On the star forest polytope

Lamia Aoudia
University of Sciences and Technology Houari boumédiéne, Algeria
Email:lamia.aoudia@lip6.fr

Viet Hung Nguyen
Laboratoire : Lip6
Université de Piere et Marie Curie; Paris, France
Email: Hung.Nguyen@lip6.fr

Ali Ridha Mahjoub
Laboratoire: Lamsade
Université Paris Dauphine (Paris 9); Paris, France
mahjoub@lamsade.dauphine.fr

Méziane AIDER
University of Sciences and Technology Houari boumédiéne, Algeria
Email: maider@usthb.dz

***Abstract*—A star forest is a collection of vertex-disjoint trees of depth at most 1, and its size is the number of leaves in all its components. A spanning star forest of a given graph G is a spanning subgraph of G that is also star forest. The spanning star forest problem (SSFP for short) [10] is to find maximum size spanning star forest of given graph. Let define some graph $G = (V, E)$, to every star forest we associate a vector x^F. $x^F(e) = 1$ if $e \in F$ and $x^F(e) = 0$ otherwise. x^F is the incident vector of spanning star forest F. The convex hull of all spanning star forest incident vectors is called a spanning star forest polytope, denoted $SFP(G)$. In this paper we are mainly interested on complete characterization of $SFP(G)$.**

Introduction

Let $G = (V, E)$ be an undirected graph. *A star* is a graph in which some vertex is incident with every edge of the graph (i.e. a graph of diameter at most 2). In particular, a single vertex is also a star. *A star forest* is a graph in which each component is star. Given a connected graph G in which the edges may be weighted positively. *A spanning star forest* of G is subgraph of G which is star forest spanning the vertices of G. Note that the spanning star forest can contain isolated vertices. The size of spanning star forest F of G is defined to be the number of edges of F if G is unweighted and the total weight if the edges of F if G is weighted. We are interested in the problem of finding a maximum weight spanning star forest in G. As we can take isolated vertices, any maximum weight star forest can be extended without additional weight to a spanning star forest. Hence, without loss of generality, we shall focus on the problem of finding a Maximum Weight Star Forest $MWSFP$.

$MWSFP$ is NP-hard already for the case G unweighted. In fact, in this case since G is connected, for any maximum size spanning star forest F of G, we can see that F does not contain any isolated vertex. Hence, for each star of F we designate *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. *A dominating set* of a graph is a subset of the vertices such that every other vertex adjacent to a vertex in the dominating set. Observe that in a spanning star forest solution, each vertex is either a center or adjacent to a center. Hence a set of centers form a dominating set of the graph. The size of the maximum spanning star forest is thus, the number of vertices minus the size of the minimum dominating set. Consequently, computing the maximum spanning star forest of a graph is NP-hard because computing the minimum dominating set is NP-hard.

The spanning star forest problem has found applications in computational biology. Nguyen *et al.* in [10] use the spanning star forest problem to give an algorithm for the problem of aligning multiple genomic sequences, which is a basic bioinformatics task in comparative genomics. The spanning star forest problem and its directed version have found applications in the comparison of phylogenetic trees [10].

Little is known about the mathematical formulation even less polyhedral investigations of the problem $MWSFP$. The first work is given by Nguyen in [11], where an integer formulation for $MWSFP$ is given. He also, investigates the facet structure of the problem. In the same paper [11], a complete description of the problem $MWSFP$ 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 are the edges of G and two vertices of $L(G)$ are adjacent if and only if their corresponding edges are incident in G. For $x \in \mathbb{R}^m$, given any $F \subseteq E$, let $x(F)$ denote $\sum_{e\in F} x_e$. For $x \in \mathbb{R}^n$, 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. Let $\mathcal{P}4$ (respectively $\mathcal{C}_3$) denote the set of the simple paths (resp. cycles) of length 3 in G. *A tree* is a connected graph with no cycle. *A star* is a tree with a diameter at most 2. *A star forest* is a graph whose each connected component is a star, or equivalently, a star forest is a graph without paths and cycles of length 3. With any star forest F, we associate an incident vector $x^F \in \mathbb{R}^m$ defined by

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

Let $SFP(G)$ be the convex hull of the incidence vectors of all star forests of G. Nguyen in [11] gives the following integer formulation for $SFP(G)$:

$$0 \leq x_e \leq 1 \text{ for all } e \in E \tag{1}$$

$$x(P) \leq 2 \text{ for all } P \in \mathcal{P}_4 \tag{2}$$

$$x(E(C)) \leq 2 \text{ for all } C \in \mathcal{C}_3 \tag{3}$$

$$x \text{ is integer} \tag{4}$$

The inequalities (1) are called *trivial inequalities*. The inequalities (2) called 3-*path inequalities* dismiss the 3−paths. The inequalities (3) called 3-*cycles inequalities* discard the 3-cycles in the forest.

In this paper we rely on work of Nguyen and relationship of the spanning star forest polytope with dominating set polytope

978-1-4799-6773-5/14/$31.00 ©2014 IEEE

[5] to give a complete description of $SFP(G)$ when G is a cycle. In section 2 introduce a 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 prove the complete description of $SFP(G)$ when G is cycle. In section 4 we give linear time algorithm for $MWSFP$ when G is a cycle. The last section is devoted to an extended formulation based on the algorithm given in the precedent section.

I. New facet defining inequalities and separation

In [11] Nguyen introduce a facet defining inequality to the $SFP(G)$ which is called cycle inequality. It is given as follows:

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

where $|C| \geq 4$ and $|C|$ is not multiple of 3.

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 with n' vertices $1, 2, \ldots n'$ numbered clockwise and n' edges $e_i = (i, i+1)$ for $i = 1, \ldots, n'-1$, and $e_{n'} = (n', 1)$. Let $\mathcal{C}$ be the collection of all cycles C in G. 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 $E_W = \{m_1, m_2, \ldots, m_p\}$.

Then we define the matching-cycle inequalities as follows:

$$x(E(C)) + x(E_W) \leq 2\sum_{i=0}^{p-1} k_i + \lfloor \frac{3p}{2} \rfloor, \forall C \in \mathcal{C}, \forall W \in \mathcal{W} \tag{6}$$

Consider $SF(G) = \{F \subset E :$ maximal star forest of $G\}$, where *a maximal star forest* is a star forest that cannot be extended by adding one more edge.

Let $a^T x \leq \alpha$ be a valid inequality for $SFP(G)$, note by SF_a the star forests tight respectively to α, $SF_a = \{F \in SF(G) : a^T x^F = \alpha\}$.

Remark 1. *Let $F^* \in SF(G)$ and $m_i, m_{i+1} \in M$, if F^* contains m_i, m_{i+1} then*

- *If F^* contains m_i+1 and $m_{i+1}-1$, then the star forest F^* can include up to $2k_i$ edges in $C(w_i+1, w_{i+1})$*
- *If F^* contains at most one of the edges $\{m_i + 1, m_{i+1} - 1\}$ then F^* includes at most $2k_i - 1$ edges in $C(w_i + 1, w_{i+1})$*

Proof: Because we have $|C(w_i+1, w_{i+1})| = 3k_i$, we can partition $C(w_i+1, w_{i+1})$ into k_i 3-paths. Suppose that m_i and m_{i+1} are taken.

(i) If we take the edges m_i+1 and $m_{i+1}-1$ then for each 3-path, we can take the first and the third (last) edges. By this way we form a saturated star forest in $C(w_i, w_{i+1}+1)$, that makes the possibility to take $2k_i$ edges in $C(w_i+1, w_{i+1})$.

(ii) If we take at most one edge between m_i+1 and $m_{i+1}-1$, suppose we take w_i+1, then to form a star forest in $C(w_i+1, w_{i+1}-1)$ we can take the first and the third edges of each 3-path in $C(w_i, w_{i+1})$ without taking the last edge $(m_{i+1}-1)$ of the last 3-path. Then we can take $2(k_i-1)+1 = 2k_i-1$. We get the same thing if we suppose to take $m_{i+1}-1$ in the same way.

∎

From this remark we elaborate some conditions must be verified by a star forest which verify (6) at equality.

Lemma 1. *The star forest tight with the inequality* (6) *may be organized as follows*

(a) *Choose some edge $m_l \in E_w$ and form $\frac{p-1}{2}$ pairs (m_i, m_{i+1}) for all $i = (l+1), i = (l+3), \ldots, i = (l+p-2)$.*

(b) *For each pair (m_i, m_{i+1}) we have $F^* \cap C(w_i, w_{i+1})$ includes the two edges of E_w, m_i and m_{i+1} and $2k_i$ edges in $C(w_i+1, w_{i+1}-1)$.*

(c) *For the $\frac{p-1}{2} - 1$ pairs (m_j, m_{j+1}) , where $j \neq l$, $j \neq l-1$ and $j \neq i$ for $i = l+1, i = l+3, \ldots, i = l+p-2$, we have $F^* \cap C(w_j, w_{j+1})$ includes $2k_j - 1$ edges in $C(w_i+1, w_{i+1}-1)$*

(d) *For the edges in $C(w_{l-1}, w_l)$ and $C(w_l, w_{l+1})$*
   - (i) *If $m_l \in F^*$ then $F^* \cap C(w_{l-1}+1, w_l-1)$ contains $2k_{l-1}-1$ edges and $F^* \cap C(w_l+1, w_{l+1}-1)$ contains $2k_l - 1$.*
   - (ii) *If $m_l \neq F^*$ then $F^* \cap C(w_{l-1}+1, w_l-1)$ contains $2k_{l-1}$ edges and $F^* \cap C(w_l+1, w_{l+1}-1)$ contains $2k_l$ edges.*

Proof: Let F' be a star forest tight with the inequality (6) not organized as (a) ∎

Now we can prove the following result.

Theorem 2. *The matching-cycle inequalities* (6) *define a facet for the $SFP(G)$ for p odd.*

Proof: 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:

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

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).

Consider $ax \leq \alpha$, valid inequality to the polytope $SFP(G)$ that is verified at equality by the same star forest verifying (6) at equality.

Let $\mathbf{F_a}$ the set of star forests those are tight with $ax \leq \alpha$ and (6) in the same time. let $F_a \in \mathbf{F_a}$ by 1 we have

- **If** $m_l \in F_a$
  Then $F_a \cap C(w_l+1, w_{l+1})$ includes $2k_l - 1$ edges and $F_a \cap C(w_{l-1}+1, w_l)$ includes $2k_{l-1} - 1$ edges.
  By fixing all edges of the forest F_a not in $C(w_l + 1, w_{l+1})$ and construct other star forests by the shift of the two edges of F_a in $C(w_l+1, w_{l+1})$. (shifting is possible because $F_a \cap C(w_l+1, w_{l+1})$ includes $2k_l - 1$ edges). we obtain these equalities:

$$\begin{aligned} a(m_l+1) &= a(m_l+2) \\ a(m_l+3) &= a(m_l+t) \text{ for all } t = 4, \ldots, 3k_l - 1 \\ a(m_l+3k_l) &= a(m_{l+1}+1) \end{aligned}$$

  By fixing all edges of the forest F_a not in $C(w_{l-1} + 1, w_l)$ and construct other star forests by the shift of the two edges of F_a in $C(w_{l-1}+1, w_l)$

$$\begin{aligned} a(m_l-1) &= a(m_l-2) \\ a(m_l-3) &= a(m_l-t) \text{ for all } t = 4, \ldots, 3k_{l-1} - 1 \\ a(m_l-3k_{l-1}) &= a(m_{l-1}-1) \end{aligned}$$

  by applying the same construction to all edges of E_W we obtain these equalities for all $l \in \{1, 2; \ldots, p\}$
  By considering $C(w_l + 1, w_{l+1})$ according to m_{l+1} we have:

$$\begin{aligned} a(m_{l+1}-1) &= a(m_{l+1}-2) \\ a(m_{l+1}-3) &= a(m_{l+1}-t) \text{ for all } t = 4, \ldots, 3k_{l-1} - 1 \end{aligned}$$

  we knew that : $m_{l+1} - t = m_l + (k_l - t)$ for all $t = 1, \ldots, k_l$
  then:

$$\begin{aligned} a(m_l+1) &= a(m_l+t) = a_0^k \text{ for all} t = 2, \ldots, 3k_l \\ a_0^l &= a_0^{l+1} \end{aligned}$$

  and this is true for all $l \in \{1, 2, \ldots, p\}$
  Let F_a^l the forests those contain m_l then $a^T x^{F_a^l} = \alpha$ because $F_a \cap C(w_l+1, w_{l+1})$ includes $2k_l - 1$ edges) we can choose F_a^l without the edges m_l+1 and m_l-1

- **If** $m_l \notin F_a$
  $F_a \cap C(w_l + 1, w_{l+1})$ includes $2k_l$ edges and $F_a \cap C(w_{l-1} + 1, w_l$ includes $2k_{l-1}$ edges.
  Let $F_a^{\bar{l}}$ the forest which don't contain m_l then we can choose $F_a^{\bar{l}} = F_a^l \setminus m_l \cup \{m_l + 1, m_l - 1\}$
  Then we have:

$$a(m_l) = a(m_l+1) + a(m_l-1) = a_0^k + a_0^{l-1} = 2a_0^l$$

  We summarize:

$$\begin{aligned} a(m_i) &= 2a_0 \text{ for all } i = 1, \ldots, p \\ a(e) &= a_0 \text{ for all } e \in C \setminus M \end{aligned}$$

Now we have to calculate α. Consider a star forest $F^* \in \mathbf{F_a}$ containing all edges of E_W, then we have:

$$\begin{aligned} \alpha &= a^T x^{F^*} \\ &= \sum_{i=0}^{p-1} 2k_i \times a_0 - [\frac{p-1}{2} - 1 + 2]a_0 + p(2a_0) \\ &= (2\sum_{i=0}^{p-1} k_i + \lfloor \frac{3p}{2} \rfloor) \times a_0 \end{aligned}$$

By fixing $a_0 = 0$ we obtain $a(e) = 0$ for all $e \in E$. We conclude that $a^T x \leq \alpha$ is equivalent to (6), proving (6) defines a facet for the star forest polytope. ■

Theorem 3. *The matching-cycle inequalities* (6) *can be separated in polynomial time.*

Proof: Note that $|C(w_i+1, w_{i+1})| = 3k_i$ for $i = 1, \ldots, p$, we can get the following.

Let $e = uv \in E(C)$, consider $y(u, v) = 1 - x(u)$ for all $u \in C$.

Consider a cycle $C = (1, 2, \ldots, n)$ on a $W = \{w_1, \ldots, w_p\}$. $(|w_i, w_{i+1}|) = 3k_i$. let define $x(e) = 1$ if $v \in D$ and $x(e) = 0$ otherwise. where D is the dominating set.

Let define y as follows, for all $e \in E(C)$, $e = uv$. $y(uv) = 1 - x(u)$ for all $u \in C$.

By replacing $y(uv) = 1 - x(u)$ in (6) by $1 - x(u)$ we get

$$x(C) + x(W) \geq \sum_{i=1}^{p} k_i + \lceil \frac{p}{2} \rceil$$

It is known from (6) that (I) is separated in polynomial time.

■

A. Complete description of $SFP(G)$ when G is a cycle

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

Proof:

Suppose that there is a star forest which is not a maximal and which verify (6) at equality. Let $ax \leq \alpha$ the inequality representing(6). It means that $ax^F = \alpha$. Or F is not maximum, it exist $e \in E$ such us $F \cup \{e'\} = F'$ 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.*

Proof:

Let $F \subseteq E$ be a spanning star forest in G, and consider $\bar{F}$ its complementary edge-set. We proceed by Contradiction, we suppose that $\bar{F}$ is not an edge dominating set. Suppose that there is an edge in F such that, it is not dominated by no edge in $\bar{F}$, then it must be bordered by two edges in F, thus form by this way a path of length 3 in F. This contradict the fact that a star forest don't admit 3-paths and 3-cycles. ■

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.*

Proof: It is known that a cycle C is isomorphic to its line graph $L(C)$. An edge in C became a vertex in $L(C)$ and a vertex in C became an edge in $L(C)$. then if a edge e dominates a set of edges M in $E(C)$ then $l(e)$ which is the vertex representing e in $L(C)$ dominates the vertices representing the edges $M \subseteq E(C)$ in $L(C)$.

■

In the other hand, Bouchakour *et al.* [5] have shown that,

Theorem 7. *[5] When the graph G is a cycle C, the complete description of dominating set polytope is given by the following system:*

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

Remark 2. *When a graph G is cycle, the edge dominating set polytope $EDSP(G)$ is characterized completely by the following system:*

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

Proof: For each edge $e \in E(C)$, it correspond 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 7).

■

Based on these results, we prove that,

Theorem 8. *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)*.*

Proof: The proof can be derived from Lemmas 4, 5, 6 and Theorem 7. 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.

■

B. 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 [10] 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 $i = 1, \dots, n-1$. In particular, the edge e_n of weight c_n is the edge between n and 1. We transform the $MWSFP$ in G into 6 problems of finding a longest path problem 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 i_{-2}, i_{-1}, i_1, i_2 called the clones of i. The arcs are created as follows: for every vertex $1 \ge i \le n-1$ in G, there are :

- an arc $(i_{-2}, (i+1)_1)$ of cost c_i,
- an arc $(i_{-1}, (i+1)_1)$ of cost c_i, an arc $(i_{-1}, (i+1)_{-2})$ of cost 0,
- an arc $(i_1, (i+1)_2)$ of cost c_i, an arc $(i_1, (i+1)_{-1}$ of cost 0, And finally, an arc $(i_2, (i+1)_{-1})$ of cost 0.

We prove the following.

Proposition 1. *There is at most two successive arcs with the same color in an $s-t$- path in $G' = (X', A')$*

Proof: In the digraph $G' = (X', A')$ as it is defined, a vertex i_x express that there is $|x|$ red arcs if x is negative and x blue arcs if the x is positive. Or by definition of the network N we have the following possibilities:

- $x = -2$ then the vertex i_{-2} is preceded by $x = 2$ red edges. And, from $x = -2$ the unique possibility to achieve is $x = 1$ or the color of the edge in this case is blue.
- $x = -1$ then the vertex i_{-1} is preceded by $x = 1$ red edge. And, from $x = -1$ there is two possibilities. Either achieve $x = 1$ or the color of the edge in this case is blue. Or $x = -2$ in this case the color of the edge is red, then the next edge must be blue. This is the case, because from $x = -2$ we have one possibility, it is to achieve $x = 1$ and the arc is blue.
- $x = 2$ then the vertex i_2 is preceded by $x = 2$ blue edges. And, from $x = 2$ the unique possibility to achieve is $x = -1$ or the color of the edge in this case is red.
- $x = 1$ then the vertex i_1 is preceded by $x = 1$ red edge. And, from $x = 1$ there is two possibilities. Either achieve $x = -1$, the color of the edge in this case is red. Or achieve $x = 2$ in this case the color

of the edge is blue, then the next edge must be blue. This is the case, because from $x = 2$ we have one possibility, it is to achieve $x = -1$ and the arc is red.

It means that there is no three successive edges with the same color (neither in red nor in blue) ∎

Lemma 9. *A non zero cost arcs of a path from 1_x to n_x in G' form a spanning star forest in G*

Proof: By corresponding each vertex i_x in G' to a vertex i in G and each non zero arc (i_x, j_y) in the path in G' by a blue edge ij in G. Because there is at most 2 successive arcs with the same cost (zero or non zero)) this reflect the same to the graph G thus we obtain at most 2 successive edges with the same color. By this the edge set obtained verify the 3- path inequalities and if we affect to a blue edge a value 1 and to the red edge a value 2. Thus the edge set as it is defined verify the integer formulation of star forest. ∎

Remark 3. *All maximal star forest in G correspond to one of the following:*

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

Theorem 10. *The problem of finding a maximum weight star forest in G is equivalent to find 6 longest paths from 1_x to n_y in G' with several values of x, y where $x, y \in \{-2, -1, 1, 2\}$. This can be done obviously in linear time.*

Proof: Each longest paths from 1_x to n_y in graph G' define a spanning star forest in G due to lemma 9.

Suppose that we have a maximum spanning star forest F^* in G (G is a cycle). Assign to each edge in F^* a color blue and to th edge not in F^* a color red. If we denote by i an edge by $i(i+1)$ where $i = 1, \ldots n$,consider $n + 1 = 1$. From a graph G we construct a directed graph G' as it is constructed above. Then we choose a path from 1_x to n_y respecting the colors of edges in G :

- If the edge is blue we choose one of arcs $i_x i_y$ with $x \leq y$.
- If the edge is red we choose an arc $i_x i_y$ with $x \geq y$

∎

Hence a simple algorithm to find a maximum weight star forest in G is to find 6 longest paths from 1_x to n_x as specified in remark 3. This can be done obviously in linear time.

II. An extended formulation

In the following section we are interested on extended formulation of the star forest Polytope. An extended formulation was given by Baiou *et al.* [2] via facility location polytope. This formulation with an exponential number of constraints due to the g-odd cycle inequalities. A compact formulation means the formulation defined with a polynomial number of variables and polynomial number of constraints. In what follows , we are interested in such formulation for star forest polytope. To do so we define the following graph.

First, we built a graph G' from G as its defined in the precedent section. We obtain a graph with vertices i_x where $x \in \{-2, -1, 1, 2\}$.

Secondly, we construct a digraph $G" = (X", A")$ by duplicate 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 $1 \leq j \leq 6$ add a those arcs by this way:

- an arc $(s, 1_{-2}^1)$ of cost 0, an arc (n_{-1}^1, t) of cost 0,
- an arc $(s, 1_{-1}^2)$ of cost 0, an arc (n_1^2, t) of cost 0,
- an arc $(s, 1_{-1}^3)$ of cost 0, an arc (n_2^3, t) of cost 0,
- an arc $(s, 1_1^4)$ of cost C_1, an arc (n_{-1}^4, t) of cost C_n,
- an arc $(s, 1_1^5)$ of cost C_1, an arc (n_{-2}^5, t) of cost C_n,
- an arc $(s, 1_2^6)$ of cost C_1, an arc (n_1^6, t) of cost C_n,

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 size to the vertex n. This construction is done by giving a color red to the edges withe cost 0 and blue to the edges with a cost c_i (red mean the edge is not in the star forest and blue 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.

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.

$$\max \sum \phi_i^j \tag{16}$$

$$\phi(\delta^+(i_x^j)) - \Phi(\delta^-(i_x^j)) = 0 \, \forall i_x^j \in X \tag{17}$$

$$\phi(\delta^-(s)) = 1 \tag{18}$$

$$\phi(\delta^+(t)) = 1 \tag{19}$$

$$\phi(a) \geq 0 \, \forall a \in A \tag{20}$$

The graph as it is constructed above is acyclic, and because system above is totally unimodular, the solution of the system given above is integer. This formulation is an $s - t$-path extended formulation for the star forest polytope.

The set of all $s - t$ flows of value 1 in network $N = (X, A)$ defines a polyhedron Q that we call flow polyhedron. the extreme point of the polyhedron define an s-t path.

We say that the flow polyhedron Q is a flow-based extension of a given polytope P in $\mathbb{R}^d$ if there exist a linear projection $\pi : \mathbb{R}^A \longrightarrow \mathbb{R}^d$ such that $\pi(Q) = P$.

Let $N = (X, A)$ be a network where i_x^j is vertex in X those are the clones of a vertex i, and A is the set of edges. For ease of notation we denote by $i_{xy}^j = i_x^j(i+1)_y^j$ the the arc between i_x^j and i_y^j in N, and by i the edge between i and $i+1$ in G.

Let L the edge set of st-path in N Let $\phi^L \in \{0,1\}^{|A|}$ be the characteristic vector of an st-path L in N and define an extreme point of the st-flow polyhedron.

$\phi^L(i_{xy}^j) = \begin{cases} 1 & \text{if } i_{xy}^j \in L, \\ 0 & \text{otherwise.} \end{cases}$ Then the projection π : $\mathbb{R}^A \longrightarrow \mathbb{R}^d$
$\phi^L \longmapsto z$ then $\pi(\phi^L(i_{xy}^j)) = z(i)$ where
$z(i) = \begin{cases} 1 & \text{if } \exists \text{ an arc } i_{xy}^j \in L, \text{ with } x < y \\ 0 & \text{otherwise.} \end{cases}$

An edge i in G take the same color (blue or red) with an arc i_{xy}^j in a path L of N. the arc i_{xy}^j take a color blue if $x < y$ and red if $x > y$. $z(i) = 1)$ means the edge i take a color blue , the same color of an arc i_{xy}^j

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

$P = \text{ Proj}_x\{(z, \phi) \in \mathbb{R}^{|E|} \times \mathbb{R}^{|A|}; z(i) = \sum_j \sum_{xy} \phi(i_{xy}^j)\}$

Theorem 11. *The system* (17)-(20) *defines an extended formulation for the star forest polytope $SFP(G)$ when the graph G is cycle.*

Proof: An extrem point of the system (17)-(20) define an characteristic vector of an st-path in N. Or each st-path in $N = (X, A)$ is a succession of blue (non zero cost) and red (zero cost) arcs. By construction of the network N and by proposition 1, there is no two successive edge with the same color. By projecting the st-path of $N = (X, A)$ on blue edges in G we obtain a spanning star forest in G. ∎

Let $\bar{z}(i) = 1 - z(i)$

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

$\bar{P} = \text{ Proj}_{\bar{z}}\{(\bar{z}, \phi) \in \mathbb{R}^{|E|} \times \mathbb{R}^{|A|}; \bar{z}(i) = \sum_j \sum_{xy} \phi(i_{xy}^j)\}$

Theorem 12. *The system* (17)-(20) *defines an extended formulation for edge dominating polytope.*

Proof: By projecting the st-path of N on the red edge of G, we obtain an edge dominating set.

It is proved in section 2 that the complementary of star forest in cycle is an edge dominating set. Or a red edge in G means the edge is not in a star forest. Because G is cycle, red edges define an edge dominating set. ∎

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

Corollary 1. *The system* (17)-(20) *define an extended formulation for vertex dominating set.*

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

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

[3] 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.

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

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

[6] 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.

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

[8] 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.

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

[10] 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).

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