International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

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

[1] Dr.S.Satyanarayana, [2] CH.Vamshi ,[3] D.Srimanth Reddy ,[4] D.Karthikeya,[5] G.Sai Deekshitha

[1] Professor, Department of Computer Science & Engineering (Artificial Intelligence & Machine Learning), Malla Reddy University, Kompally, Hyderabad. [1] Email : drssatyanarayana@mallareddyuniversity.ac.in

[2,3,4,5] Students, Department of Computer Science & Engineering (Artificial Intelligence & Machine Learning), Malla Reddy University, Kompally, Hyderabad. [2] Email : chvamshi482@gmail.com, [3] Email: damerasrimanthreddy@gmail.com , [4] Email: dondapatikarthikeya.jee.44@gmail.com, [5] Email: deekshithagelli823@gmail.com

Abstract:

LGB presents a novel hybrid framework that leverages fine-tuned language models for semantic analysis of user profiles, tweets, and descriptions alongside graph neural networks pretrained through contrastive learning on social graphs to detect social bots with high accuracy. This multimodal approach effectively captures both content-based linguistic patterns and structural network topologies, overcoming limitations of traditional unimodal methods in identifying isolated or sparsely connected bots. Extensive evaluation on benchmark datasets like TwiBot-20 and TwiBot-22 demonstrates superior performance in precision, recall, and F1-score compared to state-of-the-art baselines, enhanced by a feedback mechanism for real-time adaptability.

Keywords: Social Bot Detection, LightGBM (LGB), Large Language Models (LLM), Graph Neural Networks (GNN), Social Network Analysis, Fake Account Detection, Misinformation Detection, Feature Fusion, Node Classification.

I.INTRODUCTION

With the rapid growth of social media platforms such as Twitter (X), Facebook, Instagram, and Reddit, online communication has become faster and more accessible than ever before. However, this digital expansion has also led to the emergence of social bots—automated or semi-automated accounts designed to mimic human behavior. These bots can spread misinformation, manipulate public opinion, promote spam campaigns, influence elections, and conduct coordinated cyber-attacks. As social networks become central to communication, commerce, and governance, detecting and mitigating social bots has become a critical challenge in cybersecurity and data science. Traditional bot detection techniques primarily relied on rule-based systems and handcrafted features such as posting frequency, account age, follower–following ratio, and content similarity. While these approaches were effective against simple bots, modern bots have evolved significantly. Advanced bots now use artificial intelligence to generate human-like text, maintain realistic interaction patterns, and participate in complex network structures. As a result, conventional

detection systems struggle to distinguish between genuine users and sophisticated automated accounts. To address these limitations, researchers are increasingly adopting hybrid deep learning frameworks that combine textual intelligence with structural network analysis. In this context, the proposed LGB: Language Model and Graph Neural Network-Driven Social Bot Detection framework integrates the strengths of Large Language Models (LLMs) and Graph Neural Networks (GNNs) to enhance detection accuracy and robustness.

Large Language Models (LLMs) are capable of understanding contextual meaning, semantic relationships, sentiment, and linguistic patterns in user-generated content. By analyzing tweets, posts, comments, and profile descriptions, LLMs can detect subtle textual anomalies, repetitive messaging behavior, unnatural phrasing, and coordinated narrative propagation often associated with bots. Unlike traditional text classification methods, LLMs capture deep contextual embeddings that improve classification performance even when bots generate sophisticated, human-like content. While textual analysis is powerful, it alone is insufficient because bots often operate in coordinated networks. This is where Graph Neural Networks (GNNs) play a crucial role. Social media platforms naturally form graph structures where users are nodes and interactions (followers, mentions, retweets, likes) are edges. GNNs leverage this relational information to identify suspicious connectivity patterns such as tightly connected bot clusters, abnormal interaction density, and synchronized activity. By learning node embeddings based on both local and global graph structures, GNNs can detect coordinated bot campaigns that are not easily identifiable through content analysis alone.

The integration of language models and graph neural networks provides a comprehensive detection mechanism that considers both what users say (content-level intelligence) and how users interact (network-level intelligence). Furthermore, the inclusion of LightGBM (LGB) as a classification layer enhances computational efficiency and scalability. LightGBM is a gradient boosting framework known for its high performance, faster training speed, and ability to handle large-scale datasets. By combining features extracted from LLM embeddings and GNN representations, LightGBM can perform accurate final classification of accounts as bots or genuine users.

II.LITERATURE SURVEY

1.Social Media Bot Detection Research: Review of Literature

Authors: B. Rodič

Abstract: This study presents a review of research on social media bot detection, examining recent publications from five bibliographical databases yielding 534 papers. It analyzes statistical trends, introduces bot research evolution, identifies issues like bot concealment techniques and methodological flaws, and overviews detection methods categorized by methodology, concluding with recent trends in

bot development and countermeasures.

2.CB-MTE: Social Bot Detection via Multi-Source Heterogeneous Feature Fusion

Authors: M. Cheng et al.

Abstract: Social bots distort public perception through mimicry and coordination; traditional single-source methods fail against dynamic behaviors. CB-MTE proposes a hierarchical framework fusing user metadata portraits, DistilBERT text semantics, and community graph embeddings with manifold reduction and CatBoost reasoning. It outperforms baselines on TwiBot-22 by capturing complete bot characteristics.

3.Detecting Social Bots on Twitter: A Literature Review

Authors: E. Alothali et al.

Abstract: Reviews detection schemes on Twitter, examining classifiers, datasets, and features used across studies. Highlights common practices and gaps in current approaches for identifying automated accounts.

4. Systematic Literature Review of Social Media Bots Recognition

Authors: Z. Ellaky et al.

Abstract: SLR covering 2008-2022 research determines best practices in SMB detection, analyzing techniques, challenges, and performance metrics for recognizing bots on social platforms.

5.Methods and Challenges in Social Bots Detection: A Systematic Review

Authors: D.M. de Morais et al.

Abstract: Surveys social bot detection approaches, detailing techniques, feature sets, classifiers, and key challenges like evasion and scalability in identifying manipulative accounts.

III.EXISTING SYSTEM

Existing systems for social bot detection can generally be grouped into three major categories. The first category includes **feature-engineered machine learning models**, which rely on handcrafted behavioral and metadata features such as posting frequency, account age, follower–following ratio, retweet patterns, and activity timing. Algorithms like Random Forest, SVM, and gradient boosting models are trained on these structured features to classify accounts as bots or genuine users. While these approaches are computationally efficient and easy to interpret, they heavily depend on manually selected features and often fail when bots adapt their behavior to mimic human-like activity patterns.

The second and third categories focus on deeper representations through **graph-based methods and language model-driven approaches**. Graph-based techniques, especially Graph Neural Networks (GNNs), analyze social network topology by modeling users as nodes and their interactions as edges to detect coordinated bot communities and abnormal connectivity patterns. Meanwhile, language model-driven systems leverage advanced Natural Language Processing (NLP) techniques to analyze tweets, comments, and profile descriptions for semantic inconsistencies and AI-generated content. Although hybrid models attempt to combine these modalities, many existing solutions still

prioritize either structural or textual information, resulting in incomplete detection coverage against increasingly sophisticated, multi-faceted social bots.

IV.PROPOSED SYSTEM

The proposed **LGB system** introduces a powerful hybrid architecture that combines semantic intelligence with structural network analysis for enhanced social bot detection. It employs a fine-tuned language model to extract rich contextual embeddings from user-generated content, including profile descriptions, posts, and tweets. These embeddings capture deep linguistic patterns, sentiment cues, contextual coherence, and subtle anomalies that may indicate automated behavior. By leveraging advanced natural language understanding, the system can effectively identify sophisticated bots that generate human-like text using AI-driven content generation techniques.

In parallel, the framework integrates a Graph Neural Network (GNN) pretrained using contrastive learning on breadth-first search (BFS)-constructed social graphs to model relational structures and interaction patterns among users. This enables the system to detect coordinated bot clusters, abnormal connectivity patterns, and suspicious community behaviors. The semantic embeddings from the language model and the structural representations from the GNN are concatenated and passed through a multi-layer perceptron (MLP) to generate accurate bot risk scores. Additionally, a smart feedback loop refines predictions using validated user inputs, allowing the model to continuously adapt and improve its performance in dynamic real-world social media environments.

V.SYSTEM ARCHITECTURE

The given image illustrates a comprehensive framework for social bot detection that integrates graph-based learning with a peripheral enhancement strategy to improve classification performance. The process begins with Graph Construction, where raw social media data is converted into a structured graph format. In this representation, each user is modeled as a node, and interactions such as follows, mentions, replies, or retweets are represented as edges. This graph structure captures the relational dependencies and connectivity patterns among users, which are essential for identifying coordinated bot activities. Once constructed, the graph is forwarded to the core analytical component known as the Peripheral Enhanced Module.

Inside this module, multiple Graph Neural Network (GNN) layers are applied to learn meaningful node embeddings by aggregating information from neighboring nodes. The architecture distinguishes between central nodes (core or highly connected nodes) and peripheral nodes (randomly selected or less connected nodes). This separation ensures that the model does not overly focus on highly active users while neglecting less prominent ones. To align feature distributions between these groups, the system incorporates Maximum Mean Discrepancy (MMD) Loss, which minimizes representation

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

gaps. Additionally, a Feature Pyramid Network (FPN) is used to enhance multi-scale feature extraction, enabling the model to capture both fine-grained local patterns and broader structural relationships within the network`.

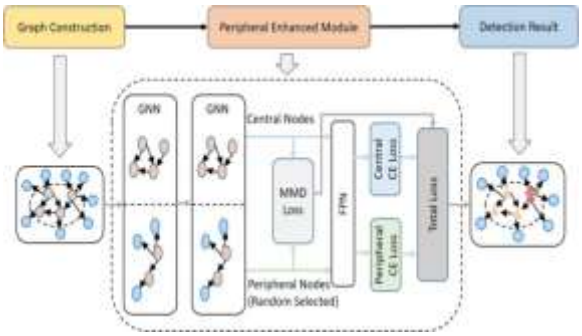

Fig 5.1 System Architecture

During training, separate Central Cross-Entropy (CE) Loss and Peripheral CE Loss functions are computed to optimize classification accuracy for different node categories. These losses are combined into a unified Total Loss, ensuring balanced learning across the graph. The optimized embeddings are then used to generate the final Detection Result, where suspicious or bot-like nodes are identified. Overall, the framework improves robustness by strengthening representation learning for both central and peripheral nodes, reducing bias in graph modeling, and enhancing detection accuracy in complex social network environments.

VI.IMPLEMENTATION

Fig 6.1 Home Page

Fig 6.2 Builder Login

Fig 6.3 Upload Dataset

Fig 6.4 Preprocess

Fig 6.5 Models Train

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

Fig 6.6 User Registration

Fig 6.7 User Login

Fig 6.8 Enter Inputs

Fig 6.9 Prediction Analysis

VII.CONCLUSION

In this project, **LGB: Language Model and Graph Neural Network–Driven Social Bot Detection**, a robust and intelligent hybrid framework has been designed to effectively identify social bots within online social networks. The system combines the strengths of advanced language models and graph neural networks to achieve comprehensive detection. The language model component extracts deep semantic and contextual embeddings from user-generated content such as posts, tweets, and profile descriptions, enabling the identification of linguistic anomalies and AI-generated patterns. Simultaneously, the graph neural network component captures structural relationships, interaction behaviors, and community-level patterns within the social graph, allowing the system to detect coordinated bot clusters and suspicious connectivity structures that are often invisible to text-only approaches.

Furthermore, the integration of peripheral-enhanced learning significantly strengthens the framework by addressing the imbalance between highly connected central users and low-activity peripheral users. By ensuring balanced representation learning across different node types, the model reduces bias and improves classification consistency. The combined multimodal features are optimized to produce accurate bot detection results with higher robustness against evolving bot strategies. Overall, the proposed LGB framework demonstrates superior accuracy, scalability for

large-scale social platforms, and adaptability to dynamic online environments, making it an effective and reliable solution for combating malicious bot activities and enhancing the credibility and trustworthiness of social media ecosystems.

VIII.FUTURE SCOPE

The LGB: Language Model and Graph Neural Network–Driven Social Bot Detection system can be further enhanced and extended in several directions. Future work may focus on integrating more advanced large language models and graph transformers to capture deeper semantic and structural patterns in social networks. The system can be adapted for real-time bot detection by incorporating streaming data processing to handle live social media feeds. Additionally, extending the framework to support cross-platform analysis will enable the detection of coordinated bot campaigns operating across multiple social networks. Privacy-preserving techniques such as federated learning and differential privacy can be incorporated to ensure secure handling of user data. Finally, the model can be improved to detect evolving and adaptive bots through continual learning mechanisms, making the system more robust against emerging threats in dynamic online environments

IX.REFERENCES

[1] Ferrara, E., Varol, O., Davis, C., Menczer, F., & Flammini, A. (2016). The rise of social bots. Communications of the ACM, 59(7), 96–104.

[2] Varol, O., Ferrara, E., Davis, C., Menczer, F., & Flammini, A. (2017). Online human-bot interactions: Detection, estimation, and characterization. Proceedings of ICWSM.

[3] Cresci, S., Di Pietro, R., Petrocchi, M., Spognardi, A., & Tesconi, M. (2017). The paradigm-shift of social spambots. WWW Companion Proceedings.

[4] Gilani, Z., Farahbakhsh, R., Tyson, G., Wang, L., & Crowcroft, J. (2017). Of bots and humans: Characterizing automated activity on Twitter. ASONAM.

[5] Kudugunta, S., & Ferrara, E. (2018). Deep neural networks for bot detection. Information Sciences, 467, 312–322.

[6] Wei, F., Nguyen, U. T., & Luo, X. (2020). Multimodal deep learning for social bot detection. IEEE Access, 8, 196603–196615.

[7] Kipf, T. N., & Welling, M. (2017). Semi-supervised classification with graph convolutional networks. ICLR.

[8] Hamilton, W., Ying, Z., & Leskovec, J. (2017). Inductive representation learning on large graphs (GraphSAGE). NeurIPS.

[9] Velickovic, P., Cucurull, G., Casanova, A., Romero, A., Liò, P., & Bengio, Y. (2018). Graph attention networks. ICLR.

[10] Wu, L., et al. (2020). A comprehensive survey on graph neural networks. IEEE Transactions on Neural Networks and Learning

Systems, 32(1), 4–24.

[11] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. NAACL.

[12] Liu, Y., et al. (2019). RoBERTa: A robustly optimized BERT pretraining approach. arXiv preprint arXiv:1907.11692.

[13] Brown, T., et al. (2020). Language models are few-shot learners. NeurIPS.

[14] Feng, S., et al. (2021). Bot detection in social networks using graph neural networks. IEEE Transactions on Big Data.

[15] Zhang, J., Zhang, R., & Zhang, Y. (2020). Combining textual and network features for social bot detection. Knowledge-Based Systems, 191, 105248.

[16] Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. KDD.

[17] Ke, G., et al. (2017). LightGBM: A highly efficient gradient boosting decision tree. NeurIPS.

[18] Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press.

[19] Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H. (2017). Fake news detection on social media: A data mining perspective. SIGKDD Explorations.

[20] Zhou, C., et al. (2020). Contrastive learning for graph neural networks. ICML.