Centro de Investigação Operacional

Min-Degree Constrained Minimum Spanning Tree Problem: Complexity, proprieties and formulations

Ana Maria de Almeida, Pedro Martins and Maurício C. Souza

CIO – Working Paper 6/2006

Min-Degree Constrained Minimum Spanning Tree Problem: Complexity, proprieties and formulations

Ana Maria de Almeida
CISUC - Centro de Informática e Sistemas da Universidade de Coimbra and
Departamento de Matemática, FCTUC
Universidade de Coimbra, Portugal
E-mail: amca@mat.uc.pt

Pedro Martins
CIO - Centro de Investigação Operacional - FC/UL and
ISCAC - Instituto Politécnico de Coimbra, Portugal
E-mail: pmartins@iscac.pt

Maurício C. Souza
Departamento de Engenharia de Produção
Universidade Federal de Minas Gerais, Brasil
E-mail: mauricio@dep.ufmg.br

June, 2006

Abstract

Given an undirected graph $G = (V, E)$ and a function $d : V \rightarrow N$, the Min-Degree Constrained Minimum Spanning Tree (md-MST) problem is to find a minimum cost spanning tree T of G where each node $i \in V$ has minimum degree $d(i)$ or is a leaf node. This problem is closely related with the well-known Degree Constrained Minimum Spanning Tree (d-MST) problem, where the degree constraint is an upper limit instead.

In this paper we prove that the md-MST problem is NP-hard and present some proprieties, namely upper and lower limits to the number of central nodes and leaf-nodes in any feasible solution to the problem. Flow based formulations are also proposed and computational experiments involving the associated LP relaxations are presented. These results indicate that, for similar formulations to both d-MST and md-MST problems, the LP versions of the d-MST stronger flow models seem to provide a better approximation to the integer polyhedron than the correspondent md-MST flow formulations (within the linear relaxation context), which seems to indicate that it might be harder to get good formulations to the later problem.

Keywords: Degree constrained spanning tree problems, complexity, single-commodity flow formulations, multicommodity flow formulations.

1 Introduction

Let $G = (V, E)$ be a connected undirected graph, where $V = \{1, \ldots, n\}$ is the set of nodes and E the set of edges. Consider that there are associated positive costs, c_e, to each edge $e \in E$. Given a positive integer valued function $d : V \rightarrow N$ on the nodes, the Min-Degree Constrained Minimum Spanning Tree (md-MST) problem consists in finding a spanning tree T of G with minimum total edge cost, given by $\sum_{e \in T} c_e$, and where each node $i \in V$ either has degree at least $d(i)$ or is a leaf node.

Figure 1 gives an example of a feasible solution to the md-MST problem, for a graph G with 10 nodes and considering $d(i) = 3$ for all $i \in V$. In this solution, all the internal nodes (in grey) have at least degree 3, while all the other nodes are leafs.

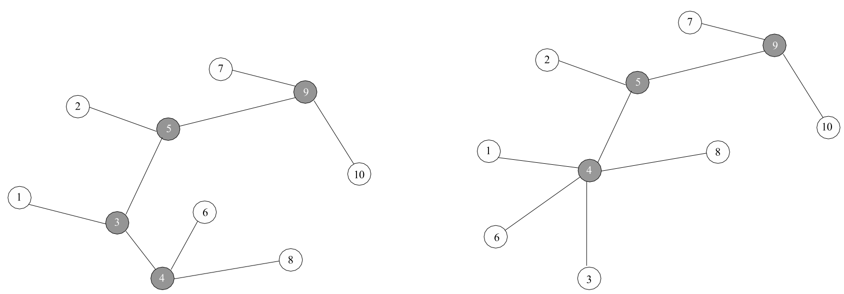

Figure 1: Example of two feasible trees to the md-MST problem, considering a graph with $n = 10$ and $d(i) = 3$ for all $i \in V$.

This problem is closely related with the well-known Degree Constrained Minimum Spanning Tree[1] (d-MST) problem, where also a minimum cost spanning tree T of G is seek, but where each node $i \in T$ must have degree at most $d(i)$. Under the same assumptions as in the previous example, Figure 2 shows a feasible solution for the d-MST problem.

Notice that the solution presented in Figure 2 is not feasible to the md-MST problem, because nodes 2, 3, 7 and 9, although being internal, have degree less than 3.

From this point on, we refer to d when all nodes have equal bound degree, i.e., when $d(i) = d$ for $d \in N$ and all $i \in V$. As usual, MST stands for minimum (cost) spanning tree.

There are various applications of the d-MST problem, namely within VLSI layout and network design (see, Monma and Shallcross [18] and Stoer [26]); in the design of electrical circuits (see, Narula and Ho [20]); in the design of road systems, where a limit is imposed on

[1] Also otherwise known as the Bounded Degree Minimum Spanning Tree.

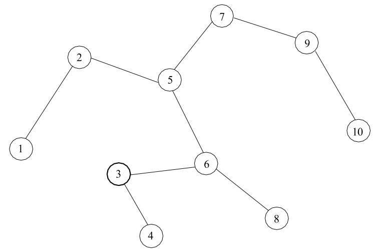

Figure 2: Feasible solution to the d-MST problem, for a graph with $n = 10$ and considering $d(i) = 3$ for all $i \in V$

the number of roads that are allowed to meet at any crossing (see, Savelsbergh and Volgenant [25]); in communication networks, where a degree constraint limits vulnerability in case of drop out of a crossing (see, Savelsbergh and Volgenant [25]) and in the design of computer communication networks (see, Gavish [10, 11]). In fact, the d-MST problem arises frequently in the design of telecommunication and energy networks. It also appears as a subproblem in the design of communication networks, transportation, sewage and plumbing.

Practical applications to the md-MST problem may occur in cases where one needs to identify a set of places (nodes) that centralize incidence with other entities (peripheral nodes), in a way that a place (node) can only assume a central status if assigned to, at least, d other places (nodes). Otherwise it must be a peripheral entity (leaf node). To satisfy this restriction, the solution should link at minimum cost all pairs of nodes, which is characterized by a spanning tree, assuming that only positive costs in the edges are being considered. Therefore, the central nodes can characterize central distribution places, or centralized communication devices, while the peripheral nodes act as individual consumers or clients.

The d-MST problem was first posed in Deo and Hakimi [6] and since then studied by many researchers, to which exact and approximate methods have been proposed, see for instance Savelsbergh and Volgenant [25], Boldon, Deo and Kumar [2], Craig, Krishnamoorthy and Palaniswami [5], Zhou and Gen [28, 29], Knowles and Corne [15], Caccetta and Hill [3], Ribeiro and Souza [23] and Andrade, Lucena and Maculan [1] for heuristic approaches; and Gavish [10], Volgenant [27] and Cacceta and Hill [3] for lower bounding approaches. Krishnamoorthy, Ernst and Sharaiha [16] provide an extensive description and comparison of approximate and exact algorithms to the d-MST problem. The decision version of a simpler problem, where all edge weights are equal, was proved to be NP-Complete by Garey and Johnson [9] (Problem ND1).

For the special case with a degree constraint for only one node, Gabow and Tarjan [8] proposed an efficient algorithm with time complexity of $O(|E| + n \log n)$.

When $d = 1$, the d-MST problem is impossible, while for $d = 2$ the problem is to find a minimum cost Hamiltonian path in G. Based on this, Garey and Johnson [9] have showed the d-MST problem to be NP-hard, so it is unlikely that a polynomially bounded algorithm exists for solving the general d-MST problem, for $2 \leq d \leq n-2$.

However, the order of complexity for the d-MST problem varies if the cost function is defined on different metric spaces. In fact, based on some computational results, Savelsbergh and Volgenant [25] indicate that the Euclidean problems were observed to be much easier than were the non-Euclidean ones. One reason that is pointed out by the authors is related with the proximity of the nodes-degree in the optimal solutions of the MST and the d-MST problems. Notice that the Euclidean d-MST problem is known to be NP-hard when $d = 2$ and $d = 3$. As previously referred, for $d = 2$ it is a generalization of the Hamiltonian path problem and for $d = 3$ it has been proved to be NP-hard by Papadimitriou and Vazirani [22], who conjectured that it is NP-hard for $d = 4$ as well. When $d = 5$, the problem can be solved in polynomial time (see, Monma and Suri [19]). Furthermore, any MST whose nodes have integer coordinates has maximum degree at most 5 (see, Papadimitriou and Vazirani [22]). Moreover, Robins and Salowe [24] studied the maximum possible vertex degree of an MST in a k-dimensional space R^k, with L_p metrics. In particular, they show that for any finite set of points in the Manhattan (rectilinear) plane there exists an MST with maximum degree of at most 4, and that for a three-dimensional Manhattan space the maximum possible degree of a minimum-degree MST is either 13 or 14.

Approximate results, relating both MST and d-MST optimal solution costs were discussed in Khuller, Raghavachari and Young [14] and Fekete et al. [7]. In the first paper the authors propose an algorithm for computing a degree-3 (degree-4) tree for points in R^2 with Euclidean distances that is within 1.5 (1.25) of an MST of the points. An extension of the algorithm finds a degree-3 tree of an arbitrary set of points in R^k within 5/3 of an MST. The authors also conjecture about even stronger ratios. Furthermore, if an MST of the points is given as part of the input, their algorithm runs in $O(n)$ time. In Fekete et al. [7], the authors show a similar result for points in R^2, considering different metric spaces, namely with L_1 and L_∞ distances. The authors also show that it is possible to construct a degree-2 tree (Hamiltonian path) for an arbitrary set of points in R^2 with Euclidean (L_2) distances that is arbitrarily close to 2 of an MST of the points. For the general problem on geometric graphs with distances induced by various L_p norms the authors propose an adoption technique that given a minimum spanning tree yields approximation algorithms with factors less than 2.

These results show that if we want to tackle hard d-MST problems, then we must concentrate in $d = 2, 3$ and 4 instances if we are dealing with Euclidean distances, or, if higher degree problems are to be addressed, then different metric spaces have to be considered. Due to this reason, some authors have proposed instances to the d-MST problem where the edge-costs are randomly generated (see, e.g., Caccetta and Hill [3] and Savelsbergh and Volgenant [25]).

This aspect has been an additional motivation towards the study of the new md-MST problem. Besides the theoretical hardness associated to both problems, the computational results presented in Section 5 show that, for similar dimensional instances and similar metric spaces, it appears to be harder to get a good characterization of the md-MST integer polyhedron. Therefore, the md-MST seems to be a very challenging network design problem.

We can also find some similarities among the md-MST problem and hub location problems, taking our central-nodes as hub-nodes. However, there are some differences, namely, in hub location we deal with a network design flow feasibility problem, while the md-MST only addresses the network design structure. On the other hand, in the md-MST we always have a spanning tree linking the central-nodes (as will be observed in Section 3), while in most hub location problems the hub nodes are fully interconnected.

In the next section we prove that the md-MST problem is NP-hard. In Section 3 we present some proprieties related with this problem. In Section 4, different formulations for the md-MST problem are proposed, considering flow based models. In the same section we also present similar formulations to the more classical d-MST problem. Computational results that compare the LP bounds produced by the various formulations are presented in Section 5. In the last section, some concluding remarks are given.

2 Complexity of the md-MST problem

Considering $d \geq 4$ and using a constructive approach, analogous to the proof of Partition into Triangles found in [9], and inspired by the NP–hardness proof found in [13], we can reduce the k–Dimensional Matching Problem (or kDM, where $k \geq 3$) to the md-MST.

The k–Dimensional Matching [21] can be stated as, given k disjoint sets, Y_i, $i = 1, \ldots, k$, each of size n, and a k–ary relation, $H \subseteq Y_1 \times Y_2 \times \ldots \times Y_k$, does there exist a set $\mathcal{M} = \{(y_{11}, \ldots, y_{1k}), \ldots, (y_{n1}, \ldots, y_{nk})\} \subseteq H$, of n ordered k-tuples, so that each element of the tuple is contained in exactly one of the sets Y_i and all tuple elements are different, i.e,

$$\forall j = 1, \ldots, n\,,\ y_{ji} \in Y_i\,,\ i = 1, \ldots, k$$

and

$$\forall l = 1, \ldots, n,\ y_{li} \neq y_{lj},\ ,i \neq j\,,\ i, j = 1, \ldots, k\,.$$

The kDM, for $k \geq 3$, is an NP-hard[21] problem thus, if we can reduce this problem to the md-MST, then the latter must also be NP-hard.

Theorem 1 *For any given integer $d \geq 4$, the md-MST problem is NP-hard.*

Proof: We begin by showing that this is true for $d = 4$. We will then end the proof generalizing this result for higher dimensions of d.

Assume that $d = 4$. Using a general instance I for the 3–Dimensional Matching Problem (3DM) [9], from which we will construct a new instance graph for the $m4$-MST Problem, G, we proceed to prove that, I has a 3–dimensional matching if and only if G has an optimal solution for the $m4$-MST, that is, a MST where each internal node has degree not smaller than 4.

Let V_1 denote the union of three disjoint sets, $V_1 = V_{11} \bigcup V_{12} \bigcup V_{13}$, where $|V_{1k}| = q$, for $k = 1, 2, 3$, and $q \geq d$. Consider also that $H = \{(a_1, b_1, c_1), \ldots, (a_N, b_N, c_N)\} \subseteq V_{11} \times V_{12} \times V_{13}$, is a ternary relation between V_1 elements, consisting of $N \geq q$ triplets. Denote by $I = (V_1, H)$ this possible instance for a 3DM problem.

Let's construct a new weighted three layered graph, $G = (V, E)$, according to the following rules:

1. $V = V_1 \bigcup \{X_1, X_2, \ldots, X_N\} \bigcup \{R\}$, where the new $N \leq q^3$ nodes in $\{X_i\,, i = 1, \ldots, N\}$ represent all the possible triplets in H;

2. $E = \{(R, X_i)\,, i = 1, \ldots, N\} \bigcup \{(X_i, a_i), (X_i, b_i), (X_i, c_i)\,, 1 \leq i \leq N\}$. That is, node R connects all the X_i nodes, $i = 1, \ldots, N$ (see figure 3). (a_i, b_i, c_i) are all possible (different) triplets in H, and therefore each a_i is connected with the respective X_i, $i = 1, \ldots, N$.

3. each edge in $\{(R, X_i)\,, i = 1, \ldots, N\}$ has an associated zero weight and each one of the remaining edges in E, that is, the ones that link the second to the third layer, has an associated unitary weight.

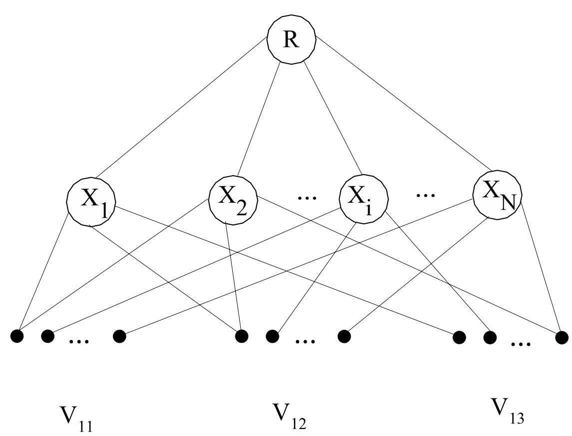

Figure 3: Example for the undirected weighted graph G_2

This construction is obviously polynomially dependent on the number of nodes in V_1 and triplets in H, and thus is of $\mathcal{O}(q + N)$. Since $N \leq q^3$, in the worst case we will have a constructions of the $\mathcal{O}(q^3)$.

A spanning tree for G exists if, at least, a perfect 3-dimensional matching exists, otherwise there should be at least one node on the bottom layer that is not connected to any one of the middle layer nodes and, therefore, graph G would be disconnected. This means that there is an optimal solution for $m4$-MST in G whenever I admits (at least) one 3-dimensional matching and vice-versa. Finally, because of the way the weights were associated to the different edges, any minimum tree for G must have total weight $w^* = 3q$.

Let's start by proving that if I admits a 3-dimensional matching then G has an optimal minimum spanning tree where the minimum degree of each node is either four (or superior) or is equal to one.

Consider then that $\mathcal{M} = \{(a_1, b_1, c_1), \ldots, (a_q, b_q, c_q)\} \subseteq H$ is a 3DM matching for I. Therefore, $|\mathcal{M}| = q \leq N$ and each and every triplet is unique in the sense that all the $3q$ nodes of V_1 are covered by the q triplets, which of course implies that all the a_i, b_i and c_i are different $(i = 1, \ldots, q)$.

Let $T = (V_2, E_T)$ be the subgraph of G where $E_T = \{(R, X_1), (R, X_2), \ldots, (R, X_N)\} \bigcup \mathcal{C}$, and $\mathcal{C}$ represents all the edges that connect the nodes on the first triplet of $\mathcal{M}$ to node X_{j_1}, the edges that connect the nodes on the second triplet of $\mathcal{M}$ to node X_{j_2}, and so on until having the nodes on the last triplet of $\mathcal{M}$ connected to node X_{j_q}. That is:

$$\mathcal{C} = \{(X_{j_1}, a_1), (X_{j_1}, b_1), (X_{j_1}, c_1), \ldots, (X_{j_q}, a_q), (X_{j_q}, b_q), (X_{j_q}, c_q)\} \ .$$

Trivially, there are no cycles in T. And, since $|\mathcal{C}| = 3|\mathcal{M}| = 3q$, we have that, $|E_T| = N + |\mathcal{C}| = N + 3q = |V| - 1$, therefore T connects all the nodes of G using $|V| - 1$ edges with no cycles being, thus, a spanning tree for G. Moreover, associated with each (X_{j_i}, a_i) we have an unitary weight and the weights associated to the edges that connect node R to the X's nodes are zero, giving thus a total tree weight of $w(T) = 0 \times N + 1 \times |\mathcal{C}| = 3q = w^*$, i.e., T is a minimum spanning tree for G.

Let $deg_T(i)$ stand for the degree for node i in T. It remains to be proven that the degree in each node of T agrees with the $m4$-MST restrictions, i.e.,

$$\forall v_i \in V_2 \, , \, deg_T(v_i) \geq d \vee deg_T(v_i) = 1$$

which is quite straightforward since,

- every node v_i on the bottom layer (nodes from V_1) has degree one (the edge incident to the node in the middle layer that represents the enumeration of the triplet where v_i stands in the matching);
- the node R has degree $N \geq q \geq d$;
- each node on the middle layer $X_{j_i}, i = 1, \ldots, q$, has degree $4 = d$ and the remaining (if any) $N - q$ nodes have degree one.

Clearly, if I admits a 3-dimensional matching than G has an optimal solution for $m4$-MST.

We will now prove the necessary condition:

Let $T^* = (V, E_{T^*})$ be an optimal solution for the $m4$-MST over G, that is, T^* is a spanning tree for G, with total weight $w(T^*) = 3q$ and each of its nodes either has degree 1 or has degree at least d.

Since T^* is a spanning tree for G, it has exactly $N + 3q$ edges and, being minimal, among these there must be exactly $3q$ with unitary weight. But this then means that there are exactly $3q$ edges incident to the $3q$ nodes on the bottom layer, which are thus leafs, and the remaining N edges are the ones that connect node R to the second (middle) layer nodes.

Now we must prove that this optimal tree implies that a 3-dimensional matching exists for I, which will be done reasoning over the degree of the middle layer nodes. As previously noted, there are exactly $3q$ links between the X_j nodes and the bottom nodes, and all the X nodes are connected to node R. Therefore, each node from the third layer can only be connected to one X_j node and each X_j is connected to 3 different bottom layer nodes, one in each subset V_{1i} $(i = 1, 2, 3)$. Thus, we must have exactly q nodes, $X_{j_1}, \ldots, X_{j_q}$, such that $deg_{T^*}(X_{j_i}) = 4$ and any remaining middle layer node must be a leaf.

In this case, each X_{j_i} represents a feasible triplet for a 3DM for I, and since there are q different triplets, I admits (at least) one 3-dimensional matching, which concludes the proof of theorem 1.

Finally, we remark that this proof is easily generalized for $d > 4$, using a reduction from $(d-1)$–Dimensional Matching, which is known to be an NP-hard problem for $d-1 \leq 2$. $\square$

3 Proprieties of the md-MST problem

We start by introducing some notation. Let $T = (V, E_T)$ be a spanning tree of G and $deg_T(i)$ the degree of node i in T. We also define T_C as the subgraph of T obtained after eliminating all the leaf-nodes and all the edges incident in those nodes in T. Hence, the nodes in T_C are represented by $V_C \subset V$ and V_1 represents the eliminated nodes from T (leaf-nodes), with $V_C \cup V_1 = V$ and $V_C \cap V_1 = \emptyset$. Therefore, $G_C = (V_C, E_C)$ is a subgraph of G with $E_C = \{\{i, j\} \in E : \ i, j \in V_C\}$. We call the nodes in V_C the central-nodes of T.

Trivially, the subgraph T_C is a spanning tree for G_C, stated in the following proposition.

Proposition 1 *The subgraph T_C is a spanning tree for G_C.*

It can also be shown that the cost of the subtree T_C of an optimal solution T^* to the md-MST problem may not correspond to a minimum cost solution for G_C. Consider the following example where an instance of the md-MST problem is given, with the correspondent costs[2] c_e appended to the edges of the trees.

[2]The correspondent cost matrix C can be found in Appendix A.

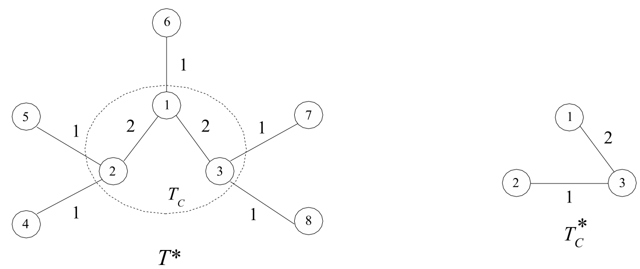

Figure 4: Optimal md-MST solution for a given graph G with $n = 8$ and $d = 3$, considering the cost matrix C presented in the Appendix A, and the optimal tree T_C^* of G_C. The labels on the edges represent the correspondent costs.

The example in Figure 4 shows that the cost of subtree T_C is equal to 4, while the cost of the optimal tree of G_C is equal to 3, confirming that subtree T_C of T^* may not correspond to the lowest cost feasible spanning tree in the subgraph G_C.

The following proposition is also easy to prove.

Proposition 2 *If $d \leq 2$, then the md-MST problem corresponds to the MST.*

In fact, when $d \leq 2$ the degree constraint has no restrictive effect. Therefore, both the md-MST and the MST problems have the same set of feasible solutions, and the problem can be solved in polynomial time.

On the other way, we can also state the following proposition.

Proposition 3 *If $n - 1 \geq d \geq \lfloor n/2 \rfloor + 1$, then any md-MST feasible tree is a star.*

Proof: It is not hard to show that condition $n - 1 \geq d \geq \lfloor n/2 \rfloor + 1$ is not violated by any spanning star of G. Then it only remains to be proven that there are only stars in the feasible set of md-MST.

Suppose that $n - 1 \geq d \geq \lfloor n/2 \rfloor + 1$ and the md-MST problem has a feasible solution T that is not a star. In this case, T has at least two central-nodes. Each of these nodes must be connected to $\lfloor n/2 \rfloor$ other, non common, leaf-nodes. This means that T must have $2\lfloor n/2 \rfloor + 2$ nodes, which is the same as $2(\lfloor n/2 \rfloor + 1)$. As $2\left(\lfloor n/2 \rfloor + 1\right) > 2(n/2) = n$, then T is not a tree; hence it cannot be a feasible md-MST solution. $\square$

This implies that, when $n - 1 \geq d \geq \lfloor n/2 \rfloor + 1$, the problem can be solved by inspection.

The following results relate the constraint on the nodes-degree with the feasibility of both d-MST and md-MST problems, namely by establishing upper and lower limits on the number of leaf-nodes in any feasible tree for each of the two problems.

Proposition 4 *If T is a feasible solution to the md-MST problem, and L its set of leaf-nodes, then $|L| \geq n - (n-2)/(d-1)$.*

Proof: Let T be a feasible spanning tree of G and $L \subset V$ its set of leaf-nodes. Using the well-known Handshaking Lemma, we have

$$\sum_{i\in L} deg_T(i) + \sum_{i\in V\setminus L} deg_T(i) = 2(n-1)$$

As $deg_T(i) = 1$ for all $i \in L$, then

$$\sum_{i\in V\setminus L} deg_T(i) = 2(n-1) - |L| \tag{1}$$

By definition, all non-leaf nodes have minimum degree at least d, then we have

$$d(n - |L|) \leq 2(n-1) - |L|$$

which is the same as (assuming that $d \geq 2$)

$$|L| \geq n - (n-2)/(d-1) \tag{2}$$

□

We can state a similar result for the d-MST problem. In this case, equality (1) holds for T a feasible d-MST solution. However, this time $deg_T(i) \leq d$ for all $i \in V \setminus L$, implying $2(n-1) - |L| \leq (n - |L|)d$, which is the same as

$$|L| \leq n - (n-2)/(d-1) \tag{3}$$

allowing to establish the following proposition.

Proposition 5 *If T is a feasible solution to the d-MST problem, and L its set of leaf-nodes, then $|L| \leq n - (n-2)/(d-1)$.*

A tighter upper bound to the number of leaf-nodes in a feasible d-MST solution is proposed by Cieslik [4], establishing that when the n nodes are defined in a Banach-Minkowski space M_d with unit ball B, then

$$|L| \leq n - \frac{n-2}{min\{d, z_d(B)\} - 1} \tag{4}$$

where $z_d(B)$ is the Hadwiger number for the unit ball B in the space M_d.

Considering that $|L|$ is an integer and any tree has at most $n-1$ leaf-nodes, then inequality (2), associated with the md-MST problem, can be rewritten as

$$n - \left\lfloor \frac{n-2}{d-1} \right\rfloor \leq |L| \leq n-1 \tag{5}$$

The same way and considering that any tree has at least 2 leaf-nodes, then inequality (3), related with the d-MST problem, can be rewritten as

$$2 \leq |L| \leq n - \left\lceil \frac{n-2}{d-1} \right\rceil \tag{6}$$

From a different perspective, we can also define lower and upper bounds to the number of central-nodes in a feasible md-MST tree T. Therefore, considering $S = V \setminus L$ as the set of central-nodes in T, we have $|S| = n - |L|$. Using inequality (5), we obtain

$$1 \leq |S| \leq \left\lfloor \frac{n-2}{d-1} \right\rfloor \tag{7}$$

This result leads to the following corollary.

Corollary 1 *The number of central-nodes in any feasible solution T to the md-MST problem is bounded by the expression* $1 \leq |S| \leq \lfloor (n-2)/(d-1) \rfloor$.

4 Formulations to the md-MST problem

A network design problem, including those that involve a tree construction, can be formulated in a number of different ways. Among others, we can think of natural and extended formulations, namely on flow based models. An interesting survey and comparison of such formulations to the MST problem can be found in [17]. In that comparison, the authors observe that better formulations (i.e., more compact and/or with a better LP bound) can be obtained by formulating network design problems in a directed graph.

Many authors have proposed formulations to the d-MST problem. Gavish [10] used a single-commodity flow model and applied a Lagrangian relaxation technique in order to obtain approximate and optimal solutions to this classical problem. Later on, in [27], [3] and [1] the authors proposed natural formulations to the same problem, defined on an undirected graph. In the first and third papers a Lagrangian relaxation technique as also been used, while in the second one, Cacceta and Hill considered a branch-and-cut procedure.

In this section we present two extended formulations to both d-MST and md-MST problems, using flow based models. For that reason and considering the already mentioned observation by Magnanti and Wolsey [17], these formulations are defined on a directed version of graph G. The new oriented graph is obtained by replacing each edge $\{i, j\}$ by two directed arcs (i, j) and (j, i), both having the same cost as the original edge. We also centralize the graph on a special node r taken from V. From this special node, only outward arcs will be kept in the graph. We denote this graph by $G_r = (V, A_r)$, where $A_r = \{(i, j) : i \in V \text{ and } j \in V \setminus \{i, r\}\}$ is the mentioned set of arcs. Node r acts as a root in any feasible arborescence of G_r, generating the whole flow sent into the network. Therefore, we assume that in any feasible solution the arcs are directed outward from the root.

Throughout this section we denote by $P(\mathrm{F})$ the set of feasible solutions of a given model F, and F_L denotes its linear programming relaxation.

Let us proceed by presenting a general formulation for both problems. Consider the set of design variables x_{ij} for all $(i, j) \in A_r$, where $x_{ij} = 1$ if arc (i, j) is in the solution, and 0 otherwise. We also define the set of node-variables k_i, being related just with the md-MST problem, where $k_i = 1$ if i is a central-node, and $k_i = 0$ when i is a leaf-node. Using these variables we can formulate both the d-MST and the md-MST problems through the following general models.

Formulation d-F		**Formulation md-F**	
$min \sum_{(i,j)\in A_r} c_{ij}x_{ij}$		$min \sum_{(i,j)\in A_r} c_{ij}x_{ij}$	(8)
$s.t. \sum_{j\in V\backslash\{r\}} x_{rj} \leq d$	(9a)	$s.t. \quad (d-1)k_r \leq \sum_{j\in V\backslash\{r\}} x_{rj} - 1 \leq (n-2)k_r$	(9b)
$\sum_{j\in V\backslash\{i,r\}} x_{ij} \leq d-1, \ i \in V\backslash\{r\}$	(10a)	$(d-1)k_i \leq \sum_{j\in V\backslash\{i,r\}} x_{ij} \leq$ $\leq (n-2)k_i, \ i \in V\backslash\{r\}$	(10b)
$x \in X_r \subset \{0,1\}^{(n-1)^2}$		$x \in X_r \subset \{0,1\}^{(n-1)^2}$	(11)
		$k_i \in \{0,1\}, \ i \in V$	(12)

In both formulations we are looking for a minimum cost spanning trees in G_r, as defined by (8) and (11), where X_r is the set of incidence vectors that characterize spanning trees of G_r. As assumed, all spanning trees are oriented outward the root node, which means that each node, out of the root, has in-degree equal to 1. The two problems are characterized by constraints of the type (9) and (10), with important differences in each case. In the first case, that is, for the d-MST problem, constraint (9a) and (10a) impose an upper limit on the number of outward arcs incident in any of the nodes from V. This upper bound is equal to d when those arcs diverge from r, defined in (9a), and equal to $d-1$ in all other cases, that is, for any other node i from $V \setminus \{r\}$, established in (10a), considering that these nodes have in-degree equal to 1.

In the second case, this time related with the md-MST problem, inequalities (9b) and (10b) define a lower and an upper bound to the number of outward arcs incident in each node i. However, and as before, there are important differences when characterizing the leaf-nodes. In fact, node r is a leaf when there is a single arc diverging from it, while any other leaf-node has a single inward arc. Therefore, when $i = r$, characterized by (9b), there is always at least one arc diverging from r, as it is the root, hence subtracting that arc, the sum of all the other outward arcs incident in r should be bounded by $(d-1)$ and $(n-2)$, for $k_r = 1$, that is, when r is a central-node. When $k_r = 0$, node r is a leaf and the second inequality in (9b) establishes that $\sum_{j\in V\backslash\{r\}} x_{rj} \leq 1$, which allows a single arc to diverge from r. On the other hand, for any spanning tree T_r of G_r, if there is more than a single arc diverging from r, then $\sum_{j\in V\backslash\{r\}} x_{rj} - 1 > 0$, which forces k_r to be 1, otherwise, if there is a single arc diverging from

r, then $\sum_{j \in V \setminus \{r\}} x_{rj} - 1 = 0$ which implies that $k_r = 0$. A similar analyze can be considered for all other nodes $i \in V \setminus \{r\}$, this time associated with constraints (10b). To these nodes there is always an inward arc incident in i. When i is a central-node, there must exist at least $(d-1)$ arcs diverging from i. When i is a leaf, then no arcs are allowed to diverge from i.

Finally, constraints (12) impose integrality on variables k_i.

The general models d-F and md-F can be used to produce various formulations by using known characterizations of spanning trees. As mentioned before, many such characterizations can be found in [17]. In the present work, we made the option to consider only single-commodity and multicommodity flow models in order to characterize set X_r. Therefore, we need to include flow variables, namely the non-negative variables y_{ij}, describing the flow that passes through arc (i,j) in the solution (for the single-commodity flow formulation); and the non-negative flow variables f_{ij}^k, specifying the amount of flow sent from the root to node k and passing through arc (i,j) (for the multicommodity flow formulation).

Using these variables, we get the following single and multicommodity flow models to characterize set X_r for all spanning trees of G_r.

Single-commodity flow model		**Multicommodity flow model**	
$X_r = \{x \in [0,1]^{(n-1)^2} :$		$X_r = \{x \in [0,1]^{(n-1)^2} :$	
$\sum_{i \in V \setminus \{j\}} x_{ij} = 1 \ , \ j \in V \setminus \{r\}$		$\sum_{i \in V \setminus \{j\}} x_{ij} = 1 \ , \ j \in V \setminus \{r\}$	(13)
$\sum_{i \in V \setminus \{j\}} y_{ij} - \sum_{i \in V \setminus \{j,r\}} y_{ji} = 1, \ j \in V \setminus \{r\}$	(14a)	$\sum_{i \in V \setminus \{j,k\}} f_{ij}^k - \sum_{i \in V \setminus \{j,r\}} f_{ji}^k = 0, \ j,k \in V \setminus \{r\}, \ j \neq k$	(14b)
		$\sum_{i \in V \setminus \{j\}} f_{ij}^j = 1 \ , \ j \in V \setminus \{r\}$	(14c)
$x_{ij} \leq y_{ij} \leq (n-1) x_{ij} \ , \ i \in V, \ j \in V \setminus \{i,r\}$	(15a)	$f_{ij}^k \leq x_{ij} \ , \ i \in V \ , \ j,k \in V \setminus \{i,r\} \ , \ j \neq k$	(15b)
		$f_{ij}^j \leq x_{ij} \ , \ i \in V \ , \ j \in V \setminus \{i,r\}$	(15c)
$x_{ij} \in \{0,1\} \ , \ i \in V \ , \ j \in V \setminus \{i,r\}$		$x_{ij} \in \{0,1\} \ , \ i \in V \ , \ j \in V \setminus \{i,r\}$	(16)
$y_{ij} \geq 0 \ , \ i \in V \ , \ j \in V \setminus \{i,r\} \ \}$	(17a)	$f_{ij}^k \geq 0 \ , \ i \in V \ , \ j,k \in V \setminus \{i,r\} \ \}$	(17b)

Equalities (13), appearing in both models, are in-degree constraints for each node $j \in V \setminus \{r\}$. Constraints (14) are flow conservation constraints, also for each node $j \in V \setminus \{r\}$. Considering that the flow in (i,j) is given by $y_{ij} = \sum_{k \in V \setminus \{i,r\}} f_{ij}^k$, then f_{ij}^k disaggregates in the final destination node the flow in y_{ij}. For this reason, constraints (14b) and (14c) can also be seen as a disaggregated version of constraints (14a). Inequalities (15) are coupling constraints and reflect the fact that, if there is any flow passing through arc (i,j), then (i,j) must be in the solution. Furthermore, if arc (i,j) does not belong to the solution, then no flow can pass through it. As before, inequalities (15b, 15c) correspond to a disaggregated version of inequalities (15a). Constraints (16) impose integrality to the arc design variables, and constraints (17) impose non-negativity to the flow variables.

Note that the value of each variable f_{ij}^k is never greater than 1. Therefore, we could have substituted constraints (17b) by $0 \leq f_{ij}^k \leq 1$. However, this is implicitly defined by inequalities (15b, 15c).

The LP relaxation version of both models can be obtained by substituting the integrality constraints (16) by the bounding constraints

$$0 \leq x_{ij} \leq 1, \text{ for all } (i,j) \in A_r \tag{18}$$

Within the LP relaxation version of the multicommodity flow model, one can think on a stronger form for the coupling constraints (15c), defined by equalities,

$$f_{ij}^j = x_{ij}\ ,\ i \in V,\ j \in V \backslash \{i, r\} \tag{19}$$

In fact, for the single unit of flow passing through arc (i, j) with destination j, we have $f_{ij}^j = 1$, which implies that arc (i, j) is in the solution. On the other hand, if (i, j) is in the solution, then one unit of flow must be sent to node j, through the single inward arc incident in j. Therefore, we can substitute inequalities (15c) by constraints (19) in the multicommodity flow model. As a consequence, constraints (14c) get redundant, after including equalities (19) in the formulation. This can be observed by summing in i all equalities (19) and substituting the resulting right-hand-side term by 1, using constraints (13).

Considering the two characterizations of the set X_r previously proposed, with constraints (19), we can define two flow formulations to each of the problems, d-MST and md-MST.

***d*-MST formulations** :

single-commodity flow :

$$d\textbf{-SCF} : \min \left\{ \sum_{(i,j)\in A_r} c_{ij}x_{ij} : (9a), (10a), (13), (14a), (15a), (16), (17a) \right\}$$

multicommodity flow :

$$d\textbf{-MCF}: \min \left\{ \sum_{(i,j)\in A_r} c_{ij}x_{ij} : (9a), (10a), (13), (14b), (15b), (16), (17b), (19) \right\}$$

***md*-MST formulations** :

single-commodity flow :

$$md\textbf{-SCF}: \min \left\{ \sum_{(i,j)\in A_r} c_{ij}x_{ij} : (9b), (10b), (12), (13), (14a), (15a), (16), (17a) \right\}$$

multicommodity flow :

$$md\text{-}\mathbf{MCF}\text{:}\ \min\left\{\sum_{(i,j)\in A_r} c_{ij}x_{ij}\ :\ (9b),(10b),(12),(13),(14b),(15b),(16),(17b),(19)\right\}$$

The LP relaxation versions of the single-commodity flow models can be obtained by substituting the integrality constraints (17) by the bounding constraints (18). However, among the multicommodity flow models, besides the mentioned substitution in the x_{ij} variables, we need also to relax the k_i variables, by changing the integrality constraints (12) by the bounding constraints

$$0 \leq k_i \leq 1\ ,\ i \in V \tag{20}$$

According to the mentioned relaxations, we designate by d-SCF$_L$, d-MCF$_L$, md-SCF$_L$ and md-MCF$_L$, the LP relaxation versions of d-SCF, d-MCF, md-SCF and md-MCF, respectively.

These relaxed versions of the two md-MST models can be showed to be equivalent to the correspondent unconstrained versions of the problem, that is, to the MST problem. This observation is based on Proposition 6. Before presenting it, let us define the two polyhedrons

$$P(\mathrm{SCF}_L) = \left\{(x,y) \in R^{2(n-1)^2}\ :\ (x,y) \text{ verifies } (13),(14a),(15a),(17a) \text{ and } (18)\right\} \text{ and}$$

$$P(\mathrm{MCF}_L) = \left\{(x,f) \in R^{(n-1)^2}{\times}R^{(n-1)^3}\ :\ (x,f) \text{ verifies } (13),(14b),(15b),(17b),(18) \text{ and } (19)\right\}$$

corresponding to the LP relaxation versions of the two mentioned MST formulations, single and multicommodity flow models, respectively. We also define the two projected polyhedrons

$$proj_{\{x,y\}}(P(md\text{-}\mathrm{SCF}_L)) = \left\{(x,y) \in R^{2(n-1)^2}\ :\ (x,y,k) \in P(md\text{-}\mathrm{SCF}_L),\ \text{for some } k \in [0,1]^n\right\} \text{ and}$$

$$proj_{\{x,f\}}(P(md\text{-}\mathrm{MCF}_L)) = \left\{(x,f) \in R^{(n-1)^2}{\times}R^{(n-1)^3}\ :\ (x,f,k) \in P(md\text{-}\mathrm{MCF}_L),\ \text{for some } k \in [0,1]^n\right\}$$

associated with models md-SCF$_L$ and md-MCF$_L$, respectively. Then, we have the following proposition

Proposition 6 *$proj_{\{x,y\}}(P(md\text{-}SCF_L)) = P(SCF_L)$ and*
$proj_{\{x,f\}}(P(md\text{-}MCF_L)) = P(MCF_L)$.

Proof: Given a solution $(x,y,k) \in P(md\text{-}\mathrm{SCF}_L)$ and a solution $(x,f,k) \in P(md\text{-}\mathrm{MCF}_L)$, the subsolutions (x,y) and (x,f) are feasible for $P(\mathrm{SCF}_L)$ and $P(\mathrm{MCF}_L)$, respectively. Hence we have $proj_{\{x,y\}}(P(md\text{-}\mathrm{SCF}_L)) \subseteq P(\mathrm{SCF}_L)$ and $proj_{\{x,f\}}(P(md\text{-}\mathrm{MCF}_L)) \subseteq P(\mathrm{MCF}_L)$.

Conversely, given a solution $(x,y) \in P(\mathrm{SCF}_L)$ and a solution $(x,f) \in P(\mathrm{MCF}_L)$, define $\sum_{j\in V\setminus\{i,r\}} x_{ij} = \theta_i$, for all $i \in V$. As $x_{ij} \in [0,1]$, then $\theta_r \leq (n-1)$ and $0 \leq \theta_i \leq (n-2)$, for $i \in V \setminus \{r\}$. Furthermore, the flow conservation constraints (14) establish that the whole flow sent out from the root is equal to $(n-1)$, that is, both solutions (x,y) and (x,f) satisfy

$\sum_{j\in V\setminus\{r\}} y_{rj} = n-1$ and $\sum_{j,k\in V\setminus\{r\}} f^k_{rj} = n-1$, which implies, using the linking constraints (15), that the two solutions verify $\sum_{j\in V\setminus\{r\}} x_{rj} \geq 1$. Hence $1 \leq \theta_r \leq n-1$. According to constraints (9b) and (10b), appearing in both md-SCF$_L$ and md-MCF$_L$ models, we have that $((d-1)k_r \leq \theta_r - 1 \leq (n-2)k_r)$ and $((d-1)k_i \leq \theta_i \leq (n-2)k_i)$, for all $i \in V \setminus \{r\}$, which is the same as

$$\frac{\theta_r - 1}{n-2} \leq k_r \leq \frac{\theta_r - 1}{d-1} \quad \text{and} \quad \frac{\theta_i}{n-2} \leq k_i \leq \frac{\theta_i}{d-1}, \text{ for all } i \in V \setminus \{r\} \tag{21}$$

As $\theta_r \in [1, n-1]$ and $\theta_i \in [0, n-2]$, for $i \in V \setminus \{r\}$, then there is a solution in the k variables with $k \in [0,1]^n$ that also satisfies (21). Consequently, $(x,y) \in proj_{\{x,y\}}(P(md\text{-SCF}_L))$ and $(x,f) \in proj_{\{x,f\}}(P(md\text{-MCF}_L))$, implying that $P(\text{SCF}_L) \subseteq proj_{\{x,y\}}(P(md\text{-SCF}_L))$ and $P(\text{MCF}_L) \subseteq proj_{\{x,f\}}(P(md\text{-MCF}_L))$. □

An outcome of Proposition 6 establishes that the lower bounds produced by both LP relaxation models md-SCF$_L$ and md-MCF$_L$ are the same as the bounds produced by the same formulations without degree constraints, that is, the bounds produced by models SCF$_L$ and MCF$_L$, respectively, to the MST problem. This means that the lower bounds produced by md-SCF$_L$ and md-MCF$_L$ are insensitive to the d parameter value.

Using Proposition 6 and the polyhedral discussion on some MST formulations proposed in Magnanti and Wolsey [17], we can state that model md-MCF$_L$ is symmetric relative to the root node, that is, the optimal solution value of md-MCF$_L$ is always the same, no matter the node selected for the root. This is based on the the equivalence between the LP relaxation versions of the multicommodity flow formulation and some known natural formulations on the undirected graph, namely the so called "natural packing" formulation (see [17]). A similar result do not apply to the md-SCF$_L$ model, as one can easily show with an example.

Using the node-variables k_i, considered in the md-MST problem formulations, we can define two other sets of non-trivial valid inequalities. These are:

1.
$$x_{ij} \leq k_i \ , \ i,j \in V\backslash\{r\}, \ i \neq j \tag{22}$$
establishing that if there is an outward arc incident in node i, then it must be a central-node, while if i is a leaf-node, then there cannot be any arc leaving node i;

2.
$$\sum_{i\in V} k_i \leq \left\lfloor \frac{n-2}{d-1} \right\rfloor \tag{23}$$
which is based on Corollary 1 presented in Section 3, defining an upper bound to the number of central-nodes in any feasible solution to the md-MST problem.

Note that constraints (22) are not defined for $i = r$, as there is at least one arc diverging from the root in any feasible solution, even when r is a leaf-node.

It can be showed, through an example, that the inequalities (22) and (23) are not redundant in both md-SCF$_L$ and md-MCF$_L$ models. However, a weaker version of inequality (23), defined by $\sum_{i\in V} k_i \leq (n-2)/(d-1)$, is implicitly contained in the two mentioned formulations to the md-MST problem, as described in the next result.

Proposition 7 *Inequality* $\sum_{i\in V} k_i \leq (n-2)/(d-1)$ *is redundant in md-SCF*$_L$ *and md-MCF*$_L$.

Proof: Consider the first inequality in (9b) and in (10b). If we sum all those inequalities, namely the one from (9b) and all in $i \in V\backslash\{r\}$ from (10b), we obtain

$$(d-1)\sum_{i\in V} k_i \leq \sum_{i\in V}\sum_{j\in V\backslash\{i,r\}} x_{ij} - 1 \tag{24}$$

Considering the sum in j of equalities (13),

$$\sum_{j\in V\backslash\{r\}}\sum_{i\in V\backslash\{j\}} x_{ij} = n-1 \tag{25}$$

and substituting (25) in (24), we obtain

$$(d-1)\sum_{i\in V} k_i \leq n-2$$

which is the same as (assuming that $d > 1$)

$$\sum_{i\in V} k_i \leq \frac{n-2}{d-1}$$

showing the intended result. □

The new constraints also indicate that the second inequality in (10b) can be dropped from md-SCF$_L$ and md-MCF$_L$. In fact, for each $j \in V$, if we sum inequalities (22) for all $i \in V \setminus \{j\}$, then we obtain the second inequalities in (10b). Therefore they can be omitted from the models, once (22) is included. Hence, the strengthen versions of both single and multicommodity flow models to the md-MST problem are defined by

md-**SCF1**$_L$: md-SCF$_L$ strength with constraints (22), *md*-**MCF1**$_L$: md-MCF$_L$ strengthen with constraints (22),
md-**SCF2**$_L$: md-SCF1$_L$ strength with inequality (23), *md*-**MCF2**$_L$: md-MCF1$_L$ strengthen with inequality (23).

Contrarily to md-MCF$_L$, the two augmented models md-MCF1$_L$ and md-MCF2$_L$ do not have the symmetry propriety, within LP relaxation. Therefore, they become sensitive to the root node selection, like all the single-commodity flow models here discussed.

As a final observation, according to Proposition 7, the two formulations md-SCF2$_L$ and md-MCF2$_L$ should only be considered when the quotient $(n-2)/(d-1)$ is not integer.

5 Computational results

Computational results for the various models proposed in Section 4 for both the d-MST and the md-MST problems are now reported. These results were obtained using a class of complete graphs with Euclidean and non-Euclidean costs. Table 1 summarizes these instances, which have long been considered by many researchers when addressing the d-MST problem ([16]).

The cost matrix for each CRD instance was taken as the Euclidean distance between the coordinates of n points, randomly generated using an uniform distribution in a square. These have been used by Narula and Ho [20], Savelsbergh and Volgenant [25], Volgenant [27], Krishnamoorthy, Ernst and Sharaiha [16] and Ribeiro and Souza [23]. The SYM instances are analogous to the CRD problems but with points generated in a higher dimensional Euclidean space. These instances have also been used in Volgenant [27], Krishnamoorthy, Ernst and Sharaiha [16] and Ribeiro and Souza [23]. In our experiments and among the CRD and SYM cases, we have only considered the first three instances in each class.

Krishnamoorthy, Ernst and Sharaiha [16] also propose another class of non-Euclidean instances, labeled SHRD. These have been particularly constructed in order to be harder than the previous ones to the d-MST problem. This class of instances has not been considered for the md-MST problem, because the correspondent optimal solutions were very easy to reach, being observed to be stars. Ribeiro and Souza [23] and Andrade, Lucena and Maculan [1] have also included this class of instances in their tests on the d-MST.

The node-degree parameter values considered in our tests are also reported in Table 1. In these tests we only use common degree constraints to all nodes, i.e, $d(i) = d$ for all $i \in V$.

			d-MST problem			md-MST problem		
Instances	n	Num. of inst.	$d = 2$	$d = 3$	$d = 5$	$d = 3$	$d = 5$	$d = 10$
CRD300, SYM300	30	3	√	√		√	√	
CRD500, SYM500	50	3	√	√		√	√	√
SHRD15	15	2	√	√	√			
SHRD20	20	2	√	√	√			
SHRD25	25	2	√	√	√			
SHRD30	30	2	√	√	√			

Table 1: Instances characteristics.

All computational experiments have been performed on a PC with a 3.2 GHz Intel Pentium-4 processor and 512 Mbytes of RAM memory, using the CPLEX 9.0 package to solve the LP and the MIP models.

We report results obtained from the LP relaxation of all models under consideration. Computational results from the branch-and-bound execution are also reported. In this case, an upper time limit of 3 hours (10800 seconds) to run the branch-and-bound has been established, and used the default parameter settings within the MIP code of CPLEX. This includes, using the dual simplex to solve all subproblems, including the starting LP model, best-bound search to select the next node to process when backtracking, and an automatic procedure for the variable selection strategy. Automatic cut generation has also been allowed, having its main phase at the root node of the search tree. For further specificities see [30].

Here, Gap represents the relative deviation percentage, being equal to $100(v^* - v)/v^*$, where v^* is the best known upper bound (or optimum when available) and v is the optimum LP solution value. Time is expressed in seconds. The upper bounds or the optimums were obtained or confirmed by the branch-and-bound. All values presented in the tables are average results taken from the instances in the same class. The correspondent individual results can be found in the Appendix B. Instances that have not reached the optimum within the 3 hours limit are also included in the average time calculation. The branch-and-bound execution time includes the root relaxation solution time.

5.1 Root node selection

Before describing our results we provide a brief discussion on the root node selection. As observed in Section 4, among all formulations here proposed, only the md-MCF$_L$ model as been proved to be insensitive to the root node selection, being a symmetric formulation. This suggests that, for all other models, such selection can be of importance to obtain good lower limits, which may influence the times to reach the optimums. In this discussion we only focus on the md-MST problem, and the question we rise is: "Should we expect to have better quality bounds, produced by the linear relaxation of our models, when the root-node coincides with a central-node? Or should it happen with a leaf-node?". In any case, we should note that in the md-MST problem, the partition of the set of nodes (among central and leaf-nodes) is not known in advance. In order the try to answer this question, we made a few experiences with a 16 nodes instance (taken from instance CRD300, using the first 16 nodes). Table 2 shows, for all possible values for the root, the LP relaxation values and times for both md-SCF$_L$ and md-MCF$_L$ models. The same table provides the times to reach the optimums through the branch-and-bound. These tests have been applied to $d = 3$ and $d = 5$. Lines in gray correspond to the central-nodes in the correspondent optimal solutions. The optimum values are 2860 and 3563, for $d = 3$ and $d = 5$, respectively.

As expected, the optimum solution values obtained from the LP relaxation of model md-SCF$_L$ vary when different nodes are selected for the root, while the bounds from md-MCF$_L$ are always the same. These results, namely those from md-SCF$_L$, do not reveal any evidence relating the type of root-node (whether is is a central or a leaf-node) and the quality of the LP bounds it reaches. In fact, when the root is a central node (lines in gray), the LP bounds obtained for the $d = 3$ case with model md-SCF$_L$ include the lowest ($r = 2$) and one of the

	***d* = 3**						***d* = 5**					
Root	*md* -SCF			*md* -MCF			*md* -SCF			*md* -MCF		
node	Linear relaxation		B-B	Linear relaxation		B-B	Linear relaxation		B-B	Linear relaxation		B-B
(*r*)	LP opt	time	time	LP opt	time	time	LP opt	time	time	LP opt	time	time
1	2115.16	0.02	4.45	2604.00	0.08	2.77	2115.16	0.00	1.28	2604.00	0.08	11.44
2	2016.79	0.00	5.91	2604.00	0.11	1.80	2016.79	0.00	2.52	2604.00	0.11	15.69
3	2020.73	0.02	8.47	2604.00	0.11	2.08	2020.73	0.02	2.25	2604.00	0.11	10.11
4	2169.03	0.00	3.27	2604.00	0.08	2.55	2169.03	0.00	1.38	2604.00	0.08	7.58
5	2118.56	0.00	2.58	2604.00	0.08	2.44	2118.56	0.00	1.75	2604.00	0.06	17.38
6	2251.83	0.00	3.33	2604.00	0.08	3.42	2251.83	0.02	1.48	2604.00	0.08	14.47
7	2118.56	0.02	4.83	2604.00	0.06	3.59	2118.56	0.00	1.72	2604.00	0.06	8.53
8	2148.33	0.02	1.83	2604.00	0.06	2.55	2148.33	0.02	1.81	2604.00	0.05	14.38
9	2148.33	0.02	2.77	2604.00	0.08	2.73	2148.33	0.00	1.45	2604.00	0.08	12.14
10	2169.03	0.00	2.20	2604.00	0.08	1.98	2169.03	0.01	1.28	2604.00	0.08	11.34
11	2169.03	0.02	2.34	2604.00	0.08	2.64	2169.03	0.00	1.64	2604.00	0.06	8.98
12	2251.83	0.02	2.39	2604.00	0.08	3.00	2251.83	0.00	1.33	2604.00	0.08	19.70
13	2115.16	0.00	4.11	2604.00	0.06	3.09	2115.16	0.00	1.83	2604.00	0.06	11.97
14	2148.33	0.00	2.52	2604.00	0.05	2.27	2148.33	0.02	1.78	2604.00	0.05	7.25
15	2143.76	0.02	3.23	2604.00	0.09	2.44	2143.76	0.02	1.64	2604.00	0.06	16.05
16	2028.36	0.00	3.83	2604.00	0.11	3.05	2028.36	0.00	1.38	2604.00	0.09	6.66

Table 2. LP relaxation bounds and times, and branch-and-bound times for models d-SCF and d-MCF for all $r \in V$, using a 16 nodes instance taken from CRD300. Times are expressed in seconds.

highest ($r = 12$) results. This conclusion can be extended to the $d = 5$ case, which doesn't give us much information to help choosing the root. For this reason and without further evidences we choose the first node for the root ($r = 1$) in all our tests.

The need to select a root node could have been avoid if other formulations were used, namely if we consider an undirected formulation to our problem. Another suggestion comes from the work of Gouveia and Telhada [12] addressing a Two-Level Network Design Problem (TLND). In this work, the authors propose a very interesting symmetric and compact formulation to the mentioned problem, characterizing the intersected polyhedra off all LP relaxation models to all possible root-nodes. A similar idea could be considered within the md-MST problem. However, we should not ignore that in the TLND we know in advance the partition of the set of nodes, that is, which nodes are primary and which are secondary nodes.

The use of a symmetric formulation, namely model md-MCF$_L$, could make us believe to be irrelevant to choose any node for the root. This is true, when we only look to the bounds produced by those models, however, this may not be irrelevant when looking for the time those models take to reach the optimums. In fact, the last column in Table 2 (column B-B times for $d = 5$) shows that the time to solve the branch-and-bound with the symmetric model md-MCF$_L$ is quite different, when different nodes are chosen for the root. Those times range from 6.66 to 19.7 seconds. Note that this may not have to do with the models in their own, but rather with the method being used to solve the integer problems. Actually, we are not using a pure branch-and-bound, but instead a branch-and-cut, benefiting from the pool of general cuts that CPLEX provides. Therefore, the original symmetric formulations used to start the method may loose this propriety after adding the mentioned cuts. Furthermore, the separation procedures used by the algorithm to identify those cuts may be sensitive to the node selected for the root. This suggests that a deeper analyze should be performed when addressing network design problems using symmetric and non-symmetric formulations.

5.2 d-MST problem results

Considering the d-MST problem, Tables 3 and 4 summarize the results obtained with both d-SCF$_L$ and d-MCF$_L$ formulations proposed in Section 4.

Type	n	d	LP relaxation				Branch-and-bound	
			Gap		Time		Time	
			d-SCF	d-MCF	d-SCF	d-MCF	d-SCF	d-MCF
CRD	30	2	10.69	0.37	0.05	9.98	379.87	11.82
CRD	30	3	16.15	0.00	0.05	5.17	1209.59	5.50
CRD	50	2	18.17	0.31	0.59	299.21	10800.00	2205.74
CRD	50	3	21.15	0.00	0.58	70.93	10800.00	73.09
SYM	30	2	16.30	0.11	0.07	9.47	129.62	26.31
SYM	30	3	16.33	0.00	0.05	1.83	97.31	2.14
SYM	50	2	18.88	0.08	0.63	351.40	10800.00	1170.14
SYM	50	3	12.82	0.00	0.64	27.18	1643.08	29.34

Table 3. Average results obtained with formulations d-SCF and d-MCF, addressing the d-MST problem and applied to CRD and SYM instances.

The results presented in Table 3 clearly show a shorter LP duality gap produced by the d-MCF$_L$ model when compared with the gap generated by the d-SCF$_L$ LP formulation. This is most evident when d increases, where the d-MCF$_L$ formulation is sufficient to reach the optimum, namely to the $d = 3$ problem and among the instances under consideration. On the other hand, model d-MCF$_L$ requires much more time to be solved, although the strong lower bound it produces seems to be an important advantage within the branch-and-bound execution, at least for the 30 and 50 nodes instances considered in our tests. In fact, among all CRD and SYM instances used, model d-SCF fails the optimum in 9 out of 24 instances, while the d-MCF model has reached them all. This is most evident for the higher dimensional problems.

It also appears to be harder to solve the d-MST for $d = 2$, confirming an observation mentioned in [10] and [1]. Note that most works addressing this problem do not present results for $d = 2$, where the d-MST corresponds to an Hamiltonian path on G.

The results obtained with the SHRD instances, reported in Table 4, confirm most of the previous observations. Remember that these instances, proposed in [16], have been constructed to be harder than the CRD and SYM. This hardness is not observed when the problem is addressed with the two flow formulations under consideration.

Type	n	d	LP relaxation				Branch-and-bound	
			Gap		Time		Time	
			d-SCF	d-MCF	d-SCF	d-MCF	d-SCF	d-MCF
SHRD	15	2	0.57	0.00	0.01	0.11	0.33	0.19
SHRD	15	3	0.59	0.00	0.01	0.07	0.08	0.09
SHRD	15	5	0.65	0.00	0.00	0.03	0.02	0.05
SHRD	20	2	0.48	0.00	0.03	1.13	2.85	1.31
SHRD	20	3	0.60	0.00	0.02	0.12	1.19	0.31
SHRD	20	5	0.63	0.00	0.01	0.16	0.24	0.23
SHRD	25	2	0.45	0.00	0.03	2.97	10.78	3.13
SHRD	25	3	0.42	0.00	0.03	1.14	2.89	1.29
SHRD	25	5	0.44	0.00	0.03	0.24	0.26	0.42
SHRD	30	2	0.14	0.00	0.05	7.16	25.50	7.46
SHRD	30	3	0.24	0.00	0.06	7.18	5.18	7.51
SHRD	30	5	0.25	0.00	0.06	1.34	0.97	1.66

Table 4. Average results obtained with formulations d-SCF and d-MCF, addressing the d-MST problem and applied to the SHRD instances.

While Krishnamoorthy, Ernst and Sharaiha [16] solve the CRD and SYM instances much faster than we do, this is no longer observed for the SHRD class. For this class of instances, our results are similar than those reported by Andrade, Lucena and Maculan in [1]. These authors do not present results for the CRD and SYM instances, although solving much higher dimensional problems. Our work is not devoted to the d-MST problem but rather to the new md-MST. In fact, we use the more classical problem to emphasize what we believe to be the empirical hardness of the new problem. For this reason, we do not promote a deeper comparison of our results with those reported in other works in the literature, within the d-MST problem.

5.3 md-MST problem results

Now we consider the md-MST problem, to which summarized results are presented in Tables 5 and 6. We start by analyzing the LP relaxation versions of the six formulations proposed in Section 4, namely md-SCF, md-SCF1, md-SCF2, md-MCF, md-MCF1 and md-MCF2. These tests involve the CRD and SYM instances, for $n = 30$ and 50, as reported in Table 1.

The average results presented in Table 5 show very high duality gaps produced by both md-SCF$_L$ and md-MCF$_L$ models. This can be explained by Proposition 6, which shows the linear relaxation of the two mentioned formulations to the md-MST problem to produce the same lower limits as the correspondent formulations to the MST problem, without degree constraints. Therefore, we should expect poor bounds from these models, especially when d increases, that is, when the md-MST withdraws the MST problem. The insensitivity of models md-SCF$_L$ and md-MCF$_L$ to the d parameter value, induced by Proposition 6, can also be observed in Tables B3 and B4 in the Appendix.

Type	n	d	LP relaxation											
			Gap						Time					
			md-SCF	*md*-SCF1	*md*-SCF2	*md*-MCF	*md*-MCF1	*md*-MCF2	*md*-SCF	*md*-SCF1	*md*-SCF2	*md*-MCF	*md*-MCF1	*md*-MCF2
CRD	30	3	24.95	16.54		10.05	5.18		0.06	0.12		4.20	16.76	
CRD	30	5	39.77	11.38		27.81	8.28		0.06	0.20		3.99	30.30	
CRD	50	3	30.56	15.02		11.92	6.53		0.59	2.56		72.16	428.52	
CRD	50	5	45.69	12.86		31.21	10.04		0.59	2.14		71.92	381.00	
CRD	50	10	59.63	8.82	7.95	48.83	8.35	7.50	0.59	3.73	2.91	71.71	581.04	453.21
SYM	30	3	26.07	9.98		15.37	3.67		0.06	0.09		0.98	9.78	
SYM	30	5	49.02	8.50		41.68	6.90		0.06	0.14		0.96	17.26	
SYM	50	3	22.96	8.80		12.57	3.38		0.66	2.56		55.82	281.24	
SYM	50	5	50.39	13.68		43.70	9.71		0.67	3.32		42.29	393.57	
SYM	50	10	76.44	17.73	16.11	73.28	17.24	15.62	0.66	4.36	3.17	47.31	552.10	339.89

Table 5. Average results obtained with the LP relaxation of formulations *md*-SCF, *md*-SCF1, *md*-SCF2, *md*-MCF1, *md*-MCF2, for the *md*-MST problem and applied to CRD and SYM instances.

Another interesting observation involves the results taken with models md-SCF1$_L$ and md-MCF1$_L$, corresponding to the md-SCF$_L$ and md-MCF$_L$ formulations strengthened with constraints (22). These inequalities seem to be very effective, especially for higher values of d. This is because constraints (22) imply $\max_{j\in V\setminus\{i,r\}} \{x_{ij}\} \leq k_i$ (for all $i \in V \setminus \{r\}$), meaning that, when i is a central-node, variables k_i are being "pumped up" and forced to take higher values. As a consequence and still within linear relaxation, the linking constraints (10b) become tighter, thus more effective to answer the md-MST problem. The added constraints also allow formulation md-SCF1$_L$ to produce much better lower limits than md-MCF$_L$, requiring much smaller execution times. In fact, the number of constraints in the augmented model md-SCF1 is $\mathcal{O}(n^2)$, being smaller than the $\mathcal{O}(n^3)$ constraints in md-MCF.

The gaps obtained with models md-SCF2$_L$ and md-MCF2$_L$ indicate that the single constraint (23) effectively cuts the md-SCF1$_L$ and md-MCF1$_L$ polyhedrons. This inequality as only been applied for $d = 10$, as explained at the end of Section 4, being able to shorten the gaps in both strength LP formulations. As observed in Table 6 and for $d = 10$, the added constraints (22) and (23) allowed model md-SCF2$_L$ to reach the optimums much faster than the weaker models md-SCF$_L$ and md-SCF1$_L$. In general, the "lighter" single-commodity flow formulations have been able to solved the problem much faster than the more disaggregated multicommodity flow models, specially for higher values of d.

It is important to mention that the md-MST problem has not been proved to be NP-hard for $d = 3$ (see Section 2). However, the duality gaps observed to this particular case are still very high, even among the stronger models, bringing some curiosity about its theoretical hardness conjecture.

Looking to the average execution times presented in Table 6, it appears to be more difficult

Type	n	d	Branch-and-bound					
			Time					
			md-SCF	*md*-SCF1	*md*-SCF2	*md*-MCF	*md*-MCF1	*md*-MCF2
CRD	30	3	7996.47	5685.58		1060.26	2693.06	
CRD	30	5	490.66	589.54		8680.02	7256.71	
CRD	50	3	10800.00	10800.00		10800.00	10800.00	
CRD	50	5	10800.00	10800.00		10800.00	10800.00	
CRD	50	10	4194.63	5799.30	614.01	10800.00	10800.00	10800.00
SYM	30	3	14.14	18.35		94.35	85.44	
SYM	30	5	17.11	21.63		859.91	584.04	
SYM	50	3	4682.57	5532.25		4585.68	4465.08	
SYM	50	5	7946.45	7344.15		10800.00	10800.00	
SYM	50	10	2333.61	2971.98	460.68	10800.00	10800.00	10800.00

Table 6. Average results obtained with the branch-and-bound, when applied to md-SCF, md-SCF1, md-SCF2, md-MCF, md-MCF1 and md-MCF2, addressing the md-MST problem and applied to CRD and SYM instances.

to reach the optimums among the CRD instances. Remember that these instances have the nodes located in the plane and distances are Euclidean. This observation may have other causes than just the problem itself, where the type of formulations being used may also be of influence. Note that similar results can be observed in problem d-MST (see Table 3), contradicting those reported in [16].

In general, and comparing this problem with the previous one, it appears to be harder to reach the optimum in a md-MST problem, rather than in a d-MST, at least among the instances under consideration and using flow based formulations. This difference is particularly notorious with the multicommodity flow model, with which very strong lower bounds were obtained for the d-MST problem, using linear relaxation. Most of those bounds were proved to be optimums. On the contrary, even with the strengthened multicommodity flow formulations, the md-MST problem still exhibits very high LP gaps.

To conclude, note that among the higher dimensional instances, with $n = 50$, many optimums are still unconfirmed. Specially within the CRD class. For $d = 10$, the multicommodity flow formulations have not been able to reach any optimum, while the md-SCF2 model showed to be the more effective formulation, reaching the goal in all instances, much faster than the other models (see Table B5 in Appendix B).

6 Conclusions

This paper presents a new degree constrained spanning tree problem, involving a minimum degree constraint on the nodes. The new md-MST problem is closely related with the well known d-MST, where the degree constraint is an upper limit instead.

We have discussed the md-MST theoretical complexity, showing the problem to be NP-hard for $d \geq 4$, being open for $d = 3$. Some proprieties have been presented, namely defining upper and lower limits to the number of central-nodes (or to the number of leaf-nodes) in any md-MST feasible solution.

Flow based formulations to this problem and to the more classic d-MST were also described. When considering the md-MST, the lower bounds produced by the LP relaxation of the stronger multicommodity flow formulation is very far from the best known upper bounds (or optimums when available). In fact, those bounds coincide with the ones produced by the unconstrained version of the problem (MST), using a similar formulation, as proved in Section 4. The same type of model reach the optimum in most of the cases, when applied to the d-MST problem, still within LP relaxation. This observation may indicate that we are dealing with a difficult problem from an empirical stand point, or that the models being used are not sufficient to approximate the md-MST integer polyhedron. In fact, even after strengthening the multicommodity flow formulations with additional cuts, described in Section 4, the LP relaxation results still kept a significant duality gap.

All these results indicate that there is still work to do on the md-MST problem. One possible idea consists in using other characterizations of the set X of all spanning trees, namely considering natural formulations. Another thought involves diminishing the dimension of our models, namely by reducing the number of variables. Some of these suggestions may possibly allow to address the higher dimensional instances, that have not been considered in this paper.

It has also been observed that the "lighter" single-commodity flow formulations become very competitive when strengthen with the additional constraints $x_{ij} \leq k_i$ (for all $i, j \in V, i \neq j$) and $\sum_{i \in V} k_i \leq \left\lfloor \frac{n-2}{d-1} \right\rfloor$. The later inequality has been derived from a propriety stated in Section 3.

References

[1] R. Andrade, A. Lucena and N. Maculan, Using Lagrangian dual information to generate degree constrained spanning trees, *Discrete Applied Mathematics*, 154(5): 703-717, 2006.

[2] B. Boldon, N. Deo and N. Kumar, Minimum-weight degree-constrained spanning tree problem: Heuristics and implementation on an SIMD parallel machine, *Parallel Computing*, 22: 369-382, 1996.

[3] L. Caccetta and S.P. Hill, A branch and cut method for the degree-constrained minimum spanning tree problem, *Networks*, 37(2): 74-83, 2001.

[4] D. Cieslik, The vertex degrees of minimum spanning trees, *European Journal of Operational Research*, 125: 278-282, 2000.

[5] G. Craig, M. Krishnamoorthy and M. Palaniswami, Comparison of heuristic algorithms for the degree constrained minimum spanning tree, In I.H. Osman and J.P. Kelly (eds.), Metaheuristics: Theory and Applications, *Kluwer, Boston*, 1996.

[6] N. Deo and S.L. Hakimi, The shortest generalized Hamiltonian tree, In Proceedings of the 6th Annual Alberton Conference, 879-888, 1968.

[7] S.P. Fekete, S. Khuller, M. Klemmstein, B. Raghavachari and N. Young, A network-flow technique for finding low-weight bounded-degree spanning trees, *Journal of Algorithms*, 24: 310-324, 1997.

[8] H.N. Gabow and R.E. Tarjan, Efficient algorithms for a family of matroid intersection problems, *Journal of Algorithms*, 5: 80-131, 1984.

[9] M.R. Garey and D.S. Johnson, Computers and intractability: A guide to the theory of NP-completeness, *W.H. Freeman, San Francisco*, 1979.

[10] B. Gavish, Topological design of centralized computer networks: Formulations and algorithms, *Networks*, 12: 355-377, 1982.

[11] B. Gavish, Topological design of centralized computer networks: the overall design problem, *European Journal of Operational Research*, 58: 149-172, 1992.

[12] L. Gouveia and J. Telhada, An augmented arborescence formulation for the two-level network design problem, *Annals Operations Research*, 106(1-4): 47-61, 2001.

[13] C. Imielińska, B. Kalantari, and L. Khachiyan A greedy heuristic for a minimum-weight forest problem, *Operations Research Letters*, 14: 65-71, 1993.

[14] S. Khuller, B. Raghavachari and N. Young, Low degree spanning trees of small weight, *SIAM Journal on Computing*, 25: 355-368, 1996.

[15] J.D. Knowles and D.W. Corne, A new evolutionary approach to the degree-constrained minimum spanning tree problem, *IEEE Transactions on Evolutionary Computation*, 4(2): 125-134, 2000.

[16] M. Krishnamoorthy, A.T. Ernst and Y.M. Sharaiha, Comparison of algorithms for the degree constrained spanning tree, *Journal of Heuristics*, 7: 587-611, 2001.

[17] T. Magnanti and L. Wolsey, Optimal Trees, In M.O. Ball, T.L. Magnanti, C.L. Monma and G.L. Nemhauser (eds.), Network Models, Handbooks in Operations Research and Management Science, Vol.7, *North-Holland, Amsterdam*, Chapter 9: 503-615, 1995.

[18] C.L. Monma and D. Shallcross, Methods for designing communication networks with certain two-connected survivability constraints, *Operations Research*, 37: 531-541, 1989.

[19] C. Monma and S. Suri, Transitions in geometric minimum spanning trees, *Discrete Computational Geometry*, 8: 265-293, 1992.

[20] S.C. Narula and C.A. Ho, Degree constrained minimum spanning tree, *Computers and Operations Research*, 7: 239-249, 1980.

[21] C.H. Papadimitriou, Computational Complexity, *Addinson–Wesley*, 1994

[22] C.H. Papadimitriou and U.V. Vazirani, On two geometric problems related to the traveling salesman problem, *Journal of Algorithms*, 5: 231-246, 1984.

[23] C.C. Ribeiro and M.C. Souza, Variable neighborhood search for the degree-constrained minimum spanning tree problem, *Discrete Applied Mathematics*, 118(1-2): 43-54, 2002.

[24] G. Robins and J.S. Salowe, Low-degree minimum spanning trees, *Discrete Computational Geometry*, 14: 151-166, 1995.

[25] M. Savelsbergh and T. Volgenant, Edge exchanges in the degree-constrained minimum spanning tree problem, *Computers and Operations Research*, 12: 341-348, 1985.

[26] M. Stoer, Design of survivable networks, Lecture Notes on Mathematics, *Springer, Heidelberg*, Vol.1531: 1992.

[27] T. Volgenant, A Lagrangean approach to the DCMST problem, *European Journal of Operational Research*, 39: 325-331, 1989.

[28] G. Zhou and M. Gen, Approach to the degree-constrained minimum spanning tree problem using genetic algorithms, *Engineering Design and Automation*, 3(2): 156-165, 1997.

[29] G. Zhou and M. Gen, A note on genetic algorithms for degree constrained minimum spanning tree problems, *Networks*, 30: 105-109, 1997.

[30] ILOG/CPLEX 9.0, Reference Manual, 1997.

Appendix A

Cost matrix C used in the example in Figure 4.

$$C = \begin{bmatrix} - & 2 & 2 & 3 & 3 & 1 & 3 & 3 \\ & - & 1 & 1 & 1 & 3 & 3 & 3 \\ & & - & 3 & 3 & 3 & 1 & 1 \\ & & & - & 3 & 3 & 3 & 3 \\ & & & & - & 3 & 3 & 3 \\ & & & & & - & 3 & 3 \\ & & & & & & - & 3 \\ & & & & & & & - \end{bmatrix}$$

Appendix B

The next five tables describe the individual results taken from the computational experiments proposed in Section 5.

The following notation was used in the tables:

1. Underlined values in column "Opt/UB" correspond to the lowest upper bound found. All other values in this column are known optimums;
2. Underlined values in the columns below "B&B times" indicate that the optimum has not been reached, and the underlined value is the maximum allowed execution time of the branch-and-bound;
3. Time is expressed in seconds.

Note that the branch-and-bound execution time includes the root relaxation solution time.

***d*-MST**				Opt/UB	LP relaxation optimums		LP relaxation times		B&B times	
Type	*n*	*d*	*I*		*d*-SCF_L	*d*-MCF_L	*d*-SCF_L	*d*-MCF_L	*d*-SCF	*d*-MCF
CRD	30	2	1	3822	3517.95	3822.00	0.05	16.42	60.31	16.75
			2	3618	3101.56	3577.50	0.05	4.78	193.80	9.66
			3	4221	3805.93	4221.00	0.06	8.75	885.50	9.06
Average gaps and times					*10.688*	*0.373*	*0.05*	*9.98*	*379.87*	*11.82*
SYM	30	2	1	1376	1172.24	1376.00	0.05	8.44	35.20	8.75
			2	1637	1360.29	1637.00	0.06	8.31	70.86	8.63
			3	1973	1633.73	1966.50	0.09	11.66	282.80	61.56
Average gaps and times					*16.302*	*0.110*	*0.07*	*9.47*	*129.62*	*26.31*
CRD	30	3	1	3634	3157.68	3634.00	0.05	7.03	655.64	7.39
			2	3277	2732.32	3277.00	0.06	3.95	506.14	4.28
			3	4001	3251.49	4001.00	0.05	4.52	2466.99	4.84
Average gaps and times					*16.154*	*0.000*	*0.05*	*5.17*	*1209.59*	*5.50*
SYM	30	3	1	1012	787.16	1012.00	0.05	2.39	267.92	2.70
			2	1285	1074.02	1285.00	0.05	2.34	12.06	2.66
			3	1311	1175.38	1311.00	0.05	0.77	11.94	1.06
Average gaps and times					*16.327*	*0.000*	*0.05*	*1.83*	*97.31*	*2.14*
CRD	50	2	1	5312	4503.50	5312.00	0.56	227.48	10800.00	229.69
			2	5553	4744.94	5536.00	0.64	301.03	10800.00	468.11
			3	5480	4124.00	5445.50	0.56	369.13	10800.00	5919.42
Average gaps and times					*18.172*	*0.312*	*0.59*	*299.21*	*10800.00*	*2205.74*
SYM	50	2	1	1759	1462.29	1754.91	0.59	543.08	10800.00	2994.94
			2	1586	1298.39	1586.00	0.63	284.27	10800.00	286.45
			3	2116	1658.38	2116.00	0.67	226.86	10800.00	229.03
Average gaps and times					*18.876*	*0.078*	*0.63*	*351.40*	*10800.00*	*1170.14*
CRD	50	3	1	4931	3980.74	4931.00	0.58	120.27	10800.00	122.42
			2	5126	4232.66	5126.00	0.63	54.13	10800.00	56.30
			3	4898	3588.42	4898.00	0.52	38.38	10800.00	40.55
Average gaps and times					*21.145*	*0.000*	*0.58*	*70.93*	*10800.00*	*73.09*
SYM	50	3	1	1156	993.49	1156.00	0.61	15.14	1783.56	17.30
			2	1106	952.07	1106.00	0.63	8.28	1450.13	10.45
			3	1459	1305.91	1459.00	0.69	58.11	1695.56	60.27
Average gaps and times					*12.823*	*0.000*	*0.64*	*27.18*	*1643.08*	*29.34*

Table B1. *d*-MST results to the CRD and SYM instances.

***d*-MST**				Opt	LP relaxation optimums		LP relaxation times		B&B times	
Type	n	d	I		d-SCF$_L$	d-MCF$_L$	d-SCF$_L$	d-MCF$_L$	d-SCF	d-MCF
SHRD	15	2	1	895	888.47	895.00	0.01	0.14	0.38	0.27
			2	904	900.23	904.00	0.00	0.08	0.27	0.11
Average gaps and times					*0.573*	*0.000*	*0.01*	*0.11*	*0.33*	*0.19*
SHRD	15	3	1	582	579.60	582.00	0.01	0.08	0.05	0.09
			2	597	592.48	597.00	0.00	0.05	0.11	0.09
Average gaps and times					*0.585*	*0.000*	*0.01*	*0.07*	*0.08*	*0.09*
SHRD	15	5	1	339	336.43	339.00	0.00	0.02	0.01	0.05
			2	332	330.22	332.00	0.00	0.03	0.03	0.05
Average gaps and times					*0.647*	*0.000*	*0.00*	*0.03*	*0.02*	*0.05*
SHRD	20	2	1	1679	1670.08	1679.00	0.03	1.38	4.48	1.66
			2	1698	1690.78	1698.00	0.02	0.88	1.22	0.95
Average gaps and times					*0.478*	*0.000*	*0.03*	*1.13*	*2.85*	*1.31*
SHRD	20	3	1	1088	1082.62	1088.00	0.01	0.05	0.94	0.36
			2	1092	1084.36	1092.00	0.02	0.19	1.44	0.25
Average gaps and times					*0.597*	*0.000*	*0.02*	*0.12*	*1.19*	*0.31*
SHRD	20	5	1	627	624.51	627.00	0.01	0.22	0.27	0.28
			2	629	623.61	629.00	0.00	0.09	0.20	0.17
Average gaps and times					*0.627*	*0.000*	*0.01*	*0.16*	*0.24*	*0.23*
SHRD	25	2	1	2703	2694.76	2703.00	0.03	1.34	2.70	1.50
			2	2714	2697.92	2714.00	0.03	4.59	18.86	4.76
Average gaps and times					*0.449*	*0.000*	*0.03*	*2.97*	*10.78*	*3.13*
SHRD	25	3	1	1745	1739.52	1745.00	0.03	0.91	0.86	1.06
			2	1756	1746.68	1756.00	0.03	1.36	4.91	1.52
Average gaps and times					*0.422*	*0.000*	*0.03*	*1.14*	*2.89*	*1.29*
SHRD	25	5	1	999	995.25	999.00	0.02	0.23	0.17	0.42
			2	1016	1010.86	1016.00	0.03	0.25	0.34	0.41
Average gaps and times					*0.441*	*0.000*	*0.03*	*0.24*	*0.26*	*0.42*
SHRD	30	2	1	3992	3985.94	3992.00	0.05	5.41	40.95	5.70
			2	3990	3985.04	3990.00	0.05	8.91	10.05	9.22
Average gaps and times					*0.138*	*0.000*	*0.05*	*7.16*	*25.50*	*7.46*
SHRD	30	3	1	2592	2583.59	2592.00	0.06	6.67	8.22	7.00
			2	2585	2580.87	2585.00	0.06	7.69	2.14	8.02
Average gaps and times					*0.242*	*0.000*	*0.06*	*7.18*	*5.18*	*7.51*
SHRD	30	5	1	1504	1497.55	1504.00	0.06	0.73	1.59	1.06
			2	1474	1473.03	1474.00	0.06	1.94	0.34	2.25
Average gaps and times					*0.247*	*0.000*	*0.06*	*1.34*	*0.97*	*1.66*

Table B2. d-MST results to the SHRD instances.

md-MST				Opt	LP relaxation optimums				LP relaxation times				B&B times			
Type	n	d	I		*md*-SCF_L	*md*-$SCF1_L$	*md*-MCF_L	*md*-$MCF1_L$	*md*-SCF_L	*md*-$SCF1_L$	*md*-MCF_L	*md*-$MCF1_L$	*md*-SCF	*md*-SCF1	*md*-MCF	*md*-MCF1
CRD	30	3	1	4026	3140.54	3496.36	3634.00	3761.65	0.06	0.09	5.47	19.31	5664.58	1699.91	1301.20	2006.22
			2	3793	2711.92	3060.07	3277.00	3601.50	0.06	0.13	3.56	20.88	10800.00	10800.00	1289.38	5545.13
			3	4293	3248.14	3557.32	4001.00	4124.50	0.06	0.13	3.56	10.28	7524.84	4556.84	590.19	527.84
Average gaps and times					*24.945*	*16.538*	*10.047*	*5.180*	*0.06*	*0.12*	*4.20*	*16.76*	*7996.47*	*5685.58*	*1060.26*	*2693.06*
SYM	30	3	1	1197	775.04	1039.63	958.00	1112.70	0.06	0.11	1.48	16.39	24.03	22.13	183.17	165.25
			2	1435	1069.76	1302.42	1219.00	1395.25	0.05	0.08	0.47	3.94	9.53	18.22	52.25	46.42
			3	1408	1161.66	1301.54	1252.00	1391.33	0.06	0.09	0.98	9.00	8.86	14.69	47.63	44.64
Average gaps and times					*26.066*	*9.982*	*15.366*	*3.666*	*0.06*	*0.09*	*0.98*	*9.78*	*14.14*	*18.35*	*94.35*	*85.44*
CRD	30	5	1	5026	3140.54	4530.87	3634.00	4626.44	0.06	0.28	4.92	37.75	566.09	806.53	10800.00	9026.89
			2	4648	2711.92	4039.77	3277.00	4294.35	0.06	0.14	3.52	27.97	424.45	359.09	4440.06	1943.23
			3	5425	3248.14	4817.95	4001.00	4922.35	0.06	0.19	3.52	25.17	481.45	603.00	10800.00	10800.00
Average gaps and times					*39.765*	*11376*	*27.814*	*8.275*	*0.06*	*0.20*	*3.99*	*30.30*	*490.66*	*589.54*	*8680.02*	*7256.71*
SYM	30	5	1	1765	775.04	1554.94	958.00	1598.76	0.06	0.16	1.48	29.30	16.53	11.84	1244.56	501.22
			2	2090	1069.76	1896.20	1219.00	1935.14	0.06	0.11	0.44	7.44	31.33	49.02	1079.91	1056.88
			3	2008	1161.66	1921.40	1252.00	1930.10	0.05	0.16	0.97	15.05	3.47	4.03	255.25	194.03
Average gaps and times					*49.017*	*8.496*	*41.682*	*6.903*	*0.06*	*0.14*	*0.96*	*17.26*	*17.11*	*21.63*	*859.91*	*584.04*

Table B3. *md*-MST results to the instances with $n = 30$.

***md*-MST**				Opt/ $\underline{\text{UB}}$	LP relaxation optimums						LP relaxation times					
Type	n	d	I		md-SCF$_L$	md-SCF1$_L$	md-SCF2$_L$	md-MCF$_L$	md-MCF1$_L$	md-MCF2$_L$	md-SCF$_L$	md-SCF1$_L$	md-SCF2$_L$	md-MCF$_L$	md-MCF1$_L$	md-MCF2$_L$
CRD	50	3	1	$\underline{5512}$	3980.74	4710.49		4931.00	5217.94		0.59	2.17		119.00	348.52	
			2	$\underline{5835}$	4225.13	4853.49		5126.00	5365.43		0.63	2.55		66.22	506.91	
			3	$\underline{5635}$	3588.42	4863.22		4898.00	5286.08		0.55	2.95		31.25	430.13	
Average gaps and times					*30.458*	*15.019*		*11.772*	*6.360*		*0.59*	*2.56*		*72.16*	*428.52*	
SYM	50	3	1	1278	949.02	1147.37		1098.00	1223.50		0.64	2.77		22.89	471.17	
			2	1178	910.17	1055.31		1045.00	1127.50		0.66	1.91		16.75	120.74	
			3	1615	1285.51	1521.98		1416.00	1589.50		0.69	3.00		127.83	251.80	
Average gaps and times					*22.960*	*8.799*		*12.566*	*3.377*		*0.66*	*2.56*		*55.82*	*281.24*	
CRD	50	5	1	$\underline{6946}$	3980.74	6018.53		4931.00	6310.88		0.59	2.31		119.31	415.30	
			2	$\underline{7294}$	4225.13	6564.61		5126.00	6692.43		0.66	1.84		65.78	310.47	
			3	$\underline{7525}$	3588.42	6378.72		4898.00	6568.32		0.53	2.28		30.66	417.22	
Average gaps and times					*45.238*	*12.862*		*30.590*	*9.191*		*0.59*	*2.14*		*71.92*	*381.00*	
SYM	50	5	1	$\underline{2054}$	949.02	1697.88		1098.00	1748.77		0.64	3.31		18.78	395.72	
			2	1760	910.17	1526.33		1045.00	1647.89		0.67	3.58		13.20	388.49	
			3	2525	1285.51	2261.70		1416.00	2325.89		0.69	3.08		94.88	396.51	
Average gaps and times					*50.390*	*13.681*		*43.696*	*9.705*		*0.67*	*3.32*		*42.29*	*393.57*	
CRD	50	10	1	9633	3980.74	8462.83	8462.83	4931.00	8539.50	8542.17	0.59	2.94	3.20	118.75	549.38	582.19
			2	9743	4225.13	9067.12	9195.62	5126.00	9077.27	9202.11	0.63	3.81	3.19	65.69	588.73	343.16
			3	9855	3588.42	9128.36	9256.29	4898.00	9179.73	9301.64	0.55	4.45	2.33	30.69	605.02	434.28
Average gaps and times					*59.633*	*8.819*	*7.947*	*48.833*	*8.346*	*7.497*	*0.59*	*3.73*	*2.91*	*71.71*	*581.04*	*453.21*
SYM	50	10	1	4121	949.02	3439.77	3498.10	1098.00	3442.06	3500.07	0.64	5.47	3.50	18.55	540.16	311.95
			2	4166	910.17	3343.44	3387.89	1045.00	3401.09	3444.14	0.64	3.81	2.80	12.97	559.31	290.55
			3	4979	1285.51	4136.34	4255.73	1416.00	4138.13	4258.41	0.69	3.80	3.20	110.41	556.84	417.17
Average gaps and times					*76.435*	*17.733*	*16.106*	*73.278*	*17.241*	*15.622*	*0.66*	*4.36*	*3.17*	*47.31*	*552.10*	*339.89*

Table B4. md-MST results to the CRD and SYM instances with $n = 50$, involving the LP relaxation information of the models under consideration.

md-MST				Opt/UB	LP relaxation optimums						B&B times					
Type	n	d	I		*md*-SCF_L	*md*-$SCF1_L$	*md*-$SCF2_L$	*md*-MCF_L	*md*-$MCF1_L$	*md*-$MCF2_L$	*md*-SCF_L	*md*-$SCF1_L$	*md*-$SCF2_L$	*md*-MCF_L	*md*-$MCF1_L$	*md*-$MCF2_L$
CRD	50	3	1	5512	3980.74	4710.49	—	4931.00	5217.94	—	10800.00	10800.00	—	10800.00	10800.00	—
			2	5835	4225.13	4853.49	—	5126.00	5365.43	—	10800.00	10800.00	—	10800.00	10800.00	—
			3	5635	3588.42	4863.22	—	4898.00	5286.08	—	10800.00	10800.00	—	10800.00	10800.00	—
Average gaps and times					*30.458*	*15.019*	—	*11.772*	*6.360*	—	*10800.00*	*10800.00*	—	*10800.00*	*10800.00*	—
SYM	50	3	1	1278	949.02	1147.37	—	1098.00	1223.50	—	7648.38	4909.45	—	8796.09	9571.03	—
			2	1178	910.17	1055.31	—	1045.00	1127.50	—	5685.80	10800.00	—	2254.31	2684.00	—
			3	1615	1285.51	1521.98	—	1416.00	1589.50	—	713.53	887.31	—	2706.63	1140.22	—
Average gaps and times					*22.960*	*8.799*	—	*12.566*	*3.377*	—	*4682.57*	*5532.25*	—	*4585.68*	*4465.08*	—
CRD	50	5	1	6946	3980.74	6018.53	—	4931.00	6310.88	—	10800.00	10800.00	—	10800.00	10800.00	—
			2	7294	4225.13	6564.61	—	5126.00	6692.43	—	10800.00	10800.00	—	10800.00	10800.00	—
			3	7525	3588.42	6378.72	—	4898.00	6568.32	—	10800.00	10800.00	—	10800.00	10800.00	—
Average gaps and times					*45.238*	*12.862*	—	*30.590*	*9.191*	—	*10800.00*	*10800.00*	—	*10800.00*	*10800.00*	—
SYM	50	5	1	2054	949.02	1697.88	—	1098.00	1748.77	—	10800.00	10800.00	—	10800.00	10800.00	—
			2	1760	910.17	1526.33	—	1045.00	1647.89	—	2239.36	2228.25	—	10800.00	10800.00	—
			3	2525	1285.51	2261.70	—	1416.00	2325.89	—	10800.00	9004.20	—	10800.00	10800.00	—
Average gaps and times					*50.390*	*13.681*	—	*43.696*	*9.705*	—	*7946.45*	*7344.15*	—	*10800.00*	*10800.00*	—
CRD	50	10	1	9633	3980.74	8462.83	8462.83	4931.00	8539.50	8542.17	4072.59	4147.39	828.23	10800.00	10800.00	10800.00
			2	9743	4225.13	9067.12	9195.62	5126.00	9077.27	9202.11	5083.28	10800.00	547.72	10800.00	10800.00	10800.00
			3	9855	3588.42	9128.36	9256.29	4898.00	9179.73	9301.64	3428.01	2450.50	466.09	10800.00	10800.00	10800.00
Average gaps and times					*59.633*	*8.819*	*7.947*	*48.833*	*8.346*	*7.497*	*4194.63*	*5799.30*	*614.01*	*10800.00*	*10800.00*	*10800.00*
SYM	50	10	1	4121	949.02	3439.77	3498.10	1098.00	3442.06	3500.07	1728.45	1110.00	318.86	10800.00	10800.00	10800.00
			2	4166	910.17	3343.44	3387.89	1045.00	3401.09	3444.14	1821.20	4509.52	424.45	10800.00	10800.00	10800.00
			3	4979	1285.51	4136.34	4255.73	1416.00	4138.13	4258.41	3451.19	3296.41	638.73	10800.00	10800.00	10800.00
Average gaps and times					*76.435*	*17.733*	*16.106*	*73.278*	*17.241*	*15.622*	*2333.61*	*2971.98*	*460.68*	*10800.00*	*10800.00*	*10800.00*

Table B5. *md*-MST results to the CRD and SYM instances with $n = 50$, involving the branch-and-bound information using the models under consideration.