# A Cutting Plane Method for Least Cost Influence Maximization

Cheng-Lung Chen[1] , Eduardo L. Pasiliao[2] , and Vladimir Boginski[1(✉)]

[1] University of Central Florida, Orlando, FL 32816, USA
vladimir.boginski@ucf.edu
[2] Air Force Research Laboratory, Eglin AFB, FL 32542, USA

**Abstract.** We study the least cost influence maximization problem, which has potential applications in social network analysis, as well as in other types of networks. The focus of this paper is on mixed-integer programming (MIP) techniques for the considered problem. The standard arc-based MIP formulation contains a substructure that is a relaxation of the mixed 0-1 knapsack polyhedron. We give a new exponential class of facet-defining inequalities from this substructure and an exact polynomial time separation algorithm for the inequalities. We report preliminary computational results to illustrate the effect of these inequalities.

**Keywords:** Mixed-integer programming · Influence maximization · Social networks · Valid inequalities

## 1 Introduction

Intricate connections between entities in many natural and man-made systems form large complex networks. Of particular interest in the area of network science is gaining insight into the dynamic behavior of spreading or influence processes in complex networks. For instance, in social network analytics, optimal initiation of the processes of spreading information, opinions, and/or influence, may play an important role in designing competitive marketing strategies. Accordingly, there is an increasing trend in studying influence and information propagation in social networks (see, e.g., [4,12]). Granovetter [7] propose the linear threshold model to describe the propagation process in social network, in which the resistance of an individual to influence and influence strength to others are quantified as threshold and influence factor, respectively. The term "active" is adopted to represent the state of individual behavior being influenced if the summation of influence factors from all the connections in social network exceeds the threshold. There are many variants of this problem related to optimally determining the most influential nodes (people), in order to trigger the propagation process and reach a desired penetration rate. Kempe et al. [11] consider the *Influence*

*Maximization Problem* (IMP), which they formulate as a discrete stochastic optimization problem. They adopt two models for diffusion processes, namely, the *linear threshold* and the *independent cascade* models. The goal is to activate some users initially and use them to influence as many other users as possible by the end of the propagation process. They show that it is NP-hard to both approximate and solve the problem to optimality. Another similar problem introduced by Chen [3] is referred to as the *Target Set Selection Problem* (TSSP). In TSSP, the decision is to find the minimum number of users required initially in order to activate the entire network through the propagation process. Chen showed that the problem is NP-hard to approximate and gives a polylogarithmic lower bound on the approximation ratio. Recently, a new problem named *Least Cost Influence Maximization Problem* (LCIM) has been introduced in [5]: it involves the combination of individual incentives (e.g., discounts, payments, free sample products) with peer influence together to activate nodes and prompt influence propagation in a social network. The goal of LCIM is to determine the required minimum cost of partial incentives given to the key opinion leaders.

Despite the fact that the aforementioned problems share certain similarities, the challenges of finding an exact optimal solution can be very different when these problems are formulated by mathematical optimization models. In this paper, we consider the LCIM problem and formulate it as a mixed-integer programming problem to study its polyhedral structure. We assume that all the parameters are deterministic and the influence propagation occurs in discrete time steps. From a practical point of view, the assumption of deterministic linear threshold depends on the accuracy of estimation of influence factor and threshold parameters. Machine learning and data mining techniques may enable one to obtain accurate predictions on those parameters from massive amounts of data available nowadays. A similar assumption on deterministic linear threshold model can be found in [10], where the authors consider targeted and budgeted influence maximization in social networks and give an iterative greedy algorithm to solve the problem. Most of the previous studies on social network optimization problems mainly focus on developing heuristic and approximation algorithms. Existing studies on exact integer programming methods for influence maximization problems are relatively limited. Raghavan and Zhang [15] study the Weighted Target Set Selection problem (WTSSP) in which each node is associated with a unique cost in the objective function for initial activation. They give a compact and tight extended formulation for WTSSP on tree graphs and later show it is also tight on directed acyclic graphs. To apply this extended formulation to general graphs, they design a branch-and-cut algorithm that includes a separation for cycle elimination constraints. Wu and Küçükyavuz [17] study the two-stage stochastic influence maximization problem where the second-stage cost function is submodular. They develop a delayed constrained generation algorithm with strong optimality cuts that utilizes the submodularity

and demonstrate its effectiveness in extensive computational results. Nannini et al. [14] propose a branch-and-cut algorithm and heuristic branch-cut-and-price algorithms for robust influence maximization, where node thresholds and arc influence factors are subject to budget uncertainty. They show that optimization for a worst-case scenario robust solution is NP-hard. Fischetti et al. [6] present a novel set covering formulation for generalized LCIM. They propose strengthened generalized propagation inequalities and show that they dominate the cycle elimination constraints in the original formulation. A price-cut-and-branch algorithm with heuristic separation for the proposed inequalities and column generation is given to deal with the exponential number of variables and constraints. Günneç et al. [9] establish the computational complexity for LCIM based on the reduction from the independent set problem. In particular, when 100% penetration rate is not required, they show that LCIM is NP-hard on arbitrary graphs and bipartite graphs for both equal and unequal influence. For the 100% penetration rate, the optimization of LCIM with unequal influence on a tree remains NP-hard. On the other hand, LCIM with equal influence on a tree with the 100% penetration rate is shown to be polynomially solvable. They give a greedy algorithm and a total unimodular formulation for this special case. In the subsequent paper, Günneç et al. [8] extend their total unimodular formulation for LCIM on a tree to an arbitrary graph. To ensure the solution is acyclic, they give several pre-processing steps and separation for cycle elimination constraints in the branch-and-cut algorithm.

## 1.1  Notation and Problem Definition

For convenience, we use the notation $[n] = \{1, \cdots, n\}$ and subscripts to indicate the elements of a vector. The $n$-dimensional $j$th unit vector is denoted as $e_j$. For a set $Q \subseteq \mathbb{R}^n$, we use conv $(Q)$ to denote its convex hull of solutions.

Formally, a given network (e.g., a social network) is represented by a directed graph $G = (V, A)$, where the set of nodes $V$ with cardinality $n$ may correspond to the set of people and set of arcs $A$ with cardinality $m$ indicates the connection and influence direction between the people in the network. Each node $i \in V$ has threshold $h_i$ and each arc $(i, j) \in A$ is associated with an influence weight $d_{ij}$. The coverage (penetration) rate is denoted by $\tau$, where $0 < \tau \leq 1$, and the neighborhood of node $i$ is denoted by $N_i := \{j \in V : (j, i) \in A\}$. We assume that $d_{ij}$ and $h_i$ are positive integers such that $\max\{d_{ji} : j \in N_i\} < h_i$ for all $i \in V$ to omit trivial cases. All nodes are assumed inactive initially and nodes remain active once influences from neighbors and incentives reach the threshold. For each node $i \in V$, let continuous variables $x_i$ be the amount of partial incentives given to user $i$, binary variables $y_{ij}$ indicate whether influence is exerted from node $i$ to $j$, and binary variables $z_i$ indicate whether node $i$ is activated. The arc-based formulation of LCIM is given by

$$\min_{x,y,z} \quad \sum_{i \in V} x_i$$

$$x_i + \sum_{j \in N_i} d_{ji} y_{ji} \geq h_i z_i \quad \forall i \in V \tag{1}$$

$$z_i \geq y_{ij} \quad \forall (i,j) \in A \text{ s.t. } (j,i) \notin A \tag{2}$$

$$\sum_{i \in V} z_i \geq \lceil \tau n \rceil \tag{3}$$

$$\sum_{(i,j) \in C} y_{ij} \leq \sum_{i \in V(C) \setminus \{k\}} z_i \quad \forall k \in V(C), \forall \text{ cycles } C \subseteq A \tag{4}$$

$$x \in \mathbb{R}^n_+$$

$$y \in \mathbb{B}^m, z \in \mathbb{B}^n.$$

Node propagation constraints (1) evaluate the total incoming influence from neighbor plus the incentives given to a node. Constraints (2) ensure that arc $(i,j)$ exerts influence if node $i$ is activated. The minimum coverage constraints (3) describe the number of nodes that need to be activated given a predetermined penetration rate $\tau$. The generalized cycle elimination constraints (4) where $V(C) = \{i \in V : (i,j) \in C\}$ cut off solutions that form a cycle as the induced optimal influence propagation graph is supposed to be acyclic. Note that the arc-based formulation proposed by [2] is different from this paper as the influence weights are coming solely from their neighbors without incentives. Günneç et al. [8] and Günneç et al. [9] on the other hand, consider the arc-based formulation with time index. Finally, Fischetti et al. [6]. adopt this arc-based formulation for computational performance comparison but the possible values of incentives are represented by a set of binary variables.

## 1.2   Main Contribution

Our main contribution can be summarized as follows: We give a class of valid inequalities derived from the substructure of the model that describes the propagation via deterministic linear threshold model. The substructure can be transformed to the mixed 0-1 knapsack polyhedron with additional binary restriction on partial knapsack size. Hence, it is a relaxation containing known valid inequalities from mixed 0-1 knapsack set studied by Marchand and Wolsey [13]. We introduce a new class of valid inequalities and give an exact polynomial separation algorithm for them. We also show that by exploiting the result of our separation algorithm, the inequalities proposed in [13] with heuristic separation only, can now be separated exactly as well.

## 2 Valid Inequalities in LCIM Based on Mixed 0-1 Knapsack Polyhedron

To develop a strong formulation for LCIM, we study the polyhedral structure of constraints (1). Assume $N_i$ is nonempty with cardinality $t_i$ and $\sum_{i \in V} t_i = m$. For $i \in [n]$, let

$$\mathcal{X}_i = \left\{ (x_i, y, z_i) \in \mathbb{R}_+ \times \mathbb{B}^{t_i} \times \mathbb{B} : x_i + \sum_{j \in N_i} d_{ji} y_{ji} \geq h z_i \right\}.$$

The set $\mathcal{X}_i$ describes the node propagation in LCIM, which can be regarded as a mixing set with a binary variable on the right-hand side value. Any inequality that is facet-defining for conv $(\mathcal{X}_i)$ is facet-defining for conv $(\cap_{i \in [n]} \mathcal{X}_i)$ as well. Therefore, we now consider a single node propagation by dropping the subscript $i$ and obtain the following set

$$\mathcal{X} = \left\{ (x, y, z) \in \mathbb{R}_+ \times \mathbb{B}^t \times \mathbb{B} : x + \sum_{j \in N} d_j y_j \geq h z \right\}.$$

Observe that the set $\mathcal{X}$ contains a mixed 0-1 knapsack structure. Let set $\overline{\mathcal{X}}$ be obtained from $\mathcal{X}$ by setting $\overline{y}_j = 1 - y_j$, $j \in N$ and $z = 1$. Then we obtain the mixed 0-1 knapsack set $\overline{\mathcal{X}}$ with weight $d_j$ for each item $j \in N$ and the capacity of knapsack $\left( \sum_{j \in N} d_j - h \right)$ plus an unbounded continuous variable $x$ in the following

$$\overline{\mathcal{X}} = \left\{ (x, \overline{y}, z) \in \mathbb{R}_+ \times \mathbb{B}^t \times \{1\} : \sum_{j \in N} d_j \overline{y}_j \leq \left( \sum_{j \in N} d_j - h \right) + x \right\}.$$

Such set can be interpreted as a special case of traditional 0-1 knapsack problem where the knapsack size is expanded with additional capacity. Marchand and Wolsey [13] propose two classes of valid inequalities for $\overline{\mathcal{X}}$ based on mixed-integer rounding and lifting function, namely, the continuous cover inequalities and continuous reverse cover inequalities, and they can immediately be used to strengthen the formulation of LCIM as $\overline{\mathcal{X}} \subset \mathcal{X}$.

**Proposition 1** [13]**.** *Let index $k$, set $S \subseteq N$ and set $T \subseteq N$ be a $(k, S, T)$ cover pair that satisfies (i) $S \cap T = \{k\}$, $S \cup T = N$, (ii) $\pi = h + \sum_{j \in S} d_j - \sum_{j \in N} d_j > 0$, and $h + \sum_{j \in S \setminus \{k\}} d_j - \sum_{j \in N} d_j < 0$, (iii)$\rho = \sum_{j \in T} d_j - h > 0$, and $\sum_{j \in T \setminus \{k\}} d_j - h < 0$. Note that these conditions also imply $\pi + \rho = d_k > 0$. Let $r_S = \min\{j \in S : d_j > \pi\}$ where $d_j \in S$ are in non-decreasing order such that $d_1 \geq d_2 \geq \cdots \geq d_{r_S}$. Similarly, let $r_T = \min\{j \in T : d_j > \rho\}$ where $d_j \in T$ are in non-decreasing order such that $d_1 \geq d_2 \geq \cdots \geq d_{r_T}$. In addition, let $D_0^S = D_0^T = 0$, $D_j^S = \sum_{\ell=1}^j d_\ell, j \in [r_S]$, $D_j^T = \sum_{\ell=1}^j d_\ell, j \in [r_T]$. Then the*

*following continuous cover and continuous reverse cover inequalities are valid for $\mathcal{X}$.*

$$x + \sum_{j \in S} \min\{\pi, d_j\} y_j + \sum_{j \in T \setminus \{k\}} \phi_S(d_j) y_j \geq \left( \min\{\pi, d_k\} + \sum_{j \in T \setminus \{k\}} \phi_S(d_j) \right) z \tag{5}$$

$$\text{and} \quad x + \sum_{j \in T} \max\{0, d_j - \rho\} y_j + \sum_{j \in S \setminus \{k\}} \psi_T(d_j) y_j \geq \left( \sum_{j \in T} \max\{0, d_j - \rho\} \right) z \tag{6}$$

*where*

$$\phi_S(g) = \begin{cases} (j-1)\pi & D_{j-1}^S \leq g \leq D_j^S - \pi, \quad j \in [r_S] \\ (j-1)\pi + g - D_j^S + \pi & D_j^S - \pi \leq g \leq D_j^S, \quad j \in [r_S - 1] \\ (r_S - 1)\pi + g - D_{r_S}^S + \pi & D_{r_S}^S - \pi \leq g, \end{cases} \tag{7}$$

*and*

$$\psi_T(g) = \begin{cases} g - j\rho & D_j^T \leq g \leq D_{j+1}^T - \rho, \quad j \in [r_T - 1] \cup \{0\} \\ D_j^T - j\rho & D_j^T - \rho \leq g \leq D_j^T, \quad j \in [r_T - 1] \\ D_{r_T}^T - \rho r_T & D_{r_T}^T - \rho \leq g. \end{cases} \tag{8}$$

*Proof.* If $z = 0$, both inequalities (5) and (6) are trivially satisfied. Otherwise, the validity and facet proof of both inequalities directly follows from [13].

**Example 1.** Let $d = (7, 6, 5, 4)$ and $h = 8$, we list the facet-defining inequalities from each $(k, S, T)$ pair of inequality (5) and (6) in Table 1. For example, for $k = 1$, $S = \{1, 2, 4\}$ and $T = \{1, 3\}$, we have $\pi = 3$, $\rho = 4$, $r_S = 3$, and $r_T = 2$. Then the lifting function $\phi_S$ is given by

$$\phi_S(g) = \begin{cases} 0 & 0 \leq g \leq 4 \\ g - 4 & 4 \leq g \leq 7 \\ 3 & 7 \leq g \leq 10 \\ g - 7 & 10 \leq g \leq 13 \\ 6 & 13 \leq g \leq 14 \\ g - 8 & 14 \leq g \end{cases}$$

Hence the coefficient of $y_3$ is $\phi_S(d_3) = \phi_S(5) = 5 - 4 = 1$.

Essentially, the continuous cover inequalities (5) and continuous reverse cover inequalities (6) are not sufficient to describe conv $(\mathcal{X})$, as the additional binary variable $z$ creates new extreme points. Furthermore, no exact separation algorithm for inequalities (5) and (6) has been proposed yet. Next we introduce a new class of valid inequalities for $\mathcal{X}$ that utilizes the concept of minimal influencing set. We use the similar definition of minimal influencing set from [6], which we include here for the reader's convenience:

**Table 1.** Continuous cover and continuous reverse cover inequalities of Example 1

| $x + 7y_1 + 6y_2 + 5y_3 + 4y_4 \geq 8z$ | | |
|---|---|---|
| Index $k$ | Set | Facet-defining inequality |
| 2 | $S = \{2, 3, 4\}, T = \{1, 2\}$ | $x + y_1 + y_2 + y_3 + y_4 \geq 2z$ |
| 1 | $S = \{1, 3, 4\}, T = \{1, 2\}$ | $x + 2y_1 + y_2 + 2y_3 + 2y_4 \geq 3z$ |
| 1 | $S = \{1, 2, 4\}, T = \{1, 3\}$ | $x + 3y_1 + 3y_2 + y_3 + 3y_4 \geq 4z$ |
| 1 | $S = \{1, 2, 3\}, T = \{1, 4\}$ | $x + 4y_1 + 4y_2 + 4y_3 + y_4 \geq 5z$ |
| 2 | $S = \{1, 2, 4\}, T = \{2, 3\}$ | $x + 4y_1 + 3y_2 + 2y_3 + 3y_4 \geq 5z$ |
| 2 | $S = \{1, 2, 3\}, T = \{2, 4\}$ | $x + 5y_1 + 4y_2 + 4y_3 + 2y_4 \geq 6z$ |
| 3 | $S = \{1, 2, 3\}, T = \{3, 4\}$ | $x + 6y_1 + 5y_2 + 4y_3 + 3y_4 \geq 7z$ |

**Definition 1** [6]. *Let $p_i \in [h_i - 1] \cup \{0\}$ be an incentive payment to node $i \in V$ and $M \subseteq N_i$ be a set of active neighbors of node $i \in V$, such that $p_i + \sum_{j \in M} d_{ji} = h_i$. We say $M$ is a minimal influencing set for node $i \in V$ if and only if for a fixed incentive payment $\overline{p}_i$, it satisfies $\overline{p}_i + \sum_{j \in M} d_{ji} = h_i$ and $\overline{p}_i + \sum_{j \in M \setminus \{k\}} d_{ji} < h_i$ for any $k \in M$. In other words, a strict subset of $M$ with the same incentive payment are not sufficient to activate node $i$. For each node $i \in V$, let $\Omega_i \subseteq N_i$ be the superset of all minimal influencing sets.*

**Theorem 1.** *Let $M \subseteq N$ be a minimum influencing subset with an incentive payment $p > 0$. The minimal influencing subset inequality*

$$x + \sum_{j \in N \setminus M} \min\{d_j, p\} y_j \geq pz \qquad (9)$$

*is valid for $\mathcal{X}$.*

*Proof.* If $z = 0$ then inequality (9) is trivially satisfied. If $y_j = 0$ for all $j \in N \setminus M$, then either $x = 0$ for $z = 0$ or $x = p$ for $z = 1$. Assume that none of these cases hold, given a $p > 0$, rewrite the left term of the inequality in $\mathcal{X}$ in the following form

$$x + \sum_{j \in N} d_j y_j$$

$$= x + \sum_{j \in N \setminus M: d_j \leq p} d_j y_j + p \sum_{j \in N \setminus M: d_j > p} y_j + \sum_{j \in M} d_j y_j \geq h,$$

which implies

$$x + \sum_{j \in N \setminus M: d_j \leq p} d_j y_j + p \sum_{j \in N \setminus M: d_j > p} y_j \geq h - \sum_{j \in M} d_j y_j \geq h - \sum_{j \in M} d_j = p.$$

**Theorem 2.** *Inequality* (9) *is facet-defining for* conv $(\mathcal{X})$ *if and only if* $p >$
*0. Moreover, for a given* $i \in V$ *and a set* $N_i$, *for each* $M \subseteq N_i$ *such that*
$h_i - \sum_{j\in M} d_{ji} = p_i > 0$, *the minimal influencing subset inequality*

$$x_i + \sum_{j \in N_i \setminus M} \min\{d_{ji}, p_i\} y_{ji} \geq p_i z_i \tag{10}$$

*is facet-defining for* conv $(\cap_{i \in [n]} \mathcal{X}_i)$.

*Proof.* Note that $\mathcal{X}$ is full-dimensional and contains the origin. If $p = 0$, the
inequality (9) reduces to $x \geq 0$, therefore $p > 0$ is a necessary and sufficient
facet condition. To show that inequalities (9) is facet-defining for $\mathcal{X}$, we exhibit
$t + 1$ linearly independent points on the face defined by inequality (9). Consider
the two feasible points where $x^0 = z^0 = 0$, $x^1 = h - d_j$, $z^1 = 1$, $y_j^0 = y_j^1 = $
1 if $j \in M$ and $y_j^0 = y_j^1 = 0$ otherwise. Next, for a fixed $j \in M$ and for
each $k \in N \setminus M$, consider the feasible points $(x^k, y_j^k, z^k) = (0, y_j^0 + e_k, 1)$. It
is straightforward to verify that these $t + 1$ points are linearly independent and
satisfy inequality (9) at equality. The second part of this theorem directly follows
the above by considering $(x_i^0, y_{ji}^0, z_i^0) = (0, e_j, 0)$ and $(x_i^1, y_{ji}^1, z_i^1) = (h_i - d_{ji}, 1, 1)$
if $j \in M$, $y_{ji}^0 = y_{ji}^1 = 0$ otherwise, there are $2n$ points in this form for $i \in V$.
Also, consider the $m - 1$ points $(x_i^k, y_{ji}^k, z_i^k) = (0, y_{ji}^0 + e_k, 1)$ for $i \in V$, a fixed
$j \in M$ and for each $k \in N_i \setminus M$. These $2n + m - 1$ points on the face defined
by inequality (10) are linearly independent, therefore inequality (10) is facet-
defining for conv $(\cap_{i \in [n]} \mathcal{X}_i)$.

**Example 1 (Continued).** The facet-defining inequalities of (9) for Example
1 are listed in Table 2

**Table 2.** Minimal influencing subset inequalities of Example 1

| $x + 7y_1 + 6y_2 + 5y_3 + 4y_4 \geq 8z$ | |
|---|---|
| Set | Facet-defining inequality |
| $M = \{1\}$ | $x + y_2 + y_3 + y_4 \geq z$ |
| $M = \{2\}$ | $x + 2y_1 + 2y_3 + 2y_4 \geq 2z$ |
| $M = \{3\}$ | $x + 3y_1 + 3y_2 + 3y_4 \geq 3z$ |
| $M = \{4\}$ | $x + 4y_1 + 4y_2 + 4y_3 \geq 4z$ |

Although inequalities (5), (6) and (9) define a large number of facets for conv $(\mathcal{X})$,
they are not sufficient to completely describe conv $(\mathcal{X})$ in its original space of
variables. Particularly, the following inequality is valid and facet-defining for this
example but cannot be obtained through inequalities (5), (6) or (9):

$$x + 3y_1 + 2y_2 + 2y_3 + 2y_4 \geq 4z.$$

## 2.1 Separation of Minimal Influencing Subset Inequalities

In this section, we give an exact polynomial time separation algorithm for finding the most violated minimal influencing subset inequality. From inequality (10), we observe that finding the most violated inequality for a given fractional solution $(x^*, y^*, z^*) \in \mathbb{R}_+^{2n+m}$ consists of choosing a set $M \subseteq N_i$ such that $p_i z_i - \sum_{j \in N_i \setminus M} \min\{d_{ji}, p_i\} y_{ji}$ is maximized. Let $t := \max\{|N_i| : i \in V\}$.

**Theorem 3.** *Given a fractional solution* $(x^*, y^*, z^*) \in \mathbb{R}_+^{2n+m}$ *from solving LCIM, there exists an* $O(nt \log t)$ *separation algorithm for inequality* (10).

*Proof.* Recall that a violated cut can be found if

$$p_i \left( z_i^* - \sum_{j \in N_i \setminus M : d_{ji} > p_i} y_{ji}^* \right) - \sum_{j \in N_i \setminus M : d_{ji} \leq p_i} d_{ji} y_{ji}^* > x_i^*,$$

which implies that it suffices to consider $y_{ji}^*$ for some $j \in N_i$ such that $z_i^* - \sum_{j \in N_i} y_{ji}^* > 0$ and $p_i > 0$. To do so, we sort $y_{ji}^*$ in a non-decreasing order for $j \in N_i$ with indices $j_1, j_2, \cdots, j_t$ such that $y_{j_1 i}^* \leq y_{j_2 i}^* \leq \cdots \leq y_{j_t i}^*$. For $j_1 \leq j_r \leq j_t$, we sum up first $r$ elements, then we check if $z_i^* - \sum_{\ell=1}^{r} y_{j_\ell i}^* > 0$ and $p_i' = h_i - \sum_{\ell=r+1}^{t} d_{j_\ell i} > 0$, until $z_i^* - \sum_{\ell=1}^{r+1} y_{j_\ell i}^* < 0$. These $r$ elements constitute the subset $M$ and $N_i \setminus M$ simultaneously and ensure $z_i^* - \sum_{j \in N_i \setminus M} y_{ji}^* > 0$ and $p_i > 0$ in order to generate a violated cut. The set $M$ that corresponds to the most violated cut can be determined by evaluating $\max \left\{ 0, p_i'(z_i^* - \sum_{\ell=1}^{r} y_{j_\ell i}^*) : r \in [1, t] \right\}$. If $\max \left\{ 0, p_i'(z_i^* - \sum_{\ell=1}^{r} y_{j_\ell i}^*) : r \in [1, t] \right\} = 0$, then there are no violated cuts. The sorting process runs in $O(t \log t)$ time and the evaluation takes $O(t)$ time, since we have to check for every node $i \in V$; thus, overall the separation algorithm runs in $O(nt \log t)$ time.

**Example 2.** Consider a directed tree graph where $V = \{1, 2, 3, 4, 5\}$ and $A = \{(1, 5), (2, 5), (3, 5), (4, 5)\}$. Assume the influence weight vector $\mathbf{d} = \langle 7, 6, 5, 4 \rangle$ and $h_5 = 8$. Let $\tau = 0.2$, the linear programming relaxation solution is $\mathbf{x}^* = \langle 0.53, 0, 0, 0, 0 \rangle$, $\mathbf{z}^* = \langle 0.53, 0, 0, 0, 0.47 \rangle$ and $\mathbf{y}^* = \langle 0.53, 0, 0, 0 \rangle$. To generate inequality (10) for node 5, we sort $\mathbf{y}^*$ in a non-decreasing order and compute $z_5^* - \sum_{\ell=1}^{r} y_{j_\ell 5}^*$ for $r \in [4]$. In this example, when $r = 3$, we have $M = \{2, 3, 4\}$ and $p_5 = 8 - 7 = 1$, therefore

$$x_5 + y_{25} + y_{35} + y_{45} \geq z_5$$

cut off this fractional solution.

## 2.2 Separation for Continuous Cover and Continuous Reverse Cover Inequalities

Until now we give an exact polynomial separation algorithm for inequalities (10). Next, we show that a violated continuous cover inequality for conv $\left( \cap_{i \in [n]} \mathcal{X}_i \right)$ can be identified by the result of Theorem 3. First, we establish the relationship between sets $S$ and $M$ formally.

**Lemma 1.** *Given $p = h - \sum_{j \in M} d_j > 0$, if there exists $k \in N \backslash M$ such that $\sum_{j \in M \cup \{k\}} d_j > h$, then $p = \pi$, $S = N \backslash M$, $\sum_{j \in M \cup \{k\}} d_j - h = \rho$ and $T = M \cup \{k\}$.*

*Proof.* First we arrange the term in the definition of $p$, let

$$p = h - \sum_{j \in M} d_j = h + \sum_{j \in N \backslash M} d_j - \sum_{j \in N} d_j.$$

Now, suppose there exists an element $k \in N \backslash M$ such that $\sum_{j \in M \cup \{k\}} d_j > h$. Since we have $\{M \cup \{k\}\} \cap N \backslash M = \{k\}$ and $\{M \cup \{k\}\} \cup N \backslash M = N$, it is clear that $S = N \backslash M$ and $T = M \cup \{k\}$ from Proposition 1. Note that $p$ is not necessary equal to $\pi$ as the range of $p$ contains 0.

Following Lemma 1, we give a theorem on how to determine a violated continuous cover inequality efficiently by using the information of the set $M$. Let $\hat{t} = \max\{|S| : S \subset N_i, i \in V\}$.

**Theorem 4.** *Given a fractional solution $(x^*, y^*, z^*) \in \mathbb{R}_+^{2n+m}$ from solving LCIM and a set $M$ corresponding to a violated inequality (10) for a fixed node $i \in V$, the most violated continuous cover inequality can be separated in $O(n\hat{t})$ time, if there exists any.*

*Proof.* Note that here we add an index $i$ to inequalities (5) similar to (10) for LCIM. Recall that inequality (10) is violated if

$$p_i \left( z_i^* - \sum_{j \in N_i \backslash M : d_{ji} > p_i} y_{ji}^* \right) - \sum_{j \in N_i \backslash M : d_{ji} \leq p_i} d_{ji} y_{ji}^* > x_i^*,$$

or equivalently by Lemma 1,

$$\pi_i z_i^* - \pi_i \sum_{j \in S : d_{ji} > \pi_i} y_{ji}^* - \sum_{j \in S : d_{ji} \leq \pi_i} d_{ji} y_{ji}^* > x_i^*.$$

Now, a continuous cover inequality for a fixed node $i \in V$ and $k \in S \cap T$ is violated if

$$\min\{\pi_i, d_{ki}\} z_i^* + \sum_{j \in T \backslash \{k\}} \phi_S(d_{ji})(z_i^* - y_{ji}^*) - \sum_{j \in S} \min\{\pi_i, d_{ji}\} y_{ji}^* > x_i^*.$$

Suppose $d_{ki} \geq \pi_i$, then the left term of the continuous cover inequality can be further written as

$$\pi_i z_i^* + \sum_{j \in N \backslash S} \phi_S(d_{ji})(z_i^* - y_{ji}^*) - \pi_i \sum_{j \in S : d_{ji} > \pi_i} y_{ji}^* - \sum_{j \in S : d_{ji} \leq \pi_i} d_{ji} y_{ji}^*.$$

Since $(z_i^* - y_{ji}^*) \geq 0$ holds and the lifting function $\phi_S$ is nonnegative, the left term of the continuous cover inequality clearly violates the current solution

$(x^*, y^*, z^*)$ when inequality (10) is violated. Otherwise, we need to compute $d_{ki} z_i^* + \sum_{j \in N \setminus S} \phi_S(d_{ji})(z_i^* - y_{ji}^*)$ to determine if it violates the current fractional solution. It takes $O(\hat{t})$ steps to compare $d_{ki}$ and $\pi_i$ for some $k \in S$ and for a fixed $i \in V$, hence, overall the complexity is $O(n\hat{t})$ to evaluate every node. In addition, the proof also suggests that $\pi_i < d_{ki}$ for $k \in S$ is necessary and sufficient to generate a violated continuous cover inequality.

**Corollary 1.** *Using the result of Theorem 3, the most violated continuous reverse cover inequality can be separated in $O(n\hat{t})$ time, if there exists any.*

## 3   Preliminary Computational Results

In this section, we report the preliminary computational results obtained by applying the aforementioned techniques on network instances from Fischetti et al. [6]. In particular, the data instances are generated based on directed small-world (SW) graphs [16], with node set $V \in \{50, 75, 100\}$ and average node degree $k \in \{4, 8, 12, 16\}$. The influence factor $d_{ij}$ for all $(i, j) \in A$ are generated uniformly randomly in $\{1, \cdots, 10\}$. For each node $i \in V$, the threshold $h_i = \max\{1, \min\{\eta_i, \sum_{j \in N_i} d_{ji}\}\}$, where $\eta_i$ is a random variable follows normal distribution with mean $0.7 \sum_{j \in N_i} d_{ji}$ and variance $\frac{\sum_{j \in N_i} d_{ji}}{|N_i|}$. The data instances are available at http://mario.ruthmair.at/wp-content/uploads/2020/04/socnet-instances-v2.zip. Here, we take five SW instances with $n = 50$, $m = 200$, where the average node degree is 4 and the connection probability between nodes is 0.1. We let $\tau$ be 0.1. The experiments are performed on a Quad-Core Intel Core i7 machine with 3.1 GHz and the memory limit is 16 GB. The computation time limit is set to 3600 s. The model and branch-and-cut algorithm are implemented in Python 3 with the Python-MIP package [1]. Gurobi 9.0.1 is used as the optimization solver. The minimum subset inequalities are separated and added to the branch-and-bound nodes dynamically, while the generalized cycle elimination constraints are implemented as lazy constraints. In Table 3, we report the final gap, number of user cuts and lazy constraints added, overall computational time, and time spent on the separation routine. Based on these small-scale computations, the results

**Table 3.** Computational results for SW-50-200 instances from [6].

| Instance # | User cuts | Lazy constraints | Overall time (sec) | SEP time (sec) | Gap |
|---|---|---|---|---|---|
| 1 | 3 | 1101 | 2.44 | 2.14 | 0 |
| 2 | 72 | 5734 | 29.21 | 21.31 | 0 |
| 3 | 29 | 174 | 34.65 | 23.54 | 0 |
| 4 | 195 | 6857 | 26.6 | 15.78 | 0 |
| 5 | 192 | 623 | 1.59 | 1.24 | 0 |

appear encouraging in the sense that the application of the proposed techniques allows one to find solutions with zero gap in a reasonable time. Thus, we believe that these approaches should be further addressed in larger-scale computational experiments.

## 4    Conclusion

We study the polyhedral structure of least cost influence maximization problem where the influence propagation is based on deterministic linear threshold model. In the process we exploit existing results on mixed 0-1 knapsack polyhedron and present a new class of valid inequalities for the influence propagation constraint in a single-node relaxation. We show that even for a small instance, these facet-defining inequalities are not sufficient to describe the convex hull. We propose an exact separation for the new valid inequalities and take advantage of the result to separate the inequalities proposed by [13]. The preliminary computations demonstrate the separation routine does not consume too much time in the experiments. Promising future research works include the development of a branch-and-cut algorithm that utilizes our proposed inequalities together with some pre-processing enhancements to reduce the computational burden on large social network instances.

## References

1. https://www.python-mip.com/
2. Ackerman, E., Ben-Zwi, O., Wolfovitz, G.: Combinatorial model and bounds for target set selection. Theor. Comput. Sci. **411**(44–46), 4017–4022 (2010)
3. Chen, N.: On the approximability of influence in social networks. SIAM J. Discret. Math. **23**(3), 1400–1415 (2009)
4. Chen, W., Lakshmanan, L.V., Castillo, C.: Information and influence propagation in social networks. Synth. Lect. Data Manag. **5**(4), 1–177 (2013)
5. Demaine, E.D., et al.: How to influence people with partial incentives. In: Proceedings of the 23rd International Conference on World Wide Web, pp. 937–948 (2014)
6. Fischetti, M., Kahr, M., Leitner, M., Monaci, M., Ruthmair, M.: Least cost influence propagation in (social) networks. Math. Program. **170**(1), 293–325 (2018)
7. Granovetter, M.: Threshold models of collective behavior. Am. J. Sociol. **83**(6), 1420–1443 (1978)
8. Günneç, D., Raghavan, S., Zhang, R.: A branch-and-cut approach for the least cost influence problem on social networks. Networks **76**(1), 84–105 (2020)
9. Günneç, D., Raghavan, S., Zhang, R.: Least-cost influence maximization on social networks. INFORMS J. Comput. **32**(2), 289–302 (2020)
10. Gursoy, F., Günneç, D.: Influence maximization in social networks under deterministic linear threshold model. Knowl.-Based Syst. **161**, 111–123 (2018)
11. Kempe, D., Kleinberg, J., Tardos, É.: Maximizing the spread of influence through a social network. In: Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 137–146 (2003)

12. Kempe, D., Kleinberg, J., Tardos, E.: Maximizing the spread of influence through a social network. Theory Comput. **11**(4), 105–147 (2015)
13. Marchand, H., Wolsey, L.A.: The 0–1 knapsack problem with a single continuous variable. Math. Program. **85**(1), 15–33 (1999)
14. Nannicini, G., Sartor, G., Traversi, E., Wolfler Calvo, R.: An exact algorithm for robust influence maximization. Math. Program. **183**(1), 419–453 (2020)
15. Raghavan, S., Zhang, R.: A branch-and-cut approach for the weighted target set selection problem on social networks. Inf. J. Optim. **1**(4), 304–322 (2019)
16. Watts, D.J., Strogatz, S.H.: Collective dynamics of 'small-world' networks. Nature **393**(6684), 440–442 (1998)
17. Wu, H.H., Küçükyavuz, S.: A two-stage stochastic programming approach for influence maximization in social networks. Comput. Optim. Appl. **69**(3), 563–595 (2018)