Neurocomputing 170 (2015) 384–392

Contents lists available at ScienceDirect

Neurocomputing

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

Community detection based on local topological information and its application in power grid

Zengqiang Chen [a,b,*], Zheng Xie [a], Qing Zhang [b]

[a] Tianjin Key Laboratory of Intelligent Robotics, College of Computer & Control Engineering, Nankai University, Tianjin 300071, China
[b] College of Science, Civil Aviation University of China, Tianjin 300300, China

ARTICLE INFO

Article history:
Received 5 July 2014
Received in revised form
27 April 2015
Accepted 27 April 2015
Available online 30 June 2015

Keywords:
Power grid
Complex network
Community detection
Local similarity
Local topological information

ABSTRACT

Community structure detection algorithm is employed in a vast amount of study to partition a network into some loosely coupled sub-networks of smaller scale. It is an effective tool to analyze and control some large-scale networks such as power grids. This paper proposes a novel algorithm based on local similarity to detect the community structure in complex network. Firstly, a new similarity index between nodes is defined to model the topological closeness of local connections in networks. Then nodes sharing high similarity are gathered to form the community structure. The results suggest the emergence of the bridging nodes and kernels within the community detection process. This proposed method performs well when it is introduced to seek out the actual community structure, kernels and bridging nodes in some benchmark networks. Thirdly, the proposed algorithm lends itself to many applications, such as detecting communities in several IEEE standard power grids and investigating the roles of bridging nodes in cascading failures. Experimentally, the proposed algorithm outperforms some other similarity indices and clustering methods. Finally, a detailed comparison helps us get the conclusion that the traditional label propagation algorithm is a special case of the proposed algorithm.

1. Introduction

With the rapid development of power industry, the scale of power grids is increasing constantly, and the architecture becomes more and more complex. Meanwhile, the analysis and control in regard to the power systems pose challenges to researchers. Now, this subject gets much hints from the development of complex network theory, such as the network control and synchronization, small-world and scale-free characteristics, community structure, cascading failures and networks robustness [1].

Communities in networks, which are local structures of dense inner connections and sparse links between them, help to solve the complex problem. The large scale networks are divided into some loosely coupled sub-networks of much smaller scale, which is easier to control. This divide-and-conquer strategy improves the robustness of the system and helps to analyze and control large-scale systems. In power grids, community structure is frequently used in reactive power network partition [2], coherency-based dynamic equivalence [3], and power system restoration [4], and so on.

Previous studies have detected several approaches to detect communities in networks, such as hierarchical clustering methods [5,6], modularity optimization methods [7,8], and some other algorithms [9–11]. Besides, the machine learning method is also deemed as an effective tool in network partition. Wang et al. introduce several similarity indices as feature matrix and use nonnegative matrix factorization to determine the final clustering [12]. Leng et al. propose an active semi-supervised community detection algorithm with label propagation by selecting some informative nodes as labeled ones [13]. These network partition methods can be applied into power grids as a complement to traditional methods which are only based on geographical and administrative partition. Ni et al. adopt division and agglomeration strategy to partition the power grids [2]. The division step using voltage sensitivity matrix confirms the primary structure of the areas, and the agglomeration step based on modularity optimization balances reactive power in place within the community. Lin et al. detect the community structure by removing the edges which have biggest betweennesses in sequence in a power grid [4]. The detected hierarchical structure helps to restore the power subsystems quickly and process them in a right parallel manner. By introducing the reactive power balance of the network partitions, Pan et al. redefine the modular index and adopt the improved Louvain hierarchical algorithm in power grids [14]. Also they discover the pilot nodes which can well reflect and significantly affect the voltage level within the partition. Chen et al. propose an efficient similarity-based clustering method which performs well in power grids [15]. Guo et al. view a node's controllability sensitivity as its coordinate in space and map the

* Corresponding author at: College of Computer and Control Engineering, Nankai University, Tianjin 300071, China.

http://dx.doi.org/10.1016/j.neucom.2015.04.093
0925-2312/

power grid into the coordinate space. Then they apply traditional hierarchical agglomerative clustering to search the grid's community structure [16]. Recently, more attention is attracted to the problem of grid partition, to enhance the robustness of the power system. Pahwa et al. propose a constraint programming formulation for optimal islanding in power grids and give two grid partition methods based on modularity for islanding large-scale systems [17]. Mehrjerdi et al. apply spectral k-way partitioning formulation in grids and propose fuzzy secondary voltage method to avoid propagation of disturbances between regions [18]. Cotilla et al. view the power grids partition as a multi-attribute objective optimization problem which consider the factor of electrical distances, cluster connectedness and so on. Then they solve it with a hybrid algorithm of K-means and evolutionary computational [19].

In this paper we propose an alternative method to detect community structure. Inspired by the consideration that similar nodes tend to gather together to form a community, we design a novel node similarity index and assign nodes sharing high similarities to the same communities. In the partition process, the bridging nodes and kernels can be detected in a heuristic way. Experiments on some benchmark networks verify the effectiveness of the algorithm. Then we apply the algorithm to some power grids, and the comparison with other employed methods indicates that our algorithm can help to get community structure in a more effective way than others. Exploration of nodes' influence on cascading failure gives us much guidance to supervise the power grids.

The rest of the paper is organized as follows. In Section 2, the community detection process is introduced, which is composed of similarity index definition, network partition and some other supplements. In Section 3, the computation cost is analyzed. In Section 4, the algorithm is verified by experiments on benchmark networks. In Section 5, the algorithm is applied in several standard power grids and compared with different indices and clustering methods. Finally, in Section 6 we summarize the work and provide some inspirations for future study.

2. A novel community detection

The basic idea of the paper is that nodes are more likely to join the community with which they share the highest similarity. So similarity index after normalization can be viewed as the probability of nodes belonging to communities. Then, how to design a similarity index to measure the closeness between nodes? How to group nodes sharing high similarities into communities? Following parts will provide corresponding solutions.

2.1. Similarity index

When considering network topology information, the so-called similar nodes mean that there are many paths between them. By employing an adjacent matrix A to represent the network topological information, many indices have been developed to measure the node similarity [20,21].

Common neighbors index (CN) is defined to describe the similarity by the number of two nodes' common neighbors, namely

$$similarity(x,y) = |\Gamma(x) \cap \Gamma(y)|, \tag{1}$$

where $\Gamma(x)$ denotes the neighbor set of node X and $|X|$ denotes the number of elements in set X. CN can be easily calculated, however, the direct links and high order paths information about the network are ignored here.

Katz index is developed as the ensemble of all paths between nodes, and the definition is given by

$$similarity = \beta A + \beta A^2 + \beta A^3 + \cdots = (I - \beta A)^{-1} - I, \tag{2}$$

where parameter $\beta < 1$ controls different influence of paths. Note that β should be smaller than the reciprocal of the maximum of the eigenvalues of matrix A to ensure the convergence of (2), and we often set $\beta = 0.01$. This index considers the global information of the network. However, real networks are often of huge size and very sparse. It is time consuming to calculate Katz index.

Then Local Paths index (LP) is proposed as a trade-off, namely

$$similarity = A^2 + \beta A^3. \tag{3}$$

It contains more topology information than CN, and incurs less computation cost than Katz. However, LP is proposed specially to predict links between disconnected nodes, which leave out the consideration of deeming direct connection as similarity between the end points. Taking the example in Fig. 1. With LP index, the similarity of indirectly connected nodes D and B is 1.01, greater than that 0.03 of directly connected nodes D and E. When applying LP in community detection, node D is more likely to join the community where node B rather than E is in. This goes against intuition. So LP index is not suitable for the community detection.

To solve these problems mentioned above, this paper combines advantages of LP and Katz index and proposes a new local similarity index called LS for short, namely

$$similarity = A + \beta A^2 + \beta A^3. \tag{4}$$

Parameter β can be adjusted appropriately to fit the community detection well. Its influence on community detection will be shown in details in Section 4. This index calculates the number of paths through which node X can reach node Y within three steps, which indirectly reflects connections between a node's neighbors. Like the community structure, this similarity metric can also reflect local close connections well.

2.2. Network partition

Among clustering methods based on similarity, K-means and K-medoids are widely used ones [22]. Both the two algorithms consider the similarity between a node and a community as the similarity between the node and the center of the community. Then they assign each node to the nearest center to form the community structure. The difference is that K-means clustering calculates the means of nodes coordinates, while K-medoids selects medoids, the nodes which have the biggest similarities with other nodes in the clusters, as community centers.

However, the above methods fail to detect community structure in networks. K-means is applicable to attribute data described by high dimension vectors, but invalid for relational data in networks. K-medoids clustering only focuses attention on the centers of communities, neglecting the connections between non-centers. For example, a node has only one edge with a

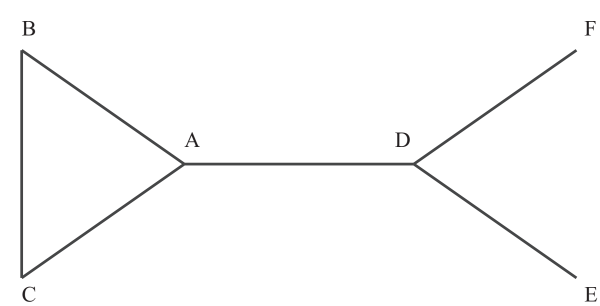

Fig. 1. An example network to illustrate the drawback of LP index.