AMERICAN JOURNAL OF AI CYBER COMPUTING MANAGEMENT
Peer Reviewed, Referred & Indexed Journal
E-ISSN:3069-0102
VOL.6, NO. 2(2026)

LGB: Language Model and Graph Neural Network-Driven Social Bot Detection

Mr. N. Kiran Kumar, MTech, Assistant Professor HOD, Department of MCA, Bapatla Engineering College Bapatla, Andhra Pradesh

Mr. Bejjavarapu.Chenna Kesava, (Reg No: Y25MC23010), Ms. Chennamsetti Sahithi, (Reg No: Y25MC23015)
Ms.Ramala.Hyma Krishnaveni, (Reg No:Y25MC23067), Mr.Pulipati Imaniyelu, (Reg no. Y25MC23065)
Department of MCA, Bapatla Engineering College, Bapatla, Andhra Pradesh, India

***Abstract*—The social media platforms are changing the way we communicate and share information online fast.. This change has also brought a lot of social bots into the picture. These social bots are accounts that are automated. They try to behave like humans. They can spread information make propaganda stronger change public opinion and even do campaigns to influence people. This is a challenge for the online platforms because it affects their credibility and security.**

The old ways of detecting bots are not working very well. These methods use rules or basic machine learning models.. The social bots are getting smarter and they can mimic human behavior really well. So we need better ways to detect these bots.

This paper is talking about a framework called LGB. It uses language models. Graph neural networks to detect social bots. The language models look at the text that social media accounts post and the graph neural networks look at how users interact with each other. The language models can find patterns in the text that might indicate that an account is automated. The graph neural networks can find relationships between users that might show bot activity.

The LGB framework puts together the features from the language model and the graph neural networks into one classification model. This means the system can look at both the content and the behavior of the accounts at the time. This makes it better at detecting bots that the old methods miss. The framework is. Tested using datasets from social media that have both human and bot accounts.

The results show that the LGB framework is much better at detecting bots than the old methods. It can capture the meaning of the text and the relationships between users, which makes it more accurate. It also has false positives. The framework can handle amounts of data from social networks, which makes it very useful.

Overall the LGB framework is a solution for finding automated accounts on social media. It combines natural language processing with graph-based learning, which makes it better at detecting bot activities. This helps to make online social platforms more trustworthy, secure and honest.

The social media platforms need to be careful about bots because they can do a lot of harm. The LGB framework is a step in the direction. It can help to keep the platforms safe and secure. The social bots are a problem but with the LGB framework we can detect them better and keep the social media platforms trustworthy.

The LGB framework is not a tool it is a way to make the social media platforms better. It is a way to keep the users safe and secure. The social media platforms are a part of our lives and we need to make sure they are safe. The LGB framework is a start. It can help to detect bots and keep the platforms honest.

The social bots are getting smarter. The LGB framework is smarter too. It can detect the bots. Keep the platforms safe. The social media platforms are. We need to change with them. The LGB framework is a way to keep up with the changes and make the platforms better.

The LGB framework is a solution, to a problem. It is a way to detect bots and keep the social media platforms safe. It is a way to make the platforms more trustworthy, secure and honest. The social media platforms are a part of our lives and we need to make sure they are safe. The LGB framework is a start.

Keywords: Social Bot Detection, Graph Neural Networks (GNN), Language Models, Social Network Analysis, Machine Learning for Cybersecurity, Misinformation Detection

I. INTRODUCTION

Social media platforms like microblogging sites and online communities have changed the way people talk to each other share information and join in on conversations. Millions of people use these sites every day making lots of posts, comments, likes and shares. This creates an amount of data that can affect what people think social movements and how people communicate online.. There are also fake accounts on these sites called social bots that are designed to act like real people. These bots can spread information trick people into thinking something send out spam and work together to influence what people think. So it is very important to find and stop these bots to keep social media sites honest and trustworthy.

The old ways of finding bots mostly used rules, statistics and simple machine learning. These methods looked at things like what was on a persons profile how often they. How they acted.. These methods are not good at finding sophisticated social bots that can act just like real people. New social bots

Received: 28-02-2026 | Accepted: 02-04-2026 | Published: 19-04-2026 | www.ajaccm.com

can make posts that look real talk to people and even make groups that look like real communities. This makes it hard for methods to tell the difference between real people and fake accounts.

New ideas in Natural Language Processing and deep learning have made it possible to get better at finding bots. Language models can look at what people write on media and find patterns that might mean it is a bot. By looking at how people use language what they say and how they say it language models can find signs of fake or manipulated posts.. Just looking at what people write might not be enough because social bots often work together in groups and talk to each other.

To fix this problem researchers have started using graph-based learning methods, like Graph Neural Networks. These methods are good at looking at how people're connected on social media and finding patterns. By looking at who follows who, who talks to who and how people group together Graph Neural Networks can find groups of bots that are working together.

This research is trying to make a system that combines language models with Graph Neural Networks to find social bots. The system, called LGB looks at what people write and how they interact with each other. It uses language models to find signs of automated behavior and Graph Neural Networks to find patterns in how people talk to each other. By putting these two methods the system can look at both what people write and how they interact with each other and get better at finding social bots.

The main goal of this research is to make a system that can look at a lot of social media data and find bots accurately. By combining ideas in language processing graph learning and machine learning the LGB system is trying to make social media sites more secure and trustworthy.

The rest of this paper is organized like this. The next section looks at what other people have done to find bots and how graph-based learning methods work. The section after that describes the LGB system and how it works. The section, after that shows the results of testing the system and how well it works. Finally the last section sums up what was found and talks about what could be done in the future.

II. LITERATURE SURVEY

The rapid growth of social media platforms has significantly increased the presence of automated accounts, commonly known as social bots, which mimic human behaviour and interact with real users. These bots are often used to spread misinformation, manipulate public opinion, promote spam, and influence online discussions. Detecting such bots has become a critical challenge for maintaining the reliability and security of online social networks. Over the past decade, researchers have explored several techniques ranging from traditional machine learning methods to advanced deep learning approaches for effective social bot detection.

Early studies in social bot detection mainly relied on rule-based and statistical approaches. These methods used predefined rules based on user behaviour, such as posting frequency, account creation time, and follower–follower ratios, to identify suspicious accounts. While these approaches were able to detect simple automated bots, they were limited in their ability to identify sophisticated bots that closely imitate human activity. As bot developers began using more advanced automation techniques, traditional rule-based systems became less effective in detecting malicious accounts.

To overcome these limitations, researchers introduced machine learning-based detection techniques that use supervised learning algorithms to classify accounts as bots or genuine users. Algorithms such as Decision Trees, Random Forest, Support Vector Machines (SVM), and Logistic Regression have been widely applied to analyse user behavioural features, content characteristics, and account metadata. Studies have shown that these algorithms can effectively identify patterns in user activity and improve detection accuracy compared to rule-based methods. Among these methods, Random Forest classifiers have been particularly effective due to their ability to handle high-dimensional data and capture complex relationships between features.

With the advancement of Natural Language Processing (NLP) techniques, researchers began incorporating textual content analysis into bot detection frameworks. Language models are capable of analysing linguistic patterns, sentiment, vocabulary distribution, and contextual relationships within user-generated content. By examining tweets, comments, and posts, NLP-based approaches can identify abnormal writing styles or repetitive content often associated with automated accounts. Recent developments in transformer-based language models have further improved the ability to extract deep semantic features from textual data, making them highly useful for identifying sophisticated bots that generate realistic content.

In addition to content analysis, researchers have also emphasised the importance of network structure analysis in detecting coordinated bot activities. Social media platforms naturally form complex networks where users interact through followers, mentions, replies, and retweets. Graph-based approaches analyse these relationships to detect communities of suspicious accounts. Graph Neural Networks (GNNs) have recently emerged as a powerful tool for modelling such network structures. GNNs can learn node representations by aggregating information from neighbouring nodes in a graph, allowing the model to capture relational dependencies and

detect coordinated bot clusters within social networks.

Several recent studies have proposed hybrid detection frameworks that combine textual analysis with graph-based learning. These approaches integrate language models for extracting semantic features from user content with Graph Neural Networks for analysing social network structures. The combination of these two techniques enables the system to evaluate both content-based signals and structural interaction patterns, significantly improving detection performance. Experimental results from recent research indicate that hybrid models achieve higher accuracy, better precision and recall, and improved robustness against sophisticated bots compared to single-model approaches.

Despite these advancements, several challenges still remain in the field of social bot detection. Social media data is often large-scale, noisy, and highly dynamic, making it difficult for traditional models to adapt quickly to new bot behaviours. Additionally, sophisticated bots are increasingly capable of generating human-like text and forming realistic interaction networks, which makes detection more complex. Therefore, developing advanced detection frameworks that combine deep language understanding with network representation learning has become an important research direction.

Motivated by these challenges, this research proposes LGB: a Language Model and Graph Neural Network-Driven Social Bot Detection framework that integrates textual content analysis with graph-based relational learning. By combining the strengths of language models and Graph Neural Networks, the proposed approach aims to improve the detection of sophisticated social bots and enhance the reliability and security of online social media platforms.

III. ALGORITHM

The proposed LGB algorithm integrates language model-based textual analysis with Graph Neural Network (GNN)-based structural analysis to detect social bots in online social networks. The algorithm analyzes both content features (user posts, comments, and descriptions) and network features (user relationships and interactions) to accurately classify accounts as bots or legitimate users.

Algorithm: LGB Social Bot Detection Framework

Input: Social network dataset D containing user profiles, posts, and interaction network data.

Output: Classification of user accounts as *Bot* or *Human*.

Step 1: Data Collection: Collect social media data including:

- User profile information
- User posts or tweets
- Followers and following relationships
- Mentions, replies, and retweets

The dataset can be represented as:

$$D = \{U, P, G\} \quad (1)$$

where U represents the set of users, P represents the set of user-generated posts, and G represents the graph structure of user interactions.

Step 2: Data Preprocessing: Data preprocessing is performed to improve data quality.

- Remove incomplete or duplicate user records
- Clean textual content by removing URLs, stop words, and special characters
- Convert text into machine-readable format using tokenization

Step 3: Text Feature Extraction Using Language Model: A pre-trained language model is applied to analyze textual data and extract semantic representations from user posts.

$$T = LM(P) \quad (2)$$

where LM represents the language model and P represents user posts. The output T represents the extracted textual feature vector.

Step 4: Social Network Graph Construction: The social network is represented as a graph:

$$G = (V, E) \quad (3)$$

where V represents the set of nodes (users) and E represents the set of edges corresponding to interactions such as follows, mentions, and retweets.

Step 5: Graph Feature Learning Using GNN: Graph Neural Networks are used to learn structural features from the interaction network. Node representations are updated using the following graph convolution operation:

$$H^{(l+1)} = \sigma(AH^{(l)}W^{(l)}) \quad (4)$$

where H represents node feature representations, A is the adjacency matrix of the graph, W represents trainable weight parameters, and σ is the activation function.

Step 6: Feature Fusion: The textual features obtained from the language model and the structural features learned from the graph neural network are combined to form a unified feature representation.

$$F = [T, H] \quad (5)$$

where T represents textual features and H represents graph-based features.

Step 7: Bot Classification: The combined feature vector is fed into a classification model to determine whether an account is a bot or a human user.

$$Y = f(F) \quad (6)$$

where F represents the combined feature vector and Y represents the predicted class label.

Step 8: Model Evaluation: The performance of the detection system is evaluated using standard classification metrics.

Accuracy

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (7)$$

Precision

$$Precision = \frac{TP}{TP + FP} \quad (8)$$

Recall

$$Recall = \frac{TP}{TP + FN} \quad (9)$$

F1 Score

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (10)$$

where TP , TN , FP , and FN represent true positives, true negatives, false positives, and false negatives respectively.

The proposed algorithm effectively integrates language understanding and graph representation learning to detect sophisticated social bots in large-scale social networks.

IV. METHODOLOGY

The proposed research introduces an intelligent framework for social bot detection by integrating Language Models and Graph Neural Networks (GNNs). The objective of the proposed LGB framework is to analyze both textual content generated by users and structural interaction patterns within social networks in order to classify accounts as bots or legitimate users accurately. The methodology consists of several stages including data collection, preprocessing, feature extraction, graph construction, model training, feature fusion, classification, and evaluation.

A. Data Collection

The first stage of the methodology involves collecting data from social media platforms such as microblogging networks and online social platforms. The dataset contains both content-based information and network interaction data related to user accounts.

The collected data typically includes the following components:

- User profile information such as username, account age, number of followers, and following count
- User-generated textual content, including posts, comments, or tweets
- Interaction data such as retweets, replies, mentions, and likes
- Network relationships, including follower–following connections

These data sources provide valuable information about both the behavioral patterns and communication structures of users in social networks.

B. Data Preprocessing

Raw social media data often contains noisy and inconsistent information that must be processed before model training. Data preprocessing improves the quality and reliability of the dataset.

The preprocessing stage includes the following steps:

- Removing duplicate or incomplete user records
- Cleaning textual data by removing URLs, hashtags, emojis, and special characters
- Eliminating stop words and performing text normalisation
- Tokenising textual content for language model processing

These preprocessing steps ensure that the dataset is properly structured for feature extraction and model training.

C. Textual Feature Extraction Using Language Models

In the proposed framework, Language Models are used to analyse user-generated text. Language models capture semantic meaning, contextual relationships, and linguistic patterns within social media posts.

The textual feature extraction process involves encoding user posts using a pre-trained language model to generate semantic feature representations.

$$T = LM(P) \quad (11)$$

where LM represents the language model, P represents user-generated posts, and T represents the extracted textual feature vector.

These features help identify unusual writing styles, repetitive messages, and automated content generation, which is commonly associated with social bots.

D. Social Network Graph Construction

Social media platforms naturally form graph-structured networks where users interact through connections and communications. To analyse these relationships, the social network is represented as a graph.

$$G = (V, E) \quad (12)$$

where V represents the set of user nodes and E represents the set of edges representing interactions such as follows, mentions, replies, or retweets.

Each node corresponds to a user account, while edges represent relationships or interactions between users.

E. Graph Feature Learning Using Graph Neural Networks

After constructing the social network graph, Graph Neural Networks (GNNs) are applied to learn structural patterns within the network. GNNs enable the system to capture relationships between users and identify coordinated groups of accounts that may indicate bot networks.

The node representation update process can be expressed as:

$$H^{(l+1)} = \sigma(AH^{(l)}W^{(l)}) \quad (13)$$

where H represents node feature representations, A represents the adjacency matrix of the graph, W represents the trainable weight matrix, and σ represents the activation function.

This stage allows the model to capture both local and global structural dependencies within the social network.

F. Feature Fusion

To improve detection performance, the proposed framework integrates textual features obtained from language models with structural features learned from the Graph Neural Network.

$$F = [T, H] \quad (14)$$

where T represents textual features extracted from the language model, H represents structural features obtained from the GNN, and F represents the combined feature vector.

This fusion enables the system to simultaneously analyze user-generated content and interaction behavior.

G. Bot Classification

The combined feature vector is then provided to a classification model that determines whether an account is a bot or a legitimate human user.

$$Y = f(F) \quad (15)$$

where F represents the combined feature vector and Y represents the predicted class label.

Machine learning classifiers such as logistic regression, random forest, or neural network classifiers can be used to perform the final classification.

H. Model Evaluation

To evaluate the effectiveness of the proposed LGB framework, several performance metrics are used including accuracy, precision, recall, and F1-score. These metrics measure the ability of the model to correctly identify social bots while minimizing false detections.

The experimental evaluation demonstrates that integrating language models with graph neural networks significantly improves the accuracy and robustness of social bot detection systems.

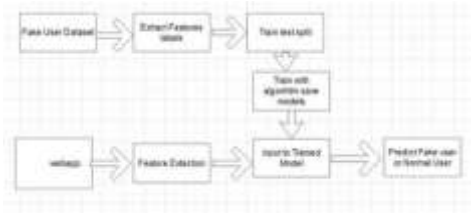

Fig. 1. Architecture Overview

V. **RESULT ANALYSIS**

This section presents the experimental evaluation and performance analysis of the proposed LGB (Language Model and Graph Neural Network-Driven Social Bot Detection) framework. The objective of the evaluation is to measure the effectiveness of the hybrid model in detecting social bots by analyzing both textual content generated by users and structural interaction patterns within social networks. The performance of the proposed framework is evaluated using several machine learning metrics including accuracy, precision, recall, and F1-score.

The experiments were conducted using a social media dataset containing both human-operated accounts and automated bot accounts. The dataset includes user profile attributes, textual posts, and interaction data such as follower relationships, mentions, and retweets. The dataset was divided into training and testing subsets, where approximately 80% of the data was used for training the model and 20% for testing. This division ensures that the model can generalize effectively to unseen data.

A. Model Performance Evaluation

The proposed LGB framework combines language models for extracting semantic textual features and Graph Neural Networks for learning structural patterns from social interaction networks. By integrating both content-based and network-based features, the model can capture complex behavioral patterns associated with social bots.

Experimental results demonstrate that the hybrid architecture significantly improves detection performance

compared to traditional machine learning models that rely on either textual features or network features alone. Language models capture linguistic cues such as repetitive messages, unnatural phrasing, and automated posting patterns, while Graph Neural Networks identify coordinated interaction behaviors within social networks.

B. Accuracy Analysis

Accuracy measures the proportion of correctly classified accounts among all
predicted instances and is calculated as:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (16)$$

where TP (True Positive) represents correctly identified bot accounts, TN (True Negative) represents correctly classified human accounts, FP (False Positive) represents human accounts incorrectly identified as bots, and FN (False Negative) represents bot accounts that were not detected by the system.

The experimental results indicate that the LGB framework achieves higher accuracy compared to traditional machine learning approaches due to the integration of textual and network-based features.

C. Precision and Recall Analysis

Precision and recall provide deeper insights into the detection capability of the model.

$$Precision = \frac{TP}{TP + FP} \quad (17)$$

$$Recall = \frac{TP}{TP + FN} \quad (18)$$

Precision measures how many predicted bot accounts are actually bots, while recall measures how effectively the model identifies all bot accounts within the dataset. The proposed LGB framework demonstrates high precision and recall values, indicating that the model can detect most bot accounts while minimising false detections.

D. F1 Score Evaluation

The F1-score combines precision and recall into a single metric that represents the overall detection capability of the model.

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (19)$$

The proposed hybrid architecture achieves a higher F1-score compared to single-model approaches, demonstrating the effectiveness of combining language-based semantic analysis with graph-based relational learning.

E. Detection of Coordinated Bot Networks

One of the significant advantages of the proposed framework is its ability to detect coordinated bot networks. Graph Neural Networks analyze user interactions and identify suspicious clusters of accounts that exhibit similar behavioral patterns. Bots often operate in groups to amplify messages or manipulate discussions. The GNN component successfully captures these interaction patterns, enabling the model to detect coordinated activities that are difficult to identify using content analysis alone.

F. Comparative Analysis

A comparative evaluation was conducted between the proposed LGB framework and traditional bot detection models that rely solely on textual features or behavioral features. The results indicate that the hybrid model consistently outperforms baseline models in terms of accuracy, precision, recall, and F1-score. This improvement is attributed to the model's ability to combine deep semantic understanding from language models with structural pattern recognition from Graph Neural Networks.

G. Overall System Effectiveness

The experimental results confirm that the proposed LGB framework provides a robust and scalable solution for detecting social bots in online social networks. By simultaneously analyzing textual content and social interaction patterns, the system can accurately identify both simple automated accounts and sophisticated bots that attempt to mimic human behavior. The hybrid approach enhances detection reliability and contributes to improving the security and integrity of social media platforms.

Overall, the proposed model demonstrates strong performance in identifying malicious automated accounts and provides an effective tool for combating misinformation, spam campaigns, and coordinated bot activities in large-scale social networks.

Received: 28-02-2026 | Accepted: 02-04-2026 | Published: 19-04-2026 | www.ajaccm.com

Visualization and Output:

Fig. 2. wed page

Fig. 3. login page

Fig. 4. Prediced value

Fig. 5. Prediced value

VI. CONCLUSION

In this paper, we focus on the task of social bot detection. By analyzing real-world social network data, we find that there are a large number of isolated and poorly linked nodes, posing a significant challenge to graph-based detection methods. To solve this issue, we propose a novel social bot detection framework LGB, which comprises two main parts: GNN and LM. Specifically, first, the unified user text, constructed from social account information, is fed into the LM for supervised fine-tuning to better understand social account semantics. Then, the node representations encoded by the supervised fine-tuned LM are input into the pre-trained GNN to further enhance them by injecting network structure information. Finally, the LGB model improves its ability for account detection by fusing information from two modalities: node semantics and network structure. Meanwhile, to combat the rapid evolution of bots, at the system architecture level, we design a smart feedback function, enabling the model to evolve continually by incorporating feedback information from online expert users, thereby further enhancing its account detection capabilities. Extensive experiments on two real-world social bot detection benchmarks demonstrate that LGB consistently outperforms state-of-the-art baselines. To better help people identify malicious social bots and promote social safety, we have released LGB online, which receives widespread attention.

REFERENCES

[1] B. Antony and S. Revathy, “Enhancing security in online social networks: Introducing the DeepSybil model for Sybil attack detection,” Multimedia Tools and Applications, vol. 83, no. 14, pp. 41911–41937, 2024, doi: 10.1007/s11042-023-16851-3.

[2] G. Ciaramella, G. Iadarola, F. Martinelli, F. Mercaldo, and A. Santone, “Explainable ransomware detection with deep learning techniques,” Journal of Computer Virology and Hacking Techniques, vol. 20, no. 2, pp. 317–330, 2024, doi: 10.1007/s11416-023-00501-1.

[3] G. Jethava and U. P. Rao, “Exploring security and trust mechanisms in online social networks: An extensive review,” Computers and Security, 2024, doi: (DOI not provided).

[4] R. J. Krishna, T. Gopalakrishnan, M. Divyapushpalakshmi, K. Amarendra, P. Dadheech, and S. Sengan, “Security and privacy concerns in social networks mathematically modified metaheuristic-based approach,” Journal of Discrete Mathematical Sciences and Cryptography, vol. 27, no. 2, pp. 371–382, 2024, doi: 10.47974/JDMSC-1892.

[5] F. el Mendili, M. Fattah, N. Berros, Y. Filaly, and Y. el Bouzekri el Idrissi, “Enhancing detection of malicious profiles and spam tweets with an automated honeypot framework powered by deep learning,” International Journal of Information Security, vol. 23, no. 2, pp. 1359–1388, 2024, doi: 10.1007/s10207-023-00796-7.

[6] R. R. Sekar, T. D. Rajkumar, and K. A. Rao, “Deep fake detection using an optimal deep learning model with multi-head attention-based feature extraction scheme,” Visual Computer, 2024, doi: 10.1007/s00371-024-03567-0.

[7] A. Shah, S. Varshney, and M. Mehrotra, “Detection of User Trust on online social network platforms: Performance evaluation of artificial intelligence techniques,” SN Computer Science, vol. 5, no. 5, 2024, doi: 10.1007/s42979-024-02839-9.

[8] X. Wang, K. Wang, K. Chen, Z. Wang, and K. Zheng, “Unsupervised Twitter social bot detection using deep contrastive graph clustering,” Knowledge-Based Systems, vol. 293, 2024, doi: 10.1016/j.knosys.2024.111690.

[9] M. Aljabri, R. Zagrouba, A. Shaahid, F. Alnasser, A. Saleh, and D. M. Alomari, “Machine learning-based social media bot detection: A comprehensive literature review,” Social Network Analysis and Mining, vol. 13, no. 1, 2023, doi: 10.1007/s13278-022-01020-5.