



Community Discovery Algorithm Based on Multi-Relationship Embedding

Dongming Chen, Mingshuo Nie, Jie Wang and Dongqi Wang*

Software College, Northeastern University, Shenyang, 110169, China

*Corresponding Author: Dongqi Wang. Email: wangdq@swc.neu.edu.cn

Received: 23 August 2022; Accepted: 21 November 2022

Abstract: Complex systems in the real world often can be modeled as network structures, and community discovery algorithms for complex networks enable researchers to understand the internal structure and implicit information of networks. Existing community discovery algorithms are usually designed for single-layer networks or single-interaction relationships and do not consider the attribute information of nodes. However, many real-world networks consist of multiple types of nodes and edges, and there may be rich semantic information on nodes and edges. The methods for single-layer networks cannot effectively tackle multi-layer information, multi-relationship information, and attribute information. This paper proposes a community discovery algorithm based on multi-relationship embedding. The proposed algorithm first models the nodes in the network to obtain the embedding matrix for each node relationship type and generates the node embedding matrix for each specific relationship type in the network by node encoder. The node embedding matrix is provided as input for aggregating the node embedding matrix of each specific relationship type using a Graph Convolutional Network (GCN) to obtain the final node embedding matrix. This strategy allows capturing of rich structural and attributes information in multi-relational networks. Experiments were conducted on different datasets with baselines, and the results show that the proposed algorithm obtains significant performance improvement in community discovery, node clustering, and similarity search tasks, and compared to the baseline with the best performance, the proposed algorithm achieves an average improvement of 3.1% on Macro-F1 and 4.7% on Micro-F1, which proves the effectiveness of the proposed algorithm.

Keywords: Network representation learning; multi-relationship; node encoder; attribute information

1 Introduction

The analysis of multilayer networks is crucial in the research of complex networks. Network structures that are modeled based on real-world complex systems can usually be represented using multilayer networks, i.e., the interaction processes between node entities in a network structure can be described as many different types of relationships, and the interaction processes of these different relationship types can be represented using multilayer networks that represent each of these network layers as one interaction type. These different



This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

interaction types constitute a multilayer network that can more accurately describe complex systems in the real world. Multipath networks [1], multilevel networks [2], interdependent networks [3], and multidimensional networks [4] can usually be considered multilayer networks [5].

The information mining for multilayer networks is based on the research of single-layer networks, and the information mining for multilayer networks can further mine the implicit information in multilayer networks by fully considering the complex relationships between different network layers. The communities in the multi-layer networks consist of a set of strong connections of nodes in the layer. Community structures appear frequently in many real-world complex systems, such as sociological, biological, and transportation systems [6]. The community discovery algorithms in multilayer network information mining tasks can identify the community structure in the multi-layer network and thus determine the set of strong connection nodes. Although nodes with different relationship types can form various single-layer networks independently, there are usually correlations among these single-layer networks. In citation networks, references vary from paper to paper, but researchers can usually infer the research topic of a paper by analyzing the published topics of the paper's authors, thus allowing the prediction of the paper's topic. Furthermore, the nodes in the network can also contain some attribute information, which plays a key role in many domains. For example, the abstract information of a paper in a citation network can help infer the topic of the paper. Therefore, network representation learning should consider not only the structural information of nodes but also the attribute information of nodes.

Recently, most such studies have been based on homogeneous networks, using network models with a single relationship between nodes and achieving certain results, but ignoring the information implied in the network structure [7]. Furthermore, existing researches seldom discuss the attribute embedding of nodes and generally has weak global feature modeling capabilities. The fact that networks exist in the real world is inherently multiple, i.e., nodes have multiple types of relationships. In the citation network, papers have multiple types of relationships with each other, including authorship relationships between papers and citation relationships between papers. In movie networks, different movies have the same director or use the same actors, thus creating connections between different movies. At present, the research for multilayer network representation learning usually focuses on the fusion of network layers without considering the nodes' attribute features themselves. The use of label information of nodes also poses a challenge in terms of algorithmic complexity. Furthermore, traditional community discovery methods in topology-based multilayer network analysis require complete alignment of nodes between layers, and the complexity of the algorithms is usually high.

In this paper, we propose the Community Discovery Algorithm Based on Multi-relationship Embedding (CDBMRE). This paper uses a network representation learning method based on a Graph Convolutional Network (GCN) to solve the problem of community discovery in multilayer networks. The network representation learning method can obtain low-dimensional vector representations of nodes from the network, preserving their node attribute feature information while preserving the node structure information in the network. First, the nodes in the network are modeled to obtain the embedding matrix $h^{(r)}$ of the nodes. Then, the encoder g^r generates the node embedding matrix $H^{(r)}$ for each specific relation type in the network $G^{(r)}$. Finally, the node embedding matrix $H^{(r)}$ for each specific relation type is integrated using GCN to obtain multiple nodes embedding matrices $S^{(r)}$ for specific relation types. Experimental results on real-world network datasets show that the proposed algorithm has a high accuracy of community discovery.

Overall, the main contributions of the proposed algorithm are as follows.

- 1) The proposed algorithm overcomes the limitations of single-layer networks and fuses the attribute information of nodes into the learning process of network embedding.

- 2) The proposed algorithm enables the creation of specific embedding matrices for nodes and node relationship types and aggregates the node embedding matrix for each specific relationship type with GCN. This strategy allows capturing of rich structural and attributes information in multi-relational networks.
- 3) The proposed algorithm obtains significant performance improvement in community discovery, node clustering, and similarity search tasks, which proves the effectiveness of the proposed algorithm.

2 Related Work

Community detection is one of the most active topics in the field of graph mining and network science [8], where the community structure can represent the implicit structure in the network [9]. Community discovery algorithms can find the most reasonable division of communities in a network by the observed topology, thus providing support for researchers to analyze the network topology. Traditional community discovery algorithms based on network topology analysis include Girvan-Newman (GN) algorithm [10], Fast-Newman (FN) algorithm [11], Louvain algorithm [12], etc.

The random walk-based network representation learning method [13] and the graph convolutional network-based network representation learning method [14] are both able to learn node embeddings in the network by aggregating the neighborhood features of nodes, and by extracting the structural information of the network and mapping the nodes in the network to a low-dimensional continuous space while maintaining the structural features of the network and the node attribute information, to achieve community discovery using node embeddings. Information Fusion in Multilayer Network Embedding (IFMNE) [15] is a method of random walk of multiple information, this method learns node information and multi-type relational information into a unified space, and combines node structure information with network topology information. The accuracy of this method has significant advantages for the link prediction task on multilayer networks. Peng et al. [16] propose a reinforced neighborhood selection-guided multi-relational graph neural network architecture that is capable of guiding the creation of multi-relational graphs with a reinforcement learning method and maintaining relational dependencies in the learning process. The method enables the learning of more discriminative node embeddings with enhanced interpretability. Furthermore, Nie et al. [17] have summarized several solutions to graph structure data mining with reinforcement learning methods. The method proposed by Pourhabibi et al. [18] can capture the position of a given node in a set of neighboring anchor sets and the type of connections between nodes in the anchor sets to learn the structural representation of nodes and relations simultaneously. The method has better community discovery performance on a terrorist dataset and a synthetic network. Analyzing and mining useful knowledge in networks has always been studied by researchers, an approach to learning network embeddings of low-dimensional vector representations of nodes in networks can effectively solve various network tasks among various network mining techniques.

3 Preliminaries

Network representation learning: The network representation learning methods aim to learn a low-dimensional vector representation of the nodes in the graph while preserving the structure of the network as well as various other properties, such as information about the attributes of the nodes, structural roles, and labeling of the nodes. Given an attribute network $G = (V, E, X)$ and a set of adjacency matrices A , the task of network embedding is to learn a d -dimensional vector representation $z_i \in Z \in R^{n \times d}$ for each node $v_i \in V$ without using any labels of the nodes.

Multi-relational networks: Multi-relational networks are networks $G^{|r|} = (V, E, X)$, where $G^r = (V, E^{(r)}, X)$ represents a graph with relation type $r \in R$, node v is the set of n nodes,

$E = \cup_{r \in R} E^{(r)} \subseteq v \times v$ is the set of all edges of relation type $r \in R$. $X \in \mathbb{R}^{n \times f}$ is a matrix encoding the node attribute information of n nodes. $|R| > 1$ indicates a multilayer network. $|R| = 1$ indicates a single-layer network. For a given network G , $A = (A^{(1)}, A^{(2)}, \dots, A^{(|R|)})$ is a set of adjacency matrices, where $A^{(r)} \in \{0, 1\}^{|V| \times |V|}$ is the adjacency matrix of network G .

4 Methodology

Network representation learning is a method of converting a high-dimensional density matrix into a low-dimensional dense vector. Using the method of network representation learning can reduce the dimensionality of the structural information in the network while preserving the structural information of the nodes in the network. The CDBMRE algorithm proposed in this paper combines multiple types of relationship embeddings between nodes in the network and considers multiple relationship types between node attributes along with network structure information. The overall framework of the CDBMRE algorithm is shown in Fig. 1.

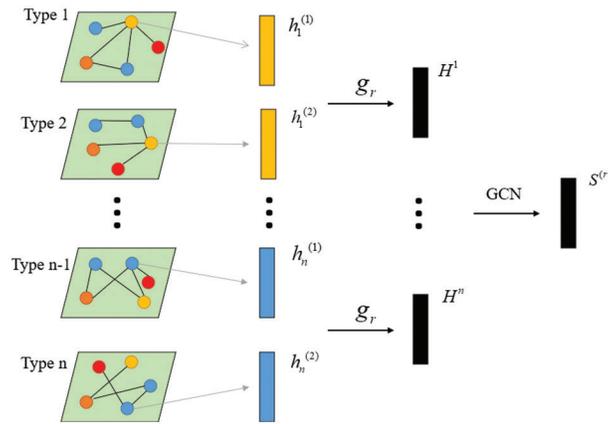


Figure 1: The framework of the CDBMRE algorithm

The algorithm proposed uses GCN to capture the potential network structure information and attribute information in multilayer networks, where GCN considers the information of higher-order neighbor nodes in the network when learning the embedding representation, and the network sparsity problem is solved by using the GCN method to obtain the structure information and node attribute information in multilayer networks. The convolution method of GCN is shown in Eq. (1).

$$H^{(l+1)} = f(H^{(l)}, A|W^{(l)}) \quad (1)$$

where $H^{(l)}$ is the input of layer l and $H^{(l+1)}$ is the output of layer l . The network G is with n nodes, each node has its features, and the features of these nodes form the $n \times d$ -dimensional matrix X . The relationship between each node also forms the $n \times n$ -dimensional adjacency matrix A . GCN is a neural network layer, and the propagation between different layers is shown in Eq. (2).

$$H^{(l+1)} = \sigma \left(\hat{D}^{-\frac{1}{2}} \hat{A} \hat{D}^{-\frac{1}{2}} H^{(l)} W^{(l)} \right) \quad (2)$$

where $\hat{A} = A + I$ and I is a unit matrix.

The real-world networks usually contain topological structure information and attribute information of the nodes. In this paper, the node attribute information of nodes is used in the representation of the relationship types of learning nodes to obtain the node embedding matrix $h^{(r)}$ for each node relationship type $r \in R$. Specifically, this paper defines the interaction between different types of nodes based on the idea of meta-paths, using a meta-path [19] obtained from the modeling of the network containing the node relationship types processes.

The meta-path is defined on the network as $T_G = (A, R)$, i.e., $\{A_{R_1}^1, A_{R_2}^2, \dots, A_{R_l}^l, A^{(l+1)}\}$, that is, between node types $A_1, A_{(l+1)}$ defines a combinatorial relation $R = \{R_1 \cdot R_2 \cdot \dots \cdot R_l\}$, where $\{\cdot\}$ represents the combinatorial operation between relations. For example, there are six types of movie information network: User, Group, Movie, Director, Actor, and Type. The director can be connected by the meta-path “Movie-Director-Movie (M-D-M)”, and also by the meta-path “Actor-Movie-Director-Movie-Actor (A-M-D-M-A)”. The different meta-paths have different meanings: M-D-M means that the director directed two movies with similar styles, while A-M-D-M-A means that different actors acted in the same movie directed by the same director. In addition, in the citation network, the relationship type between Author (A) and Paper (P) can be expressed as P-A-P, and the P-A-P relationship can be that the same author published two papers. The embedding matrix for a specific node relationship type is shown schematically in Fig. 2.

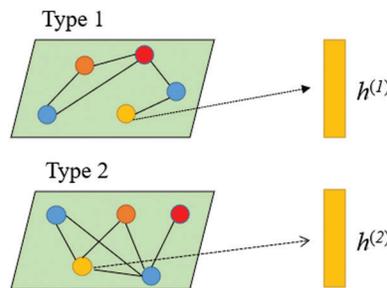


Figure 2: Node embedding matrix $h^{(r)}$

The algorithm proposed in this paper first models the nodes and obtains the node embedding matrix for each node in the network. Then a node encoder $g_r: R^{n \times f} \rightarrow R^{n \times n} \rightarrow R^{n \times d}$ is introduced for a node-specific relationship type, and the purpose of using the node encoder g_r will be to obtain the node embedding matrix $h^{(r)}$ for each node. The node embedding matrix $H^{(r)}$ is generated for each specific relation type in a network G by using the node encoder g_r . The computational procedure for generating the node embedding matrix for each specific relationship type using the node encoder g_r is shown in Fig. 3.

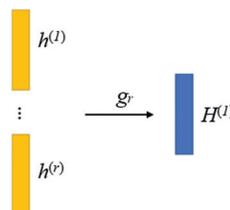


Figure 3: Embedding matrix $H^{(r)}$

The node embedding matrix $H^{(r)}$ for each specific relation type is taken as input and the node embedding matrix for multiple specific relation types is integrated by using the method of GCN with $s^{(r)}$ as output, where GCN is calculated by integrating the node embedding matrix $H^{(r)}$ for each specific relation type by Eq. (3).

$$H^{(r)} = g_r(X, A^{(r)}|W^{(r)}) = \sigma\left(\hat{D}_r^{-\frac{1}{2}}\hat{A}^{(r)}\hat{D}_r^{-\frac{1}{2}}XW^{(r)}\right) \quad (3)$$

where $\hat{A}^{(r)} = A^{(r)} + wI_n$, $\hat{D}_{ii} = \sum_j \hat{A}_{ij}$. $W^{(r)} \in \mathbb{R}^{f \times d}$ is the trainable weight matrix of decoder g_r for each node-specific relation type, σ is a nonlinear activation function. In $\hat{A}^{(r)} = A^{(r)} + wI_n$, the weights $w \in R$ are introduced to control the weights of the connected edges between nodes, and when the weights w are larger, it indicates that the nodes themselves play a more important role in generating their embeddings in the network, but this reduces the importance of the nodes with their neighboring nodes.

In this paper, we use $s^{(r)}$ to represent the network $G^{(r)}$ using GCN to integrate the node embedding matrix $H^{(r)}$ for each specific relationship type after obtaining the embedding matrix for multiple specific relationship types between nodes using a readout function *Readout*: $\mathbb{R}^{n \times d} \rightarrow \mathbb{R}^d$ is represented as $s^{(r)}$.

$$s^{(r)} = \text{Readout}\left(H^{(r)}\right) = \sigma\left(\frac{1}{n} \sum_{i=1}^n h_i^{(r)}\right) \quad (4)$$

where σ is a nonlinear activation function and $h_i^{(r)}$ denotes the i -th row vector of the matrix $H^{(r)}$. The cross-entropy of node-specific relationship types is calculated based on the node-embedding matrix $H^{(r)}$ and $s^{(r)}$ for a given node-specific relationship type.

$$\mathcal{L}^{(r)} = \sum_{v_i \in v} \log D\left(h_i^{(r)}, s^{(r)}\right) + \sum_{j=1}^n \log\left(1 - D\left(\tilde{h}_i^{(r)}, s^{(r)}\right)\right) \quad (5)$$

In this paper, when using GCN for community discovery tasks, we found that the inability to discern the importance of nodes among the many neighbors of the target node led to a large amount of noisy data. For example, in the citation network, the relationship type between the author and the paper is P-A-P, while the relationship type between the paper and subject (S) is P-S-P. Although these two relationship types are similar and both have the attribute paper, P-A-P and P-S-P belong to two completely different node relationship types. To distinguish such relationship types which are similar but different in reality, this paper uses the method of attention mechanism to distinguish these different kinds of types of node-specific relationship types, and uses $a_i^{(r)}$ to denote the importance of relationship type r in generating the final vector representation of node v_i .

Finally, the node embedding matrix $h^{(r)}$ of the node relation type $r \in R$ in the network is obtained, and the node embedding matrix $H^{(r)}$ of each specific relation type in the network $G^{(r)}$ is generated using the node encoder g_r . $s^{(r)}$ indicates that the network $G^{(r)}$ uses GCN to integrate the node embedding matrix $H^{(r)}$ of each node embedding matrix $H^{(r)}$ of a specific relation type using GCN to obtain the node embedding matrix $H^{(r)}$ of a specific relation type with global information in the network $G^{(r)}$. However, each matrix $H^{(r)}$ is trained independently for each $r \in R$. Therefore, these node embedding matrices $H^{(r)}$ contain only relevant attribute information about each relationship type and do not take advantage of the multilayer nature of the network. To address this one problem, this paper improves the quality of embeddings by jointly integrating embeddings from different relationship types using a consensus embedding matrix $Z \in \mathbb{R}^{n \times d}$ on which the consensus matrix allows the node embedding matrix $H^{(r)}$ for each specific relationship type to agree. Specifically, this paper uses a consensus regularization framework, which consists of a regularizer that minimizes the divergence between the original set of node embeddings, i.e., $\{\tilde{H}^{(r)}|r \in R\}$, and a consensus embedding matrix, and another regularizer that maximally damages the divergence between node embeddings, $\{H^{(r)}|r \in R\}$, which can be expressed as Eq. (6).

$$l_{cs} = \left[Z - Q\left(H^{(r)} | r \in R\right) \right]^2 - \left[Z - Q(H^{(r)} | r \in R) \right]^2 \quad (6)$$

where Q is an aggregation function that combines a set of node-embedding matrices $H^{(r)}$ of a particular relationship type from multiple relationship types into a single matrix H , i.e., $H \in \mathbb{R}^{n \times d}$. Q is a pooling method that can handle any permutation-invariant input. The loss function is shown in Eq. (7).

$$\mathfrak{S} = \sum_{r \in R} \mathcal{L}^{(r)} + \alpha l_{cs} + \beta \|\Theta\|^2 \quad (7)$$

The pseudo-code of the proposed algorithm is shown in Algorithm 1.

Algorithm 1: CDBMRE

Input: Network $G = (V, E, X)$.

Output: The embedding matrix Z for each node.

- 1) Node embedding matrix $H^{(r)}$ for node relation type $r \in R$.
 - 2) Node encoder g_r generates a node embedding matrix $H^{(r)}$ for each relation specific type in network $G^{(r)}$.
 - 3) GCN integrates the node embedding matrix $H^{(r)}$ for each relation-specific type.
 - 4) Multiple relation-specific types of node embedding matrices $s^{(r)}$.
 - 5) Embedding Matrix $Z \in \mathbb{R}^{n \times d}$.
-

The algorithm proposed in this paper first models the nodes in the network to obtain the node embedding matrix $h^{(r)}$ for each node relationship type $r \in R$. Based on the idea of meta-path, the node embedding matrix $H^{(r)}$ for each specific relationship type in the network $G^{(r)}$ is generated by the node encoder g_r . Embedding matrix $H^{(r)}$ as input, and using GCN to integrate the node embedding matrix $H^{(r)}$ of each specific relation type to get multiple node embedding matrices $s^{(r)}$ of specific relation types to learn the low-dimensional vector representation of nodes in the network.

5 Experiments

Graph structured data mining methods can be applied in several real applications. In this paper, we use the aforementioned Association for Computing Machinery (ACM)¹, Internet Movie Database (IMDB)², and Digital Bibliography & Library Project (DBLP)³ datasets on the community discovery task, and use Macro-F1 [20] and Micro-F1 [20] as evaluation metrics for the experiments, the values of Macro-F1 and Micro-F1 provide the worst value when they are close to 0 and the best value when they are close to 1. The Deepwalk algorithm [21], Node2vec algorithm [13], Deep Graph Infomax (DGI) algorithm [22], and Multiplex Network Embedding (MNE) algorithm [23] are used as the comparison experimental algorithms. Deepwalk and Node2vec are popular graph embedding methods and are considered as baselines in many graph embedding methods. DGI and MNE are the latest multi-relational embedding methods and have better performance and efficiency. Our experiments include (1) Community Discovery and (2) Node Clustering and Similarity Search.

5.1 Datasets

In this paper, we conduct experiments on three multi-relational networks, and the details of the datasets used are shown in Table 1, where Relations (A-B) is the relationship between node A and node B. Num.A is

¹ <https://www.biendata.xyz/hgb/#/datasets>.

² <https://www.kaggle.com/datasets/karrimba/movie-metadatasv>.

³ <http://web.cs.ucla.edu/~yzsun/data/>.

the number of node A, Num.B is the number of node B, Num.A-B is the number of connected edges between node A and node B, and Relation type is the node-specific relationship type.

Table 1: Experimental dataset

Dataset	Relations (A-B)	Num.A	Num.B	Num.A-B	Relation type
ACM	Paper-Author	3025	5835	9744	P-A-P
	Paper-Subject	3025	56	3025	P-S-P
IMDB	Movie-Actor	3550	4441	10650	M-A-M
	Movie-Director	3550	1726	3550	M-D-M
DBLP	Paper-Author	7907	1960	14238	P-A-P
	Paper-Paper	7907	7907	10522	P-P-P
	Author-Term	1960	1975	57269	P-A-T-A-P

- DBLP: this is an integrated database system of computer-based English literature in the field of computing for the results of research with authors as the core, including published papers in international journals and conferences, etc.
- ACM: this dataset uses papers published in the International Conference on Knowledge Discovery and Data Mining (KDD), the International Conference on Special Interest Group on Management of Data (SIGMOD), the International Conference on Special Interest Group on Data Communication (SIGCOMM), the International Conference on Mobile Computing and Networking (MobiCOMM), and the International Conference on Very Large Data Bases (VLDB), and classifies papers into three categories (database, wireless communication, and data mining). It includes 3025 papers (P), 5835 authors (A), and 56 subjects (S).
- IMDB: the movies in this dataset are divided into three categories (action, comedy, drama) according to their genres, which contains 3550 movies (M), 4441 actors (A), and 1726 directors (D).

5.2 Experimental Results

5.2.1 Community Discovery

According to [Table 2](#), it can be found that the experimental results of the CDBMRE algorithm are better than those of the comparison algorithms on all three data sets, and the values of Macro-F1 and Micro-F1 reach the optimal results. Compared to the DGI algorithm, the CDBMRE algorithm has a smaller improvement in the case of using Macro-F1 and Micro-F1 as evaluation metrics for the experimental results.

Table 2: Experimental results

Metric	ACM		IMDB		DBLP	
	Macro-F1	Micro-F1	Macro-F1	Micro-F1	Macro-F1	Micro-F1
Deepwalk	0.739	0.748	0.532	0.550	0.533	0.537
Node2vec	0.741	0.749	0.533	0.548	0.543	0.547
DGI	0.881	0.883	0.598	0.606	0.723	0.720
MNE	0.792	0.797	0.552	0.574	0.566	0.562
<i>CDBMRE</i>	0.898	0.918	0.664	0.652	0.798	0.781

In the ACM dataset, the CDBMRE algorithm improves 1.9% in the Macro-F1 value and 1.6% in the Micro-F1 value over the DGI algorithm; in the IMDB dataset the CDBMRE algorithm improves 0.75% in the Micro-F1 value over the DGI algorithm. This is because the DGI algorithm relies on maximizing local mutual information, which aims to capture the local node characteristics of the global information of the network, while the CDBMRE algorithm proposed in this paper, firstly, obtains node-specific relationship type embeddings, and secondly, integrates multiple types of relationship embeddings between nodes in the network using encoders as well as network representation learning GCN, which considers the network structure information while also Multiple types of relationships between node attributes are considered, so the experimental results of the CDBMRE algorithm on the dataset are slightly better than the DGI algorithm.

Both the Deepwalk algorithm and Node2vec algorithm use the random walk to generate node sequences and use Skip-gram to learn node embeddings, however, they do not utilize the attribute information of the nodes in the network. Therefore, the Deepwalk and Node2vec algorithms are less effective in the experiments on the three datasets, with lower values of Macro-F1 and Micro-F1 than the CDBMRE algorithm. The MNE algorithm jointly models multiple networks by introducing a common embedding and additional embeddings for each node relationship type and does not utilize the attribute information of the nodes in the three datasets. experiments, its Macro-F1 and Micro-F1 values are lower than the algorithm proposed in this paper.

5.2.2 Node Clustering and Similarity Search

In this paper, we also design implementations to evaluate the performance of the CDBMRE algorithm based on node clustering and similarity search. In the node clustering task, the Normalized Mutual Information (NMI) [24] is used as an evaluation metric; in the node similarity search task, the cosine similarity score of the node embedding between all node pairs is calculated, and for each node, the nodes are ranked according to the similarity score, and the percentage of the top 5 nodes (Sim@5) that belong to the same class is calculated.

Experiments are done on the four comparison algorithms, Deepwalk, Node2vec, DGI, MNE, and the CDBMRE algorithm proposed in this paper on ACM, IMDB, and DBLP datasets. The algorithms proposed in this paper outperform the comparison algorithms in two evaluation metrics, NMI and Sim@5. The experimental results on the ACM, IMDB, and DBLP datasets are shown in Figs. 4–6.

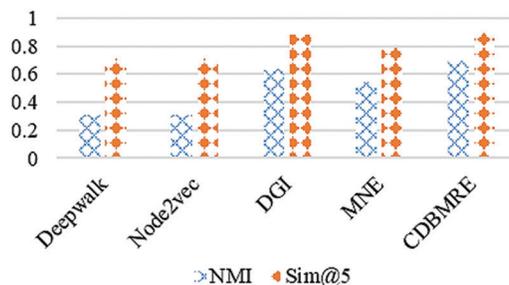


Figure 4: Experimental results on ACM

CDBMRE algorithm experiments on three datasets show that the values of NMI and Sim@5 are higher than those of the DGI algorithm. In the IMDB dataset, the NMI values obtained by the CDBMRE algorithm improved by 2.7% compared to the DGI algorithm, and in the DBLP dataset, it improved by 2%. The CDBMRE algorithm showed an excellent improvement in the NMI and Sim@5 values in all three

datasets compared to the Deepwalk algorithm, the Node2vec algorithm, and the MNE algorithm. The effectiveness of the CDBMRE algorithm is demonstrated.

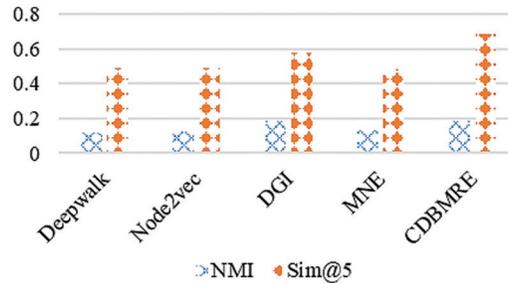


Figure 5: Experimental results on IMDB

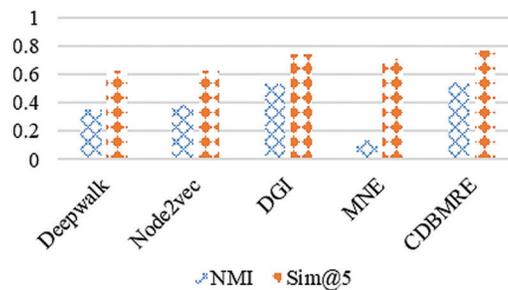


Figure 6: Experimental results on DBLP

Overall, the experimental results on ACM, IMDB, and DBLP datasets show that the CDBMRE algorithm has the best performance. The use of the network representation learning method can learn the low-dimensional vector representation of the nodes in the network, which preserves the node structure information in the network while also retaining the node attribute information, proving that the CDBMRE algorithm proposed in this chapter is effective.

6 Conclusion

Community discovery algorithms in multi-layer networks are an important direction in the research of complex networks. Community discovery algorithms based on multi-relational interaction information can mine the implicit structure and information of networks based on the interaction information between different types of nodes in multi-layer networks, and help researchers to further understand the formation process and community structure of networks. In reality, there are rich types of relationships between network nodes, and the nodes in the network not only have structural features, but also nodes have attribute information. To learn the consensus representation of a node, not only the structural information of the node, but also the attribute information of the node should be considered. In this paper, we use a network representation learning approach to solve the problem of community discovery in multilayer networks. There are multiple types of relationships between nodes in the network, and the nodes in the network are modeled to obtain node embeddings for each specific relationship type. This paper uses the GCN approach to integrate multiple types of relationship embeddings between nodes in the network. By experimenting with the comparison algorithms on three real network datasets, the results show that the

CDBMRE algorithm obtains better node embedding representations and achieves an average improvement of 3.1% on Macro-F1 and 4.7% on Micro-F1.

Funding Statement: This work was supported by the Key Technologies Research and Development Program of Liaoning Province in China under Grant 2021JH1/10400079 and the Fundamental Research Funds for the Central Universities under Grant 2217002.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

- [1] D. Zhang, Q. Liu, L. Chen, W. Xu and K. Wang, "Multi-layer based multi-path routing algorithm for maximizing spectrum availability," *Wireless Networks*, vol. 24, no. 3, pp. 897–909, 2018.
- [2] S. Aref, L. Dinh, R. Rezapour and J. Diesner, "Multilevel structural evaluation of signed directed social networks based on balance theory," *Scientific Reports*, vol. 10, no. 1, pp. 1–12, 2020.
- [3] S. V. Buldyrev, R. Parshani, G. Paul, H. E. Stanley and S. Havlin, "Catastrophic cascade of failures in interdependent networks," *Nature*, vol. 464, no. 7291, pp. 1025–1028, 2010.
- [4] M. Berlingerio, M. Coscia, F. Giannotti, A. Monreale and D. Pedreschi, "The pursuit of hubbiness: Analysis of hubs in large multidimensional networks," *Journal of Computational Science*, vol. 2, no. 3, pp. 223–237, 2011.
- [5] M. Yuvaraj, A. K. Dey, V. Lyubchich, Y. R. Gel and H. V. Poor, "Topological clustering of multilayer networks," *Proceedings of the National Academy of Sciences*, vol. 118, no. 21, pp. e2019994118, 2021.
- [6] X. Huang, D. Chen, T. Ren and D. Wang, "A survey of community detection methods in multilayer networks," *Data Mining and Knowledge Discovery*, vol. 35, no. 1, pp. 1–45, 2021.
- [7] X. Huang, "Research and implementation of community detection for multi-relational networks," M.S. Theses, Northeastern University, China, 2015.
- [8] D. Chen, M. Nie, J. Wang, Y. Kong, D. Wang *et al.*, "Community detection based on graph representation learning in evolutionary networks," *Applied Sciences*, vol. 11, no. 10, pp. 4497, 2021.
- [9] D. Chen, M. Nie, J. Yan, D. Wang and Q. Gan, "Network representation learning algorithm based on complete subgraph folding," *Mathematics*, vol. 10, no. 4, pp. 581, 2022.
- [10] M. E. J. Newman and M. Girvan, "Finding and evaluating community structure in networks," *Physical Review E*, vol. 69, no. 2, pp. 026113, 2004.
- [11] M. E. J. Newman, "Fast algorithm for detecting community structure in networks," *Physical Review E*, vol. 69, no. 6, pp. 066133, 2004.
- [12] V. D. Blondel, J. L. Guillaume, R. Lambiotte and E. Lefebvre, "Fast unfolding of communities in large networks," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2008, no. 10, pp. P10008, 2008.
- [13] A. Grover and J. Leskovec, "Node2vec: Scalable feature learning for networks," in *Proc. of ACM SIGKDD*, San Francisco, CA, USA, pp. 855–864, 2016.
- [14] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," arXiv Preprint, arXiv:1609.02907, 2016.
- [15] G. Yan, Z. Li, H. Luo, Y. Wang, W. Chang *et al.*, "Multilayer network representation learning method based on random walk of multiple information," *IEEE Access*, vol. 9, pp. 53178–53189, 2021.
- [16] H. Peng, R. Zhang, Y. Dou, R. Yang, J. Zhang *et al.*, "Reinforced neighborhood selection guided multi-relational graph neural networks," *ACM Transactions on Information Systems (TOIS)*, vol. 40, no. 4, pp. 1–46, 2021.
- [17] M. Nie, D. Chen and D. Wang, "Reinforcement learning on graphs: A survey," arXiv Preprint, arXiv:2204.06127, 2022.
- [18] T. Pourhabibi, K. L. Ong, Y. L. Boo and B. H. Kam, "Detecting covert communities in multi-layer networks: A network embedding approach," *Future Generation Computer Systems*, vol. 124, pp. 467–479, 2021.
- [19] D. Liu, L. Li and Z. Ma, "A community detection algorithm for heterogeneous information networks," *IEEE Access*, vol. 8, pp. 195655–195663, 2020.

- [20] M. Sokolova and G. Lapalme, “A systematic analysis of performance measures for classification tasks,” *Information Processing & Management*, vol. 45, no. 4, pp. 427–437, 2009.
- [21] B. Perozzi, R. Al-Rfou and S. Skiena, “Deepwalk: Online learning of social representations,” in *Proc. of ACM SIGKDD*, Sydney, AUS, pp. 701–710, 2014.
- [22] P. Velickovic, W. Fedus, W. L. Hamilton, P. Liò, Y. Bengio *et al.*, “Deep graph infomax,” in *Proc. of ICLR*, New Orleans, USA, pp. 4, 2019.
- [23] H. Zhang, L. Qiu, L. Yi and Y. Song, “Scalable multiplex network embedding,” in *Proc. of IJCAI*, Stockholm, SE, pp. 3082–3088, 2018.
- [24] K. Asmi, D. Lotfi and A. Abarda, “The greedy coupled-seeds expansion method for the overlapping community detection in social networks,” *Computing*, vol. 104, no. 2, pp. 295–313, 2022.