Theoretical Computer Science 556 (2014) 13–24

Contents lists available at ScienceDirect

Theoretical Computer Science

www.elsevier.com/locate/tcs

Mining hidden links in social networks to achieve equilibrium ☆

Huan Ma [a], Zaixin Lu [b], Deying Li [a,*], Yuqing Zhu [c], Lidan Fan [d], Weili Wu [d]

[a] School of Information, Renmin University of China, 100872, PR China
[b] NSF Center for Research on Complex Networks, Texas Southern University, Houston, TX, 77004, USA
[c] Computer Science Department, California State University Los Angeles, Los Angeles, CA, 90032, USA
[d] Department of Computer Science, University of Texas at Dallas, Richardson, TX, 75080, USA

ARTICLE INFO

Article history:
Received 18 February 2014
Received in revised form 29 July 2014
Accepted 17 August 2014
Available online 22 August 2014

Keywords:
Mining hidden links
Vertex pair proximity
Approximation algorithm
Nash equilibrium
Social networks

ABSTRACT

Although more connections between individuals in a social network can be identified with the development of high techniques, to obtain the complete relation information between individuals is still hard due to complex structure and individual privacy. However, the social networks have communities. In our work, we aim at mining the invisible or missing relations between individuals within a community in social networks. We propose our algorithm according to the fact that the individuals exist in communities satisfying Nash equilibrium, which is borrowed from game-theoretic concepts often used in economic researches. Each hidden relation is explored through the individual's loyalty to their community. To the best of our knowledge, this is the first work that studies the problem of mining hidden links from the aspect of Nash equilibrium. Eventually we confirm our approach's superiority from extensive experiments over real-world social networks.

1. Introduction

Along with increasing popularity of large, complex networks in computer science and physical domains, the social networks have received a considerable amount of attention from researchers for their theoretical interests and practical importance in real-world applications. In the social networks, the nodes represent individuals, and the links (edges) denote the interactions between the nodes. The most significant feature of social network is "community structure", which means the connections between individuals are dense within the same community while sparse across distinct communities, that is individuals in the same community share more common attributes, and identifying community structure and learning the microscopical relation between individuals in a community can help us to understand individuals' behaviors well, which has wide applications in social marketing [1,2], urban development [3], criminology [4] and so forth.

Take advantage of online networks, a large volume of data can be obtained by people, however, in the process of collecting, gathering or recording information, some information may be lost for the complex relations between individuals, the data of information from real-world networks is often incomplete and inaccurate, here we call the lost or missing informa-

---

☆ This research was supported in part by the National Natural Science Foundation of China under grant 91124001, the Fundamental Research Funds for the Central Universities, and the Research Funds of Renmin University of China under grant 10XNJ032 and Shenzhen Strategic Emerging Industries Program with Grant No. ZDSY20120613125016389, and was also supported in part by National Science Foundation of USA under grants CNS 1016320 and CCF 0829993.

* Corresponding author.
E-mail address: deyingli@ruc.edu.cn (D. Li).

http://dx.doi.org/10.1016/j.tcs.2014.08.006

tion as hidden links. Particularly, most information is distributed within communities, and some information is distributed among communities, as a result, the hidden links often reside within communities. To get insight into microscopical individual behaviors, it is of great value for us to mine these lost links more accurately. Since the connections across communities are sparse, in our problem, we assume these links will never be lost. In other words, they are observable to the public. Obviously, our problem is established on the existing community structure.

With respect to the community detection problem, up to now, a large volume of works have been published, the readers may refer to [5–10]. Especially, Newman et al. proposed "modularity" in [5], which is used to measure the quality of identified communities. The larger the modularity, the higher the quality of the community structure. Most subsequent efforts are focused on the optimization of modularity.

To efficiently explore the hidden links between individuals within a community, we start from the aspect of community formation, which was interpreted as a game framework by Chen et al. in [11]. They explained this phenomenon from the point of human nature. Each agent is rational and selfish, and their behaviors of joining communities are associated with a gain function and a loss function, therefore, they will choose to join a community that maximizes their total interest (the gain of joining the community minus the loss of joining the community). Obeying this rule, all agents finally find their corresponding communities (each agent can join more than one community at the same time) with satisfying Nash equilibrium. Witnessing this phenomenon, we aim to apply the property into our problem by reasonable and proper modifications, which is different from [12], although they find the missing links in community, they don't use the property of community to find the links, and they use the common neighbor method.

In this paper, we assume that the community structure in a social network has been identified through existing techniques in community detection, and these communities reach Nash equilibrium status. Here, the concept of Nash equilibrium is a concrete measurement based on the neighbor distribution of each node, which is different from the one in [11]. Moreover, each individual is required to belong to one community, that is, the communities are disjoint. In each community, if no link is lost, then every individual has at least the same number of in-community neighbors as that of out-community neighbors. However, when some links between members within communities are missed, some individuals may become active, meaning that they have tendency to leave their current communities and join other communities. Therefore, we try to search the lost links from these active individuals. As for determining which links are the possible ones adjacent to these active individuals, we modify the techniques used in link prediction problem, which is a long-standing problem in social networks and has received extensive attentions from researchers.

Enormous efforts have been devoted to link prediction problem, please refer [13–19] for details. The primary goals of these works mainly focus on predicting missing links or future potential links between individuals over the entire network, and the techniques they often used are related with node topology structure or node attributes like age, hobby, location and so on. They have no target nodes and predict the links based on nodes in the entire network. However, in our work, we firstly identify the active nodes as the priority targets. Then, possible existing links adjacent to them are explored. In addition, in predicting links, we combine common neighbors within p-hops to measure the closeness between vertices.

To summarize, to our best knowledge, our paper is the first one to mine the hidden links from the aspect of game-theoretic framework. the main contribution is as follows.

- We prove the $\mathcal{NP}$-hardness of finding a Nash equilibrium when a graph has to be partitioned into two communities and the function used to measure the equilibrium of each node is only based on the degrees in communities.
- We propose a novel measurement modified from Nash equilibrium in [11] to determine the loyalty of each individual to their own communities, and define all the members who are observed having tendency to leave their current community as *active individuals*.
- By adopting link prediction measurements common neighbors [20] and $Act(\cdot)$ of each node which represents the probability that a node will leave its own community, we search possible hidden links adjacent to these *active individuals* in communities.

The remaining of this paper is organized as follows. In Section 2 we will review the existing work about community detection. In Section 3, we will formally define the game-theoretic framework in mining hidden links. In Subsection 3.1, we will give the concrete concept of Nash equilibrium in our paper and several approaches in link prediction problem. Moreover, we will prove the $\mathcal{NP}$-hardness of finding a Nash equilibrium on the simplest case where only degree is used as the measurement. In Subsection 3.4, we will present the formal definition of our problem and several relative terms. The Equilibrium based Hidden Link Mining (EHLM) Algorithm will be described and analyzed in Section 4. In Section 5, we will show the experimental settings and results. Finally, we will introduce possible future researches and give the conclusion in Section 6. Note that in this paper, we restrict our attention to unweighted and undirected networks.

2. Relative work in community identification

Large networks present attractive common properties, such as power-law degree distributions [21], high network transitivity [22] and the outstanding feature "community structure", meaning that edges between vertices in the same community are dense and sparse between different communities [23]. Identifying community structure in a social network provides people information to understand individual behaviors.