Information Sciences 317 (2015) 278–294

Contents lists available at ScienceDirect

Information Sciences

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

Density-based modularity for evaluating community structure in bipartite networks

Yongcheng Xu [a], Ling Chen [a,b,*], Bin Li [a,b], Wei liu [a]

[a] Department of Computer Science, Yangzhou University, Yangzhou 225009, China
[b] State Key Lab of Novel Software Tech., Nanjing University, Nanjing 210093, China

ARTICLE INFO

Article history:
Received 8 August 2014
Received in revised form 20 April 2015
Accepted 25 April 2015
Available online 2 May 2015

Keywords:
Bipartite network
Community
Modularity

ABSTRACT

A bipartite network is an important type of complex network in human social activities. Newman defined modularity as a measurement for evaluating community structure in unipartite networks. Due to the success of modularity in unipartite networks, bipartite modularities were developed according to different understandings of community in bipartite networks. However, these modularity measurements are subject to resolution limits that could reduce the quality of community partitioning. These modularity measurements contain an intrinsic scale that depends on the total size of links and ignores the number of nodes in a bipartite network. In this paper, we first illustrate such resolution limits of traditional bipartite modularities using several examples of bipartite networks. Next, we propose a quantitative measurement called density-based modularity to evaluate community partitioning in bipartite networks. We verify that optimization of the density-based modularity proposed has no resolution limit. By optimizing this density-based modularity, we can partition the network into the appropriate communities. Experiments on synthetic and real-world bipartite networks verify the accuracy and reliability of our bipartite density-based modularity.

1. Introduction

Various types of real-world complex systems could be modeled as complex networks [27], in which nodes represent the objects and edges represent the interactions among these objects. A bipartite network is an important type of complex network in real-world systems, in which the nodes are divided into two types. Fig. 1 shows the actor-film bipartite network. In the network, there are two types of nodes, which are colored black and yellow, respectively. The black nodes represent actors, whereas the yellow nodes represent films. In a bipartite network, all edges only connect nodes of different types; no nodes of the same type are connected. Many real-world networks are naturally bipartite networks, such as the scientists-papers cooperation network [21,30], actor-films network [23], disease-gene network [7,11,39], club members-activities network [31], reader-books network [20], artistic–music network [16], and investors-company network [35], etc.

An important feature of real-world networks is community structure [1,19,28,34,36]. In a network, a community is a cluster of nodes that are densely connected to one another and sparsely connected to nodes outside the cluster. Because many networks demonstrate a community structure, the characterization and detection of this community structure have

* Corresponding author at: Department of Computer Science, Yangzhou University, Yangzhou 225009, China.
E-mail addresses: ychxu08@126.com (Y. Xu), yzulchen@163.com (L. Chen).

http://dx.doi.org/10.1016/j.ins.2015.04.049
0020-0255/

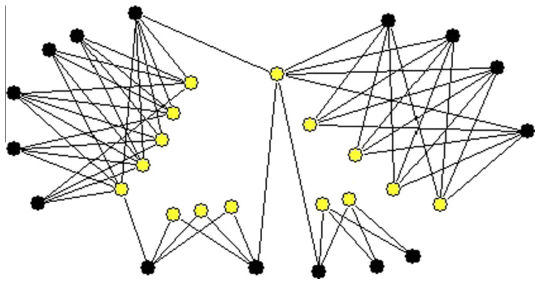

Fig. 1. The actor-film bipartite network.

attracted great interest in the area of complex network analysis. Moreover, because communities are relatively independent of one another structurally, it is believed that each of them may correspond to a fundamental functional unit. Identifying and analyzing these communities from a large network provides a means for the functional dissection of a network and sheds light on its organizational principles.

To evaluate the qualities of detected communities from unipartite networks, Newman [26] introduced a measurement called modularity, which compares the number of links inside a community with the expected value of a randomized graph of the same size and degree sequence. Based on the concept of modularity, the problem of community detection becomes equivalent to modularity optimization. Therefore, several methods have been proposed to find the optimal solution with the lowest computational cost [3,6,10,15,17,33,32,40]. Fortunato and Barthélemy [12] found that such modularity contains an intrinsic scale that depends on the total size of links in a network. Clusters smaller than this scale may not be resolved, even in the extreme case in which they are complete graphs. Li et al. [22] proposed an improved measurement that was claimed to be superior to the widely used modularity concept developed by Newman. The authors also proved the equivalence of their measurement with the objective function of the kernel k means. Although Newman's modularity suffers from the resolution limit problem, it has been widely accepted as a de facto standard.

In a bipartite network, nodes can be divided into two types such that no two nodes of the same type are adjacent. In recent years, many methods have been proposed to detect communities in bipartite networks [4,8,9,18,24]. To evaluate the quality of the results of community detection in a bipartite network, some researchers have extended the measurement of modularity to bipartite networks [2,14,25,37]. According to different understandings of bipartite network community, different versions of bipartite modularity have been proposed by various researchers, such as Guimera et al. [14], Barber [2], and Murata [25]. These bipartite modularities are based on different assumptions regarding communities in bipartite networks. Therefore, their quantification remains a controversial issue. Two aspects greatly complicate this problem. One is that the size heterogeneity of communities often greatly affects the measure of community. The other aspect is that, even in a specific network, the generation mechanism or link degree may vary greatly.

Fortunato and Barthélemy [12] claimed that modularity for a unipartite network suffers resolution limits that affect the quality of community partitioning. We found that such resolution limits also exist in the modularity measurements proposed for bipartite networks. When using those modularity measurements for bipartite networks, the communities that are smaller than the scale of bipartite networks may not be resolved, even in the extreme case in which they are complete bipartite subgraphs.

In this work, we first demonstrate that the existing modularity measurements for bipartite network community partitioning, such as Murata's modularity and Barber's modularity, suffer from resolution limits. Using certain bipartite networks as examples, we illustrate that Murata's and Barber's measurements cannot ensure that optimal partitioning can yield the highest modularity value. To overcome this limit, we propose a quantitative measurement for evaluating community partitioning in bipartite networks based on the concept of the average bipartite modularity degree. We refer to this quantitative measurement as density-based bipartite modularity. Through theoretical analysis and empirical tests on synthetic and real-world networks, we show that the presented density-based modularity overcomes the resolution limit in bipartite community structure detection.

The remainder of this paper is organized as follows. Section 2 provides a brief review of existing studies on community detection in bipartite networks. Section 3 presents reported measurements pertaining to community partitioning in bipartite networks. Section 4 illustrates the resolution limit of these measurements using three examples of bipartite networks. Section 5 proposes a density-based modularity for evaluating community partitioning in bipartite networks. Section 6 illustrates the rationality and reliability of the density-based bipartite modularity using three examples. Section 7 formulates the optimization of density-based modularity as a nonlinear program. Section 8 presents and analyzes the experimental results obtained for both synthetic and real bipartite networks. Section 9 presents the paper's conclusions.

2. Related works

Community structure detection is closely related to the graph-partitioning problem in graph theory, in which nodes of a graph are divided into communities by finding the best edge cuts of the graph based on several edge cut objectives [5], such