Physica A 407 (2014) 7–14

Contents lists available at ScienceDirect

Physica A

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

Uncovering overlapping community structures by the key bi-community and intimate degree in bipartite networks

Yaozu Cui, Xingyuan Wang *

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

HIGHLIGHTS

- Introduce the key bi-communities, the free-nodes and the intimate degree *IND*.
- Propose the algorithm 1 and the algorithm 2 for detecting the community structures in the bipartite networks.
- Find overlapping vertices.
- Give excellent experimental results.

ARTICLE INFO

Article history:
Received 31 December 2013
Received in revised form 3 March 2014
Available online 28 March 2014

Keywords:
Bipartite networks
Community structures
Key bi-communities
Free-nodes

ABSTRACT

Although many successful algorithms have been designed to discover community structures in network, most of them are dedicated to disjoint and non-overlapping communities. Very few of them are intended to discover overlapping communities, particularly the detection of such communities have hardly been explored in the bipartite networks. In this paper, a novel algorithm is proposed to detect overlapping community structures in bipartite networks. After analyzing the topological properties in bipartite networks, the key bi-communities and free-nodes are introduced. Firstly, some key bi-communities and free-nodes are extracted from the original bipartite networks. Then the free-nodes are allocated into a certain key bi-community by some given rules. In addition, the proposed algorithm successfully finds overlapping vertices between communities. Experimental results using some real-world networks data show that the performance of the proposed algorithm is satisfactory.

1. Introduction

The natural systems containing a large amount of entities and interaction among entities can be modeled as complex networks [1–3], such as the internet, the World Wide Web, social networks and biochemical networks. In other words, any network can be regarded as a graph, where a vertex (or node) represents an entity of the complex networks, and an edge represents a link between nodes in a defined relationship of the complex networks. Although these complex networks come from very different domains, such as physics, mathematics, biology, and medical and computer [4–6], the community structure is a more common property of many complex networks [7,8]. A community is made of a group or cluster of nodes within which the links between nodes are densely connected to each other, but between which they are sparsely connected with other communities [9–11].

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

http://dx.doi.org/10.1016/j.physa.2014.03.077

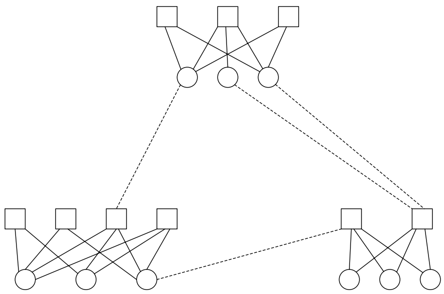
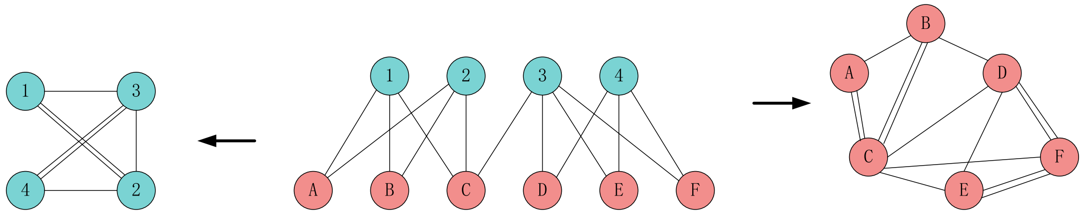


Fig. 1. A schematic representation of a bipartite network with the community structures. There are three communities of dense internal links (solid lines) with sparse connections (dot lines) among them.



(a) The projecting results of scientists.

(b) The schematic diagram of the scientist-collaboration bipartite network.

(c) The projecting results of papers.

Fig. 2. A small scientist-collaboration bipartite network and its projections. The numbers represent scientists and the letters represent the co-authored papers.

According to the number of the node type in complex networks, complex networks can be divided into one-mode networks and bipartite networks. Along with researching physical significance and some properties of the complex networks, it was found that many real-world networks appear natural bipartite structure, such as the artistic-collaboration networks [12,13], the scientists-collaboration network [14–16], the football teamers-team network [17], and the protein–protein interaction networks [18,19]. A bipartite network is a network with the two non-overlapping types of nodes and the edges only connecting a pair of nodes which belong to different types. According to the community structures in the one-type networks, the community structures in the bipartite networks can also be defined according to the connection of relative density [20]. The communities are a set of nodes in the bipartite networks, in which the internal connection between the nodes are more intensive, but between which the nodes connection is sparse as shown in Fig. 1.

To search the properties of the bipartite networks, a kind of traditional approach is adopted to project them into one-mode networks using un-weighted projection or weighted projection. For example, Newman empirically researched on the scientist-collaboration network in Refs. [12–14], in which two different types of nodes represent scientists and their papers respectively. And Newman projected this bipartite network into the one-mode and un-weighted network. Two scientists have edged when they have a co-authored paper during projection shown in Fig. 2. In addition, he pointed out that these un-bipartite projections cannot fully reflect the cooperation intensity between scientists. The one-mode projection of a bipartite network loses some information of the original bipartite networks, brings an inflation of the number of edges and adds some information which does not belong to the original bipartite networks [21].

At present, another kind of methods is proposed to detect the community structures directly based on the original bipartite networks. In Ref. [20], the clustering method detecting community in the bipartite networks is proposed, which based the edge clustering coefficient in the bipartite networks. The process of this method is similar to GN [22] algorithm which detects the community structures in the one-mode networks. This method recursively removes edges with the minimum of the edge clustering coefficient until the modularity of the bipartite networks with local maximum. But this method determines the number of community according to the modularity. In 2008 Lehmann et al. [23] define a k-clique community in the bipartite networks extending from Palla's definition [11]. And they propose a biclique detection method allowing the overlapping community structures. But the adjustable parameters (a, b) of this method are not easy to determine in different networks.

Consequently, the main contributions of this paper pay more attention to detect overlapping community structures directly in the bipartite networks. The real-world bipartite networks from different domains can give us different news. For example, the community structures show that how many scientists prefer to collaborate in what kind of researching areas in the scientific researching cooperation network. In this paper, a new algorithm based the key bi-communities and free-nodes are proposed to detect overlapping community structures directly from the bipartite networks. The rest of the