Journal of Intelligent & Fuzzy Systems 37 (2019) 7965–7976
DOI:10.3233/JIFS-190320
IOS Press

Overlapping Community Detection in Bipartite Networks using a Micro-bipartite Network Model: Bi-EgoNet

Furong Chang[a,b], Bofeng Zhang[a,*], Yue Zhao[a], Songxian Wu[c], Guobing Zou[a] and Sen Niu[d]
[a] *School of Computer Engineering and Science, Shanghai University, Shanghai, China*
[b] *School of Computer Science and Technology, Kashi University, Xinjiang, China*
[c] *School of Mathematics and Statistics, Kashi University, Xinjiang, China*
[d] *School of Computer and Information Engineering, Shanghai Polytechnic University, Shanghai, China*

Abstract. A bipartite network is a special kind of complex network that consists of two different types of nodes with edges existing only between the different node types. There are numerous real-world examples of bipartite networks, such as scientific collaboration networks and film-actor networks, among many others. Detecting the community structure of bipartite networks not only contributes to a deeper understanding of their hidden structure, but also lays the foundation for research into the personalized recommendation technology. Most existing algorithms, however, only focus on the detection of non-overlapping community structures while ignoring overlapping community structures. In this study, we developed a micro-bipartite network model, Bi-EgoNet along with an algorithm called Overlapping Community Detection using Bi-EgoNet (OCDBEN). This algorithm first extracts the sub-bi-community set from each Bi-EgoNet using similarity within the bipartite network and then constructs a global community structure by merging the sub-bi-communities using the double-merger strategy. We evaluated the OCDBEN algorithm with several synthetic and real-world bipartite networks and compared it with existing state-of-the-art algorithms. The experimental results demonstrated that OCDBEN outperformed existing algorithms in both accuracy and effectiveness.

Keywords: Overlapping community, bipartite networks, complex network

1. Introduction

Community detection in complex networks have received significant research interest in recent years. Many methods have been proposed for one-mode networks [11, 18, 20, 21, 42]. Community detection uses topological features of networks, such as Chang's method's use of Friend Intimacy [11] and Li's method's use of the maximizing likelihood function [21]. In addition to the topological features, researchers also consider other influence factors, such as social behavior and semantics [18] and background information [20]. Based on these community detecting methods, other researchers have proposed personalized recommender systems, including the Community-based Collaborative Filtering Recommender System (CCFRS) [32] and the Community-based Hashtag Recommender System (CHRS) [33].

Recently, research into bipartite networks has received attention. Bipartite networks are found in various fields including a scientific collaboration network of papers and authors [24], a movie-actor network of movies and actors [2], and the metabolic network [14] of reactions and metabolites. A bipartite

*Corresponding author. Bofeng Zhang, School of Computer Engineering and Science, Shanghai University, Shanghai, China. E-mail: bfzhang@shu.edu.cn.

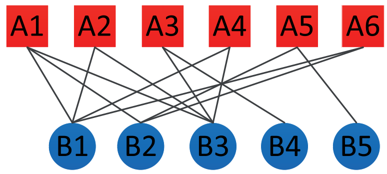


Fig. 1. A bipartite network example of South African companies. Red nodes ($A1, A2, ..., A6$) represent persons, and blue nodes ($B1, B2, ..., B5$) represent companies. Each of the 13 edges between persons and companies indicates that the person has a leadership position in the company.

network (also known as a two-mode or affiliate network) is an essential kind of complex network having two types of nodes and edges only between the different types of nodes. Fig. 1 shows a bipartite network example of South African companies [35] with shared leadership relations between persons and companies. The node types are indicated by different colors.

The community structure of bipartite networks features significant characteristics with tightly connected nodes within communities (intra-communities) and sparely connected between communities (inter-communities). Detecting community structure is helpful to excavate the hidden structured information in the network. In most real networks, an entity (nodes) has multiple attributes that means the entity can belong to more than one communities at the same time. However, in the traditional community detection algorithm, there are few researches on bipartite network overlapping community structure. So detecting overlapping communities in bipartite networks has more important significance and application value.

Recently, a large number of algorithms have been proposed to detect overlapping community structure in bipartite networks [8, 25, 26, 15]. These algorithms fall into two main categories: One method projects bipartite networks onto one-mode networks. Newman et al. [8, 40], however, have proven that the method of one-mode projection does not fully reflect the relationship strength of these nodes. One-mode projection results in a loss of information from the original bipartite network and the addition of information that does not belong to the original bipartite network. The second method addresses bipartite networks directly. This method simplifies the capture of essential network features compared to one-mode projection.

EgoNet [13] (also known as Egocentric Network) is a micro one-mode network model consisting of a single node and its neighborhood. The single node is

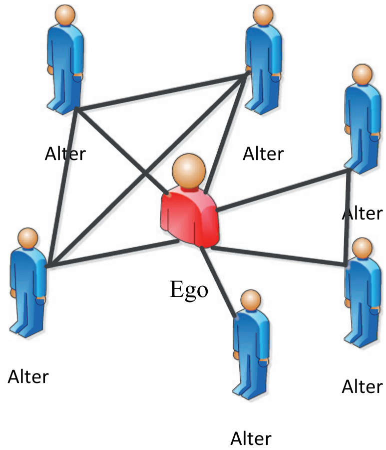


Fig. 2. A sample EgoNet. The red person is called the Ego, and the blue persons are called Alters. the Ego, Alters, and relationships among these persons form an EgoNet.

called the Ego. Its neighbors are called Alters. Edges among of these nodes are called relationships. Fig. 2 depicts a typical EgoNet. Any node in a one-mode network can be the kernel of an EgoNet. Any one-mode network is composed of more than one EgoNet. EgoNet effectively enables detection of overlapping communities in one-mode networks [1, 5]. Inspired by EgoNet, we have designed a micro-bipartite network model named Bi-EgoNet. Within our method, we introduce a new OCDBEN (Overlapping Community Detection using Bi-EgoNet) algorithm to address bipartite networks directly from the micro-perspective. Our main contributions in this work are as follows:

– **Bi-EgoNet, a micro-bipartite network model.** We extend the micro-one-mode model, EgoNet, to a micro-bipartite network model. We believe Bi-EgoNet is the first to analyze bipartite networks from a micro-perspective.
– **A new OCDBEN algorithm.** OCDBEN is an overlapping community detection algorithm based on Bi-EgoNet. Compared to state-of-the-art traditional algorithms, our experiments show OCDBEN to be more accurate and effective.
– **A new measurement method of modularity,** EQ_b EQ_b extends the Q_b measure and evaluates the modularity of overlapping community structures.

In Section 2, we discuss various algorithms used for community detection in bipartite networks. In Section 3, we describe our proposed algorithm, OCDBEN, in detail. In Section 4. we introduce three evaluation methods: EQ_b, bipartite partition density (Density), and normalized mutual information (NMI). In Section 5, we introduce two overlapping community detection methods: Cui's method and Wang's method. In Section 6, we present our experimental results and evaluation of the OCDBEN algorithm with several synthetic and real networks, and compare with the methods of Cui [41] and Wang [40]. Finally, we present our conclusions in Section 7.

2. Related work

As mentioned previously, community detection algorithms in bipartite networks fall into two categories: those that project bipartite networks into two one-mode networks and those that address bipartite networks directly.

Projecting a bipartite network into two one-mode networks using either weighted or unweighted projection is a widely used method. S.Dhillon [9] presented a new spectral co-clustering algorithm to solve a bipartite graph partitioning problem between documents and words. It used the second left and right singular vectors of an appropriate scaled word-document matrix to yield good bipartitions. Zhou et al. [36] were inspired by network-based resource-allocation dynamics and developed a weighted projection method for bipartite networks. A personal recommendation method was then proposed based on this method. Meghanathan et al. [29] proposed a community detection method based on spectral decomposition. This method projected nodes of different dimensions onto specific dimensions. It then used this projection to obtain the smallest eigenvalue and its corresponding eigenvector to detect communities in direct and indirect bipartite networks.

The second category addresses bipartite networks directly. Barber [25] proposed an algorithm called BRIM (bipartite, recursively induced modules). This technique extended the modularity measure proposed by Newman [23] to a bipartite modularity (Q_b). It then used Q_b to identify some key properties of modularity matrix B and to detect community structures. Liu et al. [39] combined the LP (Label Propagation) and BRIM methods in a fast algorithm called LP&BRIM. This algorithm generated better community structures by recursively inducing division between the two types of nodes in bipartite networks. Larremore et al. [19] presented a bipartite stochastic block model (biSBM) to solve the community detection problem. This model explicitly included vertex type information was easily extended to k-partite networks. Others have proposed a number of algorithms designed to detect the overlapping community structure of bipartite networks. Classic methods such as that proposed by Cui et al. [41] introduced the key bi-community and free node concepts and proposed a novel algorithm for the detection of overlapping community structure in bipartite networks. Using this method, Wang et al. [40] defined the concept of intimate degree, and used it to obtain core communities with overlapping communities which were revealed via the merging rule. Li et al. [43] proposed a new quantitative function called bipartite partition density. Bipartite networks can be partitioned into reasonable overlapping communities by maximizing this quantitative function. Meanwhile, they also developed a heuristic adapted label propagation algorithm (BiLPA) in order to optimize the bipartite partition density in large-scale bipartite networks.

Table 1 shows the conceptual differences among these methods relative to our algorithm OCDBEN.

These methods described thus far all detect community structure from a macro perspective. Detecting community structure from the macro perspective, however, invariably misses local features between nodes of bipartite networks. To address this problem, we propose our new algorithm, OCDBEN, that detects community structure in bipartite networks from a micro perspective. OCDBEN extracts sub-bi-community structure from a micro perspective and merges sub-bi-communities from a macro perspective. This algorithm not only takes into consideration local features between network nodes and extracts sub-bi-communities using local features, but it also incorporates the advantages associated with a macro perspective.

3. Overlapping community detection using Bi-EgoNet (OCDBEN)

We first consider a bipartite network *BG(A, B, E)* without self loops and multiple edges between any given pair of nodes. A and B represent two types of node sets, m is the number of A-type nodes, and n is the number of B-type nodes. we also define the set of edges $E = \{e_{a,b} | a \in A, b \in B\}$.

3.1. Related definitions

Definition 1. Given a bipartite network *BG(A, B, E)*, *Nei(a)* represents the neighboring node set of A-type

Table 1
A comparison of bipartite community detection algorithm

Feature	Method of processing	Object function	A priori number of communities	Analysis view
SiDhillon's method [9]	one-mode Projection	Eigenvector	Yes	Macro
Meghannthann's method [29]	one-mode Projection	Eigenvector and Eigenvalue	Yes	Macro
BRIM [23]	Addressing directly	Q_b	No	Macro
LP&BRIM [39]	Addressing directly	Maximum Likelihood	Yes	Macro
BiSBM [25]	Addressing directly	Q_b	No	Macro
Cui's method [41]	Addressing directly	Q_b	No	Macro
Wang's method [40]	Addressing directly	Q_b	No	Macro
Li's method [43]	Addressing directly	Density	No	Macro
OCDBEN	Addressing directly	EQ_b	No	Micro

node a. Likewise, *Nei(b)* represents the neighboring node set of any B-type node b. We express these as:

$$Nei(a) = \{b | b \in B, e_{a,b} \in E\}, \tag{1}$$

$$Nei(b) = \{a | a \in A, e_{a,b} \in E\}. \tag{2}$$

Definition 2. Given a bipartite network *BG(A, B, E)*, *(a, b)* represents a node pair linked by edge $e_{a,b}$. *NP(a)* represents the node pair set of any A-type a. Similarly, *NP(b)* represents the node pair set of any B-type b. We express these as

$$NP(a) = \{(a, b) | b \in B, e_{a,b} \in E\}, \tag{3}$$

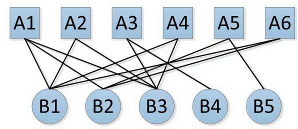
$$NP(b) = \{(a, b) | a \in A, e_{a,b} \in E\}. \tag{4}$$

Definition 3. Given a one-mode network *G(V, E)*, $EN_v (\forall v \in V)$ represents a micro one-mode network model EgoNet of v. This model consists of v, *Nei(v)* and the edges between these nodes. v is referred to as the Ego, and each node in *Nei(v)* is referred to as an Alter.

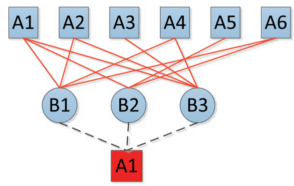
Inspired by the definition of EgoNet, we refer to the Ego of our Bi-EgoNet as the bi-Ego. Each Alter of a Bi-EgoNet is referred to as a bi-Alter. We define a Bi-EgoNet as follows.

Definition 4. Given a bipartite network *BG(A, B, E)*, *bi-EN(a, b)* is a micro-bipartite network model Bi-EgoNet of node pair *(a, b)*. This model consists of node pair *(a, b)*, the neighbor node pairs *NeiNP(a, b)* of *(a, b)* and the edges between these nodes. The node pair *(a, b)* is known as the bi-Ego. Each node pair in *NeiNP(a, b)* is known as a bi-Alter. Fig. 3 depicts a sample of a Bi-EgoNet. We express this mathematically as

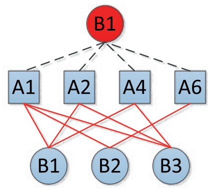
$$\begin{aligned} NeiNP(a, b) = \{&NP(y) \cup NP(x) - (a, b) | \\ &y \in Nei(a), x \in Nei(b), e_{x,y} \in E\} \end{aligned} \tag{5}$$



(a) South African Companies



(b) NP(B_y) ($B_y \in$ Nei(A1))



(c) NP(A_x) ($A_x \in$ Nei(B1))

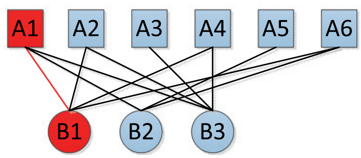


Fig. 3. A sample Bi-EgoNet. (a) The bipartite network of South African Companies; (b) The neighboring node pair set of *A1*; (c) The neighboring node pair set of *B1*. (d) The Bi-EgoNet composed of (A1, B1), (a), and (b).

Definition 5. Given a Bi-EgoNet *bi-EN(a, b)*, *subBC* represents a sub-bi-community of *bi-EN(a, b)*. The connection strength between bi-Ego and bi-Alters in $subBC_{a,b}$ is greater than that between bi-Ego and other the bi-Alters in *bi-EN(a, b)* and is represented by *subBC(LA(a), LB(b), LBe)*.

Definition 6. Given a bipartite network *BG(A, B, Be)*, *GBC* represents the global bi-community structure that has higher density intra-communities and lower density inter-communities.

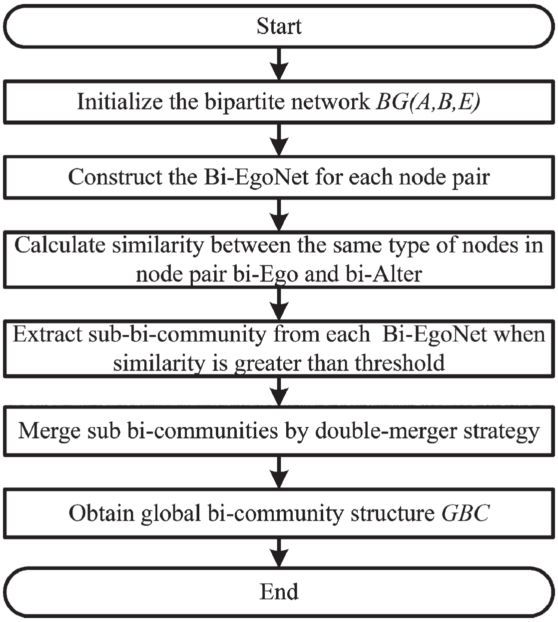


Fig. 4. OCDBEN flowchart.

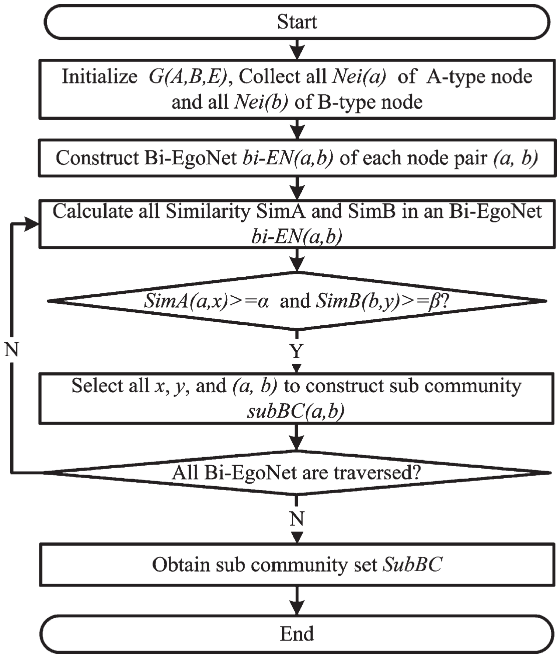


Fig. 5. Flow chart of the first stage of OCDBEN.

3.2. OCDBEN

In this section, we describe in detail the two-stage execution process of our OCDBEN algorithm. The first stage is the extraction of sub-bi-communities set *subBC* (Definition 5), using the similarity of each Bi-EgoNet (Definition 4) of a bipartite network. The second stage is the merger of the *subBCs* using a double-merger strategy to GBC (Definition 6). The OCDBEN flowchart is shown in Fig. 4.

3.2.1. Extracting sub-bi-communities using similarity from each Bi-EgoNet

In this stage, the similarity calculations between a bi-Ego and each bi-Alter are a very important part of extracting sub-bi-community from Bi-EgoNet (Definition 5).

For a Bi-EgoNet *bi-EN(a, b)*, similarity refers to the level of similarity for the same type of nodes between the bi-Ego and each bi-Alter. The definition of similarity is shown in formulas (6) and (7).

$$SimA(a, x) = \frac{|\ Nei(a) \cap Nei(x)\ |}{\sqrt{|\ Nei(a)\ | * |\ Nei(x)\ |}}, \quad (6)$$

$$SimB(b, y) = \frac{|\ Nei(b) \cap Nei(y)\ |}{\sqrt{|\ Nei(b)\ | * |\ Nei(y)\ |}}, \quad (7)$$

where x represents A-type node of bi-Alter, y represents B-type node of bi-Alter. *SimA(a, x)* represents the similarity level between a and x, Analogously, *SimB(b, y)* represents the similarity level between b and y.

Algorithm 1 Extracting sub-bi-communities using similarity from each Bi-EgoNet

---

Input: Bipartite network: *BG(A, B, E)*;
Output: Sub-bi-community set:*subBC*;
1: Traverse the bipartite network *BG*, collecting node pair set *NP* (Equations (3) and (4)) and neighbor node set *Nei* (Equations (1) and (2));
2: Construct the Bi-EgoNet set *bi-EN(a,b)* for each node pair *(a,b)*;
3: Calculate all similarity sets *SimA* (Equation (6)) and *SimB* (Equation (7)) in a Bi-EgoNet;
4: Extract sub-bi-community $subBC_{a,b}$ from the Bi-EgoNet when $SimA > \alpha$ and $SimB > \beta$;
5: Repeat steps 3 and 4 until all Bi-EgoNets are traversed;
6: **return** *subBC*;

The pseudo-code of the stage is shown in Algorithm 1. And Fig. 5 shows the flow chart for the first stage.

3.2.2. Merging sub-bi-communities using the double-merger strategy

In this stage, we apply a double-merger strategy to merge the sub-bi-communities. The double-merger strategy contains two different approaches for merging the sub-bi-communities. The first strategy is applied that only when both the similarity of the two same type nodes between the two sub-bi-communities are greater than a threshold. In this case, we consider the two sub-bi-communities to belonging to the same community and able to be merged. The second strategy is applied when there exists a large difference in the number between the two types of nodes, the nodes in the large group are highly similar, and the nodes in the smaller group are not very similar. In this case, when the similarity is greater than a threshold, we can merge the two sub-bi-communities.

Given two sub-bi-communities *subBC1* and *subBC2*, *(a, b)* and *(x, y)* are bi-Egos of the two sub-bi-communities, respectively. The contact ratio *CR* represents their level of overlap between the two sub-bi-communities. The first measure of the contact ratio, called *CR1*, which includes *CR1_A* and *CR1_B*) is shown in Equations (8) and (9):

$$CR1\_A(subBC1, subBC2) = \frac{|\ LA(a) \cap LA(x)\ |}{|\ LA(a) \cup LA(x)\ |}, \tag{8}$$

$$CR1\_B(subBC1, subBC2) = \frac{|\ LB(b) \cap LB(y)\ |}{|\ LB(b) \cup LB(y)\ |}, \tag{9}$$

where *LA(a)* represents A-type node set in sub-bi-community *subBC1*. *LB(b)* represents B-type node set in sub-bi-community *subBC1*. *CR1_A* represents the contact ratio of A-type node between *LA(a)* and *LA(x)*, and *CR1_B* represents the contact ratio of B-type node between *LB(b)* and *LB(y)*.

The other measurement of the contact ratio, called *CR2*, is given by:

$$CR2(subBC1, subBC2) = \frac{|\ LA(a) \cap LA(x)\ | + |\ LB(b) \cap LB(y)\ |}{|\ LA(a) \cup LA(x)\ | + |\ LB(b) \cup LB(y)\ |}, \tag{10}$$

where *CR2* represents the contact ratio of the hybrid-type node between the two sub-bi-communities *subBC1* and *subBC2*.

Algorithm 2 presents the pseudo-code for the merging stage. Fig. 6 shows the flow chart for the second stage.

The value of the parameter in OCDBEN was set as $\alpha = N(A)/N(E) \pm 0.1$, $\beta = N(B)/N(E) \pm 0.1$, $\gamma = 0.5 \pm 0.1$ and $\omega = 0.5 \pm 0.1$. To introduce the implementation of the OCDBEN algorithm, we use the South African companies network [35] (Fig. 1) as an example. Table 2 lists the detailed descriptions of 13 Bi-EgoNets in the South African companies network, including the Bi-EgoNet label, the bi-Ego, the set of bi-Alters, and the set *subBC*. In these 13 Bi-EgoNets, we set the similarity thresholds α and β to 0.5 and 0.4, respectively. We then extract the sub-bi-community sets *subBC*. Finally, we obtain three global bi-communities—{*A1, A2, A4, A6, B1, B2, B3*}, {*A3, B3, B4*}, {*A5, B2, B5*}—using double-merger strategy setting the thresholds of *CR1* and *CR2* as $\gamma = 0.6$ and $\omega = 0.6$,

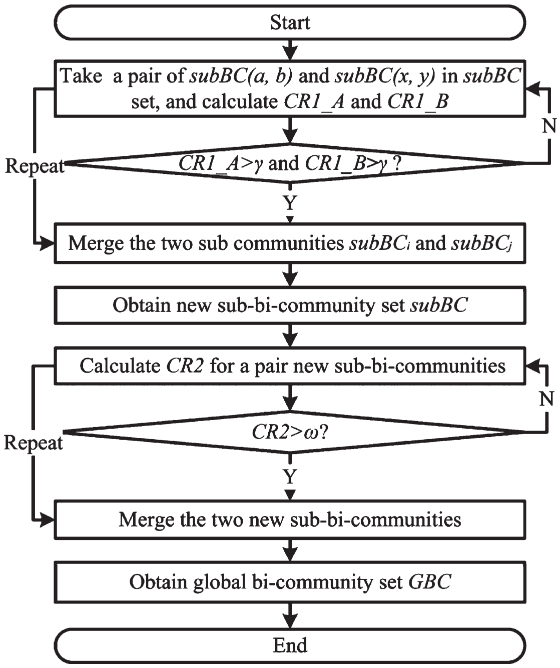


Fig. 6. The second stage of OCDBEN flowchart.

Algorithm 2 Merging sub-bi-communities using the double-merger strategy

Input: Sub-bi-community set *subBC*;
Output: Global bi-community set *GBC*;
1: For a pair of sub-bi-communities, calculate*CR1_A* (Equation (8)) and *CR1_B*(Equation (9));
2: Merge the pair of sub-bi-communities when *CR1_A*>γ and *CR1_B*>γ;
3: Repeat steps 1 and 2 until no sub-bi-community pairs satisfy the conditions, thereby forming a new sub-bi-community set *subBC*;
4: For a pair of new sub-bi-communities, calculate *CR2*(Equation (10));
5: Merge the pair of new sub-bi-communities when *CR2*>ω;
6: Repeat steps 4 and 5 until no new sub-bi-community pairs satisfy *CR2*>ω, thereby forming a new global bi-community set *GBC*;
7: **return** *GBC*;

Table 2
The Bi-EgoNet dataset of the South African Companies network

Bi-EgoNet	bi-Ego	Bi-Alters	Sub-bi-communities
1	(A1,B1)	(A2,A3,A4,A5,A6)(B2,B3)	(A2,A4,A6,A1)(B2,B3,B1)
2	(A1,B2)	(A2,A3,A4,A5,A6)(B2,B3)	(A1,A6)(B1,B2)
3	(A1,B3)	(A2,A3,A4,A5,A6)(B2,B3,B4)	(A1,A2,A4)(B1,B3)
4	(A2,B1)	(A1,A3,A4,A6)(B2,B3)	(A1,A2,A4)(B1,B3)
5	(A2,B3)	(A1,A3,A4,A6)(B1,B2,B4)	(A1,A2,A4)(B1,B3)
6	(A3,B3)	(A1,A2,A4)(B1,B2)	(A1,A2,A4)(B1,B3)
7	(A3,B4)	(A1,A2,A4)(B3)	(A3)(B3,B4)
8	(A4,B1)	(A1,A2,A3)(B2,B3)	(A1,A2,A4)(B1,B3)
9	(A4,B3)	(A1,A2,A3,A6)(B1,B2,B4)	(A1,A2,A4)(B1,B3)
10	(A5,B2)	(A1,A6)(B5)	(A5)(B2,B5)
11	(A5,B5)	(A1,A6)(B2)	(A5)(B2,B5)
12	(A6,B1)	(A1,A2,A4,A5)(B2,B3)	(A1,A6)(B1,B2)
13	(A6,B2)	(A1,A5)(B1,B3,B5)	(A1,A6)(B1,B2)

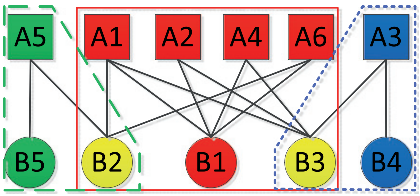


Fig. 7. The community structure of the South African companies network. Red nodes, blue nodes, and green nodes represent three distinct communities. Yellow nodes (B2 and B3) are overlapping nodes.

respectively. And *B2* and *B3* are overlapping nodes. The community structure is shown in Fig. 7.

3.3. Complexity analysis

OCDBEN consists of two main stages. We assume that there are m A-type and n B-type node in a bipartite network. In the sub-bi-communities extraction stage, the computational complexity of constructing the Bi-EgoNets for each pair of nodes in a bipartite network is $O(m*n)$. The computational complexity of extracting the sub-communities from each Bi-EgoNet is $O(m^2*n^2)$. In the merging stage, the maximum associated complexity is $O(m^2*n^2)$. Thus, the worst-case time complexity of OCDBEN algorithm is $O(m^2*n^2)$.

4. Evaluation criteria

In this section, we first extended Q_b [25] to quantify the overlapping community structure of bipartite networks. We then introduce the bipartite partition density (Density) [43] and *NMI* [22] measures.

4.1. Extended modularity of bipartite networks EQ_b

Modularity measures the structural strength of a network community. Given an unweighted, undirected bipartite network *BG(A, B, E)*, Barber [25] proposed the Q_b measure and applied it to non-overlapping community structures. The formula for modularity Q_b is

$$Q_b = \frac{1}{M}\sum_{c\in C}\left(\sum_{i=1}^{m}\sum_{j=1}^{n}\delta_{i,c}\delta_{j,c}\left(A_{ij}-\frac{d_i d_j}{M}\right)\right), \tag{11}$$

where M is the number of edges. c is a community in the network community C, m is the number of A-type nodes. n is the number of B-type nodes, and $\delta_{i,c}$ indicates whether node i belongs to community c. The value of $\delta_{i,c}$ is 1 when the node i belongs to the community c and 0 otherwise. A_{ij} is an adjacency matrix element. If there is an edge between i and j, then $A_{ij}=1$, otherwise $A_{ij}=0$. d_i is the degree of node i.

Since nodes in the real world may belong to more than one community, we extend the definition of $\delta_{i,c}$ to become a membership coefficient reflecting how much node i belongs to the community c. Shen et al. proposed the concept of the membership coefficient [34]. They point out that the membership coefficient should be normalized, such as $0 \leq \psi_{i,c} \leq 1, \forall i \in A$, $\forall c \in C$ and $\sum_{c\in C}\psi(i,c)=1$. That means that if the node i belongs to k communities, the membership coefficient $\psi(i,c)$ of node i in each community is $\frac{1}{k}$. With the membership coefficient $\psi(i,c)$, the modularity of overlapping community structure can be measured according to

$$EQ_b = \frac{1}{M}\sum_{c\in P}\left(\sum_{i=1}^{m}\sum_{j=1}^{n}\psi_{i,c}\psi_{j,c}\left(A_{ij}-\frac{d_i d_j}{M}\right)\right) \tag{12}$$

4.2. Bipartite partition density (Density)

Li et al. [43] proposed a new quantitative function called bipartite partition density (Density) for community detection in bipartite networks and defined it as according to:

$$Density = \frac{1}{M}\sum_{c\in C}\frac{M_c^2}{\mid LA_c \mid * \mid LB_c \mid}, \tag{13}$$

where M is the number of edges, C represents the community partition of a bipartite network, M_c is the number of edges in sub-bi-community c, and $|LA_c|$ indicates the number of A-type node in sub-bi-community c. The bipartite partition density can be used to detect overlapping community structures.

4.3. Normalized mutual information index (NMI)

Mcdaid et al. [22] extended the traditional normalization of variation, enabling it to evaluate the overlap between community partition. The method compares the experimental result to the standard partition. The higher the value of *NMI*, the more similar the experimental result is to the standard partition. For a pair of communities X and Y expressed as matrices of cluster membership, their associated *NMI* is shown in Equation (14). The *NMI* always falls into the range 0-1.

$$NMI = 1-\frac{1}{2}\left(\frac{H(X \mid Y)}{H(X)}+\frac{H(Y \mid X)}{H(Y)}\right). \tag{14}$$

where $H(X_i)$ is called information entropy of X_i, it represent the probability that the nodes in the network belong to community X_i. $H(X_i|Y_j)$ is called conditional entropy, it represents the degree of difference between X_i and Y_j.

5. Comparison methods

As you known, OCDBEN is an algorithm which detects overlapping community structure using similarity from a micro view in bipartite networks. Both Cui's method and Wang's method can also detect overlapping community structure using different similarity (intimacy degree) from a macro view in bipartite networks. And these three methods have similarities and differences. Thus we compare OCDBEN with the two methods, the experimental results are very valuable and practicable. These two methods are described in detail below.

Cui's method Cui et al. [41] defined two key concepts: key bi-communities and free node. In this algorithm, first, sorting nodes in order of increasing of node degree. Then constructing the basic key bi-community for each node, and merging these nodes using the intimacy degree between the basic key bi-communities of the each two nodes to form key bi-communities. Next, find out these nodes which do not belong to any key bi-communities, these nodes and their neighbors are free node. Finally, distribute these free nodes to key bi-communities.

Wang's method Wang et al. [40] defined two parameters to show the intimacy relationships between the same type nodes and heterogeneous nodes, respectively. In Wang's method, it first finds and expands core communities using intimacy relationships of the same type nodes. Then sub communities were obtained by merging the other type nodes to core communities using intimacy relationships between heterogeneous nodes. Lastly, final community structure is obtained by merging these sub communities.

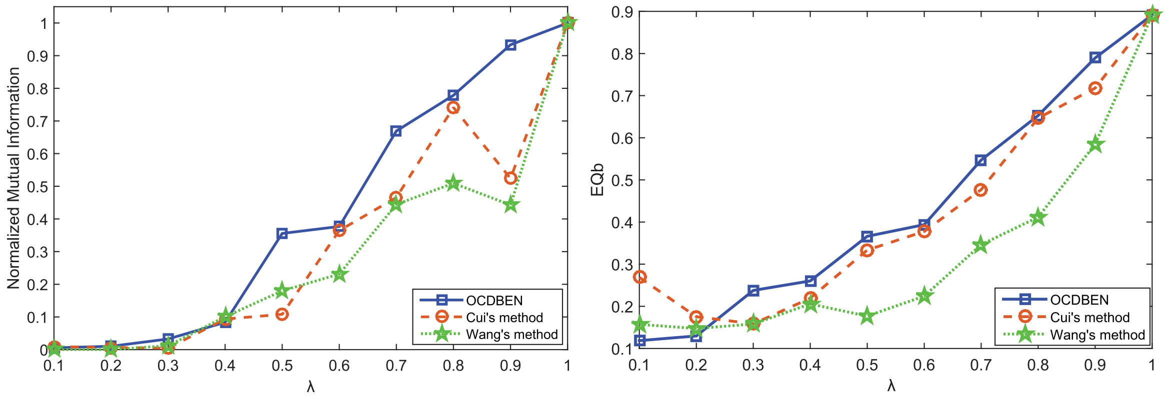
6. Experimental design and results

In this section, we present the result of our evaluation of the accuracy and effectiveness of the OCDBEN algorithm on several synthetic and real-world networks. two studies ([40, 41]) have documented the strong effectiveness of the methods of Cui and Wang at detecting overlapping community structure in bipartite networks, we compared OCDBEN with these two methods.

6.1. Synthetic bipartite networks

We used synthetic networks to evaluate the accuracy of OCDBEN. Using the synthetic network generation model proposed by Larremore et al. [19], each synthetic network consisted of four communities with equal numbers of nodes. Each community was made up of A-type nodes and B-type nodes, and λ is called the mixing parameter, represents noise ratio. It ranged from 0 (all noise) to 1 (no noise).

In our experiments, we selected synthetic networks with 256 nodes. λ ranged from 0.1 to 1, and



(a) NMI (b) EQ_b

Fig. 8. Comparison results in synthetic networks.

Table 3
Details of the 14 datasets

Name	m	n	Edge($\|E\|$)	$<k>$	Description
SAC	6	5	13	2.36	South African Companies [35]
SW	14	18	89	5.56	Southern women network [3]
Club	25	15	95	4.75	Club membership [6,17]
CL	20	24	99	4.5	Corporate leadership network [31]
D-US	50	9	225	7.63	Divorce in the United States [10]
AR	136	5	160	2.27	American revolution network [4]
DT-200	200	395	877	2.95	Dutch Top 200 network [37]
GP	314	360	1225	3.79	Graph product network [38]
Malaria	297	806	2965	5.38	Malaria gene substring network [21]
Crime	829	551	1476	2.14	Crime network [7]
PCD	680	739	3690	1.75	Protein complex-drug network [27]
Dutch	3811	937	5220	2.2	Dutch network [37]
M-UT	4009	16528	43760	4.26	Movie user-tag network [12]
Col	16727	22015	58595	3.02	Collaboration network [16, 28]

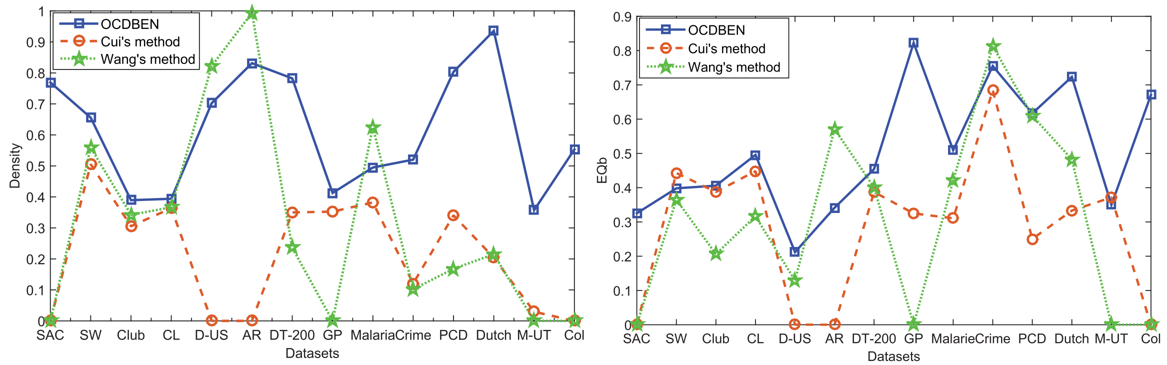
Table 4
Number of communities detected by the tested algorithms for different real-world datasets

C_num	SAC	SW	Club	CL	D-US	AR	DT-200	GP	Malaria	Crime	PCD	Dutch	M-UT	Col
OCDBEN	3	2	17	4	4	5	164	82	106	191	143	751	1708	5215
Cui's method	1	2	3	4	–	–	170	72	71	84	81	264	–	–
Wang's method	1	4	14	12	6	5	61	–	174	38	40	238	–	–

Table 5
Computation time (ms) Required by the three algorithms for the 13 real-world datasets

Name	SAC	SW	Club	CL	D-US	AR	DT-200	GP	Malaria	Crime	PCD	Dutch	M-UT	Col
OCDBEN	12	12	16	18	93	26	170	116	666	399	1348	3654	226579	206644
Cui's method	11	10	12	12	–	–	248	368	321	486	1067	17928	–	–
Wang's method	15	20	15	15	206	437	1965	387	839	1012	1904	11200323	–	–

the average degree was 4. Each synthetic network contained 128 A-type nodes and 128 B-type nodes. We applied OCDBEN, Cui's method, and Wang's method to these synthetic networks. The *NMI* and EQ_b results for the synthetic networks are shown in Fig. 8. In Fig. 8 (a) and (b), as λ declined, *NMI* values also gradually decreased. All of the algorithms accurately detected overlapping community structures in these synthetic networks when λ was 1. None of the methods detected community structures when λ was 0.1. When λ was between 0.1 and 1, the *NMI* and of the ODBCEN algorithm was greater than the *NMI* values of the other 2 methods for most synthetic networks. Thus, the community structures detected by



(a) Density (b) EQ_b

Fig. 9. Comparison chart of evaluation results of different methods.

our algorithm were closer to the actual community partition than the structure detected by the methods of Cui and Wang.

6.2. Real-world bipartite networks

To further verify the OCDBEN algorithm, we tested OCDBEN with 14 real-world datasets and compared the results with those from the methods Cui and Wang. The detailed information of 14 real-world datasets is shown in Table 3. In Table 3, *m* and *n* represent the number of each node type. $|E|$ is the number of edges. $<k>$ is the average degree.

Table 4 shows experimental results with the number of communities *C_num* detected in each dataset. OCDBEN effectively detected overlapping community structures in all 14 bipartite networks from different domains. Both of Cui's and Wang's methods are failed in dataset SAC, the main reason is that the number of nodes in SAC is too small for the two methods to divide it into smaller communities. In datasets D-US and AR, Cui's method are failed to detected overlapping community structures. Datasets D-US and AR share the most important feature that the number of nodes of one type is much smaller than the number of nodes of the other type, with communities formed from a single node of one type with multiple nodes of the other type. However, Cui's method considered each single node and its neighbors to belong to a free node set. Thus, Cui's method failed to detect the community structure in D-US and AR. Similarly, Wang's method detected community structures effectively in most datasets, with the exception of dataset GP. Since the regulation of flexibility in Wang's method is relatively weak (the value of parameter is 0.5), it was very difficult for this technique to detect community structures in the GP dataset. Finally, the numbers of nodes in M-UT and Col were both large, leading to the failure of Cui's and Wang's methods to complete in a reasonable amount of time (no exceed 5 hours).

We implemented the detection algorithms using the Java programming language on a personal computer with an Intel i5-3210M, 2.5 GHz processor, 4.0 GB of memory, and the Windows 10 operating system. Table 5 shows computation time required for each algorithm and dataset. As the network size grew, so did the computation time for all the three methods. However, OCDBEN's time grew far more slowly than the methods Cui and wang.

We also evaluated the overlapping community structures of OCDBEN by all three methods using, *Density* and EQ_b in Figs. 9 (a) and (b). Fig. 9 (a) shows that the *Density* values from OCDBEN were all superior to those of Cui's method. The *Density* values of OCDBEN with these datasets were also better than those of Wang's method, with the exception of the *D-US*, *AR* and *Malaria* for which the *Density* values from OCDBEN were 0.02, 0.1625 and 0.1302 less, respectively, than those of Wang's method. The average *Density* values of OCDBEN were 0.2776 and 0.1341 higher than those of Cui's and Wang's methods, respectively.

Similarity, Fig. 9(b) shows that the EQ_b values from OCDBEN were better than those of Cui's method, with the exception of SW, in which the EQ_b of OCDBEN was 0.0447 less than that of Cui's method. Likewise, the EQ_b values from OCDBEN in these datasets were better than those of Wang's method, with exception of *AR* and *Crime* datasets. The EQ_b values of OCDBEN in the two datasets were 0.23 and 0.0583 less, respectively, than those of

Wang's method. The average EQ_b values of OCDBEN were 0.1231 and 0.0937 which were greater than those of Cui's and Wang's methods, respectively.

7. Conclusions

As is known to all, detecting community structure lays the foundation of personalized recommendation and other related applications in bipartite networks. In real life, most networks have overlapping community structure. If only considering the non-overlapping community structure of the network, the application scope of the algorithm will be greatly reduced. EgoNet is a micro one-mode network model, can be used to analyze and detect overlapping communities in one-mode networks from a micro view. In this paper, our contribution to this field of research is threefold. First, we have extended EgoNet, and proposed Bi-EgoNet for analyzing bipartite network structures from a micro perspective. Second, we have created the OCDBEN algorithm, which is based on Bi-EgoNet, to detect overlapping community structures. Third, we have introduced an evaluation method, EQ_b, to determine the modularity of overlapping community structures in bipartite networks. We performed tests with synthetic and real-world networks to validate the new OCDBEN algorithm. Experimental results indicate that OCDBEN detected meaningful community structures in the original bipartite networks. The accuracy and effectiveness of OCDBEN were superior to those of other state-of-the-art algorithms. In the future, we will conduct further research in the following directions: to prove the characteristics of Bi-EgoNet using mathematical and statistical methods, to find a more efficient community detection method based on Bi-EgoNet for two-mode network, and to apply Bi-EgoNet to a recommendation model.

Acknowledgment

This work was partially supported by the Xinjiang Natural Science Foundation (No. 2016D01B010). We thank LetPub (www.letPub.com) for its linguistic assistance during the preparation of this manuscript.

References

[1] A. Epasto, S. Lattanzi, V. Mirrokni, I.O. Sebe, A. Taei and S. Verma, Ego-net community mining applied to friend suggestion, *Proceedings of the Vldb Endowmen* **9**(4) (2015), 324-335. doi: 10.14778/2856318.2856327.

[2] A.F. Liu, C.H. Fu, Z.P. Zhang, H. Chang and D.R. He, An empirical statistical investigation on chinese mainland movie network, *Complex Systems and Complexity Science* **4**(3) (2007), 10–17. doi: 10.1016/S1872-2040(07)60079-6.

[3] Allison Davis, Burleigh B. Gardner and Mary R. Gardner. *Deep South; a Social Anthropological Study of Caste and Class*, in: The University of Chicago Press, Chicago, 1941. http://konect.uni-koblenz.de/networks/opsahl-southernwomen.

[4] American revolution network dataset – KONECT, April, 2017, http://konect.uni-koblenz.de/networks/brunson_revolution.

[5] B. Abrahao, S. Soundarajan, J. Hopcroft and R. Kleinberg, A separability framework for analyzing community structure, *Acm Transactions on Knowledge Discovery from Data* **8**(1) (2014), 101–129. doi: 10.1145/2527231.

[6] *Club membership network dataset–KONECT*, April, 2017, http://konect.uni-koblenz.de/networks/brunson_club-membership.

[7] *Crime network dataset – KONECT*, April, 2017. http://konect.uni-koblenz.de/networks/moreno_crime.

[8] D.B. Larremore, A. Clauset and A.Z. Jacobs, Efficiently inferring community structure in bipartite networks, *Physical Review E* **90**(1) (2014), 012805. doi: 10.1103/PhysRevE.90.012805.

[9] I.S. Dhillon, *Co-clustering documents and words using bipartite spectral graph partitioning*, in: Acm Sigkdd International Conference on Knowledge Discovery & Data Mining, 2001, pp. 269–274. doi: 10.1145/502512.502550.

[10] *Divorce in US*, http://vlado.fmf.uni-lj.si/pub/networks/data/2mode/divorce.net.

[11] Furong Chang, Bofeng Zhang, Haiyan Li, Mingqing Huang, Bingchun Li and Yue Zhao, Discovering overlapping communities in ego-nets using friend intimacy, *Journal of Intelligent & Fuzzy Systems* 2019 (Preprint): 1-9.

[12] GroupLens Research. MovieLens data sets, October, 2006, http://www.grouplens.org/node/73.

[13] Gupta, X. Yan and K. Lerman, Structural properties of ego networks, in: SBP (2015), *Lecture Notes in Computer Science* **9021** (2015), 55-64. doi: 10.1007/978-3-319-16268-3_6.

[14] H. Jeong, B. Tombor, R. Albert, Z.N. Oltvai and A.-L. Barabasi, The large scale organization of metabolic networks, *Nature* **407** (2000), 651–654. doi: 10.1038/35036627.

[15] H.L. Sun, E. Ch'Ng, X. Yong, J.M. Garibaldi, S. See and D.B. Chen, A fast community detection method in bipartite networks by distance dynamics, *Physica A* **496**(15) (2018), 108–120. doi: 10.1016/j.physa.2017.12.099.

[16] Jérôme Kunegis, *KONECT–The Koblenz Network Collection*, in: Proc. Int. Conf. on World Wide Web Companion, (2013), pp. 1343–1350, http://userpages.uni-koblenz.de/~kunegis/paper/kunegis-koblenz-network-collection.pdf.

[17] Katherine Faust, Centrality in affiliation networks, *Social Networks* **19**(2) (1997), 157–191. doi: 10.1016/S0378-8733(96)00300-0.

[18] S. Kianian, M.R. Khayyambashi and N. Movahhedinia, FuSeO: Fuzzy semantic overlapping community detection, *Journal of Intelligent & Fuzzy Systems* **32**(6) (2017), 3987-3998.

[19] D.B. Larremore, A. Clauset and A.Z. Jacobs, Efficiently inferring community structure in bipartite networks, *Physical Review E* **90**(1) (2014), 012805, doi: 10.1103/PhysRevE.90.012805.

[20] H. Li, Detecting fuzzy network communities based on semi-supervised label propagation, *Journal of Intelligent & Fuzzy Systems* **31**(6) (2016), 2887-2893.

[21] H.J. Li and J. Xiang, Explore of the fuzzy community structure integrating the directed line graph and likelihood optimization, *Journal of Intelligent & Fuzzy Systems* **32**(6) (2017), 4503-4511.

[22] Mcdaid, F. Aaron, Greene Derek and Hurley Neil. Normalized Mutual Information to evaluate overlapping community finding algorithms, *Computer Science* 2011.

[23] M.E. Newman and M. Girvan, Finding and evaluating community structure in networks, *Physical Review E* **69**(2) (2004), 026113. doi: 10.1103/PhysRevE.69.026113.

[24] M.G. Everett and S.P. Borgatti, The dual-projection approach for two-mode networks, *Social Networks* **35**(2) (2013), 204–210. doi: 10.1016/j.socnet.2012.05.004.

[25] M.J. Barber, Modularity and community detection in bipartite networks, *Physical Review E* **76**(2) (2007), 066102. doi: 10.1103/PhysRevE.76.066102.

[26] M. Newman, The structure of scientific collaboration networks, *Proceedings of the National Academy of Sciences of the United States of America* **98**(2) (2001), 404–409.

[27] J.C. Nacher, J.M. Schwartz, Modularity in protein complex and drug interactions reveals new polypharmacological properties, *PloS one* **7**(1) (2012), e30028.

[28] Newman and E. J Mark, The Structure of Scientific Collaboration Networks, *Proceedings of the National Academy of Sciences* **98**(2) (2001), 404–409.

[29] N. Meghanathan, Use of eigenvalues and eigenvectors to analyze bipartivity of network graphs, in: International Conference on Wireless and Mobile Networks, 2014, pp. 221-230. doi: 10.5121/csit.2014.41218.

[30] P. Pesantezcabrera and A. Kalyanaraman, Efficient Detection of Communities in Biological Bipartite Networks, *IEEE/ACM Transactions on Computational Biology & Bioinformatics* **16**(1) (2017), 258–271. doi: 10.1109/TCBB.2017.2765319.

[31] Roy Barnes and Tracy Burkett, Structural redundancy and multiplicity in corporate networks, *International Network for Social Network Analysis* **30**(2) (2010). http://konect.uni-koblenz.de/networks/brunson_corporate-leadership.

[32] C. Sharma and P. Bedi, Community based hashtag recommender system (CHRS) for twitter, *Journal of Intelligent & Fuzzy Systems* **34**(3) (2018), 1511-1519.

[33] C. Sharma and P. Bedi, CCFRS–Community based Collaborative Filtering Recommender System, *Journal of Intelligent & Fuzzy Systems* **32**(4) (2017), 2987-2995.

[34] Shen Hua Wei, Cheng Xue Qi and Guo Jia Feng, Quantifying and identifying the overlapping com-munity structure in networks, *Journal of Statistical Mechanics Theory & Experiment* **7**(7) (2009), 07042.

[35] South african companies network dataset – KONECT, April, 2017, http://konect.uni-koblenz.de/networks/brunson_south-africa.

[36] T. Zhou, J. Ren, M. Medo and Y.C. Zhang, Bipartite network projection and personal recommendation, *Physical Review E* **76**(2) (2007), 046115. doi: 10.1103/PhysRevE.76.

[37] W. de Nooy, *Ringen om de macht*, in: Wilco Dekker&Ben van Raaij, De elite. De Volkskrant Top 200 van invloedrijkste Nederlanders. Amsterdam: Meulenhoff, 2006, pp. 85–94.

[38] W. Imrich and S. Klavzar, *Product graphs: Structure and recognition*, in: John Wiley & Sons, New York, USA, 2000.

[39] X. Liu and T. Murata, Community detection in large-scale bipartite networks, *Transactions of the Japanese Society for Artificial Intelligence* **1**(1) (2010), 50–57. doi:10.1145/1348549.1348552.

[40] X. Wang and X. Qin, Asymmetric intimacy and algorithm for detecting communities in bipartite networks, *Physica A* **462** (2016), 569-578. doi: 10.1016/j.physa.2016.06.096.

[41] Y. Cui and X. Wang, Uncovering overlapping community structures by the key bi-community and intimate degree in bipartite networks, *Physica A* **407** (2014), 7–14. doi: 10.1016/j.physa.2014.03.077.

[42] Zhao Zhongyinga, Zheng Shaoqianga, Li Chao, Sun Jinqing, Chang Liang and Chiclana Francisco, A comparative study on community detection methods in complex networks, *Journal of Intelligent & Fuzzy Systems* **35**(1) (2018), 1077-1086.

[43] Z. Li, R.S. Wang, S. Zhang and X.S. Zhang, Quantitative function and algorithm for community detection in bipartite networks, *Information Sciences* **367** (2016), 874–889. doi:10.1016/j.ins.2016.07.024.