Network Representation Learning Algorithm Based on Community Folding

Dongming Chen, Mingshuo Nie, Jiarui Yan, Jiangnan Meng, Dongqi Wang*

Software College, Northeastern University, China
chendm@mail.neu.edu.cn, niemingshuo@stumail.neu.edu.cn, yjryanjiarui@163.com, 1901268@stu.neu.edu.cn, wangdq@swc.neu.edu.cn

Abstract

Network representation learning is a machine learning method that maps network topology and node information into low-dimensional vector space, which can reduce the temporal and spatial complexity of downstream network data mining such as node classification and graph clustering. This paper addresses the problem that neighborhood information-based network representation learning algorithm ignores the global topological information of the network. We propose the Network Representation Learning Algorithm Based on Community Folding (CF-NRL) considering the influence of community structure on the global topology of the network. Each community of the target network is regarded as a folding unit, the same network representation learning algorithm is used to learn the vector representation of the nodes on the folding network and the target network, then the vector representations are spliced correspondingly to obtain the final vector representation of the node. Experimental results show the excellent performance of the proposed algorithm.

Keywords: Network representation learning, Community detection, Network folding

1 Introduction

Network representation learning methods are becoming increasingly popular for capturing complex relationships, and these methods enable the capture of features of the network structure, but the existing network representation learning algorithms based on the random walk have certain limitations [1-4]. Among them, DeepWalk algorithm [5] and Node2Vec algorithm [6] estimate the vector representation of the destination node according to the set finite step size walk sequence. In addition, LINE algorithm [7] obtains the neighborhood information of nodes according to the similarity of nodes, but only considers second-order neighbors at most. The essence of this algorithm is to estimate the node representation by obtaining the neighborhood information of the destination node. This algorithm only considers the local topology information of the network and ignores the global topology information of the network.

To solve the problem of poor acquisition of the global topology information of the network by such algorithms. This paper considers that the network can be reduced according to certain fixed structures, and the network representation of the original network's higher-order structure can be obtained by employing this folding network. Then the representation of the higher-order network is fused with that of the original network to obtain a network representation that considers both local information and global information. Therefore, it is critical to find a structure that can effectively represent the global topology of the network.

The real-world networks are often featured as communities, and this structure can provide a new perspective, which can observe the whole network from the global perspective. Community structure has practical significance in the real-world network. In the social network, people in the community share the same interests. In an ecosystem network, a community may reflect subsystems of the ecosystem, etc.

Community structures, defined as groups of nodes that are more densely connected than with the rest of the network, are widely existed in many real-world complex systems, such as sociology, biology, transportation systems, and so on [8]. The community structure is obvious in the global topology of the network viewed from a global perspective. The distribution of community can represent the overall structure, it can be clearly seen which parts the network is roughly divided into, and the global topology of the network can be grasped macroscopically Therefore, it is feasible to use community structure information as a way to obtain global network information.

Therefore, this paper folds the network into a folding network with each community structure as a folding unit by finding the community structure in the network and employs DeepWalk algorithm, Node2Vec algorithm, and LINE algorithm to estimate the network representation of node representation based on node neighborhood information to learn the representation of the folding network. Then the representation of the folding network is fused with that of the original network to obtain a network representation that integrates global information and local information.

We employ a community detection algorithm based on node influence when obtaining community structure to obtain better community detection results as input to the folding network. Compared with the baselines, this proposed algorithm can obtain better community detection results and make the folding effect more optimal.

Overall, our paper makes the following contributions:

(1) We propose to leverage the structure of communities in networks to augment the performance of network representation learning.
(2) We employ a multiscale method of stitching network

*Corresponding Author: Dongqi Wang; E-mail: wangdq@swc.neu.edu.cn
DOI: 10.53106/160792642022032302020

representations to improve the effectiveness of network representation learning.

(3) Experimental results show that the proposed algorithm provides superior performance and is capable of effectively representing the original results and node information of the network.

2 Related work

Community detection has become the main method for finding out how the structure of a network relates to the behavior of a system [8-9]. As an efficient technique for revealing the underlying structure, community detection has been employed in many applications, such as searching for potential friends in social networks [10], recommendation of products for users [11-12], social opinion analysis [13], etc. Furthermore, more and more researchers are focusing on improving the performance of community detection algorithms with different methods. Liu et al. [14] propose a new Direction Optimizing Label Propagation Algorithm (DOLPA) that relies on the use of frontiers and alternates between label push and label pull operations to enhance the performance of the standard Label Propagation Algorithm (LPA). Cheng et al. [15] propose an efficient local expansion-based overlapping community detection algorithm using local-neighborhood information (OCLN) for scalability issues. Liao et al. [16] propose a novel gravitation-based algorithm (GBA) which is inspired by the theory of galaxy evolution based on Newton's law of universal gravitation to simulate the process of community evolution. Furthermore, several network representations learning methods employ deep learning models from other domains [17-20].

3 Method

3.1 Community Detection

Nodes in the network mutually influence each other and some nodes with greater influence usually have a higher influence on the neighboring nodes, which can form the community structure with them as the central node. Communities unavoidably cross and overlap with each other, and some nodes belong to more than one community at the same time. To which community these nodes belong, the multi-level neighbor information is considered as a reference in the community detection process.

For community detection, we first select influential nodes in the network as attractive nodes as attractive nodes through the node centrality index, and these nodes are used as the initial nodes of the community. Furthermore, the second-order neighborhood information is considered in the processing of overlapping nodes. Finally, the results of community detection are readjusted to get the most reasonable community detection results.

The pseudocode for the algorithm is shown as Algorithm 1. The overall steps of community detection are summarized as follows:

Step 1: Initialize network *G* according to the selected centrality index. Among them, centrality indexes are selected as Degree Centrality (DC), Betweenness Centrality (BC), Closeness centrality (CC), Subgraph Centrality (SC), and Eigenvector Centrality (EC).

Step 2: Select the node with the greatest influence in the undivided node set in the current network, take this node as the core node of the initial community, and divide the neighbor nodes of this node into the community. If overlapping nodes are encountered, mark the overlapping nodes according to the community with the largest number of community labels in the first-order neighbors and second-order neighbors of the overlapping nodes

Step 3: Repeat Step 2 until all nodes in the network have been divided.

Step 4: Perform traversal of the community results after division, and re-divide the nodes that are not reasonably divided. The community marker that appears the most times among all the neighborhood community markers of the unreasonable node is the new community marker of the node.

The purpose of designing the community detection algorithm is to obtain better community detection results compared with the classical community detection algorithm, which can serve as the input of the network representation learning algorithm based on community folding.

Aiming to evaluate the performance of community detection, we selected four network datasets with different types and granularity, including social networks, collaborative networks, and simulation networks, to evidence the effectiveness of community detection with different scales of data sets.

By selecting Clique Percolation Method (CPM) [21], Louvain [22], Local Fitness Maximization (LFM) [23], Expectation Maximization (EM) [24], and Label Propagation Algorithm (LPA) [25] algorithms to compare with the ICD algorithm proposed in this paper, the results found by the community are measured by the modularity index. The experimental results are shown in Figure 1.

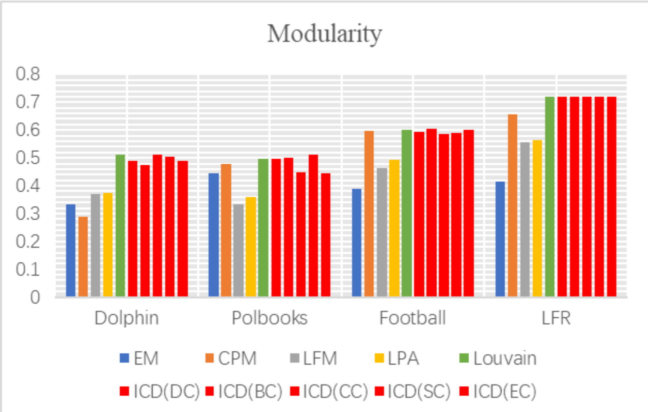

Figure 1. Comparison of modularity

We select five different metrics for the evaluation of the proposed algorithm and the modularity calculated based on the different metrics outperforms most traditional algorithms, which proves that this idea of community detection is available. In addition, for each dataset, one or two of the results corresponding to the five centrality indicators will have a value of modularity higher than that of all the comparison algorithms, so this algorithm has certain effectiveness.

The purpose of this algorithm is to obtain better community detection results through different node centrality indexes and obtain the community detection results of the group with the highest degree of modularity as algorithm input.

Algorithm 1. ICD

```
Input: network G
Output: Community detection results C
//Initialize network
I = {}; //Storage nodes and their influence
for node in V do
    I[node] = node.inf;
V_i = V; //Node combination for the Round i iteration
while V_i do
    node = max(I); //Node is currently the most influential node
    node.group = currentgroup;
    for i in node.neighnors do
        i.group = currentgroup;
    V_{i+1} = V_i .remove(C_i);
    if V_i == V_{i+1} then
        break;
    i++ ;
for i in V do
    neigh = [ ];
    if i.overlap == 1 then //overlapping nodes are processed
    for j in i.neighbors do
        neigh.append(j.group);
        for k in j.neighbors do
            neigh.append(k.group);
            i.group = max(count(neigh));
for i in V do //Unreasonable nodes are redivided
    neigh = [ ];
    for j in i.neighbors do
        neigh.append(j.group);
        i.group = max(count(neigh));
return C
```

3.2 Folding Process

The proposed community detection algorithm ICD algorithm is used to carry out community detection on the original network *G*. The original network is divided into N disjoint subsets, each subset represents a community, and each community is a folding unit. Add each folding cell in the network as a com node to the empty graph *M*. Each node in the folding network *M* corresponds to a community in the original network *G*, and the pair of connected nodes in the original network *G* is mapped to the folding network *M*. If the nodes between the communities in the original network *G* have edges, connect the com nodes in *M*. There is no regulation on the number of edges between communities, as long as there are edges in the original network *G*, the resulting folding network *M* is the community structure diagram of the original network *G*.

Algorithm 2. FLOD-C

```
Input: network G
Output: floding network M
Com = ICD(G)
M = nx.Graph(); //Create a new network
for com in Com do
    M.add_nodes(com)
for com_i in Com do
    for com_j in Com do
        for node_1 in com_i do
            for node_2 in com_j do
                if (com_i,com_j) in G.edges then
                    M.add_edges(com_i,com_j)
return M
```

The steps and pseudocode of the folding process are as follows:

Step 1: The ICD algorithm is used for community detection on the original network *G*, and the results of community detection are added to the Com set, where $Com=\{com_1,...,com_n\}$, and each com node represents a community.

Step 2: A new empty graph *M* is created to represent the folding network, and each element in the *Com* set is added to the empty graph *M*.

Step 3: Traverse the original network *G*. If there is a connection between node com_i and node com_j in the folding network *M* in the original network *G*, then a connection between node com_i and node com_j is added to the empty graph *M*.

3.3 Algorithm Process

The results of using community findings are obtained on the original network *G*, and each community is used as a folding unit to fold the network, and the folding network *M* is obtained. Using the same network representation learning algorithm for the original network *G* and the folding network *M* to obtain two sets of vectors, and the two sets of vectors are corresponding to each other. Then, the network representation learning algorithm CF-NRL based on community folding is proposed.

The overall flow of the algorithm is shown in Figure 2.

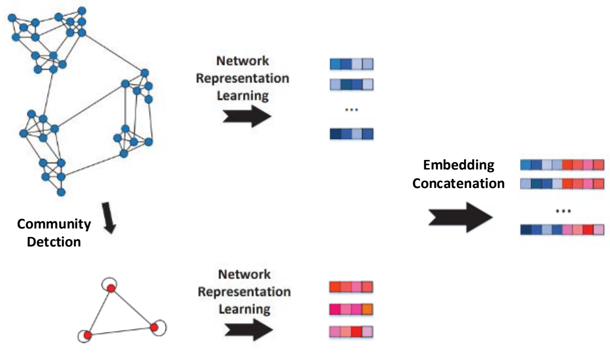

Figure 2. The flowchart of the CF-NRL algorithm

The original network is defined as *G* (*V*, *E*), where *V* represents the set of points and *E* represents the sum of edge sets. The ICD algorithm proposed in Section 4 is used to carry out community detection, and the community in the network is returned. The result of community detection is defined as a set $Com = \{com_1, ..., com_n\}$, where com_i represents a community. The algorithm steps are as follows:

Step 1: Divide the collection. The ICD algorithm is used to divide the nodes in the original network G into several disjoint subsets, which constitute the *Comset*.

Step 2: Fold the network. According to the folding rules, the network is folded into network M and stored in edgelist format.

Step 3: Represent the network. Network representation is learned on M and G respectively using the DeepWalk algorithm, Node2Vec algorithm, and LINE algorithm to obtain the *com_embedding* and *G_embedding*.

Step 4: Stitch vector. Corresponding to the original network G, the *com_embedding* and *G_embedding* are stitched together to obtain the embedding.

4 Experiments

To measure the performance of the CF-NRL algorithm, two experiments of community detection and multi-label classification are carried out here to illustrate.

4.1 Datasets

In the task of community detection, four datasets are used: Books About Us Politics [26], American College Football [27], Political Blogs [26], and Women [28]. Three datasets, namely, Books About Us Politics, American College Football, and American Political Blog Network, are used in the multi-label classification task.

PolBooks. Divide the politically related Books on Amazon into a network of liberal, centrist, and conservative Books by a political faction in the US. Nodes in the network represent political books, and edges between books represent buyers of both books.

Football. A National College Football tournament in which 115 teams play in 12 divisions. The nodes represent teams from different schools, and the edges represent two schools that have played a match.

PolBlogs. Records the relationships between links to US Political Blogs in 2005, with nodes in the network representing both liberal and conservative Political affiliations, some from the 2004 US presidential campaign.

Women: A dichotomy network in which nodes are divided into Women and activities, including 18 Women and 14 activities. It is collected in 1930 and represents the participation of Women in social activities in the Southern Club of Davis in the United States.

The basic information of the dataset is shown in Table 1, where ***Nodes*** represent the number of nodes, ***Edges*** represent the number of edges, **<*k*>** represents the average degree, and **<*d*>** represents the average shortest path length of the network.

Table 1. Datasets for verifying CF-NRL algorithm

Dataset	*Nodes*	*Edges*	<*k*>	<*d*>
Women	32	89	5.56	2.3
PolBooks	105	441	8.4	3.07
Football	115	613	10.66	2.5
PolBlogs	1224	16715	27.31	2.72

4.2 Community Detection

CF-NRL algorithm is used to characterize the network on the dataset with real labels, and *k*-means algorithm is used to cluster the results of the characterization, that is, community detection is carried out from the perspective of network representation. Using the NMI index to evaluate the results of community detection, the higher the Normalized Mutual Information (NMI) value is, the results of community detection are closer to the real communities, indicating higher quality network representation.

If the community labels of the target network are known, the performance of the community detection algorithm can be evaluated by comparing the community detection results with the true community attribution labels of the nodes. NMI is a metric commonly employed to evaluate the results of community detection experiments and is defined as:

$$NMI(A,B) = \frac{-2 \cdot \sum_{i=1}^{C_A} \sum_{j=1}^{C_B} N_{ij} \frac{N_{ij} N}{N_i N_j}}{\sum_{i=1}^{C_A} N_i \log \frac{C_i}{N} + \sum_{j=1}^{C_B} N_j \log \frac{C_j}{N}} \quad \textbf{(1)}$$

where A and B denote the truth community labels and the community detection results obtained using the algorithm, respectively, C_a and C_b denote the number of communities in A and B, respectively, N_{ij} denotes the elements in the confusion matrix, N_i and N_j denote the sum of the elements in row i and column j of the confusion matrix, respectively. n denotes the number of nodes in the network, i.e., the dimensions of the confusion matrix.

The experiment compares three classic network representation learning algorithms, DeepWalk, Node2Vec, and LINE, respectively. CF-NRL(DW) and DeepWalk, CF-NRL(N2V) and Node2Vec, and CF-NRL(LINE) and LINE are compared respectively, and the experimental environment and algorithm parameters of each group of comparison experiments are consistent. The experimental comparison results are shown in Table 2.

Table 2. Comparison of NMI value between CF-NRL algorithm and contrast algorithm

Methods	PolBooks	Football	PolBlogs	Women
DeepWalk	0.6066	0.9308	0.7500	0.1421
CF-NRL(DW)	**0.6213**	**0.9449**	**0.7639**	**0.257**
Node2Vec	0.6147	0.9314	0.7613	0.0985
CF-NRL(N2V)	**0.6216**	**0.9407**	**0.7725**	**0.1116**
Line	0.0754	0.3325	0.6367	0.054
CF-NRL(Line)	**0.4335**	**0.8848**	**0.7042**	**0.069**

It can be seen from the data in Table 2 that the NMI value of the CF-NRL algorithm is better than that of the comparison algorithm in all the comparison experiments. Among them, the improved range of the DeepWalk algorithm and Node2Vec algorithm based on random walk is limited, while the improvement effect of the LINE algorithm is more significant. The validity of CF-NRL algorithm is illustrated.

4.3 Multi-label Classification

In the multi-label classification task, three datasets, Polbooks, Football, and PolBlogs, are used. Some nodes are randomly selected as the training set and the rest nodes as the test set. The classification results were compared by calculating Macro-F1 value and Micro-F1 value. F1 value is calculated from recall rate and accuracy and is a common indicator to measure the quality of classification results.

The task compares three network representation learning algorithms, DeepWalk, Node2Vec, and LINE. CF-NRL (DW) and DeepWalk, CF-NRL (N2V) and Node2Vec, and CF-NRL (LINE) and LINE were compared respectively. 10%-90% data are selected as the training set for each group of experiments, and the F1 value is taken as the average value of 10 results.

The F1 value of the comparison experiment between CF-NRL (DW) and DeepWalk is shown in the form of a polyline graph, as shown in Figure 3. In the figure, the red polyline represents CF-NRL (DW) and the blue polyline represents DeepWalk. It can be found that CF-NRL (DW) algorithm is higher than DeepWalk algorithm in Football. PolBlogs dataset is slightly lower than the comparison algorithm only when the training data is 10%, while the remaining proportion of training data is significantly higher than the DeepWalk algorithm. PolBooks dataset is higher than the comparison algorithm when the proportion of training set is 10%-50%, and there is no significant difference between the two when the proportion of training set is 60%-90%. Overall, in this group of experiments, the CF-NRL(DW) algorithm is superior to the DeepWalk algorithm in the multi-label classification task.

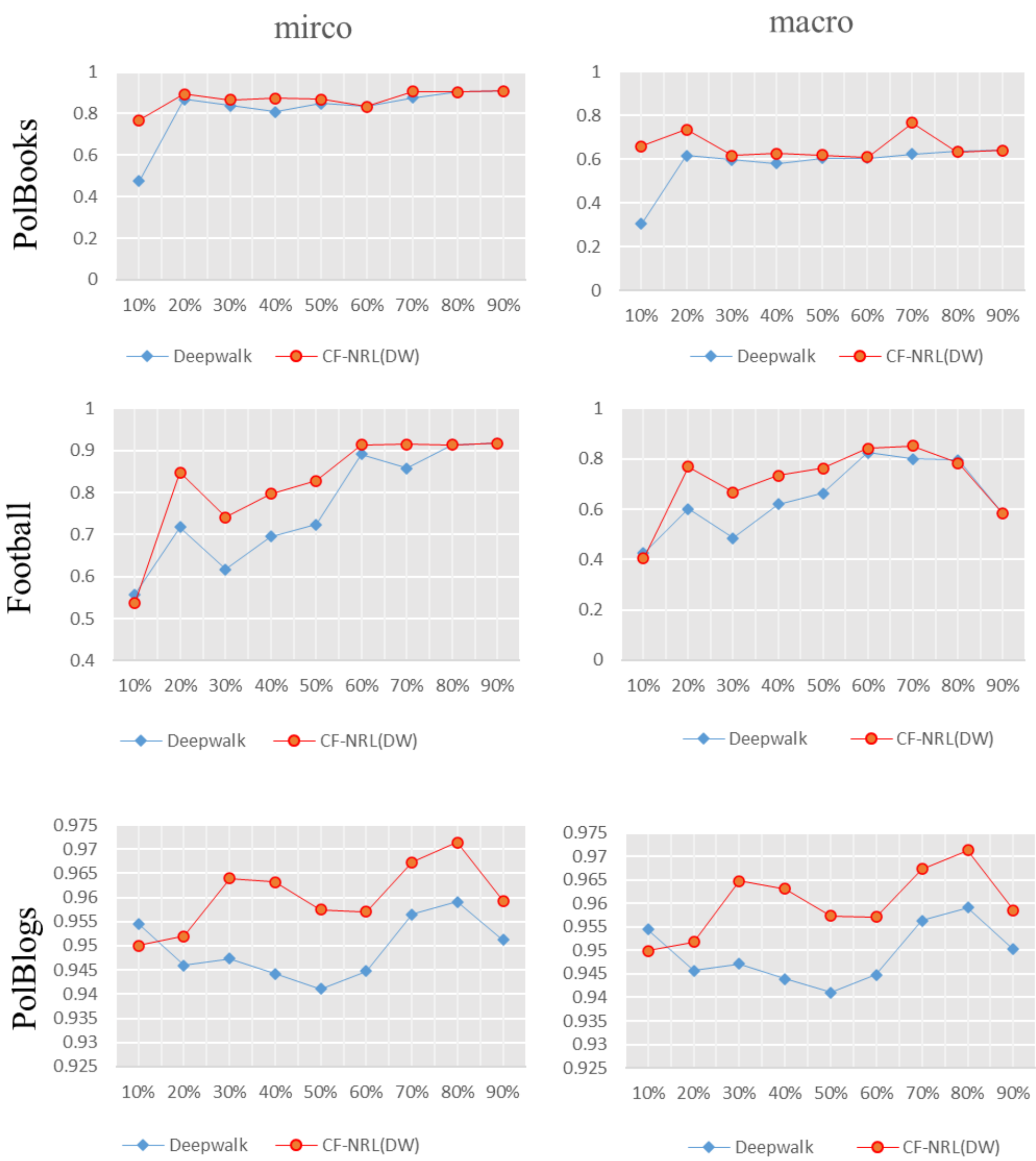

Figure 3. Comparison of F1 values of the multi-label classification experiment results of the CF-NRL (DW) algorithm and DeepWalk algorithm

The F1 value of the comparison experiment between CF-NRL (N2V) and Node2Vec is shown in Figure 4. In the figure, the red polyline is the F1 value of CF-NRL (N2V), and the blue polyline is the F1 value of Node2Vec. It is not difficult to see that CF-NRL (N2V) algorithm is higher than Node2Vec algorithm when the proportion of training set is 10%-90% on Football dataset. The improvement effect of the PolBooks dataset is not obvious, and the F1 value is greater than the comparison algorithm. PolBlogs dataset is slightly lower than the comparison algorithm only when the training data is 10% and 90%, while the remaining proportion of training data is significantly higher than Node2vec algorithm. This group of experiments can conclude that CF-NRL (N2V) algorithm is superior to Node2Vec algorithm in the multi-label classification task.

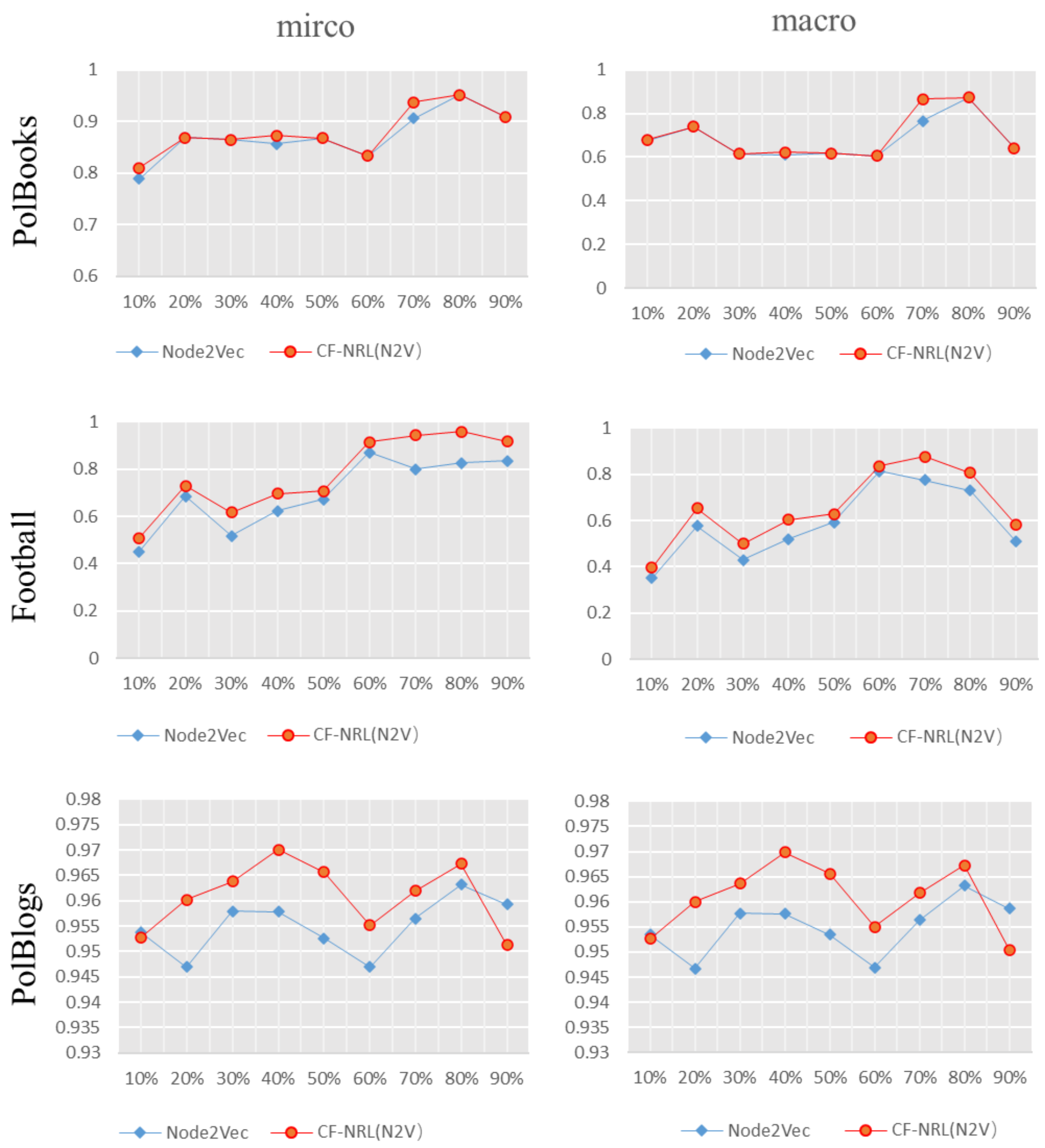

Figure 4. Comparison of F1 values of the multi-label classification experiment results of the CF-NRL (N2V) algorithm and Node2Vec algorithm

Comparison experiment of F1 value between CF-NRL (LINE) and LINE is shown in Figure 5 below. In the figure, the red polyline is the F1 value of CF-NRL (LINE), while the blue polyline is the F1 value of LINE. Compared with the previous two groups of experiments, it can be seen that CF-NRL (LINE) significantly improved Macro-F1 and Micro-F1 values of the three datasets in different proportions of training sets compared with the LINE algorithm. It can be said that the idea of this algorithm is more suitable for LINE algorithm compared with DeepWalk and Node2Vec.

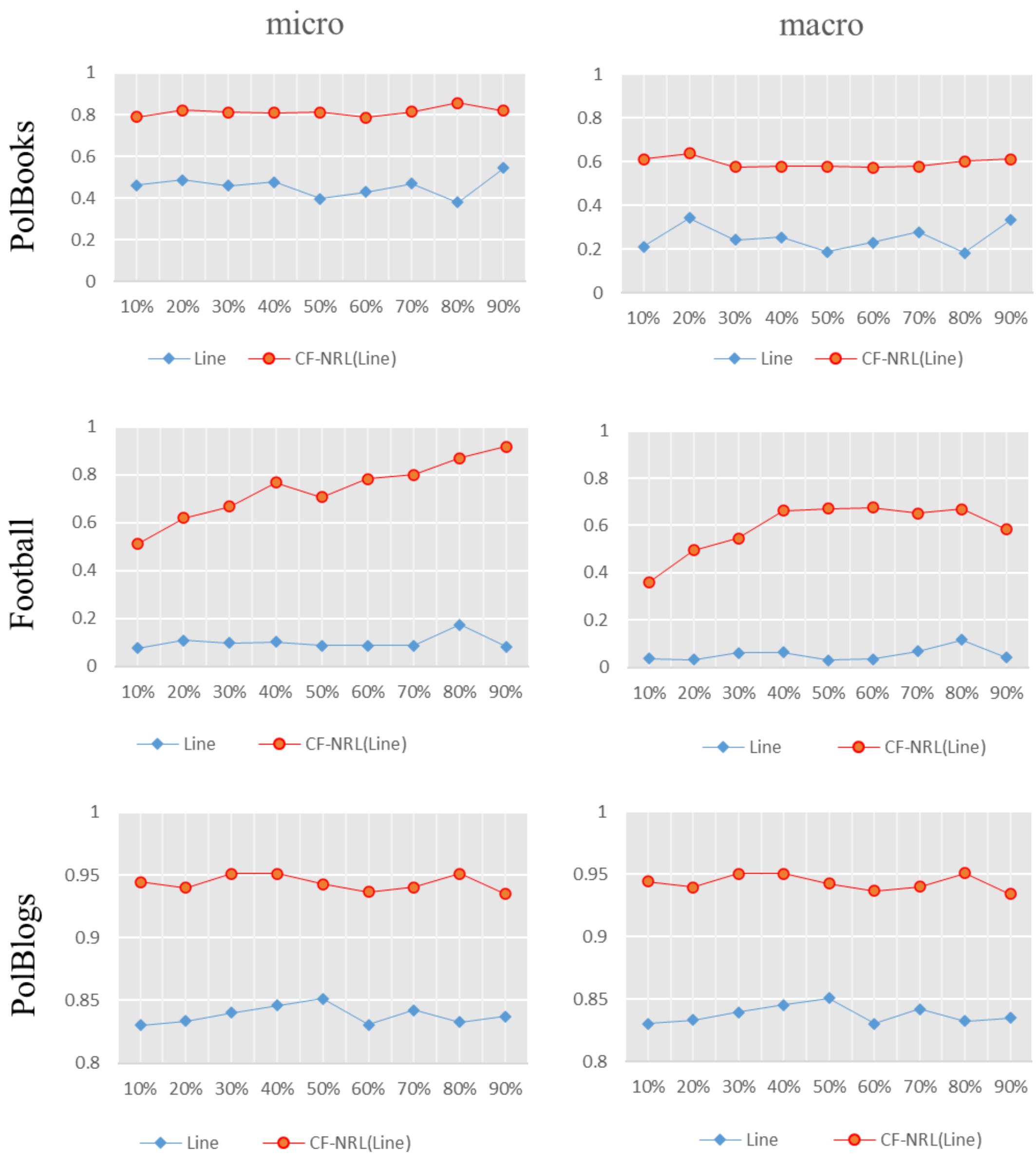

Figure 5. Comparison of F1 values of the multi-label classification experiment results of the CF-NRL (LINE) algorithm and LINE algorithm

5 Conclusion

The algorithm folds the original network into the smaller-scale networks by finding the complete subgraphs in the network and using the complete subgraphs as folding units to obtain the global topology information of the original network through this folding network. The original network and the collapsed network are learned separately for network representation, and the two sets of representations obtained are fused to obtain a network representation that considers both global and local information.

The algorithm is validated on community detection and multi-label classification tasks. In the community detection task, a binary network and three homogeneous networks are selected as experimental data, and three groups of comparative experiments are set up: CF-NRL (DW) and DeepWalk, CF-NRL (N2V) and Node2Vec, and CF-NRL (LINE) and LINE. The NMI values of the proposed algorithm are all higher than that of the contrast algorithm, and the LINE algorithm is greatly improved. Three datasets of different scales are selected in the multi-label classification task for comparison of F1 values, among which the improvement effect is significant between CF-NRL (LINE) and LINE, and the comparison between CF-NRL (DW) and DeepWalk, and CF-NRL (N2V) and Node2Vec also had certain advantages.

For future research, we will focus more on employing the proposed methods in the paper for heterogeneous and dynamic graphs to fit the complex and multiple network structures at present.

Acknowledgments

This paper is supported by the Key Technologies Research and Development Program of Liaoning Province, No. 2021JH1/10400079.

References

[1] J. S. Qi, X. Liang, Z. Y. Li, Y. F. Chen, Y. Xu, Representation Learning of Large-Scale Complex Information Network: Concepts, Methods and Challenges, *Chinese Journal of Computers*, Vol. 41, No. 10, pp. 2394-2420, October, 2018.

[2] B. Li, D. Pi, Network representation learning: a systematic literature review, *Neural Computing & Applications*, Vol. 32, No. 21, pp. 16647-16679, November, 2020.

[3] D. Zhang, J. Yin, X. Zhu, C. Zhang, Network Representation Learning: A Survey, *IEEE Transactions on Big Data*, Vol. 6, No. 1, pp. 3-28, March, 2020.

[4] Q. Zhou, C. Wang, Q. Li, AttrHIN: Network Representation Learning Method for Heterogeneous Information Network, *IEEE Access*, Vol. 9, pp. 127397-127406, September, 2021.

[5] B. Perozzi, R. Al-Rfou, S, Skiena, Deepwalk: online learning of social representations, *2014 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, New York, USA, 2014, pp. 701-710.

[6] A. Grover, J. Leskovec, Node2vec: scalable feature learning for networks, *2016 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Francisco, CA, USA, 2016, pp. 855-864.

[7] J. Tang, M. Qu, M. Wang, M. Zhang, J. Yan, Q. Mei, LINE: Large-scale Information Network Embedding, *24th International Conference on World Wide Web*, Florence, Italy, 2015, pp. 1067-1077.

[8] X. Huang, D. Chen, T Ren, D. Wang, A survey of community detection methods in multilayer networks, *Data Mining and Knowledge Discovery*, Vol. 35, No. 1, pp. 1-45, January, 2021.

[9] D. Chen, M. Nie, J. Wang, Y. Kong, D. Wang, X. Huang, Community Detection Based on Graph Representation Learning in Evolutionary Networks, *Applied Sciences*, Vol. 11, No. 10, Article No. 4497, May, 2021.

[10] J. Zhu, B. Wang, B. Wu, W. Zhang, Emotional community detection in social network, *IEICE Transactions on Information and Systems*, Vol. E100.D, No. 10, pp. 2515-2525, October, 2017.

[11] C. Li, Y. Zhang, A personalized recommendation algorithm based on large-scale real micro-blog data, *Neural Computing & Applications*, No. 32, No. 15, pp. 11245-11252, August, 2020.

[12] D. Wang, M. Nie, D. Chen, L. Wan, X. Huang, Node Similarity Index and Community Identification in Bipartite Networks, *Journal of Internet Technology*, Vol. 22, No. 3, pp. 673-684, May, 2021.

[13] D. Wang, J. Li, K. Xu, Y. Wu, Sentiment community detection: exploring sentiments and relationships in social networks, *Electronic Commerce Research*, Vol. 17, No. 1, pp. 103-132, March, 2017.

[14] X. T. Liu, M. Halappanavar, K. J. Barker, A. Lumsdaine, A. H. Gebremedhin, Direction-optimizing label propagation and its application to community detection, *the 17th ACM International Conference on Computing Frontiers (CF '20)*, Catania, Sicily, Italy, 2020, pp. 192-201.

[15] F. Cheng, C. Wang, X. Zhang, Y. Yang, A Local-Neighborhood Information Based Overlapping Community Detection Algorithm for Large-Scale Complex Networks, *IEEE/ACM Transactions on Networking*, Vol. 29, No. 2, pp. 543-556, April, 2021.

[16] M. Liao, Y. Liang, R. Wang, The Gravitation-based Algorithm for Community Detecting in Large-scale Social Networks, *3rd International Conference on Algorithms, Computing and Artificial Intelligence (ACAI 2020)*, Senya, China, 2020, Article No. 53, pp. 1-6.

[17] S. Bharati, P. Podder, M. R. H. Mondal, Hybrid deep learning for detecting lung diseases from X-ray images, *Informatics in Medicine Unlocked*, Vol. 20, Article No. 100391, 2020.

[18] M. R. H. Mondal, S. Bharati, P. Podder, Diagnosis of COVID-19 Using Machine Learning and Deep Learning: A Review, *Current Medical Imaging*, Vol. 17, No. 12, pp. 1403-1418, 2021.

[19] S. Bharati, P, Podder, M. R. H. Mondal, V. B. Surya Prasath, CO-ResNet: Optimized ResNet model for COVID-19 diagnosis from X-ray images, *International Journal of Hybrid Intelligent Systems*, Vol. 17, No. 1-2, pp. 71-85, July, 2021.

[20] A. Khamparia, S. Bharati, P. Podder, D. Gupta, A. Khanna, K. T. Phung, D. N. H. Thanh, Diagnosis of breast cancer based on modern mammography using hybrid transfer learning, *Multidimensional Systems and Signal Processing*, Vol. 32, No. 2, pp. 747-765, April, 2021.

[21] S. J. Martin, C. P. Reutelingsperger, A. J. McGahon, J. A. Radar, R. C. van Schie, D. M. LaFace, D. R. Green, Early redistribution of plasma membrane phosphatidylserine is a general feature of apoptosis regardless of the initiating stimulus: inhibition by overexpression of bcl-2 and abl, *Journal of Experimental Medicine*, Vol. 182, No. 5, pp. 1545-1556, November, 1995.

[22] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, E. Lefebvre, Fast unfolding of communities in large networks, *Journal of Statistical Mechanics: Theory and Experiment*, Vol. 2008, Article No. P10008, October, 2008.

[23] A. Lancichinetti, S. Fortunato, J. Kertész, Detecting the overlapping and hierarchical community structure in complex networks, *New Journal of Physics*, Vol. 11, No. 3, Article No. 033015, March, 2009.

[24] J. Rabiner, B. Juang, An Introduction to Hidden Markov Models, *IEEE ASSP Magazine*, Vol. 3, No. 1, pp. 4-16, January, 1986.

[25] U. N. Raghavan, R. Albert, S. Kumara, Near linear time algorithm to detect community structures in large-scale networks, *Physical Review E*, Vol. 76, No. 3 Pt 2, Article No. 036106, September, 2007.

[26] L. A. Adamic, N. Glance, The political blogosphere and the 2004 US Election, *KDD05: The Eleventh ACM SIGKDD International Conference on Knowledge*

Discovery and Data Mining, Chicago, Illinois, USA, 2005, pp. 36-43.
[27] M. Girvan, M. E. J. Newman, Community structure in social and biological networks, *the national academy of sciences*, Vol. 99, No. 12, pp. 7821-7826, June, 2002.
[28] A. Davis, B. B. Gardner, M. R. Gardner, *Deep South; A Social Anthropological Study of Caste and Class*, University of Chicago Press, 1941.

Biographies

Dongming Chen is a professor at Software College, Northeastern University, China. He is a member of ACM, IEEE Computer Society, senior member of China Computer Federation (CCF), and Senior member of China Institute of Communications (CIC). His research interests include complex networks, social network analysis, and manufacturing big data analysis.

Mingshuo Nie is a doctoral candidate in the Software College, Northeastern University. His research interests include graph neural networks, community detection, link prediction, and graph reinforcement learning.

Jiarui Yan graduated from Software College, Northeastern University, China. He was awarded a Master's degree in 2021. His research interests include graph embedding and community detection.

Jiangnan Meng is a master’s candidate in Software College, Northeastern University, China. Her research interests include complex networks, link prediction, social network analysis, and graph reinforcement learning.

Dongqi Wang is a lecturer in Software College, Northeastern University, China. He was awarded a Ph.D. degree in 2011. His research areas include network security and social network analysis. He is also a member of CCF and ACM.