Available online at www.sciencedirect.com
ScienceDirect
IFAC PapersOnLine 50-1 (2017) 14465–14470

Minimizing the Influence Propagation in Social Networks for Linear Threshold Models

Lan Yang [*,**] **Alessandro Giua** [**,****] **Zhiwu Li** [***,*]

[*] *SEME, Xidian University, Xi'an 710071, China*
[**] *Aix Marseille Univ, Universite de Toulon, CNRS, ENSAM, LSIS, Marseille, France*
[***] *ISE, Macau University of Science and Technology, Taipa, Macau*
[****] *DIEE, University of Cagliari, Cagliari, Italy*
(e-mail: yanglan3009@gmail.com, alessandro.giua@univ-amu.fr, giua@diee.unica.it, zhwli@xidian.edu.cn).

Abstract: Innovation or information propagation in social networks has been widely studied in recent years. Most of the previous works are focused on solving the problem of influence maximization, which aims to identify a small subset of early adopters in a social network to maximize the influence propagation under a given diffusion model. In this paper, motivated by practical scenarios, we propose two different influence minimization problems. We consider a Linear Threshold diffusion model and provide a general solution to the first problem solving a linear integer programming. For the second problem, we provide a technique to search for an optimal solution that works only in particular cases and discuss a simple heuristic to find a solution in the general case.

Keywords: Social network, Optimization, Influence propagation, Linear Threshold model.

1. INTRODUCTION

With the rapid growth of information and communication technology during the last two decades, people are actively using networks for getting information, exchanging ideas, and even adopting new products. From a psychology perspective, it is well understood that an individual's idea or behavior is highly influenced by its neighbors or friends. Motivated by this, the study of influence propagation finds several applications in real-world life including viral marketing, the spread of rumors or memes, trust, the adoption of innovations in organizations, opinion dynamics, etc. In order to model the propagation of an idea or innovation through a network, Kempe et al. (2003) proposed two main diffusion models, namely the *Independent Cascade model* and the *Linear Threshold model*. They consist in directed graphs where each node can be either active (if it has adopted the innovation) or inactive (if it has not adopted the innovation). The innovation propagates in the network in a *progressive* fashion, i.e., nodes can only switch from inactive to active, but not in the opposite direction.

The *influence maximization* problem has a clear practical motivation in many applications, e.g., viral marketing (Domingos and Richardson, 2001). It aims to identify a small subset of initial adopters (seed set) in a social network to maximize the influence propagation under a given diffusion model and has been widely studied in the literature. Kempe et al. (2003) formalize this problem as an optimization problem, showing that it is NP-hard for both the Independent Cascade model and the Linear Threshold model, and present a greedy algorithm that can reach a good approximation of the optimal solution. Later, several improved algorithms (Leskovec et al., 2007; Chen et al., 2009, 2010; Goyal et al., 2011; Ramasuri and Narahari, 2011; Liu et al., 2014; Song et al., 2015) were presented to solve the influence maximization problem by balancing the running time and the influence spread of algorithms, trying to make them scalable to large datasets.

For the Linear Threshold model, Rosa and Giua (2013) provided a linear algebraic characterization of the set of final adopters corresponding to a given seed set. The set of final adopters — which is the complement of the maximal cohesive set not containing nodes in the seed set — can be computed solving an Integer Programming Problem (IPP). They also used this approach to solve some problems of influence maximization over a finite horizon. Unfortunately, the main drawback is that the number of decision variables is too large, being of order $n \times K$, where n is the size of the net and K is the length of the finite horizon.

In this paper we also consider a Linear Threshold diffusion model. Differently from the works in (Budak et al., 2011) and (He et al., 2012) which study the limitation of the misinformation and influence blocking maximization respectively under a competitive circumstance where both the good and bad information co-exist, we address problems related to *influence minimization*, that so far have

[*] This work was supported in part by the National Natural Science Foundation of China under Grant No. 61374068 and 61673309, and the Science and Technology Development Fund, MSAR, under Grant No. 078/2015/A3.

10.1016/j.ifacol.2017.08.2293

not received much attention. In particular we consider two scenarios that we believe have practical relevance in different application domains.

Scenario 1. Consider a company that must cut the supply to some of its customers since current demand exceeds its capacity. Cutting the supply will damage the reputation of the company and this bad reputation may propagate in the networks of customers. The company wants to determine a suitable set of customers, whose supply will be cut, so as to minimize this damage spread.

Scenario 2. Consider a hacker who wants to spread a virus to a set of servers V_{target} but can only directly infect servers in a given set V_{init}. The hacker aims to find a suitable subset of V_{init} such that starting from this seed the infection will propagate to all nodes in V_{target}. On the other hand, she aims to minimize the number of nodes affected that do not belong to the target set, to avoid attracting too much attention or creating unnecessary damage to the network.

We formalize two optimization problems that generalize these scenarios. Following the characterization of Rosa and Giua (2013), we provide a general solution to the first problem solving a linear IPP. For the second problem, we provide a technique to search for an optimal solution that works only in particular cases and discuss a simple heuristic to find a solution in the general case. In addition, for the influence minimization problems, we tested the proposed algorithms on two real-world datasets and compared their performance with respect to other simple heuristic approaches based on nodes' degrees and centrality that are commonly used in the literatures to estimate a node's influence.

The rest of this paper is organized as follows. Section 2 reviews the Linear Threshold model and its properties. Section 3 proposes two optimization problems related to influence minimization and their solutions. Section 4 presents a series of experimental results. Conclusions and directions for future research are discussed in Section 5.

2. BACKGROUND

2.1 Linear Threshold model

First, we introduce the Linear Threshold model to describe the diffusion of innovations in social networks. Table 1 lists the notation that will be used extensively in the rest of this paper.

Table 1. Notation Explanation

Notation	Description
$G = (V, E)$	A directed graph with node set V and edge set E
n	The size of G, i.e., the number of nodes in G
m	The number of directed edges in G
θ_i	Threshold value of node $i \in V$
N_i	The in-neighbor set of node $i \in V$
ϕ_0	The seed set
ϕ_t	The set of nodes that become active at step t
$\Phi_\infty(\phi_0)$	The set of final adopters given that ϕ_0 is the seed set
ϕ_0^*	The optimal seed set for influence minimization

Network structure We consider a directed graph $G = (V, E)$, where $V = \{1, 2, ..., n\}$ is the set of nodes involved in a network and E is the set of edges where $(i, j) \in E$. Besides, we assign a *threshold value* $\theta_i \in [0, 1]$ to each node i. The thresholds θ_i intuitively represent the different tendencies of nodes to adopt the innovation when their neighbors do (Kempe et al., 2003).

We define the *in-neighbor* set of node $i \in V$ as $N_i = \{j | (j, i) \in E\}$.

The *adjacency matrix* A is a square $n \times n$ matrix. Its element $A_{i,j} = 1$ if there is an edge from node i to node j, otherwise 0. No self-loops are allowed, i.e., $A_{i,i} = 0$.

We define the *in-neighbor scaled adjacency matrix* $\bar{A} \in [0, 1]^{n \times n}$ as follows:

$$\bar{A}_{i,j} = \frac{A_{i,j}}{|N_j|}$$

Let $\Theta = diag([\theta_1, \theta_2, \cdots, \theta_n])$ be the *threshold matrix* whose diagonal elements are the thresholds of the nodes and other elements are equal to 0.

Activation process Let ϕ_0 be the *seed set* which represents a set of agents (in this paper, we will use node, agent, and individual interchangeably) that are initially activated at step $t = 0$. The activation from the seed set propagates to the network step by step. We denote ϕ_t the set of nodes which are *activated* at step t. The set of nodes *active* at step t, i.e., those which are activated in the interval $[0, t]$, is denoted by

$$\Phi_t = \bigcup_{k=0}^{t} \phi_k.$$

By definition, we have $\Phi_0 = \phi_0$.

The activation process is described as follows. At each step $t = 1, 2, \ldots,$ a non active node i becomes active iff the fraction of its neighbors that are active at the previous step is at least θ_i, i.e.,

$$i \in \phi_t \iff \frac{|\Phi_{t-1} \cap N_i|}{|N_i|} \geq \theta_i \qquad (\forall i \in V \setminus \Phi_{t-1}). \tag{1}$$

The evolution proceeds until no more individuals adopt the innovation, and we define the set of *final adopters* as follows:

$$\Phi_\infty(\phi_0) = \bigcup_{i=0}^{\infty} \phi_i.$$

Note that we write $\Phi_\infty(\phi_0)$ to highlight that if the network structure $G = (V, E)$ and each node's threshold $\theta_i (i \in V)$ are given, the propagation process depends on the seed set ϕ_0.

2.2 Cohesiveness

Definition 2.1. (Acemoglu et al., 2011) A subset $X \subset V$ is called a *cohesive set* if for all $i \in X$ it holds:

$$\frac{|X \cap N_i|}{|N_i|} > 1 - \theta_i \tag{2}$$

◇

This states that a set $X \subset V$ is cohesive if $\forall i \in X$ the fraction of its neighbors which are still in set X is strictly larger than $1-\theta_i$, or equivalently say that the fraction of its neighbors that are not in set X is smaller than its specific threshold, i.e., $\frac{|V \setminus X \cap N_i|}{|N_i|} < \theta_i$.

Note that if $\phi_0 \cap X = \emptyset$, it holds that $\forall t \geq 0$, $\phi_t \cap X = \emptyset$, or in other words, if there is no individual in X adopting the innovation initially, then no individual in X will adopt the innovation at the following steps.

Lemma 2.1. (Acemoglu et al., 2011) Given a network G with seed set $\phi_0 \in V$, let $M \in V \setminus \phi_0$ be the *maximal cohesive set* contained in $V \setminus \phi_0$. The final adopter set is:

$$\Phi_\infty(\phi_0) = V \setminus M \tag{3}$$

◇

Lemma 2.1 gives an immediate way to compute the final adopters that does not require to determine the evolution of the network.

Definition 2.2. Given a set $X \in V$, the corresponding *characteristic vector* $\boldsymbol{x} \in \{0,1\}^n$ satisfies that $x_i = 1$ if $i \in X$, otherwise $x_i = 0$. ◇

Lemma 2.2. (Rosa and Giua, 2013) A set $X \subset V$ is cohesive if and only if its characteristic vector $\boldsymbol{x}$ for all $i \in X$ satisfies

$$\boldsymbol{x}^T \bar{A}(\cdot, i) \geq 1 - \bar{\theta}_i$$

where

$$\bar{\theta}_i = \begin{cases} \theta_i - \dfrac{1}{|N_i|}, & \text{if } \theta_i \cdot |N_i| \in \mathbb{N} \\ \theta_i, & \text{if } \theta_i \cdot |N_i| \notin \mathbb{N} \end{cases} \tag{4}$$

◇

According to Equation (4), we define the *updated threshold matrix* as $\bar{\Theta} = \text{diag}([\bar{\theta}_1, \bar{\theta}_2, \cdots, \bar{\theta}_n])$ whose diagonal elements are the newly defined thresholds of nodes $\bar{\theta}_i$ and other elements are equal to 0.

Proposition 2.1. (Rosa and Giua, 2013) Given a Linear Threshold model represented by a social graph $G = (V, E)$ with n nodes and thresholds value $\{\theta_i\}_{i \in V}$, let $\phi_0 \subset V$ be a seed set with characteristic vector $\boldsymbol{y}$. The maximal cohesive set M contained in $V \setminus \phi_0$ has a characteristic vector $\boldsymbol{x}^*$ that is the solution of the following IPP:

$$\begin{aligned} \max\ & \mathbf{1}^T \cdot \boldsymbol{x} && (IPP-1) \\ \text{s.t. } & \mathbf{1} - \boldsymbol{x} \geq \boldsymbol{y} \\ & [I - \bar{\Theta} - \bar{A}^T] \cdot \boldsymbol{x} \leq 0 \\ & \boldsymbol{x} \in \{0,1\}^n. \end{aligned}$$

Hence, the set of final adopters is $\Phi_\infty(\phi_0) = \{i \in V | x_i^* = 0\}$. ◇

3. INFLUENCE MINIMIZATION PROBLEMS AND SOLUTIONS

In this section, we introduce two different influence minimization problems generalizing the two scenarios discussed in the introduction and discuss approaches for their solution. Note that all proofs of our results are omitted for the sake of space.

3.1 Scenario 1

In the first scenario discussed in the introduction, a company must cancel the supply offered to some of its customers because the current demand exceeds its capacity. We consider a *supply vector* $\boldsymbol{w} = [w_1, w_2, \cdots, w_n] \in \mathbb{R}_+^n$, where w_i denotes the supply customer $i \in V$ receives. The total supply must be reduced of a quantity $k \in \mathbb{R}_+$.

Cutting the supply of a customer will lower the consideration of the company and this bad reputation may propagate in the networks of customers. We consider a *damage vector* $\boldsymbol{c} = [c_1, c_2, \cdots, c_n] \in \mathbb{R}_+^n$ where c_i denotes the reputation damage due to customer $i \in V$ having a low consideration of the company. The company wants to determine a suitable set of customers, whose supply will be cut, so as to minimize this damage spread. We formalize this problem as follows.

Problem 3.1. Given a diffusion model represented by a social graph $G = (V, E)$ with n nodes, let $\boldsymbol{c} \in \mathbb{R}_+^n$ be a cost vector, $\boldsymbol{w} \in \mathbb{R}_+^n$ a weight vector and $k \in \mathbb{R}_+$ a constant. Find a seed set ϕ_0 whose total weight is at least k, such that the total cost of the corresponding set of final adopters $\Phi_\infty(\phi_0)$ is minimized, i.e.,

$$\begin{aligned} \min\ & \sum_{i \in \Phi_\infty(\phi_0)} c_i \\ \text{s.t. } & \sum_{i \in \phi_0} w_i \geq k. \end{aligned}$$

◇

Optimal solution Based on Proposition 2.1, we present a solution to Problem 3.1 under the Linear Threshold model.

Proposition 3.1. Given a Linear Threshold model represented by a social graph $G = (V, E)$ with n nodes and thresholds value $\{\theta_i\}_{i \in V}$, let $\boldsymbol{c} \in \mathbb{R}_+^n$ be a cost vector, $\boldsymbol{w} \in \mathbb{R}_+^n$ a weight vector and $k \in \mathbb{R}_+$ a constant. Consider the following IPP with binary variables $\boldsymbol{x}$ and $\boldsymbol{y}$:

$$\begin{aligned} \max\ & \boldsymbol{c}^T \cdot \boldsymbol{x} && (IPP-2) \\ \text{s.t. } & \mathbf{1} - \boldsymbol{x} \geq \boldsymbol{y} && (a) \\ & [I - \bar{\Theta} - \bar{A}^T] \cdot \boldsymbol{x} \leq 0 && (b) \\ & \boldsymbol{w}^T \cdot \boldsymbol{y} \geq k && (c) \\ & \boldsymbol{x}, \boldsymbol{y} \in \{0,1\}^n \end{aligned}$$

and let $\boldsymbol{x}^*, \boldsymbol{y}^* \in \{0,1\}^n$ be the optimal solution of (IPP-2). Then the seed set $\phi_0^* = \{i \in V | \ y_i^* = 1\}$ is the optimal solution of Problem 3.1 and the corresponding set of final adopters is $\Phi_\infty(\phi_0^*) = \{i \in V | x_i^* = 0\}$. ◇

3.2 Scenario 2

In the second scenario discussed in the introduction, a hacker is targeting a set of nodes V_{target} but can only directly infect nodes in the set V_{init}. She wants to determine a suitable seed set such that the infection will reach all target nodes but the number of non-target nodes infected is minimal. We can formalize this problem as follows.

Problem 3.2. Given a diffusion model represented by a social graph $G = (V, E)$, let sets $V_{target} \subseteq V$ and $V_{init} \subseteq V$ be assigned. Find a seed set ϕ_0 contained in V_{init} such that the set of final adopters $\Phi_\infty(\phi_0)$ contains all nodes in V_{target} and has minimal cardinality, i.e.,

$$\begin{aligned} &\min |\Phi_\infty(\phi_0)| \\ &\text{s.t. } \Phi_\infty(\phi_0) \supseteq V_{target} \quad (a) \\ &\quad \phi_0 \subseteq V_{init}. \quad (b) \end{aligned}$$

◇

In the rest of this section we study this problem in the framework of the Linear Threshold model.

A necessary and sufficient condition for the existence of a solution. A first observation is that Problem 3.2 may have no admissible solution. The following is a necessary and sufficient condition for the existence of a solution.

Proposition 3.2. Given a Linear Threshold model represented by a social graph $G = (V, E)$ with n nodes and thresholds value $\{\theta_i\}_{i\in V}$, let sets $V_{target} \subseteq V$ and $V_{init} \subseteq V$ with characteristic vectors $\boldsymbol{v}_{init}$ and $\boldsymbol{v}_{target}$ be assigned. Consider the following IPP with binary variable $\boldsymbol{x}$:

$$\begin{aligned} &\max \mathbf{1}^T \cdot \boldsymbol{x} \qquad (IPP-3) \\ &\text{s.t. } \mathbf{1} - \boldsymbol{x} \geq \boldsymbol{v}_{init} \\ &\quad [I - \bar{\Theta} - \bar{A}^T] \cdot \boldsymbol{x} \leq 0 \\ &\quad \boldsymbol{x} \in \{0, 1\}^n \end{aligned}$$

Problem 3.2 has solution if and only if (IPP-3) has an optimal solution $\boldsymbol{x}^* \leq \mathbf{1} - \boldsymbol{v}_{target}$. ◇

A sufficient condition for optimality In the following we provide a sufficient condition for an optimal solution of Problem 3.2.

Proposition 3.3. Given a Linear Threshold model represented by a social graph $G = (V, E)$ with n nodes and thresholds value $\{\theta_i\}_{i\in V}$, let sets $V_{target} \subseteq V$ and $V_{init} \subseteq V$ with characteristic vectors $\boldsymbol{v}_{init}$ and $\boldsymbol{v}_{target}$ be assigned. Consider the following IPP with binary variable $\boldsymbol{x}$:

$$\begin{aligned} &\max \mathbf{1}^T \cdot \boldsymbol{x} \qquad (IPP-4) \\ &\text{s.t. } \mathbf{1} - \boldsymbol{x} \geq \boldsymbol{v}_{target} \\ &\quad [I - \bar{\Theta} - \bar{A}^T] \cdot \boldsymbol{x} \leq 0 \\ &\quad \boldsymbol{x} \in \{0, 1\}^n \end{aligned}$$

Let $\bar{\boldsymbol{x}}^*$ be its optimal solution and define $\boldsymbol{z} = \max\{\boldsymbol{v}_{init} - \bar{\boldsymbol{x}}^*, \mathbf{0}\}$. If the following IPP,

$$\begin{aligned} &\max \mathbf{1}^T \cdot \boldsymbol{x} \qquad (IPP-5) \\ &\text{s.t. } \mathbf{1} - \boldsymbol{x} \geq \boldsymbol{z} \\ &\quad [I - \bar{\Theta} - \bar{A}^T] \cdot \boldsymbol{x} \leq 0 \\ &\quad \boldsymbol{x} \in \{0, 1\}^n \end{aligned}$$

has optimal solution $\boldsymbol{x}^* = \bar{\boldsymbol{x}}^*$, then the seed set $\phi_0^* = \{i \in V | z_i = 1\}$ is an optimal solution of Problem 3.2. ◇

A greedy algorithm The algebraic approach discussed above for finding an optimal solution works only in particular cases. Instead, in general case in order to obtain a reasonably efficient solution to Problem 3.2, we propose a greedy algorithm described in Algorithm 1.

We start with a seed set $\phi_0 = V_{min} = V_{target} \cap V_{init}$ (line 3). If the set of final adopters corresponding to V_{min} contains all nodes in V_{target}, we stop and output ϕ_0. Otherwise, at each iteration, we select a node u from $V_{complete} = V_{init} \setminus V_{min}$ and add it to the seed set. To choose a suitable node u, we select from $V_{complete}$ all nodes

Algorithm 1 Greedy algorithm

1: **Input:** A social graph $G = (V, E)$, sets $V_{init} \subseteq V$ and $V_{target} \subseteq V$ with $\Phi_\infty(V_{init}) \supseteq V_{target}$.
2: **Output:** Seed set ϕ_0.
3: Let $V_{min} = V_{target} \cap V_{init}$.
4: Let $V_{complete} = V_{init} \setminus V_{min}$.
5: Let $\phi_0 = V_{min}$.
6: **while** $\Phi_\infty(\phi_0) \not\supseteq V_{target}$ **do**
7: $\quad S = \underset{v\in V_{complete}}{argmax} \; |\Phi_\infty(\phi_0 \cup \{v\}) \cap V_{target}|$.
8: $\quad u = \underset{i\in S}{argmin} \, |\Phi_\infty(\phi_0 \cup \{i\})|$.
9: $\quad \phi_0 \leftarrow \phi_0 \cup \{u\}$.
10: $\quad V_{complete} \leftarrow V_{complete} \setminus \{u\}$.
11: **end while**

which added to the seed set will give a maximal marginal gain in reaching the target nodes (line 7) and call this new set S. Then we choose from S node u with minimal cardinality of the set of final adopters (line 8) and add it to the seed set ϕ_0. This is repeated until the current seed set can influence all nodes in V_{target}: in this case we stop and output ϕ_0.

The time complexity of Algorithm 1 is $O(l^2 m)$ with $l = |V_{complete}|$ and $m = |E|$, since the algorithm has at most l iterations (the worst case is that when we go through all nodes in $V_{complete}$ we can reach all target nodes), in each iteration we also need to go through all nodes $w \in V_{complete}$ to find the ones with maximal marginal gain in reaching the target nodes, for each node w we need to compute its corresponding final adopters which may require a traversal of all graph edges.

A remark on Algorithm 1 is that a choice of the starting seed set $V_{min} = Z = V_{init} \cap \Phi_\infty(V_{target})$ would result in a smaller number of repetitions of the while-loop, since this starting seed set is larger than the one described in line 3 of Algorithm 1, but it requires solving (IPP-4) firstly.

4. EXPERIMENTS AND ANALYSIS

4.1 Experimental set up

We performed a series of experiments based on two real-world networks of different scales under the Linear Threshold model. NetScience and Hep-Th are from different sections of the e-print arXiv, which is the same data source used in the literatures of influence maximization. Empirical evidence suggests that coauthorship graphs are representative of typical social networks (Newman, 2006, 2001).

Here are some details about the two networks; precise statistical information of the datasets is summarized in Table 2.

NetScience: This is the co-authorship network of scientists working on network theory and experiment, as compiled by Newman (2006).

Hep-Th: This is the collaboration network of scientists posting preprints on the high-energy theory archive at www.arxiv.org, 1995-1999, as compiled by Newman (2001).

Table 2. Statistics of real datasets

Dataset	Nodes	Edges	Average degree
NetScience	1,589	2,742	3.451
Hep-Th	8,361	15,751	3.768

Our experiment unfolds in two parts. For Problem 3.1, we compare our ILP method with other heuristics. Two of these heuristics, in particular, are based on node's degrees and centrality within the network and are commonly used in the literatures studying the influence maximization problem. While in a influence maximization problem it is common to select the seed set among nodes with highest degree/centrality, since we are solving an influence minimization problem we make the opposite choice. The tested algorithms are briefly described in the following.

Random: Randomly selects a set of k nodes from the network as the seed set.
Low-Degree: A simple heuristic approach that selects a set of k nodes with the lowest degrees in the graph.
Low-PageRank: PageRank algorithm is a popular algorithm for ranking web pages (Brin and Page, 2012). We select a set of k nodes with the lowest pagerank value as the seed set. In this algorithm, we use the power method to compute every pagerank value and the restart probability of pagerank is set as 0.15. The stopping criteria is when two consecutive iterations differ for at most 10^{-5}.
Integer Linear Programming (ILP): Our integer linear programming method presented in Proposition 3.1.

For Problem 3.2, since the greedy algorithm may take several hours to run on large datasets, we tested our method on Algorithm 1 under the Linear Threshold model on dataset NetScience.

All experiments are performed using MATLAB on a 1.20 GHz Genuine Intel and 4G memory.

4.2 Experiment result for Problem 3.1

We tested the algorithms on the two networks under the Linear Threshold model in terms of the influence spread and the running time. For convenience, we consider a cost vector $\boldsymbol{c} = \mathbf{1}$ and a weight vector $\boldsymbol{w} = \mathbf{1}$. In order to have a good estimate of the efficiency of the algorithms, we simulated the process 100 times by re-choosing thresholds randomly (standard uniform distribution) from $[0, 1]$ every time and computed the average number of the final adopters. Furthermore, we select from 1 seed to 500 and 2,000 seeds for each dataset respectively. Figs. 1, and 2 show the average influence spread of the tested algorithms on datasets NetScience and Hep-Th. The running time for selecting each largest size seed set (500 for NetScience and 2,000 for Hep-Th) is illustrated in Table 3.

From the influence spread on the two datasets, we can make the following observations: (1) As expected, randomly choosing the seed set is not a good idea. Sometimes when the seed set size is large, the bad influence may propagate and cover most of the network; (2) Our integer linear programming method (ILP) outperforms the heuristics relying solely on the structure properties of the network. On network NetScience, ILP outperforms Low-Degree and Low-PageRank by 17.2% and 40.2% respectively. On network Hep-Th, ILP outperforms Low-Degree and Low-PageRank by 38.5% and 48.2% respectively; (3) Our ILP successfully selects the optimal seed set, and when the seed size is not large enough the bad influence always just exists in the seed set and will not propagate to others through the network.

However, the running time of ILP is much longer than the others when the size of the network is large since solving a large scale IPP is hard; the result of the largest seed sets is shown in Table 3.

Table 3. Running time of the tested algorithms (for largest seed set on each dataset)

Dataset	Running time (s)			
	Random	Low-Degree	Low-PageRank	ILP
NetScience	8	2.7	6	21
Hep-Th	74	112	100	4,920

4.3 Experiment result for Problem 3.2

In this part, in order to have a good estimate of the performance of the greedy algorithm, we run the process 100 times taking the average number of final adopters and re-choose the sets V_{init} (of size k_1 which is a constant) and V_{target} (of size k_2 which is a constant) randomly at each time. In order to guarantee that Problem 3.2 has a solution and obtain a feasible seed set $\phi_0 \subseteq V_{init}$ such that $\Phi_\infty(\phi_0) \supseteq V_{target}$, we randomly choose a set of nodes of size k_1 in V as V_{init} and compute $\Phi_\infty(V_{init})$, and then we randomly choose a set of nodes of size k_2 in $\Phi_\infty(V_{init})$ as V_{target}.

We introduce a coefficient ρ^{pro} which illustrates the percent of nodes that avoid being activated by our strategy but will be activated if the seed set is V_{init}.

$$\rho^{pro} = \frac{||\Phi_\infty(V_{init})| - |\Phi_\infty(\phi_0)||}{||\Phi_\infty(V_{init})| - |V_{target}||} \times 100\%$$

where $|\Phi_\infty(\phi_0)|$ is the number of final adopters activated by seed set ϕ_0 chosen by Algorithm 1.

The experiment result is listed in Table 4. In this table, $(|V_{init}|, |V_{target}|)$ is a number pair about the cardinality of sets V_{init} and V_{target}. We test three pairs for dataset NetScience. ρ^{pro}_{max} is the maximal percent (over all simulations) of nodes which avoids being activated but will be activated if the seed set is V_{init}.

Table 4. Experiment result

Dataset	$(\lvert V_{init}\rvert, \lvert V_{target}\rvert)$	ρ^{pro}	ρ^{pro}_{max}
NetScience	(10,10)	27.27%	100%
	(20,30)	21.57%	55.17%
	(30,50)	21.67%	45.00%

We can observe from Table 4 that our strategy can protect some percent of nodes from being activated. For example, when $(|V_{init}|, |V_{target}|) = (10, 10)$, 27.27% nodes avoid being activated. Besides, among all the simulations, the maximal percent of nodes avoiding being activated is 100%, i.e., the selected seed set just reaches nodes in V_{target} at the end of the process and no additional node is activated.

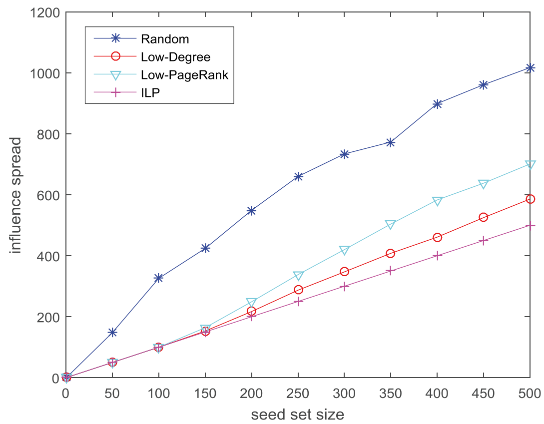

Fig. 1. Influence spread of the tested algorithms on NetScience.

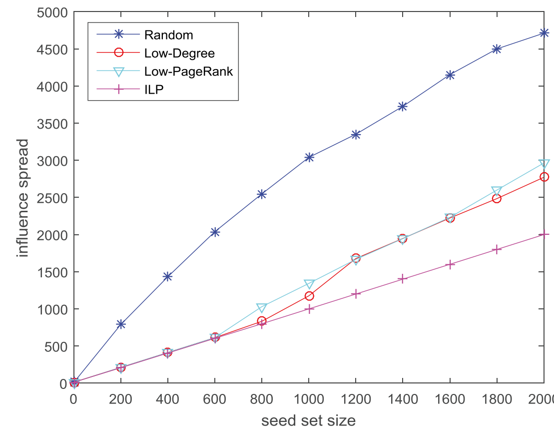

Fig. 2. Influence spread of the tested algorithms on Hep-Th.

5. CONCLUSIONS AND FUTURE WORK

In this paper, we formalize two different influence minimization problems generalizing real-world scenarios. For Problem 3.1, following the characterization of Rosa and Giua (2013), we provide a general solution solving a linear IPP and show it outperforms other standard heuristics. For Problem 3.2, we provide a technique to search for an optimal solution that works only in particular cases and discuss a simple heuristic to find a solution (possibly suboptimal) in the general case.

We are aware that IPP based solutions do not scale well for large graph: for this reason finding a technique to relax IPP to linear programming will be the objective of our future work. Furthermore, another interesting line of research could be to study the influence minimization problems in other models of diffusion, such as the Independent Cascade model.

REFERENCES

Acemoglu, D., Ozdaglar, A., and Yildiz, E. (2011). Diffusion of innovations in social networks. In *50th IEEE Conference on Decision and Control and European Control Conference*, 2329–2334. IEEE.

Brin, S. and Page, L. (2012). Reprint of: The anatomy of a large-scale hypertextual web search engine. *Computer networks*, 56(18), 3825–3833.

Budak, C., Agrawal, D., and Abbadi, A. (2011). Limiting the spread of misinformation in social networks. In *Proceedings of the 20th international conference on World wide web*, 665–674. ACM.

Chen, W., Wang, Y., and Yang, S. (2009). Efficient influence maximization in social networks. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, 199–208. ACM.

Chen, W., Yuan, Y., and Zhang, L. (2010). Scalable influence maximization in social networks under the linear threshold model. In *Data Mining (ICDM), 2010 IEEE 10th International Conference on*, 88–97. IEEE.

Domingos, P. and Richardson, M. (2001). Mining the network value of customers. In *Proceedings of the 7th ACM SIGKDD international conference on Knowledge discovery and data mining*, 57–66. ACM.

Goyal, A., LU, W., and Lakshmanan, L. (2011). Simpath: An efficient algorithm for influence maximization under the linear threshold model. In *11th IEEE International Conference on Data Mining*, 211–220. IEEE.

He, X., Song, G., Chen, W., and Jiang, Q. (2012). Influence blocking maximization in social networks under the competitive linear threshold model. In *Proceedings of the 2012 SIAM International Conference on Data Mining*, 463–474. SIAM.

Kempe, D., Kleinberg, J., and Tardos, E. (2003). Maximizing the spread of influence through a social network. In *Proceedings of the 9th ACM SIGKDD international conference on Knowledge discovery and data mining*, 137–146. ACM.

Leskovec, J., Krause, A., Guestrin, C., Faloutsos, C., VanBriesen, J., and Glance, N. (2007). Cost-effective outbreak detection in networks. In *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*, 420–429. ACM.

Liu, X., Li, M., Li, S., Peng, S., Liao, X., and Lu, X. (2014). Imgpu: Gpu-accelerated influence maximization in large-scale social networks. *IEEE Transactions on Parallel and Distributed Systems*, 25(1), 136–145.

Newman, M. (2001). The structure of scientific collaboration networks. *Proceedings of the National Academy of Sciences*, 98(2), 404–409.

Newman, M. (2006). Finding community structure in networks using the eigenvectors of matrices. *Physical review E*, 74(3), 036104.

Ramasuri, N. and Narahari, Y. (2011). A shapley value-based approach to discover influential nodes in social networks. *IEEE Transactions on Automation Science and Engineering*, 8(1), 130–147.

Rosa, D. and Giua, A. (2013). On the spread of innovation in social networks. *4th IFAC Workshop on Distributed Estimation and Control in Networked Systems*, 46(27), 322–327.

Song, G., X.B. Zhou, Y.W., and Xie, K. (2015). Influence maximization on large-scale mobile social network: a divide-and-conquer method. *IEEE Transactions on Parallel and Distributed Systems*, 26(5), 1379–1392.