Physica A 416 (2014) 467–480

Contents lists available at ScienceDirect

Physica A

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

Coritivity-based influence maximization in social networks

Yanlei Wu*, Yang Yang, Fei Jiang, Shuyuan Jin, Jin Xu

School of Electronics Engineering and Computer Science, Peking University, 100871, Beijing, People's Republic of China

HIGHLIGHTS

- Coritivity theory is applied to influence maximization problem.
- The influence spread using Coritivity algorithm is satisfactory.
- The convergence rate in Coritivity method is excitingly fast.
- Reasons of different performances of various algorithms are analyzed.

ARTICLE INFO

Article history:
Received 7 January 2014
Received in revised form 12 August 2014
Available online 16 September 2014

Keywords:
Influence maximization
Core and coritivity
Diffusion models

ABSTRACT

Influence maximization problem is about finding a small set of nodes from the social network as seed set so as to maximize the range of information diffusion. In this paper, the theory of coritivity and method of finding core nodes in networks are introduced to deal with this problem. From the perspective of network structure, core nodes are the important ones to network connectivity and is a competitive measurement of node influence. By finding the core of the network through coritivity we can finally get the initial active nodes required in the influence maximization problem. We compare this method with other conventional node-selection approaches in USAir97 and HEPTH datasets. Experimental results demonstrate that: (a) the coritivity-based method achieves large influence spread in all the diffusion models we use, and (b) the proposed method converges fast in all cases we consider.

1. Introduction

As one aspect of social network analysis, the study of information diffusion is of great value. With good knowledge of the mechanism in which beneficial message propagates through the network, we can better understand how to make it spread more widely or more quickly. Getting to know how the harmful information diffuses helps us to detect it early or effectively prevent it from large-scale spread.

The problem we study here is the information maximization problem. We use graph to model the social network. Each individual is represented as a node and edges stand for relationships between them. When one piece of information appears in the network, it can pass from one node to another through the edges between them. The information maximization problem is about how to select k nodes from the network as the initial active node set S and then propagate through the network such that we can get the most active nodes $\sigma(S)$ at the end of the diffusion. The solution of this problem can help us to achieve the best diffusion result with the least effort. Thus it has essential practical value, such as in product promotion.

To solve the influence maximization problem, researchers have done pretty much work. After Pedro Domingos and Matt Richardson [1,2] first studied the influence maximization problem as an algorithmic problem with probabilistic method,

* Corresponding author. Tel.: +86 18811418125.
E-mail address: notears2@yeah.net (Y. Wu).

http://dx.doi.org/10.1016/j.physa.2014.09.010
0378-4371/

David Kempe et al. [3] systematically formulated it as a discrete optimization problem. They came up with the greedy algorithm to get the approximately optimal solution. However, the way to get the scope of diffusion in the primary greedy algorithm is not efficient. To solve this issue, Jure Leskovec [4] put forward the "lazy-forward" method to speed up the process. Masahiro Kimura [5] presented SPM and SP1M based on IC model to efficiently calculate the scope of diffusion. Wei Chen [6] also put forward MIA and PMIA model for the similar purpose. Community-based [7] and path-based [8] methods are proposed as well. On the other hand, heuristic methods are also used to tackle the influence maximization problem. Wei Chen [9] came up with the "Degree Discount" method in independent cascade model (ICM) that improves the computing efficiency effectively and produces a diffusion result close to that of the greedy. Qingye Jiang [10] utilized the simulated annealing to heuristically get the requested nodes. Kyomin Jung [11] proposed the IRIE (Influence Rank Influence Estimation) method to tackle the problem scalably and robustly.

Besides, there are also methods that rank the node influence according to the individual nodes' structural attributes. For example, betweenness and distance centrality are simple measures of this kind. But they suffer from the intersections between the selected nodes' influence range. Therefore, many heuristics are proposed to improve the results. Linyuan Lu etc. [12] proposed the LeaderRank algorithm based on PageRank to suit it better to the social network. Bonan Hou etc. [13] combined the measure of degree, betweenness centrality together with k-core and came up with the notion of all-around nodes. An Zeng and Cheng-Jun Zhang [14] presented the MDD measure based on the k-shell method to improve the diffusion results. Jian-Guo Liu etc. [15] took into account the shortest distance from a target node to node set with highest k-core and further differentiated between the nodes with the same k-core value. Duan-Bing Chen etc. [16] took the clustering coefficient into consideration and proposed the ClusterRank algorithm.

In this paper, we propose a novel method based on coritivity theory to deal with the influence maximization problem. Coritivity is one measure of the network connectivity and reflects the network structure. Therefore, it can be utilized to measure the nodes' influence in diffusing information to some extent. So we can apply the coritivity theory to the influence maximization problem to get k initial active nodes. Our contributions are mainly as follows:

(a) The proposed method performs well in diffusion range as well as convergence rate. As far as we know, we are the first to utilize the coritivity theory to handle the influence maximization problem.
(b) We analyze the reasons for different performances of the node-selection methods in detail through comparisons.

The rest of this paper is organized as follows: in Section 2, we formulate the basic concepts and ideas in coritivity theory that will be used in our algorithms; Section 3 is about the detailed algorithms regarding how to get the core of a network and then the k initial active nodes; Section 4 deals with the diffusion models we use in this paper; in Section 5 we provide the experimental results, show and analyze the performances of different node selection methods; in Section 6, we analyze the limitations of our method and propose some optimizations; in the last part, we make the conclusion and point out our future directions of study.

2. Basic coritivity theory

In real networks, there are always some entities locating at important positions or playing crucial roles. Removing these entities will lead the network to an unstable state. These entities are called cores of the network. To study cores and their effect, we introduce coritivity theory into networks.

Coritivity theory measures the importance of a set of nodes by the number of connected components showing up after deleting the nodes and their incident edges from the graph. Given an undirected unweighted connected graph G, with $V(G)$ and $E(G)$ representing the node set and the edge set, the coritivity of graph G, $h(G)$, is defined as

$$h(G) = \max\{\omega(G - S) - |S|; S \in C(G)\} \tag{1}$$

where $C(G)$ denotes the collection of cut sets of G, and $\omega(G - S)$ is the number of components of graph $G - S$. For $S \subseteq V(G)$, $G - S$ denotes the graph obtained by deleting from G node set S together with all edges incident with any node in S. $|S|$ stands for the number of nodes in S. Moreover, if $S' \in C(G)$ and satisfies

$$h(G) = \omega(G - S') - |S'|. \tag{2}$$

S' is called a core of graph G. This definition of core and coritivity implies that each network has a unique coritivity value but may have many different cores. Each core is a cut set that satisfies Eq. (2).

Coritivity is a measurement to quantify the importance of the core in networks. Naturally, given a network, the most fundamental problem is to calculate the core and coritivity. Thus, we will introduce some notions that is necessary for working out the coritivity. Again, given any undirected, unweighted and connected graph G, S is a core of G. Note that S is a set of nodes, not one specific node. Then if S^* is a non-empty subset of S, we call S^* a subcore of graph G. Moreover, if $S^* \subsetneqq S$, then S^* is a real subcore of graph G and we call $S - S^*$ a complementary core of real subcore S.

Another useful notion in coritivity theory is generalized hanging node. We use the symbol $N(v)$ to denote the neighbor node set of any node v. For node set S, we call any node v a generalized hanging node of S, if $N(v) \subseteq S$. The set of generalized hanging nodes of a subset S in graph G, $\Gamma^+(S)$, is defined as $\Gamma^+(S) = \{v; v \notin S, N(v) \subseteq S\}$. Lastly, to introduce a notion, normal subcore, which is quite useful for the coritivity algorithms in the next section, we give a theorem about the relation between core and subcore.