PHYSA: 16808 Model 3Gsc pp. 1–12 (col. fig: NIL)
ARTICLE IN PRESS
Physica A xx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Physica A

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

Q1 # The adaptive dynamic community detection algorithm based on the non-homogeneous random walking

Q2 Yu Xin [a,*], Zhi-Qiang Xie [a], Jing Yang [b]

[a] College of Computer Science and Technology, Harbin University of Science and Technology, Heilongjiang, 150001, China
[b] College of Computer Science and Technology, Harbin Engineering University, Heilongjiang, 150001, China

HIGHLIGHTS

- The proposed method resolves the issue of instability caused by fixing the random walking step.
- The proposed method considers the impact of topology environment on the random walking.
- The ANRW gives an adaptive way to adjust the detected community, to reduce the time complex.
- The ANRW is adequate to the parallel computing and large data analysis.

ARTICLE INFO

Article history:
Received 15 November 2015
Received in revised form 7 January 2016
Available online xxxx

Keywords:
Random walk
Dynamic community
Community detection

ABSTRACT

With the changing of the habit and custom, people's social activity tends to be changeable. It is required to have a community evolution analyzing method to mine the dynamic information in social network. For that, we design the random walking possibility function and the topology gain function to calculate the global influence matrix of the nodes. By the analysis of the global influence matrix, the clustering directions of the nodes can be obtained, thus the NRW (Non-Homogeneous Random Walk) method for detecting the static overlapping communities can be established. We design the ANRW (Adaptive Non-Homogeneous Random Walk) method via adapting the nodes impacted by the dynamic events based on the NRW. The ANRW combines the local community detection with dynamic adaptive adjustment to decrease the computational cost for ANRW. Furthermore, the ANRW treats the node as the calculating unity, thus the running manner of the ANRW is suitable to the parallel computing, which could meet the requirement of large dataset mining. Finally, by the experiment analysis, the efficiency of ANRW on dynamic community detection is verified.

1. Introduction

The people's social activities are variable, and the habits and hobbies are slowly changing, giving rise to the evolution through social network. Taking the case of social network sites, there exist many new registered users and disappeared users, and the relationships of users appear to be dynamic, such as the establishment of new friendships and the blanking of old friendships. Kumar [1] designed 3 policies to simulate the evolution of social network. In the generative process, the users in social network are classified into 3 groups: passive, linkers and inviters. The passive users join the network out of curiosity or at the insistence of a friend, but never engage in any significant activity. Inviters are interested in migrating an offline Q3 Q4

* Corresponding author.
E-mail addresses: xinyu@hrbeu.edu.cn (Y. Xin), xiezhiqiang@hrbust.edu.cn (Z.-Q. Xie), yangjing@hrbeu.edu.cn (J. Yang).

http://dx.doi.org/10.1016/j.physa.2016.01.025
0378-4371/

Please cite this article in press as: Y. Xin, et al., The adaptive dynamic community detection algorithm based on the non-homogeneous random walking, Physica A (2016), http://dx.doi.org/10.1016/j.physa.2016.01.025

community into an online social network, and actively recruit their friends to participate. Linkers are full participants in the growth of the online social network, and actively connect themselves to other members. Various dynamic social network structures, such as Flickr and Yahoo!360, can be simulated by adjusting the mixture coefficient. Therefore, the research on the evolution of social networks is the practical needs on social computing, much research on community evolving [2,3], community tracing [4,5] and dynamic simulation [6,7] have been proposed.

The community evolution scenarios raise new challenges to traditional clustering algorithms. On one hand, need to guarantee the detected communities has a higher rationality (a higher modularity). On the other hand, need to guarantee the communities has a dynamic adaptability to deal with the dynamic events. For that, we proposed the ANRW to deal with the two problems above. Our main contributions in this paper can be summarized as follows:

(1) A community detection method NRW considering the topology environment is proposed, which utilizes the logistic function as the possibility function for keeping on random walking at current step, realizing the non-homogeneous random walking for the *object*, resolving the issue of instability caused by fixing the random walking step.
(2) During the random walking, taking the degree of the node of the object arriving as the input, establishes the topology gain function for keeping on walking, to give the topology environment consideration into the random walking.
(3) Based on the NRW the dynamic community detection method ANRW is proposed, which adjusts the detected community aiming to the 4 dynamic events (node adding, node deleting, link adding, link deleting). As the dynamic events occurrence, the ANRW only adjusts the impacted nodes, giving a low computational cost.
(4) As the ANRW is specific to the nodes, the computational manner of ANRW is adequate to the parallel computing and large data analysis.

2. Related work

Currently, dynamic community detection methods can be classified into 4 categories: dynamic clustering, objective function optimization, representative node (community) detection, dynamic probability modeling.

(1) Dynamic clustering methods. These methods are mainly based on the EC (Evolutionary Clustering) proposed by Chakrabarti [8], seeing the time snapshot as clustering sample unit, calculating the node distribution in time snapshot t and $t-1$. Chi [9] proposed the PCQ (Preserving Cluster Quality) and (Preserving Cluster Membership) methods based on EC, where the PCQ emphasis on the cluster density and PCM on the node similarity of the detected communities. Lin proposed the FacetNet [10] method based on EC, which established the snapshot cost function according to the community distribution in time snapshot t. Kim [11] proposed the PDEM(Particle and Density based Evolutionary Clustering) improving the FacetNet, which took into account of the drawback of FacetNet on needing to preset the number of communities and not being allowed to change it.

(2) Objective function optimization methods. These methods estimated the community evolution by the structure changes, the process of which is guided by the optimization of community density function or modularity. Blondel [12] applied the FN [13] into the dynamic community detection, which is the combination of regional optimizing and hierarchical clustering. Dinh [14] suggested the MIEN (Modules Identification in Evolving Networks), which employed the incremental modularity optimizing method, compressing the network into several representative communities. Nguyen proposed the AFOCS (Adaptive Finding Overlapping Community Structure) [15] and QCA (Quick Community Adaption) [16] methods. These two methods designed 4 adjusting policies specific to the 4 dynamic events (node adding, node deleting, link adding, link deleting), and improved the community density function and modularity respectively, allowing for the dynamic community detection. Guo [17] proposed the ECSD (Evolutionary Community Structure Discovery) aiming at weighted dynamic community detection. In the process of ECSD, as the dynamic events occurrence, the increment of merged communities is treated as the criterion of community merging. Zhou [18] proposed the multiobjective optimization method for the dynamic community detection, which revises the community assignments of new or changed vertexes during network updates.

(3) Representative node (community) detection methods. These methods facilitated the adjusting process via tracing the representative nodes or communities. Zhang [19] proposed the BSP (Bulk Synchronous Parallel), which established the similarity measurement using the number of common neighbors and the connections. Duan [20] suggested the Stream-Group, which used the transition probability matrix to establish the relevance measurement between nodes, and proposed the compactness model to evaluate the regional community tightness. Bourqui [21] measured the similarity of communities between the adjacent time snapshots, and merged the communities with a high similarity. Chen [22] proposed the Representative-based community evolution tracing method, which made an assessment on the detected communities to find out the stable communities. When the network structure varies, it only traces the stable communities, by which the comparing times can be reduced. Takaffoli [23] gave an evolution analyzing method specific to the nodes, which adopts the steady-state model and impact model to recognize the stable nodes and representative nodes. Duan [24] gave an incremental K-clique community adjusting method utilizing the dynamic updating policy of DFS tree. Ma [25] proposed the CUT (Community Update and Tracking) model, which detected the seed community (clique community). As the dynamic events occurrence, the CUT only updated the affected seed communities to reduce the community detection cost.

(4) Dynamic probability modeling methods. These methods assume the community distributions in each time snapshot as the samples of latent community distribution, and establish the probability model to associate the samples with latent