Influence Minimization in Linear Threshold Networks ⋆

Lan Yang [a,d] Zhiwu Li [a,b] Alessandro Giua [c]

[a] SEME, Xidian University, Xi'an, China

[b] ISE, Macau University of Science and Technology, Taipa, Macau

[c] DIEE, University of Cagliari, Cagliari, Italy

[d] Aix Marseille University, Universite de Toulon, CNRS, LIS, Marseille, France

Abstract

The propagation of information and innovation in social networks has been widely studied in recent years. Most of the previous works focus 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. On the contrary in this paper, motivated by real-world scenarios, we propose two different influence minimization problems. We consider a linear threshold diffusion model and provide a general solution to the first problem by solving an integer linear programming problem. 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. Several simulations on real data sets are also presented.

Key words: Social networks, Optimization, Influence propagation, linear threshold model.

Published as:
L. Yang, Z.W. Li, A. Giua, "Influence Minimization in Linear Threshold Networks," *Automatica*, Vol. 100, pp. 10C16, 2019. DOI: 10.1016/j.automatica.2018.10.053

⋆ This work was supported partially by the National Natural Science Foundation of China under Grant Nos. 61673309 and 61873342, the Science and Technology Development Fund, MSAR, under Grant No. 122/2017/A3, and the Project PHC CAI YUANPEI 2017 from the Chinese Scholarship Council under Project No. 38908TK.
(Corresponding author: Zhiwu Li. Tel.: +86-29-8820-1986. Fax: +86-29-8820-2456.)

Email addresses: yanglan3009@gmail.com (Lan Yang), zhwli@xidian.edu.cn (Zhiwu Li), giua@diee.unica.it (Alessandro Giua).

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 psychological perspective, it is well understood that an individual's 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 or the adoption of innovations in organizations [1, 2, 3, 8, 16, 21], the spread of rumors or memes, trust [9], opinion dynamics [5, 14], etc.

In order to model the propagation of an idea or innovation through a network, Kempe et al. [8] proposed a *threshold model*. It consists of a directed graph where each node can be either active (if it has adopted the innovation) or inactive (if it has not adopted the innovation). A node adopts the innovation when the number of its active neighbors is beyond a given threshold. The innovation propagates in a *progressive* fashion, i.e., nodes can only switch from inactive to active but not in the opposite direction, until a set of *final adopters* is determined.

Threshold models are dynamical systems whose evolution consists in a transient phase that eventually reaches a steady state in a finite number of steps and are suited to describe collective behaviors [4]. The *linear* threshold model that we adopt in this paper provides a simplified description which is sufficiently accurate and relatively easy to analyze. We refer to [17] for a non-progressive version of this diffusion model and to [21] for a more general *multiplex* threshold network. In [13] linear threshold networks with node dynamics are considered.

The influence maximization problem in social networks aims to identify a small subset of initial adopters (called *controlled* nodes, *driver* nodes, *leaders*, or *seed* nodes) to maximize the influence propagation under a given diffusion model and has been widely studied in the literature [8, 15, 18]. The choice of the seed set can be seen as a control action that determines the network performance and has a clear practical motivation in many applications [2]. A similar approach based on node selection has also been studied from a different perspective in other applications. We mention the problem of controlling the state of leader nodes to minimize the information propagation error caused by noise on the communication links between nodes in leader-follower systems [10, 19].

Rosa and Giua [16] provided a linear algebraic characterization of the set of final adopters corresponding to a given seed set in a linear threshold model. The set of final adopters can be computed by solving an Integer Linear Programming (ILP) problem. They also used this approach to solve the influence *maximization* over a finite horizon k, but the proposed approach is rather inefficient as it requires a number of integer decision variables up to $2 \times n \times k$, where n is the number of nodes.

Just as good things (e.g., innovation, positive information, etc.) spread through a network and bring positive effects to individuals, undesirable things (e.g., bad reputation, failures, etc.) may also spread in a similar fashion. For this reason, in this paper we consider a linear threshold model and address problems related to *influence minimization*, which so far have not received much attention. We show that an ILP characterization of the steady state can be adapted to solve two meaningful classes of influence minimization problems where the control action is the choice of the seed set. This was not the case for influence maximization problems, where maximizing the cardinality of the set of final adopters (which we show is equal to minimizing a maximal cohesive set) results in a min-max integer programming of difficult solution.

A problem of this type was addressed in [9] where a contamination minimization problem was defined and solved by blocking some links in the network. However, in many realistic situations such a solution is not viable because it may be impossible for a control agent to interfere with a network structure.

We formalize two general problems called *Loss Minimization with Disruption* (LMD) and *Diffusion Minimization with Guaranteed Target* (DMGT). The LMD represents a family of influence minimization problems with a linear constraint on the seed set. On the contrary, the DMGT represents a family of influence minimization problems with a guaranteed target and restriction of the seed set to a subset of nodes. Note that both problems have a combinatorial nature and finding the optimal solution for either of them requires an exhaustive search in general.

Our main contributions are the following:

- For the LMD problem, we first provide an optimal solution by solving an integer linear programming problem and then present two heuristics to find approximate solutions based on nodes' degrees and centralities that are commonly used in the literature to estimate a node's influence.
- For the DMGT 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.
- We test the proposed algorithms on two real-world datasets with different scale and compare their performance with other heuristic approaches.

A preliminary version of this work was presented in [20]. This paper is a substantially different version where the significance of the two influence minimization problems and the challenges in solving them are clarified; the proofs of all results are given; heuristics are generalized to the case of heterogenous nodes and the issues of their convergence, scalability and performance are discussed.

2 Preliminaries

2.1 Linear threshold network

A *linear threshold network* N is a 4-tuple (V, E, θ, W) where $V = \{1, 2, ..., n\}$ is the set of nodes in the network and $E \subseteq V \times V$ is a set of directed arcs, i.e, $(i, j) \in E$ when there is an arc from node i to j. Function $\theta : V \to (0, 1]$ is a mapping that assigns a *threshold value* $\theta_i \in (0, 1]$ to each node $i \in V$. Function $W : V \times V \to (0, 1]$ is a mapping that assigns an influence weight $W(i, j) \in (0, 1]$ to each arc $(i, j) \in E$ such that $W(i, j) = 0$ if $(i, j) \notin E$ and $\sum_{i \in V} W(i, j) = 1$ for all $j \in V$. We also call this function the *weighted adjacency matrix* $W \in (0, 1]^{V \times V}$.

Each node in a network represents an individual or agent (therefore in this paper, we will use node, agent, and individual interchangeably) and the thresholds θ_i represent the different tendencies of nodes to adopt the innovation when their neighbors do [8]. We define the *in-neighbor set* of node $i \in V$ as $N_i^{in} = \{j | (j, i) \in E\}$ and the *out-neighbor set* as $N_i^{out} = \{j | (i, j) \in E\}$. Arc (i, j) in a network denotes that node j can be influenced by node i.

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

Let ϕ_0 be the *seed set* which represents a set of agents that are initially activated at step $t = 0$. The activation from the seed set propagates in 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 that have been activated at step t or at an earlier one, is denoted by $\Phi_t = \bigcup_{k=0}^{t} \phi_k$. By definition, we have $\Phi_0 = \phi_0$.

At each step $t = 1, 2, \ldots$, an inactive node i becomes active if the total influence weight of its neighbors active at step $t - 1$ is at least θ_i, i.e.,

$$i \in \phi_t \Longleftrightarrow \sum_{j \in N_i^{in} \cap \Phi_{t-1}} W(j, i) \geq \theta_i \quad (\forall i \in V \setminus \Phi_{t-1}) . \tag{1}$$

The evolution propagates until no more individuals adopt the innovation and the network reaches a steady state. Then we define the set of *final adopters* as $\Phi_\infty(\phi_0) = \bigcup_{i=0}^{\infty} \phi_i$.

2.2 Cohesiveness

In this part, we present a linear characterization of the set of final adopters based on cohesiveness.

Definition 2.1 *[1] A subset $X \subseteq V$ is called a* cohesive set *if for all $i \in X$ it holds:*

$$\sum_{j \in X \cap N_i^{in}} W(j,i) > 1 - \theta_i \tag{2}$$

Note that for a cohesive set X, if $\phi_0 \cap X = \emptyset$, then $\forall t \geq 0$, $\phi_t \cap X = \emptyset$. In other words, if no individual in X belongs to the seed set, then no individual in X will adopt the innovation at the following steps.

Lemma 2.1 *[1] Given a linear threshold network $N = (V, E, \theta, W)$ with seed set $\phi_0 \subseteq V$, let $M \subseteq V \setminus \phi_0$ 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 a direct way to compute the set of final adopters that does not require to determine the evolution of the network.

Definition 2.2 *Given a set $X \subset V$, its* characteristic vector *$\boldsymbol{x} \in \{0,1\}^n$ is such that $x_i = 1$ if $i \in X$, otherwise $x_i = 0$.*

Lemma 2.2 *[16] A set $X \subset V$ is cohesive if and only if its characteristic vector $\boldsymbol{x}$ satisfies*

$$\boldsymbol{x}^T W(\cdot, i) > 1 - \theta_i \quad (\forall i \in X). \tag{4}$$

where $W(\cdot, i)$ is the i-th column of the weighted adjacency matrix.

Proposition 2.1 *[16] Given a linear threshold network $N = (V, E, \theta, W)$ with n nodes, let $\phi_0 \subseteq V$ 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 ILP:*

$$\begin{aligned} \max_{\boldsymbol{x}} \quad & \boldsymbol{1}^T \cdot \boldsymbol{x} \qquad\qquad (ILP-1) \\ s.t. \quad & \boldsymbol{1} - \boldsymbol{x} \geq \boldsymbol{y} \\ & [I - \Theta - W^T] \cdot \boldsymbol{x} < 0 \\ & \boldsymbol{x} \in \{0,1\}^n. \end{aligned}$$

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

Note that operators $<$ and $\geq$ are intended componentwise for vectors.

3 Loss Minimization with Disruption

3.1 Problem statement

In this section, we introduce a class of influence minimization problems called *Loss Minimization with Disruption*. The idea is that the propagation of an initial perturbation has a cost. Given a *cost vector* $\boldsymbol{c} = [c_1, c_2, \cdots, c_n] \in \mathbb{R}_+^n$, where c_i denotes the cost of node $i \in V$, one aims to minimize the total cost of active nodes in the steady state. The problem is meaningful if the initial perturbation (nodes initially activated) is lower bounded: for this reason we introduce a *weight vector* $\boldsymbol{s} = [s_1, s_2, \cdots, s_n] \in \mathbb{R}_+^n$, where s_i denotes the weight of node $i \in V$ and assume that the total weight of nodes in the seed set must be greater than or equal to a pre-assigned value.

Problem 3.1 (LMD: Loss Minimization with Disruption) *Given a diffusion model represented by a network with n nodes, let $\boldsymbol{c} \in \mathbb{R}^n_+$ be a cost vector, $\boldsymbol{s} \in \mathbb{R}^n_+$ a weight vector, $k \in \mathbb{R}_+$ a lower bound on the seed set. Find a seed set ϕ_0 whose total weight is at least k, such that the total cost of the set of final adopters $\Phi_\infty(\phi_0)$ is minimized, i.e.,*

$$\begin{aligned} \min_{\phi_0} \quad & f_c(\phi_0) \stackrel{\text{def}}{=} \sum_{i \in \Phi_\infty(\phi_0)} c_i \\ s.t. \quad & \sum_{i \in \phi_0} s_i \geq k \qquad\qquad (a) \end{aligned}$$

As an example of a problem that can be framed as a LMD, consider a company that must cut the total supply to its customers of a given amount k because the current demand exceeds its production capacity. Cutting the supply s_i of customer i will produce a discontent which will 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 discontent spread. In this scenario, the weight of each node represents the supply and the cost represents the damage due to discontent.

3.2 Optimal solution

Motivated by Proposition 2.1, we address the LMD problem for linear threshold networks.

Proposition 3.1 *Given a linear threshold network $N = (V, E, \theta, W)$ with n nodes, let $\boldsymbol{c} \in \mathbb{R}^n_+$ be a cost vector, $\boldsymbol{s} \in \mathbb{R}^n_+$ a weight vector and $k \in \mathbb{R}_+$ a constant. Consider the following ILP with binary variables $\boldsymbol{x}$ and $\boldsymbol{y}$:*

$$\begin{aligned} \max_{\boldsymbol{x}} \quad & \boldsymbol{c}^T \cdot \boldsymbol{x} & (ILP-2) \\ s.t. \quad & \boldsymbol{1} - \boldsymbol{x} \geq \boldsymbol{y} & (a) \\ & [I - \Theta - W^T] \cdot \boldsymbol{x} < 0 & (b) \\ & \boldsymbol{s}^T \cdot \boldsymbol{y} \geq k & (c) \\ & \boldsymbol{x}, \boldsymbol{y} \in \{0,1\}^n \end{aligned}$$

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

PROOF. First we observe that constraint (c) ensures that a set ϕ_0 whose characteristic vector is $\boldsymbol{y}$ is an admissible solution of the LMD problem. In addition (thanks to Proposition 2.1 and the fact that $\boldsymbol{c} > \boldsymbol{0}$) constraints (a) and (b) ensure that the set M with characteristic vector $\boldsymbol{x}$ is the maximal cohesive set contained in $V \setminus \phi_0$, hence $\Phi_\infty(\phi_0)$ has characteristic vector $\boldsymbol{1} - \boldsymbol{x}$. Finally, the objective function of (ILP-2) ensures that the set ϕ_0^* with characteristic vector $\boldsymbol{y}^*$ is an optimal solution of the LMD problem. □

The number of decision variables in (ILP-2) is $2n$. This makes it computationally hard to solve when the size of the network becomes large enough.

3.3 Heuristic solutions

To reduce the computation burden of solving (ILP-2), we propose in this section two different heuristics that provide approximate solutions to the LMD problem. In particular, they are based on nodes' degrees and centralities within the network and are commonly used in the literature to address the influence maximization problem. In an influence maximization problem it is common to select the seed set among nodes with highest degree/centrality. Since we are to solve an influence minimization problem, we then make the opposite choice. The approaches are briefly described in the following.

- **General Degree** (**G-D** for short): Standard degree heuristic is a simple heuristic approach that selects a set of k nodes with the lowest out-degrees in the graph. The out-degree of a node $i \in V$ in a graph denoted by d_i is the number of its out-neighbors, i.e., $d_i = A(i,:)\mathbf{1}$. If we set $\boldsymbol{c}, \boldsymbol{s} = \mathbf{1}$, i.e., all nodes are homogenous, we can apply this heuristic directly to select k nodes.

  However, for heterogenous nodes, i.e., $\boldsymbol{c}, \boldsymbol{s} \neq \mathbf{1}$, it is hard to select a set of nodes with lowest out-degrees such that the constraint (a) of the LMD problem is exactly satisfied. Therefore, in order to make it feasible for heterogenous nodes, we make the following changes based on the standard degree heuristic.

  By taking its cost and weight into account together we define its *general out-degree* $\bar{d}_i$ as follows: $\bar{d}_i = \frac{d_i}{s_i} \cdot c_i$. We prefer the nodes with minimum degree, low cost and high weight. Then, we select the set of nodes with minimum general out-degrees until constraint (a) of the LMD problem is satisfied.
- **General PageRank** (**G-PR** for short): Standard pagerank heuristic is a popular algorithm for ranking web pages. We use the teleportation model [7] shown in Eq. (5) to compute every pagerank value and the restart probability γ of pagerank is set as 0.15. Let $\boldsymbol{p}$ denote the pagerank vector and p_i denote the pagerank value of node i.

$$\boldsymbol{p} = (1-\gamma)\hat{A}\boldsymbol{p} + \frac{\gamma}{n}\mathbf{1}, \tag{5}$$

  where $\boldsymbol{p} \in (0,1)^n$ and $\sum_{i=1}^{n} p_i = 1$, and $\hat{A}$ is the scaled adjacency matrix in which $\hat{A}(i,j) = A(i,j)/\mathbf{1}^T A(:,j)$. If all the nodes are homogenous, we select a set of k nodes with the lowest pagerank value as the seed set.

  In order to make it feasible for heterogenous nodes, we modify the pagerank as we do for the degree heuristic. Then we define the *general pagerank* of i $\bar{p}_i$ as follows: $\bar{p}_i = \frac{p_i}{s_i} \cdot c_i$. We prefer nodes with minimum pagerank value, low cost and high weight. Then, we select the set of nodes with minimum general pagerank values until constraint (a) of the LMD problem is satisfied.

Note that, both the G-D and the G-PR are guaranteed to converge (although in general to a suboptimal solution) since they are structural approaches and not depend on the system dynamics.

4 Diffusion Minimization with Guaranteed Target

4.1 Problem statement

In this section, we introduce a class of influence minimization problems called *Diffusion Minimization with Guaranteed Target*. Let V_{init} denote the set of nodes that are accessible to be controlled and V_{target} denote the set of nodes that are desired targets. The diffusion minimization with guaranteed target aims to find a subset of nodes in V_{init} spreading from which all desired targets are influenced while the diffusion is minimized.

Problem 4.1 (DMGT: Diffusion Minimization with Guaranteed Target) *Given a diffusion model represented by a network with set of nodes V, 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.,*

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

As an example of a problem that can be framed as a DMGT consider the following. Cascading failures occur frequently in a power grid as failures of power generators can trigger additional node failures, due to load imbalance or thermal effect, following a linear threshold model [6]. A cyber-attack may aim to disrupt a set V_{target} of generator nodes by directly attacking a set of vulnerable nodes in V_{init}, while minimizing the total number of disrupted nodes to avoid a total network failure.

In the rest of this section we address the DMGT problem for a linear threshold network.

4.2 Existence of a solution

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

Fact 1 *The DMGT problem has a solution if and only if* $\Phi_\infty(V_{init}) \supseteq V_{target}$.

The previous result immediately follows by observing that the largest set of final adopters that one can obtain is $\Phi_\infty(V_{init})$ corresponding to a seed set $\phi_0 = V_{init}$. This leads to the following result.

Proposition 4.1 *Given a linear threshold network* $N = (V, E, \theta, W)$ *with* n *nodes, let sets* $V_{target} \subseteq V$ *and* $V_{init} \subseteq V$ *with characteristic vectors* $\boldsymbol{v}_{target}$ *and* $\boldsymbol{v}_{init}$ *be assigned. Consider the following ILP with binary variable* $\boldsymbol{x}$*:*

$$\begin{array}{lll} \max\limits_{\boldsymbol{x}} & \boldsymbol{1}^T \cdot \boldsymbol{x} & (ILP-3) \\ s.t. & \boldsymbol{1} - \boldsymbol{x} \geq \boldsymbol{v}_{init} & \\ & [I - \Theta - W^T] \cdot \boldsymbol{x} < 0 & \\ & \boldsymbol{x} \in \{0,1\}^n & \end{array}$$

The DMGT problem has a solution if and only if (ILP-3) has an optimal solution $\boldsymbol{x}^* \leq \boldsymbol{1} - \boldsymbol{v}_{target}$.

PROOF. By Fact 1, the DMGT problem has solution if and only if $\Phi_\infty(V_{init}) \supseteq V_{target}$. By Proposition 2.1, this happens if and only if the characteristic vector of $\Phi_\infty(V_{init})$, i.e., $\mathbf{1} - \boldsymbol{x}^*$, satisfies $\mathbf{1} - \boldsymbol{x}^* \geq \boldsymbol{v}_{target}$. □

4.3 A sufficient condition for optimality

In the following we provide a sufficient condition for a solution of the DMGT problem to be optimal.

Proposition 4.2 *Given a linear threshold network* $N = (V, E, \theta, W)$ *with* n *nodes, let sets* $V_{target} \subseteq V$ *and* $V_{init} \subseteq V$ *with characteristic vectors* $\boldsymbol{v}_{target}$ *and* $\boldsymbol{v}_{init}$ *be assigned. Consider the following ILP with binary variable* $\boldsymbol{x}$*:*

$$\begin{array}{lll} \max\limits_{\boldsymbol{x}} & \boldsymbol{1}^T \cdot \boldsymbol{x} & (ILP-4) \\ s.t. & \boldsymbol{1} - \boldsymbol{x} \geq \boldsymbol{v}_{target} & \\ & [I - \Theta - W^T] \cdot \boldsymbol{x} < 0 & \\ & \boldsymbol{x} \in \{0,1\}^n & \end{array}$$

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

$$\begin{array}{lll} \max\limits_{\boldsymbol{x}} & \boldsymbol{1}^T \cdot \boldsymbol{x} & (ILP-5) \\ s.t. & \boldsymbol{1} - \boldsymbol{x} \geq \boldsymbol{z} & \\ & [I - \Theta - W^T] \cdot \boldsymbol{x} < 0 & \\ & \boldsymbol{x} \in \{0,1\}^n & \end{array}$$

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 the DMGT problem.*

PROOF. Firstly, if we remove constraint (b) in the DMGT problem, i.e., we can select any node as seed in V, obviously the optimal seed set is $\phi_0 = V_{target}$. The optimal set of final adopters can be computed by solving (ILP-4) and is $\Phi_\infty(V_{target}) = \{i \in V | \bar{x}_i^* = 0\}$ according to Proposition 2.1.

Now assume that the seed set is $Z = V_{init} \cap \Phi_\infty(V_{target})$ (its characteristic vector is $\boldsymbol{z} = \max\{\boldsymbol{v}_{init} - \bar{\boldsymbol{x}}^*, \boldsymbol{0}\}$). This set is the largest one that will not activate nodes in the complement of $\Phi_\infty(V_{target})$ and that also satisfies constraint (b) in the DMGT problem. If the optimal solution $\boldsymbol{x}^*$ of (ILP-5) is also the solution of (ILP-4), then we have $\Phi_\infty(Z) = \{i \in V | x_i^* = 0\}$ that is identical to $\Phi_\infty(V_{target})$. □

4.4 Heuristic solution

The algebraic approach discussed above for finding an optimal solution works only in particular cases. Instead, in the general case in order to obtain a reasonably efficient solution to the DMGT problem, we propose the greedy algorithm shown in Algorithm 1. We start with a seed set $\phi_0 = V_{min} = V_{target} \cap V_{init}$

Algorithm 1 Greedy algorithm for the DMGT problem

1: **Input:** A social network with set of nodes V, 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: $\quad u = \underset{u \in S}{argmin} \, |\Phi_\infty(\phi_0 \cup \{u\})|$.
9: $\quad \phi_0 \leftarrow \phi_0 \cup \{u\}$.
10: $\quad V_{complete} \leftarrow V_{complete} \setminus \{u\}$.
11: **end while**

(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 from $V_{complete} = V_{init} \setminus V_{min}$ and add it to the seed set. We select from $V_{complete}$ all the seed nodes with maximal marginal gain in reaching the target nodes (line 7) as 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 $O(l^2nd)$ where $l = |V_{complete}|$, $n = |V|$, and d is the average degree of network. In fact, the algorithm requires at most l iterations since in the worst case it needs to go through all nodes in $V_{complete}$ to reach all target nodes. At each iteration it also needs to test all l nodes in $V_{complete}$ to find the ones with maximal marginal gain. At each step a time $O(nd)$ is required to compute the set of final adopters.

5 Experiments and analysis

5.1 Experimental set up

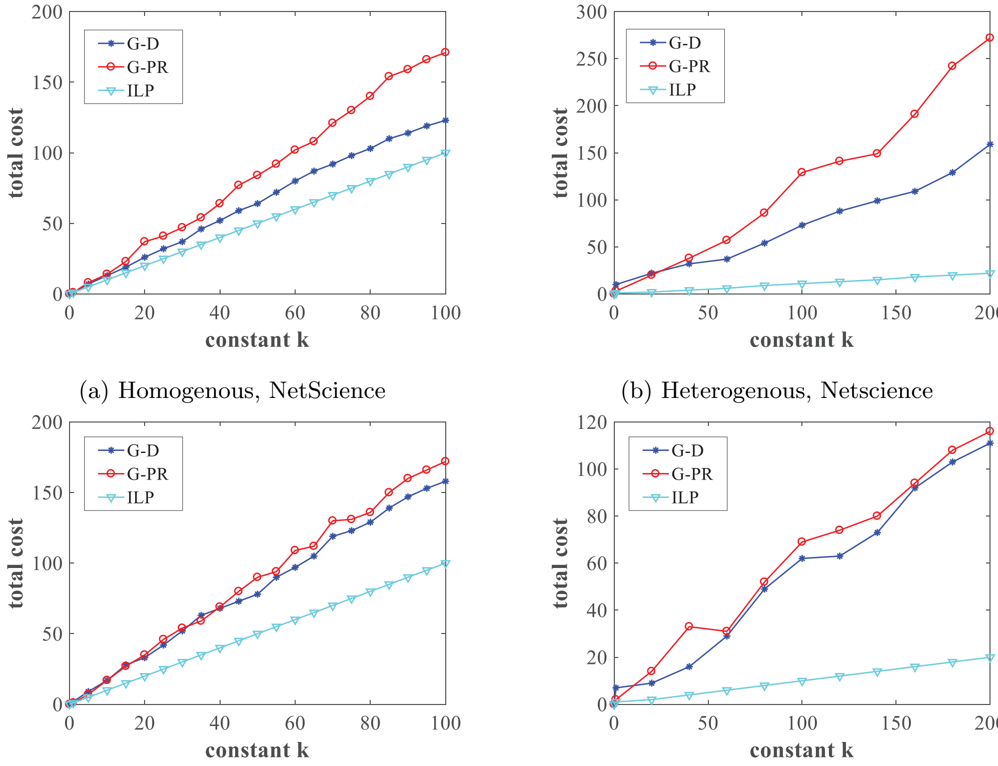
We performed a series of experiments based on two real-world networks of different scale under the linear threshold model. NetScience [12] and Hep-Th [11] are from different sections of the e-print arXiv, which is a data source commonly used in the literature for social network analysis. Empirical evidence suggests that coauthorship graphs are representative of typical social networks [11]. The precise statistical information of the datasets is summarized in Table 1. All experiments are performed using MATLAB on a 2.5 GHz Intel(R) Core(TM) i5-6300U CPU and 8GB memory.

Table 1
Statistics of real datasets

Dataset	Nodes	Edges	Average degree
NetScience	1,461	2,742	1.88
Hep-Th	7,610	15,751	2.07

5.2 Experimental results for the LMD problem

For the LMD problem, we compare our exact ILP approach proposed in Proposition 3.1 with the heuristics presented in Section 3.3. We test the algorithms on the two networks and compare them in terms of the influence spread and the running time. We consider for the cost vector $\boldsymbol{c}$ and weight vector $\boldsymbol{s}$ two cases: a) the nodes in the network are homogenous, i.e., $\boldsymbol{c} = \boldsymbol{s} = \mathbf{1}$, and b) the nodes in the network are heterogenous, i.e., $\boldsymbol{c}, \boldsymbol{s} \neq \mathbf{1}$ (in this experiment, we assume that all c_i and s_i are in [1,10]). In order to have a good estimate of the efficiency of the algorithms, we run each experiment on 100 different randomly generated linear threshold networks with the same network structure and influence weights but different thresholds randomly selected in $(0, 1]$. Then we take the average total cost over the 100 runs.



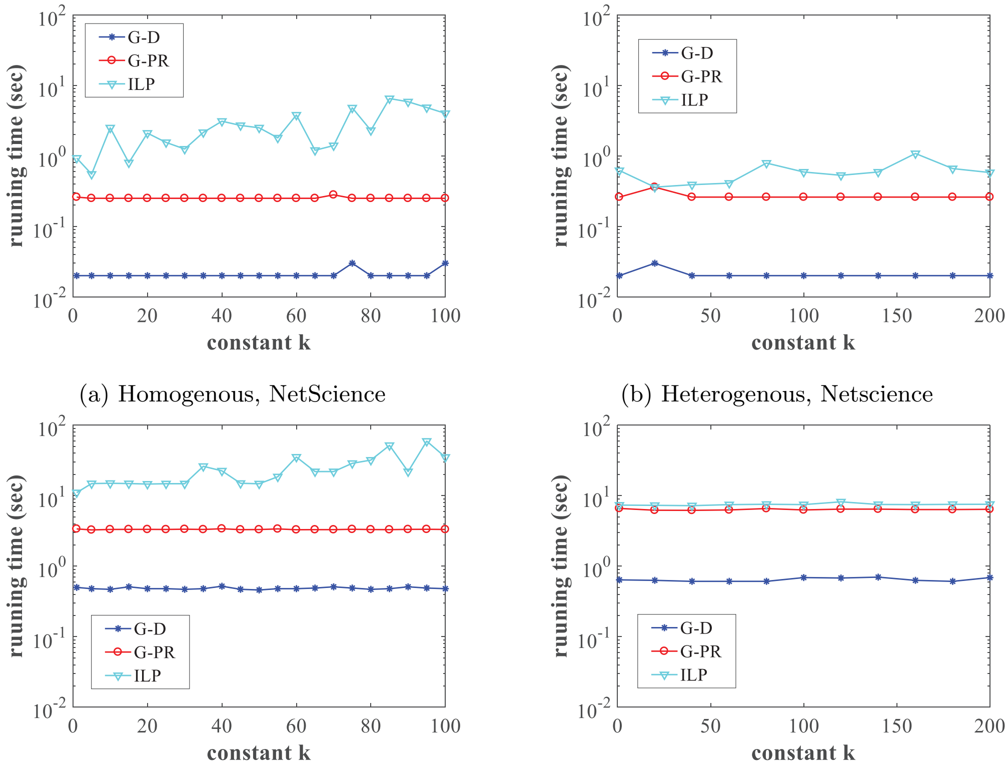
(a) Homogenous, NetScience (b) Heterogenous, Netscience

(c) Homogenous, Hep-Th (d) Heterogenous, Hep-Th

Fig. 1. Influence spread of various algorithms for the LMD problem on NetScience and Hep-Th.

Fig. 1 shows the average total cost of the tested algorithms on datasets NetScience and Hep-Th under two cases: homogenous nodes in Figs. 1a and 1c and heterogenous nodes in Figs. 1b and 1d. The average running time of the compared approaches for selecting each size k under two cases (from 1 to 100 and from 1 to 200 respectively) is illustrated in Fig. 2. From the experimental results, we can make the following observations on the effectiveness of the approaches.

- The ILP successfully selects the optimal seed set, and always outperforms the heuristics relying solely on the structure properties of the network. We define the optimality gap for the heuristics G-D and G-PR



(a) Homogenous, NetScience

(b) Heterogenous, Netscience

(c) Homogenous, Hep-Th

(d) Heterogenous, Hep-Th

Fig. 2. Comparisons of running time for the LMD problem on NetScience and Hep-Th.

as $(f_c^{heu} - f_c^{ILP})/f_c^{ILP} \times 100\%$ (where f_c^{heu} denotes the final total cost by either G-D or G-PR and f_c^{ILP} denotes the final total cost by the ILP). On NetScience with homogenous nodes (Fig. 1a), the optimality gaps for the G-D and G-PR are 28% and 64% respectively. On Hep-Th with homogenous nodes (Fig.1c), the gaps are 68% and 72% respectively. Furthermore, the gap is bigger in the heterogenous cases. On NetScience the gaps are 636% and 890% for the G-D and G-PR respectively (Fig. 1b) and they are 446% and 495% on Hep-Th (Fig. 1d). We believe this is due to the fact that low-degree (or low-centrality) nodes may be connected with nodes that have high degree (or high centrality) which may lead to further diffusion of the bad influence.

- The heuristic approaches G-D and G-PR are 1 and 2 orders of magnitude faster than the exact ILP respectively for the homogenous case in both NetScience and Hep-Th (see Figs. 2a and 2c). For the heterogenous case, the G-D and G-PR are around 20 times and twice faster than the ILP (see Figs. 2b and 2d).

Furthermore, we do an experimental study on random networks of different size under the homogenous case to test the scalability of the three approaches. The networks are generated by Software Gephi of size from 2K to 12K and the wiring probability (i.e.,the probability that there is an arc between node i and j) is set to 0.005. The result is shown in Fig. 3 for selecting 100 seeds in each network, which suggests that the G-D and G-PR are about 20 and 5 times faster than the ILP respectively. The complexity of ILP may grow large in very large scale networks. It may be possible to combine the integer linear programming approach with a divide-and-conquer scheme, that is, a large network may be divided into smaller communities according to suitable partition rules. This may lead to a reduced complexity approach for finding solutions close to the optimum.

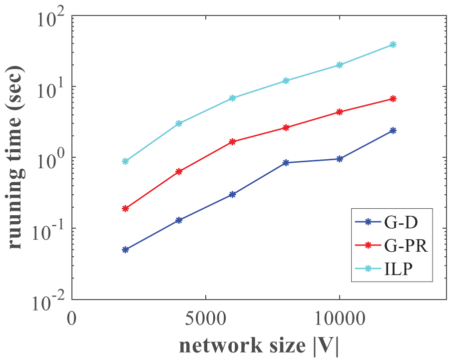


Fig. 3. Scalability of the algorithms

5.3 Experimental results for the DMGT problem

For the DMGT problem, we test the efficiency of Algorithm 1 which can find a solution in the general case on both networks NetScience and Hep-Th. In this part, in order to have a good estimate of the performance of the greedy algorithm, we also 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 the DMGT problem 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} that denotes the percent of nodes that can avoid being activated by our strategy but will be activated if the seed set is V_{init}, i.e.,

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

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 2, where ρ^{pro}_{max} is the maximal percent (over all simulations) of nodes that can avoid being activated but will be activated if the seed set is V_{init}, and $(|V_{init}|, |V_{target}|)$ is a pair showing the cardinality of sets V_{init} and V_{target}. We test three pairs for each dataset. The three pairs are for the cases $|V_{init}| = |V_{target}|$, $|V_{init}| < |V_{target}|$, and $|V_{init}| > |V_{target}|$, respectively. We can observe

Table 2
Experimental result for the DMGT problem

Dataset	$(\lvert V_{init}\rvert, \lvert V_{target}\rvert)$	ρ^{pro}	ρ^{pro}_{max}
NetScience	(10,10)	29.62%	100%
	(30,50)	22.06%	45.00%
	(40,20)	44.72%	69.23%
Hep-Th	(10,10)	28.13%	100%
	(30,50)	20.63%	36.07%
	(40,20)	44.53%	70.33%

from Table 2 that the proposed greedy strategy can efficiently protect some percent of nodes from being activated. For example, when $(|V_{init}|, |V_{target}|) = (10, 10)$, 29.62% nodes in average avoid being activated. Besides, among all the simulations, the maximal percent of nodes that can avoid 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. The algorithm performs the best for the case $|V_{init}| > |V_{target}|$. Since for the other two cases, the number of targeted nodes is bigger than or equal to that of the nodes which can be infected directly, sometimes we need to select almost all of the nodes in V_{init} such that all targeted nodes can be reached. However, how far away the solution obtained by the proposed strategy is from the optimal solution is unknown so far, and determining a lower bound of the approximate guarantee for our solution is an open problem.

6 Conclusions and future work

In this paper, we formalize two different influence minimization problems generalizing real-world scenarios. For the LMD problem, following the characterization of [16], we provide an optimal solution by solving an integer linear programming problem and show that it outperforms other general heuristics. For the DMGT problem, we provide a technique to search for an optimal solution that works in some particular cases and discuss a simple heuristic to find a solution (in general suboptimal) in the general case.

We are aware that ILP based solutions do not scale well for large networks. For this reason our future work will consist in exploring reduced complexity approach for finding solutions close to the optimum by combining the ILP with a divided-and-conquer scheme.

References

[1] Acemoglu, D., Ozdaglar, A., & Yildiz, E. (2011). Diffusion of innovations in social networks. *In Proc. of the 50th IEEE Conf. on Decision and Control and Europeen Control Conf.*, Orlando, USA (pp. 2329–2334).

[2] Domingos, P., & Richardson, M. (2001). Mining the network value of customers. *In Proc. of the 7th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data mining*, San Francisco, USA (pp. 57–66).

[3] Fazeli, A., Ajorlou, A., & Jadbabaie, A. (2017). Competitive diffusion in social networks: Quality or seeding? *IEEE Transactions on Control of Network Systems*, 4(3), 665–675.

[4] Granovetter, M. (1978). Threshold models of collective behaviors. *The American Journal of Sociology*, 83(6), 1420–1443.

[5] Ghaderi, J., & Srikant, R. (2014). Opinion dynamics in social networks with stubborn agents: Equilibrium and convergence rate. *Automatica*, 50(12), 3209–3215.

[6] Golnari, G., & Zhang, Z. L. (2015). The effect of different couplings on mitigating failure cascades in interdependent networks. *In Proc. of 2015 IEEE Conf. on Computer Communications Workshops*, Hong Kong, China (pp. 677–682).

[7] Ishii, H., & Tempo, R. (2014). The PageRank problem, multiagent consensus, and web aggregation: A systems and control viewpoint. *IEEE Control Systems*, 34(3), 34–53.

[8] Kempe, D., Kleinberg, J., & Tardos, E. (2003). Maximizing the spread of influence through a social network. *In Proc. of the 9th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data mining*, Washington, USA (pp. 137–146).

[9] Kimura, M., Saito, K., & Motoda, H. (2009). Blocking links to minimize contamination spread in a social network. *ACM Transactions on Knowledge Discovery from Data*, 3(2), Art. 9.

[10] Lin, F., Fardad, M., & Jovanovic, M. R. (2014). Algorithms for leader selection in stochastically forced consensus networks. *IEEE Transactions on Automatic Control*, 59(7), 1789–1802.

[11] Newman, M. E. (2001). The structure of scientific collaboration networks. *Proc. of the National Academy of Sciences*, 98(2), 404–409.

[12] Newman, M. E. (2006). Finding community structure in networks using the eigenvectors of matrices. *Physical Review E*, 74(3), Id: 036104.

[13] Nozari, E., & Cortes, J. (2018). Stability analysis of complex networks with linear-threshold rate dynamics. *In Proc. of the American Control Conf.*, Milwaukee, USA (pp. 191–196).

[14] Proskurnikov, A. V., Matveev, A. S., & Cao, M. (2016). Opinion dynamics in social networks with hostile camps: Consensus vs. polarization. *IEEE Transactions on Automatic Control*, 61(6), 1524–1536.

[15] Ramasuri, N., & 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.

[16] Rosa, D., & Giua, A. (2013a). On the spread of innovation in social networks. *In Proc. of the 4th IFAC Workshop on Distributed Estimation and Control in Networked Systems*, Rhine-Moselle-Hall, Germany (pp. 322–327).

[17] Rosa, D., & Giua, A. (2013b). A non-progressive model of innovation diffusion in social networks. *In Proc. of the 52nd IEEE Conf. on Decision and Control*, Florence, Italy (pp. 6202–6207).

[18] Song, G., Zhou, X., Wang, Y., & 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.

[19] Yang, W., Wang, Z., Zuo, Z., Yang, C., & Shi, H. (2016). Nodes selection strategy in cooperative tracking problem. *Automatica*, 74, 118–125.

[20] Yang, L., Giua, A., & Li, Z. (2017). Minimizing the influence propagation in social networks for linear threshold models. *In Proc. of the 20th IFAC World Congress*, Toulouse, France (pp. 14465–14470).

[21] Zhong, Y. D., Srivastava, V., & Leonard, N. E. (2017). On the linear threshold model for diffusion of innovations in multiplex social networks. *In Proc. of the 56th IEEE Conf. on Decision and Control*, Melbourne, Australia (pp. 2593–2598).