Expert Systems With Applications 55 (2016) 255–267

Contents lists available at ScienceDirect

Expert Systems With Applications

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

Maximizing influence under influence loss constraint in social networks

Yifeng Zeng[a,b,*], Xuefeng Chen[c], Gao Cong[d], Shengchao Qin[e,b], Jing Tang[b], Yanping Xiang[c]

[a] Automation Department, Xiamen University, SIming Road Xiamen, Fujian 361005, China
[b] Teesside University, UK
[c] University of Electronic Science and Technology of China, Chengdu, China
[d] Nanyang Technological University, Singapore
[e] Shenzhen University, China

ARTICLE INFO

Keywords:
Influence maximization
Influence loss
Social networks

ABSTRACT

Influence maximization is a fundamental research problem in social networks. Viral marketing, one of its applications, aims to select a small set of users to adopt a product, so that the *word-of-mouth* effect can subsequently trigger a large cascade of further adoption in social networks. The problem of influence maximization is to select a set of K nodes from a social network so that the spread of influence is maximized over the network. Previous research on mining top-K influential nodes assumes that all of the selected K nodes can propagate the influence as expected. However, some of the selected nodes may not function well in practice, which leads to *influence loss* of top-K nodes. In this paper, we study an alternative influence maximization problem which is naturally motivated by the reliability constraint of nodes in social networks. We aim to find top-K influential nodes given a threshold of influence loss due to the failure of a subset of $R(<K)$ nodes. To solve the new type of influence maximization problem, we propose an approach based on *constrained simulated annealing* and further improve its performance through efficiently estimating the influence loss. We provide experimental results over multiple real-world social networks in support. This research will further support practical applications of social networks in various domains particularly where reliability would be a main concern in a system deployment.

1. Introduction

Social networks provide an intuitive representation about individual connections and display interesting behavioral patterns across various populations of users (Wasserman & Faust, 1994). Social network analysis is attracting more and more attention from different research areas and becomes an important tool for developing intelligent systems in recommendation, crowdsourcing service and so on Domingos and Richardson (2001), Zafarani, Abbasi, and Liu (2014), Sun, Lin, and Xu (2015), Zeng et al. (2015).

The merit of a social network lies in the power of users' interaction that propagates influence of individuals toward the entire network. Such effects have been seen in many real-world applications. For example, a Tweet in *Twitter* is probably followed by hundreds even thousands of the registered users. By exploiting influence spread, a marketing campaign may target a small set of influential individuals and expect that the selected users would generate the largest influence coverage in the market. This is a general problem of influence maximization in social networks where the task is to find top-K influential nodes through influence diffusion models (Kempe, Kleinberg, & Tardos, 2003).

In an ideal circumstance, top-K nodes will spread the influence once they are selected and subsequently activated in a social network. The maximum influence can be achieved only if all of the selected nodes have successfully propagated the influence. However, the influence will be compromised when some of the nodes may not function as they are expected. For example, to market a new product, a company selects a set of retailers that are active and show interest in a similar product market. Due to the changing of financial situations, some of the retailers may not persist the marketing focus on the recommended product. Consequently the new product will not be exposed as much as it should be in the market.

* Corresponding author at: Automation Department, Xiamen University, SIming Road Xiamen, Fujian 361005, China. Tel.: +4401642284311.

E-mail addresses: Y.Zeng@tees.ac.uk (Y. Zeng), cxflovechina@gmail.com (X. Chen), gaocong@ntu.edu.sg (G. Cong), s.qin@tees.ac.uk (S. Qin), X9019186@tees.ac.uk (J. Tang), xiangyanping@gmail.com (Y. Xiang).

http://dx.doi.org/10.1016/j.eswa.2016.01.008
0957-4174/

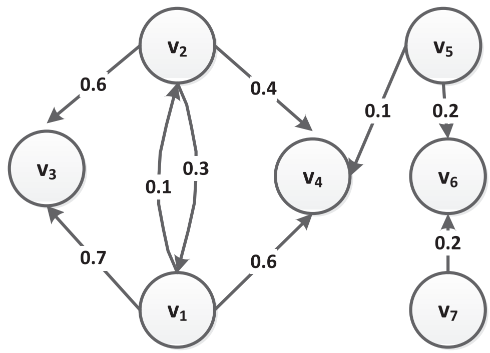

Fig. 1. A social network with influence values on the directed edges. Top-3 influential nodes are $\{v_2, v_5, v_7\}$ in the conventional influence maximization while $\{v_2, v_5, v_1\}$ are solutions to the influence maximization with a tolerable influence loss (1.5).

The *influence loss* occurs to the selected retailers from the perspective of the marketing company. Intuitively, the company may prefer to choose a set of retailers such that they are able to reach a certain level of market coverage (probably not the maximum one), and the market loss is tolerable due to possible malfunctions of the selected retailers in the campaign. As another example application of influence maximization, considering a water network, sensors deployed on the selected locations (pipe junctions) to monitor contaminant spread in the network (Ostfeld, Uber, & Salomons, 2006) as quickly as possible. The detection loss due to the possible malfunctions of sensors would lead to a disastrous effect, and thus it is vital that the detection loss shall be considered when locations are selected. In this paper, we study how the consideration of influence loss would impact the selection of top-K influential nodes in social networks.

Things become complex since the set of nodes that are proned to failure are unknown in the selection process. In addition, it is normally hard, if not impossible, to predict individual failure probabilities of nodes in a large scale of social networks, which may depend on many uncertain factors and vary from time to time. On the other hand, it is easier to estimate the number of failure nodes according to previous observations on the malfunctioning networks. For example, in a water network (Ostfeld et al., 2006), it is rather difficult to predict the failure probabilities of individual sensors, which may be affected by their environment, deployment duration, etc. However, it is easier to estimate the number of failure sensors in the future since the sensor quality is a main factor deciding its functions. Similarly, it is difficult to get failure probabilities of retailers in the marking campaign (Hajian & White, 2012) while it is more feasible to predict the number of retailers that may not perform well as expected. Hence we focus on the investigation of influence loss given the number of failure nodes in social networks.

Failure of some nodes may lead to an invisible influence loss while the loss could be significantly large due to failure of others. To act in a pessimistic way, we consider the worst case that the largest influence loss occurs to the selected nodes, and we may tolerate the loss only if it is not beyond a threshold. By computing the influence of nodes through a traditional influence diffusion model, e.g. independent cascade model as well as its improvement (Jacob, Barak, & Eitan, 2001; Liu, Cong, Zeng, Xu, & Meng, 2014a), we elaborate one example of influence loss in Fig. 1.

Example 1. Given $K = 3$, we can compute the influence for the set of nodes, $\{v_2, v_5, v_7\}$, as follows. We first compute the influence for every node in the social network and then sum the individual influence. As v_1 is only influenced by v_2, the influence is 0.3 from the entire set. For the node v_3, two paths, $v_2 -> v_1 -> v_3$ and $v_2 -> v_3$, may spread the influence from the set $\{v_2, v_5, v_7\}$. v_3 could be influenced by either of them or both. Hence, the influence is counted as: 1-(1–0.6) $\times$ (1–0.3 $\times$ 0.7)=0.684. Similarly, v_4 receives the influence from three paths: 1-(1–0.4) $\times$ (1–0.3 $\times$ 0.6) $\times$ (1–0.1)=0.5572. v_6 gets the influence: 1-(1–0.2) $\times$ (1–0.2)=0.36. Finally, the set have deterministic influence (=1) on their own nodes $\{v_2, v_5, v_7\}$. Hence the influence induced by $\{v_2, v_5, v_7\}$ is calculated as: $0.3 + 0.684 + 0.5572 + 0.36 + 3 = 4.9012$.

For the network of a small size, we can compute the influence for every set of three nodes and identify that the set of nodes, $\{v_2, v_5, v_7\}$, exhibit the maximum influence (influence value=4.9012) while nodes, $\{v_1, v_2, v_5\}$, are the second best influential ones (influence value=4.864). Assume that only one node fails in the selected set. We then compute the influence exhibited by the remaining two nodes in the set. The influence difference between the original set of three nodes and the remaining nodes is the influence loss value due to the node failure. Accordingly, the largest loss is 2.4412 for the first set when node v_2 fails; while, the largest loss is 1.2264 when v_1 fails in the second set. Given a tolerable value of influence loss (threshold=1.5), we may accept the second set of nodes, $\{v_1, v_2, v_5\}$, although they are not the top-3 influential ones in the conventional influence maximization problem.

To provide a reliable top-K solution to influence maximization problems in social networks, we study influence loss in this paper. Assume that the number of failure nodes could be predicted in influence propagation, we develop an approach to finding top-K nodes that maximize the influence spread given a threshold of the largest influence loss. We solve the task of mining top-K nodes by formulating it as one constrained optimization problem. In the context of social networks, solving such a problem is rather hard. The greedy algorithm (Kempe et al., 2003) that was extensively used to solve conventional influence maximization problems becomes problematic since feasible/optimal solutions may be prevented. Incrementally adding nodes with the largest marginal influence may simultaneously introduce potential nodes failure of which will result in an incredible influence loss. In addition, the computation of influence loss may further contribute to the solution complexity.

We propose a *Constrained Simulated Annealing* (CSA) (Wah & Wang, 1999; Wang, 2001) based technique for solving influence maximization problems with the constraint of influence loss. The approach is guaranteed to converge towards the optimum with probability one if such solutions exist. *The CSA algorithm development is not trivial in our context* since we need to design a proper penalty function that correctly encodes the influence maximization problem with influence loss constraint in an effective way. More importantly, we need to investigate sufficient conditions and practical parameter settings that guarantee the algorithmic convergence in the new problem.

To improve the efficiency of the CSA algorithm, we develop a new evaluation function that allows a fast computation of influence loss by reusing calculations of individual influence in a social network. Subsequently, we prove that the new algorithm can generate feasible solutions to the complex influence maximization problem. We conduct extensive experiments in multiple real-life social networks and demonstrate performance of the proposed algorithms.

The remainder of this paper is organized as follows. In Section 2, we review the most relevant work on influence maximization techniques. In Section 3, we formulate an influence maximization problem with influence loss constraint and prove its hardness. In Section 4, we propose the CSA based algorithm and further improve its efficiency by developing a new penalty function in CSA. The algorithmic convergence and complexity are also