Social Network Analysis Using Graph Neural Networks and Large Language Models

Jinjie Lin

Department of Mathematics and Statistics, Hainan Normal University, Hainan, China

wulingfan765@gmail.com

Abstract. Social network analysis has become essential for understanding interaction patterns, information diffusion, and community behavior in the digital era. With the rapid expansion of online platforms such as Twitter, Facebook, and Weibo, vast amounts of data are generated daily, reflecting not only individual communication but also broader societal trends. This explosion of data requires advanced computational techniques to process, analyze, and derive meaningful insights. Graph Neural Networks (GNNs) and Large Language Models (LLMs) have emerged as two of the most promising approaches for addressing the complexity inherent in social network data. GNNs excel at modeling relational structures within networks, allowing for accurate predictions of community structures, influence propagation, and link prediction. LLMs, built on transformer architectures, process large-scale unstructured textual data, enabling tasks such as sentiment analysis, misinformation detection, and influencer identification. This paper offers an in-depth overview of the use of Graph Neural Networks (GNNs), Large Language Models (LLMs), and their combined application in analyzing social networks. It provides a structured assessment of the advantages and shortcomings of these methodologies, highlights existing obstacles, and proposes potential avenues for future research to overcome them. Major suggestions involve improving computational efficiency, reducing biases in both data and model outputs, advancing the modeling of time-dependent behaviors, and developing consistent and reliable evaluation protocols for integrated models. By consolidating recent advancements in this fast-growing domain, the study emphasizes the significant potential of merging GNNs and LLMs to achieve more precise, scalable, and fair analysis of social networks.

Keywords: Graph neural networks, large language models, social network.

1. Introduction

Social networks have become integral to the way people interact, communicate, and form communities in the digital age. Platforms such as Twitter, Facebook, and LinkedIn generate massive amounts of data through user interactions, content sharing, and communication. This data holds valuable insights into human behavior, societal trends, and the spread of information, making social network analysis an essential tool for fields such as marketing, political science, and public health.

The rapid growth of social media platforms has introduced new challenges for traditional methods of social network analysis, which were primarily based on statistical models and heuristic algorithms. These traditional methods were limited in their ability to scale and handle the complexity of large, dynamic networks. To overcome these limitations, machine learning (ML) techniques, particularly deep learning, have emerged as powerful tools for analyzing social network data. Among the various deep learning approaches, two paradigms have garnered significant attention: Graph Neural Networks (GNNs) and Large Language Models (LLMs).

Graph Neural Networks (GNNs) are specifically developed to handle data structured as graphs by capturing node features through the analysis of topological connections. They have demonstrated strong performance across a range of social network applications, such as classifying nodes, predicting links between entities, and identifying community structures. By effectively representing both entities as nodes and their interactions as edges, GNNs are well-suited for capturing the intricate and interconnected patterns commonly found in social networks.

On the other hand, LLMs, built on the transformer architecture, have revolutionized natural language processing (NLP). These models have demonstrated superior performance in tasks such as sentiment analysis, text generation, and information extraction, thanks to their ability to capture long-range dependencies and contextual information in textual data. Social media platforms generate vast amounts of unstructured textual data, such as posts, comments, and messages, and LLMs are well-suited for analyzing this content to derive insights into public sentiment, trends, and influence.

Despite the individual successes of GNNs and LLMs, recent research has shifted towards integrating these two paradigms to combine their complementary strengths. Hybrid models, such as GNN-LM and Graph Neural Prompting, have shown promise in tasks such as influence detection, misinformation tracking, and dynamic network modeling, where both structural and semantic information are crucial. The integration of these models allows for more comprehensive analyses of social networks, providing deeper insights into the underlying dynamics.

However, the combination of GNNs and LLMs presents several challenges. These include high computational costs, the propagation of biases from both textual and network data, limited adaptability to evolving networks, and a lack of robust evaluation metrics to assess model performance across different tasks. Addressing these challenges is critical for realizing the full potential of hybrid models in social network analysis.

This paper aims to provide a comprehensive review of GNNs, LLMs, and their integration in social network analysis. This paper will discuss the strengths and limitations of each approach, explore hybrid models that combine GNNs and LLMs, and analyze the challenges that remain in the field. The paper will also outline future research directions to address these challenges and improve the performance and applicability of hybrid models.

2. Method

2.1. LLMs in Social Network Analysis

Large Language Models (LLMs), including models like GPT (Generative Pretrained Transformer) and BERT (Bidirectional Encoder Representations from Transformers), have significantly advanced the domain of natural language processing (NLP) by demonstrating remarkable capabilities in comprehending and producing human-like text. These models are based on the transformer framework, which employs self-attention mechanisms to effectively process and represent extensive textual data, enabling them to capture complex contextual relationships and long-distance linguistic dependencies. Thanks to this architectural foundation, LLMs achieve superior performance across a wide range of NLP applications such as sentiment analysis, text categorization, and automated translation, surpassing the limitations of conventional approaches.

In the context of social network analysis, LLMs have been applied to analyze user-generated content on platforms like Twitter, Reddit, and Facebook. One of the most significant applications of LLMs in social network analysis is influencer identification. Zhang et al. demonstrated how ChatGPT, a version of GPT, could identify key influencers in a social network based solely on textual interactions without the need for labeled data [1]. This approach leverages the zero-shot learning capabilities of LLMs, allowing them to infer influence from the content of user posts and comments.

LLMs have also been used for synthetic social network generation, as demonstrated by Park et al. in their Social Simulacra framework [2]. By training an LLM to simulate realistic user behaviors and interactions, researchers can generate synthetic communities for testing algorithms and interventions without relying on real-world data. This allows for more controlled experiments in social network analysis, particularly when studying the effects of different social dynamics or intervention strategies.

The ability of LLMs to analyze vast amounts of unstructured text also makes them valuable for misinformation detection. Given the prevalence of fake news and the rapid spread of misinformation on social media platforms, LLMs can be used to identify false narratives and detect misleading content based on linguistic patterns and context. Moreover, LLMs can help track the sentiment of users or groups within social networks, providing insights into public opinion on specific topics or events.

2.2. GNNs in Social Network Analysis

Graph Neural Networks (GNNs) are specifically developed to handle data structured in the form of graphs, with nodes denoting entities and edges capturing the interconnections among them. In contrast to conventional machine learning approaches that assume data points are isolated, GNNs utilize the topological structure of graphs by incorporating neighborhood information to generate more informative and context-aware node representations. This capability renders GNNs highly effective for applications where relational information plays a key role, including but not limited to analyzing social networks.

Graph Neural Networks (GNNs) operate through a message-passing framework, in which nodes iteratively collect and integrate information from their neighboring nodes to refine their own feature representations. By repeating this process across several layers, nodes can capture structural dependencies from increasingly distant regions of the graph. In the context of social network analysis, GNNs facilitate tasks such as community identification, link prediction—estimating likely future connections between users—and the detection of unusual or anomalous patterns within the network topology.

One notable extension of GNNs is the Graph Attention Network (GAT), introduced by Veličković et al. [3]. GATs use attention mechanisms to assign different weights to neighboring nodes, allowing the model to focus on the most influential connections. This improves the performance of GNNs in tasks such as node classification and community detection, where the importance of different relationships may vary.

Another extension, Temporal Graph Networks (TGNs), developed by Rossi et al., is designed to model the dynamics of time-evolving networks [4]. Social networks are not static; they evolve over time as new relationships are formed and old ones are dissolved. TGNs integrate temporal information into GNNs, enabling them to predict the future evolution of relationships and identify emerging trends in social networks.

Additionally, robust GNNs have been developed to handle adversarial attacks on social networks [5]. Malicious actors can manipulate network structures to distort analyses, for example, by spreading misinformation or amplifying certain voices. Robust GNNs mitigate these effects by incorporating defenses against adversarial manipulation, ensuring that social network analysis remains reliable even in the presence of malicious actors.

2.3. Hybrid Approaches

Hybrid approaches that integrate GNNs and LLMs combine the strengths of both paradigms to address more complex tasks in social network analysis. While GNNs excel at modeling the structural dependencies between nodes, LLMs provide deep semantic understanding of unstructured textual data. By integrating these two models, hybrid approaches can better capture the intricate relationships between the text and the structure of social networks.

An illustrative instance is Graph Neural Network–Language Model (GNN-LM), introduced by Zhu et al. [6], which integrates graph-based embeddings into the training framework of language models. By doing so, the model can process textual content while simultaneously accounting for the inherent network topology, thereby enabling more precise predictions in tasks such as influence analysis, sentiment classification, and community identification. Embedding GNNs within LLMs strengthens their capacity to capture user interrelations, leading to enhanced performance in social network-related applications.

Another hybrid approach is Graph Neural Prompting, introduced by Sun et al. [7], which uses graph-derived embeddings as prompts to guide the outputs of LLMs. This method allows LLMs to generate more informed responses by leveraging the structural context provided by the graph. For example, when identifying key influencers, the graph embeddings can provide additional context about the user's position in the network, enabling the LLM to produce more accurate and relevant predictions.

3. Discussion

3.1. Computational Complexity

Hybrid GNN–LLM models demand substantial computational resources. LLMs with billions of parameters already require significant hardware, and GNNs add iterative graph operations. Scaling such models to networks with millions of users remains challenging.

3.2. Model Bias

Biases in textual corpora and graph structures propagate through LLMs and GNNs. For instance, marginalized groups may appear less influential due to fewer connections, while biased text can amplify stereotypes. Hybrid models risk compounding these biases, leading to unfair predictions.

3.3. Generalization

Social networks are dynamic, with evolving behaviors and relationships. Many hybrid models rely on static data, limiting adaptability. Temporal approaches like TGNs and transfer learning strategies offer partial solutions but require further refinement.

3.4. Evaluation Gaps

Traditional metrics such as accuracy or F1-score are insufficient for hybrid models that must balance structural and semantic dimensions. Without multi-faceted benchmarks, it is difficult to assess progress fairly.

4. Future Prospects

4.1. Efficient Models

Compression techniques such as pruning and knowledge distillation can reduce computational demands [8, 9], while distributed computing enables deployment at scale.

4.2. Fairness and Bias Mitigation

Fairness-aware training and debiasing approaches should be integrated into hybrid models. Data preprocessing pipelines can also identify and filter biased inputs.

4.3. Temporal Adaptation

Incorporating temporal modeling through TGNs and fine-tuning with transfer learning can enhance adaptability to evolving networks [10]. This is crucial for applications like misinformation monitoring, where both content and structure change quickly.

4.4. Comprehensive Evaluation

Developing composite evaluation frameworks is essential. Benchmarks should integrate structural metrics (e.g., modularity, influence spread) and semantic metrics (e.g., coherence, sentiment accuracy). Such frameworks would provide more rigorous comparisons across hybrid approaches.

5. Conclusion

GNNs and LLMs have shown great potential in social network analysis, especially in handling complex network structures and rich textual data. LLMs excel at understanding and generating textual content, while GNNs are proficient at capturing the relationships between nodes and edges in networks. The combination of these two approaches provides a more comprehensive and accurate tool for dynamic social network analysis and influence prediction. However, these methods still face significant challenges, particularly in terms of computational complexity, bias control, generalization, and evaluation standards.

Future research can address these challenges through various approaches. For example, to tackle computational complexity, model compression techniques and distributed computing methods can be explored to reduce the computational load. To mitigate model bias, fairness-aware training methods and data preprocessing techniques can be developed to reduce bias in training data. Additionally, employing transfer learning and temporal GNNs can improve generalization and help models adapt to evolving social network environments. Finally, the development of new evaluation frameworks will help assess hybrid models more comprehensively in social network analysis.

Overall, the integration of GNNs and LLMs holds vast potential, not only providing innovative approaches to social network analysis but also opening new avenues for combining graph and textual data in other domains.

References

[1] Zhang Y, Wang H, Li X, et al. ChatGPT for social network analysis: A case study on influencer identification. arXiv preprint arXiv:2305.12651. 2023.

[2] Park JS, Kim J, Lee S, et al. Social simulacra: Creating populated prototypes for social computing systems. In: Proc 35th Annu ACM Symp User Interface Softw Technol (UIST). 2022. p. 1-15.

[3] Veličković P, Cucurull G, Casanova A, Romero A, Liò P, Bengio Y. Graph attention networks. In: Proc Int Conf Learn Represent (ICLR). 2018.

[4] Rossi E, Chamberlain B, Frasca F, Eynard D, Monti F, Bronstein M. Temporal graph networks for deep learning on dynamic graphs. arXiv preprint arXiv:2006.10637. 2020.

[5] Zhu D, Zhang Z, Cui P, Zhu W. Robust graph neural networks against adversarial attacks. In: Proc 25th ACM SIGKDD Int Conf Knowl Discov Data Min (KDD). 2019. p. 1399-407.

[6] Zhu Y, Xu Y, Chen J, et al. GNN-LM: A graph neural network enhanced language model. arXiv preprint arXiv:2110.14127. 2021.

[7] Sun X, Liu Q, Wang J, et al. Graph neural prompting. arXiv preprint arXiv:2310.00001. 2023.

[8] Liu Z, Sun M, Zhou T, Huang G, Darrell T. Rethinking the value of network pruning. arXiv preprint arXiv:1810.05270. 2018 Oct 11.

[9] Gou J, Yu B, Maybank SJ, Tao D. Knowledge distillation: A survey. International journal of computer vision. 2021 Jun;129 (6):1789-819.

[10] Han X, Huang Z, An B, Bai J. Adaptive transfer learning on graph neural networks. InProceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining 2021 Aug 14 (pp. 565-574).