Information Systems 46 (2014) 1–23

Contents lists available at ScienceDirect

Information Systems

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

Viral marketing for dedicated customers

Cheng Long [a,*], Raymond Chi-Wing Wong [b]

[a] Rm 4212, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong
[b] Rm 3542, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong

ARTICLE INFO

Article history:
Received 17 August 2013
Accepted 6 May 2014
Recommended by: F. Carino
Available online 21 May 2014

Keywords:
Viral marketing
Interest-Specified
Dedicated customers

ABSTRACT

Viral marketing has attracted considerable concerns in recent years due to its novel idea of leveraging the social network to propagate the awareness of products. Specifically, viral marketing first targets a limited number of users (seeds) in the social network by providing incentives, and these targeted users would then initiate the process of awareness spread by propagating the information to their friends via their social relationships. Extensive studies have been conducted for maximizing the awareness spread given the number of seeds (the *Influence Maximization* problem). However, all of them fail to consider the common scenario of viral marketing where companies hope to use as few seeds as possible yet influencing at least a certain number of users. In this paper, we propose a new problem, called *J-MIN-Seed*, whose objective is to minimize the number of seeds while at least *J* users are influenced. *J*-MIN-Seed, unfortunately, is NP-hard. Therefore, we develop an approximate algorithm which can provide error guarantees for *J*-MIN-Seed. We also observe that all existing studies on viral marketing assume that *all* users in the social network are *of interest* for the product being promoted (i.e., all users are potential consumers of the product), which, however, is not always true. Motivated by this phenomenon, we propose a new paradigm of viral marketing where the company can specify which types of users in the social network are of interest when promoting a specific product. Under this new paradigm, we re-define our *J*-MIN-Seed problem as well as the *Influence Maximization* problem and design some algorithms with provable error guarantees for the new problems. We conducted extensive experiments on real social networks which verified the effectiveness of our algorithms.

1. Introduction

Viral marketing is an advertising strategy that takes the advantage of the effect of "word-of-mouth" among the relationships of individuals to promote a product. Instead of broadcasting to a *massive* number of users directly as existing advertising methods [1] do, viral marketing targets a *limited* number of *initial* users (by providing *incentives*) and utilizes their social relationships, such as friends, families and co-workers, to further spread the awareness of the product among individuals. Each individual who gets the awareness of the product is said to be *influenced*. The number of all influenced individuals corresponds to the *influence* incurred by the initial users. According to some recent research studies [2], people tend to trust the information from their friends, relatives or families more than that from general advertising media like TVs. Hence, it is believed that viral marketing is one of the most effective marketing strategies [3]. In fact, extensive commercial instances of viral marketing succeed in real life. For example, *Nike Inc.* used social networking websites such as *orkut.com* and *facebook.com* to market products successfully [4].

The propagation process of viral marketing within a social network can be described in the following way.

* Corresponding author. Tel.: +852 9512 9172; fax: +852 2358 1477.
E-mail address: clong@cse.ust.hk (C. Long).

http://dx.doi.org/10.1016/j.is.2014.05.003
0306-4379/

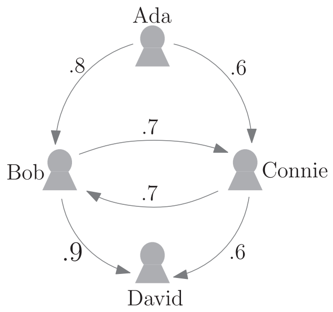

Fig. 1. Social network for *J*-MIN-Seed.

At the beginning, the advertiser selects a set of initial users and provides these users incentives so that they are willing to initiate the awareness spread of the product in the social network. We call these initial users *seeds*. Once the propagation is initiated, the information of the product *diffuses* or *spreads* via the relationships among users in the social network. A lot of models about how the above diffusion process works have been proposed [5–10]. Among them, the *Independent Cascade Model* (*IC model*) [5,6] and the *Linear Threshold Model* (*LT model*) [7,8] are the two that are widely used in the literature. In the social network, the IC model simulates the situation where for each influenced user u, each of its neighbors has a probability to be influenced by u, while the LT model captures the phenomenon where each user's tendency to become influenced increases when more of its neighbors become influenced.

1.1. Minimizing seed set

Consider the following scenario of viral marketing. A company wants to advertise a new product via viral marketing within a social network. Specifically, it hopes that at least a certain number of users, says J, in the social network must be influenced yet the number of seeds for viral marketing should be as small as possible. Clearly, the above problem can be formalized as follows. *Given a social network* $G(V,E)$*, we want to find a set of seeds such that the size of the seed set is minimized and at least J users are influenced at the end of viral marketing.* We call this problem *J-MIN-Seed*.

We use Fig. 1 to illustrate the main idea of the *J*-MIN-Seed problem. The four nodes shown in Fig. 1 represent four members in a family, namely Ada, Bob, Connie and David. In the following, we use the terms “nodes” and “users” interchangeably since they correspond to the same concept. The directed edge (u,v) with the weight of $w_{u,v}$ indicates that node u has the probability of $w_{u,v}$ to influence node v for the awareness of the product. Now, we want to find the *smallest* seed set such that *at least* 3 nodes can be influenced by this seed set. It is easy to verify that the expected influence incurred by seed set {Ada} is about 3.57[1] under the IC model and no smaller seed set can incur at least three influenced nodes. Hence, seed set {Ada} is our solution.

J-MIN-Seed can be applied to most (if not all) applications of viral marketing. Intuitively, *J*-MIN-Seed asks for the minimum cost (seeds) while satisfying an explicit requirement of revenue (influenced nodes). Clearly, in the mechanism of viral marketing, a seed and an influenced node correspond to cost and potential revenue of a company, respectively, because the company has to *pay* the seeds for incentives, while an influenced node might bring revenue to the company. In many cases, companies face the situation where the goal of revenue has been set up explicitly and the cost should be minimized. Thus, *J*-MIN-Seed meets these companies' demands.

No existing studies have been conducted for *J*-MIN-Seed on the IC model and the LT model. even though it plays an essential role in the viral marketing field. First, most existing studies related to viral marketing focus on maximizing the influence incurred by a certain number of seeds, says k [11–16]. Specifically, they aim at maximizing the number of influenced nodes when only k seeds are available. We denote this problem by *k-MAX-Influence*. Clearly, *J*-MIN-Seed and *k*-MAX-Influence have different *goals* with different given *resources*. Second, though a few studies [17,18] have been done for minimizing the number of seeds while influencing a certain number of users, which is called the *Target Set Selection* (TSS) problem, they adopt the *Deterministic Linear Threshold* (DLT) model as the underlying diffusion model. In contrast, we consider the *Independent Cascade* (IC) model and the *Linear Threshold* (LT) model as the underlying diffusion models for our *J*-MIN-Seed problem. As will be shown later, both the IC model and the LT model enjoy a nice property (*submodularity*), which, however, is not owned by the DLT model, and based on this property, we design an approximate algorithm for *J*-MIN-Seed with good error guarantees. Mainly, [17,18] provide some results about the hardness of approximating the TSS problem, which, do not apply to the *J*-MIN-Seed problem.

Naïvely, we can solve the *J*-MIN-Seed problem [19] by adapting an existing algorithm for *k*-MAX-Influence. Let k be the number of seeds. We set $k=1$ at the beginning and increment k by 1 at the end of each iteration. For each iteration, we use an existing algorithm for *k*-MAX-Influence to calculate the maximum number of nodes, denoted by I, that can be influenced by a seed set with the size equal to k. If $I \geq J$, we stop our process and return the current number k. Otherwise, we increment k by 1 and perform the next iteration. However, this naïve method is very time-consuming since it issues the existing algorithm for *k*-MAX-Influence many times for solving *J*-MIN-Seed. Note that *k*-MAX-Influence is NP-hard [12]. Any existing algorithm for *k*-MAX-Influence is computationally expensive, which results in this naïve method with a high

[1] The expected influence incurred by seed set {Ada} on Bob is $1-(1-0.8)\cdot(1-0.6\cdot 0.7)=0.884$ (note that Ada can influence Bob either directly with the probability equal to 0.8 or via Connie with the probability equal to 0.6∗0.7). Similarly, we can compute the expected influence incurred by {Ada} on other users. Overall, the influence incurred by {Ada} is equal to 3.57.