Computer Networks 123 (2017) 38–50

Contents lists available at ScienceDirect

Computer Networks

journal homepage: www.elsevier.com/locate/comnet

# Scalable influence blocking maximization in social networks under competitive independent cascade models

Peng Wu[a,b], Li Pan[a,b,*]

[a] School of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University, 800 Dong Chuan Rd, Shanghai, China
[b] National Engineering Laboratory for Information Content Analysis Technology, Shanghai Jiao Tong University, Shanghai, China

ARTICLE INFO

*Article history:*
Received 21 January 2017
Revised 2 April 2017
Accepted 4 May 2017
Available online 5 May 2017

*MSC:*
68P10
91D30

*Keywords:*
Influence blocking maximization
Competitive independent cascade model
Social networks

ABSTRACT

Bad information propagation in online social networks (OSNs) can cause undesirable effects. The opposite good information propagating competitively with bad information can restrain the propagation of bad information. In this paper, we address the Influence Blocking Maximization (IBM) problem aiming to find a set of influential people initiating good information propagation to maximize the blocking effect on the bad information propagation in OSNs. The problem is studied on two competitive propagation models describing competitive propagation processes in two classic situations in OSNs. Two models are derived from the Independent Cascade Model (ICM). Greedy algorithms for IBM problem under two competitive propagation models are slow and not scalable. Thus, we design two heuristics CMIA-H and CMIA-O based on the maximum influence arborescence (MIA) structure to efficiently solve the IBM problem under two competitive propagation models, respectively. Extensive experiments are conducted on real-world and synthetic datasets to compare the proposed algorithms with the greedy algorithms and other baseline heuristics. The results demonstrate that both CMIA-H and CMIA-O achieve matching influence blocking performance to the greedy algorithms and consistently outperform other baseline heuristics, while they are several orders of magnitude faster than the greedy algorithms.



## 1. Introduction

Online Social Networks (OSNs), such as Facebook, Twitter and Weibo, etc., have become one of the most important platforms for the propagation of news, ideas, opinions, adoption of new products, etc.. Such information propagation is generally referred as influence propagation [1]. While OSNs are very beneficial for genuine and trustworthy information dissemination, their openness also facilitates the propagation of rumors, gossips and other kinds of misinformation which could potentially cause undesirable effects and even lead to serious economic and political consequences [2–5]. For example, the rumor that “Two bombs had exploded at the White House, injuring Barack Obama” from hacked Associated Press Twitter account resulted to 10 billion USD losses [6]. Similarly, the negative information about a product propagating in OSNs can cause huge losses to a company. Misinformation and negative information are collectively called bad information in this paper.

In order to make OSNs a more reliable platform for information dissemination, it is crucial to seek effective strategies to reduce the devastating effects of bad information [2,7]. If users have accepted good information used to correct bad information, they will hardly be influenced by such bad information [1,2,7,8]. Thus the propagation of good information can restrain the propagation of corresponding bad information in OSNs [1,2,7–12]. Given a set of bad information sources that initiate bad information propagation in OSNs, we study the Influence Blocking Maximization (IBM) problem of finding a set of users from whom the good information is disseminated so that the blocked negative influence range of bad information is maximized. The problem is studied on two competitive propagation models in this paper, i.e., Multi-Campaign Independent Cascade Model (MCICM) with high-effectiveness property and Campaign-Oblivious Independent Cascade Model (COICM) [2]. These two models are extended from Independent Cascade Model (ICM) [13] which is a classic stochastic propagation model describing how single information is propagated on the network starting from a set of seed nodes [1]. MCICM with high-effectiveness property and COICM describe and model competitive propagation processes in two classic situations, respectively. It has been proved that the IBM problems under above two models are NP-hard and their objective functions are submodular [2]. Thus, greedy algorithms can achieve an approximation ratio of $1-1/e-\epsilon$ of the optimal solution, where $e$ is the base of natural logarithm, and $\epsilon$

* Corresponding author.
*E-mail addresses:* panli@sjtu.edu.cn, catking@sjtu.edu.cn (P. Wu).

http://dx.doi.org/10.1016/j.comnet.2017.05.004
1389-1286/

depends on the accuracy of the Monte–Carlo estimate of the influence spread range. However, the greedy algorithms are very slow for large networks, if we want to keep the above $\epsilon$ small by running the Monte–Carlo simulations a large number of times. Given that OSNs are usually very large and blocking bad information propagation typically requires immediate strategies before it propagates too far, we believe that the scalability issue of the greedy algorithms for IBM problems will hinder their application in containment of bad information propagation in OSNs.

In this paper, we address the scalability issue of the greedy algorithms by designing two efficient heuristic algorithms CMIA-H and CMIA-O for the IBM problem under MCICM with high-effectiveness property and COICM, respectively. The good (bad) information propagation is referred to as positive (negative) propagation and the good (bad) information spread range that measures the number of nodes accepting good (bad) information is referred to as positive (negative) influence spread. The main reason of the inefficiency of the greedy algorithms is that they spend a lot of time to estimate the negative influence spread by a large number of Monte–Carlo simulations. Inspired by [14], we approximately compute the negative influence spread in the local Maximum Influence Arborescence (MIA) structure of every node in two designed algorithms. The influence spread computation in MIA is more than three orders of magnitude faster than that by Monte–Carlo simulations. Thus two designed algorithms can be much faster than greedy algorithms. It should be noted that the influence spread computation procedures here are more complex than that in [14], since our propagation models model two competitive propagations while theirs only model a single propagation. Moreover, the computation procedures under MCICM with high-effectiveness property and COICM are different. Specifically, for MCICM with high-effectiveness property, we adopt the high-effectiveness property of the positive propagation to help compute the blocked negative influence of any node. While for COICM, we design a dynamic programming method to compute the blocked negative influence of any node. Then the node with the largest blocked negative influence is iteratively selected as the new positive seed. To test the efficiency and effectiveness of the proposed CMIA-H and CMIA-O, we conduct extensive experiments on several real-world and synthetic networks. The performance of two proposed algorithms is compared with both the greedy algorithms and several heuristic algorithms, and the results show that (a) compared with the greedy algorithms, CMIA-H and CMIA-O have matched influence blocking performance, while they are several orders of magnitude faster to run; and (b) compared with other scalable heuristics, CMIA-H and CMIA-O consistently outperform them in negative influence blocking performance with a significant margin.

The rest of the paper is organized as follows. Some related works are discussed in Section 2. Section 3 formalizes the IBM problem under the two competitive propagation models, i.e., MCICM with high-effectiveness property and COICM. In Section 4, we describe the CMIA-H algorithm for IBM problem under MCICM with high-effectiveness property, and the CMIA-O algorithm for IBM problem under COICM. The experiments results are analyzed in Section 5. Finally, Section 6 concludes the paper.

## 2. Related work

The influence maximization (IM) problem is first formulated as a combinatorial optimization problem by Kemp et al.[13]. They study the IM problem under the classic ICM and Linear Threshold Model (LTM), and propose a standard greedy algorithm with approximation guarantee of $1 - 1/e - \epsilon$. LTM is another single information stochastic propagation model. To address the scalability issue of the greedy algorithm, many follow-up studies [14–19] try to propose much more efficient heuristics for the IM problem. The scalable LDAG algorithm of [16] utilizes the linear computation property of influence in directed acyclic graphs (DAG) to efficiently approximate the influence spread of any seed set under the LTM. The scalable PMIA algorithm of [14] utilizes MIA structure to efficiently approximate the influence spread of any seed set under the ICM. Our CMIA-H algorithm and CMIA-O algorithm are also based on the efficient influence computation property of the MIA structure, but with nontrivial extensions to deal with competitive influence propagation and IBM problem.

Different from previous IM problem that aim at maximizing the influence spread of one propagation, containment of bad information aims at minimizing the negative influence spread by either blocking bad information at influential users [5,20–24], or clarifying bad information by spreading good information [1,2,7–10]. In reality, blocking rumors at influential users may incur criticism since it has risks of violating human rights [3], while spreading good information competing with bad information is more moderate. Most of works about blocking bad information by spreading good information are studied under competitive propagation models extended from either ICM or LTM. Nguyen et al. [7] study the $\beta_T^I$-Node Protectors problems which aim to find the smallest set of nodes with good information to contain the spread of misinformation under two models extended from ICM and LTM, respectively. Fan et al. [9] address the Least Cost Rumor Blocking (LCRB) problem which aims at identifying a minimal subset of individuals as initial protectors to ensure bridge ends in neighbor communities of the rumor community protected under a model from ICM. Different from IBM problem studied in this paper, above two works study the seed set minimization problem. He et al. [1] study the IBM problem under the CLTM coming from LTM and propose CLDAG algorithm based on the DAG structure to efficiently solve the problem. Budak et al. [2] study the IBM problems under the MCICM with high-effectiveness property and the COICM, respectively. They prove that the IBM problems under two propagation models are both NP-hard, and their objective functions are submodular. Recently, several NP-hard problems about enhancing the performance of communication networks have been solved by algorithms mimicking biological processes [25,26]. As for problems with submodular objective functions, it has been proved that the greedy algorithms can solve them with approximation guarantee of $1 - 1/e$ [27,28]. However, the greedy algorithms for IBM problems are slow, as they require a large number of Monte–Carlo simulations to estimate the influence spread. Motivated by the extremely low efficiency of the greedy algorithms, this paper proposes two scalable heuristic algorithms CMIA-H and CMIA-O based on MIA structure [14] for the IBM problems under MCICM with high-effectiveness property and COICM, respectively. The CLDAG algorithm [1] inspires us in designing CMIA-O algorithm for COICM, but we propose a novel influence computation procedure. The comparisons of works most related to this paper are listed in Table 1.

## 3. Propagation models and problem definition

A social network is modeled as a directed graph $G = (V, E)$, where $V$ is a set of $n$ nodes representing individuals and $E$ is a set of $m$ edges representing relationships. Each node is associated with some states indicating whether it accepts the information or not. Propagation is a state cascade process where the state changes of the nodes depend on the state changes of their in-neighbors, and will further influence the state changes of their out-neighbors. The propagation models define some rules to capture such processes on networks. Nodes accepting the information are considered as being influenced by the information. Thus information propagation can be referred to as influence propagation. Two competitive propagation models, the MCICM with high-effectiveness property and the COICM [2], are first described as follows. Let $P$ and $N$ be the