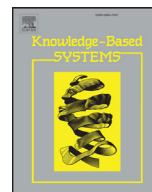




Contents lists available at ScienceDirect

Knowledge-Based Systems

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

Community-based influence maximization in social networks under a competitive linear threshold model

Arastoo Bozorgi^{a,*}, Saeed Samet^b, Johan Kwisthout^c, Todd Wareham^a

^a Department of Computer Science, Memorial University, St. John's, NL A1B 3X9, Canada

^b eHealth Research Unit, Faculty of Medicine, Memorial University, St. John's, NL A1B 3V6, Canada

^c Donders Institute for Brain, Cognition, and Behaviour, Radboud University, Nijmegen, The Netherlands

ARTICLE INFO

Article history:

Received 25 September 2016

Revised 19 July 2017

Accepted 24 July 2017

Available online xxx

Keywords:

Competitive influence maximization

Linear threshold model

Community detection

Influence maximization

Social network

ABSTRACT

The main purpose in influence maximization, which is motivated by the idea of viral marketing in social networks, is to find a subset of key users that maximize influence spread under a certain propagation model. A number of studies have been done over the past few years that try to solve this problem by considering a non-adversarial environment in which there exists only one player with no competitor. However, in real world scenarios, there is always more than one player competing with other players to influence the most nodes. This is called competitive influence maximization. Motivated by this, we try to solve the competitive influence maximization problem by proposing a new propagation model which is an extension of the Linear Threshold model and gives decision-making ability to nodes about incoming influence spread. We also propose an efficient algorithm for finding the influential nodes in a given social graph under the proposed propagation model which exploits the community structure of this graph to compute the spread of each node locally within its own community. The aim of our algorithm is to find the minimum number of seed nodes which can achieve higher spread in comparison with the spread achieved by nodes selected by other competitor. Our experiments on real world and synthetic datasets show that our approach can find influential nodes in an acceptable running time.

© 2017 Elsevier B.V. All rights reserved.

1. Introduction

The effect of online social networks (OSNs) in our daily life is undeniable as they have introduced new ways of communication and serve as a medium for propagating news, ideas, thoughts and any type of information. Such information can propagate via links between people, which leads to word-of-mouth advertising and its famous application, viral marketing. In viral marketing, the owner of a product gives free or discounted samples of a product to a group of people to gain a large number of adoptions through the word-of-mouth effect. The influence maximization problem, which is motivated by the idea of viral marketing, was introduced by Kempe et al. [1] as finding a subset $S \subseteq V$ containing of k nodes in a graph $G = (V, E)$, such that the spread of influence from S will be maximized. There exists a huge amount of work on solving the influence maximization problem [1–5]. Most of these works assume that there is only one party trying to find influential users in the social network. However, in the real world, multiple

parties typically compete simultaneously with similar products. This is called competitive influence maximization (CIM). Recently, several works have tried to solve the CIM problem [6–13] by proposing new propagation models which are extensions of Linear Threshold and Independent Cascade models [1] or the Distance-based and Wave-propagation models [14].

In this paper, we examine the CIM problem from the follower's perspective and propose a new propagation model called DCM (Decidable Competitive Model) which is an extension of the Linear Threshold model. In DCM, each node has the ability to think about the incoming influence spreads from its neighbors for d timesteps and then decide to be activated by the neighbor with the majority of adoption.

In real social networks, people interact with each other based on common interests and strong ties between themselves. Such strong ties between individuals create community structures in social networks, which in turn allow information to circulate within these networks at a high velocity. We propose an algorithm called Competitive Influence Improvement (CI2) which finds the minimum number of influential nodes within their respective communities. Closely related to our work are [15–19] which also

* Corresponding author.

E-mail addresses: ab1502@mun.ca (A. Bozorgi), saeed.samet@med.mun.ca (S. Samet), j.kwisthout@donders.ru.nl (J. Kwisthout), harold@mun.ca (T. Wareham).

exploit community structure within social networks to find influential nodes.

Contribution. Our major contributions in this research are summarized as follows:

- We propose the DCM propagation model, the primary intent of this work, which gives decision-making power to nodes based on incoming influence in a competitive version of the LT propagation model.
- We prove the *NP*-hardness of competitive influence improvement under the DCM model.
- We propose the CI2 algorithm to find the minimum number of the most influential nodes for a competitor C_2 . This algorithm uses knowledge of the nodes selected by a competitor C_1 so that C_2 can achieve more influence spread by spending less budget. Computing the spread of seed nodes is done locally inside communities of the input graph, which results in a substantial decrease in running time.
- We conduct experiments using three real and three synthetic datasets to show that CI2 can find influential nodes in an acceptable running time. Synthetic datasets are generated with the same number of nodes and edges but different community structures in order to track the effect of community structure of networks on our approach. Also, we consider the effect of the algorithm which finds the seed nodes for the first competitor on the seed nodes which will be selected by the second competitor by conducting different experiments which use well-known algorithms [15,20,21] to extract the first competitor's seed set.

Organization. In Section 2, we review some background knowledge to enable a better understanding of the upcoming concepts. In Section 3, we describe our Linear-Threshold-based propagation model, prove that competitive influence improvement under this model is *NP*-hard, and propose our CI2 algorithm. Section 4 describes the experiments performed with real and synthetic data to evaluate the proposed approach. Finally, in Section 5, we give our conclusions and directions for future work.

2. Background and related work

In [1], Kempe et al. introduced two propagation models to address the influence maximization problem, the Linear Threshold (LT) and Independent Cascade (IC) models. In both models, a threshold value $\theta \in [0, 1]$ is assigned to each node and each node can be active or inactive. Also, each edge from node u to node v has an influence weight $p_{u,v} \in (0,1]$. At first, all nodes are inactive except the nodes in set S which have been activated before as seed nodes and the propagation process is started from them. In LT, an inactive node u can be activated at time t if $f_v(S) > \theta_v$, where S stands for v 's neighbors which are activated at time $t-1$. As is mentioned in [1], the value of f_v is initialized as

$$f_v(S) = \sum_{u \in S} b_{v,u}$$

where $b_{v,u}$ is the weight of edge (v,u) . In the LT model, the sum of all edge weights between v and its neighbors should be less than 1 [1].

In IC, the activation process is the same as that in LT except that in IC, an activated node u has only one chance to activate its inactive neighbor v with probability $p_{u,v}$.

Community structure. Communities are subsets of nodes in the graph with more edges between them and fewer edges to nodes in different communities [22]. Community detection is formulated as a clustering problem – that is, given the full graph $G = (V, E)$, partition the vertex set into k subsets S_1, S_2, \dots, S_K , such that $\bigcap_{i=1}^k S_i = \emptyset$ and $\bigcup_{i=1}^k S_i = V$. A quality metric $Q(S_1, \dots, S_K)$ is

defined over the partitions and a community detection algorithm will try to find a partition that maximizes or minimizes Q depending on its nature. This is for non-overlapping community detection and one can simply remove the constraint $\bigcap_{i=1}^k S_i = \emptyset$ to get the overlapping version [23].

Competitive influence maximization. Recently, several works have tried to solve the competitive influence maximization by introducing new propagation models to simulate the competitive manner of the competitors which are mostly an extension of the LT or IC models. Some efforts such as [14,24] look to this problem from the follower's perspective, i.e. they assume that there are two competitors trying to find some influential nodes and the second competitor starts his process with knowledge of the seed nodes selected by the first competitor and tries to find some new seed nodes other than the ones selected by the first competitor to achieve more influence spread. In some other works such as [6,7] one competitor tries to block the effect of the other competitor. In [8], Lu et al. solve the competitive influence maximization problem from the host's perspective, i.e. the owner of the social network is responsible for fairly allocating some specific number of seed nodes to the competitors. In the next section, we explain [8,25] in more details as they are more related to our work.

3. Propagation model and algorithm

In this section, we introduce the DCM propagation model (which is an extension of the LT model), compare DCM with the Weighted-proportional (WPCLT) [25] and *K*-LT [8] models, and prove the *NP*-hardness of competitive influence improvement under DCM. Finally, we introduce the CI2 algorithm to find the influential nodes in a social network under our competitive propagation model.

3.1. DCM propagation model

In the DCM propagation model, each node can be in one of the following states: *inactive*, *thinking*, *active*⁺ or *active*⁻. Suppose there are two competitors who try to advertise for their products over a social network. We denote the first competitor with the + sign and the second competitor with the - sign and each node v , picks a threshold value θ_v uniformly at random from [0,1]. Let S_1 be the seed set selected by the first competitor and S_2 be the seed set selected by the second one. At first all nodes except those in the seed set are *inactive*. The activation process of node v is as follows: at time $t > 1$ if the total incoming influence weight from the in-neighbors of v which are active ($N_{active}^{in}(v)$) reaches the threshold value of v , its state changes to *thinking*, which means the state of node v would be changed with probability

$$\sum_{u \in N_{active}^{in}(v)} p_{u,v} \geq \theta_v \quad (1)$$

Node v remains in *thinking* state after this state change for d timesteps and after that, it decides to become *active*⁺ or *active*⁻ based on the maximum total incoming influence weight from its in-neighbors. Let A_{t+d}^+ be the set of in-neighbor nodes of v with state *active*⁺ and A_{t+d}^- be the set of in-neighbor nodes of v with state *active*⁻ at time $t+d$. The state of node v changes from *thinking* to *active*⁺ or *active*⁻ as follows:

$$v_{state} = \begin{cases} active^+, & \text{if } \sum_{u \in A_{t+d}^+} p_{u,v} > \sum_{u \in A_{t+d}^-} p_{u,v} \\ active^-, & \text{otherwise} \end{cases} \quad (2)$$

In the WPCLT model, which was proposed by Borodin et al. [25], the state of a node v changes to *active*⁺ with probability $\sum_{u \in A_{t-1}^+} p_{u,v} / \sum_{u \in A_{t-1}} p_{u,v}$. This means that a node v would be activated as *active*⁺ (*active*⁻) at time t with probability equal to the

Download English Version:

<https://daneshyari.com/en/article/4946072>

Download Persian Version:

<https://daneshyari.com/article/4946072>

[Daneshyari.com](https://daneshyari.com)