Chaos, Solitons and Fractals 108 (2018) 57–65

Contents lists available at ScienceDirect

Chaos, Solitons and Fractals

Nonlinear Science, and Nonequilibrium and Complex Phenomena

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

A novel trust-based community detection algorithm used in social networks

Xianhuan Chen[a,b], Chengyi Xia[c,d,*], Jin Wang[a,b]

[a] School of Management, Hefei University of Technology, Hefei 230009, Anhui, PR China
[b] Key Laboratory of Process Optimization and Intelligent Decision-Making, Ministry of Education, Hefei 230009, Anhui, PR China
[c] Tianjin Key Laboratory of Intelligence Computing and Novel Software Technology, Tianjin University of Technology, Tianjin 300384, PR China
[d] Key Laboratory of Computer Vision and System (Ministry of Education), Tianjin University of Technology, Tianjin 300384, PR China

ARTICLE INFO

Article history:
Received 2 November 2017
Revised 26 December 2017
Accepted 17 January 2018

Keywords:
Complex network
Online social network
Community structure
Detection strategy
Fitness evaluation
Trust mechanism

ABSTRACT

Mining the community structure is an important subject in the area of social network analysis, and detecting the hidden communities within the social networks will help to better understand the topological properties of the real-life networks. Meanwhile, community detection will be also helpful to monitor the public opinion, identify the opinion leaders and perform the personalized recommendation. In comparison with the simplex user ties or contents, considering the trust features from multiple users will provide a more comprehensive account of the linking relationship between users. To this end, we propose a novel non-overlapping community detection algorithm, which is based on the trust mechanism, to recognize the community structure in this paper. At first, we propose several definitions with regard to trust relationship between users to depict the trust strength, which includes the direct, indirect and mutual trust, and then the specific trust calculation method is provided to quantitatively describe the extent of trust. Secondly, starting from the trust relationship, we integrate the edge fitness and community fitness into the non-overlapping community detection and propose a novel trust-based algorithm to comprehensively leverage the trust among nodes to further mine the communities within the networks. Finally, to deeply analyze the analyze the performance, we take use of Lesmis and Gemo data sets to carry out extensive experiments, and the results show that, compared with other classical algorithms, the community based on the newly proposed algorithm features the higher trust cohesion on the condition that the structural cohesiveness of social network is fully satisfied. The current methods will be of significance to deeply understand and effectively find out the communities within realistic networks.

1. Introduction

In parallel with the rapid development of social network such as communication network, science collaboration network, online social network in the recent years, more and more scholars have realized that in the environment of social network there are not only explicit relations among users but also implicit relations demonstrated in the forms of friends similarity, demand similarity, browsing and purchasing similarity etc. The lack of face-to-face contact in online social interaction leads to the unstable construction of the social network. Meanwhile, more scholars have realized that in the environment of social networks, there are explicit relationships among users and implicit relationships demonstrated by friend similarity, demand similarity, browsing and purchasing similarity, and other characteristics. To strengthen the stability of the social network, "trust" has been considered in the community detection. Trust often refers to a subjective feeling among users that is particularly restrained by context. The trust relationships of users in social network will directly influence the intentions of target users. For example, a trust-based social network has been applied to E-business recommendation, financial fraud and other concerns [1].

Related researches on trust originate in the field of cyber security, mainly tackling the question of identification and recognition among nodes within the distributed network [2]. With the fast development of social network, scholars now begin to focus on some key questions of trust measurement and transmission in social network, and have made great progresses. Skopik et al. [3] proposed a principle-based method to construct an interaction and experience-oriented trust measurement within the online virtual community; Wang et al. [4] built a net-based trust model on the basis of Bayes Inference; Victor et al. [5] thought that in the recommended system of trust-based network, trust is a progres-

[*] Corresponding author.
E-mail address: xialooking@163.com (C. Xia).

https://doi.org/10.1016/j.chaos.2018.01.025
0960-0779/© 2018 Elsevier Ltd. All rights reserved.

sive phenomenon and put forward a dynamic modeling of trust-ambiguous relation to resolve the questions of trust degree and its progressiveness. Golbeck [6] concluded on the relations between trust and user similarity which paved a theoretical way for further service recommendation; Qiao [7] proposed a user-context-based trust algorithm, applying relevant theories in sociology and calculating the user trust via familiarity and relevance dimensions. Meanwhile, we also studied the modeling of trust by users towards software (service) in the selection process of trustworthy software (trustworthy cloud services) and displayed sorts of comprehensive trust calculating models [8,9].

Social network, as a complicated structure, is composed of different granularity recursions, in which users will spontaneously develop into abundant virtual communities in accordance with set rules or features. Virtual communities essentially reflect local features and correlations of individual behaviors in the social network, and contribute to understanding the network structures and exploring the inherent functions of complex systems. In practice, through the detection of virtual communities, we could more effectively monitor the public opinions, identify opinion leaders in groups, enhance the accuracy of advertisement launching and build a more reliable individualized recommendation system [10,11]. Meanwhile, discovering opinion leaders is beneficial to monitoring public opinions and lunch advertisements. In the individualized recommendation system, focal players are prone to adopt the suggestions and recommendations from trusted ones [1].

At present, the studies on virtual communities in social network have aroused much attention. In general, common community-detection methods could be summarized as two major types [12]: optimization-based and heuristic-principle-based. The former is to identify the potential community structure in the network by converting the virtual community detection into optimizing objective functions through designing a proper objective functions, represented by KL algorithm [13], spectral algorithm [14,15] and FN algorithm [16,17], while the latter is to be implementing by utilizing some proper heuristic principles, represented by MFC (maximum flow community) algorithm [18], HITS (hyperlink induced topic search) algorithm [19], Girvan and Newman (GN) algorithm [20] and its improved algorithm [21,22], Wu–Huberman(WH) algorithm [23] and CPM (clique percolation method) algorithm [24]. With a constant development of social networks, especially for the online social networks, the attributes and characteristics of nodes have played a more and more important role in community detection. Naresh et al. [25] raised a traditional clustering algorithm on the basis of users' multi-attributes; Dang et al. [26] designed the related attribute-similarity function for each node in consistence with users' attributes, then made the weight sum of attribute-similarity functions and modularity functions, and lastly realized the community detection by optimizing modules through Louvian algorithm; Sun [27] devised the similarity modularity functions based on the common attention and common fans, and then the maximization of modularity functions via greedy algorithm is used to discover virtual communities.

It can be drawn from the previous works that the trust measurement in the social network is mostly conducted based on interactive strength or information while an effective set completion upon the nodes similarity is often neglected. Moreover, as a significant characteristic among social network users, trust has not been widely applied in the detection of communities. Accordingly, in this paper, we suggest a more comprehensive trust calculating model, on the basis of a trust-based non-overlapping community detection method in which the trust among users is integrated into the way of finding communities so as to get a dual cohesion community with both structure and trust.

The rest of this paper is structured as follows. In Section 2, we describe and define the trust calculation in detail. We propose the new community detection algorithm: TLCDA, and carry out the related algorithm analysis and experimental results on two real-world datasets in Sections 3 and 4. At last, we end this paper with some concluding remarks in Section 4.

2. Trust calculation model

The trust relationship is a basic feature of social network users, normally in a comprehensive consideration of direct and indirect trust. In this section, we will introduce the calculation of trust degree among nodes in detail, and the direct trust degree is firstly calculated herein through the strength of ties and node similarity, then the calculation of indirect trust within distance restrictions is given.

2.1. Direct trust

In the scope of sociology and psychology, trust is generally believed to be relevant to features of oneself and able to be displayed by concrete subjective actions [28]. Therefore, it can be argued that users' actions like their active cooperation, communication and concern within social network, to some degree, reflect their trust extent. The relationship among users, as a result of subjective actions, is the most obvious and apparent embodiment of their trust degree.

Definition 1 (trust of direct relation based on tie strength). To a pair of adjacent nodes, trust obtained through their strength of ties is the direct-relation trust. The calculation of this direct relation trust degree is:

$$d\_trust\_r(u,v) = \frac{w(u,v)}{w(u)} \tag{1}$$

where $d\_trust\_r(u,v)$ is the direct relation trust degree between u and v and $d\_trust\_r(u,v) \in (0,1]$. $w(u, v)$ is the strength between u and v. $w(u)$ is the sum of strength of ties between u and its neighboring node. $w(u, v)$ has different interpretations in various applications, such as referred to as the collaborative number among users in science collaboration network, or to the interactive number among users in social network.

Recent researches find that there exist the apparent homogeneity between network nodes, i.e., similar nodes are prone to be correlated [28,29] one another. There has been a large quantity of studies on the measurement of node similarity, a simple method is to measure the similarity by calculating the number of shared neighbors between two nodes. Upon two adjacent nodes, the more their neighbors are overlapped, the higher the nodes similarity is [30]. In real social network, however, when the neighboring nodes of a certain node tend to be more, then the similarity the present node contributes to its neighboring node tends to be less. This point is clearly demonstrated in the unidirectional attention mechanisms such as micro-blog where some big stars have a great number of fans while contributing little in making a similar attention to their fans.

Definition 2 (trust of direct relation based on node similarity). For two adjacent users, the trust obtained through their similarity is the trust of direct similarity. The calculation method of this direct relation trust is:

$$d\_trust\_s(u,v) = \sum_{t \in N(u) \cap N(v)} \frac{1}{I(t)} \tag{2}$$

where $d\_trust\_s(u,v)$ is the direct similarity trust degree, $N(u)$ and $N(v)$ are the neighboring sets of u and v, respectively, and $I(t)$ is the penetration degree of t. It hereby presents the calculation of direct trust between two adjacent nodes as

$$d\_trust(u,v) = d\_trust\_r(u,v) + d\_trust\_s(u,v) \tag{3}$$