Journal of Network and Computer Applications 63 (2016) 39–49

Contents lists available at ScienceDirect

Journal of Network and Computer Applications

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

An exploration of broader influence maximization in timeliness networks with opportunistic selection

Meng Han [a], Mingyuan Yan [b], Zhipeng Cai [c], Yingshu Li [c,*]

[a] The Department of Computer Science, Georgia State University, Atlanta, GA 30303, United States
[b] Department of Computer Science and Information Systems, University of North Georgia, Dahlonega, GA 30597, United States
[c] The Department of Computer Science, Georgia State University, Atlanta, GA 30303, United States

ARTICLE INFO

Article history:
Received 5 May 2015
Received in revised form
17 November 2015
Accepted 5 January 2016
Available online 21 January 2016

Keywords:
Social networks
Influence maximization
Algorithm
Data mining

ABSTRACT

The goal of classic influence maximization in Online Social Networks (OSNs) is to maximize the spread of influence with a fixed budget constraint, e.g. the size of seed nodes is pre-determined. However, most existing works on influence maximization overlooked the information timeliness. That is, these works assume that the influence will not decay with time and the influence could be accepted immediately, which are not practical. Second, even the influence could be passed to a specific node in time, whether the influence could be delivered (influence take effect) or not is still an unknown question. Furthermore, if let the number of users who are influenced as the depth of influence and the area covered by influenced users as the breadth, most of research results only focus on the influence depth instead of the influence breadth. Timeliness, acceptance ratio and breadth are three important factors neglected before but strongly affect the real result of the influence maximization. In order to fill the gap, a novel algorithm that incorporates time delay for timeliness, opportunistic selection for acceptance ratio, and broad diffusion for influence breadth has been investigated in this paper. In our model, the breadth of influence is measured by the number of communities, and the tradeoff between depth and breadth of the influence could be balanced by a parameter φ. Empirical studies on different large real-world social networks show that high depth influence does not necessarily imply broad information diffusion. Our model, together with its solutions, not only provides better practicality but also gives a regulatory mechanism for the influence maximization. It also outperforms most of the existing classical algorithms.

Published by Elsevier Ltd.

1. Introduction

Each month, more than 1.3 billion users are active on Facebook, and 190 million unique visitors are active on Twitter. Furthermore, 48% of 18–34 year old Facebook users check their online personal web pages when they wake up, and 98% of 18–24 year old people are involved in at least one kind of social media.[1] Since customers are the most important foundation of business, Online Social Networks (OSNs) have become one of the most effective and efficient solutions for marketing and advertising. But there is still no specific answer for how to handle and utilize data from OSNs. The development of OSNs and the resultant of a huge volume of data bring both opportunities and challenges.

Influence maximization, as one of the most popular topics in OSNs, attracts a lot of interest recently. Several models have been proposed in literatures (Kempe et al., 2003; Leskovec et al., 2007) to model the influence diffusion. However, because of the complexity and diversity of social phenomenon, many important features have been ignored, resulting in no practical influence diffusion is well modeled. We are facing a lot of challenges such as timeliness, acceptance ratio and breadth while analyzing and maximizing influence in OSNs. *Timeliness* refers to the phenomena that the effect of influence would decay with time; *acceptance ratio* measures the percentage of influence which gets a response; and influence *breadth* aims at maximizing influence not only by having more users, but also by achieving a broader user distribution in reality.

In the viral marketing and media domain, it is very common that many limited-time promotions and immediacy news exist where the influence and spreading of them decay with time. During the process of advertisement promotion or marketing strategies, the fact that a message could be passed on to someone never means the message could be accepted by the receiver (acceptance means the receiver takes action or response to the message). Therefore, receiving and accepting would be two procedures of influence. From this point of view, taking the acceptance ratio into account would make the influence model more practical than the traditional naive way. The expectation of the

* Corresponding author.
E-mail addresses: mhan@cs.gsu.edu (M. Han), myan@ung.edu (M. Yan), zcai@gsu.edu (Z. Cai), yli@cs.gsu.edu (Y. Li).

[1] http://www.statisticbrain.com/facebook-statistics/

http://dx.doi.org/10.1016/j.jnca.2016.01.004
1084-8045/Published by Elsevier Ltd.

traditionally formulated influence model is considered as the depth of influence. Another important issue is how broad the influence could be propagated based on the selected source seeds: the breadth of influence. Breadth relies not only on the number of influenced nodes, but also on the size of the area that could be covered by the influenced nodes. Surprisingly, although most researchers consider the path or routing of influence spreading based on network structures, as far as we know, there is no existing work considering the range (breadth) of the influence yet. Therefore, the question appears: which one is more important for influence maximization? influence more users in depth[2] or breadth?

Let us take a conventional social network activity as an example to discuss the influence diffusion in daily life. Assume that there is one user on Facebook sharing a new song or movie. This action results in an influence diffusion process. That is, friends or followers of the action initiator will have similar behaviors – be influenced. Considering one instance, user *Mike* posts a new status "*I got a new iPhone 6 plus from Apple Store with student promotion. It is awesome!*" with pictures on Facebook. All of *Mike*'s friends and followers will get this information from their Facebook's news feed or related search results. For timeliness, the effect of this influence will be weakened as time goes on. For acceptance ratio, obviously not all the neighbors who see the post will forward it, although some of *Mike*'s friends might have already been influenced and begun to take the next step to purchase an iPhone, but some of his friends might have simply ignored this post. We consider the receiving of that post as the first step of influence, and all the users having a friend relationship with *Mike* have a probability to receive this influence. But only the neighbors who comment, forward this status, or take response action regarding this post could be considered as accepting the influence, which is the second step of the influence. For the breadth of influence, one possibility is a lot of *Mike*'s friends are studying at the same department of the same university. If we evaluate the influence ability of *Mike* in the whole social network, he might not be as good as another user *Michael*, who has fewer friends but his friends are studying in many different universities. Compared with *Mike*, *Michael* has a good chance to pass the influence much more broader than *Mike*. Thus, all the three aforementioned factors should be taken into consideration.

Additionally, how to evaluate influence in OSNs is still an open problem. Although several models have been proposed to evaluate the influence by analyzing history logs (Goyal et al., 2010) or learning users' behaviors (Zhang et al., 2013), there is still lack of literatures considering the impact between users in a timeliness model with respect to the influence decaying process and the optimistic selection for a better acceptance ratio. Therefore, different from the traditional influence models which only focus on the traditional influence expectation result or the efficiency of the algorithm (Chen et al., 2009; Goyal et al., 2013; Tang et al., 2009), we investigate the influence maximization from a much more practical and comprehensive perspective.

In this paper, we address the problem of identifying the node set which maximizes influence in practical social networks. Our model incorporates influence decay function, opportunistic selection and broader maximization accommodating to three factors: timeliness, acceptance ratio and breadth. More specifically, our contributions are summarized as follows:

1. We formulate the problem of influence maximization with opportunistic selection in a timeliness model *ICOT*. The model incorporates the timeliness feature and considers the decaying of influence diffusion.
2. We propose opportunistic selection to deal with the acceptance ratio which represents the real reception of influence propagation in practice.
3. We show the NP-hardness of the proposed problem followed by the proof of the monotone and submodular properties of the objective function. Our model is generalizable to other influence maximization problem by using a different influence diffusion model. The analysis result shows that the classical models (e.g. *IC*) are special cases of our model.
4. Considering the coverage of influence diffusion, we take the first step to explore the relationship between the breadth and depth of influence and propose the model *BICOT*. Specifically, in the extended version of our model, we use the number of communities to measure the breadth of the influence, which is novel.
5. The experiment results on several real data sets show that our solution can significantly improve the practicability and accuracy against several baseline methods. Especially on the aspect of influence spreading range.

The rest of the paper is organized as follows. Section 2 reviews the related works. Section 3 presents the preliminaries and problem definition, then we introduce our model with analysis and the algorithm in Section 4. The evaluation results based on real and synthetic data sets are shown in Section 5. Section 6 concludes the paper.

2. Related work

To maximize influence in OSNs, the *IC* model (Kempe et al., 2003) and another threshold model *LT* together with their extensions set the foundation for most of the existing cascading algorithms. Since Kempe et al. (2003) formulated the influence maximization problem as an optimization problem, a series of empirical studies have been performed on influence learning (Goyal et al., 2010; Saito et al., 2013; Zhang et al., 2014), algorithm optimizing (Tang et al., 2009; Goyal et al., 2011, 2005), scalability promoting (Wang et al., 2012; Chen et al., 2010), and influence of group conformity (Tang et al., 2013). Leskovec et al. (2007) modeled the outbreak detection problem and proved that the influence maximization problem is a special case of their new problem. A Cost-Effective Lazy Forward (*CELF*) scheme is proposed which uses the submodular property achieving 700 times speedup in selecting seed vertices compared with the basic greedy algorithm (Kempe et al., 2003). As indicated in Chen et al. (2010), *CELF* still faces the serious scalability problem. Therefore, Chen et al. proposed some new heuristics algorithms based on the arborescence structure which could handle million-sized graphs. The proposed algorithm spreads influence as the greedy algorithm while is more than six orders of magnitude faster than the greedy one. In Jung et al., the authors proposed algorithm *IRIE* where *IR* is for influence ranking and *IE* is for influence maximization in both the classical *IC* model and the extension *IC-N* model considering negative opinions (Chen et al., 2011). They claimed that their algorithms scale better than *PMIA* (Chen et al., 2010) with up to two orders of magnitude speedup and significant savings on memory usage, while maintaining the same or even better influence.

Besides the fundamental influence maximization problem and several variants mentioned above, there are two kinds of previous works related to ours: dynamic network models (He et al.) and structural analysis for influence diffusion. The phenomena of time delay in influence diffusion have been explored in statistics. Timeliness concerned by us, different from time decay, emphasizes more on the delivery time of influence. The observation in Iribarren and Moro (2009) shows that the heterogeneity of human activities has an important effect on the influence diffusion. Dinh et al. (2012) modeled influence maximization by limiting the influence of nodes that are within d hops from the seeding for

[2] Depth might result in "rendezvous problem", which is a term from mathematics to state the overcrowded of seeds selection.