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 due to consistent interactions. Such interactions are usually driven by an 'utility of will', 'sharing common interests' and etc. Interactions between persons in society offer a vivid image of dynamic distances among vertices in network science. Shao et al. thus proposed Attractor to uncover communities from underlying networks using distance dynamics [40]. The network is viewed as a dynamic system where the edge dynamics has been investigated. Due to the interactions of vertices with their neighboring vertices, the distances among them evolve over time and reach convergence in the end. Communities can be naturally discovered by removing inter-community edges.

Next, the primary steps of Attractor have been introduced. Given an undirected and unweighted network $G = (V, E)$ where V is the set of vertices and E representing the set of edges. Attractor aims to find a partition of V into C communities $V_1, V_2, \ldots, V_C$ based on distance dynamics.

1. Initially, each edge is associated with an initial distance according to Jaccard coefficient.

2. As time goes on, the distance of each edge changes due to the interactions among local neighboring vertices. The interaction patterns include influence from directly connected vertices, influences from common neighbors and exclusive neighbors.

3. In the end, all distances reach convergent states. Edges associated with zero indicate intra-community edges. Those associated with one indicating inter-community edges have been removed to form communities naturally.

Attractor has a parameter λ to determine the size and number of communities. It influences the exclusive neighbors as a threshold of positive effect or negative effect. The linear time complexity $O(|E|)$ enables it to be applied to very large sparse networks in a short time.

However, Attractor is merely applied to unipartite networks containing one type of vertices. Thus we generalize it to deal with bipartite networks in the next section.

2.2. Distance dynamics in bipartite networks

In this section, we extend Attractor to BiAttractor to deal with community detection in bipartite networks. Several preliminary definitions will be explained as following.

2.2.1. Preliminaries

$G = (U, V, E)$ is an undirected and unweighted bipartite graph, where U and V represent two sets of different types of vertices. Edge e belonging to E connects different types of vertices. There is no edge between vertices from the same set U or V. Attractor in the previous section depends on vertex neighbors and Jaccard Distance. Similarly, our method BiAttractor depends on vertex second order neighbors and a novel Local Jaccard Distance (LJD). Next, they will be introduced in detail respectively.

Definition 1 (*Neighbors of Vertex u and Vertex v*)**.** Neighborhood of vertex u consists of adjacent vertices of u. Similarly we can obtain neighborhood of vertex v. In bipartite networks, directly connected neighbors of vertex u belong to the other set V shown in Fig. 1(a). It is similar to directly connected neighbors of vertex v.

$$N(u) = \{v \mid v \epsilon V, u \epsilon U, (u, v) \epsilon E\} \tag{1}$$

$$N(v) = \{u \mid u \epsilon U, v \epsilon V, (u, v) \epsilon E\}. \tag{2}$$

Definition 2 (*Second Order Neighbors of Vertex u and Vertex v*)**.** Extension from neighbors to second order neighbors is shown in Fig. 1(b). If we define $N(u)$, neighbors of vertex u, as its adjacent vertices. Then second order neighbors $NN(u)$ are defined as neighbors of y, where y is the neighbors of u. The definition of $NN(v)$ is similar to $NN(u)$.

$$NN(u) = \{x \mid x \epsilon N(y),\ y \epsilon N(u),\ x \epsilon U, y \epsilon V \text{ and } u \epsilon U\} \tag{3}$$

$$NN(v) = \{x \mid x \epsilon N(y),\ y \epsilon N(v), x \epsilon V, y \epsilon U \text{ and } v \epsilon V\}. \tag{4}$$

Definition 3 (*Jaccard Distance*)**.** Given an undirected and unweighted bipartite graph $G = (U, V, E)$, the Jaccard Distance of vertex u and vertex v is defined as below in Eq. (5). $N(u) \cap N(v)$ is a null set if u is connected with v because $N(u)$ and $N(v)$ are different types of vertices.

$$d_{jac}(u, v) = 1 - \frac{N(u) \cap N(v)}{N(u) \cup N(v)}. \tag{5}$$

Definition 4 (*Local Jaccard Distance*)**.** Jaccard Distance works well in Attractor because common neighbors $N(u) \cap N(v)$ exist in unipartite networks. However, it is not the same case in bipartite networks. $N(u)$ and $N(v)$ are different types of vertices and $(N(u) \cap N(v)) = \varnothing$. However, u second order neighbors $NN(u)$ are the same types as v neighbors $N(v)$. Due to this point of view, we propose Local Jaccard Distance to deal with community detection in bipartite networks.

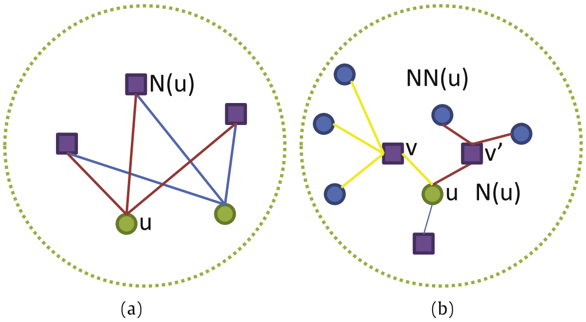


Fig. 1. (Color online) Neighbors and second order neighbors of given vertices. Circles and squares denote two types of vertices in bipartite networks. (a) Green circle vertex u and its neighbors $N(u)$ denoted by purple squares. (b) Blue circles indicate second order neighbors $NN(u)$ of u.

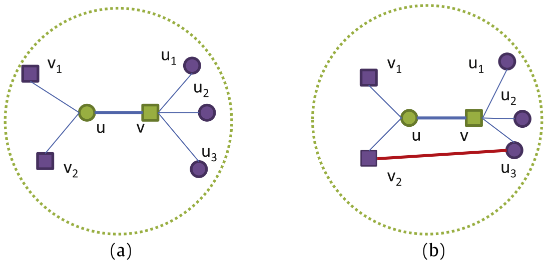


Fig. 2. (Color online) A toy sample to compare Jaccard Distance and Local Jaccard Distance in bipartite networks. (a) A sample bipartite network $G_1 = (U_1, V_1, E_1)$, where $|U_1| = 4$, $|V_1| = 3$ and $|E_1| = 6$. Vertex v_2 and u_3 are disconnected. (b) A sample bipartite network $G_2 = (U_2, V_2, E_2)$, where $|U_2| = 4$, $|V_2| = 3$ and $|E_2| = 7$. Vertex v_2 and u_3 are connected.

The Local Jaccard Distance (LJD) between vertex u and v is shown in Eq. (6). Exclusive neighbors of u is defined as $EN(u) = N(u) - \{v\}$. Similarly, we can get exclusive neighbors of v as $EN(v) = N(v) - \{u\}$. The Jaccard Distances between vertex u and exclusive neighbors of v are summed and divided by its exclusive neighbor number $|EN(v)|$. Because neighbors of u and second order neighbors of v are the same type of vertices. Similarly, Jaccard Distances from vertex v and exclusive neighbors of u are summed and divided by its exclusive neighbor number $|EN(u)|$. The mean value of those two parts is defined as LJD. In addition, second order neighbors are effective to deal with link prediction of complex networks [41,42].

$$d(u, v) = \frac{1}{2}\left(\frac{1}{|EN(v)|}\sum_{v'\epsilon EN(v)} d_{jac}(u, v') + \frac{1}{|EN(u)|}\sum_{u'\epsilon EN(u)} d_{jac}(v, u')\right). \tag{6}$$

To compare Jaccard Distance and LJD in bipartite networks, a toy sample is given below.

As shown in Fig. 2(a), $d_{jac}(u, v) = 1.000$ according to Eq. (5). Similarly, it is shown in Fig. 2(b) where $d'_{jac}(u, v) = 1.000$. Whether vertex v_2 and u_3 are connected or not do not affect d_{jac} in G_1 and d'_{jac} in G_2. But in fact the red edge between vertex v_2 and u_3 leads to shorter distance between u and v in G_2. Furthermore, we apply LJD to the same sample. As shown in Fig. 2(a), $d(u, v) = 0.708$ according to Eq. (6). Similarly, it is shown in Fig. 2(b) where $d'(u, v) = 0.590$. The $d'(u, v)$ of G_2 is smaller than $d(u, v)$ of G_1 because the red edge between vertex v_2 and u_3 makes u and v closer.

2.2.2. Interactive patterns

In this section, the interactive patterns of BiAttractor will be explained before we interpret the algorithm BiAttractor. BiAttractor is motivated by social interaction patterns in human society. People tend to have close relationships with their friends in the same community, but keep a certain distance away from strangers. Their relationships are not completely stable but change over time. Inspired by this idea, we think vertices are close to those vertices in the same community but are far away from those in different communities. Theoretically, close ties are indicated by intra-community edges and general relationships are indicated by inter-community edges.

It is necessary to identify the interactive scope before we introduce interactive patterns. It is impossible for a vertex to interact with every other vertex in the underlying network. Instead, It interacts with neighboring vertices to form communities. Local scope rather than global scope is selected in our method.

There are three different interaction patterns of Attractor including influence of directly linked vertices, influence of common neighbors and influence of exclusive neighbors [40]. However, there are no common neighbors of u and v in bipartite networks because $N(u)$ and $N(v)$ are different types of vertices according to Eq. (1). Thus, there are only two interactive patterns discussed in BiAttractor including influence of directly connected vertices and influence of exclusive neighbors.

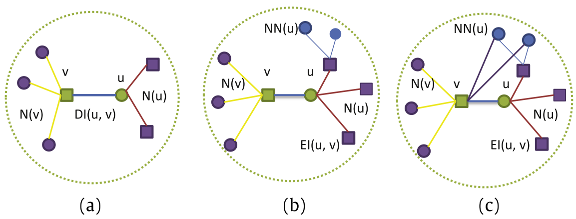


Fig. 3. (Color online) Interaction patterns of BiAttractor. Circles and squares represent different types of vertices. (a) Direct interaction between vertex u and v is highlighted in weighted solid lines. Green circle u is influenced by green square v and vice versa. (b) Negative influence between vertex u and its exclusive neighbors $N(u)$ is illustrated. The green circle u moves towards purple squares $N(u)$, but it moves far away from v. (c) Positive influence between vertex u and its exclusive neighbors $N(u)$ is displayed. The green circle u moves towards purple square $N(u)$. Because $NN(u)$ are strongly connected with the green square v, v is closer to u.

Definition 5 (*Direct Influence*)**.** The direct distance between vertex u and vertex v is defined as below. Direct influence $DI(u, v)$ leads to a phenomenon that u and v move close to each other because u and v are connected.

$$DI(u, v) = -(\frac{f(1 - d(u, v))}{deg(u)} + \frac{f(1 - d(u, v))}{deg(v)}). \tag{7}$$

It is similar to the definition proposed by Shao et al. [40], where $f(u)$ is $sin(u)$, $deg(u)$ is the degree of vertex u. The illustrative expression of this notion is shown in Fig. 3(a). We change the distance function $d(u, v)$ from Jaccard Distance to Local Jaccard Distance because Jaccard Distance always fails to denote node similarity in bipartite networks. The reason is illustrated previously when we present the definition of Local Jaccard Distance.

Definition 6 (*Influences from Exclusive Neighbors*)**.** Influence of edge (u, v) between vertex u and vertex v from exclusive neighbors is defined as Eq. (8). $d(u, v)$ is the Local Jaccard Distance, $deg(u)$ is the degree of vertex u and $f(u)$ is $sin(u)$. Dynamic distance comes from exclusive neighbors $EN(u) = N(u) - \{v\}$ and $EN(v) = N(v) - \{u\}$.

Similar to the influences from directly linked vertices, each exclusive neighbor attracts vertex u to move towards itself because they are connected. But we have no knowledge whether vertex u tends to move closer to vertex v or not. To illustrate the positive or negative influences of exclusive neighbors on the distance, a node similarity strategy is proposed by Shao et al. [40]. If vertex v is similar with exclusive neighbors of u, the movement from u to its exclusive neighbors leads to the decrease of the $d(u, v)$. Otherwise, if vertex v is not similar with exclusive neighbors of u, the movement from u to its exclusive neighbors results in increase of $d(u, v)$.

$$\begin{aligned} EI(u, v) = &- \sum_{x \epsilon EN(u)} \frac{1}{deg(u)} f(1 - d(x, u))\rho(x, u) \\ &- \sum_{y \epsilon EN(v)} \frac{1}{deg(v)} f(1 - d(y, v))\rho(y, v) \end{aligned}$$

$$\rho(x, u) = \begin{cases} 1 - d(x, v), (1 - d(x, v)) \geqslant \lambda \\ 1 - d(x, v) - \lambda, otherwise. \end{cases} \tag{8}$$

Negative influence of exclusive neighbors is shown in Fig. 3(b). Because there are no connections between v and second order neighbors of u $NN(u)$, $N(u)$ and v are not similar. Thus exclusive neighbors of u attract u to move closer to them, but u is far away from v.

Positive influence of exclusive neighbors is shown in Fig. 3(c). Due to the connections between v and second order neighbors of u, $NN(u)$, v and u are similar. Exclusive neighbors of u attract u to move towards them and u moves closer to v as well.

The motivation of introducing λ in Eq. (8) is to determine whether the influences of exclusive neighbors are positive or negative. If λ is small, exclusive neighbors of u have positive impacts to make u and v move closer. Because exclusive neighbors of u are similar with v, when λ is small. It is indicated by $\rho \geqslant 0$. Otherwise exclusive neighbors of u have negative impacts to make vertices u and v move far away. Negative influence is indicated by $\rho < 0$.

Definition 7 (*Dynamic Distance*)**.** Considering the influences from both directly connected and exclusive neighbors, we further explain the concept of dynamic distance. Dynamic distance $d^{t+1}(u, v)$ at time step $t+1$ is determined by the distance $d^t(u, v)$ at previous time step t, influences from direct interaction $DI^t(u, v)$ and exclusive neighbors $EI^t(u, v)$. The initial value of $d^t(u, v)$ is calculated by Local Jaccard Distance from Eq. (6). The distances update until they become convergent ($d^{t+1}(u, v) = 1$ or $d^{t+1}(u, v) = 0$). When all distances reach the stable states, the dynamic process terminates. Detailed algorithm can be described in Algorithm 1 in the next section.

$$d^{t+1}(u, v) = d^t(u, v) + DI^t(u, v) + EI^t(u, v). \tag{9}$$

Next, we illustrate the similarity and difference between Attractor and BiAttractor. The general framework of BiAttractor is identical with the original Attractor scheme, because we follow the same motivation idea. Both of them are motivated by social interaction patterns in human society. People tend to have close relationships with their friends in the same community, but keep a certain distance away from strangers. Thus the distances among vertices in the same community are small but those between different communities are large. When we extend the method from unipartite to bipartite networks, the following parts are modified.

Firstly, Attractor is based on the Jaccard Distance to measure the dynamic distance of any two vertices in the networks. Jaccard Distance is based on common neighbors of two given vertices. But in the study of bipartite networks we find that vertex u and vertex v are different types of vertices and their neighbors are different types of vertices as well. Thus no common neighbor exists in terms of Jaccard Distance between any two different types of vertices in bipartite networks. Further observation shows that neighbors of vertex u $N(u)$ are the same type as second order neighbors of vertex u $NN(u)$. Thus we propose Local Jaccard Distance (LJD) instead of Jaccard Distance in BiAttractor.

Secondly, due to no common neighbor issue in bipartite networks, common neighbor influence on dynamic distance is omitted. Thus another difference between Attractor and BiAttractor is that only directly linked influence and influence from exclusive neighbors are considered in BiAttractor. But Attractor has three patterns containing influence of common neighbors. Furthermore, The equations of directly linked influence and influence from exclusive neighbors come from the original method Attractor. But they are modified to use Local Jaccard Distance (LJD) instead of Jaccard Distance.

For these two reasons shown above, the dynamic distance updating equations of Attractor and BiAttractor are different. It is shown in Eq. (9) to define the updating process of BiAttractor. $DI^t(u, v)$ and $EI^t(u, v)$ depend on Local Jaccard Distance (LJD).

2.2.3. Model

In the previous section, interactive patterns are introduced as the preliminary criteria of BiAttractor. Here, we will introduce BiAttractor in detail.

Given an undirected and unweighted bipartite graph $G = (U, V, E)$, where U and V represent two sets of different types of vertices. There is no edge between vertices from the same set U or V. BiAttractor aims to identify a partition of U and V into k communities $C_1, C_2, \ldots, C_k$ where C_k contains two types of vertices from U and V. The formal description the of algorithm is illustrated in Algorithm 1.

Algorithm 1 BiAttractor

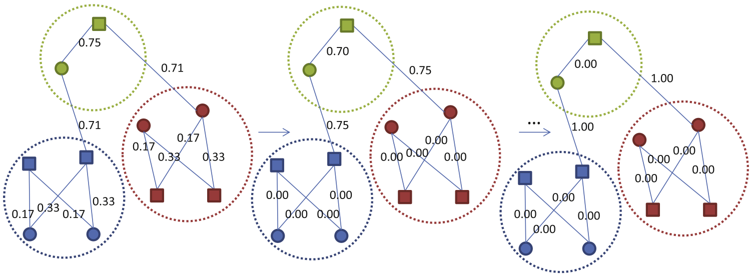
Input: Given an undirected and unweighted bipartite network $G(U, V, E)$ and cohesive parameter λ. λ is iteratively updated between [0, 1] to obtain an optimal community partition with the maximal Q_b.
Output: k two-mode communities $C_1, C_2, \ldots, C_k$.
Initialize each edge $e = \{u, v\}$ from E with an initial distance $d^0(u, v)$ by Eq. (6).
Initialize $d^0(u, x)$ where $x \in EN(u)$ by Eq. (6).
Initialize $d^0(v, y)$ where $y \in EN(v)$ by Eq. (6).
while Not convergence from all edges **do**
  if $0 < d^t(u, v) < 1$ **then**
    $d^{t+1}(u, v) = d^t(u, v) + DI^t(u, v) + EI^t(u, v)$ by Eq. (9)
  end if
  if $d^t(u, v) <= 0$ **then**
    $d^t(u, v) = 0$, d^t is convergent.
  end if
  if $d^t(u, v) >= 1$ **then**
    $d^t(u, v) = 1$, d^t is convergent.
  end if
end while
Generate the communities $C_1, C_2, \ldots, C_k$ by removing all edges with ($d^t(u, v) = 1$).

1. Initially, each edge is associated with an initial distance according to Local Jaccard Distance by Eq. (6).

2. As time goes on, the distance of each edge changes due to the interactions from directly connected vertices and exclusive neighbors according to Eq. (9). Here, the interaction patterns of BiAttractor are different from those of Attractor. The comparisons and reasons were introduced in the previous section of interactive patterns.

3. The dynamic distance d^t changes iteratively if $0 < d^t(u, v) < 1$. Edges within the same communities tend to decrease and edge between different ones tend to increase. In the end, all distances are convergent ($d^t(u, v) = 0$ or $d^t(u, v) = 1$). Edges associated with value zero indicate intra-community edges. Those associated with value one representing inter-community edges have been removed from the underlying networks to form communities naturally.

4. The motivation of introducing λ is to determine whether the influences of exclusive neighbors are positive or negative. If λ is small, exclusive neighbors of u have positive impacts to make u and v move closer. Because exclusive neighbors of u are similar with v, when λ is small. Otherwise exclusive neighbors of u have negative impacts to make vertices u and v move far away. λ is iteratively updated between [0, 1] to run the algorithm (e.g. $\Delta\lambda = 0.05$ is suggested). Due to the optimal λ, the optimal community partition is selected with the maximal modularity Q_b [7].



(a) t=1. (b) t=2. (c) t=5.

Fig. 4. (Color online) A toy sample of BiAttractor. (a) The graph is composed of three sub components by different colors where circles and squares indicate two different types of vertices. The distances attaching to edges are initialized at the first time step. (b) At the second time step, the distances are updated according to BiAttractor. (c) In the end, all intra-community edges are associated with value zero and inter-community edges are labeled with one. Removing edges associated with value one can detect communities automatically.

To better understand the algorithm BiAttractor, the basic idea is briefly presented by a toy sample in Fig. 4. Given a simple bipartite with ten vertices and eleven edges, circles and squares represent two types of vertices separately. Edge distances indicate the relationships between two vertices connected by the edge. Smaller values always denote that they are more possible to be in the same community. At the first time step, each edge is given an initial distance according to local topological structures calculated by Eq. (6) as shown in Fig. 4(a). The distances will be updated by Eq. (9) if they do not reach stable states as shown in Fig. 4(b). After all distances converge to zero or one, the communities can be partitioned from underlying network by removing edges with distance one as shown in Fig. 4(c). It is known that three communities are discovered by removing two inter-community edges.

2.2.4. *Complexity*

To study the time complexity of BiAttractor, each edge of the underlying network has been calculated once to obtain an initial value. As shown in line 3 from Alg. 1 , thus $O(|E|)$ times are considered at the first step. Next, as shown in line 4 and line 5, local topological structures require BiAttractor to obtain influences from exclusive neighbors. The time complexity is at least $O(k|E|)$ where k is the average number of exclusive neighbors. It is shown from line 6 to line 16 that after T time steps, BiAttractor converges to the final state. The complexity is $O(T|E|)$ in the iterative process. Thus the total time complexity of BiAttractor is $O(|E| + k|E| + T|E|)$. T is constant in our experiment given the underlying network G. The time complexity $O(|E| + k|E| + T|E|)$ can be written as $O(z|E|)$ where z is the usual label of network average degree. If the network is dense, the complexity is at least quadratic. Real networks are usually sparse networks and z is constant. The time complexity is linear in this case. The experimental study in the next section will demonstrate its accuracy and efficiency in both synthetic and real networks.

3. Experimental results

3.1. *Settings*

In this section, we aim to evaluate BiAttractor compared with several existing well-known algorithms to detect community in both synthetic and real bipartite networks.

Adaptive BRIM. Barber proposed BRIM (bipartite, recursively induced modules) [7,39] based on the idea of iterative maximization of modularity Q_b in bipartite network. For each iteration, Q_b is guaranteed not to decrease. However, the identified division of bipartite networks leads to a local maximum rather than a global maximum of Q_b. Meanwhile the number of modules are also determined by maximization of modularity Q_b.

LP BRIM. Liu et al. extended the work of BRIM to propose a joint method of label propagation (LP) and BRIM named LP BRIM [43]. Its time complexity is at most $O(n^2)$ (where n is the number of vertices), which is acceptable to be applied in real networks.

AsymIntimacy. Wang et al. defined asymmetric parameters for the intimate degree between the same type of vertices and different type of vertices [21]. Initially, the same type of vertices are merged as subsets due to the asymmetric intimate degree. Then another type of nodes incorporate into those subsets from previous step to form core communities. Each pair of core communities are merged if the ratio of intersection exceeds the threshold. The process continues until no more core communities can be merged. Its time complexity is $O(2n^2 + mn)$ where m is the number of edges and n is the number of vertices.

The implementations of BRIM and LP BRIM come from BiMat[1] which is a MATLAB library whose main function is the analysis of biological bipartite networks [44]. AsymIntimacy is implemented by Java 1.8 using Eclipse Mars 2. BiAttractor is implemented by C++ using Eclipse.

1 http://bimat.github.io/.

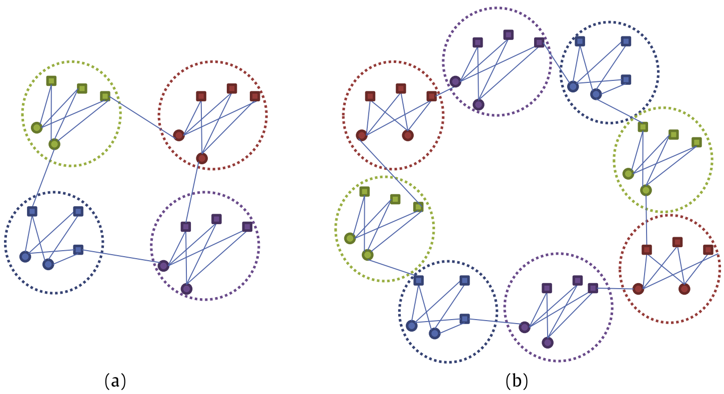


Fig. 5. (Color online) Rings of bicliques. (a) A ring of 4 bicliques, each biclique consists of two circles fully connected with three squares. (b) A ring of 8 bicliques with the same rules used in (a).

Table 1
Basic topological features of rings of bicliques.

Network	n_1	n_2	n	m	$\langle k\rangle$	C	r
4 *bicliques*	12	8	20	28	2.800	0.482	−0.5
8 *bicliques*	24	16	40	56	2.800	0.482	−0.5
16 *bicliques*	48	32	80	112	2.800	0.482	−0.5
64 *bicliques*	192	128	320	448	2.800	0.482	−0.5
128 *bicliques*	384	256	640	896	2.800	0.482	−0.5

Table 2
Performance of Adaptive BRIM, LP BRIM, BiAttractor and AsymIntimacy on rings of bicliques. NMI indicates the accuracy of two-mode community detection. N_C is the number of communities discovered by different methods.

Network	*BiAttractor*		*AdaptiveBRIM*		*LPBRIM*		*AsymIntimacy*	
	(*NMI*)	(N_C)	(*NMI*)	(N_C)	(*NMI*)	(N_C)	(*NMI*)	(N_C)
4 *bicliques*	**1.000**	**4**	**1.000**	4	**1.000**	4	0.714	4
8 *bicliques*	**1.000**	**8**	**1.000**	8	**1.000**	8	0.759	8
16 *bicliques*	**1.000**	**16**	0.934	15	0.802	13	0.785	16
64 *bicliques*	**1.000**	**64**	0.986	63	0.887	56	0.816	64
128 *bicliques*	**1.000**	**128**	0.993	127	0.900	113	0.826	128

Measures. To study and compare different methods, there are two types of measures in general. If the community divisions over underlying networks are given in advance, Normalized Mutual Information (NMI) [45] is applied to give a score from [0, 1]. Otherwise, modularity [22] is employed. The original modularity is defined only for unipartite networks. Thus Barber has extended the definition for community via bipartite networks as Q_b [7]. A higher value from [0, 1] indicates more intra-community edges than those expected by null model. But Q_b also has limitations due to the issue of resolution limit [23].

All the experiments have been carried out on a Mac laptop with Intel i5 1.4 GHz CPU and 4 GB memory.

3.2. Synthetic networks

A number of methods depend on the optimization of modularity. However, such methods might have problem of resolution limit [23]. They have limitations to detect communities smaller than a certain scale depending on the total size of the network and its inner connections. To demonstrate the effectiveness of BiAttractor, a ring of bicliques have been designed with different biclique numbers. It is shown in Fig. 5 that a ring is connected by 4 bicliques from head to tail and another ring of 8 bicliques with the same rule. Each biclique has two mode vertices and two circles are fully connected with three squares. The basic topological structure features have been summarized in Table 1. Further experiments are carried out on rings of different number of bicliques.

It is known from Table 2 that both Adaptive BRIM and LP BRIM can detect perfect communities as BiAttractor when there are 4 bicliques and 8 bicliques. However, when the network is a ring of 16 bicliques, LP BRIM obtains a result ($NMI = 0.802$) with 13 communities. Adaptive BRIM can detect 15 communities with $NMI = 0.934$. Because BiAttractor still can find perfect solution, it obtains 7.07% of improvement of accuracy against Adaptive BRIM and as much as 24.69% of improvement of accuracy over LP BRIM. Further study in a ring of 64 bicliques shows similar result. BiAttractor acquires 1.42% of accuracy improvement over Adaptive BRIM and 12.74% improvement against LP BRIM. Our last experiment was carried out in a ring

Table 3
Basic topological features of real bipartite networks in this experiment.

Network	n_1	n_2	n	m	$\langle k \rangle$	C	r
SW	18	14	32	89	5.563	0.328	−0.337
AR	136	5	141	160	2.270	0.781	−0.743
SCI	108	136	244	358	3.140	0.303	−0.171
CN	829	551	1 380	1 476	2.139	0.427	−0.166
MG	297	806	1 103	2 965	5.376	0.227	−0.300
PCD	680	739	1 419	3 690	1.746	0.407	−0.140
DW	89 356	46 215	135 571	144 342	2.129	0.447	−0.122
DP	48 833	138 839	187 672	207 268	2.209	0.514	−0.138

Table 4
Accuracy comparison of two-mode community detection methods, including Adaptive BRIM, LP BRIM, BiAttractor and AsymIntimacy. Q_b is the modularity score in bipartite networks and N_c is the number of two-mode communities.

Network	*BiAttractor*		*AdaptiveBRIM*		*LPBRIM*		*AsymIntimacy*	
	(Q_b)	(N_C)	(Q_b)	(N_C)	(Q_b)	(N_C)	(Q_b)	(N_C)
SW	**0.345**	4	**0.345**	4	0.313	2	0.333	4
AR	0.601	5	**0.602**	5	0.591	5	0.480	3
SCI	0.660	39	0.660	24	0.648	36	**0.668**	40
CN	**0.859**	132	0.798	104	0.823	203	0.821	142
MG	**0.704**	114	0.687	28	0.592	60	0.604	48
PCD	0.799	129	0.770	113	**0.806**	107	0.784	79
DW	**0.242**	30 947	–	–	–	–	–	–
DP	**0.242**	37 813	–	–	–	–	–	–

Table 5
Time consuming comparison of BiAttractor (t_1), Adaptive BRIM (t_2), LP BRIM (t_3) and AsymIntimacy (t_4) in real bipartite networks.

Network	$t_1(s)$	$t_2(s)$	$t_3(s)$	$t_4(s)$
SW	**0.192**	0.383	0.774	0.805
AR	**0.022**	1.133	1.207	1.001
SCI	**0.473**	1.753	2.459	0.749
CN	**1.009**	4.995	5.685	10.571
MG	**1.482**	1.713	8.606	5.383
PCD	**3.325**	15.092	4.771	72.664
DW	**1812.120**	–	–	–
DP	**2913.600**	–	–	–

of 128 bicliques. BiAttractor still has 0.70% of accuracy improvement over Adaptive BRIM and 11.11% improvement against LP BRIM. Comparing with AsymIntimacy, BiAttractor, Adaptive BRIM and LP BRIM always outperform it in terms of NMI on all the rings of bicliques shown in Table 2. But AsymIntimacy obtains the correct number of communities compared with Adaptive BRIM and LP BRIM. Thus modularity optimization methods such as Adaptive BRIM, LP BRIM and bottom up merging method AsymIntimacy have common issues to accurately detect small bicliques because of resolution limit. But BiAttractor can accurately detect such small bicliques.

3.3. Real networks

Next, several experimental studies will be carried out in real networks without known modular structures and Q_b has been applied to verify the accuracy. Here, real networks in our study include Southern Women Events Participation (SW), America Revolution (AR), Scotland Corporate Interlock (SCI), Crime Network (CN), Malaria and *var* Genes (MG), Protein Complex and Drug network (PCD), DBpedia Writer network (DW) and DBpedia Producer network (DP).

Southern Women events participation (SW). It is a well-known benchmark data set in community discovery from bipartite network. Southern Women network was collected by Davis et al. around Natchez, Mississippi during 1930s for the study of class and race [46]. It describes the partition of 18 women in 14 social events. Women and their social events consist a bipartite network of 32 vertices and 89 edges shown in Table 3. Next, four different community detection methods are compared to study the two-mode community detection problem.

It is shown in Table 5 that BiAttractor is the fastest method compared with others in Southern Women network. Furthermore, Both BiAttractor and Adaptive Brim obtain the optimal two-mode communities.

In this experiment, BiAttractor divides Southern Women network into 4 different size two-mode communities. As shown in Fig. 6(a) that blue community is the largest one with seven women ($W1$, $W2$, $W3$, $W4$, $W5$, $W6$, $W7$) and 6 events ($E1$, $E2$, $E3$, $E4$, $E5$, $E6$). The dark pink community on top of the blue one consists of 4 women ($W8$, $W9$, $W10$, $W16$) and 3 events ($E7$, $E8$, $E9$). The red community on the right side of the dark pink one contains 4 women ($W11$, $W12$, $W13$, $W14$)

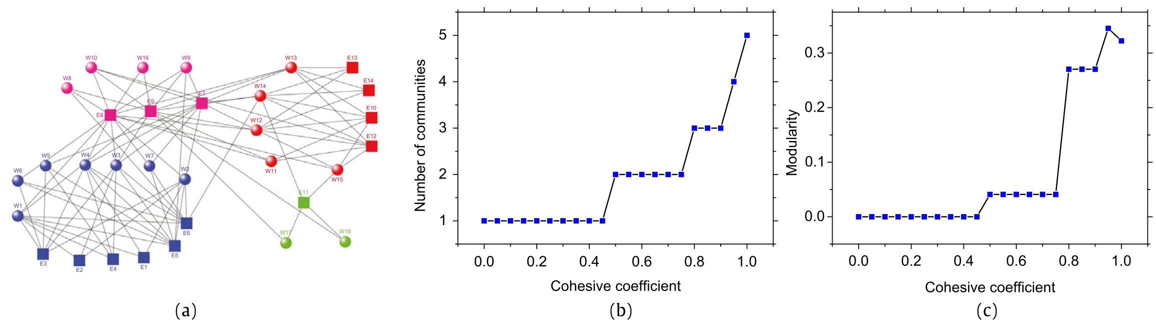


Fig. 6. (Color online) Community division of Southern Women network by BiAttractor. Circles represent 18 women and squares denote 14 events. (a) Color blue, deep pink, red and light green denote 4 communities discovered by BiAttractor ($\lambda = 0.95$) (b) Number of communities changes with respect to cohesive coefficient λ. The optimal community number is 4 with $\lambda = 0.95$. (c) Modularity Q_b changes with respect to cohesive coefficient λ. When $\lambda = 0.95$, $Q_b = 0.345$ is the optimal partition.

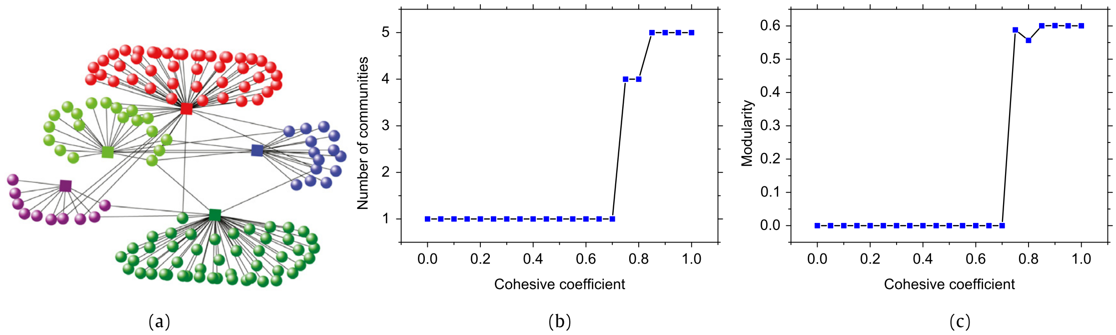


Fig. 7. (Color online) Community partition of American Revolution Network by BiAttractor. Circles represent 136 women and squares denote 5 events. (a) Color blue, dark green, purple, light green and red denote 5 communities discovered by BiAttractor. (b) Number of communities changes with respect to cohesive coefficient λ. The case with $\lambda = 0.85$ leads to the optimal partition of 5 communities. (c) Modularity Q_b changes with cohesive coefficient λ. We obtain the optimal division when $\lambda = 0.85$ and $Q_b = 0.601$.

and 4 events ($E10$, $E12$, $E13$, $E14$). The smallest one is colored light green, which consists of 2 women ($W17$, $W18$) and 1 event ($E11$).

In BiAttractor, λ is the unique user defined parameter. It is iteratively updated between [0, 1] to obtain the optimal partition of communities via bipartite networks. It is shown in Fig. 6(c) that at the first time step, $\lambda = 0$ and $Q_b = 0$. As time goes on, λ is iteratively updated with Δ (e.g. $\Delta = 0.05$ is suggested here). When $\lambda = 0.95$, we obtain the optimal community partition with $Q_b = 0.345$. It is shown in Fig. 6(b) that when $\lambda = 0.95$, Southern Women (SW) network is divided into 4 communities. They are visualized in Fig. 6(a).

American Revolution (AR). This data set contains membership information of 136 people from 5 organizations dating back to the time before the American Revolution [47].[2] The list includes well-known people such as the American activist Paul Revere. The relations between well-known people and their organizations can be described in a bipartite network. An edge between a person and an organization shows that the person is a member of this organization. Firstly, the basic topological structure features will be described in Table 3. There are 141 vertices and 160 edges in this bipartite network. Then four methods will be compared to discover community structures in this bipartite network.

It is shown in Table 5 that BiAttractor is the fastest method compared with others in American Revolution network. Furthermore, it is shown in Table 4 that Adaptive Brim obtains the best modularity score ($Q_b = 0.602$), which is slightly better than the modularity score of BiAttractor ($Q_b = 0.601$).

The result of community division by BiAttractor is shown in Fig. 7. There are 5 different communities represented in different colors. Such 5 communities have the same pattern that each community represents the specific organization and their members. In Fig. 7(a), on one hand we know that each organization locating in the center is surrounded by its own members. On the other hand, a small number of people belong to different organizations, which are represented by edges bridging on different communities. BiAttractor obtains two-mode community partition in an iterative way to seek for the optimal cohesive coefficient λ. In this experiment, $\lambda = 0.85$ denotes the optimal partition with 5 communities shown in Fig. 7(b) and the modularity score ($Q_b = 0.601$) is shown Fig. 7(c) as well.

2 http://konect.uni-koblenz.de/networks/brunson_revolution.

Scotland Corporate Interlock (SCI). As the third example, we consider a data set on corporate interlocks in Scotland in the early twentieth century [48]. This data set includes board members of Scottish companies, who held multiple directorships in the period of 1904–1905. Directorships maintain a bipartite network of 136 individuals and 108 firms.

As shown in Tables 4 and 5, BiAttractor is faster than others. AsymIntimacy obtains the best division of communities with ($Q_b = 0.668$). But the gap between BiAttractor and AsymIntimacy ($Q_b = 0.660$) is very small.

Crime Network (CN). This data set includes persons who were recorded in at least one crime case as either a victim, a witness or as a suspect [47].[3] The relationships between crime related persons and crime cases form a bipartite network of 1476 edges connecting 829 persons and 551 crime cases. As shown in Tables 4 and 5, BiAttractor is the fastest compared with others. It also outperforms others in terms of accuracy ($Q_b = 0.859$).

Malaria and *var* Genes (MG). Malaria parasite evades the human immune system via a protein camouflages encoded in *var* genes [28,49]. *Var* genes frequently recombine, which amounts to the constrained genetic substring and create novel camouflages [28]. Thus the relationships between *var* genes and their genetic substrings form a bipartite network with natural community structures. It is shown in Table 3 that MG network consists of 2965 edges connecting 297 genes and 806 substrings.

Further study on community structures by four methods can uncover modular structures of *var* genes and their substrings. As shown in Tables 4 and 5, BiAttractor obtains the best partition of communities with ($Q_b = 0.704$). BiAttractor is also the fastest method compared with others.

The Protein Complex and Drug Network (PCD). Recent studies have revealed important connections between protein complexes and key diseases. Nacher and Schwartz studied a bipartite network containing 680 drugs and 739 protein complexes to uncover the interconnectivity in molecular and human disease related systems [50]. Firstly the basic topological features are described to gain insights into PCD. It is shown in Table 4 that PCD consists of 1419 vertices and 3690 edges. From Table 4, it is known that both BiAttractor ($Q_b = 0.799$) and LP BRIM ($Q_b = 0.806$) outperform other methods. But BiAttractor is faster than other three methods.

DBpedia Writer network (DW). DBpedia Writer Network consists of 46 215 writers and their 89 356 works from DBpedia. The 144 342 edges represent collaborations between writers to produce works.[4] Differing from previous 6 networks, DW consists of hundreds of thousands of vertices and edges, which poses challenges for existing methods to discover community structures.

Further study on community structures has been carried out by Adaptive BRIM and LP BRIM firstly. Due to the memory consumption, we work on a powerful workstation with Intel Core i7 3.40 GHz CPU and 32 GB memory. Matlab program of Adaptive BRIM returned an exception of out of memory after a few hours. The same problem appeared because Matlab program LP BRIM exited abnormally. We applied AsymIntimacy to DW and obtained no result after a few hours. Then we redid the same experiment on a laptop with 4 GB memory using BiAttractor. As shown in Table 4, it obtains the result with $Q_b = 0.242$ and finished in 30.202 min. The consuming time actually consists of two parts, the first part of community detection consumes only 4.035 s. But the calculation of modularity costs most of the rest time of 30.201 min. It concludes that BiAttractor can detect the community structures in a short time. However, Adaptive BRIM, LP BRIM and AsymIntimacy fail due to the expensive consumption of memory and time.

DBpedia Producer network (DP). DBpedia Producer Network also comes from DB Pedia. This bipartite network consists of 48 833 producers and their 138 839 works.[5] Next, four methods are employed to discover communities from this large bipartite network.

BiAttractor consumes only 20.727 s to obtain the community structures but it also takes 48.56 min to calculate the modularity score ($Q_b = 0.242$) shown in Table 4. Such an experiment is carried out on a laptop of 4 GB memory. But both Adaptive BRIM and LP BRIM fail due to the out of memory problem on a workstation of Intel Core i7 3.40 GHz CPU and 32 GB memory. AsymIntimacy obtains no result after several hours due to the non-linear time complexity. It concludes that BiAttractor can deal with very large bipartite networks effectively.

4. Conclusions

We have suggested a new algorithm, BiAttractor, to detect two-mode communities in bipartite networks. It is based on the distance dynamics method Attractor by Shao et al. with extension from unipartite to bipartite networks. Distance dynamics among vertices is inspired by interactions in human society. At the very beginning, each edge is given an initial value according to novel Local Jaccard Distance (LJD). Jaccard Distance from Attractor cannot be adopted in bipartite networks because no common neighbor between two different types of vertices. Then each vertex interacts with its directly linked neighbors and exclusive neighbors in a local way. The distance of any pair of connected vertices is simulated according to influence of neighbors. As a result of local topological structure influences, the distances among vertices in the same communities tend to decrease, while those in different communities tend to increase eventually. Finally, all dynamical distances will converge to stable states. Removing intercommunity edges with maximal distances can naturally discover two-mode communities in bipartite networks.

---

3 http://konect.uni-koblenz.de/networks/moreno_crime.

4 http://konect.uni-koblenz.de/networks/dbpediawriter.

5 http://konect.uni-koblenz.de/networks/dbpediaproducer.

Further experimental study shows that it has three advantages compared with existing other methods.

Firstly, methods using global optimization usually have the problem of the resolution limit [23]. Communities smaller than a certain scale cannot always be detected accurately. The main reasons of this problem come from the global view towards the whole network and limitations of the measure modularity for bipartite networks. This new model BiAttractor takes advantages of local topological structures rather than a global view, which can detect small communities with high accuracy. Experimental study on rings of bicliques demonstrates that it outperforms other three methods in terms of NMI.

Secondly, experiment study is carried out in 6 real bipartite networks with at most thousands of vertices and edges. BiAttractor can obtain the optimal community partitions on 3 of them, including SW (Southern Women) network, CN (Crime Network) and MG (Malaria and var Genes) network. The results of the other 3 networks show that the gaps between BiAttractor and existing best methods are less than 1%. Furthermore, BiAttractor is the fastest method compared with others. It is at least 1.15 times and at most 54.86 times faster than other methods.

Thirdly, due to expensive time and memory consumption of other three methods, they don not obtain reasonable results in large networks with hundreds of thousands of vertices and edges using powerful workstation with 32 GB memory. While this new model can obtain meaningful results in a few minutes on a laptop with 4 GB memory. It concludes that BiAttractor can detect two-mode communities of large bipartite networks efficiently.

It concludes that our method BiAttractor can accurately detect two-mode communities of bipartite networks in a short time. It is faster than others because each vertex interacts with local neighboring vertices. Local interaction instead of global optimization usually leads to high efficiency. Parameter λ is iteratively selected to obtain an optimal partition of community with maximal modularity. Parameter free methods with local vertices information are simple but with poor performance. Multiple parameters methods considering global topological structures are complex and lead to high performance. Thus striking a balance between local topology and global topology, between parameter free and acceptable parameters is suggested to future development of methods. Furthermore, this paper merely explores the method in bipartite networks. In real practice there are various network types. Thus it is meaningful and valuable to explore community detection in temporal networks, heterogeneous networks with vertex profiles, multi-level networks in the near future. Discovery and visualization of such communities are also the future directions of our efforts.

Acknowledgments

This work is jointly supported by the UK Engineering and Physical Sciences Research Council (Grant Number EP/L015463/1), the National Natural Science Foundation of China (Grant Nos. 61433014 and 61673085), and the Fundamental Research for the Central Universities (Grant Nos. ZYGX2014Z002 and ZYGX2015J152).

We give special thanks to Barber for sharing BRIM implemented by Octave. We also acknowledge the support received from Big Data Research Center of University of Electronic Science and Technology of China (UESTC), Ningbo Education Bureau, Ningbo Science and Technology Bureau and China MOST.

References

[1] M.E.J. Newman, Networks: An Introduction, Oxford University Press, Oxford, 2010, p. 18.
[2] X.F. Wang, X. Li, G.R. Chen, Network Science: An Introduction, Higher Education Press, Beijing, 2012, pp. 10–15.
[3] M.E.J. Newman, The structure of scientific collaboration networks, Proc. Natl. Acad. Sci. USA 98 (2) (2001) 404–409.
[4] S. Fortunato, Community detection in graphs, Phys. Rep. 486 (3) (2010) 75–174.
[5] S. Fortunato, D. Hric, Community detection in networks: A user guide, Phys. Rep. 659 (2016) 1–44.
[6] K.I. Goh, M.E. Cusick, D. Valle, B. Childs, M. Vidal, A.L. Barabasi, The human disease network, Proc. Natl. Acad. Sci. USA 104 (21) (2007) 8685–8690.
[7] M.J. Barber, Modularity and community detection in bipartite networks, Phys. Rev. E 76 (6) (2007) 066102.
[8] G. Roger, S.P. Marta, A.N.A. Luís, Module identification in bipartite and directed networks, Phys. Rev. E 76 (3) (2007) 036102.
[9] L.Y. Tang, S.N. Li, J.H. Lin, Q. Guo, J.G. Liu, Community structure detection based on the neighbor node degree information, Internat. J. Modern Phys. C 27 (04) (2016) 1650046.
[10] H. Peng, D.D. Zhao, L. Li, J.F. Lu, J.M. Han, S.Y. Wu, An improved label propagation algorithm using average node energy in complex networks, Physica A 460 (2016) 98–104.
[11] W. Li, C. Huang, M. Wang, X. Chen, Stepping community detection algorithm based on label propagation and similarity, Physica A 472 (2017) 145–155.
[12] J.X. Yang, X.D. Zhang, Finding overlapping communities using seed set, Physica A 467 (2017) 96–106.
[13] S. Bilal, M. Abdelouahab, Evolutionary algorithm and modularity for detecting communities in networks, Physica A 473 (2017) 89–96.
[14] W. Chen, Z.M. Liu, X.R. Sun, Y.J. Wang, A game-theoretic framework to identify overlapping communities in social networks, Data Min. Knowl. Discov. 21 (2) (2010) 224–240.
[15] H.L. Sun, E. Ch'ng, X. Yong, J.M. Garibaldi, S. See, D.B. Chen, An improved game-theoretic approach to uncover overlapping communities, Internat. J. Modern Phys. C 28 (9) (2017) 1750112.
[16] G. Palla, I. Derényi, I. Farkas, T. Vicsek, Uncovering the overlapping community structure of complex networks in nature and society, Nature 435 (7043) (2005) 814–818.
[17] T.S. Evans, R. Lambiotte, Line graphs, link partitions, and overlapping communities, Phys. Rev. E 80 (1) (2009) 016105.
[18] V.D. Blondel, J.-L. Guillaume, R. Lambiotte, E. Lefebvre, Fast unfolding of communities in large networks, J. Stat. Mech. Theory Exp. 2008 (10) (2008) P10008.
[19] T. Zhou, J. Ren, M. Medo, Y.C. Zhang, Bipartite network projection and personal recommendation, Phys. Rev. E 76 (4) (2013) 046115.
[20] A. Mukherjee, M. Choudhury, N. Ganguly, Understanding how both the partitions of a bipartite network affect its one-mode projection, Physica A 390 (20) (2011) 3602–3607.
[21] X. Wang, X. Qin, Asymmetric intimacy and algorithm for detecting communities in bipartite networks, Physica A 462 (2016) 569–578.
[22] M.E.J. Newman, Modularity and community structure in networks, Proc. Natl. Acad. Sci. USA 103 (23) (2006) 8577–8582.

[23] S. Fortunato, M. Barthelemy, Resolution limit in community detection, Proc. Natl. Acad. Sci. USA 104 (1) (2007) 36–41.
[24] S. Lehmann, M. Schwartz, L.K. Hansen, Biclique communities, Phys. Rev. E 78 (1) (2008) 016108.
[25] M.E.J. Newman, Finding community structure in networks using the eigenvectors of matrices, Phys. Rev. E 74 (3) (2006) 036104.
[26] P. Zhang, J.L. Wang, X.J. Li, M.H. Li, Z.R. Di, Y. Fan, Clustering coefficient and community structure of bipartite networks, Physica A 387 (27) (2008) 6869–6875.
[27] Y. Cui, X.Y. Wang, Uncovering overlapping community structures by the key bi-community and intimate degree in bipartite networks, Physica A 407 (2014) 7–14.
[28] D.B. Larremore, A. Clauset, A.Z. Jacobs, Efficiently inferring community structure in bipartite networks, Phys. Rev. E 90 (1) (2014) 012805.
[29] Y.C. Xu, L. Chen, B. Li, W. liu, Density-based modularity for evaluating community structure in bipartite networks, Inform. Sci. 317 (2015) 278–294.
[30] J.G. Liu, L. Hou, X. Pan, Q. Guo, T. Zhou, Stability of similarity measurements for bipartite networks, Sci. Rep. 6 (2016) 18653.
[31] K. Li, Y. Pang, A unified community detection algorithm in complex network, Neurocomputing 130 (2014) 36–43.
[32] X. Han, S. Cao, Z. Shen, B. Zhang, W.X. Wang, R. Cressman, H.E. Stanley, Emergence of communities and diversity in social networks, Proc. Natl. Acad. Sci. USA 114 (11) (2017) 2887–2891.
[33] D.B. Chen, Y. Fu, M.S. Shang, A fast and efficient heuristic algorithm for detecting community structures in complex networks, Physica A 388 (13) (2009) 2741–2749.
[34] J.L. He, D.B. Chen, A fast algorithm for community detection in temporal network, Physica A 429 (2015) 87–94.
[35] J.L. He, D.B. Chen, C.J. Sun, Y. Fu, W.J. Li, Efficient stepwise detection of communities in temporal networks, Physica A 469 (2017) 438–446.
[36] D.B. Chen, M.S. Shang, Z.H. Lv, Y. Fu, Detecting overlapping communities of weighted networks via a local algorithm, Physica A 389 (19) (2010) 4177–4187.
[37] R. Shang, W.T. Zhang, L.C. Jiao, R. Stolkin, Y. Xue, A community integration strategy based on an improved modularity density increment for large-scale networks, Physica A 469 (2017) 471–485.
[38] Y. Pan, D.H. Li, J.G. Liu, J.Z. Liang, Detecting community structure in complex networks via node similarity, Physica A 389 (14) (2010) 2849–2857.
[39] M.J. Barber, M. Faria, L. Streit, Strogan, Searching for communities in bipartite networks, Stoch. Quantum Dyn. Biomol. Syst. 1021 (2008) 171–182.
[40] J.M. Shao, Z.C. Han, Q.L. Yang, T. Zhou, Community detection based on distance dynamics, in: Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Sydney, Australia, August 10-13, 2015, pp. 1075–1084.
[41] T. Zhou, L.Y. Lü, Y.C. Zhang, Predicting missing links via local information, Eur. Phys. J. B 71 (4) (2009) 623–630.
[42] L.Y. Lü, C.H. Jin, T. Zhou, Similarity index based on local paths for link prediction of complex networks, Phys. Rev. E 80 (4) (2009) 046122.
[43] X. Liu, T. Murata, Community detection in large-scale bipartite networks, Trans. J. Japan Soc. Artif. Intell. 25 (1) (2010) 16–24.
[44] C.O. Flores, T. Poisot, S. Valverde, J.S. Weitz, G. McInerny, BiMat: aMATLABpackage to facilitate the analysis of bipartite networks, Methods Ecol. Evol. 7 (1) (2016) 127–132.
[45] P. Zhang, Evaluating accuracy of community detection using the relative normalized mutual information, J. Stat. Mech. Theory Exp. 11 (2015) P11006.
[46] A. Davis, B.B. Gardner, M.R. Gardner, Deep south: A social anthropological study of caste and class, in: A Study in the Sociology of Science, The University of South Carolina Press, Columbia, South Carolina, USA, 1941.
[47] J. Kunegis, Konect: the koblenz network collection, in: Proceedings of the 22nd International Conference on World Wide Web, Rio de Janeiro, Brazil, May 13-17, 2013, pp. 1343–1350.
[48] J. Scott, M. Hughes, The Anatomy of Scottish Capital: Scottish Companies and Scottish Capital, McGill-Queen's Press-MQUP, Montreal, 1981, p. 291.
[49] D.B. Larremore, A. Clauset, C.O. Buckee, A network approach to analyzing highly recombinant malaria parasite genes, PLoS Comput. Biol. 9 (10) (2013) e1003268.
[50] J.C. Nacher, J.M. Schwartz, Modularity in protein complex and drug interactions reveals new polypharmacological properties, PLoS One 7 (1) (2012) e30028.