Physica A 460 (2016) 98–104

Contents lists available at ScienceDirect

Physica A

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

An improved label propagation algorithm using average node energy in complex networks

Hao Peng [a,1], Dandan Zhao [a,1], Lin Li [b,1], Jianfeng Lu [a], Jianmin Han [a], Songyang Wu [c,*]

[a] *College of Mathematics, Physics and Information Engineering, Zhejiang Normal University, Jinhua 321004, China*
[b] *China Electronics Standardization Institute, Beijing 100007, China*
[c] *The Third Research Institute of Ministry of Public Security, Shanghai 201204, China*

HIGHLIGHTS

- We propose an improved label propagation algorithm (LPA) to uncover overlapping community structure.
- The improved LPA can identify the bridge nodes in each iteration and then we can uncover overlapping communities when the iteration terminates.
- The introduced algorithm can effectively uncover reasonable overlapping community structures in the real-world and social networks.

ARTICLE INFO

Article history:
Received 24 October 2015
Received in revised form 11 April 2016
Available online 10 May 2016

Keywords:
Community structure
Label propagation
Average node energy

ABSTRACT

Detecting overlapping community structure can give a significant insight into structural and functional properties in complex networks. In this Letter, we propose an improved label propagation algorithm (LPA) to uncover overlapping community structure. After mapping nodes into random variables, the algorithm calculates variance of each node and the proposed average node energy. The nodes whose variances are less than a tunable threshold are regarded as bridge nodes and meanwhile changing the given threshold can uncover some latent bridge node. Simulation results in real-world and artificial networks show that the improved algorithm is efficient in revealing overlapping community structures.

1. Introduction

A wide variety of complex systems can be regarded as complex networks which are composed of vertices connected by edges [1–3], and understanding the properties of these networks is a powerful tool for analyzing complex systems. Community structure is one of the important topological property in complex networks. The communities are groups of nodes within which nodes are densely connected. Since this property gives a valuable insight into the functional behavior and topological feature of various complex systems [4], detecting and analyzing community structure has attracted great attentions in recent years.

Some existing works have been developed to uncover reasonable community structure, especially in the last few years. The Kernighan–Lin algorithm [5] and the spectral analysis algorithm [6] are two typical methods to detect communities.

* Corresponding author.
E-mail address: wusongyang@stars.org.cn (S. Wu).
1 These authors contributed equally to this work.

http://dx.doi.org/10.1016/j.physa.2016.04.042
0378-4371/

Recently, Newman and Girvan proposed a seminal algorithm to split the network into hierarchical communities using the definition of betweenness which is the fraction of the shortest paths passing on each edge [7]. Then Clauset and Newman proposed the definition of Modularity which can evaluate the partition of communities in the complex network [8]. Based on the definition of Modularity, many kinds of algorithms are proposed to uncover community structure in complex networks. For example, Fang Wei et al. improved Modularity to uncover local community structure which lead to a local expansion algorithm [9]. Liu Xu et al. proposed an agglomerative clustering algorithm which based on neighborhood similarity to reveal community structure in complex networks [10]. Gong Maoguo et al. solved the community detection as a multiobjective optimization problem using multiobjective evolutionary algorithm [11]. To uncover community structure in a near linear time, Raghavan et al. proposed a notable algorithm named label propagation algorithm (LPA) by employing a simple label propagation in each iteration [12]. The LPA identified communities after propagating label among nodes until each node owns the label shared by most of its neighbors. All the algorithms above can only generate the hard-partition of a complex network. Recently, however, many modules are proofed to be overlapping communities. Thus many of the fore-mentioned algorithms are expanded to reveal overlapping communities in complex networks. Palla et al. proposed a clique percolation method (CFinder) to uncover overlapping community structure using the detected cliques [13]. Lancichinetti et al. developed an algorithm named OSLOM to uncover overlapping communities [14]. OSLOM is a local optimization of a fitness function which expresses the statistical feature of a community structure. Gregory et al. employed a tunable parameter υ to allow each node belong to υ communities at the most and improved classic LPA to uncover overlapping community structure (COPRA) [15]. All these valuable works can detect overlapping communities in the complex network. However, when detecting overlapping scopes, they also have some limitations in function or accuracy.

In this paper, we introduce the average node energy to express the bridge nodes and the improved LPA algorithm to detect overlapping communities. Under this way, we use a given parameter δ_{ave} to calculate a threshold which labels the latent bridge nodes and any nodes whose variance are less than the threshold will not update its label in the iteration. When all the labels stop updating, the overlapping community structure and bridge nodes are unambiguous.

2. The improved LPA using average node energy

In order to describe the algorithm proposed in this paper, we first define some notations as follows. Let $G = (V, E)$ be an undirected weight network, where V is node set and E is edge set. $L = \{L_1, L_2, \ldots, L_M\}$ is the label set which presents all active labels of nodes in network. $|L_i|$ $(i = 1, 2, 3, \ldots, M)$ is the number of nodes whose label is L_i. l_i is the label set and all elements in l_i present the label obtained by node i's nearest neighbors. Let $N_i(k)$ be the node set. All nodes in $N_i(k)$ are neighbors of i and labeled by k. $|N_i|$ is the number of nodes connecting with node i. $W_i(k) = \sum_{j \in N_i(k)} w_{ij}$ where w_{ij} is weight of edge between node i and j.

2.1. Label propagation

To detect community structure in a near linear time, Raghavan et al. proposed the label propagation algorithm which can reveal community structure following a simple workflow. Each node is identified by a unique label which implies different community identifier. Then each node asynchronous or synchronous updates its label with the label shared by most of its neighbors iteratively. If more than one label to which the maximum number of its neighbors belongs, one of them is chosen randomly. When all nodes own the label shared by most of its neighbors, the iteration terminates and meanwhile nodes with the same label belong to one community. The LPA workflow is simple and can detect some reasonable communities in the near linear time. However, LPA still owns some defects which limit its performance and application. For example, the propagation nature of LPA can cause the detected communities fall into local optimum. The label feature that limits the LPA cannot uncover overlapping community structure in the complex network. Some valuable works are proposed to solve the existing problems. Leung et al. used geodesic distance to uncover the overlapping communities. However it is difficult to know the diameter of all communities in advance. COPRA allows each node associate with more than one label which is more reasonable, but the parameter υ which means each node belong to υ communities at the most is also unpredicted for many complex networks. For this reason, in this paper, we introduce an improved LPA to detect overlapping community in the complex networks.

2.2. The improvements of LPA

The classic LPA updates the label of each node with the label used by the greatest number of its neighbors. Thus all nodes choose only one label to update in each iteration, and this rule results in the hard partitions for LPA. If some bridge nodes which belong to more than one community exist in the complex network, they will oscillate among different communities in iterations. Although there is no quantitative definition for a bridge node, most researchers identify a node almost equally connecting with more than one community as a bridge node. If these nodes can be identified and do not update their labels, the overlapping communities will be uncover when the algorithm terminates.

As is shown in Fig. 1, let $W_R = \sum_{t=1}^{4} R_t$, $W_Y = \sum_{t=1}^{3} Y_t$ and $W_B = \sum_{t=1}^{3} B_t$. Then if $W_R \approx W_Y \approx W_B$, node A is a bridge node. If we associate each node with a random variable which means the sum of weight between this node and the nodes