Physica A 436 (2015) 665–677

Contents lists available at ScienceDirect

Physica A

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

# Community detection using local neighborhood in complex networks

Justine Eustace, Xingyuan Wang *, Yaozu Cui

*Faculty of Electronic Information and Electrical Engineering, Dalian University of Technology, Dalian 116024, China*

## HIGHLIGHTS

- Local neighborhood ratio is proposed in defining community structure.
- Introduce the average neighborhood ratio as threshold in associating nodes to a particular community structure.
- Presents related theorem and its proofs about the neighborhood ratio in defining community structures in complex networks.
- A new method is presented in detecting community structures in complex networks.

## ARTICLE INFO

*Article history:*
Received 10 February 2015
Received in revised form 5 April 2015
Available online 12 May 2015

*Keywords:*
Complex networks
Data mining
Community detection algorithms
Neighborhood ratio
Behavior science

## ABSTRACT

It is common to characterize community structure in complex networks using local neighborhood. Existing related methods fail to estimate the accurate number of nodes present in each community in the network. In this paper a community detection algorithm using local community neighborhood ratio function is proposed. The proposed algorithm predicts vertex association to a specific community using visited node overlapped neighbors. In the beginning, the algorithm detects local communities; then through iterations and local neighborhood ratio function, final communities are detected by merging close related local communities. Analysis of simulation results on real and artificial networks shows the proposed algorithm detects well defined communities in both networks by wide margin.


## 1. Introduction

In a network, node connection is attributed to a number of significant relations such as web page relations, social relations and biological relations. Similarly node connection can be used to define features and operations of the modeled systems [1]. Furthermore node connection can be used to model feedback control [2]. One of the most prominent feature is the community [3], which is often considered as a sub-graph with more internal connections than external connections to the rest of the network. Complex network and community detection have a wide variety of practical applications in several disciplines [4–9]. For example, in computer science, community detection can be used to identify similar community features in parallel processing, network scheduling, neural networks [10–12] and World Wide Web (www). Community detection can also be a significant tool to model non-linear systems.

Significant numbers of community detection algorithms in complex networks have been devised. Some of them can detect overlapping communities, some can detect non-overlapping communities. Others can detect both overlapping and non-overlapping communities. Of particular interest to this work is the detection of non-overlapping communities using local community approaches where a number of challenges still persist. One of the challenges is performance variation

* Corresponding author.
*E-mail addresses:* justineustace@yahoo.com (J. Eustace), wangxy@dlut.edu.cn (X. Wang), cyz3471@sina.com (Y. Cui).

http://dx.doi.org/10.1016/j.physa.2015.05.044

across network types. For example, in real and artificial networks where community boundaries are hard to define, these algorithms fail [13–15]. Another challenge is: some algorithms have limited scope and their performance depends on the position of the seed vertex. For example, community detection using local communities by Bagrow and Bolt [13] performs well when the seed node is at the center of local community [15].

To overcome the above mentioned challenges, here a new method to detect non overlapping communities in a given network is proposed. The proposed approach uses structure similarity measure to determine node membership to a particular community. These community members can be nodes or group of similar nodes. First, average neighborhood ratio is computed, then similarity measure is computed between the pair of nodes suspected to be in the same community. The qualified node pairs are then merged into a local community. Through iterations, similar local community pairs are then merged into a global community. The same measure and density is extended to compute similarity between local communities. The algorithm terminates when no more local communities can be merged into the global communities. In addition, some basic properties related to the proposed algorithm are derived and proved.

The proposed algorithm is tested and its performance is compared with existing algorithms using real and computer generated networks. Experimental results show that with great accuracy, the proposed algorithm successfully detects community in real and artificial networks. Similarly, simulation results show that; the proposed algorithm can well defined community with Normalized Mutual Information score of 0.40 under maximum mixing parameter. We can summarize our contribution presented in this paper as:

- We proposes a new definition of a community structure in complex network. The proposed definition is based network natural property which are neighborhood ratio and its corresponding average neighborhood ratio.
- We deduce some basic properties of a community structure defined using neighborhood ratio.
- We propose a method to detect community structure in complex network which is defined by using neighborhood ratio and its corresponding average neighborhood ratio.
- We evaluate the performance of the proposed method in real and artificial networks in varying nodes and community sizes.

The rest of this paper is organized as follows; Section 2, briefly outlines a list of related methods. Section 3 give formal notations and definitions used in this paper. The algorithm and methodology used in this paper are presented in Section 4. Section 5, presents the experiment results and comparison with other related works. Finally, this paper is concluded in Section 6.

## 2. Related methods

Currently, there are some algorithms which address the above mentioned challenges. Unlike overlapping communities detection methods, which can network as fuzzy system [6,16]. This work proposes a non-overlapping community detection method. Related to our work is propinquity dynamic by Zhang et al. [17]. Their work proposed a method to compute probability of a neighboring vertex pair participating in a community. The propinquity values are then used to discover a community. Although this approach has a linear computation complexity for a sparse network, it fails to define a community when the network is too large.

Similar methods which uses local community approach to detect community structure are Local Modularity [14] which at first construct local community and then using modularity, closely related local community are combined into their corresponding community. Other methods which uses local community approach to detect community structure includes, Local Optimal Community [18], Local Optimization Fitness method [19] and Local Tightness Expansion [15]. Like many other local community approach their performance is affected by the network density. A similar approach of community detection using local community is community detection using hierarchical. Here several local community are defined at different levels and thereafter merged to their corresponding global community [20–23].

For example, agglomerativE hierarchicAl clusterinG based on maximaL cliquE (EAGLE) algorithm [22] proposes agglomerative approach whereas, local communities are first defined and then using modularity [24], maximum related pair are combined. Fast unfolding of communities using modularity optimization (FMODULARITY) by Blondel et al. [25] also uses hierarchical approach. In this algorithm similar local communities are merged by using modularity approach. Like any other modularity based algorithm, EAGLE and FMODULARITY algorithm is also affected by resolution limit. Specific drawback of EAGLE algorithm is its computation complexity, in which it may be considered exponential [26].

Label propagation algorithm by Raghavan et al. [27] and its variants [16,28] uses somewhat related method to our work. In label propagation algorithm, each node is labeled with an integer identifier. Then, through iterations each node updates its label by replacing it with the label used by the greatest number of its neighbors [16]. Finally, a community is defined by merging all the vertices with the same label. The pitfall of this approach is that, it defines different communities depending on the position of the seed node in the network structure.

Vertex similarity methods like *k-core* is another approach which can be used to detect a community. We use a similar approach to define local community and global community structure. For a given $k$, a *k-core* is a sub-graph which is adjacent to at least $k$ vertices. Clique percolation algorithm [29] is a good example of an algorithm which uses k-core. It consider a community as a union of adjacent k-clique. Although it is a popular algorithm, clique percolation algorithms faces a number