Article

Overlapping Community Detection of Bipartite Networks Based on a Novel Community Density

Yubo Peng, Bofeng Zhang * and Furong Chang

School of Computer Engineering and Science, Shanghai University, Shanghai 200444, China; redefine@shu.edu.cn (Y.P.); changfurong123@163.com (F.C.)

* Correspondence: bfzhang@shu.edu.cn; Tel.: +86-18817355917

Abstract: Community detection plays an essential role in understanding network topology and mining underlying information. A bipartite network is a complex network with more important authenticity and applicability than a one-mode network in the real world. There are many communities in the network that present natural overlapping structures in the real world. However, most of the research focuses on detecting non-overlapping community structures in the bipartite network, and the resolution of the existing evaluation function for the community structure's merits are limited. So, we propose a novel function for community detection and evaluation of the bipartite network, called community density D . And based on community density, a bipartite network community detection algorithm DSNE (Density Sub-community Node-pair Extraction) is proposed, which is effective for overlapping community detection from a micro point of view. The experiments based on artificially-generated networks and real-world networks show that the DSNE algorithm is superior to some existing excellent algorithms; in comparison, the community density (D) is better than the bipartite network's modularity.

Citation: Peng, Y.; Zhang, B.; Chang, F. Overlapping Community Detection of Bipartite Networks Based on a Novel Community Density. *Future Internet* **2021**, *13*, 89. <https://doi.org/10.3390/fi13040089>

Academic Editor: Eirini Eleni Tsiropoulou

Received: 23 February 2021

Accepted: 29 March 2021

Published: 31 March 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: overlapping community; bipartite network; community detection; community density

1. Introduction

A complex network is the network with a high degree of complexity, and its complexity is mainly reflected in the large scale of the system, diverse structures, and complex types of nodes. According to the number of node types, the complex network can be classified as the one-mode network and multi-mode network. The network that contains only one type of node is called the one-mode network, such as financial networks [1], social networks [2], and the Internet [3]. The network containing more than two types of nodes in the network is called the multi-mode network. The bipartite network is one of the multi-mode networks defined as containing two types of nodes. There is no internal connection of the same type of node, but there are edges that exist only between different types of nodes. In the real world, there are various types of bipartite networks, such as protein networks [4] and scientific collaboration networks [5]. The community of a complex network is a set composed of vertices in the complex network. The vertices in the same community are closely connected, while the vertices in different communities are sparsely connected. Such as various associations generally tend to communicate with people in the same association rather than with people outside the association. Figure 1 shows a community example of a South African Companies bipartite network [6]. The areas framed by blue and yellow dotted lines represent community 1 and community 2, respectively, and the green nodes represent the overlapping nodes in the community. Because of its wide applicability and universality, the bipartite network is the hotspot of current researchers. There are various perspectives and methods around complex network research. As an important feature, we can obtain deeper topology, hidden meaning, and network information through community structure detection. In the real world, a bipartite network community has been found to solve many practical problems. For example, the study of bipartite network

mining provides a new research basis and method for the epidemic model [7]. In the field of recommendation, the research on the community structure detection of the bipartite network can provide a new method for the recommendation technology of bipartite network systems. For example, in the bipartite network recommendation system of e-commerce users and commodities, if users are divided according to their characteristics, users with different characteristics are divided into different communities, and products are recommended for users according to the relationship between users in the communities, then the recommended products will be more in line with the needs of users.

Figure 1. Bipartite network example—South Africa companies bipartite network.

According to the subordinate relationship of nodes, community structure can be divided into overlapping community structure and non-overlapping community structure. Traditional community detection algorithms mostly focus on the non-overlapping communities in the bipartite network, while few studies focus on the overlapping communities in the bipartite network. But in the real world, a node usually has more than one attribute. A node can belong to more than one community. For example, in the scientific collaboration network [5], various authors can write a paper, and an author can sign various papers. Therefore, it has more crucial practical value and scientific research value to study the detection of bipartite network overlapping communities.

Recently, with more and more researchers going deep into the field of bipartite network community detection, quite a few algorithms have emerged in the field of community detection [7–11]. In general, these algorithms can be divided into two categories. One is to project the bipartite network to the one-mode network [12–14]; the advantages of this approach can be based on the current research results of the one-mode network, which is easier to operate. However, in projecting to the one-mode network, the bipartite network would lose some original information and increase the noise. The other is to directly detect the community on the bipartite network [15]. Compared with the projection of the bipartite network, it is more challenging to operate. Still, the original network structure information can be protected to the greatest extent not to be destroyed in the projection process. However, this kind of approach mainly starts from the network's topology structure and carries out community detection from the macro perspective. It takes little consideration of the relationship between nodes and fails to discover some microscopic structural features of the bipartite network.

How to measure the merits of community detection algorithm results is another crucial problem. So far, there has not been a recognized universal evaluation index for detecting bipartite network communities. The existing evaluation indexes are mostly based on the modularity of bipartite network [11,16,17] and make some improvements to it. However,

due to the modularity resolution limitation [18], modularity cannot judge the merits and demerits of community detection results in some particular cases.

In this article, our main contribution is:

- A new evaluation index (Community Density D) is proposed, which alleviates the problem of modularity resolution limitation and evaluates the bipartite network's merits and demerits from a new point.
- A new community detection algorithm (Density Sub-community Node-pair Extraction (DSNE)) is proposed. It has the advantages of a micro-network model, adopts a new merging strategy for community merging, and can detect the bipartite network's overlapping community structure.
- The experimental results show that the proposed algorithm is superior to other algorithms' accuracy and effectiveness.

2. Related Work

Compared with the one-mode network, the bipartite network's topology is more complex, and it has stronger universality and practicability. Therefore, the research on the problem detection of the bipartite network community is more valuable. There are two main problems in the bipartite network community detection: finding the community structure more effectively, and another is how to evaluate the merits and demerits of the community detection results. In conclusion, there are mainly two kinds of research on the bipartite network's community detection algorithm, and the evaluation index of the bipartite network is mainly aimed at the improvement of the modularity of the bipartite network.

2.1. Based on the Projecting of the Bipartite Network Community Detection Algorithm

The methods of projecting bipartite networks to one-mode networks are mainly divided into unweighted one-mode projecting and weighted one-mode projecting. Unweighted one-mode projecting is simply projecting bipartite networks to one-mode networks. Newman et al. [19] converted a bipartite network (scientific collaboration network) into two one-mode networks by using a single-mode right projection. Horvat et al. [20] proposed a statistical method that extends the bipartite network projection algorithm containing a single relationship and solves multiple types of relationship projection in large datasets. Valejo et al. [21] applied the traditional layered method to the projection process of bipartite network, proposes a multi-layered method based on projection. Comar et al. [22] proposed a framework that could simultaneously detect the network structures of different communities and their communication, with advantages, such as strong adaptability and fast convergence. However, in the process of unweighted one-mode projection, network structure information, such as edge frequency, may be lost, leading to a significant decrease in the accuracy of community detection by this method. Due to the apparent defects of unweighted modular projecting, researchers turned their attention to weighted modular projecting.

The weighted one-mode projecting converts the frequency of edges in a bipartite network to the weight of edges in a one-mode network during the projection process. Liu et al. [23] presented a novel fast nonnegative matrix (F-NMTF) method for clustering two kinds of nodes in a bipartite network. Zhang et al. [12] used a structure that can describe the relationship between two kinds of nodes, the Weighted Symmetric Bipartite Matrix Factorization (WSBMF), to detect the overlapping communities in the bipartite network. Wang et al. [13] presented a new way to find and extend the second part of the network's core community. Define two parameters to represent the relationship between nodes of the same type and heterogeneous nodes, respectively. In Zhou et al. [14], based on the Louvain algorithm, the Bi-Louvain algorithm suitable for the bipartite network is proposed, and a two-stage algorithm with partial modularization and quantification of partition strength is applied to bipartite network community detection. Pesantez-Cabrera et al. [24] proposed an efficient algorithm, called Bilouvann, which implements a set of heuristics for fast and accurate community detection in bipartite networks. Compared with

non-weighted modular projecting, the weighted modular projecting method can reduce structural information loss to a large extent. Still, its disadvantage is that the accuracy of the community detection results depends on the calculation of weights, which has substantial uncertainty, affecting the use of such methods.

2.2. Processed Directly on the Bipartite Network

The advantage of direct processing of bipartite network community detection algorithm is that it can not only effectively avoid the loss of structural information in the process of unweighted projection, keep the original network structural information to the maximum extent, but also avoid the problem of increasing noise in the process of weighted projection. Liu et al. [25] proposed a new algorithm, LP&BRIM, based on label propagation (LP) algorithm and BRIM algorithm, which is suitable for large-scale bipartite network detection. Beckett et al. [26] proposed two new algorithms, LPAWB + DirtLAPWB. It can find the largest community in the network by maximizing the bipartite network's weighted modularity and searching the partition with the largest modularity. Sun et al. [27] proposed a new bipartite community Attractor algorithm, BiAttractor, to extend the distance dynamic model Attractor into the bipartite network. Wang et al. [28] proposed a bipartite network overlapping community detection algorithm (MACD-BNS). This algorithm realizes overlapping community detection by optimizing the four existing methods for calculating similar nodes' similarity. Che et al. [29] proposed a new meme algorithm, MATMCD-BN, which approximates the optimal global solution by local search through a new population initialization method and a new crossover operator. Chang et al. [30] proposed a new community detection algorithm, CBG&BEN, which is based on the complete bipartite graph (CBG) and micro-network model (Bi-EgoNet). Gmati et al. [31] proposed a new algorithm called Bi-Comdet, the main thrust of the introduced approach is that it stresses the importance of grouping two types of nodes in communities having a full connection between its node. Taguchi et al. [32] proposed a new multi-label propagation algorithm BIMLPA, which overcomes the limitations of previous propagation algorithms and has good speed and stability. Yen et al. [33] extended the stochastic block model (SBM) model to bipartite networks and proposed biSBM; the biSBM improves community detection results over general SBM when data are noisy and improves the model resolution limit by a factor of two.

2.3. Community Detection Evaluation Criteria

If there is no scientific evaluation index of community detection results, then the community detection algorithm is at a loss, so how to evaluate the merits of community detection results is another important topic in community detection research. Especially for overlapping communities, the criteria for evaluation of community structure are very complex. It is difficult to give a unified and recognized evaluation index. Barber [16] extended the modularity of the traditional one-mode network to the bipartite network. Murata et al. [17] proposed another kind of modularity of bipartite networks, which can be represented by a pair of two numerical groups in two directions, allowing one-to-many correspondence of communities of different vertex types. Suzuki et al. [34] proposed a new modularity measurement method based on M.E.J Newman, which applies to bipartite networks and non-uniform networks. Chang et al. [30] proposed a new modularity evaluation method, which extended the modularity to overlapping community detection based on M.E.J Newman's method. This evaluation criterion can evaluate the findings of overlapping communities in bipartite networks. All evaluation criteria described above are limited by modularity resolution [18]. The problem of resolution limit was first proposed by Fortunato et al. [35] in 2007. Fortunato et al. [36] find that multi-resolution modularity may not be able to prepare to identify a large number of network community structures with scale-free characteristics in reality. Specifically, multi-resolution modularity may merge some small communities into large communities and may split large communities into small communities. Li et al. [37] proposed a mass function of community partition

community density to measure the quality of community structure in a bipartite network. Still, this mass function is only applicable to the flying overlapping community structure in the bipartite network.

3. Community Density

This section is divided into two parts: the definition of community density D , reasoning D solves the modularity resolution limitation to a certain extent. We want to define community density D to measure the quality of community division is mainly due to the problems left by the history of modularity. Take the latest methods as examples. Chang et al. [30] proposed modularity EQ_b suitable for bipartite network overlapping community structure, but it is still limited by modularity resolution. Li [37] proposed Bipartite Partition Density, which can effectively alleviate the modularity resolution limitation and considers the community as a whole and cannot be used for overlapping communities. Although these methods can detect community structure, they lack to consider community structure from a micro perspective, such as the relationship between node pairs within the community. Based on this, from the micro point of view, this paper considers the relationship between the node pairs within the community and puts forward an evaluation standard suitable for the bipartite network overlapping community structure, which can effectively alleviate the modularity resolution limitation.

3.1. Key Definitions

Given a bipartite network $G = (U, V, E)$ ($U \neq \emptyset, V \neq \emptyset, U \cap V = \emptyset$) with no weight, no direction and the edges exist only between nodes of different types. $U = \{u \mid u \in U\}$: U is a one node typeset, $V = \{v \mid v \in V\}$: V is a another node typeset. a represents the number of nodes of type U , and b represents the number of nodes of type V . E represents the set of edges connected by two types of nodes, and (u, v) represents a pair of nodes connected by edge $e_{(u,v)}$. An example diagram of a bipartite network is shown in Figure 2. There are five U nodes and six V nodes, with a total of thirteen edges.

$$E = \{e_{(u,v)} \mid u \in U, v \in V\}. \quad (1)$$

Figure 2. Bipartite network example.

In a bipartite network $G = (U, V, E)$, $PN(u)$ represents the set of all node pairs containing node u , $L(PN(u))$ represents the number of node pairs u in U , and then $L(PN(u)) = L(u)$ is the total number of node pairs in the set. Likewise, $PN(v)$ represents the set of all node pairs containing node v , we let $L(PN(v))$ be the number of node pairs v in V , and then $L(PN(v)) = L(v)$ is the total number of node pairs in the set. As shown in Figure 3: For node $U1$, $PN(U1) = \{(U1, V1), (U1, V2), (U1, V3)\}$; for node $V1$, $PN(V1) = \{(U1, V1), (U2, V1), (U4, V1), (U6, V1)\}$. We express these as:

$$PN(u) = \{(u, v) \mid v \in V, e_{u,v} \in E\}, \quad (2)$$

$$PN(v) = \{(u, v) \mid u \in U, e_{u,v} \in E\}, \quad (3)$$

$$L(u) = L(PN(u)) = \sum_{j=1}^b (u, v_j), \quad (4)$$

$$L(v) = L(PN(v)) = \sum_{i=1}^a (u_i, v). \quad (5)$$

Figure 3. All node pairs containing nodes $U1$ and $V1$.

In a bipartite network $G = (U, V, E)$, $NeiPN(u, v)$ represents a set of neighbor node pairs of node pairs (u, v) .

$$NeiPN(u, v) = \{PN(p) \cup PN(q) \mid p \in PN(u), q \in PN(v), (p, q) \neq (u, v)\}. \quad (6)$$

In Figure 2, select a node pair $(U1, V1)$, and then its neighbor node pairs set is $NeiPN(U1, V1) = \{(U1, V2), (U1, V3), (U2, V1), (U2, V3), (U4, V1), (U4, V3), (U6, V1), (U6, V2)\}$. Figure 4 shows a set of neighbor node pairs of node pair $(U1, V1)$.

Figure 4. A set $NeiPN(U1, V1)$ of neighbor node pairs of node pairs $(U1, V1)$.

In a bipartite network $G = (U, V, E)$, assume that the bipartite network G is divided into X communities, where the t -th community is expressed as $G_t = (U_t, V_t, E_t)$, and $\varphi(u, v)$ represents whether the node pair (u, v) is a node pair within the community. $D_{t(m,n)}$ represents the node pair density of node pairs (m, n) in community G_t ; in $\sum_{k=1}^{L(m)} \varphi(m, v_k)$, $L(m)$ represents the total number of all V nodes connected to m , k represents the serial number of node v_k connected to m , and, when $k = 1$, it means that m is connected to V node v_1 . For node pair (m, v_k) , if they are edges $e_{(m,v_k)}$ within the t -th community, then $\varphi(m, v_k) = 1$. We express these as:

$$\varphi(u, v) = \begin{cases} 1, & e_{(u,v)} \in E_t \\ 0, & e_{(u,v)} \notin E_t \end{cases} \quad (7)$$

$$D_{t(m,n)} = \frac{\sum_{k=1}^{L(m)} \varphi(m, v_k) + \sum_{k=1}^{L(n)} \varphi(u_k, n)}{L(m) + L(n)}. \quad (8)$$

Suppose we use the bipartite network shown in Figure 2 for community detection, and the results are shown in Figure 1. There are two communities in Figure 2. In community 1, select node pair $(U1, V1)$ to calculate $D_{1(U1, V1)}$. For node pair $(U1, V1)$, $PN(U1) = \{(U1, V1), (U1, V2), (U1, V3)\}$, $PN(V1) = \{(U1, V1), (U2, V1), (U4, V1), (U6, V1)\}$, $L(U1) = 3, L(V1) = 4$. The calculation formula of $D_{1(U1, V1)}$ is shown in Formula (9).

$$\begin{aligned} D_{1(U1, V1)} &= \frac{\sum_{k=1}^{L(U1)} \varphi(U1, v_k) + \sum_{k=1}^{L(V1)} \varphi(u_k, V1)}{L(U1) + L(V1)} \\ &= \frac{\varphi(U1, V1) + \varphi(U1, V2) + \varphi(U1, V3) + \varphi(U1, V1) + \varphi(U2, V1) + \varphi(U4, V1) + \varphi(U6, V1)}{3 + 4} \\ &= \frac{1 + 1 + 1 + 1 + 1 + 1 + 0}{7} = \frac{6}{7} \end{aligned} \quad (9)$$

In the t -th community $G_t = (U_t, V_t, E_t)$, by adding and averaging the node pair density $D_{t(u_i, v_j)}$ of node pairs (u_i, v_j) , the density D_t of the t -th community can be shown as Formula (10). Among them, P_t represents the sum of all U nodes in the t -th community, Q_t represents the sum of all V nodes in the t -th community, $\sum_{i=1}^{P_t} \sum_{j=1}^{Q_t} D_{t(u_i, v_j)}$ represents the sum of node pair density of all node pairs in the t -th community, $\sum_{i=1}^{P_t} \sum_{j=1}^{Q_t} \varphi(u_i, v_j)$ represents the sum of all node pairs in the t -th community.

$$D_t = \frac{\sum_{i=1}^{P_t} \sum_{j=1}^{Q_t} D_{t(u_i, v_j)}}{\sum_{i=1}^{P_t} \sum_{j=1}^{Q_t} \varphi(u_i, v_j)}. \quad (10)$$

In community 1, $P_t = 4, Q_t = 4$, the calculation formula of D_1 is shown in Formula (11).

$$\begin{aligned} D_1 &= \frac{\sum_{i=1}^4 \sum_{j=1}^4 D_{1(u_i, v_j)}}{\sum_{i=1}^4 \sum_{j=1}^4 \varphi(u_i, v_j)} \\ &= \frac{D_{1(U1, V1)} + D_{1(U1, V2)} + D_{1(U1, V3)} + D_{1(U2, V1)} + D_{1(U2, V3)} + D_{1(U3, V3)} + D_{1(U3, V4)} + D_{1(U4, V1)} + D_{1(U4, V3)}}{1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1} \\ &= \frac{\frac{6}{7} + \frac{2}{3} + \frac{7}{7} + \frac{5}{6} + \frac{6}{6} + \frac{6}{6} + \frac{3}{3} + \frac{5}{6} + \frac{6}{6}}{9} = \frac{\frac{172}{21}}{9} \approx 0.91 \end{aligned} \quad (11)$$

Likewise, we can obtain $D_2 \approx 0.78$. After obtaining the density of a certain community G_t , the density of the whole bipartite network community D is the average density of all communities. X represents the number of communities in the bipartite network $G = (U, V, E)$.

$$D = \frac{1}{X} \sum_{t=1}^X D_t = \frac{1}{X} \sum_{t=1}^X \frac{\sum_{i=1}^{P_t} \sum_{j=1}^{Q_t} D_{t(u_i, v_j)}}{\sum_{i=1}^{P_t} \sum_{j=1}^{Q_t} \varphi(u_i, v_j)} \quad (12)$$

For the community detection shown in Figure 1, $D = \frac{1}{2} \sum_{t=1}^2 D_t = \frac{1}{2} (D_1 + D_2) \approx 0.85$.

When we use community density D as an evaluation criterion of community detection, the higher the community density is, the closer the node pairs within the community are, and the sparser the node pairs between the communities are. This trend is very consistent with the community structure required by community detection. Therefore, community density can be used to determine whether community detection results are good or bad.

3.2. Reasoning Community Density Alleviates the Modularity Resolution Limitation

The root cause of the problem of modularity resolution limitation [18] is that modularity cannot effectively identify the results of community division. This problem has existed for quite a long time, but there has been no suitable method to completely solve this problem. The community density proposed in this paper can effectively solve the modularity resolution limitation in some cases. Now, let us look at an example.

There is a bipartite network $G(U, V, E)$. This bipartite network is a complete graph, in which there are $L(u)$ nodes of type u , and $L(v)$ nodes of type v , $L(u) \geq 2$, $L(v) \geq 2$, $(L(u) - L(v)) \bmod 2 = 0$. Then, this bipartite network has NV vertices and NE edges. This type of picture is called SCBG (Special Complete Bipartite Graph).

$$NV = L(u) + L(v), \quad (13)$$

$$NE = L(u) * L(v). \quad (14)$$

The community detection of this graph yielded the following two results:

R1: Divide the result into a community G .

R2: Evenly divided into two communities $\{G_1(U_1, V_1, E_1), G_2(U_2, V_2, E_2)\}$. $U_1 \cap U_2 = 0$, $U_1 \cup U_2 = U$, $V_1 \cap V_2 = 0$, $V_1 \cup V_2 = V$. For community G_1 , $L(u)_1 = \frac{1}{2}L(u)$, $L(v)_1 = \frac{1}{2}L(v)$, the total number of edges in community G_1 is $\frac{1}{2}NE$. Among them, the number of edges in the community E_{1in} accounts for half of the total number of edges in the community G_1 , and the number of edges outside the community E_{1out} accounts for the other half. Likewise, for community G_2 , $L(u)_2 = \frac{1}{2}L(u)$, $L(v)_2 = \frac{1}{2}L(v)$, G_1 owns half of node U and node V , and G_2 owns the other half of node U and V , and the total number of edges in community G_2 is $\frac{1}{2}NE$. Among them, the number of edges in the community E_{2in} accounts for half of the total number of edges in the community G_2 , and the number of edges outside the community E_{2out} accounts for the other half.

An example of the above situation is given in Figure 5, where $L(u) = 4$ and $L(v) = 4$. Figure 5a is divided according to R1, and Figure 5b is divided according to R2.

Figure 5. When $u = 4$ and $V = 4$, the complete bipartite graph (CBG) is divided according to R1 and R2.

Barber [16] proposed the modularity Q_b suitable for bipartite networks. Assuming that a bipartite network can be expressed as $G(U, V, E)$, the formula for the modularity Q_b of the bipartite network is as Formula (15):

$$Q_b = \frac{1}{M} \sum_{c \in C} \left(\sum_{i=1}^m \sum_{j=1}^n \delta_{i,c} \delta_{j,c} \left(A_{ij} - \frac{d_i d_j}{M} \right) \right), \quad (15)$$

where M represents the number of edges in the bipartite network, C represents the set of community in the bipartite network, c means the c -th community $G_c(U_c, V_c, E_c)$, and m and n represent the number of two types of nodes in the bipartite network, respectively. $\delta_{(i,c)}$ represents the membership degree of node i to community c , and if node i belongs to community c , then c is 1, otherwise c is 0. d_i represents the degree of node i . A_{ij} represents the adjacency matrix of bipartite network, if there is an edge between i and j , $A_{ij} = 1$, else $A_{ij} = 0$.

For the two kinds of community partition results, R_1 and R_2 , the modularity is calculated, respectively. For the community structure in Figure 5a, the calculation of modularity Q_{R_1} is shown in Formula (16). For the community structure in Figure 5b, the calculation of modularity Q_{R_2} is shown in Formula (17).

$$Q_{R_1} = \frac{1}{NE} \left((L(u) + L(v)) * 1 * 1 * \left(1 - \frac{L(u) * L(v)}{NE}\right) \right) = \frac{1}{NE} * 0 = 0, \quad (16)$$

$$\begin{aligned} Q_{R_2} &= \frac{1}{NE} \left(\sum_{i=1}^{L(u_1)} \sum_{j=1}^{L(v_1)} \delta_{i,L(v)} \delta_{L(u),j} \left(1 - \frac{L(u) * L(v)}{NE}\right) + \sum_{i=1}^{L(u_2)} \sum_{j=1}^{L(v_2)} \delta_{i,L(v)} \delta_{L(u),j} \left(1 - \frac{L(u) * L(v)}{NE}\right) \right) \\ &= \frac{1}{NE} \{ E_{1_{in}} * 1 * 1 * \left(1 - \frac{L(u) * L(v)}{NE}\right) + E_{1_{out}} * 1 * 0 * \left(1 - \frac{L(u) * L(v)}{NE}\right) + E_{2_{in}} * 1 * 1 * \left(1 - \frac{L(u) * L(v)}{NE}\right) + E_{2_{out}} * 1 * 0 * \left(1 - \frac{L(u) * L(v)}{NE}\right) \} \\ &= \frac{1}{NE} (0 + 0 + 0 + 0) \\ &= 0 \end{aligned} \quad (17)$$

As Equations (16) and (17) show, the modularity of the two community detection results calculated by Barber's modularity is 0. However, R_1 and R_2 have different community detection but $Q_{R_1} = Q_{R_2}$. In R_1 , divide the network into a community. Logically, the best partition is the one where all the nodes are put in a single cluster. So, there are no cuts, no edges between clusters. So, the modularity proposed by Barber cannot effectively judge the merits and demerits of the results.

Let us use the community density proposed in this paper to calculate the two partition results. D_{R_1} represents the community density divided according to R_1 , and D_{R_2} represents the community density divided according to R_2 .

For R_1 , any node pair (u, v) in community G , all node pairs containing G are within the community, so the calculation formula of node pair (u, v) density $D_{(u,v)}$ is as follows (18).

$$D_{(u,v)} = \frac{L(n) + L(m)}{L(m) + L(n)} = 1, \quad (18)$$

$$D_{R_1} = \frac{1}{1} * \sum_{t=1}^1 \frac{L(m) * L(n)}{NE} = 1. \quad (19)$$

For R_2 , in community G_1 , for any node pair (u, v) , half of any node pair containing (u, v) is in the community, and the other half is outside the community. Similarly, the situation in community G_2 is exactly the same as in G_1 .

$$D_{1_{(u,v)}} = \frac{L(u)_1 + L(v)_1}{L(u) + L(v)} = \frac{1}{2}, \quad (20)$$

$$D_{G_1} = \frac{\frac{1}{2}(NE) * \frac{1}{2}}{\frac{1}{2}(NE)} = \frac{1}{2}, \quad (21)$$

$$D_{G_2} = D_{G_1} = \frac{1}{2}, \quad (22)$$

$$D_{R_2} = \frac{1}{2} * (D_{G_1} + D_{G_2}) = \frac{1}{2}. \quad (23)$$

After calculation, $D_{R_1} \neq D_{R_2}$, $D_{R_1} - D_{R_2} = \frac{1}{2}$. Therefore, community density can be used to evaluate the merits and demarcations of such results. Effectively alleviate the resolution limit of modularity.

4. DSNE Algorithm

4.1. Definition of Community Similarity and Bi-EgoNet

Given two communities $G_1(U_1, V_1, E_1)$ and $G_2(U_2, V_2, E_2)$, and CD_{G_1, G_2} represents the similarity of the two communities, as shown in Formula (24):

$$CD_{G_1, G_2} = \frac{|L(u_1) \cap L(u_2)| + |L(v_1) \cap L(v_2)|}{|L(u_1) \cup L(u_2)| + |L(v_1) \cup L(v_2)|}. \quad (24)$$

In the DSNE (Density Sub_community and Node-pair Extraction) algorithm, we use the Bi-EgoNet [11] network model proposed by Chang et al., which can detect the overlapping communities of the bipartite network from a microscopic point of view. Figure 6 shows an example of this model. As shown in Figure 6, in a bipartite network $G(U, V, E)$, the microscopic bipartite network model Bi-EgoNet of node pair (u_2, v_1) is composed of a central node pair (u_2, v_1) , its neighbor node pair set $NeiPN(u_2, v_1)$, and the edges between these node pairs. (u_2, v_1) is the Bi-Ego, which is the center node pair; $NeiPN(u_2, v_1)$ is the Bi-Alter, which is the each neighbor node pair.

Figure 6. Bi-EgoNet microscopic network model of node pairs (U_2, V_1) .

The algorithm is divided into two stages: the first stage is to construct a sub-community for each node pair centered on each node pair; the second stage is to combine all node pair sub-communities to get the global community structure.

4.2. Construct Sub-Communities for Each Node Pair

In this stage, it is important to calculate node pairs' density (Equation (8)). The flow of the first stage is shown in Algorithm 1.

Algorithm 1 Construct Sub-communities For Node Pair.**Input:** A bipartite network $G(U, V, E)$;**Output:** Sub-communities set: $Sub - DC$;

- 1: Input bipartite network $G(U, V, E)$;
- 2: To traverse the bipartite network. For each node pair (u_i, v_j) , (u_i, v_j) as the Bi-Ego, get the set $PN(u_i)$ of all node pairs including node u_i , get the set $PN(v_j)$ of all node pairs including node v_j ;
- 3: Calculate the set $NeiPN(u_i, v_j)$ of all adjacent node pairs of node pair (u_i, v_j) ;
- 4: Each node pair (u_i, v_j) is numbered, the first node pair is numbered as 1, the second node pair is numbered as 2, and so on, the t node pair is numbered as t , until all node pairs have been numbered;
- 5: With Bi-Ego (u_i, v_j) numbered as t as the center, the initial value of t is 1, the sub-community $G_t(U_t, V_t, E_t)$ is constructed. In the beginning, there is only one node pair (u_i, v_j) in the community;
- 6: Repeat step 5 and increment t by one each time until all Bi-Ego (u_i, v_j) as the center in the $G(U, V, E)$ has been traversed. Correspondingly, all sub-communities have been constructed;
- 7: For communities $G_t(U_t, V_t, E_t)$, Suppose we add a node pair (m, n) from the set $NeiPN(u_i, v_j)$ to this community, calculate the density $D_{t(m,n)}$ of node pair (m, n) after joining. If $D_{t(m,n)} > \alpha$, add (m, n) to the community, Otherwise, do not join;
- 8: Repeat step 7 until all node pairs in the set $NeiPN(u_i, v_j)$ have been traversed;
- 9: Repeat steps 7 and 8 until all sub-communities in the set $Sub - DC$ have been traversed;
- 10: **return** $Sub - DC$.

4.3. A New Strategy to Merge The Sub-Communities

For two communities to be merged, $G_1(U_1, V_1, E_1)$ and $G_2(U_2, V_2, E_2)$, the usual practice is to merge the node pairs and edges directly. However, this will increase the noise, such as adding some node pairs that are not very closely connected to the newly merged community. In this regard, we provide a new merging idea: starting from the perspective of node pairs, we extract the central node pairs (u_1, v_1) and (u_2, v_2) from two communities $G_1(U_1, V_1, E_1)$ and $G_2(U_2, V_2, E_2)$ to be merged, and form a new community $NG_1(U_1, V_1, E_1)$ from the two type of node pairs. We define an identifier for each sub-community $Sub_{(u_i, v_j)}(U_i, V_j, E_{u_i, v_j})$: *mark*, which indicates whether the central node pair in the sub-community has participated in the merge.

$$mark = \begin{cases} 1, & \text{has been involved in mergers} \\ 0, & \text{Not involved in mergers} \end{cases}. \quad (25)$$

The flow of the first stage is shown in Algorithm 2.

Algorithm 2 Use New Strategy To Merge Sub-communities.**Input:** Sub-communities set $Sub - DC$;**Output:** The merged community set NC ;

- 1: Sort in descending order according to the number of node pair in the sub-communities of $Sub - DC$ set, Add a mark to each sub-community with an initial value of 0;
- 2: The sub-community $G_i(U_i, V_i, E_i)$ with the current $mark = 0$ and the most node pairs is selected to merge, and the merged community is $NG_j(U_j, V_j, E_j)$. The initial value of j is 1, the central node pair of $G_i(U_i, V_i, E_i)$ is selected as the central node pair of the merged community, and the mark of the sub-community $G_i(U_i, V_i, E_i)$ is set as 1;
- 3: Calculate the similarity CD_{G_i, G_k} between sub-community $G_i(U_i, V_i, E_i)$ and other sub-communities $G_k(U_k, V_k, E_k)$, if $CD_{G_i, G_k} > \beta$, add the central node pairs of sub-community $G_k(U_k, V_k, E_k)$ to the merged community $NG_j(U_j, V_j, E_j)$, and the mark of the sub-community $G_k(U_k, V_k, E_k)$ is set as 1;
- 4: Repeat step 3 until all sub-communities have been traversed, j increment by 1;
- 5: Repeat steps 2, 3, and 4 until all sub-communities have marks of 1;
- 6: Return $NC = \{NG_p(U_p, V_p, E_p) \mid p = 1, 2, 3, \dots, j\}$.

In the DSNE algorithm, parameter α is used to judge whether the neighbor node pair (m, n) can join the sub-community $G_t(U_t, V_t, E_t)$. We believe that, if half of all the neighbor node pairs of a node pair are in the community, then the node pair should be in the community, so parameter α is set to 0.5 ± 0.1 . The β parameter is defined as the ratio of all identical node pairs in both communities $G_i(U_i, V_i, E_i)$ and $G_k(U_k, V_k, E_k)$ to the total node pairs, parameter β is set to 0.5 ± 0.1 .

Complexity Analysis

Algorithm DSNE is divided into two stages to complete, so we analyze the complexity in turn according to the stage. Assume that there are m nodes of type U and n nodes of type V in the bipartite network. In constructing sub-communities for each node pair, the time complexity required for constructing sub-communities for each central node pair is $O(m^2n^2)$, and, in merging sub-communities, the time complexity required is $O(m^2n^2)$. Therefore, $O(m^2n^2)$ is the time complexity of the algorithm. Since the algorithm does not use unknown storage space during execution, the space complexity is $O(n)$.

5. Experimental Evaluation

In this section, we choose two methods, LOCD [9] and CBG&BEN [30], as the comparison algorithm, community density (D), NMI [38], and EQ_b [11] as the evaluation indexes, and we use manually generated network [37] and the network in the real world as the dataset to evaluate DSNE algorithm. The LOCD algorithm and CBG&BEN algorithm are written in the Java programming language, while the DSNE algorithm is written in the C++ programming language and runs on a 2.3 GHz processor, 8 GB of memory, and the macOS Big Sur operating system. Finally, the superiority of the proposed evaluation index, community density (D), is verified in evaluating community detection results compared with modularity.

5.1. Contrast Algorithm

The DSNE algorithm proposed in this paper uses community density to detect the overlapping community structure in the bipartite network starting from node pairs. LOCD

algorithm uses the similarity of nodes to detect the overlapping community structure in the bipartite network from the graph's topology. CBG&BEN algorithm uses the Bi-EgoNet, a micro-network model to detect the bipartite network's overlapping community structure based on the complete dichotomous graph. These three methods have certain similarities, and the LOCD algorithm and CBG&BEN algorithm are relatively novel and excellent in community detection results. Therefore, comparing the DSNE algorithm with these two algorithms can better illustrate the DSNE algorithm's superiority. We briefly describe the two algorithms below.

Wang et al. [9] proposed LOCD algorithm for selecting the right partitions in bipartite networks in 2017. This method contains two parameters; one parameter is used to represent the similarity between nodes of the same types, and the other parameter is used to represent the similarity between nodes of different types. The two types of nodes are processed separately. Firstly, extend the sub-community with the first parameter, which is the similarity between nodes of the same type. The similarity between nodes of different types is used to merge different types of nodes with the sub-community, and many sub-communities are obtained. Then, the sub-communities are merged according to some specific merge rules to obtain the global community structure.

In 2019, Chang et al. [30] proposed a community detection algorithm, CBG&BEN, which is based on the micro-network model (Bi-EgoNet) and complete bipartite graph (CBG). First, the algorithm constructed the micro bipartite network model Bi-EgoNet for each node pair. Second, it used the complete binary graph to extract the basic complete binary graph from each Bi-EgoNet, and the set of local communities was obtained by merging rules. Third, according to specific merging rules, the global bipartite network community structure is obtained. Finally, the network's divergent nodes are collected and allocated to the corresponding global community, and the final global community structure is finally obtained.

5.2. Evaluation Criteria

5.2.1. Extended Modularity of Bipartite Networks EQ_b

Chang et al. [30] extended the modularity of bipartite network proposed by Barber [16] and proposed EQ_b , which is suitable for the evaluation of detection results in the overlapping communities of the bipartite network. The formula of EQ_b is shown in (26).

$$EQ_b = \frac{1}{M} \sum_{c \in C} \left(\sum_{i=1}^{|U|} \sum_{j=1}^{|V|} \psi_{i,c} \psi_{j,c} \left(A_{ij} - \frac{d_i d_j}{M} \right) \right). \quad (26)$$

We introduced Barber's modularity in Formula (16), Except for $\psi_{i,c}$, the meanings of other attributes in EQ_b are consistent with the modularity proposed by Barber. The membership coefficient $\psi_{i,c}$ represents the proportion of node i belonging to community c . The membership coefficient $\psi_{i,c}$ should meet the normalized characteristics, such as $0 \leq \psi_{i,c} \leq 1$ and $\sum_{c \in C} \psi_{i,c} = 1$. For example, there are a total of $m + n$ edges connected to node i , among which m edges belong to community c , and n edges do not belong to community c ; then, $\psi_{i,c} = \frac{m}{m+n}$.

5.2.2. NMI

Andrea et al. [39] proposed a standardized mutual information index NMI (Normalized Mutual Information Index) for community detection in complex networks. When we want to measure the difference between the community structure obtained by the algorithm and the known real community structure, NMI is generally used as the evaluation index. NMI applies to many types of complex networks and is also applicable to evaluating the bipartite network's detection results overlapping communities. The value of NMI

is between 0 and 1. The larger the NMI value is, the closer the algorithm's community structure is to the real community. The calculation formula of NMI is shown in (31).

$$H(X) = I(X : Y) + H(X|Y), \quad (27)$$

$$H(Y) = I(X : Y) + H(Y|X), \quad (28)$$

$$H(X, Y) = H(X|Y) + I(X : Y) + H(Y, X), \quad (29)$$

$$I(X : Y) = \frac{1}{2}(H(X) - H(X|Y) + H(Y) - H(Y|X)), \quad (30)$$

$$NMI = 1 - \frac{1}{2} \left(\frac{H(X|Y)}{H(Y)} + \frac{H(Y|X)}{H(X)} \right). \quad (31)$$

$I(X : Y)$ represents the mutual information between X and Y , $H(X)$ represents X 's entropy information, and $H(X|Y)$ represents X 's and Y 's entropy information under certain conditions. When $X = Y$, $H(X|Y) = 0$, then NMI reaches the maximum value of 1; otherwise, when X and Y are completely different, $H(X|Y) = 1$, then NMI is the minimum value of 0.

5.3. Experiments on Artificially Generated Networks

We refer to the artificially-generated network model proposed by Li [37]. The artificially-generated network model consists of three parameters: the number of network nodes N , the number of network edges $|E|$, and the noise parameter λ . In Li's artificially-generated network, each network has four communities with the same number of nodes. There are two different types of nodes, U and V in the community, and each community contains the same number of U and V nodes. The number of sides $|E| = 2N$, that is, each U node is associated with 4 V nodes. λ is the network's noise, that is, the number of sides between different communities in the network. The size of λ ranges from 0 to 1. When λ is 0, it means that the network is full of noise. When λ is 1, it means that there is no noise in the network.

In this experiment, an artificially-generated network with 256 nodes is selected. There are 128 U nodes and 128 V nodes in the network. There are four communities in the network. Each community contains 32 U nodes and 32 V nodes. There are 128 edges in each community. The initial value of λ is 0.1, increasing by 0.1 every time, increasing by nine times until the maximum value is 1. The DSNE algorithm, LOCD algorithm, and CBG&BEN algorithm are applied to the designed network, and the community detection is obtained. Table 1 shows the results of NMI evaluation.

Table 1. Normalized Mutual Information Index (NMI) evaluation of the experimental results of three methods in artificially-generated network with different indexes λ .

λ	DSNE	CBG&BEN	LOCD
0.1	0.00	0.00	0.00
0.2	0.01	0.01	0.00
0.3	0.04	0.03	0.03
0.4	0.09	0.07	0.10
0.5	0.32	0.24	0.18
0.6	0.71	0.67	0.22
0.7	0.94	0.93	0.43
0.8	0.96	0.90	0.51
0.9	0.98	0.96	0.44
1.00	1.00	1.00	1.00

As Table 1 shows, the NMI index of the DSNE algorithm is better than the other two algorithms in most cases, which indicates that the community structure found by the DSNE algorithm is closer to the real community structure than the other two algorithms. Compared with the CBG&BEN algorithm, the average NMI of the DSNE algorithm is

improved by 2.6%, compared with the LOCD algorithm, 12.5% improves the average NMI of the DSNE algorithm. In the λ decreasing from 0.6 to 0.4, the value of the NMI index showed a cliff-like downward trend. As Table 1 shows, when λ fluctuates around 0.5, the impact on the algorithm's accuracy is huge. When $\lambda \in [0.1, 0.4]$, the value of NMI does not exceed 0.1, which indicates that only less than 10% of the nodes in the community found by the algorithm correspond to the real network. Therefore, when $\lambda \in [0.1, 0.4]$, the community structures found by the three algorithms are invalid.

The decreasing speed of the algorithm represents the anti-noise ability of the algorithm. We consider setting $\lambda \in [0.5, 1]$ and comparing the decreasing speed of the three algorithms under different λ . The results are shown in Table 2. The average decreasing speed of the DSNE algorithm was 2.7% less than that of CBG&BEN and 7.3% less than that of LOCD. Therefore, the anti-noise ability of LOCD algorithm is superior to the other two algorithms.

Table 2. When $N = 256$, NMI decreasing speed with noise increasing (λ decreasing) under different algorithms.

λ	DSNE	CBG&BEN	LOCD
0.5	54.9%	64.1%	18.1%
0.6	24.4%	27.9%	48.8%
0.7	2.1%	−3.3%	15.6%
0.8	2.0%	6.25%	−15.9%
0.9	2.0%	4.0%	56.0%

In artificially-generated networks, the number of nodes directly affects network structure information. Therefore, we consider setting $\lambda \in [0.5, 1]$ and implementing the DSNE algorithm on artificially-generated networks with 64, 128, and 256 nodes. The results are shown in Figure 7.

Figure 7. Density Sub-community Node-pair Extraction (DSNE) algorithm uses NMI to evaluate experimental results under different N conditions.

Obviously, under different N , with the increase of noise, that is, during the process of λ decreasing from 1 to 0, the decrease rate of NMI is different. Table 3 shows the decrease rate of NMI under different N .

Table 3. NMI decreasing speed with noise increasing (λ decreasing) under different N .

λ	$N = 64$	$N = 128$	$N = 256$
0.5	34%	23%	57%
0.6	25%	38%	24%
0.7	31%	27%	3%
0.8	18%	15%	2%
0.9	40%	10%	2%

According to Table 3 and Figure 7, when $N = 256$, with the increase of noise, i.e., the decrease of λ , NMI decreases more slowly. When $\lambda > 0.7$, and decreases rapidly when λ hovers around 0.5. When $N = 128$, NMI's value showed a strong downward trend when λ hovered around 0.6. When $N = 64$, NMI showed a strong downward trend when λ hovered around 0.9. Therefore, we find that the increase of the number of network edges N is negatively related to NMI and λ ; that is to say, more noise is needed to make the algorithm imprecise. When the value of N is 64, 128 and 256, the corresponding average decline rates of NMI are 29.6%, 22.6%, and 17.%, which indicates that with the increasing value of N , the decline rate of NMI is getting slower and slower, that is, the stability of the algorithm is getting better. Besides, with the increase of N , there is more and more structural information in the network, and the average value of NMI is also increasing. In other words, the matching degree between the community detected by the algorithm and the real community is higher.

5.4. Experiments on Real-World Bipartite Network

In this section, we will verify the performance of the DSNE algorithm in the real world network, the CBG&BEN algorithm, and LOCD algorithm is used as the comparison algorithm, and eleven real bipartite network datasets from different domains are used as the dataset to evaluate the DSNE algorithm. The detailed information of these networks is shown in Table 4, where m and n represent the number of different types of nodes in the bipartite network, $|E|$ represents the number of edges between nodes, and $\langle k \rangle$ represents the average degree of the network. Taking the SAC dataset in Figure 2 as an example, there are six U nodes and five V nodes, and the network has a total of thirteen edges, so $m = 6$, $n = 5$, $|E| = 13$, and $\langle k \rangle = \frac{13+13}{5+6} = 2.36$.

Table 4. Details of 11 real-world networks.

Name	$ E $	$\langle k \rangle$	m	n	Description
SAC	13	2.36	6	5	South African Companies [6]
SW	89	5.56	14	18	Southern Women network [40]
CLUB	95	4.75	25	15	Club membership network [41]
CL	99	4.5	20	24	Corporate Leadership network [42]
AR	169	2.27	136	5	American revolution network [43]
DUS	225	7.73	50	9	Divorce in the United States [44]
GP	1225	3.79	314	360	Graph Product network [45]
CRIME	1476	2.14	829	551	Crime network [46]
MARIA	2965	5.38	297	806	Malaria gene substrating network [47]
PCD	3690	1.75	680	739	Protein complex-drug network [48]
COL	58,595	3.02	16,727	22,015	Collaboration network [49]

In order to explain the better understand the dataset we used in Table 4, the South African Companies Bipartite Network (SAC) is taken as an example to illustrate. Figure 2 shows the SAC dataset. Among them, the square nodes (U1 to U6) represent different leaders, the round nodes (V1 to V5) represent different companies, and the edges between nodes represent the leadership relationship between leaders and companies. The community detection results after DSNE algorithm is run are given in Figure 1. The merged community set NC contains two communities, which are, respectively, $NG_1(U_1, V_1, E_1)$

and $NG_2(U_2, V_1, E_2)$. In $NG_1(U_1, V_1, E_1)$, $U_1 = \{U1, U2, U3, U4\}$, $V_1 = \{V1, V2, V3, V4\}$, $E_1 = \{(e_{(U1,V1)}, e_{(U1,V2)}, e_{(U1,V3)}, e_{(U2,V1)}, e_{(U2,V3)}, e_{(U3,V3)}, e_{(U3,V4)}, e_{(U4,V1)}, e_{(U4,V3)})\}$. In $NG_2(U_2, V_1, E_2)$, $U_2 = \{U5, U6\}$, $V_2 = \{V1, V2, V5\}$, $E_2 = \{e_{(U5,V2)}, e_{(U5,V5)}, e_{(U6,V1)}, e_{(U6,V2)}\}$.

Table 5 shows the number of communities found in the 11 datasets by the three methods. Table 5 shows that the DSNE algorithm and CBG&BEN algorithm can effectively get the number of communities, while the LOCD algorithm cannot effectively get the community structure in some datasets. LOCD algorithm cannot effectively obtain the number of SAC communities; the main reason is that the SAC dataset contains too little network information than the LOCD algorithm, so the number of communities obtained in the SAC dataset using the LOCD algorithm is 1. In the COL dataset, the LOCD algorithm is considered ineffective because it still fails to obtain sufficient community numbers after a long time (6 h) of operation. In most datasets, the number of communities is not always the more, the better. The more the number of communities, the more the noise interference in the merger process, the less the communities involved in the merger, the lower the merger rate, and the final number of communities is too many. DSNE algorithm is superior to CBG&BEN algorithm and LOCD algorithm in the number of communities in most datasets.

Table 5. Number of communities obtained by the three algorithms in the eleven datasets.

Number of Communities	SAC	SW	CLUB	CL	AR	DUS	GP	CRIME	MARIA	PCD	COL
DSNE	2	2	5	6	4	2	75	91	101	66	4819
CBG&BEN	3	2	4	6	5	2	88	148	122	100	5215
LOCD	1	4	14	12	5	6	1	38	174	40	-

Figure 8 shows the experimental results of 3 algorithms evaluated by EQ_b in 11 datasets. From Figure 8, we can get a conclusion: the EQ_b of the DSNE algorithm is better than the other two algorithms. Because the LOCD algorithm cannot effectively detect the community structure in SAC and COL datasets, LOCD modularity in these two datasets is 0. Except for CL, PCD, and CRIM datasets, the DSNE algorithm's performance is better than the other two algorithms. In the CL dataset, DSNE is slightly 0.02 less than the CBG&BEN algorithm. In the PCD dataset, DSNE is 0.11 less than the CBG&BEN algorithm, and in the CRIME dataset, DSNE is 0.01 less than the LOCD algorithm. In conclusion, the average EQ_b of the DSNE algorithm is 5.5% higher than that of the CBG&BEN algorithm and 33.3% higher than that of LOCD algorithm.

Figure 8. Experimental results of three algorithms evaluated by EQ_b in 11 datasets.

Figure 9 shows the experimental results of 3 algorithms evaluated by D in 11 datasets. Compared with the LOCD algorithm, except for Maria, AR, and DUS datasets, the D of the DCIM algorithm is better than the LOCD algorithm. In the Maria dataset, the D of the

DSNE algorithm is 14% less than the LOCD algorithm, and in the AR dataset, the D of the DSNE algorithm is 5% less than the LOCD algorithm. In the DUS dataset, the DSNE algorithm is compared with LOCD 6% less. In SAC and GP datasets, because the LOCD algorithm could not effectively divide the community, the D of the LOCD algorithm on these two datasets is 0. Compared with the CBG&BEN algorithm, the D of DSNE is better than the CBG&BEN algorithm in all datasets. In the SW, GP, and CRIME datasets, the D of DSNE algorithm is 17%, 17%, and 12% higher than that of the CBG&BEN algorithm, respectively. In general, the DSNE algorithm's performance on D is better than that of LOCD room rate and CBG&BEN algorithm. Using D as the evaluation criterion, the average D of the DSNE algorithm is 8.6% higher than the CBG&BEN algorithm and 30.6% higher than the LOCD algorithm.

Figure 9. Experimental results of three algorithms evaluated by D in 11 datasets.

Table 6 shows the results of the comparison between D and EQ_b under different datasets using the DSNE algorithm and CBG&BEN algorithm for community detection. Since EQ_b can't evaluate the detected community structure properly, Both algorithms are 0 in SCBG using EQ_b evaluation, the D value of CBG&BEN is 0.5 more than that of DSNE because DSNE adopts a more strict strategy when conducting sub-community merger. In SW, CL, and DUS datasets, the CBG&BEN algorithm, and DSNE algorithm use EQ_b as the evaluation index, and the values are relatively close. Within the range of effective data accuracy, the EQ_b values of the two algorithms are the same. We can see from Table 6 that, when the two algorithms use D as the evaluation index, their performance is different in the above three datasets; so, when EQ_b as the evaluation standard appears, the same value in the dataset, we can use D to measure the merits and demerits of community detection. In DUS and Maria datasets, EQ_b is used as the evaluation standard, and CBG&BEN algorithm and DSNE algorithm are used for community detection. The results show that the EQ_b value of the MARIA dataset is better than that of the DUS dataset. Still, the MARIA dataset's D value is lower than that of the DUS dataset when D is used as the evaluation standard, which indicates that there are more edges between communities, but the community division is relatively broken. It is not very complete.

Table 6. DSNE algorithm and CBG&BEN algorithm are compared to use EQ_b and D as evaluation index on different datasets.

Dataset	SCBG		SW		CL		DUS		CRIME	
	EQ_b	D	EQ_b	D	EQ_b	D	EQ_b	D	EQ_b	D
CBG&BEN	0	1	0.41	0.50	0.54	0.57	0.26	0.75	0.79	0.35
DSNE	0	0.5	0.41	0.67	0.54	0.66	0.26	0.76	0.8	0.48

6. Conclusions

In this paper, we propose a new evaluation index: community density D , which can effectively solve the modularity resolution limitation to a certain extent. Besides, a novel algorithm, DSNE, is proposed, which combines the advantages of the proposed community density D to obtain the community. Among them, in the second stage of the DSNE algorithm: merging communities, we put forward a new merging idea to reduce the increase of noise during merging.

The DSNE algorithm is evaluated in the artificial and real-world networks and compared with the CBG&BEN algorithm and LOCD algorithm. Experimental results show that the DSNE algorithm can detect meaningful community structures in artificial and real-world networks, and its accuracy and effectiveness are better than LOCD and CBG&BEN algorithms. The average EQ_b of the DSNE algorithm is 5.5% higher than that of the CBG&BEN algorithm and 33.3% higher than that of the LOCD algorithm. The average D of the DSNE algorithm is 8.6% higher than the CBG&BEN algorithm and 30.6% higher than the LOCD algorithm. In the future, we will carry out further research on the following aspects: to find the optimal global solution based on the optimal local solution in the first stage of the algorithm and to find a more effective bipartite network community detection algorithm based on community density D .

Author Contributions: Conceptualization, Y.P.; Data curation, Y.P.; Formal analysis, Y.P.; Investigation, Y.P.; Methodology, Y.P.; Project administration, Y.P.; Supervision, B.Z.; Validation, Y.P.; Visualization, Y.P.; Writing—original draft, Y.P.; Writing—review & editing, B.Z. and F.C. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the National Key R&D Program of China under Grant 2017YFC0907505.

Data Availability Statement: Not Applicable, the study does not report any data.

Acknowledgments: The authors would like to thank anonymous reviewers for their fruitful feedback and comments that have helped them improve the quality of this work.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Bollmer, G.D. Community as a Financial Network: Mortgages, Citizenship, and Connectivity. *Democr. Commun.* **2011**, *24*, 39.
2. Newman, M.E.J. The structure and function of complex networks. *SIAM Rev.* **2003**, *45*, 167–256. [CrossRef]
3. Hu, Y.; Chen, H.; Zhang, P.; Li, M.; Di, Z.; Fan, Y. A New Comparative Definition of Community and Corresponding Identifying Algorithm. *arXiv* **2008**, arXiv:0802.0242.
4. Jeong, H.; Mason, S.; Barabasi, A. Oltvai ZN: Lethality and centrality in protein networks. *Nature* **2001**, *411*, 41–42. [CrossRef]
5. Everett, M.G.; Borgatti, S.P. The dual-projection approach for two-mode networks. *Soc. Netw.* **2013**, *35*, 204–210. [CrossRef]
6. South African Companies Network Dataset—KONECT. Available online: http://konect.cc/networks/brunson_south-africa/ (accessed on 29 March 2021).
7. Wu, G.; Gu, C.; Qiu, L.; Yang, H. A uniform framework of projection and community detection for one-mode network in bipartite networks. *Chin. Phys. B.* **2017**, *26*, 128901. [CrossRef]
8. Cui, Y.; Wang, X. Uncovering Overlapping Community Structures by the Key Bi-Community and Intimate Degree in Bipartite Networks. *Phys. A Stat. Mech. Its Appl.* **2014**, *407*, 7–14. [CrossRef]
9. Wang, X.; Liu, G.; Li, J. Overlapping Community Detection Based on Structural Centrality in Complex Networks. *IEEE Access* **2017**, *5*, 25258–25269. [CrossRef]
10. Feng, L.; Zhou, C.; Zhao, Q. A spectral method to find communities in bipartite networks. *Phys. A Stat. Mech. Its Appl.* **2019**, *513*, 424–437. [CrossRef]
11. Chang, F.; Zhang, B.; Zhao, Y.; Wu, S.; Niu, S. Overlapping Community Detection in Bipartite Networks using a Micro-bipartite Network Model: Bi-EgoNet. *J. Intell. Fuzzy Syst.* **2019**, *37*, 7965–7976. [CrossRef]
12. Zhang, Z.Y.; Ahn, Y.Y. Community Detection in Bipartite Networks Using Weighted Symmetric Binary Matrix Factorization. *Int. J. Mod. Phys. C* **2015**, *26*, 1550096. [CrossRef]
13. Wang, X.; Qin, X. Asymmetric Intimacy and Algorithm for Detecting Communities in Bipartite Networks. *Phys. A Stat. Mech. Its Appl.* **2016**, *462*, 569–578. [CrossRef]
14. Zhou, C.; Feng, L.; Zhao, Q. A Novel Community Detection Method in Bipartite Networks. *Phys. A Stat. Mech. Its Appl.* **2018**, *492*, 1679–1693. [CrossRef]

15. Yang, J.; McAuley, J.; Leskovec, J. Detecting Cohesive and 2-Mode Communities Indirected and Undirected Networks. In Proceedings of the 7th ACM International Conference on Web Search and Data Mining, New York, NY, USA, 24–28 February 2014; pp. 323–332. [\[CrossRef\]](#)
16. Barber, M.J. Modularity and community detection in bipartite networks. *Phys. Rev. E* **2007**, *76*, 066102. [\[CrossRef\]](#) [\[PubMed\]](#)
17. Tsuyoshi Murata, T.I. A new modularity for detecting one-to-many correspondence of communities in bipartite networks. *Adv. Complex Syst.* **2010**, *13*, 19–31. [\[CrossRef\]](#)
18. Guimerà, R.; Sales-Pardo, M.; Amaral, L.A.N. Module identification in bipartite and directed networks. *Phys. Rev. E* **2007**, *76*, 036102. [\[CrossRef\]](#)
19. Newman, M.E.J. The structure of scientific collaboration networks. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 404–409. [\[CrossRef\]](#)
20. Horvat, E.A.; Zweig, K.A. One-mode Projection of Multiplex Bipartite Graphs. In Proceedings of the IEEE/ACM International Conference on Advances in Social Networks Analysis & Mining, Istanbul, Turkey, 26–29 August 2013.
21. Valejo, A.; Ferreira, V.; Filho, G.P.R.; de Oliveira, M.C.F.; de Andrade Lopes, A. One-Mode Projection-Based Multilevel Approach for Community Detection in Bipartite Networks. Available online: <http://ceur-ws.org/Vol-2029/paper8.pdf> (accessed on 29 March 2021).
22. Comar, P.M.; Tan, P.N.; Jain, A.K. A Framework for Joint Community Detection across Multiple Related Networks. *Neurocomputing* **2012**, *76*, 93–104. [\[CrossRef\]](#)
23. Yang, L.; Tao, W.; Xin-Sheng, J.; Caixia, L.; Mingyan, X. Detecting Communities in 2-Mode Networks via Fast Nonnegative Matrix Trifactorization. Available online: <https://www.hindawi.com/journals/mpe/2015/937090/> (accessed on 29 March 2021).
24. Pesantez-Cabrera, P.; Kalyanaraman, A. Efficient Detection of Communities in Biological Bipartite Networks. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2019**, *16*, 258–271. [\[CrossRef\]](#)
25. Liu, X.; Murata, T. Community Detection in Large-scale Bipartite Networks. In Proceedings of the 2009 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology, Milan, Italy, 15–18 September 2009.
26. Beckett, S.J. Improved community detection in weighted bipartite networks. *R. Soc. Open Sci.* **2016**, *3*, 140536. [\[CrossRef\]](#)
27. Sun, H.L.; Ch'ng, E.; Yong, X.; Garibaldi, J.M.; See, S.; Chen, D.B. A Fast Community Detection Method in Bipartite Networks by Distance Dynamics. *Phys. A Stat. Mech. Its Appl.* **2018**, *496*, 108–120. [\[CrossRef\]](#)
28. Wang, X.; Liu, J. A Comparative Study of the Measures for Evaluating Community Structure in Bipartite Networks. *Inf. Sci.* **2018**, *448*, 249–262. [\[CrossRef\]](#)
29. Che, S.; Yang, W.; Wang, W. A Memetic Algorithm for Community Detection in Bipartite Networks. *IEEE Access* **2019**, *7*, 126897–126912. [\[CrossRef\]](#)
30. Chang, F.; Zhang, B.; Zhao, Y.; Wu, S.; Yoshigoe, K. Overlapping Community Detecting Based on Complete Bipartite Graphs in Micro-Bipartite Network Bi-Egonet. *IEEE Access* **2019**, *7*, 91488–91498. [\[CrossRef\]](#)
31. Gmati, H.; Mouakher, A.; Hilali-Jaghdam, I. Bi-COMDET: Community Detection in Bipartite Networks. *Procedia Comput. Sci.* **2019**, *159*, 313–322. [\[CrossRef\]](#)
32. Taguchi, H.; Murata, T.; Liu, X. BiMLPA: Community Detection in Bipartite Networks by Multi-Label Propagation. In *Proceedings of NetSci-X 2020: Sixth International Winter School and Conference on Network Science*; Masuda, N., Goh, K.I., Jia, T., Yamanoi, J., Sayama, H., Eds.; Springer: Cham, Switzerland, 2020; pp. 17–31.
33. Yen, T.C.; Larremore, D.B. Community Detection in Bipartite Networks with Stochastic Blockmodels. *Phys. Rev. E* **2020**, *102*, 032309. [\[CrossRef\]](#)
34. Suzuki, K.; Wakita, K. Extracting Multi-facet Community Structure from Bipartite Networks. In Proceedings of the 2009 International Conference on Computational Science and Engineering, Vancouver, BC, Canada, 29–31 August 2009.
35. Fortunato, S.; Barthélemy, M. Resolution Limit in Community Detection. Available online: <https://www.pnas.org/content/pnas/104/1/36.full.pdf> (accessed on 29 March 2021).
36. Lancichinetti, A.; Fortunato, S. Limits of modularity maximization in community detection. *Phys. Rev. E* **2011**, *84*, 066122. [\[CrossRef\]](#)
37. Li, Z.; Wang, R.S.; Zhang, S.; Zhang, X.S. Quantitative function and algorithm for community detection in bipartite networks. *Inf. Sci.* **2016**, *367*, 874–889. [\[CrossRef\]](#)
38. Mcdaid, A.F.; Greene, D.; Hurley, N. Normalized Mutual Information to evaluate overlapping community finding algorithms. *arXiv* **2011**, arXiv:1110.2515.
39. Lancichinetti, A.; Fortunato, S.; Kertész, J. Detecting the overlapping and hierarchical community structure in complex networks. *New J. Phys.* **2009**, *11*, 033015. [\[CrossRef\]](#)
40. Davis, A.; Gardner, B.B.; Gardner, M.R.; Pierson, D. *Deep South: A Social Anthropological Study of Caste and Class*; University of South Carolina Press: Columbia, SC, USA, 1941; Volume 2, pp. 117–118.
41. Club Membership Network Dataset-KONECT. Available online: http://www.konect.cc/networks/brunson_club-membership/ (accessed on 29 March 2021).
42. Barnes, R.; Burkett, T. Structural redundancy and multiplicity in corporate networks. *Int. Netw. Soc. Netw. Anal.* **2010**, *30*, 4–20.
43. American Revolution Network Dataset. Available online: http://konect.cc/networks/brunson_revolution/ (accessed on 29 March 2021).
44. Divorce in US. Available online: <http://vlado.fmf.uni-lj.si/pub/networks/data/2mode/divorce.net> (accessed on 29 March 2021).

-
45. Imrich, W.; Klavžar, S.; Hammack, R.H. *Product Graphs: Structure and Recognition*; JohnWiley & Sons: Hoboken, NJ, USA, 2000. Available online: <http://2017drama-st.tvbs.com.tw/4rcw/15-dr-christophe-schumm/read-product-graphs-structure-and-recognition-0471370398.pdf> (accessed on 29 March 2021).
 46. Crime Network Dataset—KONECT. Available online: http://konect.cc/networks/moreno_crime/ (accessed on 29 March 2021).
 47. Li, H.J.; Xiang, J. Explore of the fuzzy community structure integrating the directed line graph and likelihood optimization. *J. Intell. Fuzzy Syst.* **2017**, *32*, 4503–4511. [[CrossRef](#)]
 48. Nacher, J.C.; Jean-Marc, S.; John, P. Modularity in Protein Complex and Drug Interactions Reveals New Polypharmacological Properties. *PLoS ONE* **2012**, *7*, e30028. [[CrossRef](#)]
 49. Newman, M.E.J. *The Structure of Scientific Collaboration Networks*; Princeton University Press: Princeton, NJ, USA, 2011; pp. 221–226. [[CrossRef](#)]