IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 1, June 2026
Impact Factor: 8.2

Applications of Graph Neural Networks in Social Network Mining

Udayagiri Seetaram Gupta[1] and Dr. Rajeev Yadav[2]
[1]research Scholar, [2]Associate Professor
Department of Computer Applications
Sunrise University Alwar Rajasthan

Abstract: *Graph Neural Networks have emerged as an important class of deep learning methods for analyzing graph-structured data, particularly social networks in which users, organizations, posts, and interactions can naturally be represented as nodes and edges. Unlike conventional machine-learning approaches that generally require manually engineered features, GNNs learn representations by combining node attributes with information propagated through graph neighborhoods. The development of graph representation learning before and during the emergence of GNNs established the foundation for applications such as node classification, link prediction, and community analysis. DeepWalk demonstrated that social relationships could be represented through random-walk-based embeddings, while node2vec introduced a more flexible strategy for exploring network neighborhoods (Perozzi et al., 2014; Grover & Leskovec, 2016).*

Keywords: Graph Neural Networks, Social Network Mining, Node Classification, Link Prediction, Community Detection.

I. INTRODUCTION

Social networks can be naturally modeled as graphs in which users or other entities are represented by nodes and social interactions are represented by edges. Examples include friendship networks, follower networks, communication networks, collaboration networks, online discussion networks, and recommendation networks. The large scale and complex structure of these networks create substantial opportunities for data mining, but they also make conventional machine-learning approaches difficult to apply because observations are not independent. The characteristics of one user may depend strongly on the characteristics and behavior of neighboring users. Consequently, social network mining requires methods capable of incorporating both individual attributes and relational structure.

Before the development of modern Graph Neural Networks, network representation learning provided an important foundation. DeepWalk, introduced by Perozzi et al. (2014), used truncated random walks to learn latent representations of network vertices and demonstrated their usefulness for classification and other social-network tasks. The method treated random walks as analogous to sequences and learned continuous representations capable of capturing local social relationships. Node2vec subsequently extended this idea by introducing biased random walks that could explore different types of network neighborhoods, making the learned representations more flexible for tasks such as node classification and link prediction (Grover & Leskovec, 2016).

II. EVOLUTION OF GRAPH-BASED DEEP LEARNING FOR SOCIAL NETWORK MINING

The period from 2010 to 2020 witnessed a gradual transition from conventional network analysis toward representation-learning-based approaches. Early network-mining techniques relied heavily on handcrafted structural measures, matrix factorization, probabilistic models, and community-detection algorithms. Although these approaches remain useful, they often require task-specific feature engineering. Representation learning changed this situation by allowing algorithms to automatically learn low-dimensional representations from network structure.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 1, June 2026 Impact Factor: 8.2

Deep Walk was an influential milestone because it demonstrated that random walks could be treated as sequences from which meaningful social representations could be learned (Perozzi et al., 2014). Its applications included social-network classification and anomaly detection, and its experiments involved networks such as Blog Catalog, Flickr, and YouTube. Node2vec expanded this framework by controlling the exploration of neighborhoods through biased random walks, allowing the model to capture different structural roles and proximity relationships (Grover & Leskovec, 2016).

The next major development was the introduction of GCNs. Kipf and Welling (2017) proposed a scalable graph convolutional architecture in which node representations are obtained by aggregating information from neighboring nodes. Their work showed that graph convolution could combine graph topology and node features within a unified learning framework. This was particularly relevant to social networks because users frequently have both relational information and attributes such as profile information, interests, textual content, or activity patterns.

III. GNNS FOR NODE CLASSIFICATION

Node classification is one of the most extensively studied applications of GNNs in social network mining. In this task, each node has a potentially unknown label, and the objective is to predict the label using available node attributes and graph relationships. In social networks, labels may represent user interests, demographic categories, professional roles, political or topical communities, content categories, or behavioral characteristics. The major advantage of GNN-based node classification is that a node's representation is influenced not only by its own features but also by the representations of neighboring nodes.

GCN provided a fundamental framework for this purpose. Kipf and Welling (2017) showed that graph convolution can learn hidden representations encoding both local graph structure and node features. Their experiments on citation networks demonstrated strong performance compared with several existing approaches, establishing GCN as an important baseline for graph-based semi-supervised classification. The same principle can be applied to social networks: if connected users tend to exhibit related behaviors or attributes, neighborhood aggregation can improve the prediction of labels for nodes with limited labeled information.

Graph SAGE extended node classification toward inductive learning. Instead of assigning a fixed learned vector to every training node, Graph SAGE learns an aggregation function that can generate representations for unseen nodes from their features and local neighborhoods. Hamilton et al. (2017) reported experiments involving evolving information networks, including Reddit data, demonstrating the relevance of inductive representation learning to networks in which new nodes continually emerge. This characteristic is highly relevant to social-media mining because social platforms are dynamic and may contain millions of newly created accounts or posts.

IV. GNNS FOR LINK PREDICTION

Link prediction aims to determine whether an edge should exist between two nodes. In social networks, this problem can correspond to predicting future friendships, following relationships, interactions, collaborations, recommendations, or communication links. Link prediction is important because social networks evolve continuously and the ability to predict emerging relationships can support recommendation systems, community analysis, and network evolution studies.

Earlier representation-learning approaches already demonstrated the value of learned node representations for link prediction. Node2vec explicitly evaluated its representations on both node- and edge-level prediction tasks and showed that flexible neighborhood exploration could improve network prediction performance (Grover & Leskovec, 2016). However, GNN-based approaches provided a more direct mechanism for incorporating neighborhood information.

A particularly important contribution was the Variational Graph Auto-Encoder proposed by Kipf and Welling (2016). VGAE uses a graph convolutional encoder to obtain latent representations and an inner-product decoder to reconstruct relationships. The authors demonstrated the approach on link prediction in citation networks and showed that node features could be naturally incorporated into the model. The architecture provides a conceptual framework for social-network link prediction because relationships can be reconstructed from latent representations learned from both graph structure and node attributes.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 1, June 2026 Impact Factor: 8.2

Graph SAGE can also support link prediction because learned node embeddings can be combined to estimate the likelihood of edges between pairs of nodes. Its inductive design makes it particularly attractive for dynamic social networks in which new users can appear after model training. Consequently, GNN-based link prediction can move beyond simple similarity measures by learning task-specific representations from multi-hop neighborhoods.

Table 1. Major Graph Representation and GNN Methods Relevant to Social Network Mining (2014–2020)

Method	Year	Main Principle	Primary Task	Relevance to Social Networks
Deep Walk	2014	Random-walk-based representation learning	Node classification, embedding	Learns social representations from local network walks
node2vec	2016	Biased random walks	Node classification, link prediction	Captures flexible structural neighborhoods
VGAE	2016	Graph autoencoder with variational latent variables	Link prediction	Learns latent representations for reconstructing network links
GCN	2017	Graph convolution and neighborhood aggregation	Node classification	Integrates node features with graph topology
Graph SAGE	2017	Sampling and neighborhood aggregation	Node classification, representation learning	Generates embeddings for unseen nodes
DGI	2019	Mutual-information-based self-supervised learning	Node representation/classification	Useful where labeled social-network data are limited

Table 2. Application-Wise Comparison

Application	Input Information	Typical GNN/Embedding Approach	Expected Output	Social Network Application
Node Classification	Node features + neighborhood structure	GCN, Graph SAGE, DGI	Node label	User-interest or user-category prediction
Link Prediction	Existing edges + node representations	VGAE, Graph SAGE-based models	Probability of an edge	Friend/follow recommendation
Community Detection	Graph topology + node features	GCN/VGAE embeddings + clustering	Communities/groups	Identification of interest or interaction groups
Node Representation	Graph structure + attributes	Graph SAGE, GCN, DGI	Low-dimensional embedding	Search, recommendation and visualization
Anomaly Analysis	Structural and behavioral patterns	Learned graph embeddings	Anomaly score/class	Detection of unusual accounts or interactions

V. ADVANTAGES OF GNNS IN SOCIAL NETWORK MINING

One of the major strengths of GNNs is their ability to combine structure and attributes. Conventional machine-learning models may treat users as independent observations, whereas GNNs explicitly exploit relationships between users. This is particularly important in social networks because neighboring users frequently share interests, behaviors, or interaction patterns. GCNs demonstrated that graph topology and node features can be incorporated into a common representation-learning process.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 1, June 2026
Impact Factor: 8.2

A second advantage is the ability to operate under limited supervision. Social networks can contain huge numbers of unlabeled users while only a small subset has known labels. GCNs address semi-supervised classification, while DGI provides an unsupervised approach for learning useful representations that can subsequently support downstream tasks.
A third advantage is inductive generalization. In rapidly changing social networks, training a model once and applying it only to the nodes observed during training is insufficient. GraphSAGE addresses this limitation by learning aggregation functions that can generate representations for previously unseen nodes.
Finally, GNNs can potentially capture higher-order structural information. Through multiple layers of message passing, a node can receive information from increasingly distant parts of the graph. This allows representations to encode more than direct connections and can be useful for identifying structurally similar users, predicting relationships, and discovering communities.

VI. RESEARCH GAPS

The literature from 2010–2020 demonstrates substantial progress, but several research gaps remain. First, much early GNN research focused on relatively static graph benchmarks rather than continuously evolving social networks. Future systems need to integrate temporal information so that changes in relationships can be explicitly modeled. Second, many models assume that graph homophily is strong, meaning connected nodes tend to have similar properties. Real social networks can contain heterogeneous and conflicting relationships, making this assumption problematic.
Third, community detection requires greater attention to overlapping communities. A user may simultaneously belong to professional, family, educational, entertainment, and interest-based groups. A model that assigns every node to a single community may therefore provide an incomplete representation of social behavior. Fourth, social networks are often heterogeneous, containing users, posts, hashtags, pages, organizations, and other node types. Future GNN architectures need to represent these heterogeneous entities and relation types more effectively.
Finally, there is a need for improved explain ability, privacy preservation, robustness, and fairness. These issues are particularly important when GNNs are applied to large-scale platforms where automated predictions can affect recommendations, content visibility, user classification, and relationship suggestions.

VII. CONCLUSION

Graph Neural Networks have significantly changed the methodological landscape of social network mining by enabling learning directly from graph structure while simultaneously incorporating node attributes. The development began with network representation-learning approaches such as Deep Walk and node2vec, which demonstrated that meaningful low-dimensional representations could be obtained from network neighborhoods and random walks. GCNs subsequently introduced a more direct neural mechanism for aggregating information over graph neighborhoods, providing a powerful approach to semi-supervised node classification. Graph SAGE extended this paradigm to inductive representation learning, making it possible to generate representations for previously unseen nodes, which is particularly valuable for dynamic social networks. VGAE demonstrated the applicability of graph neural representations to link prediction, while DGI expanded unsupervised and self-supervised graph representation learning.
Among the three major social-network mining applications considered in this review, node classification benefits from the ability of GNNs to combine labels, node attributes, and neighborhood information; link prediction benefits from learned representations that encode structural similarity and latent relationships; and community detection benefits from embeddings that capture both local topology and broader structural patterns. Overall, the literature from 2010 to 2020 indicates a clear transition from manually engineered network features toward learned graph representations. The future development of social-network mining is likely to depend on models capable of handling dynamic, heterogeneous, large-scale, noisy, and privacy-sensitive graphs while providing interpretable and fair predictions.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 1, June 2026
Impact Factor: 8.2

REFERENCES

1. Grover, A., & Leskovec, J. (2016). *node2vec: Scalable feature learning for networks*. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 855–864.
2. Hamilton, W. L., Ying, R., & Leskovec, J. (2017). *Inductive representation learning on large graphs*. Advances in Neural Information Processing Systems, 30, 1024–1034.
3. Kipf, T. N., & Welling, M. (2016). *Variational graph auto-encoders*. Bayesian Deep Learning Workshop, NeurIPS.
4. Kipf, T. N., & Welling, M. (2017). *Semi-supervised classification with graph convolutional networks*. International Conference on Learning Representations.
5. Perozzi, B., Al-Rfou, R., & Skiena, S. (2014). *Deep Walk: Online learning of social representations*. Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 701–710.
6. Veličković, P., Fedus, W., Hamilton, W. L., Liò, P., Bengio, Y., & Hjelm, R. D. (2019). *Deep Graph Infomax*. International Conference on Learning Representations.