CaROMaD, Volume 02 (2017) : 22 − 34
http://www.laromad.usthb.dz

Cahiers de Recherche Opérationnelle
et Mathématiques de la Décision

Star forest : Linear time algorithm and extended formulation

Lamia Aoudia(1) & **Méziane Aïder**(2) & **Viet Hung Nguyen**(3)

(1,2) LaROMaD, Fac. Maths, USTHB, Pb 32, 16111 Bab Ezzouar, Algeria.

(3) LIP6; UPMC: Sorbonne Universités, France.

laoudia@usthb.dz ; maider@usthb.dz & hung.nguyen@lip6.fr

Abstract: In this paper, we introduce the maximum weight spanning star forest ($MWSFP$) which consist, given an undirected graph with weights on the edges in finding a star forest spanning the nodes of the graph with maximum weight.
We deal with the polyhedral investigation of the $MWSFP$. We present a polyhedral investigation of the $MWSFP$. More precisely, we study the facial structure of the star forest polytope, denoted by $SFP(G)$, which is the convex hull of the incidence vectors of the star forests of G. First, we give a complete characterization of the $SFP(G)$ when G is a cycle. Then we provide a linear time algorithm for the problem on cycles, which lead to a compact extended formulation for the $MWSFP$ when the graph is cycle.

Key Words: Star forest, convex hull

MSC(2010): Donner un ou deux codes MSC2010

Résumé : Dans cet article, nous présentons le problème de la forêt d'étoiles de poids maximum (MWSFP) qui consiste à, étant donné un graphe non orienté avec des poids sur les arêtes, trouver une forêt d'étoiles couvrant tousles sommets du graphe, de poids maximum. Nous présentons une étude polyédrale du MWSFP. Plus précisément, nous étudions la structure faciale du polytope de la forêt d'étoiles, notée $SFP(G)$, qui est l'enveloppe convexe des vecteurs d'incidence des forêts d'étoiles de G. Nous donnons en premier lieu une caractérisation complète de $SFP(G)$ lorsque G est un cycle. Ensuite, nous fournissons un algorithme de temps linéaire pour le problème sur les cycles, ce qui conduit à une formulation étendue compacte pour MWSFP lorsque le graphe est un cycle.

Mots clés : Forêt d'étoiles, enveloppe convexe, approche polyédrale

1 Introduction

A star is a tree where some vertex is adjacent to the other vertices of the tree. *A star forest* is a collection of vertex-disjoint stars in G.

Given a graph $G = (V, E)$ and weight on the edges, the maximum weight spanning star forest problem ($MWSFP$ for short) consist in finding a star forest spanning the vertices of given maximum weight. As isolated vertices can be taken, any star forest can be extended to a spanning star forest without additional weight. Hence, without loss of generality, we suppose that the graph without isolated vertices.

The $MWSFP$ is NP-hard [14]. In fact, it is not difficult to make a one-to-one correspondence between the spanning star forests and the dominating sets in a graph. *A dominating set* of a graph $G = (V, E)$ is a node subset $D \subseteq V$ such that every node in $V \setminus D$ is adjacent to at least one node in D. Given a star we define *the center* of the star as the vertex of degree strictly greater than one or any of the vertices if the star is an edge. Observe that in a spanning star forest, each vertex is either a center or adjacent to a center. Therefore, maximum size of spanning star forest is the number of vertices minus the cardinality of the minimum dominating set. Consequently, computing the maximum size spanning star forest of a graph is equivalent to finding a minimum size dominating set. Hence the $MWSFP$ is NP-hard.

The maximum spanning star forest problem has applications in computational biology as shown in [14]. The problem of aligning multiple genomic sequences, which is a basic bio-informatics task in comparative genomics, reduces to the $MWSFP$. Further applications appear in the comparison of phylogenetic trees [5], and in the diversity problem in the automobile industry [1].

Nguyen [15], gives an integer programming formulation for the $MWSFP$. He also, investigates the facial structure of the problem. Moreover, a complete description of the polytope associated with the $MWSFP$ when the graph is a tree is given.

Let us introduce the notations that will be used in the paper.

Let $G = (V, E)$ be a simple graph, where $|V| = n$ and $|E| = m$. *The line graph* of G is the graph $L(G) = (V', E')$ whose vertices correspond to the edges of G and two vertices of $L(G)$ are adjacent if and only if their corresponding edges are incident in G. For $x \in \mathbb{Q}^m$, given any $F \subseteq E$, let $x(F)$ denote $\sum_{e\in F} x_e$. For $x \in \mathbb{Q}^n$, given a set $S \subseteq V$, let $x(S)$ denote $\sum_{v\in S} x_v$. Given a set of vertices S, we denote by $E(S)$ the set of edges with both ends belonging to S. Let $v \in V$, *the neighborhood* of v denoted by $N(v)$ is the vertex set consisting of v and the vertices which are adjacent to v. We call a 3-*path* a simple path having 3 edges in G and a cycle of length 3. Let $\mathcal{P}4$ (respectively $\mathcal{C}_3$) denote the collection of the 3-paths (resp. 3-cycles) in G.

With any star forest F, we associate an incident vector $x^F \in \mathbb{Q}^m$ defined by

$$x_e^F = \begin{cases} 1 & \text{if } e \in F, \\ 0 & \text{otherwise.} \end{cases}$$

Let $SFP(G)$ be the convex hull of the incidence vectors of all the spanning star forests of G.

Remark 1 *[15] A star forest is a graph without* 3*-paths and* 3*-cycles.*

Based on this remark1, Nguyen in [15] proposes the following integer formulation for $MWSFP$:

$$\max \sum_{e\in E} c_e x_e$$

$$0 \leq x_e \leq 1 \text{ for all } e \in E \tag{1}$$
$$x(P) \leq 2 \text{ for all } P \in \mathcal{P}_4 \tag{2}$$
$$x(E(C)) \leq 2 \text{ for all } C \in \mathcal{C}_3 \tag{3}$$
$$x_e \in \{0,1\} \text{ for all } e \in E \tag{4}$$

Inequalities (1) are called *trivial inequalities*. The inequalities (2) called 3-*path inequalities* express the fact that any path in a star forest can not have more than two edges. The inequalities (3) called 3-*cycles inequalities* discard the 3-cycles in the forest.

In this paper we give a linear time algorithm for the problem $MWSFP$ on cycles which is the same. Moreover, we give a complete characterization and a compact extended formulation for $SFP(G)$ when G is a cycle. As a consequence of our results, a complete description of the edge dominating set polytope on cycles is derived.

The paper is organized as follows. In Section 2 we introduce new facet defining inequalities for $SFP(G)$. We prove that these inequalities, called *the matching-cycle inequalities*, can be separated in polynomial time. In Section 3 we give a complete description of $SFP(G)$ when G is a cycle. In Section 4 we provide a linear time algorithm for $MWSFP$ when G is a cycle. This algorithm can be used to give a minimum edge dominating set in graph. Section 5 is devoted to an extended formulation based on the algorithm given in the previous section.

2 Valid inequalities

Here, we introduce two classes of valid inequalities for the $SFP(G)$ in relation with the dominating set polytope.

Proposition 2 *Let C be a chordless cycle, then the inequality*

$$x(E(C)) \leq \lfloor \frac{2|C|}{3} \rfloor \tag{5}$$

is valid for the $SFP(G)$. Moreover it define a facet when , either$|C| = 3$ or $|C| \geq 4$ and $|C|$ not multiple of 3.

Preuve. Let denote by $P(e_i)$ a 3-path containing e_i as a middle edge. By (2) we have

$$x(P(e_i)) \leq 2 \text{ for all } e \in E(C)$$

By summing over $e_i \in E(C)$ we have

$$\sum_{e_i \in E(C)} x(P(e_i)) \leq 2|C|$$
$$3 \sum_{e_i \in E(C)} x(e_i) \leq 2|C|$$

hence

$$x(E(C)) \leq \lfloor \frac{2|C|}{3} \rfloor$$

■

In what follows, we introduce a large class of facet-defining inequalities for $SFP(G)$ called *matching-cycle inequalities*. Let us introduce some notations that are necessary to formulate this class of inequalities.

Let $C = \{1, 2, \ldots, n'\}$ a cycle having n' vertices $1, 2, \ldots n'$ numbered clockwise and n' edges $e_i = (i, i+1)$ for $i = 1, \ldots, n'-1$, and $e_{n'} = (n', 1)$. We denote by $C(u, v)$ the set of the edges between u and v in the clockwise sens. Let $W = \{w_1, w_2, \ldots, w_p\} \subseteq \mathcal{W}$ be a subset of p vertices with p odd and $p \geq 3$, where $|C(w_j + 1, w_{j+1})| = 3k_j$ with $k_j \geq 1$ for $j = 1 \ldots p$. Let $\mathcal{W}$ be the collection of all the vertex subsets W defined above. Let $m_j = (w_j, w_j + 1)$ for $j = 1, \ldots, p$ and let $M = \{m_1, m_2, \ldots, m_p\}$. Observe that M be a matching.

Then, we define the matching-cycle inequalities as follows:

$$x(E(C)) + x(M) \leq 2\sum_{i=1}^{p} k_i + \lfloor \frac{3p}{2} \rfloor, \forall C \in \mathcal{C}, \forall W \in \mathcal{W} \tag{6}$$

Where $\mathcal{C}$ be the collection of all cycles C in G. In the following, we state that the inequalities (5) and (6) define facets for the spanning star forest polytope $SFP(G)$.

Theorem 3 *The matching cycle inequality* (6) *defines a facet for* $SFP(G)$

Preuve. First we give the proof of validity of (6) for the $SFP(G)$. Because the validity of 3-path inequality for the $SFP(G)$, and by denoting $P(m_i)$ the 3-path with m_i as a middle edge, we have the following:

$$\begin{aligned} x(m_i) &\leq 1 \\ x(P(m_i)) &\leq 2 \\ x(P(m_i + 3t + 1)) &\leq 2 \quad \forall t = 0, \ldots, k_i - 1. \\ x(P(m_i + 3t + 3)) &\leq 2 \quad \forall t = 0, \ldots, k_i - 1. \end{aligned}$$

for $i = 1, \ldots, p$.
by summing these constraints, we get:

$$4\sum_{e \in E_W} x(e) + 2\sum_{e \in E(C) \setminus M} x(e) \leq 2\sum_{i=1}^{p} 2k_i + 3p. \tag{7}$$

As p is odd, by dividing the constraint (7) by two and rounding to the integer lower value we obtain the validity of (6) (by Chvatal-Gomory cut).

■

From now on, $G = (V, E)$ is considered as a cycle. Let $C = \{1, 2, \ldots, n\}$ a cycle having n vertices $1, 2, \ldots n$ numbered clockwise and n edges $e_i = (i, i+1)$ for $i = 1, \ldots, n-1$, and $e_n = (n, 1)$.

3 Dominating set and star forest

Lemma 4 *Any extreme point of* $SFP(G)$ *which verifies* (6) *at equality is a maximal star forest.*

Preuve. Suppose there is a star forest which is not maximal and which verify (6) at equality. Let $ax \leq \alpha$ be the inequality representing(6). This means $ax^F = \alpha$. Or F is not maximum, there exist $e \in E$ such that $F \cup \{e'\} = F'$ form a star forest in C. Thus $x(e') \neq 0$, $x(F') = x(F) + x(e')$. Or $ax^{F'} \leq \alpha$, then $ax^F + x(e') \leq \alpha$. By supposition $ax^F = \alpha$, we get $x(e') \leq 0$ which is a contradiction. ■

An edge dominating set EDS of a graph is subset of edges such that every other edge is adjacent to an edge in EDS. The following lemma establishes the link between edge dominating sets and maximal star forests in a cycle.

Lemma 5 *The complimentary of a maximal star forest F in C is an edge dominating set and vice versa.*

Preuve. Let $F \subseteq E$ be a spanning star forest in G, and let $\bar{F}$ be its complementary edge-set. Suppose that $\bar{F}$ is not an edge dominating set. Suppose that there is an edge $e' \in F$ which is not adjacent to any edge in $\bar{F}$. Hence, e' is adjacent to two edges in F, thus form a path of length 3 in F. Contradiction by remark 1. ■

Given a cycle C, we denote by $L(C)$ its line graph. Note that $L(C)$ is also a cycle. We have the following:

Lemma 6 *Any edge dominating set in C is a dominating set in $L(C)$ and vice versa.*

Preuve. It is known that a cycle C is isomorphic to its line graph $L(C)$. An edge in C is substituted by a vertex in $L(C)$ and a vertex in C is replaced by an edge in $L(C)$. If an edge e dominates a set of edges M in $E(C)$, there must exist a vertex $l(e)$ which substitute e in $L(C)$ which dominates the vertices in $L(C)$ which substitutes the edge-set $M \subseteq E(C)$. ■

Proposition 7 *Given a cycle $C = (V, E)$, let $L(C) = (L(V), L(E))$ be its line graph. Let x^e (y^v, respectively) be the edge variable (vertex variable), The following statements are equivalent:*

(i) $\alpha^t y^v \geq \beta$ *defines a facet for* $DP(L(C))$.

(ii) $\alpha^t x^e \geq \beta$ *defines a facet for* $EDP(C)$,

(iii) $\alpha^t x^e \leq \sum_{e \in E(C)} \alpha(e) - \beta$ *defines a facet for* $SFP(C)$.

Preuve.

(i)⇒(ii) Suppose (i) occures, this means, there exist $|C| - 1$ dominating set affinely independent in $L(C)$ verifying $\alpha^t y^v \geq \beta$ at equality. By lemma6, for each dominating set in $L(C)$ corresponds an edge dominating set in C, then there exists $|C| - 1$ edge dominating sets in $L(C)$ verifying $\alpha^t x^e \geq \beta$ at equality. We have to prove that they are affinely independent. This is the case because there is one to one correspondence between edge dominating set and dominating set. which means, $\alpha^t x^e \geq \beta$ defines a facet for.

(ii)⇒ (i) The same scheme is applied for the other sense (ii)⇒ (iii) For each edge dominating set $L(D)$ in $L(C)$ corresponds a complimentary which define a star forest $L(SF)$ in $L(C)$.

Let x^{ED1} an incident vector associated to an edge dominating set $ED1$, its complimentary is $\bar{x}^{ED1} = \mathbf{1}_{E(C)} - x^{ED1}$, where $\mathbf{1}(C)$ is a vector in which all components are 1. $a^t x^{ED1} = \beta$ is equivalent to $a^t(1_{E(C)} - \bar{x}^{ED1}) = \beta$ which is equivalent to $a^t \bar{x}^{ED1} = \sum_{e \in E(C)} a(e) - \beta$. (let $SF1$ the complimentary of $ED1$, then $\bar{x}^{ED1} = x^{SF1}$). Then we have $a^t x^{SF1} = \sum_{e \in E(C)} a(e) - \beta$.

Suppose there exist $|C| - 1$ affinely independent edge dominating set verifying $\alpha^t x^e \geq \beta$ at equality then there is $|C| - 1$ affinely independent complementary of edge dominating set verifying $\alpha^t x^e \leq \sum_{e \in E(C)} \alpha(e) - \beta$ at equality.

and vice versa.

(iii)⇒ (ii) We apply the same scheme

■

Bouchakour *et al.* in [7] give a complete description of dominating set polytope.

Theorem 8 *[7] When the graph G is a cycle C, the complete description of dominating set polytope is given by the following system:*

$$
\begin{aligned}
0 \leq x(v) &\leq 1 \text{ for all } v \in V && (8)\\
x(N(u)) &\geq 1 \text{ for all } u \in V && (9)\\
x(C) &\geq \lceil \frac{|C|}{3} \rceil && (10)\\
x(C) + x(W) &\geq \sum_{i=0}^{p-1} k_i + \lceil \frac{p}{2} \rceil \text{ for all } W \in \mathcal{W} && (11)
\end{aligned}
$$

Remark 9 *When a graph G is cycle, the edge dominating set polytope $EDSP(G)$ is characterized completely by the following system:*

$$
\begin{aligned}
0 \leq x(e) &\leq 1 \text{ for all } e \in E && (12)\\
x(P) &\geq 1 \text{ for all } P \in \mathcal{P}_4 && (13)\\
x(E(C)) &\geq \lceil \frac{|E(C)|}{3} \rceil && (14)\\
x(E(C)) + x(M) &\geq \sum_{i=0}^{p-1} k_i + \lceil \frac{p}{2} \rceil \text{ for all } W \in \mathcal{W} && (15)
\end{aligned}
$$

Preuve. For each edge $e \in E(C)$, there corresponds a vertex $v \in L(C)$ such that: $x(e)_C = y(v)_{L(C)}$. By replacing $x(e)$ for all $e \in E(C)$ by $y(v)$ where v is the corresponding vertex to e in the line graph $L(G)$. We obtain the system defined by the inequalities (8)-(11) which define complete characterization of dominating set when G is a cycle (from theorem 8). ■

Based on these results, we prove that,

Theorem 10 *When G is cycle, $SFP(G)$ is completely described by the trivial inequalities* (1)*, the 3-path inequalities* (2)*, the cycle inequalities* (5) *and the matching-cycle inequalities* (6).

Preuve. The proof can be derived from Lemmas 4, 5, 6 and Theorem 8. By lemma 5 the complement of an edge dominating set is spanning star forest. In cycle C, let $x(e) = 1 - z(e)$ for all $e \in E(C)$. By replacing $x(e)$ by $1 - z(e)$ in the system defined by (1), (2), (5) and (6); we obtain the system which characterize the edge dominating set,this is true when the graph is cycle. ■

4 A linear time algorithm for MWSFP when G is a cycle

To the best of our knowledge, there exists a polynomial time algorithm given in [14] to solve $MWSP$ only for the case when G is a tree. In this section, we will give a linear time algorithm solving $MWSP$ when G is a cycle.

Let us suppose that the vertices of C is numbered from 1 to n and the edges e_i of weight c_i is the edge between i and $i+1$ for $i = 1, \ldots, n-1$. In particular, the edge e_n of weight c_n is the edge between n and 1. We will transform the $MWSFP$ in G into a problem of finding a longest path in some acyclic graph G'. First, we built a graph $G' = (X', A')$ from G. For an edge $i \in G$ we create in G' four vertices called i_{-2}, i_{-1}, i_1, i_2, which are all clones of i. The arcs are created as follows: for every vertex $1 \leq i \leq n-1$ in G, there are :

- an arc $(i_{-2}, (i+1)_1)$ of cost c_i and of blue color.
- an arc $(i_{-1}, (i+1)_1)$ of cost c_i and of blue color, an arc $(i_{-1}, (i+1)_{-2})$ of cost 0 and of red color
- an arc $(i_1, (i+1)_2)$ of cost c_i of blue color, an arc $(i_1, (i+1)_{-1}$ of cost 0 and of red color, Finally, an arc $(i_2, (i+1)_{-1})$ of cost 0 and of red color.

We prove the following.

Proposition 11 *There is at most two successive arcs with the same color in any path from 1_x to n_y for some $x, y \in \{-2, -1, 1, 2\}$ in G'.*

Preuve. Given a path P from 1_x to x_y for some $x, y \in \{-2, -1, 1, 2\}$. Suppose that there is 3 consecutive blue arcs in P : $(j, k), (k, l)$ and (l, m) as for both l and m, there are two blue arcs preceding them, there exists an $1 \leq i \leq n$ such that $l = i_2$ and $m = (i+1)_2$. But by construction of G', there cannot exist the arc (l, m). Contradiction.

The same proof can be applied to prove that there cannot be 3 consecutive red arcs in P. ■

Lemma 12 *The blue arcs in any path P from 1_x to n_y in G' correspond to the edges of spanning star forest in G.*

Preuve. An arc $(i_{\bar{x}}, (i+1)_{\bar{y}}) \in P$ with $\bar{x}, \bar{y} \in \{-2, -1, 1, 2\}$ corresponds to the edge $i(i+1)$ in G. According to proposition11, the set of the blue arcs in P does not contain 3-paths. It does not contain either any 3-cycle as G is a cycle. according to remark 1. Hense, the set of the blue arcs in P corresponds to the star forest in G. ■

Lemma 13 *All maximal star forest in G correspond to one of the following:*

1. *The blue arcs of a path from* 1_{-2} *to* n_{-1} *in* G'.
2. *The blue arcs of a path from* 1_{-1} *to* n_1 *in* G'.
3. *The blue arcs of a path from* 1_{-1} *to* n_2 *in* G'.
4. *The blue arcs of a path from* 1_1 *to* n_{-1} *in* G' *+ the arc* $(n, 1)$.
5. *The blue arcs of a path from* 1_1 *to* n_{-2} *in* G' *+ the arc* $(n, 1)$.
6. *The blue arcs of a path from* 1_1 *to* n_2 *in* G' *+ the arc* $(n, 1)$.

Preuve. Let F^* be any maximal spanning star forest in G. We will construct a path P^* from 1_x to n_x in G' have the same cost as F^*.

Let us assign the blue color to each edge in F^* and the red color to each edge in $C \setminus F^*$.

We determine the vertices in P^* as follows :

For each vertex $i \in C$, if there are x red edges prior clockwise to i, then the vertex i_{-x} belongs to P^*. Otherwise, if there are x blue edges prior clockwise to i, then the vertex i_x belongs to P^*

Let us consider the path P^* (from 1_x to n_y) determined by its vertices in G'. By the construction of G', we can see that the non zero costs of P^* are exactly the blue arcs.

Hence, the cost of P^* is equal to the cost of F^*.

■ Thus, the $MWSFP$ in C can be reduced to finding 6 longest paths in G' which is acyclic. Instead of solving 6 longest paths problems, we can unify them into a unique problem of longest $s - t$ path as follows.

We construct a digraph $G" = (X", A")$ by cloning the resulting digraph $G' = (X', A')$ on 6 copies. The vertex in $G"$ is denoted by i_x^j where $1 \leq i \leq n$, $1 \leq j \leq 6$ and $x \in \{-2, -1, 1, 2\}$. The set of arc $A" = \cup_{j \in \{1,2,3,4,5,6\}} A_j$.

Finally construct a network $N(X, A)$ by adding a source s and a sink t to $G" = (V", A")$. for all $1 \leq j \leq 6$ and for some x and y add an arc between s and 1_x^j and an arc between n_y^j and t as follows:

- a red arc with a zero cost $(s, 1_{-2}^1)$, a red arc with a zero cost (n_{-1}^1, t),
- a red with a zero cost $(s, 1_{-1}^2)$, a red arc with a zero cost (n_1^2, t),
- a red arc with a zero cost $(s, 1_{-1}^3)$, a red arc with a zero cost (n_2^3, t),
- a blue arc $(s, 1_1^4)$ with a cost c_i, a blue arc $(n_{-1}^4, t)$with a cost c_i,
- a blue arc $(s, 1_1^5)$ with a cost c_i, a blue arc (n_{-2}^5, t) with a cost c_i,
- a blue arc $(s, 1_2^6)$ with a cost c_i, a blue arc (n_1^6, t) with a cost c_i,

In fact the construction of network $N(X, A)$ simulates the construction of a star forest in G from the vertex 1 and in the clockwise to the vertex n.

Red edge means the edge is not in the star forest and blue edge means that the edge is in the star forest). Then the vertex i_x^j express the fact that there are $|x|$ red edges prior to i^j if x is negative and there are x blue edge (edges) prior to i^j if x is positive.

Find the maximum weight spanning star forest on C turn out to find the longest path between s and t. Because $N = (X, A)$ is acyclic, using Bellman-Ford algorithm This is proved to be a linear.

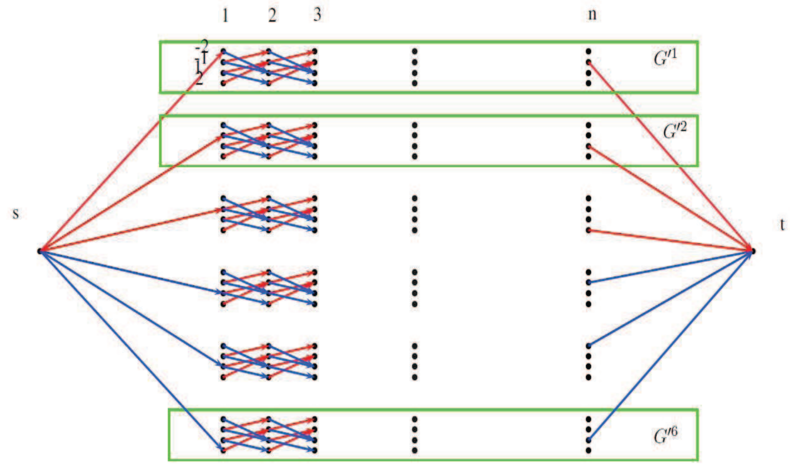

Figure 1: Network $N = (X, A)$

Corollary 14 *Given a network $N = (X, A)$ and a length function $l : A \to \mathbb{Q}$, a longest $s - t$ path can be found in time $O(|X|)$.*

Preuve. By using Bellman ford algorithm the longest path is returned after $n - 1$ iterations. In each iteration, at most 24 test are performed. ■

5 An extended formulation

In what follows, we derive from the previous scheme of section (4), a compact extended formulation.

Let $N = (X, A)$ be a a network and let a vertex set $S \subseteq X$. We denote by $\delta^+(S)$ the set of arcs of D with a tail in S and head in $X \setminus S$ and by $\delta^-(S)$ the set of arcs of D with a head in S and tail in $X \setminus S$. We write $\delta^+(v)$ $(\delta^-(v))$ instead of $\delta^+(\{v\})$ $(\delta^-(\{v\}))$.

It turns then to find the longest path between s and t. This is the solution of the following system.

$$\max \sum c(a)\phi(a) \tag{16}$$
$$\phi(\delta^+(i_x^j)) - \phi(\delta^-(i_x^j)) = 0 \, \forall i_x^j \in X \tag{17}$$
$$\phi(\delta^+(s)) - \phi(\delta^-(t)) = 1 \tag{18}$$
$$\phi(a) \geq 0 \, \forall a \in A \tag{19}$$

The graph $N = (X, A)$ as it is constructed in section 4 is acyclic, and because system above is totally unimodular, the extreme solutions of the system given above are integer. We denote by $F \in \mathbb{Q}^{|V \setminus \{s,t\}|+|A|)}$ the coefficient matrix of conservation flow equations, in other word, $F\phi = 0$ is the matrix from (17).

$$x(i(i+1)) = \sum_{j=1}^{6} \phi(i^j_{-2}(i+1)^j_1) + \phi(i^j_{-1}(i+1)^j_1) + \phi(i^j_1(i+1)^j_2). \quad (20)$$

We call the system formed by the inequalities (17)-(20) a flow based formulation of P.

Theorem 15 *ϕ is the incident vector of the longest path from s to t in $N = (X, A)$.*

Preuve. Dantzig[8] showed that the solution of the system defined by (17)-(19) is the incident vector of an $s-t$ path. Because, the objective function is to maximize the costs over the arc set A, and ϕ is the solution of the linear program defined by (17)-(19). Then, ϕ defines the longest $s-t$ path in $N = (X, A)$. ■

Lemma 16 *Let ϕ be the incident vector of path L. L contains exactly one arc through $i^j_x(i+1)^j_y$, for fixed i and for all $j \in \{1, 2, 3, 4, 5, 6\}$ and $(x, y) \in \{(-2, 1), (-1, 1), (1, 2), (1, -1), (2, -1), (-1, -2)\}$.*

Preuve. By definition of the network $N = (V, E)$, there is no arc between a vertex i^j_x and a vertex $(i+1)^{j'}_y$ for $j \neq j'$.

If path L gets through i^j_x for fixed j and fixed x it must also get through $(i+1)^j_y$, if $(x, y) \in \{(-2, 1), (-1, 1), (1, 2), (1, -1), (2, -1), (-1, -2)\}$. Because L is a path and by definition of $N = (V, E)$ the arcs are oriented from i to $(i + 1)$, L gets through exactly one arc $(i^j_x, (i+1)j_y)$ for all $j \in \{1, 2, 3, 4, 5, 6\}$ and $(x, y) \in \{(-2, 1), (-1, 1), (1, 2), (1, -1), (2, -1), (-1, -2)\}$ ■

Let Y be the set of couple (x, y). $Y = \{(-2, 1), (-1, 1), (1, 2), (1, -1), (2, -1), (-1, -2)\}$. We can partition the set Y into two sets $Y^- \cup Y^+$ where $Y^+ = \{(-2, 1), (-1, 1), (1, 2)\}$ and $Y^- = \{(1, -1), (2, -1), (-1, -2)\}$.

Remark 17 $\sum_{j=1}^{6} \sum_{(x,y)\in Y} \phi((i^i_x, (i+1)^j_y)) = 1.$

Suppose that, an arc $(i^j_x, (i+1)^j_y)$ with $(x, y) \in Y^+$ take a color blue, and an arc $(i^j_x, (i+1)^j_y)$ with $(x, y) \in Y^-$ take a color red.

Because a path L gets through exactly one arc $(i^j_x, (i+1)^j_y)$, blue $((x, y) \in Y^+)$or red $((x, y) \in Y^-)$arc, we can translate $x(i(i + 1))$ as follows:

$$x(i(i+1)) = \begin{cases} 1 & \text{if } L \text{ contains a blue arc} \\ 0 & \text{otherwise.} \end{cases}$$

Lemma 18 *Let (x, ϕ) be a basis solution defined by the system* (17)-(20)

(i) The basis solution (x, ϕ) is integer.

(ii) x is an incident vector of a star forest in C.

Preuve.

(i) The basis solution (x, ϕ) is integer.

Because ϕ is a solution of the system defined by (17)-(19), it must be integer, (due to unimodularity of the matrix associated to the system (17)-(19)).

Note that x is an integer linear combination of vectors ϕ, then it is integer. Hence the vector x, ϕ is integer.

(ii) x is an incident vector of a star forest in C.

By definition of x, the non zero component of x are those associated to passes $i(i+1)$ issued from a blue arc from path L. By lemma 12, there is at most two successive arcs with the same color in L, then there is at most two passes with color blue. This define a star forest in C.

■

Lemma 19 *Given an incident vector x of a star forest, there must exist a vector ϕ such that (x, ϕ) is a solution of the system* (17)-(20).

Preuve.

Given a cycle C and a star forest F in C, let consider x^F its incident vector. We color all edges of F by blue and the other edges not in F by red.

We can substitute each vertex $i \in V(C)$ by a vertex i_x where a negative x means there are $|x|$ red edges preceding i and a positive x means there are x blue edges preceding i. Then $x \in \{-2, -1, 1, 2\}$. And add the arcs with a weight c_i on blue arcs $(i_x, (i+1)_y)$ and "0" on the red arcs.

To represent all the possibilities we can add an index j to each vertex i_x to have i_x^j. we have 6 possibilities , $j \in \{1, 2, 3, 4, 5, 6\}$. Then we construct a network $N = (X, A)$; let ϕ an incident vector of an st-path corresponding to a star forest F.

Such $s-t$ path constructed from F is a longest path between s ant t then it must be a solution of a system (17)-(19).

Because each star forest corresponds to one $s-t$-path in $N = (X, A)$ then each edge $i(i+1)$ which is red or blue correspond to one arc $(i_x^j(i+1)_y^j)$ for $j \in \{1, 2, 3, 4, 5, 6\}$ and $(x, y) \in Y$. Blue edges are in F and red ones are not in F. Then the sum of $\phi((i_x^j(i+1)_y^j))$ over $j \in \{1, 2, 3, 4, 5, 6\}$ and $(x, y) \in Y^+$ give a the incidence vector of star forest F.

Then (x, ϕ) is the solution of a system defined by (17)-(20). ■

Theorem 20 *The system* (17)-(20) *defines an extended formulation for the star forest polytope $SFP(G)$ when the graph G is cycle.*

Preuve. It follows from lemma 18 and 19.

■

Let $\bar{x}(i(i+1)) = 1 - x(i(i+1))$,

$$\bar{x}(i(i+1)) = \sum_{j=1}^{6} \phi(i_2^j(i+1)_{-1}^j) + \phi(i_1^j(i+1)_{-1}^j) + \phi(i_{-1}^j(i+1)_{-2}^j) \tag{21}$$

Let $\bar{P} \subseteq \mathbb{R}^n$ polytope with $|E|$ vertices.

$\bar{P} =$ Proj$_{\bar{x}}\{(\bar{x}, \phi) \in \mathbb{R}^{|E|} \times \mathbb{R}^{|A|}; (\bar{x}, \phi)$ is solution of the system (17)-(21) $\}$

Theorem 21 *The system* (17)-(21) *defines an extended formulation for edge dominating polytope.*

Preuve. By lemma 5 the complementary of a star forest is an edge dominating set. The projection of $(\bar{x}, \phi)$ on $\bar{x}$. By replacing x by $\bar{x}$ in lemma 18, $\bar{x}$ is an incident vector of an edge dominating set. The same for lemma 19 from an incident vector of an edge dominating set $\bar{x}$, there must exist ϕ an incident vector of an $s-t$ path such that $(\bar{x}, \phi)$ defines a solution of the system defined by (17)-(21). ■

A cycle is isomorphic to its line graph, the edge dominating set polytope is an edge formulation of dominating set, we have the following.

Corollary 22 *The system* (17)-(21) *define an extended formulation for vertex dominating set.*

6 Conclusion

The aim of this work is to give a polynomial time complete description of star forest polytope in graph. We have established a relationship between star forest and dominating set in cycle which leads to a new facet defining inequality (matching-cycle inequality) and a complete description of $SFP(G)$ when G is a cycle. Then we have given a linear time algorithm to solve $MWSFP$ which allows us to give a flow-based extended formulation to $MWSFP$ on a cycle.

References

[1] A. Agra, D. Cardoso, O. Cerdeira, and E. Rocha, *A spanning star forest model for the diversity problem in automobile industry*, ECCO XVIII, Minsk, Belarus, May 26-28, 2005.

[2] M. Baiou and F. Barahona. Integrality of some Facility location. In *SIAM Journalof Discrete Mathematics*, pages 665–679, 2009.

[3] M. Baiou and F. Barahona. Simple extended formulation of the dominating set polytope via facility location. Technichal report of IBM research. 2012.

[4] R. Bellman, *Dynamic Programming*, Princeton Univ. Press, Princeton, N. J. 1957.

[5] V. Berry, S. Guillemot, F. Nicolas, and C. Paul.*On the Approximation of computing evolutionary trees*, in Proceding of the 11'th Annual Computing and Combinatorics Conference, Lecture Note in Computer Sciences 3595, Springer -Verlag, New York, 2005, pp. 115-125.

[6] M. Bouchakour.*Composition de graphes et le polytope des absorbants.*, PHD thesis (1996).

[7] M. Bouchakour, T. M. Contenza, C. W. lee,A. R. Mahjoub. *On the dominating set polytope.* In *European Journal of Combinatorics* (2008).

[8] G. B Dantzig, D.R Fulkerson *On the max flow Min cut Theorem of Networks* . [Research Memorandum RM-1418] The RAND Corporation, Santa Monica, California [April 15,] 1955.

[9] S. Ferneyhough, R. Haas, D. Hanson and G. MacGillivaray. *Star forests, dominating sets and Ramsey-type problems*, Discrete Mathematics, Volume 245, Issues 1-3, 255-262.

[10] S. Fiorini and K. Paskovich. *Uncapacitated Flow-based Extended Formulations.* arXiv:1306.3119 [math.OC]

[11] M. M. Hou, P. Berman, L. X. Zhang, and W. Miller, *Controlling size when aligning multiple genomic sequences with duplications*, in Proc. of WABI, pp. 138-149, Zurich, 2006.

[12] A. R. Mahjoub, *Polytope des absorbants dans une classe de graphes à seuil*, Annals of Discrete Mathematics 17 (1983) pp. 443-452.

[13] J. Moravek, *A note upon minimal path problem*, Journal of Mathematical Analysis and applications 30(1970) 702-717.

[14] C. T. Nguyen, J. Shen, M. Hou, L. Sheng,W. Miller, L. Zhang: Approximating the spanning star forest problem and its applications to genomic sequence alignment. SIAM Journal of Computing. 38(3), 946-962 (2008).

[15] V. H. Nguyen. *The maximum weight spanning star forest problem: polyhedral and algorithmic results on trees and cycles* , technical report.