Physica A 413 (2014) 565–576

Contents lists available at ScienceDirect

Physica A

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

Evolutionary community structure discovery in dynamic weighted networks

Chonghui Guo *, Jiajia Wang, Zhen Zhang

Institute of Systems Engineering, Dalian University of Technology, Dalian 116024, China

HIGHLIGHTS

- We propose a novel evolutionary community structure discovery algorithm.
- This algorithm is for dynamic weighted networks whose number of nodes and communities is changing.
- The number of communities can be found automatically by our algorithm.
- Two measures for evolutionary community structure discovery are proposed.
- This algorithm performs more effectively and feasibly in synthetic and real-word complex networks.

ARTICLE INFO

Article history:
Received 18 February 2014
Received in revised form 23 May 2014
Available online 9 July 2014

Keywords:
Dynamic networks
Evolution
Community structure
Weighted networks

ABSTRACT

Detecting evolutionary community structure in dynamic weighted networks is important for understanding the structure and functions of networks. In this paper, an algorithm which considers the historic community structure of networks is developed to detect evolutionary community structure in dynamic weighted networks. In the proposed algorithm, two measures are proposed to determine whether to add a node to a community and whether to merge two communities to form a new community. The proposed algorithm can automatically discover evolutionary community structure in weighted networks whose number of nodes and communities is changing over time and does not need to determine the number of communities in advance. The algorithm is tested using a synthetic network and two real-word complex networks. Experimental results demonstrate that the proposed algorithm can discover evolutionary community structure in dynamic weighted networks effectively.

1. Introduction

Many complex systems can be described as networks, such as social networks, technological networks, and biological networks. One of the common properties of these networks is community structure, which refers to partitioning the nodes in a network into some communities. The connections within a community are dense, while the connections between different communities are sparse [1–6]. Finding community structure in networks has received increasing attention from scholars, because it is crucial for understanding the behavior of the system and the properties of individual communities which may be quite different.

* Corresponding author. Tel.: +86 411 84708007.
E-mail addresses: dlutguo@dlut.edu.cn, guochonghui@tsinghua.org.cn (C. Guo), jjwang@mail.dlut.edu.cn (J. Wang), zhangzhen@mail.dlut.edu.cn (Z. Zhang).

http://dx.doi.org/10.1016/j.physa.2014.07.004

Research on methods for discovering groups in static networks has a long history. Various community detecting algorithms have been proposed, including spectral partitioning [7,8], information-theory based algorithm [9], Markov clustering [10], and so on. In addition, hierarchical clustering [3,7,11] is one of the classical methods, which has been applied to social and biological networks. Newman and Girvan [7] proposed a spectral bisection method based on modularity matrix and showed good results for bisections. However this method is less accurate when the number of communities is larger than two. A detailed introduction about these methods can be found in Refs. [12,13].

However, in many complex networks, the interactions between entities evolve dynamically over time. Let us take Facebook as an example. Users can register on the Facebook, add or delete "friends" and can also communicate with "friends". This can construct a dynamic network in which users are nodes and the frequency of communication is the weight of edges. After a period of time, some users may cancel the Facebook account and decrease the frequency of communications between "friends". Under such circumstances, the number of nodes, the weight of edges and the number of edges are changing with time. Similarity, new forms of social networks can be observed in phone calls, E-mail exchanges and scientific research cooperation. As dynamic features are crucial to understand complex networks, much research has been conducted on community structure discovery in dynamic networks.

In general, algorithms for detecting community structure in dynamic networks can be divided into two parts. On one hand, a two-stage strategy is used, in which community structure is independently extracted at consecutive timesteps. Evolutionary characteristics are introduced to trace the same community at consecutive timesteps and explain the difference between these community structure over time. In recent years, some two-stage algorithms are proposed to detect evolutionary community structure [14–17]. Gergely et al. [14] developed a new algorithm based on clique percolation, that allows, for the first time, to investigate the time dependence of overlapping communities on a large scale and as such, to uncover basic relationships characterizing community evolution. Duan et al. [15] focused on the community mining including community discovery and change-point detection on dynamic weighted directed graphs. In these studies, community extraction and community evolution are analyzed in two separate stages. Historic community structure, which contains valuable information related to current community structure, is not taken into account.

On the other hand, evolutionary clustering is a useful approach to discover evolution community structure, which processes the data at each timestep to generate a sequence of clustering. To obtain stable and consistent clustering results, the clustering results at each timestep not only should remain faithful to the current data as much as possible, but also should not shift dramatically from one timestep to the next timestep. To obtain evolutionary clustering results, many methods have been proposed [18–22]. Traditional clustering revisited for an evolutionary setting refers to initial evolutionary clustering efforts that modify existing static clustering algorithms (such as k-means) to remember previous data states [18]. Spectral clustering introduces and minimizes clustering cost functions by combining graph-based measures, such as a normalized graph cut, with temporal smoothness regularization terms [19]. Block-model approximation acts in multimode networks with cross-mode interactions among nodes as interactions among blocks comprising nodes of the same mode (i.e., single-mode communities) [20]. Nonnegative-matrix/tensor factorization based methods discover communities as latent variables by jointly maximizing the fit to the observed data and temporal evolution [21].

To our knowledge, Chakrabati et al. [16] are the first to address evolutionary clustering problems in data mining literature, where the cluster membership at timestep t is influenced by the clusters at timestep $t-1$. It captures continuity with respect to previously learned model through the notion of temporal smoothness. The temporal smoothness is incorporated into the objective function to be optimized during the learning process. Based on similar ideas, Chi et al. [17] proposed the first evolutionary spectral clustering algorithm by incorporating the temporal smoothness constraint into the solution. In particular, they proposed two aspects of model quality or, equivalently, model cost, i.e. snapshot cost which captures the quality of the clustering learned at each time interval and temporal cost which measures the similarity between a clustering learned at a particular timestep with that learned at the previous timestep. The evolutionary clustering problem is transformed into the optimization problem of finding a sequence of models that minimizes the overall cost.

The above mentioned two types of methods have some common weaknesses. First, it is assumed that the number of nodes and clusters remain unchanged over time. Second, the number of the clusters should be known in advance. Finally, both of the methods cannot be applied to weighted networks whose edges have different strengths. Besides, just as Newman pointed out [8], spectral partitioning is a poor approach to detect natural community structure in real-world networks, because cut sizes do not accurately reflect the intuitive concept of network communities, which are simply the right thing to optimize.

Inspired by evolutionary clustering and the temporal smoothness, this paper proposes a new algorithm to discover evolutionary community structure in weighted dynamic networks. In this paper, we focus on weighted networks whose number of nodes, weight of edges and number of communities are different from one timestep to the next timestep. The main strategies are as follows. First, evolutionary matrices are built as the input which consider the previous community structure, and then an initial community with a node whose node strength is maximum is discovered. Afterwards, the community is expanded by adding nodes that can improve the quality of the community. Finally, the communities whose numbers of nodes are smaller than a threshold are merged to improve the total quality.

The rest of this paper is organized as follows. In Section 2, we describe some concepts related to evolutionary community structure discovery. In Section 3, we propose two measures for evolutionary community discovery and describe the details of the evolutionary community structure discovery algorithm. In Section 4, the algorithm is applied to three datasets and the experimental results are provided. Finally, we conclude this paper in Section 5.