Information Sciences 355–356 (2016) 110–126

Contents lists available at ScienceDirect

Information Sciences

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

Finding top-*k* influential users in social networks under the structural diversity model[☆]

Wenzheng Xu[a,*], Weifa Liang[b], Xiaola Lin[c], Jeffrey Xu Yu[d]

[a] College of Computer Science, Sichuan University, Chengdu, 610065, PR China
[b] Research School of Computer Science, The Australian National University, Canberra, ACT 0200, Australia
[c] School of Data and Computer Science, Sun Yat-Sen University, Guangzhou, 510006, PR China
[d] Department of Systems Engineering and Engineering Management, The Chinese University of Hong Kong, Hong Kong

ARTICLE INFO

Article history:
Received 24 June 2015
Revised 10 March 2016
Accepted 16 March 2016
Available online 24 March 2016

Keywords:
Influence maximization
Structural diversity model
Social networks
Approximation algorithm

ABSTRACT

The influence maximization problem in a large-scale social network is to identify a few influential users such that their influence on the other users in the network is maximized, under a given influence propagation model. One common assumption adopted by two popular influence propagation models is that a user is more likely to be influenced if more his/her friends have already been influenced. This assumption recently however was challenged to be over simplified and inaccurate, as influence propagation process typically is much more complex than that, and the social decision of a user depends more subtly on the network structure, rather than how many his/her influenced friends. Instead, it has been shown that a user is very likely to be influenced by structural diversities of his/her friends. In this paper, we first formulate a novel influence maximization problem under this new structural diversity model. We then propose a constant approximation algorithm for the problem. We finally evaluate the effectiveness of the proposed algorithm by extensive experimental simulations, using different real datasets. Experimental results show that the users identified from a social network by the proposed algorithm have much larger influence than that found by existing algorithms.

© 2016 Elsevier Inc. All rights reserved.

1. Introduction

The last decade experienced the exponential growth of a variety of online social networks such as Facebook, Twitter, LinkedIn, etc [20]. A recent snapshot of the friendship network Facebook showed that there are over 1 billion users in it [14]. Not only are these social networks effective tools for people to connect their friends and share interests and backgrounds, but also they now become powerful information dissemination and marketing platforms to allow information, ideas, fads, and political opinions to spread to a large population economically via the so called “word-of-mouth” exchanges [1,2,15,23,25,31,32,34,35]. For example, suppose that one company would like to market a new product and hopes the product will be adopted by a large fraction of users in a social network. To market the product economically, the company initially targets a few “influential” users in the network by giving them free product samples. These users then recommend the product to their friends, and some of their friends will accept the product and recommend the product to their friends,

☆ This document is a collaborative effort.

* Corresponding author. Tel.: +86 13880337745.
E-mail addresses: wenzheng.xu3@gmail.com (W. Xu), wliang@cs.anu.edu.au (W. Liang), linxl@mail.sysu.edu.cn (X. Lin), yu@se.cuhk.edu.hk (J.X. Yu).

http://dx.doi.org/10.1016/j.ins.2016.03.029
0020-0255/© 2016 Elsevier Inc. All rights reserved.

and so on. As a result, this triggers a cascade of influence propagation and many users in the network will accept the product ultimately [21]. This marketing strategy usually is referred to as *viral marketing*. One well-known real story of viral marketing is the commercial success of the Hotmail company in the early 1990s, which made the company become the number-one e-mail provider within only 18 months [19]. Thus, a fundamental research topic in social network sciences is to effectively and efficiently identify a very few influential individuals in a large-scale social network for information diffusion.

In their seminal paper, Kempe et al. [21] formalized information diffusion in a large social network as the *influence maximization problem*, which is defined as follows. Given an integer k (as the budget), the problem is to find k "seeds" (i.e., source nodes) in the network such that the expected number of activated nodes eventually by these k seeds is maximized, assuming that an influence propagation model is given, where the activation of a node means that the node accepts the recommendation. Following their work, several studies have been conducted in the past several years [3–10,16–18,21,22,30,33]. Most of these studies adopt the two popular influence propagation models: *the independent cascade model* and *the linear threshold model* [8,21]. One common property of these two models is that for a given user, the more his/her friends recommend a product to the user, the more likely the user will accept the product and recommend the product to his/her friends. This influence propagation is similar to the epidemic diseases spread, i.e., the probability of a user being influenced monotonically grows with the number of his/her friends whom have already been affected. However, these two models have recently been challenged by Ugander et al. [28] to be over simplified and thus inaccurate, as influence propagation process typically is more complex, and the social decision of a user depends more subtly on the network structure, rather than how many his/her influenced friends. Instead, they studied two influence propagation processes in Facebook: the process whereby a person joins Facebook in response to an invitation email from an existing Facebook user; and the process that a user becomes an engaged user after joining Facebook. They found through empirical analysis that the chance of a user accepting a recommendation is positively correlated with the number of connected components in the induced graph by the neighbors, rather than the number of neighbors, of the user, where each connected component represents a distinct social context of the user in the network and the multiplicity of social contexts is referred to as *the structural diversity*. They showed that a person is more likely to join Facebook if he/she receives more invitation emails from his/her friends with distinct social contexts, e.g., families, workmates, classmates, etc. On the other hand, surprisingly, they also demonstrated that once the number of connected components is controlled (or fixed), a person is less likely, rather than more likely, to join Facebook if more friends invite him/her. They concluded neither the number of friends inviting the user nor the number of connections among his/her friends will determine his/her acceptance probability. Instead, it is the number of connected components (structural diversity) derived by his/her neighbors that captures his/her acceptance probability. We term this model as the structural diversity model, which has been empirically demonstrated to be able to accurately capture the propagation progress of influence.

Motivated by the seminal work of Ugander et al. [28], in this paper we study the influence maximization problem under the structural diversity model, where the decision of a user depends on different groups of neighbors, rather than the number of neighbors. It thus poses a challenging problem, that is, how to incorporate the structural diversity of each user into the influence maximization problem. Existing algorithms for the problem under the two widely adopted influence propagation models thus are not applicable, and new algorithm is desperately needed. In this paper, we propose a probabilistic approximation algorithm for the problem under the structural diversity model. Interestingly, we show that the problem is equivalent to a slightly different influence maximization problem in an auxiliary graph under the independent cascade model. The contributions of the paper can be summarized as follows.

- We first formulate a novel influence maximization problem under the structural diversity model, in which a user is more likely to accept a recommendation if more his/her friends with distinct social contexts make the recommendation to the user.
- We then devise a $(1-\frac{1}{e}-\epsilon)$-approximation algorithm with probability $1-\alpha$ for the problem, which takes $O(\epsilon^{-2}\alpha^{-1}n^2(m+n)k^3\log k)$ time, improving the time complexity $O(\epsilon^{-2}n^3(m+n)k^3\log\frac{nk}{\alpha})$ of the state-of-the-art approximation algorithm by a factor of αn [8] (pp.43), where e is the base of the natural logarithm, both ϵ and α are given constants with $0 < \epsilon < 1$ and $0 < \alpha < 1$, and n and m are the number of nodes and links in the social network, respectively.
- We finally evaluate the effectiveness of the proposed algorithm by extensive experimental simulations, using different real datasets. Experimental results show that the proposed algorithm outperforms existing algorithms.

The rest of the paper is organized as follows. Section 2 reviews related work. Section 3 introduces preliminaries and formulates the influence maximization problem under the structural diversity model. Section 4 devises a probabilistic approximation algorithm for the problem, and analyzes its time complexity, approximation ratio, and success probability. Section 5 empirically evaluates the proposed algorithm, and Section 6 concludes the paper and points out potential future work.

2. Related work

In this section, we review related work on the influence maximization problem under different influence propagation models. Domingos et al. [13,26] were the first to study the influence maximization problem as an algorithmic problem, and proposed an influence propagation model which specifies a joint distribution over all nodes' behavior globally. Kempe

Download English Version:

https://daneshyari.com/en/article/392581

Download Persian Version:

https://daneshyari.com/article/392581

Daneshyari.com