq 1, & \text{for all } S \subset V \text{ s.t. } W \cap S \neq \phi \neq W \setminus S & (6) \\
& \sum_{e\in\delta(i)} x_e = 0, & \text{for all } i \in V \setminus W & (7) \\
& 0 \leq x_e \leq 1, & \text{for all } e \in E & (8)
\end{array}
$$

Let us consider the following relaxation of problem LP2.

Problem LP3:

$$
\begin{array}{rlll}
Z_3^*(W) = \min & \sum_{e\in E} c_e x_e & & \\
\text{s.t.} & x(\delta(S)) \geq 1, & \text{for all } S \subset V \text{ s.t. } W \cap S \neq \phi \neq W \setminus S & (6) \\
& 0 \leq x_e, & \text{for all } e \in E & (9)
\end{array}
$$

Thus we omitted constraint (7) and relaxed constraint (8).

The following result is a straightforward consequence of the parsimonius property, if we choose r_{ij}=1 if $i, j \in S$ and 0 otherwise, and D=$V \setminus W$.

Lemma 4 *The optimal solution values to problems LP2 and LP3 are the same, that is*

$$Z_2^*(W) = Z_3^*(W).$$

■

Consider the following problem:

Problem IP4:

$$
\begin{array}{rll}
Z_4 = \text{minimize} & \sum_{e\in E} c_e x_e & \\
\text{subject to} & x(\delta(S)) \geq 1, \quad \text{for all } S \subset V,\ \phi \neq S \neq V & (10) \\
& x_e \in \{0,1\}, \quad \text{for all } e \in E & (11)
\end{array}
$$

Clearly, it is the integer programming formulation of the MST (minimum spanning tree) problem. Let LP4 be the LP relaxation of this formulation, that is we simply replace the constraint (11) by a new constraint $(11^{'})$, where

$(11^{'})$ $\quad 0 \leq x_e \leq 1, \quad$ for all $e \in E$

Denote by Z_4^* the value of the optimal solution of the LP4.

Proposition 5

$$L^T(V) \leq (2 - \frac{2}{|V|})Z_4^*.$$

Proof. Let S be a proper nonempty subset of V. Denote by $\overline{S}$ the complement of S in V that is $\overline{S} = V \setminus S$. Constraints (10) imply that for all $i \in V$

$$1 \leq \sum_{e\in\delta(i)} x_e$$

and summing over all $i \in \overline{S}$, we get

$$|\overline{S}| \leq \sum_{i\in\overline{S}} \sum_{e\in\delta(i)} x_e = 2 \sum_{e\in E(\overline{S})} x_e + \sum_{e\in\delta(S)} x_e$$

Adding constraint (10) gives

$$|\overline{S}| + 1 \leq \sum_{i\in\overline{S}} \sum_{e\in\delta(i)} x_e + \sum_{e\in\delta(S)} x_e = 2 \sum_{e\in E(\overline{S})} x_e + 2 \sum_{e\in\delta(S)} x_e$$

Now, it easy to see that

$$|\overline{S}| + 1 = \frac{|\overline{S}| + 1}{|\overline{S}|}|\overline{S}| \geq \frac{|V|}{|V| - 1}|\overline{S}| = \frac{|V|}{|V| - 1}(|V| - |S|).$$

Therefore the problem:

$$\begin{array}{lll} \min & \sum_{e\in E} c_e x_e & \\ \text{s.t.} & \sum_{e\notin E(S)} x_e \geq \frac{|V|}{2(|V|-1)}(|V|-|S|), \text{ for all } S\subset V,\ \phi\neq S\neq V & (12) \\ & x_e \geq 0, \qquad\qquad \text{for all } e\in E & (13) \end{array}$$

is a valid relaxation of the LP4 problem. Combining the results of Fulkerson [2], Edmonds [1] and Tutte [10], we see that the extremal points of the polyhedra determined by constraints (12) and (13) are $\frac{|V|}{2(|V|-1)}$ multiples of spanning trees.

Therefore, the length of the minimum spanning tree on V is no longer than $(2-\frac{2}{|V|})Z_4^*$. ■

Let $W \subset V$. We can easily modify Proposition 5 to obtain:

Proposition 6

$$L^T(W) \leq (2-\frac{2}{|W|})Z_2^*(W).$$

Proof. Since for any feasible solution to LP2, $e \notin E(W)$ implies $x_e = 0$, we can use Proposition 5 to prove the inequality. ■

5 Performance Bounds

Let $(y^*, x^*, Z_1^*) = ((y_i^*)_{i=1}^n, (x_e^*)_{e\in E}, Z_1^*)$ be the optimal solution to the LP relaxation fo the GMST problem. Define

$$\widehat{x}_e = \rho x_e^*$$

$$\widehat{y}_i = \begin{cases} 1 & \text{if } y_i^* \geq \frac{1}{\rho} \\ 0 & \text{otherwise} \end{cases}$$

$W^* = \left\{ i \in V | y_i^* \geq \frac{1}{\rho} \right\} = \{ i \in V | \widehat{y}_i = 1 \}$. Because we need only one vertex from every cluster we delete extra vertices from W^* and consider $W^{'} \subset W^*$ such that $|W^{'}|$ = m and $W^{'}$ consists of exactly one vertex from every cluster.

Since LP1 is the LP relaxation of the problem IP1, we have

$$Z_1^* \leq Z_1$$

Now let us show that $(\widehat{x}_e)_{e\in E}$ is a feasible solution to LP3. Indeed, $\widehat{x}_e \geq 0$ for all $e \in E$, hence condition (9) is satisfied. Let $S \subset V$ be such that $W^{'} \cap S \neq$

$\phi \neq W^{'} \setminus S$ and choose some $i \in W^{'} \cap S$. Hence $\widehat{y}_i = 1$ and $y_i^* \geq \frac{1}{\rho}$. Then we have

$$\sum_{e\in\delta(S)} \widehat{x}_e = \rho \sum_{e\in\delta(S)} x_e^* \geq \rho y_i^* \geq \rho\frac{1}{\rho} = 1$$

by definition of $\widehat{x}_e$ and the fact that the x_e^* are solution to LP1. Hence the $\widehat{x}_e$ satisfy constraint (6) in LP3.

Therefore,

$$\begin{aligned} APP &= L^T(W^{'}) \leq (2 - \frac{2}{|W'|})Z_2^*(W^{'}) = (2 - \frac{2}{|W'|})Z_3^*(W^{'}) \\ &\leq (2 - \frac{2}{|W'|})\sum_{e\in E} c_e\widehat{x}_e = (2 - \frac{2}{|W'|})\rho \sum_{e\in E} c_e x_e^* = (2 - \frac{2}{|W'|})\rho Z_1^* \\ &\leq (2 - \frac{2}{|W'|})\rho Z_1 = (2 - \frac{2}{|W'|})\rho OPT. \end{aligned}$$

And since $W^{'} \subset V$, that is m $= |W^{'}| \leq |V| =$ n, we have proved the following.

Theorem 7 *The performance ratio of the algorithm "Approximate GMST problem" for approximating the optimum solution to the GMST problem satisfies:*

$$\frac{APP}{OPT} \leq (2 - \frac{2}{n})\rho.$$

■

Note: One can easily generalize the algorithm and its analysis to the case when, in addition to distances between edges, there is a cost associated with each vertex. In fact, we can show that, in this case,

$$\begin{aligned} APP &= L_{APP} + C_{APP} \leq \rho(2 - \frac{2}{n})L_{OPT} + \rho C_{OPT} \\ &\leq \rho(2 - \frac{2}{n})(L_{OPT} + C_{OPT}) \leq \rho(2 - \frac{2}{n})OPT. \end{aligned}$$

References

[1] J. Edmonds, Minimum partition of a matroid into independent sets. Journal of Research of the National Bureau of Standards B, 69B: 67-72, 1965.

[2] D.R. Fulkerson, Blocking and anti-bloking polyhedra. Mathematical Programming, 1(2): 168-194, Nov. 1971.

[3] M.X. Goemans and D.J. Bertsimas, Surviveble networks, linear programming relaxations and the parsimonious property. Mathematical Programming, 60: 145-166, 1993.

[4] R.E. Gomory and T.C. Hu, Synthesis of a communication network. SIAM Journal on Applied Mathematics 9 (1961) 551-570.

[5] N.K. Karmarkar, A new polynomial-time algorithm for linear programming. Combinatorica, 4: 373-395, 1984.

[6] L. Lovasz, On some connectivity properties of Eulerian graphs. Acta Mathematica Academiae Scientiarum Hungaricae 28 (1976) 129-138.

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

[8] P.C. Pop, U. Faigle, W. Kern, G. Still, Relaxation Methods for the Generalized Minimum Spanning Tree Problem, submitted.

[9] P. Slavik, On the Approximation of the Generalized Traveling Salesman Problem, submitted.

[10] W.T. Tutte, On the problem of decomposing a graph into n connected factors. Journal of London Mathematical Society, 36: 221-230, 1961.