2nd INTERNATIONAL CONFERENCE ON
Artificial Intelligence
31 May - 1 June 2026

IICAI
Second International Interdisciplinary Conference on Artificial Intelligence

A Comprehensive Review of Big Data Analytics Approaches in Social Networks Using Graph Neural Networks: Methods, Challenges, and Future Directions

Mohammad Javad Khedmati Yavand*, Hussein Saberi , Mohammad Ali Javadzade .
*- Corresponding author: PhD's student in artificial intelligence and robotics, Faculty of artificial intelligence and cognitive sciences, Imam Hossein Comprehensive University, Tehran, Iran
2,3- Assistant Professor, Faculty of artificial intelligence and cognitive sciences, Imam Hossein Comprehensive University, Tehran, Iran
mjkhedmati@ihu.ac.ir

Abstract

The rapid growth of social media generates massive, complex, and interdependent data that traditional machine learning methods cannot analyze effectively due to its non-Euclidean and relational nature. Graph Neural Networks (GNNs) have emerged as powerful tools for extracting meaningful patterns from large-scale social networks by leveraging message-passing and neighborhood aggregation. This study presents a systematic review of foundational GNN concepts, major architectures, and their principal applications in social network big data analytics. We provide a detailed, quantitative comparison of state-of-the-art methods for tasks such as node classification, link prediction, community detection, sentiment analysis, and fake news detection. The study also investigates critical challenges, including scalability, noise, privacy, and limited generalizability, along with their algorithmic implications. Finally, we highlight emerging research directions, such as the integration of GNNs with Large Language Models (LLMs), dynamic graph learning, federated learning, and improved model explainability.

The purpose of the research

The primary objective of this research is to provide a systematic and comprehensive review of Graph Neural Network (GNN) approaches for big data analytics in social networks. This study aims to:
1. Identify and categorize the fundamental concepts and major architectures of GNNs specifically designed for social network analysis.
2. Provide a detailed quantitative comparison of state-of-the-art GNN methods across key application areas including node classification, link prediction, community detection, sentiment analysis, and fake news detection.
3. Analyze the algorithmic implications of critical challenges such as scalability, noise and incomplete data, privacy concerns, and over-smoothing in deep GNN architectures.
4. Evaluate existing solutions to these challenges through quantitative performance metrics and complexity analysis.
5. Identify emerging research directions that integrate GNNs with Large Language Models, dynamic graph learning, federated learning, and explainable AI to guide future investigations in this rapidly evolving field.

By achieving these objectives, this review aims to present researchers with a consolidated, updated view of advances, limitations, and promising future directions for GNN-based analysis in large-scale social networks.

Research method

This study employs a systematic literature review methodology to comprehensively analyze GNN applications in social network big data analytics. The research process includes four phases: (1) systematic collection of over 200 papers from major databases (IEEE Xplore, ACM, arXiv) focusing on 2017-2025, (2) categorization of GNN architectures (GCN, GAT, GraphSAGE, GIN) and applications (node classification, link prediction, community detection, sentiment analysis, fake news detection), (3) quantitative comparative analysis using performance metrics including Accuracy, AUC, NMI, and F1-score across benchmark datasets (Cora, PubMed, Reddit, Twitter, Facebook), and (4) algorithmic challenge analysis examining scalability, noise handling, privacy preservation, over-smoothing, and generalizability with quantitative evaluation of solution effectiveness and trade-offs.

Practical or simulation results

Quantitative evaluation demonstrates that GNNs significantly outperform traditional methods across all social network analysis tasks. For node classification, GCN achieves 81.5% accuracy on Cora (+11.5% over MLP), GAT improves to 83.0%, and GraphSAGE reaches 95% accuracy on Twitter spam detection. For link prediction, VGAE achieves 94% AUC on Cora (+8% over spectral methods). Community detection using GAEC achieves NMI of 0.85 on Reddit. Sentiment analysis with HetSANN reaches 88% accuracy on Twitter (+6-8% over text-only models). Fake news detection using Bi-GNN exceeds 90% accuracy on FakeNewsNet. For scalability, neighbor sampling reduces complexity from O(degree) to O(K) with minimal accuracy loss, while Cluster-GCN reduces memory from O(N) to O(batch) with <1% accuracy trade-off. DropEdge improves robustness by 5-8% against noisy graph structures. For privacy, FedGNN enables training without raw data sharing, and FedGNN 2.0 reduces communication overhead by 40%. JK-Net and PairNorm enable deep GNN training (6-8 layers) with 8-12% accuracy improvement over standard architectures.

conclusion

This systematic review confirms that GNNs achieve superior performance in social network big data analytics through their unique capability to model complex graph relationships, with significant improvements over traditional methods: 11.5-13% higher accuracy in node classification, 8% higher AUC in link prediction, 6-8% improvement in sentiment analysis, and 15-20% better information propagation prediction. Key challenges including scalability, noise sensitivity, privacy concerns, and over-smoothing have been addressed through innovative solutions: neighbor sampling and Cluster-GCN for scalability (O(degree)→O(K) complexity reduction), DropEdge for robustness (5-8% improvement), federated learning and differential privacy for privacy preservation, and JK-Net/PairNorm for deep GNN training (8-12% accuracy gain). Future research directions integrating GNNs with Large Language Models, dynamic graph learning, and enhanced explainability promise to revolutionize social network analysis, enabling more efficient, robust, and ethically-aware understanding of human behavior and social phenomena in the digital age.

References

[1] W. L. Hamilton, Graph Representation Learning, Synthesis Lectures on Artificial Intelligence and Machine Learning, vol. 14, no. 3. Morgan & Claypool, 2020, pp. 1-159.
[2] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," in Proc. International Conference on Learning Representations (ICLR), Toulon, France, Apr. 2017, pp. 1-14.
[3] J. Zhou et al., "Graph neural networks: A review of methods and applications," AI Open, vol. 1, pp. 57-81, 2020, doi: 10.1016/j.aiopen.2020.100004.
[4] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Liò, and Y. Bengio, "Graph attention networks," in Proc. International Conference on Learning Representations (ICLR), Vancouver, Canada, Apr. 2018, pp. 1-12.
[5] W. L. Hamilton, R. Ying, and J. Leskovec, "Inductive representation learning on large graphs," in Proc. Advances in Neural Information Processing Systems (NeurIPS), Long Beach, CA, USA, Dec. 2017, pp. 1024-1034.
[6] K. Xu, W. Hu, J. Leskovec, and S. Jegelka, "How powerful are graph neural networks?," in Proc. International Conference on Learning Representations (ICLR), New Orleans, LA, USA, May 2019, pp. 1-17.
[7] A. Madani et al., "BERT-GraphSAGE: Hybrid approach to spam detection," Journal of Big Data, vol. 12, no. 1, pp. 1-20, May 2025, doi: 10.1186/s40537-025-01176-9.
[8] T. N. Kipf and M. Welling, "Variational graph auto-encoders," arXiv preprint arXiv:1611.07308, Nov. 2016.
[9] L. Chen et al., "Unsupervised graph attention autoencoder clustering-oriented for community detection," Data Science and Management, vol. 8, no. 2, pp. 123-140, May 2025, doi: 10.1007/s41060-025-00800-4.
,….