Physica A 483 (2017) 156–170

Contents lists available at ScienceDirect

Physica A

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

Detecting community structure in networks via consensus dynamics and spatial transformation

Bo Yang [a,*], He He [a], Xiaoming Hu [b]

[a] School of Automation, Wuhan University of Technology, Wuhan 430070, China

[b] Optimization and Systems Theory, KTH Royal Institute of Technology, Stockholm 10044, Sweden

HIGHLIGHTS

- We detect the community structure via the dynamical evolution of nodes.
- We measure the similarity by consensus dynamics and spatial transformation.
- The topological relations are translated into the distances in Euclidean space.
- The proposed algorithm operates on computer-generated and real-world networks.

ARTICLE INFO

Article history:
Received 6 September 2016
Received in revised form 28 February 2017
Available online 2 May 2017

Keywords:
Networks
Community detection
Consensus
Clustering
Spatial transformation

ABSTRACT

We present a novel clustering algorithm for community detection, based on the dynamics towards consensus and spatial transformation. The community detection problem is translated to a clustering problem in the N-dimensional Euclidean space by three stages: (1) the dynamics running on a network is emulated to a procedure of gas diffusion in a finite space; (2) the pressure distribution vectors are used to describe the influence that each node exerts on the whole network; (3) the similarity measures between two nodes are quantified in the N-dimensional Euclidean space by k-Nearest Neighbors method. After such steps, we could merge clusters according to their similarity distances and show the community structure of a network by a hierarchical clustering tree. Tests on several benchmark networks are presented and the results show the effectiveness and reliability of our algorithm.

1. Introduction

The science of complex networks has been a hot area of research of the scientists' community in the past decades [1–4]. Researchers have applied complex networks successfully on many fields as diverse as the internet [5], neural networks [6], social networks [7], electricity grids [8] and biological organizations [9]. With the deepening of the study on complex networks, it was found that the distribution of edges is not characterized by random, but organized by some certain rules. One of these distinctive rules existed in many complex networks is "community structure", the division of nodes into groups within which the network connections are dense, while between which there are only very sparse connections. In general, communities are groups of nodes which probably share common properties or play similar roles. So the community structure is closely related to the organization and function of a network.

* Corresponding author.
E-mail address: boboboy@gmail.com (B. Yang).

http://dx.doi.org/10.1016/j.physa.2017.04.098

The community detection is to find an appropriate partition of identifying communities in a network. It provides help to reveal the deeper structure of the entire network and the functional modules. When communities are detected, we can simplify the network by shrinking all nodes of a community to one new node, and recognizes the whole outline of the network. As detecting communities exerts so much positive influence on the understanding of complex systems, scientists have spent a large effort on it and proposed a number of methods already. Santo Fortunato [10] has made an exposition of most major methods, including graph partitioning [11], spectral algorithm [12], modularity-based method [13,14], and the hierarchical clustering algorithms [15].

The hierarchical clustering in sociology has a long history and becomes a common use in detecting communities, in which clusters are iteratively merged or split according to their similarity. So the key point of any hierarchical clustering method is the definition of a similarity measure between nodes. The similarity measure is usually limited to the local information of each node, such as the number of common neighbors, the geographic distance between nodes, and the Pearson coefficients. Hence it is of significant importance to develop a proper similarity measure that utilize the global information in the network of interest.

On the other hand, the networks' dynamics is an important object as the agents of most complex systems in real world are dynamical. Hence many physicists study the relationship between the structure and the dynamics on complex networks, and solve the community detection problem by dynamic methods such as spin models [16], random walks [17], and synchronization [18]. Additionally, He et al. [19] established a connection between consensus dynamics and the community structure in networks and propose two methods for community detection by observing dynamical matrices. Compared with traditional methods, methods employing process running on a graph detect communities by the underlying structure of data evolution. It is helpful for researchers to understand how the topological structure and the dynamical process interact and influence each other.

Embarking from the above thoughtfulness, we seek the organization of communities by a novel clustering method, which characterizes the similarity between nodes via consensus dynamics and spatial transformation. To address this aim, this paper firstly proposes a physical model that the consensus process running on a network is emulated to a procedure of gas diffusion in a finite space. Then the definition of a similarity measure between nodes is presented by the k-Nearest Neighbors method [20]. Then we could merge groups of nodes according to their similarity distances and obtain a hierarchical clustering tree which represents the community organization of the network. The major contributions of the algorithm are: (1) the community detection problem is translated to a clustering analysis problem in Euclidean space; (2) it measures the similarity between nodes by consensus dynamics and k-Nearest Neighbors method; (3) it correctly detects communities by the combination of consensus dynamics and spatial transformation.

The rest is organized as follows: A brief summary of some preliminaries is provided in Section 2. Section 3 presents our algorithm for community detection. In Section 4, we test our algorithm on computer-generated and real-world networks. Finally, some conclusions and future work are obtained in Section 5.

2. Preliminaries

The theoretical framework presented in this work relies on some basic concepts in graph theory, clustering method and modularity. Now we review some essential background in this section.

Let $G = (V, E)$ be an undirected graph of order N with the set of nodes $V = \{v_1, v_2, \ldots, v_N\}$, set of edges $E \subseteq V \times V$. The node indexes belong to a limited index set $\ell = \{1, 2, \ldots, N\}$. An edge between node v_i and node v_j is denoted by $e_{ij} = (v_i, v_j)$. We say that node v_i is a neighbor of node v_j if $e_{ij} \in E$, and the set of neighbors of node v_i is denoted by $N_i = \{j \in \ell, (v_i, v_j) \in E\}$. The adjacency matrix $\boldsymbol{A}$ is a basic expression of graph G in graph theory and computer science, which is defined by

$$A_{ij} = \begin{cases} 1 & e_{ij} \in E \\ 0 & \text{otherwise}. \end{cases} \tag{1}$$

Additionally, we assume $A_{ii} = 0$ for all $i \in \ell$.

As communities are usually groups of nodes which share common properties or play similar roles, it is feasible to classify similar nodes into a community. In our algorithm we measure the similarity by combining the consensus dynamics and the k-Nearest Neighbors procedure. Then we iteratively merge clusters of nodes according to such similarity: (1) compute the similarity measure for all node pairs, and assign each node to a cluster of its own; (2) find the pair of clusters with the highest similarity and join them together into a single clusters; (3) compute the similarity between the new composite cluster and all others using single linkage clustering; (4) repeat steps (2) and (3) until all nodes have formed a single cluster. The above procedure produces a hierarchical decomposition of a network into a set of nested clusters, which could be visualized in the form of a hierarchical clustering tree.

Since a larger modularity indicates a stronger community structure [13], the community structure of a network is identified by selecting the optimal cut on the hierarchical clustering tree according to the value of modularity. For a particular division of a network into C clusters, the modularity can be defined by

$$Q = \sum_{s=1}^{C} (q_s - a_s^2) \tag{2}$$