Physica A 469 (2017) 471–485

Contents lists available at ScienceDirect

Physica A

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

A community integration strategy based on an improved modularity density increment for large-scale networks

Ronghua Shang [a,*], Weitong Zhang [a], Licheng Jiao [a], Rustam Stolkin [b], Yu Xue [c]

[a] *Key Laboratory of Intelligent Perception and Image Understanding of Ministry of Education of China, Xidian University, Xi'an, 710071, China*

[b] *Extreme Robotics Lab, University of Birmingham, UK*

[c] *School of Computer and Software, Nanjing University of Information Science & Technology, Nanjing, China*

HIGHLIGHTS

- This paper presents a community integration strategy for large-scale networks.
- We search the core node in the network namely the potential community center.
- We arrange the communities according to the external connections in descending order.
- We propose an improved modularity density increment as the objective function.
- We add global judgment in the process of the local integration.

ARTICLE INFO

Article history:
Received 3 July 2016
Received in revised form 20 October 2016
Available online 22 November 2016

Keywords:
Large-scale network
Local core node
Improved modularity density increment
Community integration

ABSTRACT

This paper presents a community integration strategy for large-scale networks, based on pre-partitioning, followed by optimization of an improved modularity density increment ΔD. Our proposed method initially searches for local core nodes in the network, i.e. potential community centers, and expands these communities to include neighbor nodes which have sufficiently high similarity with the core node. In this way, we can effectively exploit the information of the node and structure of the network, to accurately pre-partition the network into communities. Next, we arrange these pre-partitioned communities according to their external connections in descending order. In this way, we can ensure that communities with greater influence are prioritized during the process of community integration. At the same time, this paper proposes an improved modularity density increment ΔD, and shows how to use this as an objective function during the community integration optimization process. During the process of community consolidation, those neighbor communities with few external connections are prioritized for merging, thereby avoiding the fusion errors. Finally, we incorporate global reasoning into the process of local integration. We calculate and compare the improved modularity density increment of each pair of communities, to determine whether or not they should be integrated, effectively improve the accuracy of community integration. Experimental results show that our proposed algorithm can obtain superior community classification results on 5 large-scale networks, as compared with 8 other well known algorithms from the literature.

* Corresponding author.
E-mail address: rhshang@mail.xidian.edu.cn (R. Shang).

http://dx.doi.org/10.1016/j.physa.2016.11.066

1. Introduction

In order to solve problems of "big data" in real world applications, complex network models, composed of nodes and edges, have been widely used and studied [1–3]. A variety of methods have been used to help reveal topological properties of complex networks, including social networks, information networks and biological networks, and have helped to advance the empirical analysis and modeling of complex networks [4–6]. Developing methods for community detection is critically important in order to understand the network topology and the dynamic behavior of the network [7].

In recent years, a considerable number of algorithms for community detection have been proposed and improved. In 2002, Newman and Girvan proposed a method (GN algorithm) [8] to decompose the network by removing edges. The GN algorithm decomposes the network hierarchically, and has been shown to accurately divide community structures in many cases. However, this method lacks an effective measure for determining whether or not the result is the optimal partition. To overcome this problem, Newman and Girvan later proposed a modularity function Q [9] as a standard for evaluating community detection results. In 2004, Newman proposed a fast algorithm (Fast Newman algorithm) [10], which optimized the Q value directly through merging nodes. However, the Newman Fast algorithm is an example of a greedy algorithm, and greedy algorithms, in general, tend not to converge on globally optimal solutions. Raghavan et al. [11] proposed a community detection method using label propagation technology in 2007. This algorithm is simple, easy to perform and yields good classification results. It also has the advantage of linear time complexity, which makes it especially suitable for applications in large-scale network community detection. However, the conventional label propagation algorithm has strong randomness and weak robustness, which can cause small communities to be influenced or erroneously annexed during the process of label propagation. Lin et al. proposed a new label propagation algorithm [12] (CKLPA) based on community core in 2014, which improved the randomness of the label propagation algorithm and the accuracy of community detection. The LPAm algorithm [13], proposed by Barber and Clark et al., is similar to the original LPA algorithm, however it incorporates and improved node label update strategy to optimize local modularity. This method ensures that each node label update always moves in the direction of increasing local modularity in each iteration. The final optimization results of the algorithm correspond to the optimization of the modularity. However, this method is vulnerable to convergence on local optima. To overcome the problem of local optima convergence, Liu et al. [14] proposed a multi-step greedy fusion strategy, similar to the MSG algorithm [15], which demonstrated improved results over previous methods. In 2008, Li et al. [16] proposed a new community detection evaluation index, the Modularity Density D function, for the resolution problem of modularity. Since then, a variety of community discovery algorithms have been proposed, based on using D as an objective function for optimization. In 2011, Gong et al. proposed the Meme-Net algorithm [17] using D as the optimization function, and adopted a local search strategy hill climbing method. They proposed a way of detecting different resolutions of networks by adjusting the λ parameter within their objective function. Unfortunately, this algorithm has high computational complexity, and it therefore does not scale well for detecting communities within very large-scale networks. Gong et al. proposed the MOEA/D-Net algorithm [18] in 2012. They treated both parts of the modularity density function, ratio association and ratio cut, as joint objective functions, and decomposed the resulting multi-objective optimization problem into a series of scalar optimization sub problems. As with the previously discussed method, this algorithm is unsuited for community detection in large-scale networks because of its high computational complexity. 2016, Shang et al. proposed the APMOEA algorithm [19], which combined the AP and MOEA algorithms. Similar to the MOEA/D-Net algorithm, APMOEA also utilized both parts of the modularity density function, ratio association and ratio cut, as joint objective functions, and demonstrated good results of community division. However most of these algorithms are based on evolutionary algorithms for numerical optimization, with slow convergence, poor local search ability, making them less suitable for large-scale network community detection.

In contrast, this paper presents a community integration strategy specifically designed for large-scale networks. It is based on pre-partition and the improved modularity density increment ΔD, which incorporates global information into the local integration process. Initially, each node is regarded as an independent community. Next, we search for a local core node in the network, i.e. a potential community center, and then we grow this community by incorporating neighboring nodes which share a high similarity with the core node. In this way, we can effectively exploit information of both the node and the structure of the network, in order to accurately pre-partition the network. Next, we arrange the communities acquired from the pre-partition according to their external connections in descending order. In this way, we can ensure that communities with greater influence are prioritized during the process of community integration. Last, we propose an improved modularity density increment ΔD based on the modularity density as the objective function. In the process of community consolidation, the neighbor communities with least external connections are handled first, thereby avoiding fusion errors. Finally, we calculate the improved modularity density increment of each pair of communities to determine whether they should be integrated, so as to effectively improve the accuracy of community integration.

2. Related works

In this section, we introduce several relevant definitions and algorithms that will be used later in the paper, mainly including: local core node; node similarity function; community integration strategy; and modularity density function.