Physica A 491 (2018) 361–376

Contents lists available at ScienceDirect

Physica A

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

Partition network into communities based on group action on sets

Shu-Cheng Lin [a], Han-Wen Tuan [b], Cheng-Tan Tung [c], Peterson Julian [d,*]

[a] Department of Hotel Management, Lee-Ming Institute of Technology, Taiwan
[b] Department of Computer Science and Information Management, Hungkuang University, Taiwan
[c] Department of Information Management, Central Police University, Taiwan
[d] Department of Traffic Science, Central Police University, Taiwan

HIGHLIGHTS

- We use modularity values to solve the incomplete algorithm of Zhang et al. (2011).
- Our revised algorithm reduced the computation amount of modularity values.
- We provide our intersection–union operation to replace the GAS algorithm.
- The GAS algorithm of Zhang et al. (2011) will result in too many cores in the initial stage and then too many communities.
- We point out the Burnside Theorem, cited in Zhang et al. (2011) that is no use to decide the number of communities.

ARTICLE INFO

Article history:
Received 2 April 2016
Received in revised form 12 June 2017
Available online 29 September 2017

Keywords:
Community in network
Intersection–union operation
Group action on sets (GAS)
r-cycle

ABSTRACT

In this paper an improved algorithm is provided to detect communities within a network based on group action on sets (GAS). Modularity has been used as the criterion to revise the results of three previous papers, deriving a better method of partition for the network of Karate club. We developed a new method to replace the complicated GAS to achieve the same effect as GAS. Through four examples, we demonstrated that our revised approach reduced the computation amount of modularity values. Based on a branch marked example, a detailed example is provided by us to illustrate that there is too many cores in the initial stage of GAS approach to induce too many communities in the final partition. The findings shown here, will allow scholars to understand using GAS algorithm to partition a network into communities is an unreliable method.

1. Introduction

A group of people who share common values or interests can form a human community. A human community, thereupon, can represent a type of political power, social power, consumption tendency, consumption behavior transference, commercial opportunity, new demand formation, value and preference change, or new type social- or scientific-oriented cohesion. All living organisms can be attributed to a community. For instance, each type of bacteria or virus can be classified into different communities, each of them having a unique set of survival condition, parasitical conditions, influential objects and extents, proliferation speed and likely mutation. The identification process of various communities has been the focus of researchers, creating different approaches for different purposes. In order to study communities in social, computer,

* Corresponding author.
E-mail address: una211@sun4.cpu.edu.tw (P. Julian).

https://doi.org/10.1016/j.physa.2017.08.124
0378-4371/

metabolic, and regulatory networks, Newman [1] presented a mathematical technique to investigate the community structure which differentiates dense links from sparse links among numerous networks of scientific interest. Radicchi et al. [2] proposed a quantitative local algorithm for the identification of community structures, claiming that it performs better than existing approaches and contributes to the application of large-scale technological and biological systems. By focusing on community structures with node-joined networks in tightly knit groups, among which there are only looser connections, Girvan and Newman [3] developed a method for detecting such communities built around the idea of using centrality indices to find the boundaries of those communities. In the application of living organic communities, Risatti et al. [4] used hybridization probes to complement various phylogenetic groups and directly determined the abundance and distribution of sulfate-reducing bacterial populations in a microbial non-overlapping mat community. In the pursuit of disclosing the modular structure of complex systems, Palla et al. [5] proposed a clique percolation for analyzing and masterminding the main statistical features of the interwoven sets of communities. Donetti and Muñoz [6] used spectral properties of the graph, Laplacian matrix, and hierarchical-clustering techniques to maximize the modularity of the output. It turns out that the algorithm they presented was quite encouraging in terms of accuracy and computer running time for detecting and analyzing communities and modular structures in complex networks. Other essential means for identifying and detecting communities include local techniques [7,8] betweenness-based methods [9], and so on. Zhang et al. [10] proposed a GAS algorithm to detect communities in highly-clustered networks, such that they utilized the community separability to reveal the inherent structure of a network and the modularity to evaluate the quality of the community division. They claimed that GAS algorithm was shown to be more accurate and effective in the detection of communities in high clustered networks. The aim of this study is to point out seven relevant questionable results of Zhang's work [10] and provide improvements. The first, the algorithm presented by Zhang et al. [10] only considered one of the many possibilities in real world situations, when building up communities from a particular network while our explanation exhausts all aspects when carrying out the GAS processes of community division of Zhang's work [10]. The second, their approach in the formation of targeted communities through the fulfillment of adding isolated orbits into non-isolated orbits will imply many possible partitions. The third, the GAS to generate a group from a finite set will be a tedious task and then we provide a simplified method to obtain the same effect. Therefore, this study provides a new and simplified method to achieve goal-finding. The fourth, we show that their method to compute the difference between community separability and modularity is unnecessary. The fifth, following the branch marked example recommended by Girvan and Newman [3], this study shows a GAS algorithm of lower than 1% to get the predesigned communities. The sixth, we pointed out that the Burnside Theorem (cited in Zhang et al. [10]), which is useless to predict the number of communities in a network. The seventh, based on the branch marked example with our 3-2-3 simplification, we compared the results that derived by GAS method of Zhang et al. [10] and that of Newman [11] to indicate that there are 9 communities by GAS and 4 communities by Newman [11]. By our performance estimation Zhang et al. derived 34.4% desired partition less than 75% derived by Newman [11]. These findings will help researchers choose algorithms to derive communities from a network.

2. Review of Zhang et al. [10]

We will first explain in detail how Zhang et al. [10] applied GAS of nodes to obtain the cores of communities and then merged isolated nodes into existing cores as a base to derive a partition of the network into communities. For this part, we will point out two questionable results in Zhang et al. [10]: (a) The proof for non-overlapping after GAS is missing, and (b) Their GAS algorithm may imply many possible partitions.

An r-cycle (r click) in a network is defined as a subset of r-nodes, such that there is a link between every two nodes in this subset.

In Zhang et al. [10], they selected a number, r from $\{3, 4, 5\}$ and collected all r-cycles in the network as base elements, denoted as X, to develop a finite group, denoted as $G = \langle X \rangle$. They then applied group G to act on set of nodes to derive the orbit, denoted as $\Theta(x)$, of a node, x, as follows

$$\Theta(x) = \{gx : g \in G\}. \tag{1}$$

The set of nodes become a disjoint union of obits.

However, Zhang et al. [10] did not provide a proof for the disjoint union. Therefore, we will show that GAS implies the disjoint union of orbits.

Lemma 1. *If $\Theta(x)$ and $\Theta(y)$ intersect, then $\Theta(x) = \Theta(y)$.*

Proof. Let z be an element in the intersection of $\Theta(x)$ and $\Theta(y)$, then there are two elements g_1 and g_2 in the group G such that

$$z = g_1 x = g_2 y. \tag{2}$$

For an arbitrary point in the orbit $\Theta(x)$, $g_3 x$, we will show that there is an element, g_4, in the group G that satisfies

$$g_3 x = g_4 y. \tag{3}$$