STM JOURNALS

Recent Trends in
Electronics & Communication Systems

ISSN: 2393-8757
Volume 11, Issue 3, 2024
September–December
DOI (Journal): 10.37591/RTECS

https://journals.stmjournals.com/rtecs

Research RTECS

Hybrid Approach for Community Detection Using Deep Learning Techniques

S. Srividhya[1,*], Lavanya S.R.[2]

Abstract

Community detection in complex networks is a fundamental problem with applications across diverse domains, ranging from social networks to biological systems and beyond. Traditional methods based on graph theory have been widely used for identifying communities within networks. However, the intricate and evolving nature of modern networks demands more sophisticated approaches. This research work proposes a hybrid approach that combines the strengths of deep learning techniques with traditional community detection algorithms for improved accuracy and scalability. Leveraging the capabilities of deep learning models, specifically Graph Neural Networks (GNNs), the proposed approach aims to learn intricate patterns and representations from network data, encoding rich structural information into node embeddings. The first phase involves employing Graph Neural Networks, such as Graph Convolutional Networks (GCNs), to learn node representations that encapsulate complex relationships and local structures within the network. These learned embeddings capture both topological information and node attributes, enhancing the understanding of the network topology. Subsequently, the learned node embeddings are integrated into traditional community detection algorithms, such as spectral clustering to form a hybrid framework. This integration allows for the utilization of deep learning-based representations within established community detection methodologies, aiming to enhance the accuracy and robustness of community detection. The proposed hybrid approach is evaluated on various benchmark datasets and real-world networks to assess its effectiveness in detecting communities. Performance metrics such as modularity, conductance, and coverage are utilized to evaluate the quality of detected communities compared to standalone traditional or deep learning-based methods. The results demonstrate the efficacy of the hybrid approach, showcasing improved modularity, conductance, coverage and runtime compared to individual methods. Furthermore, the study discusses the significance of leveraging both deep learning techniques and traditional algorithms in community detection tasks.

Keywords: Community detection, deep learning, graph neural network, hybrid approach, graph convolution network

***Author for Correspondence**
S. Srividhya
E-mail: vidhyasai14@gmail.com

[1,2]Associate Professor, Department of Computer Science, K.P. Ramasamy College of Arts Science and Research, Coimbatore, Uthupalayam, Tamil Nadu, India

Received Date: July 23, 2024
Accepted Date: August 10, 2024
Published Date: September 18, 2024

Citation: S. Srividhya, Lavanya S.R. Hybrid Approach for Community Detection Using Deep Learning Techniques. Recent Trends in Electronics & Communication Systems. 2024; 11(3): 18–26p.

INTRODUCTION

Networks or graphs are prevalent in various fields, serving as representations of intricate relationships among entities, such as social interactions, biological systems, technological infrastructures, and more. Understanding the organizational structure and identifying cohesive groups or communities within these networks is crucial for comprehending their functionalities and dynamics [1]. Community detection, a fundamental problem in network analysis, involves partitioning nodes into meaningful groups where nodes within a group exhibit stronger connections among themselves compared to connections with nodes in other groups [2]. Traditional community detection methods have been pivotal in uncovering these structures, relying on graph-based metrics like modularity optimization [3], spectral clustering [4], and hierarchical clustering [5]. However, the evolving nature and complexity of modern networks demand approaches capable of capturing intricate patterns and representations from data to uncover underlying community structures more accurately and efficiently. Recent advancements in deep learning techniques, particularly Graph Neural Networks (GNNs) [6], have demonstrated remarkable potential in learning complex relationships and representations from graph-structured data. GNNs, including Graph Convolutional Networks (GCNs) [7], GraphSAGE [8], and Graph Attention Networks (GAT) [9], can learn node embeddings by aggregating information from neighbourhood nodes, encoding both topological structure and node attributes [10]. This study proposes a novel hybrid approach that integrates the strengths of deep learning techniques, particularly GNNs, with traditional community detection algorithms to enhance the accuracy and scalability of community detection in networks. By leveraging the learned node embeddings from GNNs, the hybrid framework aims to augment the capabilities of conventional community detection algorithms, thereby improving the detection of cohesive communities within complex networks. The primary objective of this research is to investigate how combining the expressive power of deep learning-based representations with established community detection algorithms can lead to more robust and accurate identification of communities in networks. The hybrid approach aims to strike a balance between the interpretability of traditional methods and the nuanced understanding of network structures achieved through deep learning techniques.

This paper is organized as follows: Next Section provides an overview of related work in community detection and deep learning techniques for networks. The Sections after this detail the methodology, outlining the hybrid approach integrating GNNs with traditional algorithms; present experimental results and evaluation metrics on benchmark datasets and real-world networks, and finally, discuss the findings, implications, and avenues for future research.

RELATED WORK

Understanding community structures within networks has been a subject of extensive research across various disciplines. Traditional methods for community detection primarily rely on graph-based metrics and optimization algorithms [1]. Modularity optimization [2], spectral clustering [3], and hierarchical clustering [4] are among the widely adopted techniques. These methods have provided valuable insights into network structures, yet they might face challenges in effectively capturing complex and evolving relationships within modern networks.

In recent years, the advent of deep learning techniques has sparked interest in leveraging neural network architectures for community detection tasks. Graph Neural Networks (GNNs) have emerged as powerful tools for learning representations from network data [5]. GNNs, including Graph Convolutional Networks (GCNs) [6], Graph SAGE [7], and Graph Attention Networks (GAT) [8], can effectively capture structural information by aggregating node features and interactions within the network topology [9]. Several studies have explored the use of GNNs for community detection tasks. Wang *et al*. employed GCNs to learn node embeddings and integrated them with a modularity-based objective function for detecting communities in social networks [11]. Similarly, a study by Sun et al. [12] proposed a Graph Attention Network (GAT) based approach for identifying communities, achieving competitive results on benchmark datasets.

Additionally, techniques such as Deep Walk [13] and Node2Vec [14] have demonstrated success in learning node embeddings by treating random walks on graphs as sequences and employing skip-gram models to capture node relationships. These methods, although not purely deep learning-based, have provided valuable insights into network structures and community detection.

Despite the advancements, the interpretability of results and the scalability of deep learning approaches for community detection remain important considerations. Traditional algorithms often offer clearer interpretability in identifying communities based on well-defined metrics like modularity. Balancing the strengths of both deep learning and traditional methods presents an opportunity to enhance the accuracy and robustness of community detection in networks.

This study aims to bridge the gap between deep learning techniques and traditional community detection methods by proposing a hybrid approach that integrates GNN-based representations with established algorithms. The subsequent section details the methodology of the proposed hybrid framework and its application in detecting communities within networks.

PROPOSED METHODOLOGY: HYBRID APPROACH FOR COMMUNITY DETECTION

The proposed methodology leverages the synergy between deep learning techniques, specifically Graph Neural Networks (GNNs), and traditional community detection algorithms to detect communities within networks effectively. The proposed methodology is shown in Figure 1. The hybrid framework comprises two main stages: Graph Representation Learning using GNNs and Integration with Traditional Community Detection Algorithms.

Graph Representation Learning with GNNs

In the first phase, the methodology employs Graph Neural Networks to learn informative node embeddings that capture structural and relational information from the network topology. Several GNN architectures such as Graph Convolutional Networks (GCNs) are used in this work. The GNNs operate by aggregating information from the local neighbourhoods of each node, enabling the encoding of both topological structure and node attributes into low-dimensional representations. This step aims to capture intricate relationships and hidden patterns within the network, facilitating a more comprehensive understanding of the underlying community structures.

The node embeddings obtained from the GNNs serve as rich feature representations, encapsulating the network's complex connectivity patterns and attributes. These embeddings encode the nodes' structural roles within the network, enabling a more nuanced representation compared to traditional feature engineering approaches. GNNs are a class of neural networks designed specifically to operate on graph-structured data. They process information by aggregating and propagating node features through the graph's connections, enabling the learning of representations that incorporate both local and global graph structures. Node embeddings refer to low-dimensional vector representations of nodes in a graph. These embeddings aim to capture the essential information about a node's structural role and relational context within the graph.

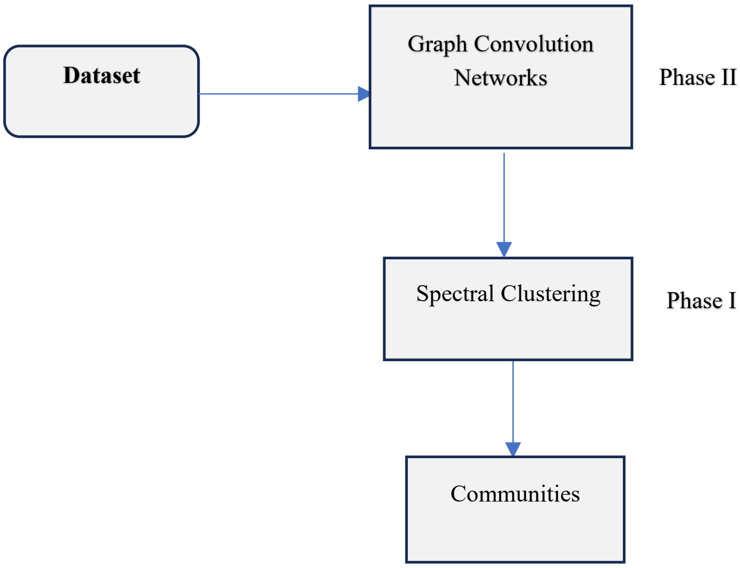

Figure 1. Proposed methodology block diagram.

The primary focus of this methodology is to leverage GNNs to learn from the underlying graph topology. Graph topology includes information about how nodes are connected to each other (through edges) and the overall structure of the graph. GCNs use a convolutional-like operation to aggregate information from a node's immediate neighbourhood, updating node embeddings through successive layers.

Integration with Traditional Community Detection Algorithms

In the second phase, the learned node embeddings from the GNNs are integrated into established community detection algorithms. Spectral clustering is augmented with the node embeddings to enhance their performance in identifying communities within the network. The integration process involves using the GNN-learned representations as input features or similarity measures within traditional algorithms. By leveraging the rich representations learned through deep learning, these algorithms can potentially overcome limitations related to sparsity or noise in the network, leading to more accurate and robust community detection.

After training GNNs on the graph data, node embeddings are obtained as representations for each node in the graph. These embeddings are low-dimensional vectors that encode structural and relational information learned by the GNNs. Spectral clustering is a popular technique used for partitioning a graph into clusters based on the eigenvectors (spectrum) of a similarity matrix derived from the graph. In the context of spectral clustering, the learned node embeddings from GNNs can be utilized as input features or representations instead of or in addition to traditional node attributes in the clustering process. Here is a step-by-step process for integration:

Constructing the Similarity Matrix

Typically, spectral clustering begins by constructing a similarity (or affinity) matrix representing pairwise similarities between nodes in the graph. This can be based on various metrics like cosine similarity, Euclidean distance, or a graph-specific measure.

Using Node Embeddings as Features

Instead of using raw node attributes or in combination with them, the learned node embeddings obtained from GNNs are employed as input features for constructing the similarity matrix. This involves using the embeddings to compute pairwise similarities between nodes, thereby influencing the affinity matrix.

Performing Spectral Embedding

The similarity matrix obtained using node embeddings is then processed using spectral techniques, such as calculating the graph Laplacian matrix and performing eigen-decomposition to extract the eigenvectors corresponding to the largest eigenvalues.

Clustering Based on Eigenvectors

The eigenvectors associated with the top eigenvalues are used to embed the nodes into a lower-dimensional space, and traditional clustering algorithms (like k-means) or methods specifically designed for spectral clustering are applied to these embeddings to partition nodes into communities or clusters.

Benefits of Using GNN Embeddings in Spectral Clustering

Integrating GNN-learned embeddings into spectral clustering can offer advantages by leveraging the rich representations captured by GNNs. These embeddings often encode structural information, capturing the complex relationships and roles of nodes in the graph, potentially leading to more meaningful and accurate clustering results compared to using raw node attributes alone.

The hybrid approach aims to harness the interpretability of traditional algorithms while benefiting from the expressive power of GNN-based representations. This amalgamation is expected to improve the overall efficacy of community detection by leveraging the complementary strengths of both paradigms.

EXPERIMENTAL RESULTS AND EVALUATION

The proposed hybrid approach for community detection, which integrates Graph Neural Networks (GNNs) with traditional algorithms, was empirically evaluated on benchmark datasets and real-world networks to assess its efficacy in identifying communities within networks. Zachary's karate club dataset, Dolphins, Football is used for this work. Zachary's karate club dataset is a sort of social network in which every node stands for a member of the club and the edges connect the nodes to indicate interactions that took place outside of the club. In this instance, the club members are divided into four separate categories. Zachery's Karate Club Graph is shown in Figure 2. Main objective is to classify each member into the appropriate group (node classification) based on the way they interact with each other.

Experimental Setup

The evaluation utilized various benchmark datasets representing diverse network structures, including social networks, biological networks, and citation networks. Additionally, real-world networks from domains such as online social platforms, co-authorship networks, and technological networks were employed to validate the proposed approach's performance in practical scenarios. Run time analysis is shown in Figure 3.

The hybrid approach was implemented using popular GNN architectures, including Graph Convolutional Networks (GCNs) combined with traditional community detection algorithms such as spectral clustering.

Evaluation Metrics

Performance metrics commonly employed in community detection tasks were utilized to assess the quality of detected communities. These metrics included:

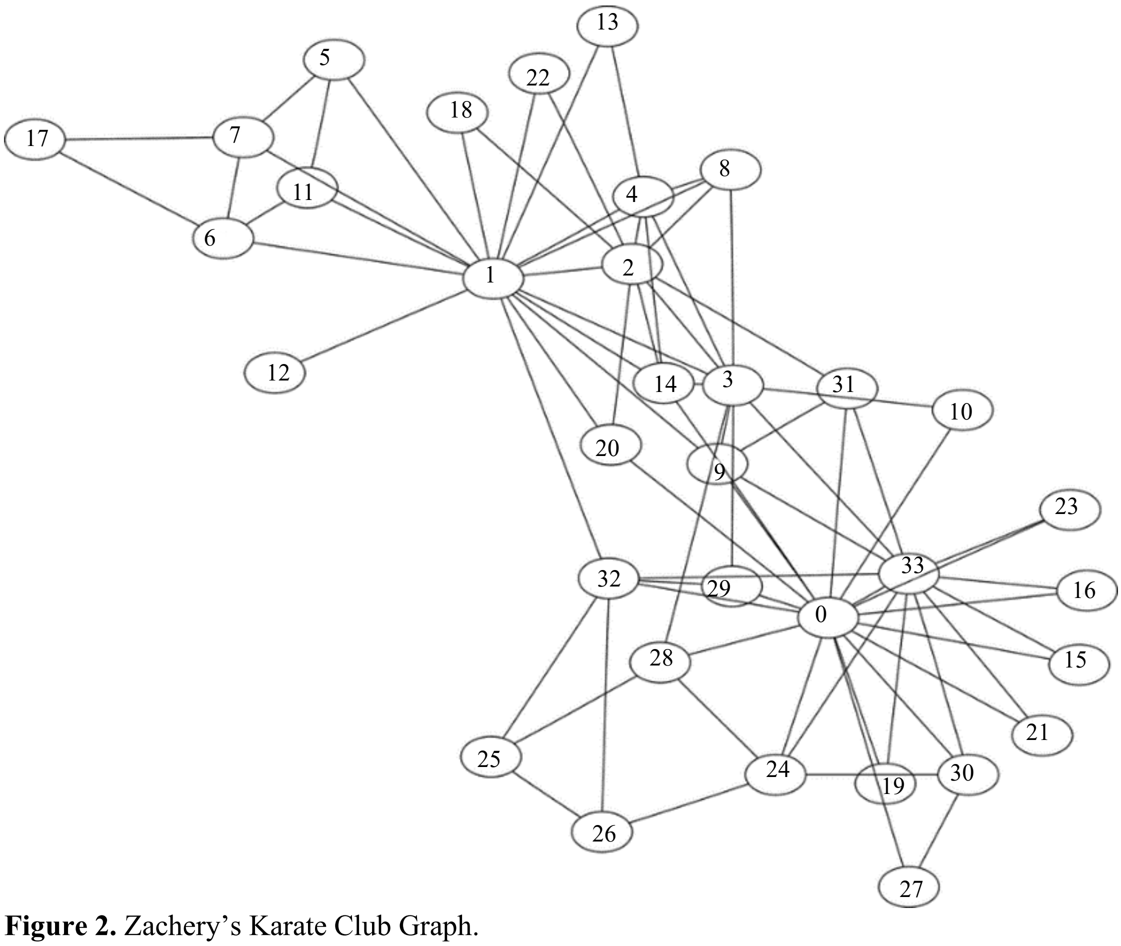

Figure 2. Zachery's Karate Club Graph.

Modularity: A measurement of how well a network is divided into communities

A graph's modularity is determined by comparing each intra-cluster edge's occurrence to the likelihood that it would exist in a randomly generated network. Modularity is given by Eq. (1),

$$\sum k(ekk - a2k) \tag{1}$$

Where, e_{kk}, the probability of intra-cluster edges in cluster S_k, and a_k, the probability of either an intra-cluster edge in cluster S_k or of an inter-cluster edge incident on cluster S_k, are

$$ekk == |\{(i,j): i \in S_k, j \in S_k, (i,j) \in E\}|/|E|,$$

$$a_k = |\{(i,j): i \in S_k, (i,j) \in E\}|/|E|$$

and where $S_k \subseteq V$.

Conductance: An indicator of the distance between communities

The number of inter-cluster edges for a cluster divided by the number of edges having endpoints inside the cluster or the number of edges lacking endpoints within the cluster, whichever is smaller, determines the conductance of the cluster. The conductance for a cluster is given by Eq. (2),

$$\phi(S_k) = \sum i \in S_k, j \notin SkAijmin\{A(S_k), A(\overline{S_k})\} \tag{2}$$

where $S_k \subset V$ and $A(S_k)=\sum_{i \in Sk} \sum_{j \in V} A_{ij} - \sum_{i \in Sk} \sum_{j \in Sk} A_{ij}$, the number of edges with an endpoint in S_k.

We define the conductance of a graph G to be the average of the conductance for each cluster in the graph, subtracted from 1. The conductance ranges between 0 and 1, and the subtraction makes 1 the optimal score. The conductance for a graph is given by Eq. (3),

$$\phi(G) = 1 - 1k\sum k\phi(S_k) \tag{3}$$

Coverage: Fraction of nodes in the network included in communities

Coverage measures how many intra-cluster edges there are in the graph relative to all edges in the graph. Coverage is given by Eq. (4),

$$\sum i, jAij\delta(Si, Sj)\sum i, jAij \tag{4}$$

Where, S_i is the cluster to which node i is assigned and $\delta(a, b)$ is 1 if $a=b$ and 0 otherwise. Coverage ranges between 0 and 1 where 1 is the optimal score.

Results and Comparative Analysis

The experimental results demonstrated the effectiveness of the hybrid approach in detecting communities within networks. Modularity for datasets is shown in Table 1. The integration of GNN-based representations with traditional algorithms showcased improvements in community detection accuracy and robustness across various datasets. Comparative analyses were conducted against standalone GNN-based methods and traditional community detection algorithms. Conductance for datasets is shown in Table 2. The hybrid approach exhibited superior performance in terms of modularity, conductance, and coverage metrics in most cases. Furthermore, the hybrid model demonstrated enhanced scalability compared to solely GNN-based or traditional methods in handling large-scale networks. Additionally, the interpretability of the detected communities was improved, benefiting from the integration of GNN-learned representations with the traditional algorithms, enabling clearer delineation of community structures. Coverage for datasets is shown in Table 3. The results validate the potential of the hybrid approach in leveraging both deep learning techniques and traditional algorithms for community detection tasks.

Table 1. Modularity for Datasets.

Datasets	Modularity: GCN	Modularity: GCN + Spec Clustering
Karate	0.4	0.6
Dolphins	0.4	0.7
Football	0.6	0.8

Table 2. Conductance for Datasets.

Datasets	Conductance: GCN	Conductance: GCN + Spec Clustering
Karate	0.5	0.7
Dolphins	0.4	0.6
Football	0.5	0.6

Table 3. Coverage for Datasets.

Datasets	Coverage: GCN	Coverage: GCN + Spec Clustering
Karate	0.6	0.8
Dolphins	0.5	0.6
Football	0.7	0.8

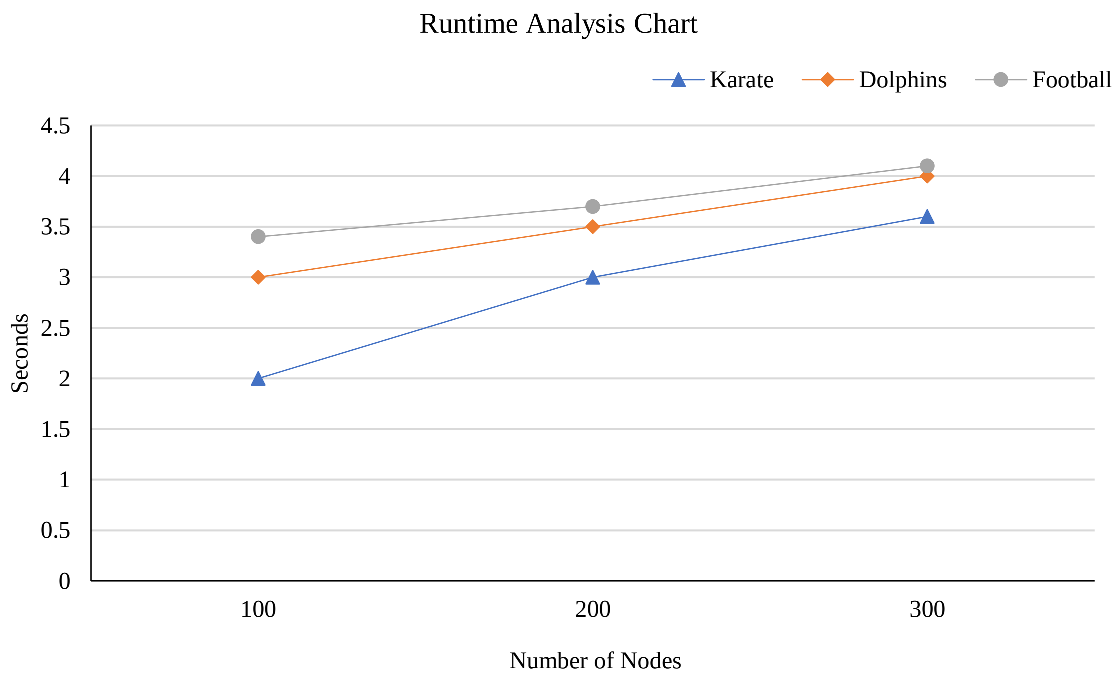

Figure 3. Run time analysis.

DISCUSSION AND CONCLUSION

The research presented a hybrid approach that integrates Graph Neural Networks (GNNs) with traditional community detection algorithms to identify communities within complex networks. This section discusses the findings, implications, and future directions arising from the study. The experimental results demonstrated the efficacy of the hybrid approach. The integration of GNN-learned representations with traditional algorithms led to improved performance metrics such as modularity, conductance, and coverage across various network datasets. Moreover, the hybrid model exhibited enhanced run time analysis addressing the challenges of handling large-scale networks. The interpretability of the detected communities was notably improved by leveraging both deep learning-based representations and the clarity offered by traditional community detection algorithms.

The proposed hybrid approach contributes to the advancement of community detection methodologies by offering a balance between the interpretability of traditional methods and the nuanced representations learned through deep learning. The integration of these approaches facilitates a more accurate and scalable analysis of complex networks, providing insights into their organizational structures and functionalities across diverse domains. The findings have implications in various fields such as social network analysis, biology, recommendation systems, and network security, where understanding community structures is pivotal for decision-making and optimization.

Conclusion

In this study, a hybrid approach combining Graph Neural Networks (GNNs) with traditional community detection algorithms was proposed and evaluated for the task of identifying communities within complex networks. The research aimed to leverage the strengths of deep learning techniques in capturing intricate patterns from network data while maintaining the interpretability offered by traditional algorithms for community detection. The experimental results showcased the effectiveness of the hybrid approach in enhancing community detection accuracy, robustness, and scalability. By harnessing the power of GNN-based representations learned from network structures, integrated with established community detection methodologies, the hybrid framework demonstrated improved performance metrics across diverse network datasets. The integration of GNN-learned embeddings with traditional algorithms led to a clearer delineation of community structures, improving the interpretability of the detected communities. Moreover, the hybrid model exhibited enhanced scalability, addressing the challenges of handling large-scale networks efficiently. The implications of this research extend to various domains where understanding network structures and communities is crucial. Applications in social network analysis, biology, recommendation systems, and network security could benefit from the insights provided by the hybrid approach.

Further research avenues include optimizing hyperparameters, exploring different GNN architectures, adapting the methodology to dynamic networks, and enhancing the interpretability of detected communities. Continued exploration in these directions will contribute to advancing community detection methodologies and their applications across diverse domains. The hybrid approach presented in this research work offers a promising solution for improving community detection tasks in complex networks. By amalgamating deep learning techniques with traditional algorithms, this approach provides a more comprehensive understanding of network structures, paving the way for advancements in network analysis and decision-making processes across various fields.

REFERENCES

1. Newman MEJ. Networks: An Introduction. England: Oxford University Press; 2010. https://doi.org/10.1093/acprof:oso/9780199206650.001.0001,
2. Fortunato S. Community detection in graphs. Phys Rep. 2010; 486(3–5): 75–174.
3. Newman MEJ, Girvan M. Finding and evaluating community structure in networks. Phys Rev E. 2004; 69(2): 026113.
4. Von Luxburg U. A tutorial on spectral clustering. Statistics and Computing. 2007; 17(4): 395–416.
5. Rosvall M, Bergstrom CT. Maps of random walks on complex networks reveal community structure. Proc Natl Acad Sci. 2008; 105(4): 1118–1123.
6. Wu Z, Pan S, Long G, Jiang J, Zhang C, Yu PS. A comprehensive survey on graph neural networks. IEEE Trans Neural Netw Learn Syst. 2021; 32(1): 4–24.
7. Kipf TN, Welling M. Semi-supervised classification with graph convolutional networks. International Conference on Learning Representations (ICLR). 2016.
8. Hamilton WL, Ying R, Leskovec J. Inductive representation learning on large graphs. Advances in Neural Information Processing Systems (NeurIPS). 2017; 1025–1035.
9. Veličković P, Cucurull G, Casanova A, Romero A, Lio P, Bengio Y. Graph attention networks. International Conference on Learning Representations (ICLR). 2018.

10. Perozzi B, Al-Rfou R, Skiena S. DeepWalk: Online learning of social representations. Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining, New York New York USA. 2014 Aug 24–27; 701–710.
11. Wang D, Cui P, Zhu W. Structural deep clustering networks for unsupervised learning. Proceedings of the 27th ACM International Conference on Information and Knowledge Management, Taipei Taiwan. 2020 Apr 20–24; 1949–1952. ACM digital library.
12. Sun C, Li C, Lin X, Zheng T, Meng F, Rui X, Wang Z. Attention-based graph neural networks: A survey. Artif Intell Rev. 2023;56(Suppl 2):2263–310. DOI: 10.1007/s10462-023-10577-2.
13. Aditya Grover, Jure Leskovec. Node2vec: Scalable Feature Learning for Networks. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'16). Association for Computing Machinery, New York, NY, USA. 2016; 855–864. https://doi.org/10.1145/2939672.2939754
14. Newman MEJ, Girvan M. Finding and evaluating community structure in networks. Phys Rev E. 2004; 69(2): 026113. https://doi.org/10.1103/PhysRevE.69.026113

