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.giu@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.

© 2017, IFAC (International Federation of Automatic Control) Hosting by Elsevier Ltd. All rights reserved.

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...
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 $V_{\text{target}}$ but can only directly infect servers in a given set $V_{\text{init}}$. The hacker aims to find a suitable subset of $V_{\text{init}}$ such that starting from this seed the infection will propagate to all nodes in $V_{\text{target}}$. 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.

| 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 $G$ |
| $\theta_i$ | Threshold value of node $i \in V$ |
| $N_i$ | The in-neighbor set of node $i \in V$ |
| $\phi$ | The seed set |
| $\phi_t$ | The set of nodes that become active at step $t$ |
| $\Phi_\infty(\phi_0)$ | The set of final adopters given that $\phi_0$ is the seed set |
| $\phi_0^*$ | 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 $\theta_i$ 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 $\tilde{A} \in [0,1]^{n \times n}$ as follows:

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

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

Activation process Let $\phi_0$ 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 $\phi_t$ 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 $\theta_i$, i.e.,

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

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 $\phi_0$.

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:

$$\left| \frac{|X \cap N_i|}{|N_i|} \right| > 1 - \theta_i$$

\(\diamond\)
دریافت فوری متن کامل مقاله

امکان دانلود نسخه تمام متن مقالات انگلیسی
امکان دانلود نسخه ترجمه شده مقالات
پذیرش سفارش ترجمه تخصصی
امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
امکان دانلود رایگان ۲ صفحه اول هر مقاله
امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
دانلود فوری مقاله پس از پرداخت آنلاین
پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات