

The Minimum Representation Spanning Tree Problem

Elio David Zaldivar Linares ^{a,1}, Luiz Satoru Ochi ^a,
Thiago Gouveia Da Silva ^{a,b} and Ruben Interian ^a

^a *Instituto de Computação, Universidade Federal Fluminense, Niterói-RJ, Brazil*

^b *Instituto Federal de Educação Ciência e Tecnologia da Paraíba, João Pessoa-PB, Brazil*

Abstract

This work presents the *Minimum Representation Spanning Tree Problem* (MRSTP) and some of its particular features. We prove the MRSTP is NP-Hard, propose a mathematical formulation, an effective constructive heuristic and a GRASP based algorithm for the problem. Furthermore, the concept of monochromatic maximal subtree is introduced. Finally, we compare and discuss the performance of the exact and heuristic methods proposed.

Keywords: spanning tree, edge-labeled graph, metaheuristic

1 Introduction

The *minimum labelling spanning tree problem* [2], [3] (MLSTP) is a NP-Complete problem that aims to find the minimum number of labels necessary to generate a spanning tree of the input graph. The main application of the MLSTP is to reduce the number of resources managed in some network design

¹ Corresponding author: ezaldivar@id.uff.br

[4]. To do so, the MLSTP assumes that the operational costs of the nodes is static, while the operational cost of the network is proportional to its number of labels. On the other hand, edge-labelled graphs can also be used to describe network designs in which the overall cost is the sum of the operational costs of the nodes, where the operational cost of a node is defined as the number of different labels represented in it. The MLSTP is not suitable to address this new kind of situation, what justifies the introduction of the minimum representation spanning tree problem. Formally:

Definition 1.1 An edge-labeled graph (ELG) is an undirected simple graph $G = (V, E, L)$ such that V the set of vertices, E the set of edges, L is the function that assigns to each edge $e \in E$ a unique label $L(e)$.

Definition 1.2 Given an ELG $G = (V, E, L)$, the representation set of a vertex $v \in V$, denoted by $R_G(v)$, is the set of labels associated to edges that are incident to v .

The representation number $r_G(v)$ of a vertex v is the number of elements in $R_G(v)$.

Definition 1.3 Given an ELG $G = (V, E, L)$, the *Minimum Representation Spanning Tree Problem* (MRSTP) aims to find a spanning tree T of G such that the function $R(T) = \sum_{v \in V} r_T(v)$ is minimized.

Remark that, since T is a tree, $c_T(v) \geq 1, \forall v \in V$. It is convenient to use the *representation function* $Rt(T) = \sum_{v \in V} (r_T(v) - 1)$ as the objective function for the MRSTP.

The two problems (MLSTP and MRSTP) can be applied in the same environments, that is, for the same type of network structures. MLSTP assumes that the operational cost of nodes is fixed and that the operational cost of network is directly proportional to the number of labels.

However, labeled graphs can also, in a second scenario, describe networks where the operational cost of the network is the sum of the operational costs of the nodes. In this case, the operating cost of a node is assumed to be directly proportional to the number of different labels used on that node, so that MRSTP can correctly model this scenario.

One application of the MRSTP is when it is necessary to make investments on the nodes of a network, for instance for buying software licenses or hardware to enable the use of new protocols. The problem also has applications in areas such as transport projects, network design, telecommunication systems, and distribution of energy.

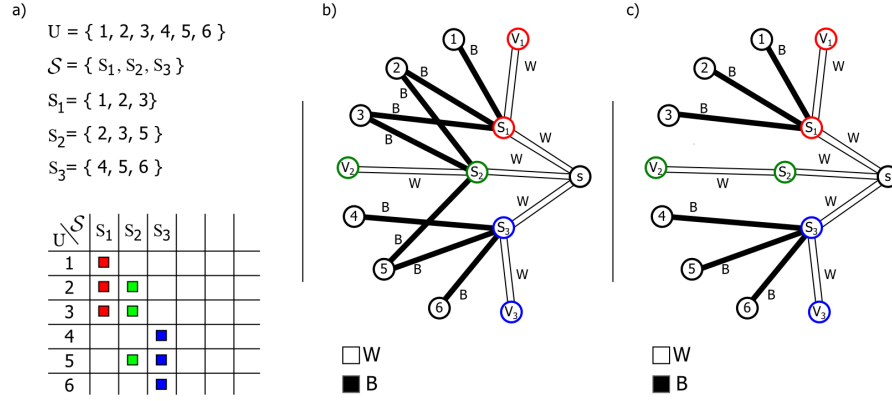


Fig. 1. Example of an instance used during the proof of NP-completeness of BRST problem.

As far as we know, the MRSTP hasn't been previously studied. We now prove that the MRSTP is NP-Hard. First, a decision version, called *Bounded Representation Spanning Tree* (BRST), for the MRSTP is defined and shown to be NP-Complete.

Definition 1.4 Given an ELG $G = (V, E, L)$ and a positive integer K , the goal of BRST problem is to determine if there is a spanning tree T for G such that $Rt(T) \leq K$.

Theorem 1.5 *BRST is NP-Complete, even when the number of labels is two.*

Proof. It is easy to see that $BRST \in NP$ since a non-deterministic algorithm need only guess a subset of edges and check in polynomial time whether these edges connect all vertices and whether $Rt(T)$ is the appropriate value.

NP-Completeness is proved by reducing the minimum set covering problem to the BRST problem. Let $U = \{a_1, a_2, a_3, \dots, a_n\}$ and $S_1 = \{a_{i_1}, a_{j_1}, a_{k_1}\}, S_2 = \{a_{i_2}, a_{j_2}, a_{k_2}\}, S_3 = \{a_{i_3}, a_{j_3}, a_{k_3}\}, \dots, S_m = \{a_{i_m}, a_{j_m}, a_{k_m}\}$ be subsets. The NP-Complete minimum set covering problem is to decide whether exists a set of at most K subsets that cover all the element in U . From U and the S_i 's, we will build a graph $G = (V, E)$ such that if G has a spanning tree (ST) with representation K , then there is a cover for S with K subsets.

The construction of G is as follows. G contains the following vertices: the element nodes $\{a_1, a_2, a_3, \dots, a_n\}$, the subset nodes $\{S_1, S_2, S_3, \dots, S_m\}$, the clones of the subset nodes $\{V_1, V_2, V_3, \dots, V_m\}$ and a special node s . The edges set contains the special edges $\{(s, S_i) | i = 1, \dots, m\}$, the edge between original and his clone for each subset node $\{(S_i, V_i) | i = 1, \dots, m\}$ and the covering edges $\{(S_p, a_{l_p}) | p = 1, \dots, m; l \in \{i, j, k\}\}$. All the covering edges have label

B and the others edges have label W . It can be shown that the construction can be accomplished in polynomial time and a minimum covering of size K exists if G has a ST with representation function value K . $\square$

Since we only used two labels in the proof we can conclude that BRSTP is NP-Complete for graphs that have two or more labels. If the number of labels is one, then the representation of each vertex is also one in the graph, and therefore it is also one in any spanning tree. In this case, we have a “yes” instance if $K \geq 0$.

2 Mathematical Formulation for the MRSTP

We use the following decision variables. Let x_e be a binary variable that is set to 1 if the edge $e \in E$ is in the final solution and 0 otherwise. Let z_l^v be a binary variable that is set to 1 if the label $l \in L$ is represented in the vertex $v \in V$ and 0 otherwise. The integer programming model is presented below:

$$\text{Minimize } -|V| + \sum_{v \in V} \sum_{l \in L} z_l^v \quad (3.1)$$

$$\text{s.t. } \sum_{x_e \in \delta(S)} x_e \geq 1, \forall S \subset V, S \neq \emptyset \quad (3.2)$$

$$z_{L(e)}^u \geq x_e, u \in \{v, w\}, \forall e = \{v, w\} \in E \quad (3.3)$$

$$\sum_{x_e \in E} x_e = |V| - 1 \quad (3.4)$$

$$z_l^v, x_e \in \{0, 1\}, \forall l \in L, v \in V, e \in E \quad (3.5)$$

In the model, (3.1) is the formulation of Rt , (3.2) and (3.4) guarantees connectivity and no cycles in the solution, (3.3) compute Rt and (3.5) declare the domain of the variables.

3 An alternative definition for the MRSTP

Definition 3.1 Let $G = (V, E, L)$ be an ELG. A monochromatic subtree (MS) of G is an acyclic connected subgraph of G induced by a set of edges with the same label.

Definition 3.2 We say a monochromatic subtree $T = (V', E', L)$, $V' \subseteq V$, $E' \subseteq E$, is maximal if does not exist a monochromatic subtree $T' = (V'', E'', L)$ such that $E' \subset E''$.

Theorem 3.3 Any edge-labeled tree T satisfies $|M(T)| = Rt(T) + 1$, where $M(T)$ is the set of maximal monochromatic subtrees of T .

We omit the proof of this Theorem due to the space limitations. Taking into account this result, the MRSTP can be formulated alternatively in the following way:

Definition 3.4 Let $G = (V, E, L)$ be an ELG, and let $M(T)$ be the set of maximal monochromatic subtrees of a tree T . The goal of *Minimum Representation Spanning Tree Problem* (MRSTP) is to find the spanning tree T of G such that $|M(T)|$ is minimized.

4 GRASP based metaheuristic

This section presents a new **constructive heuristic** for the MRSTP. It is an adaptation of the *Maximum Vertex Cover Algorithm* (MVCA) heuristic [5]. We propose to use the *Minimum Labelling Steiner Problem* (MLSteiner) [1].

First we define the MLSteiner as follows. Given $G = (V, E, Q, L)$ an edge-labeled graph and $Q \subset V$ a set of terminal nodes of G , the MLSteiner aims to find a tree $T = (V', E', L')$, $Q \subseteq V' \subseteq V$ spanning all the vertices of Q , such that $|L'|$ is minimized. So we can show how to reduce an instance of the MRSTP to one of the MLSteiner.

Let $G = (V, E, L)$ be an instance of BRST where $L(e)$ is the label of the edge e (that is: $L = \{(e, L(e)) | e \in E\}$), we build a instance of BLSteiner (decision version of MLSteiner) $G' = (V + \{s_e | e \in E\}, E', V, L')$ such that G has a MRST T with $R(T) = K$ if and only if G' has a MLSteiner T' with K labels.

First we consider a bijective function $f : V \times \text{Img}(L) \rightarrow \text{Img}(L')$. So we can complete G' by setting the set E' and the set L' as follows:

$$E' = \bigcup_{\{v,w\} \in E} \{\{v, s_{\{v,w\}}\}, \{w, s_{\{v,w\}}\}\} \quad (4.1)$$

$$L' = \bigcup_{e \in E} \{(\{v(e), s_e\}, f(v(e), L(e))), (\{w(e), s_e\}, f(w(e), L(e)))\} \quad (4.2)$$

It can be easily checked that G' admits a spanning tree with labelling K if and only if G have spanning tree T with $R(T) = K$. Now we can apply the MVCA idea. At each iteration we choose a node v and a label l for generating an edge set where all elements are incident to v and have label l .

Our algorithm begins with an empty tree and keep inserting edges until the spanning tree is completed. At every step, from all possibilities that minimize the representation, it chooses the option that inserts the biggest branch, where a branch is a set of edges with the same label and a common vertex. The result of this construction is the Maximum Branch Cover Algorithm (MBCA) (Algorithm 1).

Algorithm 1 $MBCA(G = (V, E, L))$

```

1:  $T \leftarrow (V', E', L'), E' \leftarrow \emptyset, V' \leftarrow \emptyset, L' \leftarrow \emptyset$ 
2:  $U \leftarrow \{(v, l) \mid \exists e = \{v, w\} \in E \text{ and } L(e) = l\}$ 
3: while  $|E'| < |V| - 1$  do
4:   for  $(v, l) \in U$  do
5:      $B_l^v \leftarrow \{e \in E \mid v \in e, L(e) = l, e \notin E' \text{ and } e \text{ doesn't form cycle in } T\}$ 
6:     while  $(\exists \{\{v, w_1\}, \{v, w_2\}\} \subseteq B_l^v \mid \exists \text{ path from } w_1 \text{ to } w_2 \text{ in } T)$  do
7:       if  $(l \text{ is represented in } w_1 \text{ inside } T)$  then
8:          $B_l^v = B_l^v \setminus \{\{v, w_2\}\}$ 
9:       else
10:         $B_l^v = B_l^v \setminus \{\{v, w_1\}\}$ 
11:      end if
12:    end while
13:     $N_l^v \leftarrow \{w \mid \{v, w\} \in B_l^v\} \cup \{v\}$ 
14:     $Drt_l^v \leftarrow Rt((V' \cup N_l^v, E' \cup B_l^v, L' \cup (B_l^v \times \{l\}))) - Rt((V', E', L'))$ 
15:  end for
16:   $drtmin \leftarrow \min_{(v,l) \in U, B_l^v \neq \emptyset} Drt_l^v$ 
17:   $(vmax, lmax) \leftarrow \arg \max_{(v,l) \in U, Drt_l^v = drtmin} |B_l^v|$ 
18:   $E' \cup = B_{lmax}^{vmax}$ 
19:   $V' \cup = N_{lmax}^{vmax}$ 
20:   $L' \cup = B_{lmax}^{vmax} \times \{lmax\}$ 
21:   $U - = \{(vmax, lmax)\}$ 
22: end while
23: return  $T$ 

```

For the **local search heuristic**, we propose the representation metric defined for edges as follows:

$$Rt(\{v, w\}) = \begin{cases} 2, & \text{if } c(v) \text{ and } c(w) \text{ decrease by edge elimination} \\ 0, & \text{if neither decrease} \\ 1, & \text{otherwise} \end{cases}$$

The main idea is to delete edges and replace them using MBCA algo-

rithm. Different alternatives by which edges can be removed define different neighborhoods:

- (i) By label: Edges with less voted labels are deleted and replaced by MBCA.
- (ii) By subtree: Edges from a subtree that have the higher average representation are deleted and replaced by MBCA.
- (iii) By branch: Edges that belongs to the maximal monochromatic subtree branch are deleted and replaced by MBCA.
- (iv) By node: Node with higher average representation in the incident edges is isolated and the edges are replaced by MBCA.
- (v) By expensive and balanced edges: Edges with representation 2 that connect two subtrees that have similar size are deleted and replaced by MBCA.
- (vi) By leaf edges: Edges with leaf nodes and representation 1 are deleted and replaced by MBCA.

We use the Variable Neighborhood Descent(VND) framework proposed in Algorithm 2.

Algorithm 2 $VND(y)$

```

1:  $k \leftarrow -1, j \leftarrow 0$ 
2: while  $j < 6$  do
3:    $k \leftarrow (k + 1) \% 6$ 
4:    $y' \leftarrow Search(N_{k+1}, y)$ 
5:   if  $y'$  is better than  $y$  then
6:      $y \leftarrow y'$ 
7:      $j \leftarrow 0$ 
8:   else
9:      $j \leftarrow j + 1$ 
10:  end if
11: end while
12: return  $y$ 

```

The *Search* procedure used in Algorithm 2 uses the first-improving tactic, as shown in Algorithm 3.

5 GRASP computational results

We present a comparative table of the Integer Programming Model (IPM) implemented in CPLEX solver version 12.5.1 versus MBCA based GRASP

Algorithm 3 *Search*(N_k, y)

```
1:  $y' \leftarrow y, S' \leftarrow \emptyset$ ,  
2: repeat  
3:    $S' \leftarrow \arg \max_{S \in N_k(y', S')} F(N_k, S)$   
4:    $y'' \leftarrow \text{complete } y' - S' \text{ with } MBCA$   
5:   if  $y''$  is better than  $y'$  then  
6:      $y' \leftarrow y''$   
7:     return  $y'$   
8:   end if  
9: until there is no more improvement  
10: return  $y'$ 
```

metaheuristic. We execute computational experiments in a computer with an Intel(R) Core(TM)2 Duo T6600 @ 2,20GHz processor and 4 Gb of RAM memory. We use a set of simple graph instances well known for the *Minimum Labelling Spanning Tree Problem* generated in Cerulli et al. [1].

Since the running time for the exact method was restricted to two hours, we managed to identify instances in which the IPM does not give the optimal value. From the analysis of Table 1 we can observe that the exact method is better than GRASP only for low density instances (denoted ld, when $D = 0.2$). Therefore GRASP is better for medium (md, $D = 0.5$) and high (hd, $D = 0.8$) density instances. To achieve the best performance of GRASP we set $\alpha = 0.01$. For harder instances IPM was not able to find feasible solutions after two hours of running time. All instances that were optimally solved by the IPM present less than two hours of running time.

6 Summary and Remarks

We introduce and define the *Minimum Representation Spanning Tree Problem* (MRSTP) and auxiliary concepts. We also give the mathematical formulation and prove the NP-Hardness of the problem. We show it to be, as far as we know, the first problem with labeled edges that is NP-Hard with only two labels. We also propose a MVCA based constructive heuristic MBCA and design a neighborhood structure. Finally, we compare the performance of a GRASP metaheuristic and an integer programming model (IPM). We conclude recommending to use IPM in low density and small instances, and GRASP otherwise.

As future work, we recommend to improve exact methods, testing thinner mathematical formulations.

Table 1
Computational Results

Instance				IPM		GRASP	
Name	n	l	d	Rt	t(s)	Rt	t(s)
hd201	20	20	hd	2	0.11	2	0.28
md201	20	20	md	2	0.04	2	0.06
hd300	30	30	hd	4	7200.02	4	1.09
md300	30	30	md	4	0.53	4	0.34
ld301	30	30	ld	15	0.68	15	1.59
hd400	40	40	hd	6	7200.02	5	15.55
ld500	50	50	ld	23	6.49	23	6.18
ld501	50	50	ld	25	7200.01	25	71.71
md508	50	50	md	10	7200.03	9	32.42
hd508	50	50	hd	10	7200.01	5	49.30
ld1000	100	100	ld	34	847.50	37	33.47
ld1007	100	100	ld	42	7200.07	40	47.97
hd1009	100	100	hd	70	7200.90	14	841.38
md1007	100	100	md	37	7200.05	23	902.75
md1001	100	100	md	45	7200.25	23	76.76

References

- [1] Cerulli, R., Fink, A., Gentili, M., and Vo, S. *Extensions of the minimum labelling spanning tree problem* Journal of Telecommunications and Information Technology 4 (2006), pp. 3945.
- [2] Chang, R.-S. and S.-J. Leu, *The Minimum Labelling Spanning Tree* Information Processing Letters, 63.5 (1997), 277-282.
- [3] Chwatal, A. M., G. R. Raidl e K. Oberlechner. *Solving the Minimum Label Spanning Tree Problem by Mathematical Programming Techniques*. Adv. Operations Research 2011 (2011) (pp. 10, 18, 19, 23, 25, 26, 35, 60).
- [4] Silva, T. G. , E. V. Queiroga, G. F. de Sousa Filho, L. A. F. Cabral and L. S. Ochi, *Abordagem híbrida para o problema da árvore geradora com rotulação mínima* Anais do XLVI Simpósio Brasileiro de Pesquisa Operacional (2014) 1823-1833.
- [5] Xiong, Y., B. Golden and E. Wasil, *Improved heuristics for the minimum label spanning tree problem* Evolutionary Computation, IEEE Transactions, v.10, n.6 (2006) 700-703.