Physica A 496 (2018) 108–120

Contents lists available at ScienceDirect

Physica A

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

A fast community detection method in bipartite networks by distance dynamics

Hong-liang Sun [a,b], Eugene Ch'ng [a], Xi Yong [c], Jonathan M. Garibaldi [d], Simon See [e,f], Duan-bing Chen [g,*]

[a] NVIDIA Joint-Lab on Mixed Reality, International Doctoral Innovation Centre, PR China
[b] School of Computer Science, University of Nottingham, Ningbo, 315100, PR China
[c] Water Information Center, Ministry of Water Resources, Beijing, 100053, PR China
[d] School of Computer Science, University of Nottingham, Nottingham, NG8 1BB, UK
[e] NVIDIA AI Technology Centre, NVIDIA, Singapore 138522, Singapore
[f] Center for High Performance Computing, Shanghai Jiao Tong University, Shanghai 200240, PR China
[g] Web Sciences Center, Big Data Research Center, University of Electronic Science and Technology of China, Chengdu, 611731, PR China

HIGHLIGHTS

- Out method extends distance dynamics model from *unipartite networks* to *bipartite networks*.
- BiAttractor is 51.5 times faster than Adaptive BRIM, 54.86 times faster than LP Brim and 45.5 times faster than AsymIntimacy in American Revolution network (AR) with hundreds of vertices and edges.
- Our method can detect small communities with high accuracy with respect to resolution limit.

ARTICLE INFO

Article history:
Received 1 August 2017
Received in revised form 24 November 2017
Available online 30 December 2017

Keywords:
Node similarity
Community detection
Large bipartite networks

ABSTRACT

Many real bipartite networks are found to be divided into two-mode communities. In this paper, we formulate a new two-mode community detection algorithm BiAttractor. It is based on distance dynamics model Attractor proposed by Shao et al. with extension from unipartite to bipartite networks. Since Jaccard coefficient of distance dynamics model is incapable to measure distances of different types of vertices in bipartite networks, our main contribution is to extend distance dynamics model from unipartite to bipartite networks using a novel measure Local Jaccard Distance (LJD). Furthermore, distances between different types of vertices are not affected by common neighbors in the original method. This new idea makes clear assumptions and yields interpretable results in linear time complexity $O(|E|)$ in sparse networks, where $|E|$ is the number of edges. Experiments on synthetic networks demonstrate it is capable to overcome resolution limit compared with existing other methods. Further research on real networks shows that this model can accurately detect interpretable community structures in a short time.

1. Introduction

The science of networks is a fundamental discipline across biology, social sciences, computer science and other fields. Networks represent various complex systems in different disciplines [1–3]. Networks inferred from complex systems consist

* Corresponding author.
E-mail address: dbchen@uestc.edu.cn (D.-b. Chen).

https://doi.org/10.1016/j.physa.2017.12.099

of vertices and edges, which represent entities and relationships. A pair of vertices is connected by an edge if they have a certain relationship [4,5]. Biological scientists seek to understand the associations between all known phenotypes and disease genes from a network of disorders and disease genes [6]. While social scientists study the behavioral patterns of different groups of users from online social networks of acquaintanceships. Many other examples come from politics, economics, marketing, computer science, transportation and etc.

A number of research focus on *unipartite network* or *one-mode network*, which contains only one type of vertices. However, real-world networks usually contain multiple types of vertices. The simplest case is *bipartite* or *two-mode network* containing two different kinds of vertices. Connections within bipartite only occur between different types of vertices and there are no edges between the same type of vertices [7].

A key property of most bipartite networks is community structures, where networks are divided into groups of vertices and edges. Community is modular structure of underlying networks where there are dense connections between vertices within the same group yet loose connections between different groups. Discovery of communities can detect coarse-grained sub-networks from underlying networks, which provides a mesoscopic view differing from network-level macroscopic and vertex-level microscopic views. On the other hand, different communities have distinct structural properties, thus global average properties are insufficient to present features of community structures [8]. Community also provides a better way to understand modular structures in bipartite networks. There are two kinds of ideas with respect to bipartite community. Members are considered as the same type or different type within the same group.

A number of methods have been proposed to identify unipartite communities with different assumptions [4], such as non-negative matrix decomposition [9], label propagation [10,11], expansion from seed sets [12], evolutionary method [13], game-theoretic approaches [14,15], line graphs [16,17], modularity optimization [18] and etc. Thus one simple idea comes from the projection method by Zhou et al. [19], which transforms bipartite networks to unipartite networks, then existing community detection methods from unipartite networks can be naturally employed [20]. However, researchers argued that projection methods might lead to the incomplete information problem because only one type of vertices have been applied yet another type of them are lost after the projection [21]. Thus various methods have been developed to maintain two types of vertices after the divisions of communities. Barber firstly proposed a bipartite modularity [7] extending from unipartite modularity [22], then BRIM algorithm has been developed to induce two independent parts of vertices into modular structures. However, modularity from bipartite networks has limitations of resolution issue [23] because small communities cannot be accurately detected with high modularity scores. Lehmann et al. presented a method for detecting biclique communities based on an extension of the k-clique community detection algorithm [24]. It remains all of the advantages of the k-clique algorithm and provides a level of flexibility by incorporating independent clique thresholds.

Numerous methods have been proposed from distinct assumptions, such as eigenvectors of matrices [25], modularity optimization [8], clustering coefficient [26], intimate degree [27], stochastic block model [28], density based modularity [29], asymmetric intimacy [21] and etc. Li et al. proposed an unified community detection method based on vertex similarity probability [30] to deal with both unipartite and bipartite networks together [31]. Although a number of methods have been proposed to detect communities, the mechanism governs the formation of communities has not been well understood. Wang et al. seek to understand the emergence of communities and diversity [32]. More detailed descriptions are beyond the scope of this paper. Interested readers can refer to review papers from Fortunato [4,5].

As the size of network increases rapidly, striking a balance between accuracy and performance has been significant in practice. Current research methods have paid little attention, except for the case of unipartite networks [33–37], to community discovery in large bipartite networks with hundreds of thousands vertices and edges. Pan et al. suggested an accurate and efficient method to discover communities in large unipartite networks using node similarity [30,38]. However, more methods are needed to deal with large bipartite networks efficiently. One of possible reasons of existing works comes from that the time complexity is at least quadratic, which always takes several hours to deal with large bipartite networks. They also have the problem of resolution limit, which leads to inaccurate detection of small communities [23].

In this paper, a novel method using distance dynamics has been proposed to detect two-mode communities in large bipartite networks. It is inspired by interactions in human society such that there are more interactions within the same community but less between different ones. It has time complexity $O(|E|)$ in sparse networks and obtains accurate partition of communities as well. Experiments demonstrate that it is faster than other methods in real sparse networks with thousands of vertices and edges. It also obtains at most 7.64% improvement of accuracy compared with Adaptive BRIM [39].

In Section 2, we will discuss community detection by dynamic distance in unipartite networks. Then a varied new method has been proposed to extend original method from unipartite to bipartite networks. In Section 3, our new method has been evaluated in both synthetic and real networks, especially in large bipartite networks with hundreds of thousands of vertices and edges. Finally, it concludes in Section 4 to summarize the main contributions and future directions.

2. Community discovery in bipartite networks

2.1. Distance dynamics in unipartite networks

In this section, we briefly describe the algorithm Attractor to discover communities of unipartite networks using distance dynamics. Attractor is the foundation of our proposed BiAttractor in the next section. The philosophical idea of Attractor is inspired by the view of community formation from sociology. A community of friends is usually established and intensified