RAPPORT DE RECHERCHE

L R I

UNE FORMULATION POLYNOMIALE POUR LE PROBLEME STOCHASTIQUE DE LA FORET DE POIDS MAXIMUM

ADASME P / ANDRADE R / LETOURNEL M / LISSER A

Unité Mixte de Recherche 8623
CNRS-Université Paris Sud – LRI

11/2012

Rapport de Recherche N° 1554

CNRS – Université de Paris Sud
Centre d'Orsay
LABORATOIRE DE RECHERCHE EN INFORMATIQUE
Bâtiment 650
91405 ORSAY Cedex (France)

Laboratoire de Recherche en Informatique, Université de Paris-Sud
Bât. 650, 91405, Orsay Cedex France.

Rapport de Recherche

Titre

Une formulation polynomiale pour le problème stochastique de la forêt de poids maximum [1]

Auteurs

Pablo Adasme[2]
Rafael Andrade[3]
Marc Letournel et Abdel Lisser[4]

2012

[1]Projet réalisé en collaboration avec l'Université Fédérale de Ceará (Fortaleza - Brésil) et avec l'Université de Santiago de Chile. Recherche financée par le Conseil National de Recherche Scientifique (CNPq/Brésil – Proc. 201347/2011-3) et par la Commission Nationale Chilienne pour la Recherche Scientifique et Technologique (CONICYT/Chili - Proc. 79100020).

[2]Departamento de Ingenieria Electrica, Universidad de Santiago de Chile, Avenida Ecuador 3519, Santiago, Chile. Email: pablo.adasme@usach.cl

[3]Departamento de Estatística e Matemática Aplicada, Universidade Federal do Ceará, Campus do Pici - Bloco 910, 60455-760, Fortaleza, Ceará, Brasil. Email: rca@lia.ufc.br

[4]Laboratoire de Recherche en Informatique, Université de Paris-Sud, Bât. 650, 91405, Orsay Cedex France. Email: {marc.letournel,abdel.lisser}@lri.fr

Résumé : Le problème stochastique de la forêt de poids maximum (SMWF) est une extension de sa version déterministe classique où certaines arêtes présentent des poids incertains. Différemment du cas déterministe, la complexité du problème SMWF est inconnue. Curieusement, sa formulation en fonction des contraintes classiques d'élimination de sous-tour n'a pas la propriété TDI et, en raison du nombre exponentiel de contraintes de cette formulation, seules des instances de petites dimensions peuvent être traitées. Ainsi, nous développons une nouvelle formulation compacte pour le problème SMWF. Elle est basée sur les travaux de Martin (1991) pour le polytope de l'arbre couvrant avec une adaptation pour traiter des forêts dans des graphes non complets. Nous donnons une preuve de la validité de la nouvelle formulation, tout en utilisant un nouveau théorème caractérisant les forêts en graphes. Nous rapportons quelques expériences numériques pour ce problème.
Mots-clés: formulation compacte, problème stochastique de la forêt de poids maximum.

A polynomial formulation for the stochastic maximum weight forest problem

Pablo Adasme[1], Rafael Andrade[2], Marc Letournel[3] and Abdel Lisser[3]

[1] Departamento de Ingenieria Electrica, Universidad de Santiago de Chile, Avenida Ecuador 3519, Santiago, Chile.
pablo.adasme@usach.cl

[2] Departamento de Estatística e Matemática Aplicada, Universidade Federal do Ceará, Campus do Pici - Bloco 910, 60455-760, Fortaleza, Ceará, Brasil.
rca@lia.ufc.br

[3] Laboratoire de Recherche en Informatique, Université de Paris-Sud XI, Bâtiment 650, 91405, Orsay Cedex France.
{marc.letournel,abdel.lisser}@lri.fr

Abstract. The stochastic maximum weight forest (SMWF) problem is an extension of the classic deterministic version [6] where some edges present uncertain weights. Different from the deterministic case, the complexity of the SMWF problem is unknown. Surprisingly, its formulation based on sub-tour elimination constraints [2] has not the TDI property [3] and due to the exponential number of constraints of this formulation, only instances with small dimensions can be handled. Thus, we develop a new compact extended formulation for the SMWF problem. It is based on the work of [4] for the spanning tree polytope with a straightforward adaptation for dealing with forests in non complete graphs. We give a proof of the correctness of the new formulation based on a new theorem characterizing forests in graphs. Numerical results are reported and show that the model with exponential number of constraints can treat only instances with up to 15 nodes in complete graphs, while the new polynomial model can treat instances with up to 40 nodes, considering few scenarios of uncertain edge weights.

Keywords: combinatorial optimization, compact extended formulation, stochastic maximum weight forest problem

1 Introduction

This work concerns the two stage stochastic maximum weight forest (SMWF) problem [3]. It is a variant of the maximum weight forest (MWF) problem where some edges present uncertain weights. While the MWF problem is polynomial [6], there is no similar result concerning the complexity of the SMWF problem.

The problem can be described as follows. Consider $G = (V, E_D \cup E_S)$ a non directed graph with set of nodes V and set of weighted edges $E := E_D \cup E_S$, where edges in E_D and E_S have deterministic and uncertain weights, respectively, with $E_D \cap E_S = \emptyset$. Consider a finite set $S = \{1, 2, \cdots, P\}$ of possible scenarios for the uncertain edges in E_S and let π_s be the probability associated with a given scenario $s \in S$, with $\sum_{s=1}^{P} \pi_s = 1$ and $\pi_s \geq 0$, for all $s \in S$. Let $c_{uv}^s \in \mathbb{R}$ represent the weight of an edge $uv \in E$ in scenario $s \in S$, with $c_{uv}^1 = c_{uv}^2 = \cdots = c_{uv}^P$ given for all $uv \in E_D$ and c_{uv}^s, for all $s \in S$, randomly generated a priori for all $uv \in E_S$. The SMWF problem consists in determining a forest of G, one for each $s \in S$, sharing the same deterministic edges in E_D. As objective function, we want to maximize the sum of the common deterministic edges weights plus the expectation over all scenarios, of the weights associated with the uncertain edges whose union with the deterministic ones keeps the forest structure in each scenario.

As contributions of this work, we propose a new compact extended formulation for the SMWF problem based on the spanning tree polytope of complete undirected graphs [4, 5]. We extend the model proposed in [4] for the SMWF and prove the correctness of the new formulation. For this, we introduce a new characterization theorem for forests in graphs. With this compact extended formulation we can now provide optimal integer solutions for instances with up to 40 nodes. The compact formulation describing trees [4] is intended to give integer optimal solutions, thus allowing to develop a decomposition algorithm for the SMWF problem. However, here we do not report the decomposition scheme. But this will be done in a full version of this paper, as well as characterizing theoretically which type of SMWF instances are easy or difficult to solve.

The remaining of this paper is organized as follows. In Section 2 we present standard formulations for the SMWF and the forest characterization theorem used to prove the correctness of the compact extended formulation presented in Section 3. In Section 4 we provide preliminary experiments for the proposed models. Finally, Section 5 concludes this paper and present some works in progress.

2 Problem formulation

Let represent a forest of G in scenario $s \in S$ by a vector $(x_D^s, x_S^s) = x^s \in \{0,1\}^{|E|}$, where $x_{uv}^s = 1$ if uv belongs to the forest of that scenario, and $x_{uv}^s = 0$, otherwise. A mathematical model for the SMWF problem is

$$(P_1) \quad \max \sum_{uv \in E_D} c_{uv}^1 x_{uv}^1 + \sum_{uv \in E_S,\, s \in S} \pi_s c_{uv}^s x_{uv}^s \tag{1}$$

$$s.t. \quad \sum_{uv \in E(H)} x_{uv}^s \leq |H| - 1, \; \forall \, H \subset V, \; \forall \, s \in S \tag{2}$$

$$x_D^1 = x_D^2 = \cdots = x_D^P \tag{3}$$

$$x_{uv}^s \in \{0,1\}, \forall \, uv \in E, \;\; \forall \, s \in S \tag{4}$$

where $E(H)$ stands for the set of edges with both extremities in H, such that $|H| \geq 3$. Constraints (2) impose the non existence of cycles in any solution and (3) indicate that the deterministic edges in an optimal solution are the same for all scenarios. Commonly, (3) are referred to as non anticipativity constraints. As each block of cycle elimination constraints (2), one for each scenario, has an exponential number of constraints in (P_1), we can expect to tackle only instances with a small number of nodes and scenarios by using this formulation. Note that we can reduce the number of variables in (P_1) to obtain the following equivalent formulation [3]

$$(P_2) \quad \max \sum_{j \in E_D} c_j^1 y_j + \sum_{s \in S} \pi_s \left(\sum_{j \in E_S} c_j^s x_j^s \right) \tag{5}$$

$$s.t. \sum_{j \in E(H) \cap E_D} y_j + \sum_{j \in E(H) \cap E_S} x_j^s \leq |H| - 1, \ \forall\, H \subset V, \ \forall\, s \in S \tag{6}$$

$$x_j^s \in \{0,1\}, \ \forall\, j \in E_S, \ \ \forall\, s \in S \tag{7}$$

$$y_j \in \{0,1\}, \ \forall\, j \in E_D \tag{8}$$

where $y \in \{0,1\}^N$ replace the $x_D^1 = x_D^2 = \cdots = x_D^P$ deterministic variables and $x^s \in \{0,1\}^M$, for all $s \in S$, are as defined above, with $|E_D| = N$ and $|E_S| = M$.

Proposition 1. *The maximum weight forest problem (i.e. the SMWF with $|S| = 1$) is polynomially solvable by a greedy algorithm [2].*

Proposition 2. *The SMWF formulation is not totally dual integral (TDI).*

In [3] the authors present a small example for a six nodes graph and three scenarios of edges weights where the linear relaxed solution is not integer. This proves Proposition 2. The fact that the problem may present fractional linear relaxed solution makes it difficult to deal with.

Now let, for a given node $k \in V$ and a given scenario $s \in S$, F_k^s represent an abstract orientation of the edges of a general subgraph F of G such that the node k in the scenario s is taken as a referential in F. In this case, if an edge $(v,u) \in E$ belongs to F, then the referential node k observes u preceding $v \neq k$ or $v \neq k$ preceding u, but not both, in F_k^s.

Theorem 1. *Let F^s be a subgraph of G for scenario $s \in S$. For all $k \in V$, if there exist independent abstract orientations F_k^s of the edges of F^s, for $k = 1, \cdots, |V|$, verifying simultaneously the following conditions in each F_k^s*

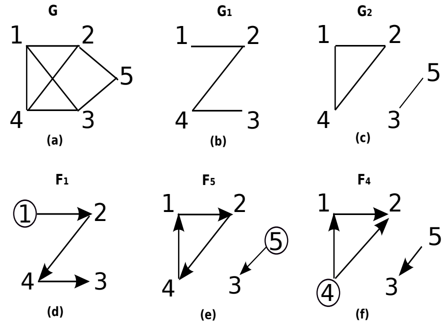
1. *There is no arc entering each referential node k;*
2. *There is at most one arc entering a node $u \neq k$;*

then F^s is an acyclic forest.

Consider Figure 1 to help following the proof of Theorem 1.

Proof. Suppose that F^s contains a cycle $C = (V(C), E(C))$ for some scenario s. We show that F^s cannot satisfy both conditions above for all $k \in V$ in scenario $s \in S$. To see this, first consider the nodes not in $V(C)$ as referential nodes and assume without loss of generality that they do not induce a cycle in F^s. For these nodes, if the arcs of the cycle C were symbolically oriented in the clockwise sense, both conditions 1 and 2 should be satisfied. But when we consider the nodes in

Fig. 1. A graph G in (a) with two subgraphs G_1 in (b) and G_2 in (c) with some orientations for them.



$V(C)$ as referential in any abstract orientation of the edges of $E(C)$, there are at least two arcs connected to any node of $V(C)$ (two leaving or two entering or one leaving and the other entering each node). We distinguish two possible situations. First, if all arcs connected to any referential node of C leave that node, then at least one other node in this cycle presents two arcs entering it, which violates the condition 2. Second, if the arcs of C were symbolically oriented in the clockwise sense, then the condition 1 should be violated for all referential node in C. Thus, if F^s contain a cycle, there is no possibility of orienting the edges of F_k^s, $k = 1, \cdots, |V|$, in order to satisfy the conditions 1 and 2 simultaneously. Therefore F^s must be a forest, thus concluding the proof. □

The reader may notice that when F^s is a forest, then always exist abstract orientations of the edges of F_k^s, $k = 1, \cdots, |V|$, satisfying the two conditions of Theorem 1.

In the Figure 1(d), F_1 is a possible orientation of the edges of G_1 with the node 1 as referential. The scenario indexing is omitted in this figure and referential nodes are pointed by a circle. In (e), F_5 is a possible orientation of the edges of G_2 that respects the conditions 1 and 2 of the Theorem 1 with the node 5 as referential. However, the one F_4 for G_2 in (f) violates the condition 2 (the node 2 has two arcs entering it).

Corollary 1. *If subgraph F^s in the Theorem 1 is such that $|E(F^s)| = |V(G)|-1$, then F^s is a spanning tree of G in scenario s.*

Below we present a new model for this problem by taking into account the above theorem and some ideas originally introduced in [4].

3 Polynomial formulation

Consider the sets $E := E_S \cup E_D$ and $E^+ := E \cup \{(u, v) \in V \times V \mid u \neq v,\ (u, v) \notin E\}$. Note that E^+ contains all possible edges for G. Let, for all scenario $s \in S$,

$x^s \in \{0,1\}^{|E^+|}$ be the characteristic vector of a general solution (forest) F^s in scenario s for the SMWF problem with coordinates x^s_{uv} for all $uv \in E^+$. Define, for every $s \in S$, $k \in V$ and $(i,j) \in E^+$, binary variables λ^s_{kij} and λ^s_{kji}, such that $\lambda^s_{kij} = 1$ if edge (i,j) belongs to the solution of the scenario s and, for the referential node k, node j precedes node i in an abstract orientation F^s_k of the edges of F^s; and $\lambda^s_{kij} = 0$, otherwise. The SMWF problem can be modeled as

$$(P_3) \quad \max \sum_{uv \in E_D} c^1_{uv} x^1_{uv} + \sum_{uv \in E_S,\, s \in S} \pi_s c^s_{uv} x^s_{uv} \tag{9}$$

$$s.t. \quad \lambda^s_{kij} + \lambda^s_{kji} = x^s_{ij}, \quad \forall\, s \in S,\ \forall\, k \in V,\ \forall\, ij \in E \tag{10}$$

$$\sum_{j \in V - \{i\}} \lambda^s_{kij} \leq 1, \quad \forall\, s \in S,\ \forall\, i, k \in V,\ i \neq k \tag{11}$$

$$\lambda^s_{kkj} = 0, \quad \forall\, s \in S,\ \forall\, j, k \in V,\ k \neq j \tag{12}$$

$$x^s_{uv} = 0, \quad \forall\, s \in S,\ \forall\, uv \in E^+ - E \tag{13}$$

$$x^1_{E_D} = x^2_{E_D} = \cdots = x^P_{E_D} \tag{14}$$

$$\lambda \in \{0,1\}^{|V \times E^+ \times S|} \tag{15}$$

$$x \in \{0,1\}^{|S \times E^+|} \tag{16}$$

Constraints (10) state that if an edge (i,j) is in a solution of the scenario s (i.e. $x^s_{ij} = 1$), then for all referential node $k \in V$ either j precedes i or i precedes j in F^s_k. Constraints (11) limit the number of predecessors of any node i in F^s_k to at most one, with $i \neq k$, in order to satisfy the condition 2 of the Theorem 1. Constraints (12) state that no node j can precede the referential node k in F^s_k (the condition 1 of the Theorem 1). Constraints (13) fix at zero all the corresponding variables related to the extra edges we add to make G a complete graph. Note that the λ variables of each scenario (those associated with existing edges of E) induce an abstract orientation (one for each referential k) of the edges present in any feasible solution for problem (P_3) and these orientations satisfy the two conditions of the Theorem 1. Equalities in (14) are the non anticipativity constraints and (15) and (16) express the domain of the problem variables.

Proposition 3. *The projection on the space of the x variables of an optimal solution $(\bar{\lambda}, \bar{x})$ for (P_3), if it is non empty, is a set of forests, one for each scenario, maximizing (9).*

Proof. Suppose, by contradiction, that F^s is feasible for (P_3) and contains a cycle $C^s = (V(C^s), E(C^s))$ in some scenario $s \in S$ induced by the variables $\bar{x}^s$. For all the edges uv in $E(C^s)$ we have that $\bar{x}^s_{uv} = 1$. Thus, by (10), we must have $\lambda^s_{kuv} + \lambda^s_{kvu} = 1$, for all $k \in V(C^s)$ and for all $uv \in E(C^s)$. In particular, these λ^s must verify (11) and (12). But (see the proof of the Theorem 1) finding one evaluation for these variables means orienting the edges in $E(C^s)$ such that every referential node $k \in V(C^s)$ have no incoming arc and the remaining nodes in $V(C^s) - \{k\}$ have at most one incoming arc. Therefore, this is not possible and thus the related constraints correspond to an infeasible system of linear inequalities, contradicting the assumption that F^s is feasible for (P_3). The reader may notice that when a solution F^s is a forest, then there exist variables $\bar{\lambda}^s$ and $\bar{x}^s$ whose values correspond to such solution. From the edges in F^s we obtain the $\bar{x}^s$ values and if we apply an algorithm to remove iteratively the leaves from every sub-tree of the forest F^s we can identify which node can be considered as sub referential. If this deletion process reaches an edge, any extremity of this edge can be considered as the sub-referential node $\hat{k}$ to obtain the required orientation with respect to the main referential node k. This orientation can be obtained from the $\hat{k}$-rooted pending sub-tree structure. The solution optimality follows from the objective function in (9). □

4 Computational experiments

Numerical experiments have been carried out on a Pentium IV, 1 GHz with 2G-RAM under windows XP. The source codes are generated with Matlab and the optimization models are solved by CPLEX 12. We report preliminary results for randomly and arbitrarily generated instances. We generate the random input data as follows. The edge weights for the first 21 instances in Table 1 are distributed uniformly into the interval $[1, 100]$ and uncertain edge weights are made negative with probability 1/2 in each scenario. The scenarios probabilities π_s are chosen randomly from the interval $(0, 1)$ and then normalized so that $\sum_{s \in S} \pi_s = 1$. We use arbitrarily up to 100 scenarios of uncertain edge weights. For the remaining instances (numbered from 22 to 26) we create them arbitrarily trying to make the problem difficult to solve (this explains the reduced number of scenarios and dimensions of these instances). In the next tables, (RP_2) and (RP_3) correspond to the linear relaxations of the models (P_2) and (P_3), respectively.

The dimensions of the instances (i.e. the number of constraints and of variables) for these models are in the Table 1. The first column of this table identifies the instance number. Columns 2-5 provide the parameters $|V|$, $N = |E_D|$, $M = |E_S|$ and $P = |S|$. Columns 6-7 and 8-9 show an estimate of the order of the number of constraints and variables for (RP_2) and (RP_3), respectively. We exclude variable fixing and non negativity bounds, i.e. considering only the constraints (6) for (RP_2) (the variables upper bounds are induced by these constraints) and constraints (10), (11), (14) with the variables upper bounds (15) and (16) for (RP_3). For the model (RP_2), the number of variables is of the order of $\mathcal{O}(N + MP)$ and the number of constraints, $\mathcal{O}(2^{|V|}P)$. For the model (P_3), the number of variables and constraints are of the order of $\mathcal{O}(P|V|^3)P$. In this table we have an idea of how fast the number of constraints grows for (RP_2). This growth is highly affected by the increase in the number of scenarios and nodes. Observe that for $|V| \geq 10$, the number of constraints of the instances in the model (RP_2) became larger than the ones in the model (RP_3).

In Table 2 we report numerical results for instances presenting all uncertain edges weights with positive values. Thus, the scenarios solutions are spanning trees in each instance. Column *Inst.* identifies the instance from the Table 1.

Table 1. Instances dimensions

Inst.	Instance parameters				(RP_2)		(RP_3)	
	$\lvert V\rvert$	N	M	P	# Constr.	# Var.	# Constr.	# Var.
1	5	5	5	5	130	30	675	550
2	5	5	5	50	1300	255	6750	5500
3	5	5	5	100	2600	505	13500	11000
4	8	14	14	5	1235	84	2730	2380
5	8	14	14	50	12350	714	27300	23800
6	8	14	14	100	24700	1414	54600	47600
7	10	22	23	5	5065	137	5285	4725
8	10	22	23	50	50650	1172	52850	47250
9	10	22	23	100	101300	2322	105700	94500
10	12	33	33	5	20415	198	9075	8250
11	12	33	33	50	204150	1683	90750	82500
12	12	33	33	100	408300	3333	181500	165000
13	15	52	53	5	163760	317	17585	16275
14	15	52	53	50	1637600	2702	175850	162750
15	15	52	53	100	3275200	5352	351700	325500
16	20	95	95	5	5242775	570	41325	38950
17	20	95	95	10	10485550	1045	82650	77900
18	20	95	95	25	26213875	2470	206625	194750
19	30	217	218	5	5368708965	1307	138110	132675
20	30	217	218	10	10737417930	2397	276220	265350
21	40	390	390	5	5497558138675	2340	325650	315900
22	6	3	6	3	171	21	684	585
23	6	3	6	3	171	21	684	585
24	8	4	10	4	988	44	2144	1904
25	8	4	11	5	1235	59	2680	2380
26	14	6	30	4	65476	126	11308	10556

Columns 2-3, 4-5 and 6-7 give the optimal solution value and cpu time in seconds for (P_2), (RP_2) and (RP_3), respectively.

In Table 3 we report numerical results for the same instances of the Table 1, but where some uncertain edges are allowed to have negative weights. Thus, the solutions are intended to be sets of forests in each instance, one for each scenario. The legend is the same as for the Table 2. Almost a half of the edges of each instance present negative weights.

In Table 4 we report numerical results for the arbitrarily generated instances of the Table 1. These instances intend to show the problem difficulty since the algorithm we use to create random instances is not capable to find instances presenting non integer relaxed solutions. The technique we use for obtaining them is omitted here, but it will be reported in a full version of this work. These results suggest that the SMWF problem may be NP-hard.

Concerning the results from the Tables 2 and 3, we mainly have that the optimal relaxed solutions are all integer. This means that these random generated instances proved easy to solve and it seems extremely complicated that random instances do not present this behavior. It is not surprisingly that the model (P_2) is very limited in solving those instance with more than 13 nodes. However, generally solving their corresponding linear relaxation (RP_2) takes less cpu time than by using the linear relaxation (RP_3). Note that CPLEX finds integer solutions for (RP_2) in smaller cpu time than for (P_2) for these instances.

Table 2. Numerical results for random generated instances where all uncertain edges have positive weights in every scenario.

Inst.	(P_2)	cpu(s)	(RP_2)	cpu(s)	(RP_3)	cpu(s)
1	323.3891	0.36	323.3891	0.34	323.3891	0.39
2	333.2423	0.40	333.2423	0.37	333.2423	1.20
3	259.5762	0.51	259.5762	0.42	259.5762	3.18
4	572.0971	0.46	572.0971	0.39	572.0971	0.51
5	616.8961	3.01	616.8961	0.95	616.8961	14.39
6	570.0392	5.71	570.0392	2.62	570.0392	24.09
7	821.7794	1.21	821.7794	0.75	821.7794	0.70
8	806.2184	17.35	806.2184	6.92	806.2184	84.34
9	789.5104	33.01	789.5104	17.70	789.5104	220.50
10	940.7488	5.50	940.7488	2.64	940.7488	1.53
11	1005.6553	146.51	1005.6553	48.43	1005.6553	381.70
12	983.3620	268.81	983.3620	150.89	983.3620	2102.62
13	1302.1674	64.23	1302.1674	35.17	1302.1674	4.15
14	-	-	-	-	1301.8292	2861.45
15	-	-	-	-	1265.1313	21331.84
16	-	-	-	-	1783.3778	73.07
17	-	-	-	-	1792.8146	1399.15
18	-	-	-	-	1763.0296	10344.45
19	-	-	-	-	2765.6342	4123.75
20	-	-	-	-	2760.3163	33121.92
21	-	-	-	-	3787.2354	35400.62

"-": No solution found due to CPLEX shortage memory.

Table 3. Numerical results for random generated instances where about a half of the uncertain edges have negative weights in each scenario.

Inst.	(P_2)	cpu(s)	(RP_2)	cpu(s)	(RP_3)	cpu(s)
1	163.0000	0.34	163.0000	0.32	163.0000	0.36
2	178.6012	0.36	178.6012	0.36	178.6012	1.50
3	176.4886	0.41	176.4886	0.41	176.4886	2.94
4	460.0961	0.42	460.0961	0.39	460.0961	0.53
5	395.5182	0.81	395.5182	0.75	395.5182	3.98
6	460.7503	3.38	460.7503	1.36	460.7503	12.91
7	585.3537	0.73	585.3537	0.69	585.3537	0.78
8	589.5790	5.22	589.5790	2.66	589.5790	6.81
9	612.3388	14.06	612.3388	5.50	612.3388	27.20
10	808.9535	2.97	808.9535	2.36	808.9535	0.92
11	809.6658	44.58	809.6658	19.00	809.6658	28.12
12	855.6202	107.17	855.6202	37.00	855.6202	57.48
13	1182.2429	44.23	1182.2429	27.05	1182.2429	2.83
14	-	-	-	-	1142.6433	161.94
15	-	-	-	-	1108.7019	1025.52
16	-	-	-	-	1616.3587	15.67
17	-	-	-	-	1586.4263	37.13
18	-	-	-	-	1579.3486	759.53
19	-	-	-	-	2537.2853	254.66
20	-	-	-	-	2649.1216	3425.19
21	-	-	-	-	3540.0685	4682.42

"-": No solution found due to CPLEX shortage memory.

Table 4. Numerical results for arbitrarily generated instances. All relaxed solutions are fractional even for the model (P_2) containing all possible sub-tour elimination constraints.

Inst.	(P_2)	cpu(s)	(RP_2)	cpu(s)	(P_3)	cpu(s)	(RP_3)	cpu(s)
22	59	0.450	61.500	0.328	59	0.015	61.500	0.002
23	41	1.484	43.500	0.422	41	0.011	43.500	0.010
24	94	0.515	97.333	0.391	94	0.028	97.333	0.006
25	118	0.406	120.500	0.391	118	0.032	120.500	0.006
26	202	14.375	204.500	11.563	202	0.110	204.500	0.044

Results from Tables 4 are very important because they prove that the model e.g. (P_2) is not a TDI system. Moreover, they allow to obtain an idea of the solution structure that makes the SMWF problem difficult and give some insights about the complexity of this problem (this is an open question). Note that despite the reduced dimensions of these instance, and the fact we do not have many instances to perform exhaustive numerical experiments, the new compact model obtains all their optimal solutions in considerable fewer cpu time than the model (P_2).

5 Conclusion

In this paper we propose a polynomial size formulation for the stochastic maximum weight forest problem. This formulation is based on the one for the spanning tree polytope of complete undirected graphs [4]. We extend the model in [4] to deal with forests in non complete graphs. For this, we propose a new theorem characterizing forests in undirected graphs that is important to prove the correctness of the new model.

The numerical experiments evidence that this problem can be NP-hard. Thus, further research will be dedicated to answer this open question concerning the complexity of the SMWF problem.

References

1. Conforti, M. and Cornuéjols, G. and Zambelli, G. Extended Formulations in Combinatorial Optimization. 4OR - A Quarterly Journal of Operations Research 8: 1–48 (2010).
2. Edmonds, Jack : Matroids and the greedy algorithm. Mathematical Programming, 1: 127–136 (1971).
3. Letournel, Marc and Lisser, Abdel and Schultz, Rüdiger : Is the polytope associated with a two stage stochastic level problem TDI? Laboratoire de Recherche en Informatique, Universit de Paris Sud, Rapport de Recherche N. 1546 (2010).
4. Martin, R. K.: Using separation algorithms to generate mixed integer model reformulations. Operations Research Letters 10(3): 119–128 (1991).
5. Yannakakis Mihalis: Expressing Combinatorial Optimization Problems by Linear Programs. Journal of Computer and System Sciences, 43(3): 441-466 (1991).

6. Pulleyblank, W.R. Polyhedral combinatorics, In Nemhauser et al., editor, Optimization, volume 1 of Handbooks in Operations Research and Management Science, North-Holland, chapter 5, 371–446 (1989).