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 5, July 2026
Impact Factor: 8.2

Graph Neural Networks for Social Influence Prediction and Information Diffusion Analysis

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 deep-learning framework for analyzing complex relational data, particularly social networks in which users, interactions, communities, and information cascades can naturally be represented as graphs. Social influence prediction aims to identify users who are likely to influence others, while information diffusion analysis focuses on understanding and predicting how information propagates through interconnected users. Earlier approaches often relied on handcrafted network features, statistical diffusion models, or conventional machine-learning techniques.*

Keywords: Graph Neural Networks, Social Influence, Information Diffusion, Social Networks, Graph Convolutional Networks.

I. INTRODUCTION

Social networking platforms have transformed the way information is generated, shared, and consumed. Platforms such as Twitter, Facebook, Weibo, and other online communities generate highly interconnected networks in which users influence one another through following, friendship, communication, sharing, commenting, and reposting behaviors. Information introduced by one user can propagate through several levels of connected users, producing what is commonly described as an information cascade. Understanding these processes is important for applications such as viral marketing, recommendation systems, public-opinion analysis, misinformation detection, emergency communication, and online advertising. The central difficulty is that social influence is not determined simply by the number of connections a user has; rather, influence depends on network position, characteristics of neighboring users, previous interactions, community structure, temporal patterns, and the characteristics of the information being propagated.

Traditional social-network analysis generally represented users through manually constructed structural indicators such as degree, centrality, clustering coefficient, shortest paths, and community membership. Although these measures remain useful, they may fail to capture the high-dimensional and nonlinear relationships present in large-scale social networks. Representation-learning methods attempted to overcome this limitation by learning compact vector representations of network nodes. The emergence of deep graph learning subsequently enabled models to directly process graph structures and aggregate information from neighboring nodes. GCN research showed that graph-structured data could be processed through localized neighborhood propagation, providing an important foundation for applying neural networks to social-network analysis.

II. CONCEPTUAL FOUNDATION OF SOCIAL INFLUENCE PREDICTION

Social influence refers to the process through which the behavior, decisions, attitudes, or actions of one individual affect another individual within a social network. In an online environment, influence may be reflected through actions such as sharing a post, clicking a link, commenting, retweeting, following another user, or adopting a particular behavior. Social influence prediction therefore attempts to estimate whether a particular user will influence another user or whether an individual will participate in a future information cascade.

DOI: 10.48175/568

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 5, July 2026
Impact Factor: 8.2

The graph representation of a social network provides a natural mathematical framework for this problem. Users can be represented as nodes, while social relationships or interactions can be represented as edges. Each node may contain attributes such as activity level, historical behavior, profile characteristics, or learned embeddings. Similarly, edges may contain information about interaction frequency, relationship strength, communication direction, or temporal characteristics. A GNN processes these elements through neighborhood aggregation, enabling a node representation to incorporate information from connected users.

This approach is particularly valuable because influence is inherently relational. A user's behavior may depend not only on individual characteristics but also on the behavior of friends, followers, communities, and previous information adopters. GNNs are designed precisely for this type of relational dependency. Instead of treating every user as an independent observation, they learn representations in which neighboring nodes contribute to the representation of a target node.

III. EVOLUTION OF GRAPH-BASED LEARNING FOR SOCIAL NETWORKS, 2010–2020

The development of GNN-based social influence analysis can be understood as a progression from network representation methods toward increasingly sophisticated end-to-end neural architectures. Early research in the 2010s emphasized network embeddings and representation learning. These methods attempted to transform graph structures into low-dimensional vectors while preserving relationships between nodes. Such representations could subsequently be used for classification, recommendation, link prediction, and influence-related tasks.

A major transition occurred with the development of graph convolutional approaches. Rather than independently learning an embedding for every node, GCN-style methods aggregate information from neighboring nodes. This mechanism makes it possible to incorporate local structural context directly into learned representations. Reviews of GCN research emphasize that graphs naturally capture relationships in areas including social analysis and that graph convolution provides a mechanism for exploiting those structural relations.

Attention-based graph learning further improved this framework by allowing models to assign different importance to neighboring nodes. This is especially relevant to social influence because not every social connection contributes equally to diffusion. A highly influential neighbor may provide substantially more predictive information than a weak or inactive connection. GAT-type models therefore provide a mechanism for learning relationship-specific importance rather than applying identical aggregation weights to all neighbors.

IV. GRAPH NEURAL NETWORKS AND SOCIAL INFLUENCE PREDICTION

A typical GNN-based social influence model begins by constructing a graph representing the social environment. Each node is associated with an initial feature vector, while edges represent relationships between users. During graph propagation, information from neighboring nodes is aggregated and combined with the target node's existing representation. Repeated propagation allows the model to capture increasingly broad structural contexts.

For social influence prediction, the learned representation can subsequently be used for binary classification, multilabel prediction, ranking, or probability estimation. For example, the model may estimate whether user uu will activate user vv after receiving particular information. The prediction can incorporate node attributes, local network structure, previous interactions, and information-related features.

The advantage of this approach is that the model does not need to manually determine which network features are most informative. Instead, the neural architecture learns useful representations from observed examples. Attention mechanisms can further identify which neighbors are particularly important. Evidence from social-influence experiments has shown that GAT-based approaches can outperform simpler logistic-regression, GCN, and other baseline approaches on datasets such as Weibo and Twitter, illustrating the potential value of adaptive neighborhood weighting.

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 5, July 2026
Impact Factor: 8.2

V. INFORMATION DIFFUSION AND CASCADE PREDICTION

Information diffusion describes the process through which information moves from one user to another through a network. When a user initially publishes information and other users subsequently share or respond to it, a cascade is created. Such cascades can be represented as temporal graphs, trees, or subgraphs embedded within the larger social network.

The prediction of cascade growth is challenging because diffusion is affected by both network structure and temporal dynamics. A cascade that begins with a small number of users can sometimes become extremely popular, while another cascade with apparently similar initial characteristics may stop quickly. Traditional approaches often attempted to solve this problem by designing handcrafted features based on cascade size, structural properties, and user characteristics. DeepCas represented an important methodological shift by learning the representation of the cascade itself through an end-to-end deep-learning framework. The study found that representing the cascade graph as a whole was important and reported improvements over feature-based, node-embedding, and graph-kernel methods.

The development of diffusion-prediction research around 2016–2017 included embedded cascade models, topological recurrent neural networks, DeepHawkes, and other neural approaches. These studies demonstrate a broader transition toward models that learn directly from diffusion structure rather than relying solely on manually defined indicators.

VI. DEEP LEARNING FOR INFORMATION CASCADE PREDICTION

Deep learning expanded information diffusion research by allowing models to automatically learn representations of diffusion processes. DeepCas is a notable example from the 2017 period. Instead of requiring researchers to manually define numerous cascade characteristics, it learned representations of cascade graphs in an end-to-end manner. Its results suggested that representing the entire cascade can be more informative than relying only on individual node embeddings.

Other research during this period explored temporal and topological information. Topological recurrent neural networks attempted to incorporate diffusion topology, while DeepHawkes combined neural representation learning with cascade dynamics. This development is important because information diffusion is simultaneously structural and temporal: the model must understand both who influences whom and when activation occurs.

By 2020, CoupledGNN approaches had begun explicitly modeling the interaction between node activation states and influence propagation. Such approaches aimed to capture cascading effects more naturally by coupling graph representations with the successive activation process.

VII. COMPARATIVE DEVELOPMENT OF MAJOR APPROACHES

Period	Approach	Main Idea	Social Influence/Diffusion Contribution	Major Limitation
2010–2013	Network analysis and handcrafted features	Use degree, centrality, community and structural measures	Provides interpretable structural indicators	Requires manual feature engineering
2013–2016	Network representation learning	Learn low-dimensional node representations	Captures latent relationships among users	Often focuses on individual nodes
2016	Embedded cascade models	Learn representations of diffusion cascades	Connects graph structure with diffusion prediction	Temporal complexity can remain difficult
2017	Deep Cas	End-to-end cascade representation	Predicts future cascade growth from cascade structure	Primarily focused on cascade prediction

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 5, July 2026 Impact Factor: 8.2

2017	Topological/temporal neural models	Combine diffusion topology with sequence modeling	Captures structural and temporal diffusion patterns	More complex training requirements
2017–2019	GCN-based models	Aggregate information from graph neighborhoods	Learns relational representations for influence prediction	Conventional aggregation may treat neighbors similarly
2018–2020	GAT-based models	Attention-weighted neighborhood aggregation	Learns different importance for different neighbors	Computational and interpretability challenges
2019	Graph Diffusion Convolution	Uses broader graph diffusion instead of only one-hop propagation	Captures more extensive graph context	Greater computational complexity
2020	Coupled GNN approaches	Couple activation states with influence propagation	Models cascading activation more directly	Requires rich diffusion and network data

Graph diffusion convolution introduced another important direction in 2019 by extending graph convolution beyond strictly direct one-hop neighborhoods. The approach used generalized graph diffusion, including heat-kernel and personalized-PageRank-style diffusion, to address limitations associated with noisy or arbitrarily defined edges.

VIII. METHODOLOGICAL FRAMEWORK

A research framework for GNN-based social influence prediction generally consists of five major stages: data collection, graph construction, feature representation, graph learning, and prediction evaluation. In the data-collection stage, information can be obtained from social-network interactions, including following relationships, friendship connections, reposts, comments, mentions, and historical diffusion events. The resulting data are converted into a graph in which users are represented as nodes and relationships are represented as edges.

The second stage involves feature construction. Node features may include activity frequency, historical engagement, user-level behavioral characteristics, and learned embeddings. Edge features may represent interaction frequency, relationship strength, or temporal information. The graph is then supplied to a GNN architecture such as GCN or GAT.

During graph learning, neighborhood information is aggregated to generate latent node representations. These representations can then be passed to a prediction layer. For influence prediction, the output may represent the probability that one user influences another. For diffusion analysis, the output may represent predicted cascade size, future adoption, popularity, or probability of further propagation.

Evaluation can involve classification measures such as accuracy, precision, recall, F1-score, and AUC for user-level influence prediction. For cascade-size prediction, regression metrics such as mean absolute error, root mean squared error, or correlation-based measures may be more appropriate. The choice of metric should correspond to the research objective and the distribution of the target variable.

IX. DISCUSSION

The literature from 2010 to 2020 shows a clear methodological progression from handcrafted network analysis toward learned graph representations. Early approaches established that network topology is important for understanding social influence, while representation learning demonstrated that latent structural characteristics could be learned automatically. The emergence of GCNs provided a general framework for aggregating neighborhood information, and GATs introduced adaptive attention over social relationships.

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 5, July 2026
Impact Factor: 8.2

Information diffusion research developed along a complementary trajectory. DeepCas demonstrated the importance of learning cascade-level representations rather than depending exclusively on individual node embeddings. Research during 2017 also included topological recurrent models and other neural approaches for modeling diffusion, indicating that the field increasingly recognized the importance of combining network structure with temporal information.

By the end of the 2010–2020 period, the convergence of graph representation learning, attention mechanisms, and diffusion modeling created a strong foundation for modern social influence prediction. The use of graph diffusion mechanisms in 2019 further demonstrated that information propagation within a graph does not necessarily need to be restricted to immediate neighbors. The 2020 development of coupled GNN approaches represented another step toward directly modeling the cascading relationship between node activation and information propagation.

X. CONCLUSION

Graph Neural Networks provide a powerful framework for social influence prediction and information diffusion analysis because social networks are inherently graph-structured. The research trajectory between 2010 and 2020 demonstrates a movement from manually designed network features toward end-to-end learning of structural, relational, and diffusion representations. GCNs made it possible to incorporate neighborhood information into node representations, while GATs improved this process by learning differentiated importance among neighboring users. Cascade-oriented approaches such as DeepCas demonstrated that complete diffusion structures can provide valuable information for predicting future cascade growth.

The overall evidence indicates that GNN-based methods are particularly valuable when social influence depends on complex relationships and multi-hop network structure. Nevertheless, important challenges remain, including scalability, temporal dynamics, noisy data, interpretability, and distinguishing predictive association from causal influence. Future research should therefore focus on dynamic and heterogeneous GNNs capable of jointly modeling users, content, relationships, time, and diffusion events. Such developments can contribute to more accurate prediction of information cascades while also improving understanding of the mechanisms underlying social influence.

REFERENCES

1. Bourigault, S., Lamprier, S., & Gallinari, P. (2016). Representation learning for information diffusion through social networks: An embedded cascade model. *Proceedings of the Ninth ACM International Conference on Web Search and Data Mining*.
2. Cao, Q., Shen, H., Cen, K., Ouyang, W., & Cheng, X. (2017). DeepHawkes: Bridging the gap between prediction and understanding of information cascades. *Proceedings of the 26th ACM International Conference on Information and Knowledge Management*.
3. Cao, Q., Shen, H., Gao, J., Wei, B., Cheng, X., et al. (2020). Popularity prediction on social platforms with coupled graph neural networks. *Proceedings of the 13th International Conference on Web Search and Data Mining*, 70–78.
4. Gao, S., Pang, H., Gallinari, P., Guo, J., & Kato, N. (2017). A novel embedding method for information diffusion prediction in social network big data. *IEEE Transactions on Industrial Informatics*.
5. Gasteiger, J., Weißenberger, S., & Günnemann, S. (2019). Diffusion improves graph learning. *Advances in Neural Information Processing Systems, 32*.
6. Li, C., Ma, J., Guo, X., & Mei, Q. (2017). DeepCas: An end-to-end predictor of information cascades. *Proceedings of the 26th International Conference on World Wide Web*, 577–586.
7. Wang, J., Zheng, V. W., Liu, Z., & Chang, K. C.-C. (2017). Topological recurrent neural network for diffusion prediction. *Proceedings of the IEEE International Conference on Data Mining*.
8. Wang, Y., Shen, H., Liu, S., Gao, J., & Cheng, X. (2017). Cascade dynamics modeling with attention-based recurrent neural network. *Proceedings of IJCAI*.

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 5, July 2026
Impact Factor: 8.2

9. Zhang, S., Tong, H., Xu, J., & Maciejewski, R. (2019). Graph convolutional networks: A comprehensive review. *Computational Social Networks, 6*, 11.
10. Zhang, Y., Lyu, T., & Zhang, Y. (2017). Hierarchical community-level information diffusion modeling in social networks. *Proceedings of SIGIR*.