Published by :
https://www.ijert.org/
An International Peer-Reviewed Journal

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 05 , May - 2026

Anomalous Behavior Detection in Social Networks Using Graph Neural Networks

Dhanashekar Pattem
Dept. of Computer Engineering SIES Graduate School of Technology
Nerul, Navi Mumbai, India

Dipak Patil
Dept. of Computer Engineering SIES Graduate School of Technology
Nerul, Navi Mumbai, India

Geeten Parab
Dept. of Computer Engineering SIES Graduate School of Technology
Nerul, Navi Mumbai, India

Pranav B
Dept. of Computer Engineering SIES Graduate School of Technology
Nerul, Navi Mumbai, India

Shraddha Kawji
Dept. of Computer Engineering SIES Graduate School of Technology
Nerul, Navi Mumbai, India

***Abstract* - A dual-model Graph Neural Network (GNN) framework is proposed for identifying malicious actors and coordinated botnets in social networks. By modeling user interactions as heterogeneous graphs, the system leverages a supervised Graph Attention Network (GAT) to classify known bot patterns and an unsupervised Graph Autoencoder (GAE) to detect structural anomalies based on reconstruction error. Multi-modal user representations are constructed using numerical profile features, categorical metadata, and RoBERTa-based semantic embeddings to capture intent beyond simple activity metrics.**

The integration of the Louvain modularity algorithm isolates dense clusters of coordinated accounts, which are then analyzed through a large language model to generate natural language summaries of behavioral and semantic characteristics.

Preliminary evaluations on the TwiBot-22 and Cresci-2015 datasets demonstrate the system's capability to achieve high detection accuracy while providing interpretable insights into evolving attack strategies. The analysis further highlights the trade-offs between structural accuracy, resource overhead, and the necessity for scalable, automated security in modern social network environments.

Index Terms—Graph Neural Networks, GAT, GAE, Louvain Modularity, Coordinated Botnets, Anomaly Detection

I. INTRODUCTION

Social networks have become critical communication infrastructures, enabling large-scale information exchange and social interaction. However, their openness also makes them vulnerable to malicious activities such as automated bots, spam networks, coordinated misinformation campaigns, and compromised accounts.

Traditional detection mechanisms, including rule-based filters and honeypot systems, were designed for conventional network security settings. While useful in controlled environments, they struggle to scale in dynamic social platforms and often rely on static thresholds that can be easily bypassed by adaptive adversaries. As attackers increasingly mimic legitimate user behavior, isolated user-level analysis
becomes insufficient.

Modern social threats are inherently collective. Botnets and coordinated groups manipulate trends and engagement metrics by forming dense relational structures. Detecting such behavior requires analyzing the network as an interconnected system rather than as independent user profiles.

To address these challenges, this work proposes a multi-layered anomalous behavior detection framework based on graph-centric modeling. Users are represented as nodes and interactions as directed edges, enabling analysis of the relational fabric of the network. A dual Graph Neural Network (GNN) architecture is employed, combining a supervised Graph Attention Network (GAT) for bot classification and an unsupervised Graph Autoencoder (GAE) for structural anomaly detection.

The framework further incorporates Louvain community detection, statistical thresholding, and AI-driven semantic profiling to identify and interpret coordinated malicious groups.

II. OUR CONTRIBUTION

This study makes several contributions to the field of social network anomaly detection by transforming a conceptual graph-based model into a practical, multi-layered security framework. First, we design a unified graph-centric architecture that captures both individual behavioral signals and collective relational patterns within large-scale social networks. Unlike traditional user-level approaches, the proposed system analyzes the structural fabric of interactions to detect coordinated malicious activity.

Second, we implement a dual Graph Neural Network framework that integrates supervised and unsupervised learning. The supervised Graph Attention Network (GAT) enables accurate classification of known malicious entities by assigning higher importance to influential neighboring nodes in the interaction graph. In parallel, the unsupervised Graph Autoencoder (GAE) learns normal structural patterns and detects deviations through

reconstruction error, allowing identification of previously

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 05 , May - 2026

un-seen or zero-day anomalies.

Third, the framework extends beyond individual detection by incorporating Louvain community detection and statistical thresholding to uncover coordinated botnets and dense anomalous clusters. Communities whose mean reconstruction error exceeds a defined statistical boundary are flagged as potential organized threats rather than isolated anomalies.

Finally, we enhance interpretability and operational usability through semantic feature modeling using RoBERTa embeddings and AI-driven profiling of suspicious communities. The complete system is deployed through a Streamlit-based dashboard, enabling scalable, analysis across multiple detection pillars, including bot detection, spammer identification, and coordinated anomaly discovery.

III. RESEARCH GAPS

Despite extensive research on malicious behavior detection, several critical gaps remain unresolved. A major limitation of traditional detection systems is their reliance on isolated user-level analysis. Rule-based and classical machine learning models often ignore relational dependencies, making them ineffective against coordinated attacks that exploit network structure.

Supervised learning approaches require large labeled datasets, which are expensive to obtain and often suffer from severe class imbalance. In real-world social networks, malicious accounts represent only a small fraction of users, leading to biased classifiers that struggle to generalize to rare or emerging attack patterns [5], [6]. Moreover, supervised models are inherently limited in their ability to detect previously unseen anomalies.

Unsupervised methods such as Isolation Forests and graph autoencoders offer improved generalization but introduce new challenges. These models may incorrectly flag legitimate high-activity users, such as journalists or influencers, as anomalous. Additionally, many unsupervised approaches lack interpretability, making it difficult for analysts to trust or validate detection results [8].

Finally, most existing systems operate on static snapshots of social networks and fail to handle concept drift effectively. Attack strategies evolve rapidly, requiring adaptive models that can learn from changing behavior patterns without frequent retraining. Addressing these gaps motivates the proposed dual GNN-based framework.

IV. METHODOLOGY

This section outlines the overall methodology used for anomalous behavior detection in social networks. The approach models the platform as a heterogeneous graph to capture both individual user characteristics and collective interaction patterns. It combines multi-modal feature extraction, graph construction, and a dual Graph Neural Network (GNN) framework to detect both known and previously unseen malicious entities.

The pipeline begins with data preprocessing, followed by semantic feature extraction using transformer-based models. A graph structure is then built to represent relational dependencies among users. Finally, supervised and unsupervised GNN models operate in parallel to perform bot classification and anomaly detection. This modular design makes the system scalable, adaptable, and robust against evolving attack strategies.

A. Data Ingestion and Preprocessing

The proposed system utilizes the TwiBot-22 dataset, which provides a comprehensive benchmark for social network anomaly detection. The dataset includes user profile information, social relationships, and textual content such as tweets and user descriptions. These heterogeneous data sources enable a realistic modeling of user behavior in modern social platforms.

During preprocessing, numerical attributes such as follower count, following count, and tweet frequency are normalized to reduce scale variations. Categorical attributes, including account verification status and profile flags, are encoded into suitable numerical representations. Missing or inconsistent entries are handled to ensure data integrity. This preprocessing stage ensures that the input data is well-structured and suitable for downstream feature extraction and graph-based learning.

B. Multi-Modal Feature Engineering using RoBERTa

To capture semantic behavior, textual data from user descriptions and tweets is processed using a pre-trained RoBERTa model. The transformer generates contextual embeddings that help the system recognize subtle linguistic patterns often used by malicious accounts.

These text embeddings are combined with numerical and categorical features to create a unified feature vector for each user. This multi-modal representation allows the model to analyze activity patterns, profile credibility, and content intent together, reducing dependence on manual rules and improving resilience against content-based manipulation.

C. Graph Construction and Representation

The social network is modeled as a heterogeneous directed graph, where users are represented as nodes and interactions such as follows, retweets, and mentions are represented as edges. This structure preserves relational dependencies and enables the detection of coordinated behavior that cannot be identified through isolated user analysis.

Feature vectors generated during preprocessing are assigned to their respective nodes, and edge directionality is maintained to reflect asymmetric social relationships. The graph is implemented using PyTorch Geometric to support efficient message passing and scalable Graph Neural Network operations.

D. Supervised Bot and Human Detection using GAT

The first component of the detection framework focuses on identifying known malicious accounts. A Graph Attention Network (GAT) is trained on labeled data to classify users as bots or humans. Unlike traditional classifiers, the GAT considers both user features and their relational context within the network. Through its attention mechanism, the model assigns greater importance to influential neighbors, enabling it to detect coordinated bot behavior more effectively. This

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 05 , May - 2026

graph-aware learning improves robustness against organized campaigns that operate through interconnected accounts.

E. Unsupervised Anomaly and Community Detection using GAE

The second component focuses on detecting unseen or evolving threats. A Graph Autoencoder (GAE) is trained to learn normal human interaction patterns within the network. By reconstructing the graph structure, the model assigns anomaly scores based on reconstruction error, flagging users whose behavior significantly deviates from typical patterns.

To identify coordinated activity, Louvain community detection is used to cluster densely connected users. Communities with unusually high average anomaly scores are marked as potential coordinated groups rather than isolated anomalies. For better interpretability, the semantic content of flagged communities is analyzed through word clouds and large language model (LLM) summaries. This helps provide a clear understanding of the intent and behavior behind suspicious clusters, making the framework both robust and explainable.

F. System Architecture

The overall workflow of the proposed anomalous behavior detection framework is illustrated in Fig. 1. The architecture highlights the integration of multi-modal feature extraction, graph construction, and the dual GNN-based detection engine.

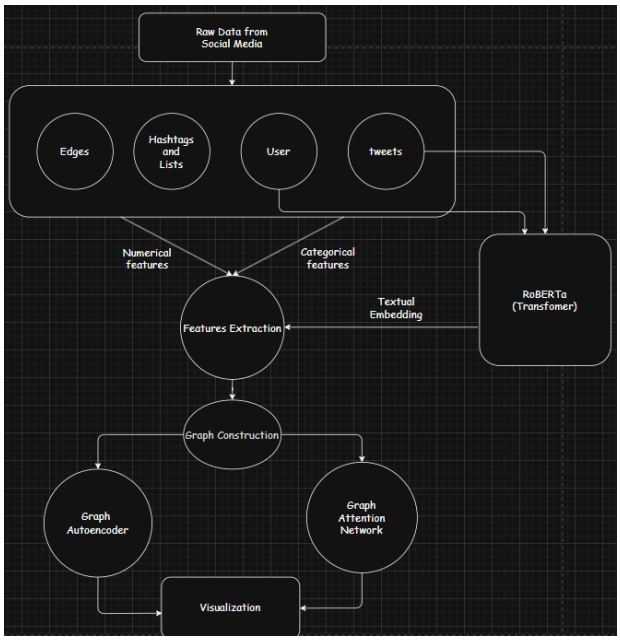

Fig. 1: Overall architecture of the proposed dual GNN-based anomalous behavior detection system.

As shown in Fig. 1, preprocessed user data is first transformed into multi-modal node features and embedded into a graph structure. The supervised GAT and unsupervised GAE operate on this graph in parallel, producing classification labels and anomaly scores respectively. The outputs are then aggregated to provide a comprehensive assessment of malicious behavior within the social network.

V. RESULTS AND DISCUSSION

The effectiveness of the proposed framework is evaluated through qualitative visualization and behavioral analysis of detected communities.

To assess representation quality, t-SNE is applied to both raw node features and GNN-generated embeddings. Fig. 2 shows the visualization of original node features, where bots and humans exhibit significant overlap. This indicates that raw attributes alone are insufficient for clear separation.

In contrast, Fig. 3 presents the t-SNE visualization after GNN processing. A noticeably improved separation between bots and humans can be observed. This demonstrates that graph-based learning successfully captures relational dependencies and coordinated interaction patterns that are not visible in isolated feature space.

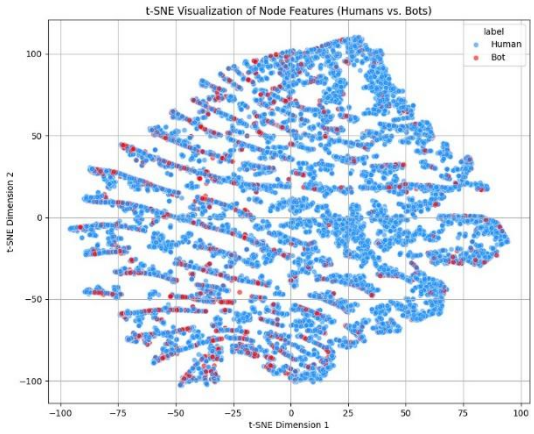

Fig. 2: t-SNE visualization of raw node features showing overlap between bots and humans.

Beyond individual detection, community-level analysis further validates the framework. Fig. 4 illustrates the behavioral footprint and semantic profiling of a detected anomalous community. The radar chart highlights abnormal activity patterns compared to the human baseline, while the word cloud summarizes dominant semantic themes within the cluster.

This combined structural and semantic analysis confirms that the framework not only detects anomalous accounts but also identifies coordinated groups with shared behavioral and content characteristics.

Overall, the results demonstrate that integrating graph-based learning, community detection, and semantic profiling significantly enhances detection robustness compared to feature-only approaches.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 05 , May - 2026

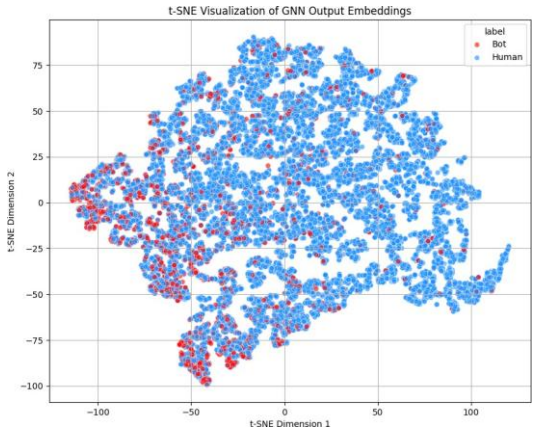

Fig. 3: t-SNE visualization of GNN output embeddings show-ng improved separation between bots and humans.

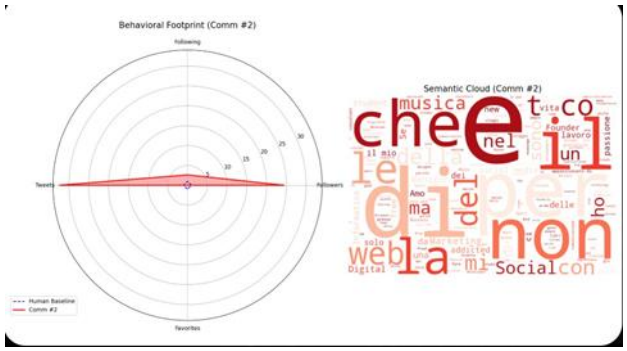

Fig. 4: Behavioral and semantic analysis of a detected anomalous community.

VI. CONCLUSION

This paper presented a graph-based anomalous behavior detection framework designed to identify malicious and coordinated activity in social networks. By modeling users and their interactions as a relational graph, the system enables structural analysis of collective behavior rather than relying solely on isolated user attributes. The integration of supervised Graph Attention Networks (GAT) and unsupervised Graph Autoencoders (GAE) allows the framework to detect both known bot accounts and previously unseen anomalous patterns within the network.

The evaluation demonstrated that graph-based embeddings significantly improve separability between malicious and legitimate users compared to raw feature representations. In addition, community-level analysis using Louvain clustering and statistical thresholding enables the identification of coordinated groups rather than isolated anomalies. The incorporation of semantic profiling further enhances interpretability, providing meaningful insights into the behavioral intent of suspicious clusters.

Despite these strengths, certain challenges remain. The effectiveness of graph-based models depends on data quality and completeness of relational information. Highly adaptive adversaries may attempt to mimic normal structural patterns, requiring continuous refinement of detection strategies. Furthermore, maintaining scalability in extremely large dynamic networks demands efficient optimization and resource management.

In conclusion, the proposed framework demonstrates that combining structural graph learning, anomaly detection, and semantic analysis can significantly strengthen social network security. By integrating relational intelligence with interpretable community profiling, the system offers a practical and scalable approach to detecting modern coordinated threats in large-scale social environments.

ACKNOWLEDGMENT

We wish to extend our sincere thanks to Dr. Aparna Bannore, our respected Head of Department, for her support, insightful guidance, and continual encouragement during the duration of the project. Additionally, we express our heartfelt gratitude to all the faculty members of the Computer Engineering Department for their essential contributions, prompt help, and unwavering support.

This paper was reviewed and refined with the assistance of AI tools for grammar checking and formatting alignment. The core research and experimental analysis were conducted by the authors.

REFERENCES

[1] E. Alatawi and U. Albalawi, "Harnessing AI for cyber defense: Honeypot-driven intrusion detection systems," *Symmetry*, vol. 17, no. 5, pp. 1–18, 2025.

[2] Z. Moric´, V. akic´, and D. Regvart, "Advancing cybersecurity with honeypots and deception strategies," *Informatics*, vol. 12, no. 1, pp. 1–22, 2025.

[3] G. Wagener, R. State, A. Dulaunoy, and T. Engel, "Self-adaptive high-interaction honeypots driven by game theory," in *Proc. Int. Symp. Self-Stabilizing Systems*, Springer, 2009, pp. 741–755.

[4] S. Machmeier, "Honeypot implementation in a cloud environment," *arXiv preprint arXiv:2301.00710*, 2023.

[5] Z. Ding and M. Fei, "An anomaly detection approach based on isolation forest algorithm for streaming data using sliding window," *IFAC Proceedings Volumes*, vol. 46, no. 20, pp. 12–17, 2013.

[6] M. Heigl, K. A. Anand, A. Urmann, D. Fiala, M. Schramm, and R. Hable, "On the improvement of the isolation forest algorithm for outlier detection with streaming data," *Electronics*, vol. 10, no. 13, pp. 1–20, 2021.

[7] J. J. Liu, G. W. Cassales, F. T. Liu, B. Pfahringer, and A. Bifet, "Streaming isolation forest," in *Proc. Pacific-Asia Conf. Knowledge Discovery and Data Mining*, Springer, 2025, pp. 95–107.

[8] F. Moomtaheen, S. S. Bagui, S. C. Bagui, and D. Mink, "Extended isolation forest for intrusion detection in Zeek data," *Information*, vol. 15, no. 7, pp. 1–18, 2024.

[9] A. Javadpour, F. Ja'fari, T. Taleb, M. Shojafar, and C. Benza¨ıd, "A comprehensive survey on cyber deception techniques to improve honeypot performance," *Computers & Security*, vol. 140, pp. 1–28, 2024.

[10] S. Feng, H. Wan, N. Wang, J. Li, and M. Wang, "TwiBot-22: Towards graph-based Twitter bot detection," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2022, pp. 1–15.