ARTICLE IN PRESS

Physica A ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

Physica A

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

Consensus-based methodology for detection communities in multilayered networks

Amir-Mohsen Karimi-Majd [a], Mohammad Fathian [a,*], Masoud Makrehchi [b,1]

[a] School of Industrial Engineering, Iran University of Science and Technology, Narmak, Tehran, Iran
[b] Department of Electrical, Computer, and Software Engineering, University of Ontario Institute of Technology UOIT, Oshawa, ON, Canada

HIGHLIGHTS

- Finding communities in multilayer networks, at each layer and in parallel.
- Using a consensus clustering approach to aggregate layers.
- Different behavior of users in different layers could be detected.
- The methodology would be able to handle missing values.

ARTICLE INFO

Article history:
Received 26 May 2017
Received in revised form 21 October 2017
Available online xxxx

Keywords:
Community detection
Multilayered network
Consensus-based clustering
Heterogeneous networks
Social network analysis

ABSTRACT

Finding groups of network users who are densely related with each other has emerged as an interesting problem in the area of social network analysis. These groups or so-called communities would be hidden behind the behavior of users. Most studies assume that such behavior could be understood by focusing on user interfaces, their behavioral attributes or a combination of these network layers (i.e., interfaces with their attributes). They also assume that all network layers refer to the same behavior. However, in real-life networks, users' behavior in one layer may differ from their behavior in another one. In order to cope with these issues, this article proposes a consensus-based community detection approach (CBC). CBC finds communities among nodes at each layer, in parallel. Then, the results of layers should be aggregated using a consensus clustering method. This means that different behavior could be detected and used in the analysis. As for other significant advantages, the methodology would be able to handle missing values. Three experiments on real-life and computer-generated datasets have been conducted in order to evaluate the performance of CBC. The results indicate superiority and stability of CBC in comparison to other approaches.

1. Introduction

Community detection has been earlier defined as finding groups of nodes with many edges among themselves. Today, network communities refer to groups of people who interact with each other through various ways: sharing a video, commenting on a photo, liking a post, and so on. Moreover, they could interact not only with people they know but also with those who do not know them. They also differ from each other in terms of preferences, opinions and social behavior.

* Corresponding author.
E-mail addresses: karimimajd@alumni.iust.ac.ir (A.-M. Karimi-Majd), fathian@iust.ac.ir (M. Fathian), masoud.makrehchi@uoit.ca (M. Makrehchi).
[1] Address: 2000 Simcoe Street North Oshawa, Ontario L1H 7K4 Office: ENG 1024.

https://doi.org/10.1016/j.physa.2017.11.130
0378-4371/

Detecting communities, in the presence of such network heterogeneities, seems to become more problematic but more interesting than in the past.

Due to the interest, researchers have done some research in detecting communities in heterogeneous networks. As the first article, Hanisch et al. [1] focused on biological networks associated with gene expressions. Their approach of finding communities was to model such network as a dual-layer network (i.e., relationships and gene expressions), and then they simultaneously cluster the layers. This idea would be useful in many applications because all types of network layers and real-world cases could be classified into two groups: links and attributes. Most articles in this area have also adopted a similar approach [2–8]. Their common characteristic is that they focused on simultaneous analysis of two layers: links of objects and one type of attribute layers (e.g. topics, demographic attributes, contents, and users' behavior attributes). However, since real networks include several layers of different types, these methods suffer from a limited number of admissible layers. For instance, if a network includes two or more link (or attribute) layers, these methods will fail. Another group of articles modeled network as a combination of two or more link layers, without any attribute layer [9–11]. The main drawback of these methods is that they fail to directly analyze an attribute layer. Moreover, both group methods assume that all network layers imply same behavior. In other words, their main idea is that if we move toward finding communities in a layer, we will necessarily approach them to find communities in other layer(s) as well. However, in real networks, users' behavior may be different from layer to layer.

In order to cope with these issues, this article proposes a new methodology, called Consensus-based Community (CBC) detection. CBC relies on three ideas: (1) Multilayered network consisting of several distinct layers (2) each layer could be classified into two groups: link or attribute layer, thus, underlied patterns (i.e., communities) of each layer could be found by means of community detection (for link layers) or a clustering (for attribute layers) method, and (3) extracted patterns could be aggregated using a consensus approach. Based on these ideas, CBC finds communities among nodes at each layer, in parallel. The results of layers should be aggregated using a consensus clustering method. This means that the number of layers does not matter. Moreover, each layer would be separately analyzed, thus different behavior could be detected and used in the analysis. As for other significant advantages, the methodology would be able to calculate the effects of each layer on the final result, to determine which layers should be considered for detecting communities and to handle missing values.

In addition to the advantages of CBC, the performance of this methodology should be evaluated through some experiments. To do so, this article conducts nine experiments, seven on simulated datasets and two on real-life datasets, using a ranking method inspired by Yang and Leskovec [12]. In this sense, "different methods" would be evaluated and ranked based on their results achieved in terms of some metrics.

The remainder of this article is organized as follows. Background is presented in Section 2. Section 3 provides a comprehensive description of CBC and its components. We provide our evaluation methodology and the results obtained from the experiments in Section 4. Section 5 presents our concluding remarks.

2. Background

Network could be modeled as a set of nodes with their relationships. Real-life networks include meaningful patterns, so-called community, defined as a dense part in terms of relationships among nodes. Detecting such patterns leads to understand a network's structures and functionality. To do so, one can define a scoring function that measures density of detected sub-structure. This approach seems to be similar to clustering. Thus, a large number of methods developed for detecting communities, directly or indirectly, refer to clustering methods. Spectral clustering [13] and Hierarchical clustering [14] are two well-known branches of clustering, which are frequently used in the area of community detection. Although clustering and community detection are alternatively used to refer to the same concept, clustering would mainly be employed for discovering natural groups of objects based on their attributes and not on their relationships.

In the literature on community detection, most of the methods focused on homogeneous networks (i.e., networks that own only simple nodes with simple relationships). These methods could be classified into seven main groups as proposed by Fortunato [15] — "Traditional methods, Divisive algorithms, Modularity-based methods, Dynamic algorithms, Statistical inference, Spectral algorithms and Alternative methods". Although the number of these methods is very high, real-life networks are not basically homogeneous but heterogeneous. Network heterogeneity may be because of four causes: (1) node differences in terms of their attributes (e.g. age of users), (2) edges' difference in terms of their performance (e.g., friendship or sharing a video), (3) a combination of them, and (4) node differences in terms of their types (e.g., user, video, comment, and so on). The last one is called multi-mode networks [16]. Three other would be grouped into multilayered networks, which are in focus in this article. The other names of multilayered networks are as follows: multi-dimensional networks [16], multiplex networks [17], and multi-relational networks [18].

In order to clearly describe multilayered networks, Fig. 1 illustrates a very small network as an example. In this figure, layers of a three-layer network including fourteen nodes have been separately depicted. The first layer, the so-called Links, contains links among nodes. The next layer, so-called Attributes, shows different attribute values of nodes by means of different colors. The third layer, so-called Comments, shows which node has commented on a post of another person. This figure also indicates that the problem of community detection in multilayered network would be much more complex than in single-layer (or homogeneous) network.

Based on the authors' knowledge, the first research in the area of community detection in multilayered network has been done by Hanisch et al. [1] in 2002. They developed a co-clustering method for finding communities in biological networks,

Please cite this article in press as: A.-M. Karimi-Majd, et al., Consensus-based methodology for detection communities in multilayered networks, Physica A (2017), https://doi.org/10.1016/j.physa.2017.11.130.