INTERNATIONAL TRANSACTIONS IN OPERATIONAL RESEARCH

Intl. Trans. in Op. Res. 19 (2012) 323–352
DOI: 10.1111/j.1475-3995.2011.00830.x

Min-degree constrained minimum spanning tree problem: complexity, properties, and formulations

Ana Maria de Almeida[a], Pedro Martins[b] and Maurício C. de Souza[c]

[a] *CISUC—Centro de Informática e Sistemas da Universidade de Coimbra and Departamento de Matemática, FCTUC, Universidade de Coimbra, Portugal*
[b] *CIO—Centro de Investigação Operacional - FC/UL and ISCAC—Instituto Politécnico de Coimbra, Portugal*
[c] *Departamento de Engenharia de Produção, Universidade Federal de Minas Gerais, Brasil*
E-mail: amca@mat.uc.pt [Ana Maria de Almeida]; pmartins@iscac.pt [Pedro Martins]; mauricio@dep.ufmg.br [Maurício C. de Souza]

Received 26 July 2009; received in revised form 18 August 2011; accepted 23 August 2011

Abstract

This paper addresses a new combinatorial problem, *the Min-Degree Constrained Minimum Spanning Tree (md-MST)*, that can be stated as: given a weighted undirected graph $G = (V, E)$ with positive costs on the edges and a node-degrees function $d : V \to N$, the *md*-MST is to find a minimum cost spanning tree T of G, where each node i of T either has at least a degree of $d(i)$ or is a leaf node. This problem is closely related to the well-known Degree Constrained Minimum Spanning Tree (d-MST) problem, where the degree constraint is an upper bound instead.

The general NP-hardness for the *md*-MST is established and some properties related to the feasibility of the solutions for this problem are presented, in particular we prove some bounds on the number of internal and leaf nodes. Flow-based formulations are proposed and computational experiments involving the associated Linear Programming (LP) relaxations are presented.

Keywords: degree constrained spanning tree problems; computational complexity; single-commodity flow formulations; multicommodity flow formulations

1. Introduction

Let $G = (V, E)$ be a connected weighted undirected graph, where $V = \{1, \ldots, n\}$ is the set of nodes, E the set of edges, and there are positive costs, c_e, associated to each edge $e \in E$. Denote by $T = (V, E_T)$ a spanning tree of G and $deg_T(i)$ the degree of node i in T. Given a positive integer valued function $d : V \to N$, the *Min-Degree Constrained Minimum Spanning Tree (md-MST)* problem is to find a spanning tree T of G with minimum total edge cost, such that each

Published by Blackwell Publishing, 9600 Garsington Road, Oxford, OX4 2DQ, UK and 350 Main St, Malden, MA 02148, USA.

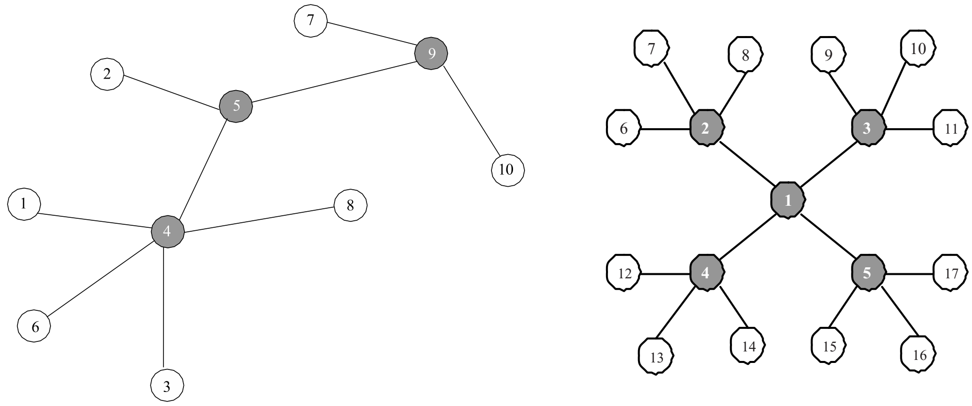

Fig. 1. Two feasible solutions of the *md*-MST problem: for all $i \in V$, $d(i) = 3$ in the solution on the left, while $d(i) = 4$ in the solution on the right.

node i in the tree either has a degree of at least $d(i)$ or is a leaf node, that is, either $deg_T(i) \geq d(i)$ or $deg_T(i) = 1$. As usual, total edge cost of T is given by $\sum_{e \in E_T} c_e$.

Figure 1 shows two feasible solutions of the *md*-MST problem. Considering the same value for all $i \in V$, the solution on the left has degree bound of $d(i) = 3$, and the one on the right has degree bound of $d(i) = 4$. If node 2 is removed from the solution from the left, then it becomes infeasible, because node 5 would have degree less than 3 without being a leaf.

This problem is closely related to the well-known Degree Constrained Minimum Spanning Tree (d-MST) problem (also known as the Bounded Degree Minimum Spanning Tree), where a minimum cost spanning tree T of G is also sought, but where each node $i \in T$ must have a degree of at most $d(i)$.

In what follows, we refer to degree d when all nodes have equal bound degree restrictions, that is, when $d(i) = d$ for all $i \in V$. As usual, MST stands for Minimum (cost) Spanning Tree.

The *md*-MST was first introduced in Almeida et al. (2006). In that work, we provided a broader discussion on the relation between the *md*-MST and the d-MST problems. In this paper, we concentrate on our research results on the *md*-MST problem.

Practical applications to the *md*-MST problem may occur in cases where one needs to identify a set of locations (nodes) that should act as central incidence nodes for other entities (peripheral nodes), in such a way that a location (node) can only assume a central status if it has been assigned to, at least, d other entities (nodes). Otherwise, it must be a terminal 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 on the edges are being considered. Therefore, the internal nodes in an *md*-MST solution can represent central distribution places or centralized communication devices, while the terminal or leaf nodes act as individual consumers or clients.

When $d = 1$, the d-MST problem is impossible to solve, while for $d = 2$ the problem is to find a minimum cost Hamiltonian path in G. Based on this, Garey and Johnson (1979) have shown 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 (see, e.g., Papadimitriou and Vazirani, 1984; Savelsbergh and Volgenant, 1985; Monma and Suri, 1992; Robins and Salowe, 1995).

Besides the theoretical hardness associated to both problems, the comparison results discussed in Almeida et al. (2006) suggest 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, in hub location we deal with a network design flow feasibility problem, while the *md*-MST only addresses the network design structure. Furthermore, in the *md*-MST we always have a spanning tree linking the central-nodes (as will be observed in Section 2), while in most hub location problems the hub nodes are fully interconnected.

The *md*-MST can also be related to the minimum spanning tree problem with a constraint on the number of leaves, where a given fixed number of leaves ($k \in \{1, \dots, n\}$) is imposed on any feasible tree (see, e.g., Fernandes and Gouveia, 1998).

The next section presents some properties and bounds to characterize *md*-MST feasible solutions. In Section 3, we prove that the general *md*-MST problem is NP-hard. In Section 4, different formulations for the *md*-MST are proposed, considering directed flow-based models. Undirected models are also given. Computational results that compare the LP bounds produced by the various directed formulations are presented in Section 5. The last section is devoted to concluding remarks.

2. Bounding characterization of feasible *md*-MST solutions

In what follows, we define T_C as the subgraph of a spanning tree T of G obtained after eliminating all the leaf-nodes and all the edges incident to those nodes in T. Hence, the nodes in T_C are in the set $V_C \subset V$ and V_L is the set of nodes eliminated from T (leaf-nodes), with $V_C \cup V_L = V$ and $V_C \cap V_L = \emptyset$. 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. However, the cost of the subtree T_C of an optimal solution T^* to the *md*-MST problem might not be a minimum cost spanning tree for G_C. Consider the following example where an instance of the *md*-MST problem is given, where the edges of the tree are labeled with the corresponding costs c_e (The corresponding cost matrix C can be found in Appendix A.).

The example in Fig. 2 shows that the cost of subtree T_C is equal to 4, while the minimum cost spanning tree of G_C is equal to 3.

The following results relate the constraint on the node-degrees with the feasibility of both d-MST and *md*-MST instances, namely by establishing upper and lower bounds on the number of leaf-nodes in any feasible spanning tree for each problem. We assume that $|V| > 2$.

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

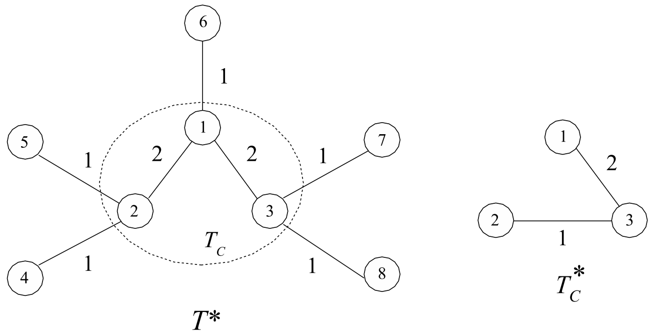

Fig. 2. The optimal *md*-MST tree T^* for a given graph G (see Appendix A) with $n = 8$ and $d = 3$, and the optimal spanning tree T_C^* of G_C. The labels on the edges represent their associated costs.

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

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

since $deg_T(i) = 1$ for all $i \in L$.

By definition, all nonleaf nodes of T have minimum degree at least d, then we have

$$d(n - |L|) \le 2(n-1) - |L|,$$

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

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

□

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

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

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

$$n - \left\lfloor \frac{n-2}{d-1} \right\rfloor \le |L| \le n - 1. \tag{4}$$

From a different perspective, we can also define lower and upper bounds on the number of central nodes in any feasible *md*-MST T. Therefore, considering $S = V \setminus L$ as the set of central nodes in T, we have $|S| = n - |L|$. Using inequality (2), we obtain

$$1 \le |S| \le \left\lfloor \frac{n-2}{d-1} \right\rfloor. \tag{5}$$

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* $1 \leq |S| \leq \lfloor (n-2)/(d-1) \rfloor$.

3. The computational complexity of the *md*-MST problem

This section establishes the complexity classification for the general *md*-MST problem. We begin by analyzing two particular cases for this problem and then proceed to a more general complexity classification.

3.1. Particular characterizations

We introduce two particular *md*-MST instances (by restricting the values of d) which can be easily solved.

Proposition 2. *If $d \leq 2$, the md-MST and MST problems are equivalent.*

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.

Additionally, considering that we need at least d nodes adjacent to any central node, then if d is larger than $|V|/2$, any *md*-MST feasible solution has no more than a single central node, as established by 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.*

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

3.2. Complexity of the general md-MST Problem

Let us consider the case that $\lfloor n/2 \rfloor \geq d \geq 4$. Using a constructive approach, analogous to the proof of Partition into Triangles found in Garey and Johnson (1979), and inspired by the NP-hardness proof in Imielińska et al. (1993), we reduce the k-Dimensional Matching Problem (kDM) to the *md*-MST Problem.

The kDM (where $k \geq 3$) in Papadimitriou (1994) 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 a 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

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

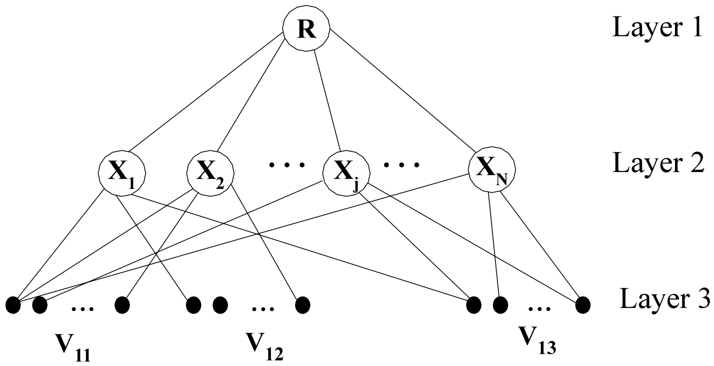

Fig. 3. Example for the undirected weighted graph G.

The kDM, for $k \geq 3$, is an NP-hard (Papadimitriou, 1994) 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 $\lfloor n/2 \rfloor \geq d \geq 4$, the general md-MST problem is NP-hard.*

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

Assume that $d = 4$. Using a general instance I for the 3-Dimensional Matching Problem (3DM; Garey and Johnson, 1979), we construct a new instance graph G for the m4-MST Problem. Next, we will prove that I has a 3D matching if and only if G has an optimal solution for the m4-MST, that is, an 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 pair which stands as a possible instantiation for a 3DM problem.

Now we 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 each of the 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 Fig. 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.

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 the construction will run in time $\mathcal{O}(q^3)$.

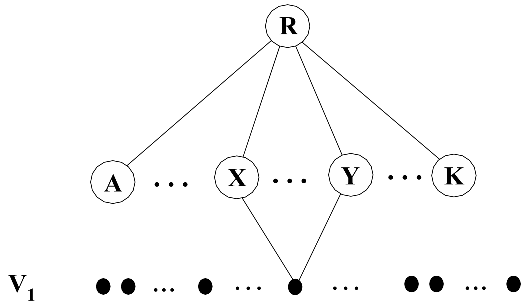

Fig. 4. Example that implies that a spanning tree for G exists only if I has one perfect 3DM.

Note that, because of the weights associated to the different edges, any optimal minimum cost spanning tree for G must have total weight $w^* = 3q$.

A spanning tree T for G exists if, at least, a perfect 3D matching exists. Otherwise, there should be at least one node on the bottom layer that either is not connected to any one of the middle layer nodes and, therefore, graph G would be disconnected, or there would be two edges from different Xs connecting the same third layer node within T, and therefore T would have a cycle and could not be a tree (see Fig. 4). This means that there is a feasible solution for m4-MST in G whenever I admits (at least) one 3D matching and vice-versa.

Let us now prove that if I admits a 3D matching then G has an optimal md-MST tree, that is a minimum spanning tree where the minimum degree for 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 perfect 3DM matching for I. Therefore, $|\mathcal{M}| = q \leq N$ and each triplet is unique in the sense that all the $3q$ nodes of V_1 are covered by the q triplets, which implies that all the a_i, b_i, and c_i are different ($i = 1, \ldots, q$).

Let $T = (V, E_T)$ 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 the nodes of the last triplet of $\mathcal{M}$ are connected to node X_{j_q}. That is:

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

and the sets X_{j_1} to X_{j_q} form a perfect matching.

Trivially, there are no cycles in T. Since $|\mathcal{C}| = 3q$ we have $|E_T| = N + 3q = |V| - 1$ and thus T is a spanning tree for G. Moreover, the total tree weight is $w(T) = 0 \times N + 1 \times |\mathcal{C}| = 3q = w^*$, that is, T is a minimum spanning tree for G.

It remains to be proven that $\forall v_i \in V, deg_T(v_i) \geq 4 \vee deg_T(v_i) = 1$, that is, the degree of each node in T agrees with the m4-MST restrictions, which is quite straightforward since:

- every node from V_1 (the ones in the bottom layer) has degree one;
- the node R has degree $N \geq q \geq d = 4$;

- each node X_{j_i}, $i = 1, \ldots, q$, has degree $d = 4$ and the remaining $N - q$ nodes on the middle layer have degree one.

Clearly, if I admits a 3D matching then 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 a degree of at least $d = 4$.

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 leaves, and the remaining N edges are the ones that connect node R to the second (middle) layer nodes. Therefore, each node from the third layer can only be connected to one X_j node and each X_j is connected to three 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 3D matching, which concludes the proof of Theorem 1.

Finally, we remark that this proof is easily generalized for $\lfloor n/2 \rfloor \geq d > 4$, using a reduction from $(d-1)$-Dimensional Matching[1], which is known to be an NP-hard problem for $d - 1 \geq 2$. □

Based on the test results obtained (Section 5.2), we believe that even for $d = 3$ the problem is still NP-hard, that is,

Conjecture 1. *For $d = 3$, the general md-MST problem is NP-hard.*

This thus remains as a challenge to be dealt with in the near future.

Notice that the *md*-MST problem can be looked at, in a very loose way, as the d-MST problem in a reverse fashion. Interestingly enough, the complexity results here presented seem also to be the reverse of the complexity conclusions for the Bounded Degree Problem!

4. Formulations for the *md*-MST problem

Network design problems, 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 for the MST problem can be found in Magnanti and Wolsey (1995). 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.

A number of formulations have also been proposed for the d-MST problem, including a single-commodity flow model in Gavish (1982) and natural formulations in Volgenant (1989), Caccetta and Hill (2001) and Andrade et al. (2006).

[1]Note that, for the d-Dimensional Matching to be well defined, we must have $d \leq \lfloor n/2 \rfloor$, where n is the cardinality of the d disjoint sets.

In Section 4.1, we present two extended formulations for the *md*-MST problem using flow-based models. For that reason and considering the already mentioned observation by Magnanti and Wolsey (1995), 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 consider the directed graph rooted at a node r taken from V. There are no incoming arcs incident to r. 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 out from the root.

In Section 4.2, we present an undirected formulation for the *md*-MST problem, based on the well-known packing constraints for characterizing spanning trees (see Magnanti and Wolsey, 1995).

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.

4.1. Directed flow formulations

Given the previously introduced oriented graph G_r, we 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, where $k_i = 1$ if i is a central node, and $k_i = 0$ when i is a leaf-node. A similar set of node-variables has been used in Fernandes and Gouveia (1998) to formulate the minimum spanning tree problem with a constraint on the number of leaves. Hence, we propose the following general formulation for the *md*-MST problem.

Formulation *md*-F

$$min \sum_{(i,j)\in A_r} c_{ij} x_{ij} \tag{6}$$

$$s.t. \ \ (d-1)k_r \leq \sum_{j\in V\setminus\{r\}} x_{rj} - 1 \leq (n-2)k_r \tag{7}$$

$$(d-1)k_i \leq \sum_{j\in V\setminus\{i,r\}} x_{ij} \leq (n-2)k_i\,, \quad i \in V\setminus\{r\} \tag{8}$$

$$x \in X_r \subset \{0,1\}^{(n-1)^2} \tag{9}$$

$$k_i \in \{0,1\}\,, \quad i \in V. \tag{10}$$

The model seeks for a minimum cost spanning tree in G_r as defined by (6) and (9), 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, except the root, has in-degree equal to 1.

Inequalities (7) and (8) define a lower and an upper bound to the number of outward arcs incident to each node i. Every leaf-node has a single incoming arc, except if node r is a leaf when there is a single arc outgoing from it. Therefore, when $i = r$, characterized by (7), there is at least one arc outgoing from node r. In the case $k_r = 1$, r is a central-node and the sum of all outgoing arcs minus 1 is bounded by $(d - 1)$ and $(n - 2)$. Otherwise, node r is a leaf and the second inequality in (7) establishes that $\sum_{j \in V \setminus \{r\}} x_{rj} \leq 1$, 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 outgoing arc from node r then $\sum_{j \in V \setminus \{r\}} x_{rj} - 1 > 0$, which forces k_r to be 1, otherwise, if there is a single outgoing arc from node r, then $\sum_{j \in V \setminus \{r\}} x_{rj} - 1 = 0$, implying that $k_r = 0$. A similar analysis can be considered for all other nodes $i \in V \setminus \{r\}$, this time associated with constraints (8). To all these nodes $i \in V \setminus \{r\}$, there is always one incoming arc. When i is a central-node, there are at least $(d - 1)$ outgoing arcs from it. When i is a leaf there is no outgoing arc from it.

Finally, constraints (10) impose integrality on variables k_i.

The general model *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 Magnanti and Wolsey (1995). In this subsection, 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. For the single-commodity flow formulation nonnegative variables y_{ij} describe the flow that passes through arc (i, j). For the multicommodity flow formulation nonnegative flow variables f_{ij}^k describe the amount of flow sent from the root to node k passing through arc (i, j).

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		**Multi-commodity flow model**	
$X_r = \{x \in [0,1]^{(n-1)^2}$:		$X_r = \{x \in [0,1]^{(n-1)^2}$:	
$\sum_{i \in V \setminus \{j\}} x_{ij} = 1,\ j \in V \setminus \{r\}$		$\sum_{i \in V \setminus \{j\}} x_{ij} = 1,\ j \in V \setminus \{r\}$	(11)
$\sum_{i \in V \setminus \{j\}} y_{ij} - \sum_{i \in V \setminus \{j,r\}} y_{ji} = 1,\ j \in V \setminus \{r\}$	(12a)	$\sum_{i \in V \setminus \{j,k\}} f_{ij}^k - \sum_{i \in V \setminus \{j,r\}} f_{ji}^k = 0,\ j, k \in V \setminus \{r\},\ j \neq k$	(12b)
		$\sum_{i \in V \setminus \{j\}} f_{ij}^j = 1,\ j \in V \setminus \{r\}$	(12c)
$x_{ij} \leq y_{ij} \leq (n-1)x_{ij},\ i \in V,\ j \in V \setminus \{i, r\}$	(13a)	$f_{ij}^k \leq x_{ij},\ i \in V,\ j, k \in V \setminus \{i, r\},\ j \neq k$	(13b)
		$f_{ij}^j \leq x_{ij},\ i \in V,\ j \in V \setminus \{i, r\}$	(13c)
$x_{ij} \in \{0, 1\},\ i \in V,\ j \in V \setminus \{i, r\}$		$x_{ij} \in \{0, 1\},\ i \in V,\ j \in V \setminus \{i, r\}$	(14)
$y_{ij} \geq 0,\ i \in V,\ j \in V \setminus \{i, r\}\ \}$	(15a)	$f_{ij}^k \geq 0,\ i \in V,\ j, k \in V \setminus \{i, r\}\ \}$	(15b)

Equalities (11), appearing in both models, are in-degree constraints for each node $j \in V \setminus \{r\}$. Equalities (12) 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 the possible unitary flow in variables f_{ij}^k, for each destination node k, represent a disaggregation of the whole flow in y_{ij}. For this reason, constraints

(12b) and (12c) can also be seen as a disaggregated version of constraints (12a). Inequalities (13) 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 (13b, 13c) correspond to a disaggregated version of inequalities (13a). Constraints (14) impose integrality to the arc design variables, and constraints (15) impose nonnegativity to the flow variables.

Note that the value of each variable f_{ij}^k is never greater than 1. Therefore, we could have set constraints (15b) as $0 \le f_{ij}^k \le 1$. However, this is implicitly defined by inequalities (13b, 13c).

The LP relaxation version of both models can be obtained by substituting the integrality constraints (14) by the bounding constraints

$$0 \le x_{ij} \le 1, \quad \text{for all } (i, j) \in A_r. \tag{16}$$

Within the LP relaxation version of the multicommodity flow model, one can think of a different form for the coupling constraints (13c) by considering equalities (17) instead.

$$f_{ij}^j = x_{ij}, \quad i \in V, \; j \in V \setminus \{i, r\}. \tag{17}$$

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 incoming arc incident to j. Therefore, we can substitute inequalities (13c) by constraints (17) in the multicommodity flow model. The advantage is that constraints (12c) become redundant after including equalities (17) in the formulation. This can be observed by summing in i all equalities (17) and substituting the resulting right-hand side term by 1 using constraints (11). The mentioned substitution of constraints (13c) by (17) may seem that we are strengthening model *md*-MCF$_L$ (in LP relaxation). However, as shown in Gouveia and Lopes (2000), this is not true because equalities (17) are implicitly contained in the relaxed polyhedron (11)–(15b).

Considering the two characterizations of the set X_r previously proposed, with constraints (17), we can define two flow formulations for problem *md*-MST.

single-commodity flow:

$$md\text{-}\mathbf{SCF}\text{: min} \left\{ \sum_{(i,j)\in A_r} c_{ij}x_{ij} \; : \; (7), (8), (10), (11), (12a), (13a), (14), (15a) \right\}$$

multi-commodity flow:

$$md\text{-}\mathbf{MCF}\text{: min} \left\{ \sum_{(i,j)\in A_r} c_{ij}x_{ij} \; : \; (7), (8), (10), (11), (12b), (13b), (14), (15b), (17) \right\}.$$

The LP relaxation versions of both flow models can be obtained by substituting the integrality constraints (10) and (14) by the bounding constraints (16) and (18), respectively.

$$0 \le k_i \le 1, \quad i \in V. \tag{18}$$

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

These relaxed versions of the two *md*-MST models can be showed to be equivalent to the corresponding unconstrained versions of the problem, that is, to the MST problem. This observation is based on Proposition 4. Before presenting it, let us define the two polyhedrons

$$P(\mathrm{SCF}_L) = \left\{(x, y) \in R^{2(n-1)^2} \ : \ (x, y) \text{ satisfies (11), (12a), (13a), (15a) and (16)}\right\} \quad \text{and}$$

$$P(\mathrm{MCF}_L) = \left\{(x, f) \in R^{(n-1)^2} \times R^{(n-1)^3} \ : \ (x, f) \text{ satisfies (11), (12b), (13b), (15b), (16) and (17)}\right\}$$

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

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

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

associated with models md-SCF_L and md-MCF_L, respectively. Then we have the following proposition:

Proposition 4. *$proj_{\{x,y\}}(P(md\text{-}SCF_L)) = P(SCF_L)$ and $proj_{\{x,f\}}(P(md\text{-}MCF_L)) = P(MCF_L)$.*

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

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 (12) 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 $\sum_{j,k \in V \setminus \{r\}} f_{rj}^k = n - 1$, which implies, using the linking constraints (13), 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 (7) and (8), appearing in both md-SCF_L and md-MCF_L models, we have $((d-1)k_r \leq \theta_r - 1 \leq (n-2)k_r)$ 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

$$\frac{\theta_r - 1}{n-2} \leq k_r \leq \frac{\theta_r - 1}{d-1} \quad \text{and} \quad \frac{\theta_i}{n-2} \leq k_i \leq \frac{\theta_i}{d-1}, \ \text{for all } i \in V \setminus \{r\}. \tag{19}$$

As $\theta_r \in [1, n-1]$ and $\theta_i \in [0, n-2]$ for $i \in V \setminus \{r\}$, there is a solution in the k variables with $k \in [0, 1]^n$ that also satisfies (19). 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 $P(\mathrm{SCF}_L) \subseteq proj_{\{x,y\}}(P(md\text{-SCF}_L))$ and $P(\mathrm{MCF}_L) \subseteq proj_{\{x,f\}}(P(md\text{-MCF}_L))$. □

It follows from Proposition 4 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, for 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 4 and the polyhedral discussion on some MST formulations proposed in Magnanti and Wolsey (1995), we can state that model md-MCF$_L$ is symmetric concerning the root node, that is, the optimal solution value of md-MCF$_L$ is always the same, no matter which node is selected for the root. This is based on 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 Magnanti and Wolsey, 1995). A similar result does 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 nontrivial valid inequalities. These are:

(1) Constraints linking design and node variables

$$x_{ij} \leq k_i \ , \ i, j \in V \backslash \{r\}, \ i \neq j \tag{20}$$

establishing that if arc (i, j) is in the solution, then there is an outward arc incident to node i that forces it to be a central node; otherwise, if i is a leaf-node, then there cannot be any arc leaving node i; and for the root node case the inequalities

$$x_{rj} \leq k_r + k_j \ , \ j \in V \backslash \{r\} \tag{21}$$

establish that if arc (r, j) is in the solution, then either r is central or it is linked to a central node. This last set (21) has been reported in Martins and de Souza (2009).

(2) A different constraint involving just node variables is

$$\sum_{i \in V} k_i \leq \left\lfloor \frac{n-2}{d-1} \right\rfloor \tag{22}$$

which is based on Corollary 1 presented in Section 2, defining an upper bound to the number of central-nodes in any feasible solution to the md-MST problem.

Note that constraints (21) do not follow the (20) form as there is at least one outgoing arc from the root in any feasible solution, even when r is a leaf-node.

It can be shown by an example that inequalities (20), (21), and (22) are not redundant in both md-SCF$_L$ and md-MCF$_L$ models. However, a weaker version of inequality (22), defined by $\sum_{i \in V} k_i \leq (n-2)/(d-1)$, is redundant in the two mentioned formulations for the md-MST problem, as described in the next result.

Proposition 5. *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 (7) and (8). If we sum all those inequalities, namely the one from (7) and all for $i \in V\backslash\{r\}$ from (8), we obtain

$$(d-1)\sum_{i\in V} k_i \le \sum_{i\in V}\sum_{j\in V\backslash\{i,r\}} x_{ij} - 1. \tag{23}$$

Considering the sum in j of equalities (11),

$$\sum_{j\in V\backslash\{r\}}\sum_{i\in V\backslash\{j\}} x_{ij} = n-1 \tag{24}$$

and substituting (24) in (23), we obtain

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

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

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

showing the intended result. □

The new constraints also indicate that the second inequality in (8) can be dropped from md-SCF$_L$ and md-MCF$_L$. In fact, for each $j \in V$, if we sum inequalities (20) for all $i \in V \setminus \{j\}$, then we obtain the second inequalities in (8). Therefore, they can be omitted from the models, once (20) are included. In fact, inequalities (20) can be seen as a disaggregated version of the mentioned inequalities (8). Hence, the strengthened versions of both single- and multicommodity flow models to the md-MST problem are defined by

- md-**SCF1**$_L$: md-SCF$_L$ strengthened with constraints (20) and (21);
- md-**MCF1**$_L$: md-MCF$_L$ strengthened with constraints (20) and (21);
- md-**SCF2**$_L$: md-SCF1$_L$ strengthened with inequality (22);
- md-**MCF2**$_L$: md-MCF1$_L$ strengthened with inequality (22).

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

As a final observation, according to Proposition 5, the two formulations md-SCF2$_L$ and md-MCF2$_L$ should only be considered when the quotient $(n-2)/(d-1)$ is not integer.

4.2. *Undirected flow formulations*

Considering the original undirected graph $G = (V, E)$, we also propose an undirected formulation to the md-MST problem. In this case, we use undirected edge design binary variables w_e taking value 1 if edge e is in the solution, and value 0 otherwise. Node-variables k_i (for $i \in V$) keep the definition from the directed models.

This formulation resorts to the so-called "natural packing" constraints for characterizing spanning trees on the undirected graph $G = (V, E)$ (see, Magnanti and Wolsey, 1995), leading to the following natural model for the *md*-MST.

Formulation *md*-NP

$$min \quad \sum_{e \in E} c_e w_e \tag{25}$$

$$s.t. \quad (d-1)k_i \leq \sum_{e \in \delta(i)} w_e - 1 \leq (n-2)k_i \,, \quad i \in V \tag{26}$$

$$\sum_{e \in E} w_e = n - 1 \tag{27}$$

$$\sum_{e \in E(S)} w_e \leq |S| - 1 \,, \quad \text{for all } S \subset V \text{ and } |S| \geq 2 \tag{28}$$

$$w_e \in \{0, 1\} \,, \quad e \in E \tag{29}$$

$$k_i \in \{0, 1\} \,, \quad i \in V. \tag{30}$$

The cost of edge $e \in E$ is denoted by c_e, being equal to c_{ij} for $e = \{i, j\}$. The set of edges incident to node i is represented by $\delta(i)$, and $E(S)$ is the set of edges in the subgraph induced by $S \subset V$. Furthermore, both directed and undirected design variables can be related by

$$x_{rj} = w_e \,, \quad j \in V \backslash \{r\} \quad \text{and} \quad e = \{r, j\} \tag{31}$$

$$x_{ij} + x_{ji} = w_e \,, \quad i, j \in V \backslash \{r\} \,, \quad i < j \quad \text{and} \quad e = \{i, j\}. \tag{32}$$

Inequalities (26) characterize the node bounding constraints, forcing that at least d edges are incident to i whenever i is a central-node, and imposing a single edge to be incident to i whenever it is a leaf-node. These are the undirected version of inequalities (7), (8).

Constraints (27)–(29) characterize all spanning trees of G, forcing each solution to have $(n-1)$ edges (equality (27)) and to be cycle free (inequalities (28)). In fact, inequalities (28) are also known as subtour elimination constraints.

As for the multicommodity flow spanning trees' characterization, represented by $X_r = \{x \in [0, 1]^{(n-1)^2} :$ (11), (12b), (12c) (13b), (13c), (14), and (15b)$\}$, the linear programming relaxation of (27)–(29) characterize the convex hull of the spanning tree polytope (see, Magnanti and Wolsey, 1995). Furthermore, constraints (26) can be obtained using inequalities (7), (8), the equalities (11), and the variable transformations (31), (32). This means that the lower bound produced by the linear programming relaxation of model *md*-NP is not better than the one produce by model *md*-MCF$_L$. In fact, all inequalities (28) are redundant in *md*-MCF$_L$. A similar observation does not hold for

md-SCF$_L$. In this case, constraints (28) can still improve the lower bound produced by md-SCF$_L$, and the same can be observed with the strengthened models md-SCF1$_L$ or md-SCF2$_L$.

Considering that (28) involve an exponential number of inequalities, we propose strengthening model md-SCF2$_L$ only with a subset of constraints (28), namely those with $|S| = 2$. Using the variable transformation defined in (32), the mentioned subset of constraints is characterized by

$$x_{ij} + x_{ji} \leq 1 \ , \ i, j \in V \backslash \{r\} \ , \ i < j. \tag{33}$$

We denote by md-SCF3$_L$ the md-SCF2$_L$ model strengthened with inequalities (33). Considering Proposition 5, inequality (22) is only included when the quotient $(n - 2)/(d - 1)$ is not integer. The option for strengthening model md-SCF2$_L$ just with constraints (33), and avoiding the larger family (28), has also been followed in other works addressing spanning trees problems (Ljubic et al., 2006; Moura, 2009; Gouveia and Moura, 2011; Gouveia et al., 2011). As mentioned in these papers, constraints (33) have been observed to be very effective when added to the SCF$_L$ model, producing almost the same lower bound as model MCF$_L$ (see, Moura, 2009). As we will show later in Section 5.2, for the instances of class SYM, the LP relaxation of the SCF-based models strengthened with constraints (33) produce almost the same quality lower bounds as the associated MCF models.

5. Computational results

Computational results for the various models proposed in Section 4 are now reported. These results were obtained using a class of complete graphs with Euclidean costs in various dimensional spaces. These are the CRD and SYM instances that have long been considered by many researchers when addressing the d-MST problem (see, e.g., Krishnamoorthy et al., 2001; Ribeiro and de Souza, 2002).

The cost matrix for each CRD instance was taken as the Euclidean distance between the coordinates of n points, randomly generated using a uniform distribution in a square. The SYM instances are analogous to the CRD problems but with points generated in a higher dimensional Euclidean space. We have only considered the first three instances in each class.

Common degree constraints to all nodes have been considered in all tests, that is, $d(i) = d$ for all $i \in V$. These are $d = 3, 5$ for the $n = 30$ instances, and $d = 3, 5, 10$ for the $n = 50$ instances.

All computational experiments have been performed on a PC with a 3.2 GHz Intel Pentium-4 processor and 512 Mb of RAM, using the `CPLEX` 9.0 package to solve the LP and the Mixed Integer Programming (MIP) models.

We report results obtained from the LP relaxation of all models under discussion. Computational results from the branch-and-bound execution are also reported. In this case, an upper time limit of 3 hours (10,800 seconds) to run the branch-and-bound has been established. The default parameter settings within the `MIP` code of `CPLEX` have been applied including the usage of the dual simplex to solve all subproblems, 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 ILOG/CPLEX (1997).

Gap values in the tables represent 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 optima were obtained or confirmed by the branch-and-bound algorithm. All values presented in the tables are average results taken from the instances in the same class. The corresponding individual results can be found in the Appendix C. 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

We first provide a brief discussion on the root node selection. As observed in Section 4, the md-MCF$_L$ model has been proved to be insensitive to the root node selection, being a symmetric formulation. However, for model md-SCF$_L$, this selection can be of importance as it can result in different LP bounds, which may influence the times to reach the integer optimums. So the following question arises: "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 to try to answer this question, we made a few experiments with a 16-node instance (taken from instance CRD300, using the first 16 nodes). Table 1 shows, for all possible root nodes, 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 $d = 5$. The central-nodes in the optimal solutions correspond to the lines highlighted in boldface. The optimal objective 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 it 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 highest ($r = 12$) results. This conclusion can be extended to the $d = 5$ case, which does not give us much information to help choosing the root.

On the other hand, model md-MCF$_L$ is a symmetric formulation, which means that it should be irrelevant which node to choose to be the root. This is correct when we only look to the bounds produced by the model (in LP), however, this may not be irrelevant when looking for the time the model takes to reach the optimum. In fact, the last column in Table 1 (column B-B times for $d = 5$) shows that the times to solve the branch-and-bound with the symmetric model md-MCF$_L$ are quite different when different nodes are chosen for the root. Those times range from 6.66 to 19.7 seconds. A reason for this is that 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 lose this property after adding the mentioned cuts. Furthermore, the separation procedures used by the algorithm to identify those cuts may be sensitive to the node selected as root.

In order to further stress the effect of selecting different nodes for rooting the tree, we solved instances CRD300 and SYM300 for $d = 3$ using formulations md-SCF, md-SCF3, md-MCF, and

Table 1
LP relaxation bounds and times as well as branch-and-bound times for models *md*-SCF and *md*-MCF for all $r \in V$, using a 16 nodes instance taken from CRD300. Times are expressed in seconds.

Root	$d = 3$						$d = 5$					
node	*md*-SCF			*md*-MCF			*md*-SCF			*md*-MCF		
	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

md-MCF1, and analyzed lower bounds and times produced by the LP relaxation as well as branch-and-bound execution times for all possible choices for the root node, that is, considering $r = i$ for all $i \in V$. These results are reported in Tables B1 and B2 of Appendix B, for the CRD300 and the SYM300 instances, respectively. In each column, we highlighted the three best values (or more than three when multiple equal observations exist in the third position). These values are shaded. In this case, we raise the following question: "Is there a node that returns the best results for all formulations?" The answer seems to be negative, otherwise we would observe straight shaded lines in Tables B1 and B2, revealing nodes behaving consistently better in all experiments. In fact, for most selected (shaded) values involving models *md*-SCF, *md*-SCF3, and *md*-MCF1, choosing a root node based on the LP relaxation optimum may not lead to the shortest time to reach the optimum by the branch-and-bound. Furthermore, in most cases, the best choice for rooting the tree is not consistent among different formulations. Also, and confirming an observation stressed in the previous paragraph, the symmetry property of model *md*-MCF$_L$ is lost in model *md*-MCF1$_L$, that is, after strengthening formulation *md*-MCF$_L$. In fact, it is curious to see that the times required by the branch-and-bound for the two strongest formulations (models *md*-SCF3 and *md*-MCF1), reported in Tables B1 and B2, exhibit very large standard-deviations when compared with the associated average times, especially for the CRD300 tests.

So, considering all these arguments and without further evidences we chose the first node of every instance as the root ($r = 1$) in all our tests. In fact, a similar option has been followed by other authors when addressing undirected tree based problems using directed models (Duhamel et al., 2007; Moura, 2009; Gouveia and Moura, 2010, 2011).

The need to select a node to root the tree could have been avoided if other formulations were used, namely if we consider an undirected formulation for our problem. Another suggestion comes from the work of Gouveia and Telhada (2001) addressing a Two-Level Network Design Problem (TLND). In this work, the authors propose a very interesting symmetric and compact formulation for the mentioned problem, characterizing the intersecting polyhedra of all LP relaxation models to all possible root-nodes. However, in the TLND problem, we know in advance a partition for the set of nodes, that is, V is a priori partitioned into primary and secondary nodes.

5.2. *Results for the md-MST problem*

Our computational tests for the *md*-MST problem are summarized in Tables 2–4. We start by analyzing the LP relaxation versions of the seven formulations proposed in Section 4, namely *md*-SCF$_L$, *md*-SCF1$_L$, *md*-SCF2$_L$, *md*-SCF3$_L$, *md*-MCF$_L$, *md*-MCF1$_L$, and *md*-MCF2$_L$. These tests involve the CRD and SYM instances, for $n = 30$ and 50.

In several classes of instances, averaged gaps presented in Table 2 show very high duality gaps produced by both *md*-SCF$_L$ and *md*-MCF$_L$ models. This can be explained by Proposition 4, which establishes that the linear relaxation of the two mentioned formulations produces the same LP value optimums as the corresponding formulations for the MST problem without degree constraints. Therefore, we should expect poor bounds from these models, especially when d increases, that is, following the *md*-MST deviation from the MST problem. The insensitivity of models *md*-SCF$_L$ and *md*-MCF$_L$ to the d parameter value can also be observed in Tables C1 and C3 in Appendix C.

Table 2
Average gaps (%) produced by the LP relaxation of formulations *md*-SCF, *md*-SCF1, *md*-SCF2, *md*-SCF3, *md*-MCF1, and *md*-MCF2

Type	n	d	LP relaxation						
			Gap						
			md-SCF	*md*-SCF1	*md*-SCF2	*md*-SCF3	*md*-MCF	*md*-MCF1	*md*-MCF2
CRD	30	3	24.95	16.30		11.12	10.05	5.05	
CRD	30	5	39.77	9.56		9.12	27.81	6.55	
CRD	50	3	30.36	13.41		9.20	11.65	5.01	
CRD	50	5	44.82	11.31		10.00	30.08	7.49	
CRD	50	10	59.63	6.69	4.91	4.91	48.83	5.98	4.24
SYM	30	3	26.07	8.65		2.65	15.37	2.61	
SYM	30	5	49.02	6.83		5.50	41.68	5.28	
SYM	50	3	22.96	8.68		3.58	12.57	3.06	
SYM	50	5	50.39	10.82		8.17	43.70	7.58	
SYM	50	10	76.44	14.80	11.57	11.56	73.28	14.56	11.40

Another interesting observation involves the results produced by the models md-SCF1_L and md-MCF1_L, obtained through the strengthening constraints (20) and (21). These inequalities seem to be very effective, especially for higher values of d. This is because constraints (20) 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 forced to take higher values in the linear programming relaxation. As a consequence, the first set of linking constraints (8) become tighter, leading to a tighter linear relaxation. For most of the mentioned instances, the added constraints also allow formulation md-SCF1_L to produce much better lower limits than md-MCF$_L$ requiring smaller execution times. In fact, the number of constraints in the augmented model md-SCF1 is $\mathcal{O}(n^2)$ while in the md-MCF$_L$ there are $\mathcal{O}(n^3)$ constraints.

The gaps obtained with models md-SCF2_L and md-MCF2_L indicate that the single constraint (22) effectively cuts the md-SCF1_L and md-MCF1_L polyhedrons. This inequality has only been applied for $d = 10$, as explained at the end of Section 4, being able to reduce the gaps in both strengthened LP formulations. As observed in Table 3 and for $d = 10$, the added constraints (20)–(22) 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 solve the problem much faster than the more disaggregated multicommodity flow models, especially for higher values of d.

Table 2 also shows that the lower limits generated by model md-SCF3_L (that include the two-cycle subtour elimination constraints (33)) can still reduce model's md-SCF2_L gaps, especially for smaller values of d and for the SYM class instances. These limits were almost the same as those generated by the associated MCF-based models for the mentioned set of instances. A similar observation does not hold for the CRD class and 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 3). However, the duality gaps observed in this particular case are still very high, even among the stronger models, bringing some curiosity about its theoretical hardness conjecture.

Table 3
Average execution times (in seconds) to solve the LP relaxation of formulations *md*-SCF, *md*-SCF1, *md*-SCF2, *md*-SCF3, *md*-MCF1, and *md*-MCF2

Type	n	d	LP relaxation						
			Time						
			md-SCF	*md*-SCFl	*md*-SCF2	*md*-SCF3	*md*-MCF	*md*-MCFl	*md*-MCF2
CRD	30	3	0.06	0.12		0.13	4.20	16.99	
CRD	30	5	0.06	0.18		0.23	3.99	30.09	
CRD	50	3	0.59	2.25		2.24	72.16	270.36	
CRD	50	5	0.59	1.92		2.67	71.92	376.70	
CRD	50	10	0.59	3.76	2.93	4.22	71.71	383.62	377.00
SYM	30	3	0.06	0.09		0.08	0.98	6.11	
SYM	30	5	0.06	0.11		0.17	0.96	14.05	
SYM	50	3	0.66	2.65		0.76	55.82	302.55	
SYM	50	5	0.67	1.66		0.80	42.29	172.72	
SYM	50	10	0.66	2.80	1.14	2.43	47.31	168.35	125.67

Table 4
Average execution times (in seconds) to solve the branch-and-bound algorithm using models *md*-SCF, *md*-SCF1, *md*-SCF2, *md*-SCF3, *md*-MCF, *md*-MCF1, and *md*-MCF2

Type	n	d	Branch-and-bound						
			Time						
			md-SCF	*md*-SCFl	*md*-SCF2	*md*-SCF3	*md*-MCF	*md*-MCFl	*md*-MCF2
CRD	30	3	7996.47	8484.25		602.60	1060.26	918.65	
CRD	30	5	490.66	489.93		224.48	8680.02	7948.18	
CRD	50	3	10,800.00	10,800.00		10800.00	10,800.00	10,800.00	
CRD	50	5	10,800.00	10800.00		10800.00	10,800.00	10,800.00	
CRD	50	10	4194.63	6103.76	586.89	608.76	10,800.00	10,800.00	10,800.00
SYM	30	3	14.14	19.29		0.69	94.35	62.84	
SYM	30	5	17.11	22.27		4.88	859.91	826.75	
SYM	50	3	4682.57	3399.18		15.03	4585.68	3015.12	
SYM	50	5	7946.45	7129.13		172.58	10,800.00	10,800.00	
SYM	50	10	2333.61	6284.66	1127.57	460.83	10,800.00	10,800.00	10,800.00

Looking at the average execution times presented in Table 4, it appears to be more difficult to reach the optimums among the CRD instances. Remember that these instances have the nodes located in the plane and distances are Euclidean.

To conclude, note that among the higher dimensional instances with $n = 50$, many optimums are still unconfirmed, especially within the CRD class. For $d = 10$, the multicommodity flow formulations have not been able to reach any optimum, while the *md*-SCF3 model showed to be the most effective formulation, reaching the goal in all instances much faster than the other models (see Tables C2 and C4 in Appendix C). In addition, model *md*-SCF3 was able to strongly reduce the times to reach the integer optimums, especially for the SYM class instances. In fact, even with the strengthened multicommodity flow formulations the *md*-MST problem still exhibits very high LP gaps. This suggests that the *md*-MST is still requiring further polyhedral knowledge.

6. Conclusions

This paper addresses a new degree constrained spanning tree problem, involving a minimum degree constraint on the nodes. This problem is closely related to 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 $\lfloor n/2 \rfloor \geq d \geq 4$, being open for $d = 3$. Some properties have been presented, namely defining upper and lower bounds to the number of central nodes (or to the number of leaf-nodes) in any *md*-MST feasible solution.

Flow-based formulations were proposed for the *md*-MST. The corresponding computational tests showed that the lower bounds produced by the LP relaxation of the multi-commodity flow formulation are 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. This observation may indicate that we are dealing with a difficult problem from an empirical point of view, 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 inequalities as described in Section 4, the LP relaxation results still kept a significant duality gap.

It has also been observed that the "lighter" single-commodity flow formulations become very competitive when strengthened with the additional constraints $x_{ij} \leq k_i$ (for all $i, j \in V \setminus \{r\}, i \neq j$), $x_{rj} \leq k_r + k_j$ (for all $j \in V \setminus \{r\}$), $x_{ij} + x_{ji} \leq 1$ (for all $i, j \in V \setminus \{r\}$ and $i < j$), and $\sum_{i \in V} k_i \leq \left\lfloor \frac{n-2}{d-1} \right\rfloor$. The latter inequality has been derived from a property stated in Section 2.

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_r of all spanning trees, namely considering natural formulations, for example the one described in Section 4.2. Another thought involves reducing 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.

References

Almeida, A.M., Martins, P., de Souza, M.C., 2006. Min-degree constrained minimum spanning tree problem: complexity, properties and formulations. Operations Research Center, University of Lisbon, Working paper no 6. Available at `http://cio.fc.ul.pt/files/6.2006.pdf`

Andrade, R., Lucena, A., Maculan, N., 2006. Using Lagrangian dual information to generate degree constrained spanning trees. *Discrete Applied Mathematics* 154, 5, 703–717.

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

Duhamel, C., Gouveia, L., Moura, P., de Souza, M., 2007. Minimum cost degree constrained spanning trees with node-degree costs. Operations Research Center, University of Lisbon, Working Paper no 8.

Fernandes, L.M., Gouveia, L., 1998. Minimal spanning trees with a constraint on the number of leaves. *European Journal of Operational Research* 104, 250–261.

Garey, M.R., Johnson, D.S., 1979. *Computers and Intractability: A Guide to the Theory of NP-Completeness*. W.H. Freeman, San Francisco, CA.

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

Gouveia, L., Lopes, M.J., 2000. Valid inequalities for non-unit demand capacitated spanning tree problems with flow costs. *European Journal of Operational Research* 121, 394–411.

Gouveia, L., Moura, P., 2010. Spanning trees with node degree dependent costs and knapsack reformulation. *Electronic Notes in Discrete Mathematics* 36, 1, 985–992.

Gouveia, L., Moura, P., 2011. Enhancing discretized formulations: the knapsack reformulation and the star reformulation. *TOP.* DOI: 10.1007/s11750-011-0212-x.

Gouveia, L., Moura, P., de Sousa, A., 2011. Prize collecting Steiner trees with node degree dependent costs. *Computers and Operations Research* 38, 1, 234–245.

Gouveia, L., Telhada, J., 2001. An augmented arborescence formulation for the two-level network design problem. *Annals Operations Research* 106, 1–4, 47–61.

ILOG/CPLEX 9.0. Reference Manual, 1997.

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

Krishnamoorthy, M., Ernst, A.T., Sharaiha, Y.M., 2001. Comparison of algorithms for the degree constrained spanning tree. *Journal of Heuristics* 7, 587–611.

Ljubic, I., Weiskircher, R., Pferschy, U., Klau, G.W., Mutzel, P., Fischetti, M., 2006. An algorithmic framework for the exact solution of the prize-collecting Steiner tree problem. *Mathematical Programming, Series B* 105, 427–449.

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

Martins, P., de Souza, M.C., 2009. VNS and second order heuristics for the min-degree constrained minimum spanning tree problem. *Computers and Operations Research* 36, 11, 2969–2982.

Monma, C., Suri, S., 1992. Transitions in geometric minimum spanning trees. *Discrete Computational Geometry* 8, 265–293.

Moura, P., 2009. Problema da árvore de suporte de custo mínimo com restrição de grau e custos associados aos nodos. PhD Thesis, Faculty of Science, University of Lisbon, Portugal (in Portuguese).

Papadimitriou, C.H., 1994. *Computational Complexity.* Addison-Wesley, Reading, MA.

Papadimitriou, C.H., Vazirani, U.V., 1984. On two geometric problems related to the traveling salesman problem. *Journal of Algorithms* 5, 231–246.

Ribeiro, C.C., de Souza, M.C., 2002. Variable neighborhood search for the degree-constrained minimum spanning tree problem. *Discrete Applied Mathematics* 118, 1–2, 43–54.

Robins, G., Salowe, J.S., 1995. Low-degree minimum spanning trees. *Discrete Computational Geometry* 14, 151–166.

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

Volgenant, T., 1989. A Lagrangean approach to the DCMST problem. *European Journal of Operational Research* 39, 325–331.

Appendix A

Cost matrix C used in the example in Fig. 2.

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

Appendix B

Tables B1 and B2 report the results addressing the root node selection computational experiments described in Section 5.1, involving instances CRD300 and SYM300 with $n = 30$ and $d = 3$. The tests reported in Tables B1 and B2 are conducted for each possible root node $r = i$, for all $i \in V$. Tests involve models *md*-SCF, *md*-SCF3, *md*-MCF, and *md*-MCF1, considering both the linear programming relaxation and the branch-and-bound computational results.

The last four lines in the tables represent the average, the standard-deviation, the minimum, and the maximum, among all observations reported in each column.

Appendix C

Tables C1–C4 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.

Table B1
Root node selection experiments for instance CRD300 with $d = 3$

Root	*md*-SCF			*md*-SCF3			*md*-MCF			*md*-MCF1		
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	3140.54	0.06	5664.58	3649.65	0.14	306.78	3634.00	5.47	1301.20	3764.33	23.73	940.58
2	3054.93	0.06	5164.30	3644.88	0.20	269.11	3634.00	2.95	771.02	3819.50	10.33	316.36
3	3106.13	0.06	3305.87	3692.01	0.11	113.03	3634.00	2.09	820.77	3819.83	11.19	394.76
4	3148.26	0.08	1385.79	3690.28	0.11	88.61	3634.00	4.20	1315.41	3849.33	14.77	493.90
5	3142.30	0.06	2691.65	3649.65	0.11	282.04	3634.00	6.33	2293.67	3764.33	21.19	6106.21
6	3191.17	0.06	847.17	3722.27	0.09	57.39	3634.00	3.22	1376.02	3880.19	12.41	299.15
7	3142.30	0.06	1778.86	3649.54	0.11	261.82	3634.00	6.22	2554.83	3761.65	22.19	907.94
8	3143.21	0.06	1060.58	3698.81	0.11	126.47	3634.00	3.91	1015.63	3822.61	16.48	406.18
9	3119.41	0.08	1520.58	3717.47	0.13	75.50	3634.00	2.86	400.59	3843.35	21.89	336.38
10	3145.16	0.06	1084.98	3696.26	0.09	120.70	3634.00	5.08	1366.56	3861.09	14.09	427.38
11	3145.16	0.06	6604.82	3690.28	0.13	67.73	3634.00	4.84	1063.81	3849.33	19.59	364.24
12	3191.17	0.06	380.64	3708.69	0.09	44.82	3634.00	3.91	538.01	3866.79	11.86	314.79
13	3140.54	0.08	1618.56	3649.65	0.13	378.86	3634.00	6.00	1270.58	3764.33	22.05	1442.49
14	3143.21	0.14	4384.31	3717.47	0.13	60.56	3634.00	3.61	740.41	3840.98	19.28	426.33
15	3137.72	0.05	1960.89	3697.82	0.11	136.73	3634.00	4.80	1168.42	3826.70	24.50	933.98
16	3056.53	0.06	6088.91	3638.05	0.09	226.00	3634.00	4.80	893.86	3759.72	26.66	1149.31
17	3106.13	0.06	4142.08	3717.06	0.13	110.30	3634.00	1.89	498.33	3829.90	12.72	228.44
18	3148.26	0.06	2673.91	3692.34	0.09	80.08	3634.00	3.73	571.09	3851.75	17.34	491.12
19	3072.17	0.14	7311.65	3690.57	0.08	704.38	3634.00	4.52	459.17	3833.54	12.36	343.59
20	3099.21	0.13	9017.41	3671.22	0.09	379.84	3634.00	4.38	433.98	3809.08	12.14	294.39
21	3126.13	0.06	1832.31	3708.19	0.11	150.80	3634.00	2.30	500.56	3819.90	14.09	255.26
22	3040.31	0.08	10800.00	3650.96	0.14	809.84	3634.00	4.25	899.99	3786.50	21.03	708.46
23	3104.58	0.06	795.58	3730.35	0.09	76.75	3634.00	2.45	337.45	3851.29	14.52	358.46
24	3126.13	0.16	1840.32	3706.87	0.11	72.86	3634.00	2.34	404.39	3825.45	15.17	289.36
25	3171.15	0.06	1631.74	3671.93	0.08	187.00	3634.00	6.17	1038.84	3802.13	26.63	555.47
26	3056.53	0.06	10800.00	3644.50	0.09	475.76	3634.00	4.36	4110.84	3764.33	18.59	957.03
27	3218.50	0.06	8510.32	3671.93	0.08	511.06	3634.00	6.74	1535.52	3802.13	22.28	942.18
28	3099.21	0.08	2208.22	3675.24	0.09	115.07	3634.00	5.02	938.36	3803.58	8.58	460.85
29	3130.16	0.14	1990.19	3662.84	0.14	181.70	3634.00	4.53	1095.77	3817.83	22.17	875.05
30	3218.50	0.06	4744.43	3671.93	0.09	122.92	3634.00	6.08	1295.77	3802.13	21.94	542.30
Average	3128.82	0.08	3794.69	3682.62	0.11	219.82	3634.00	4.30	1100.36	3816.45	17.73	752.06
Stand-dev	44.92	0.03	3040.31	27.27	0.03	193.11	0.00	1.38	771.57	34.28	5.14	1057.42
Minimum	3040.31	0.05	380.64	3638.05	0.08	44.82	3634.00	1.89	337.45	3759.72	8.58	228.44
Maximum	3218.50	0.16	10800.00	3730.35	0.20	809.84	3634.00	6.74	4110.84	3880.19	26.66	6106.21

Table B2
Root node selection experiments for instance SYM300 with $d = 3$

Root	*md*-SCF			*md*-SCF3			*md*-MCF			*md*-MCFl		
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	775.04	0.06	24.03	1148.50	0.03	0.91	958.00	1.48	183.17	1148.63	8.42	112.89
2	775.04	0.08	61.66	1123.42	0.09	1.20	958.00	2.48	239.54	1125.00	12.59	159.70
3	789.50	0.06	99.38	1123.91	0.06	1.92	958.00	2.67	330.17	1133.33	21.67	392.33
4	768.93	0.14	42.44	1117.94	0.11	1.89	958.00	1.86	254.21	1119.97	22.41	256.03
5	795.34	0.06	57.60	1096.15	0.06	3.44	958.00	6.16	420.57	1098.32	19.06	394.80
6	775.04	0.08	35.44	1133.06	0.19	1.59	958.00	1.83	125.82	1135.35	16.73	280.81
7	789.50	0.06	32.48	1098.64	0.09	1.75	958.00	4.03	233.19	1100.40	17.80	257.91
8	768.41	0.06	33.44	1113.63	0.03	1.20	958.00	2.33	291.87	1113.79	16.20	232.70
9	790.83	0.06	60.22	1080.25	0.03	1.73	958.00	3.05	440.28	1080.25	14.47	263.84
10	769.43	0.16	89.96	1102.08	0.08	3.03	958.00	6.44	627.80	1105.17	20.66	366.70
11	762.24	0.06	48.08	1099.86	0.03	1.47	958.00	4.61	353.66	1099.86	13.45	515.95
12	774.83	0.06	67.36	1116.50	0.06	2.63	958.00	1.77	381.20	1116.50	8.77	356.91
13	756.20	0.06	57.10	1099.92	0.08	3.31	958.00	2.28	466.34	1100.13	15.48	237.42
14	768.89	0.06	53.72	1108.00	0.05	1.16	958.00	1.91	334.35	1108.00	11.72	191.95
15	790 83	0.06	83.22	1105.25	0.03	1.48	958.00	2.41	284.31	1105.25	8.08	205.09
16	789.50	0.17	32.54	1116.23	0.06	4.48	958.00	3.38	370.26	1117.55	22.67	374.63
17	768.93	0.06	32.66	1117.74	0.09	1.59	958.00	1.77	181.62	1119.79	17.61	283.75
18	765.34	0.05	56.82	1123.14	0.05	2.72	958.00	2.70	395.10	1123.14	11.33	224.48
19	779.72	0.06	65.70	1082.58	0.03	3.42	958.00	4.13	397.85	1082.95	12.33	224.25
20	762.75	0.06	41.96	1097.08	0.09	5.16	958.00	4.00	315.23	1100.50	16.69	228.16
21	774.69	0.13	50.74	1108.22	0.08	3.41	958.00	2.36	246.92	1109.60	16.14	279.47
22	795.34	0.06	91.14	1113.08	0.06	2.55	958.00	4.19	253.13	1113.75	17.66	267.28
23	757.71	0.06	41.80	1081.75	0.03	2.89	958.00	2.66	317.66	1081.75	15.27	281.34
24	769.43	0.06	91.60	1093.54	0.06	2.31	958.00	7.41	477.63	1094.75	16.44	271.05
25	765.51	0.06	56.56	1101.50	0.03	1.09	958.00	2.52	291.42	1102.50	12.81	237.42
26	753.77	0.06	77.62	1082.92	0.03	1.83	958.00	3.69	352.53	1082.92	13.11	264.45
27	776.41	0.17	27.74	1104.65	0.06	3.55	958.00	4.27	223.52	1108.42	13.25	263.36
28	779.72	0.06	66.24	1112.00	0.03	2.30	958.00	2.98	328.46	1112.00	14.75	155.94
29	775.04	0.06	40.74	1133.77	0.08	0.88	958.00	1.44	147.29	1135.35	12.06	191.89
30	762.83	0.08	67.76	1102.31	0.03	1.36	958.00	1.67	167.99	1102.31	13.86	216.59
Average	774.22	0.08	56.26	1107.92	0.06	2.28	958.00	3.15	314.44	1109.24	15.12	266.30
Stand-dev	11.67	0.04	20.82	16.20	0.03	1.08	0.00	1.51	110.50	16.71	3.85	82.63
Minimum	753.77	0.05	24.03	1080.25	0.03	0.88	958.00	1.44	125.82	1080.25	8.08	112.89
Maximum	795.34	0.17	99.38	1148.50	0.19	5.16	958.00	7.41	627.80	1148.63	22.67	515.95

Table C1
md-MST results to the CRD and SYM instances with $n = 30$, involving the LP relaxation information of the models under consideration

md-MST				Opt	LP relaxation optimums					LP relaxation times				
Type	n	d	I		*md*-SCF_L	*md*-$SCF1_L$	*md*-$SCF3_L$	*md*-MCF_L	*md*-$MCF1_L$	*md*-SCF_L	*md*-$SCF1_L$	*md*-$SCF3_L$	*md*-MCF_L	*md*-$MCF1_L$
CRD	30	3	1	4026	3140.54	3498.54	3649.65	3634.00	3764.33	0.06	0.13	0.14	5.47	23.73
			2	3793	2711.92	3069.10	3227.50	3277.00	3613.67	0.06	0.11	0.13	3.56	19.70
			3	4293	3248.14	3575.31	3902.49	4001.00	4124.50	0.06	0.13	0.13	3.56	7.53
Average gaps and times					*24.945*	*16.301*	*11.118*	*10.047*	*5.052*	*0.06*	*0.72*	*0.13*	*4.20*	*16.99*
SYM	30	3	1	1197	775.04	1087.10	1148.50	958.00	1148.63	0.06	0.11	0.03	1.48	8.42
			2	1435	1069.76	1302.42	1395.06	1219.00	1395.25	0.05	0.08	0.09	0.47	3.91
			3	1408	1161.66	1301.89	1392.27	1252.00	1393.67	0.06	0.09	0.13	0.98	6.00
Average gaps and times					*26.066*	*8.652*	*2.651*	*15.366*	*2.670*	*0.06*	*0.09*	*0.08*	*0.98*	*6.11*
CRD	30	5	1	5026	3140.54	4554.44	4564.85	3634.00	4626.96	0.06	0.19	0.22	4.92	20.09
			2	4648	2711.92	4173.99	4216.06	3277.00	4437.56	0.06	0.16	0.20	3.52	28.11
			3	5425	3248.14	4931.87	4942.51	4001.00	5034.61	0.06	0.19	0.27	3.52	42.08
Average gaps and times					*39.765*	*9.557*	*9.121*	*27.814*	*6.554*	*0.06*	*0.18*	*0.23*	*3.99*	*30.09*
SYM	30	5	1	1765	775.04	1603.42	1645.57	958.00	1647.62	0.06	0.14	0.23	1.48	12.28
			2	2090	1069.76	1909.91	1928.27	1219.00	1939.46	0.06	0.09	0.09	0.44	9.61
			3	2008	1161.66	1953.33	1967.92	1252.00	1968.30	0.05	0.09	0.20	0.97	20.25
Average gaps and times					*49.017*	*6.831*	*5.500*	*41.682*	*5.277*	*0.06*	*0.11*	*0.17*	*0.96*	*14.05*

Table C2
md-MST results to the CRD and SYM instances with $n = 30$, involving the branch-and-bound information using the models under consideration

md-MST				Opt	LP relaxation optimums					B&B times				
Type	n	d	I		*md*-SCF_L	*md*-$SCF1_L$	*md*-$SCF3_L$	*md*-MCF_L	*md*-$MCF1_L$	*md*-SCF	*md*-SCF1	*md*-SCF3	*md*-MCF	*md*-MCF1
CRD	30	3	1	4026	3140.54	3498.54	3649.65	3634.00	3764.33	5664.58	3852.75	305.97	1301.20	940.58
			2	3793	2711.92	3069.10	3227.50	3277.00	3613.67	<u>10800.00</u>	<u>10800.00</u>	1471.53	1289.38	1420.98
			3	4293	3248.14	3575.31	3902.49	4001.00	4124.50	7524.84	<u>10800.00</u>	30.30	590.19	394.38
Average gaps and times					*24.945*	*16.301*	*11.118*	*10.047*	*5.051*	*7996.47*	*8484.25*	*602.60*	*1060.26*	*918.65*
SYM	30	3	1	1197	775.04	1087.10	1148.50	958.00	1148.63	24.03	12.45	0.92	183.17	112.89
			2	1435	1069.76	1302.42	1395.06	1219.00	1395.25	9.53	21.20	0.97	52.25	40.13
			3	1408	1161.66	1301.89	1392.27	1252.00	1393.67	8.86	24.22	0.19	47.63	35.50
Average gaps and times					*26.066*	*8.652*	*2.651*	*15.366*	*2.610*	*14.14*	*19.29*	*0.69*	*94.35*	*62.84*
CRD	30	5	1	5026	3140.54	4554.44	4564.85	3634.00	4626.96	566.09	647.16	352.77	<u>10800.00</u>	<u>10800.00</u>
			2	4648	2711.92	4173.99	4216.06	3277.00	4437.56	424.45	358.89	53.87	4440.06	2244.55
			3	5425	3248.14	4931.87	4942.51	4001.00	5034.61	481.45	463.75	266.81	<u>10800.00</u>	<u>10800.00</u>
Average gaps and times					*39.765*	*9.557*	*9.121*	*27.814*	*6.554*	*490.66*	*489.93*	*224.48*	*8680.02*	*7948.18*
SYM	30	5	1	1765	775.04	1603.42	1645.57	958.00	1647.62	16.53	35.80	8.94	1244.56	1911.27
			2	2090	1069.76	1909.91	1928.27	1219.00	1939.46	31.33	28.97	4.63	1079.91	465.58
			3	2008	1161.66	1953.33	1967.92	1252.00	1968.30	3.47	2.03	1.06	255.25	103.39
Average gaps and times					*49.017*	*6.831*	*5.500*	*41.682*	*5.277*	*17.11*	*22.27*	*4.88*	*859.91*	*826.75*

Table C3
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							LP relaxation times						
Type	n	d	I		$md\text{-SCF}_L$	$md\text{-SCF1}_L$	$md\text{-SCF2}_L$	$md\text{-SCF3}_L$	$md\text{-MCF}_L$	$md\text{-MCF1}_L$	$md\text{-MCF2}_L$	$md\text{-SCF}_L$	$md\text{-SCF1}_L$	$md\text{-SCF2}_L$	$md\text{-SCF3}_L$	$md\text{-MCF}_L$	$md\text{-MCF1}_L$	$md\text{-MCF2}_L$
CRD	50	3	1	5512	3980.74	4710.49		4939.24	4931.00	5230.48		0.59	1.95		2.70	119.00	139.97	
			2	5827	4225.13	4965.24		5269.93	5126.00	5456.25		0.63	2.50		2.44	66.22	250.42	
			3	5590	3588.42	4981.03		5162.14	4898.00	5391.18		0.55	2.30		1.58	31.25	420.70	
Average gaps and times					*30.359*	*13.408*		*9.202*	*11.650*	*5.009*		*0.59*	*2.25*		*2.24*	*72.16*	*270.36*	
SYM	50	3	1	1278	949.02	1150.98		1230.65	1098.00	1235.84		0.64	3.08		1.38	22.89	541.66	
			2	1178	910.17	1056.28		1120.25	1045.00	1127.50		0.66	2.11		0.09	16.75	144.70	
			3	1615	1285.51	1521.98		1580.39	1416.00	1589.50		0.69	2.75		0.80	127.83	221.30	
Average gaps and times					*22.960*	*8.677*		*3.583*	*72.566*	*3.055*		*0.66*	*2.65*		*0.76*	*55.82*	*302.55*	
CRD	50	5	1	6908	3980.74	6050.31		6069.36	4931.00	6337.79		0.59	1.92		2.95	119.31	351.41	
			2	7204	4225.13	6640.76		6667.33	5126.00	6763.00		0.66	2.03		2.47	65.78	329.41	
			3	7285	3588.42	6506.27		6526.55	4898.00	6694.90		0.53	1.81		2.58	30.66	449.28	
Average gaps and times					*44.822*	*11.308*		*10.000*	*30.077*	*7.492*		*0.59*	*1.92*		*2.67*	*71.92*	*376.70*	
SYM	50	5	1	2054	949.02	1800.88		1840.23	1098.00	1840.23		0.64	1.77		0.41	18.78	99.72	
			2	1760	910.17	1571.90		1640.48	1045.00	1662.87		0.67	1.78		1.53	13.20	260.30	
			3	2525	1285.51	2286.16		2340.57	1416.00	2352.78		0.69	1.42		0.47	94.88	158.13	
Average gaps and times					*50.390*	*10.823*		*8.168*	*43.696*	*7.582*		*0.67*	*1.66*		*0.80*	*42.29*	*172.72*	
CRD	50	10	1	9633	3980.74	8806.41	8991.51	8991.51	4931.00	8944.67	9120.18	0.59	3.44	3.41	4.66	118.75	461.42	433.16
			2	9743	4225.13	9212.75	9369.00	9369.00	5126.00	9226.52	9391.19	0.63	4.80	3.11	4.31	65.69	357.38	382.41
			3	9855	3588.42	9258.72	9438.42	9438.42	4898.00	9314.69	9482.42	0.55	3.05	2.27	3.69	30.69	332.05	315.44
Average gaps and times					*59.633*	*6.691*	*4.908*	*4.908*	*48.833*	*5.976*	*4.238*	*0.59*	*3.76*	*2.93*	*4.22*	*71.71*	*383.62*	*377.00*
SYM	50	10	1	4121	949.02	3607.24	3724.49	3724.49	1098.00	3607.24	3724.49	0.64	3.06	0.73	1.22	18.55	110.95	89.03
			2	4166	910.17	3467.72	3627.67	3628.90	1045.00	3495.81	3647.60	0.64	1.92	0.77	2.34	12.97	191.56	143.14
			3	4979	1285.51	4223.18	4373.42	4373.42	1416.00	4225.91	4374.45	0.69	3.42	1.91	3.73	110.41	202.53	144.84
Average gaps and times					*76.435*	*14.803*	*11.569*	*11.559*	*73.278*	*14.560*	*11.402*	*0.66*	*2.80*	*1.14*	*2.43*	*47.31*	*168.35*	*125.67*

Table C4
md-MST results to the CRD and SYM instances with $n = 50$, involving the branch-and-bound information using 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-SCF3$_L$	md-MCF$_L$	md-MCF1$_L$	md-MCF2$_L$	md-SCF$_L$	md-SCF1$_L$	md-SCF2$_L$	md-SCF3$_L$	md-MCF$_L$	md-MCF1$_L$	md-MCF2$_L$
CRD	50	3	1	5512	3980.74	4710.49		4939.24	4931.00	5230.48		10800.00	10800.00		10800.00	10800.00	10800.00	
			2	5827	4225.13	4965.24		5269.93	5126.00	5456.25		10800.00	10800.00		10800.00	10800.00	10800.00	
			3	5590	3588.42	4981.03		5162.14	4898.00	5391.18		10800.00	10800.00		10800.00	10800.00	10800.00	
Average gaps and times					*30.359*	*13.408*		*9.202*	*11.650*	*5.009*		*10800.00*	*10800.00*		*10800.00*	*10800.00*	*10800.00*	
SYM	50	3	1	1278	949.02	1150.98		1230.65	1098.00	1235.84		7648.38	4778.80		24.28	B796.09	4021.45	
			2	1178	910.17	1056.28		1120.25	1045.00	1127.50		5685.80	3888.14		8.50	2254.31	3025.95	
			3	1615	1285.51	1521.98		1580.39	1416.00	1589.50		713.53	1530.61		12.32	2706.63	1997.97	
Average gaps and times					*22.960*	*8.677*		*3.583*	*12.566*	*3.055*		*4682.57*	*3399.18*		*75.03*	*4585.68*	*1015.12*	
CRD	50	5	1	6908	3980.74	6050.31		6069.36	4931.00	6337.79		10800.00	10800.00		10800.00	10800.00	10800.00	
			2	7204	4225.13	6640.76		6667.33	5126.00	6763.00		10800.00	10800.00		10800.00	10800.00	10800.00	
			3	7285	3588.42	6506.27		6526.55	4898.00	6694.90		10800.00	10800.00		10800.00	10800.00	10800.00	
Average gaps and times					*44.822*	*11.308*		*10.000*	*30.077*	*7.492*		*10800.00*	*10800.00*		*10800.00*	*10800.00*	*10800.00*	
SYM	50	5	1	2054	949.02	1800.88		1840.23	1098.00	1840.23		10800.00	6646.47		138.35	10800.00	10800.00	
			2	1760	910.17	1571.90		1640.48	1045.00	1662.87		2239.36	3940.92		115.23	10800.00	10800.00	
			3	2525	1285.51	2286.16		2340.57	1416.00	2352.78		10800.00	10800.00		264.17	10800.00	10800.00	
Average gaps and times					*50.390*	*70.823*		*8.168*	*43.696*	*7.582*		*7946.45*	*7129.13*		*172.58*	*10800.00*	*10800.00*	
CRD	50	10	1	9633	3980.74	8806.41	8991.51	8991.51	4931.00	B944.67	9120.18	4072.59	5443.88	417.66	526.52	10800.00	10800.00	10800.00
			2	9743	4225.13	9212.75	9369.00	9369.00	5126.00	9226.52	9391.19	5083.28	5285.30	696.55	755.01	10800.00	10800.00	10800.00
			3	9855	3588.42	9258.72	9438.42	9438.42	4898.00	9314.69	9482.42	3428.01	7582.11	646.45	544.75	10800.00	10800.00	10800.00
Average gaps and times					*59.633*	*6.691*	*4.908*	*4.908*	*48.833*	*5.976*	*4.238*	*4194.63*	*6103.76*	*586.89*	*608.76*	*10800.00*	*10800.00*	*10800.00*
SYM	50	10	1	4121	949.02	3607.24	3724.49	3724.49	1098.00	3607.24	3724.49	1728.45	3093.78	268.98	61.49	10800.00	10800.00	10800.00
			2	4166	910.17	3467.72	3627.67	3628.90	1045.00	3495.81	3647.60	1821.20	4960.19	534.63	147.24	10800.00	10800.00	10800.00
			3	4979	1285.51	4223.18	4373.42	4373.42	1416.00	4225.91	4374.45	3451.19	10800.00	2579.11	1173.76	10800.00	10800.00	10800.00
Average gaps and times					*76.435*	*14.803*	*11.569*	*11.559*	*73.278*	*14.560*	*11.402*	*2333.61*	*6284.66*	*1127.57*	*460.83*	*10800.00*	*10800.00*	*10800.00*