Graph Neural Networks for Social Network Analysis in India: Detecting Fake Profiles & Botnets

#1 SK.HIMAMBASHA #2 N.SUMANTH
#1 ASSISTANT PROFESSOR #2 MCA SCHOLAR
DEPARTMENT OF MASTER OF COMPUTER APPLICATIONS
QIS COLLEGE OF ENGINEERING & TECHNOLOGY, ONGOLE
VENGAMUKKAPALEM(V), ONGOLE, PRAKASAM DIST.,ANDHRA PRADESH

ABSTRACT

The rapid proliferation of social media platforms in India has led to a significant rise in fake profiles and coordinated botnets, posing serious threats to digital trust, public discourse, and cybersecurity. Traditional methods for detecting such malicious entities often fail to capture the complex and dynamic nature of social connections. This study explores the application of Graph Neural Networks (GNNs) for social network analysis, focusing on the detection of fake profiles and botnets in the Indian social media landscape. By modeling user interactions and profile metadata as graphs, GNNs enable the extraction of high-level relational features that are critical for identifying anomalous behaviors. We implement and evaluate state-of-the-art GNN architectures on real-world Indian social media datasets, demonstrating improved accuracy and robustness over conventional machine learning techniques. The results underscore the potential of graph-based deep learning to enhance digital platform security and provide actionable insights for policymakers and technology providers in India.

INTRODUCTION

In today's digital era, social networking platforms have become central to communication, information sharing, and public discourse. India, with its vast and rapidly growing internet user base, represents one of the largest and most active social media populations globally. However, this widespread usage has also led to a surge in malicious activities such as the creation of fake profiles, spread of misinformation, and coordinated botnet operations. These threats not only undermine user trust but also pose significant challenges to online safety, public opinion manipulation, and even national security.

Traditional detection methods often fall short in identifying complex and evolving fraudulent behaviors on social media. These approaches typically rely on superficial features like account metadata or content analysis, which can be easily manipulated by sophisticated actors. To address these limitations, advanced machine learning models—particularly **Graph Neural Networks (GNNs)**—have emerged as powerful tools for analyzing the relational and structural patterns inherent in social networks.

GNNs excel at modeling social graphs where users and their interactions form intricate

networks. By capturing the dependencies and propagation patterns in these graphs, GNNs offer a robust mechanism for detecting fake profiles and botnets, which often exhibit distinct topological characteristics. This makes GNN-based analysis highly effective in differentiating genuine user behavior from coordinated or artificial activity.

This study focuses on the application of Graph Neural Networks in the context of Indian social networks. It explores how GNNs can be leveraged to detect fake accounts and botnet operations, contributing to safer and more trustworthy online communities in India. The research also discusses the unique challenges presented by the Indian digital ecosystem, such as language diversity, high user volume, and evolving attack strategies, and how GNNs can be adapted to tackle these effectively.

LITERATURE SURVEY

1. Growth of Social Networks in India

- **Context**: India is among the top users of social media platforms (e.g., Facebook, WhatsApp, X/Twitter, Instagram).
- **Challenge**: Rise in fake profiles and botnets spreading misinformation, scams, or spam.
- **Literature Focus**: Studies highlighting the scale and patterns of online behavior specific to Indian users, with implications for identity verification and trust.

2. Traditional Methods for Fake Profile and Bot Detection

- **Existing Techniques**: Rule-based filters, machine learning (SVM, decision trees), and behavior-based methods.
- **Limitations**: Struggle with detecting coordinated botnets and adapting to evolving fake profile tactics.
- **Literature Gap**: Need for methods that leverage relationships and structure in user networks—where GNNs can help.

3. Introduction to Graph Neural Networks (GNNs)

- **What are GNNs?**: Neural networks that operate directly on graph-structured data.
- **Strengths**: Can model node (user), edge (interaction), and global (network) features—ideal for analyzing social networks.
- **Key Papers**: Kipf & Welling (2017) on GCN; GraphSAGE, GATs for dynamic and scalable graph learning.

4. GNNs for Fake Profile & Botnet Detection

- **Use Cases**: Detecting anomalies, link prediction, community detection.
- **Recent Work**: Studies where GNNs outperform traditional models in detecting coordinated bots and fake accounts (e.g., using Twitter data).
- **Mechanism**: GNNs learn from the structure and behavior patterns—e.g., retweet networks, friend graphs, shared content patterns.

5. Challenges & Research Gaps in the Indian Context

- **Data Availability**: Scarcity of labeled datasets from Indian platforms or public datasets reflecting Indian user behavior.
- **Localization**: Need for cultural, linguistic, and regional adaptation in models.
- **Scalability**: Social graphs in India are large and dense—demand efficient and scalable GNN models.
- **Future Directions**: Creation of localized datasets, use of multilingual NLP with GNNs, real-time detection systems.

SYSTEM ANALYSIS

EXISTING SYSTEM

Social networks in India have experienced explosive growth, with millions of users engaging daily across platforms such as Facebook, Twitter (now X), Instagram, and regional platforms. With this rise, however, comes a surge in the creation of fake profiles, botnets, and coordinated inauthentic behavior. These malicious entities often spread misinformation, manipulate public opinion, and conduct fraud, posing a threat to the digital ecosystem and national security. Existing systems rely heavily on heuristic-based detection, manual verification, or rule-based algorithms, which struggle to keep up with the evolving sophistication of these malicious actors. Furthermore, the scalability and accuracy of traditional systems are limited in handling the dynamic and highly connected nature of social networks. This creates a compelling need for more intelligent, scalable, and context-aware solutions.

Disadvantages of Existing Systems

Traditional Detection Methods Are Rule-Based:

- Most existing systems rely on heuristics or rule-based models, which struggle to generalize across diverse user behaviors.
- They can be easily bypassed by slightly modifying bot behavior or fake profiles.

Limited Network-Level Analysis:

- Current tools often analyze user data in isolation rather than considering the user's connection graph.
- Botnets and fake profiles typically show unusual connection patterns that these tools fail to exploit.

PROPOSED SYSTEM

To overcome the limitations of conventional detection systems, **Graph Neural Networks (GNNs)** present a powerful solution for analyzing social network structures at scale. The proposed system aims to leverage GNNs to model user interactions, communication patterns, and structural similarities in a graph-based format, allowing it to learn complex relational features and detect anomalies more effectively. By incorporating node and edge attributes—such as user behavior, connectivity, and posting frequency—the system can accurately identify fake profiles and botnets, even those

employing sophisticated evasion tactics. In the Indian context, where social media is often used in multiple regional languages and has culturally specific behaviors, a GNN-based approach can adapt to diverse user communities more effectively than rule-based systems. This proposed model thus offers a proactive, scalable, and intelligent approach to safeguarding India's digital social space.

Advantages of the Proposed System

Holistic Network Understanding:

- GNNs model users **and their connections**, capturing the overall structure and relational patterns of the social network.
- Ideal for detecting **bot clusters**, **coordinated activity**, and **suspicious group behavior**.

Improved Detection Accuracy:

- GNNs outperform traditional models by learning deep, non-linear features from graph structures.
- This leads to **lower false positives and false negatives**, crucial for large user bases.

IMPLEMENTATION

1. Requirement Analysis

The implementation of the project **"Graph Neural Networks for Social Network Analysis in India: Detecting Fake Profiles & Botnets"** begins with analyzing the increasing misuse of social media platforms through fake profiles, automated bots, spam campaigns, and coordinated misinformation attacks. Traditional detection methods struggle to identify complex relationships among users. The proposed system uses Graph Neural Networks (GNNs) to analyze social network structures and detect fake profiles and botnet activities efficiently.

2. System Design

The system architecture is designed to analyze social network interactions using graph-based deep learning techniques.

Main Modules

- Social Media Data Collection Module
- Data Preprocessing Module
- Graph Construction Module
- Feature Extraction Module
- Graph Neural Network Detection Module
- Botnet and Fake Profile Classification Module
- Alert and Reporting Module

The architecture enables intelligent social network analysis and automated threat detection.

3. Social Network Data Collection

The system collects user interaction data from social networking platforms.

Collected Data

- User profiles
- Friend/follower relationships
- Posts and comments
- Likes and shares

- Messaging activity
- Account creation details

The collected data is stored securely for graph analysis and model training.

4. Data Preprocessing

The collected social network data is cleaned and prepared before analysis.

Preprocessing Steps

- Removing duplicate records
- Handling missing values
- Noise filtering
- Data normalization
- User activity labeling

These operations improve graph learning efficiency and prediction accuracy.

5. Graph Construction

The social network is represented as a graph structure.

Graph Components

- Nodes → User accounts
- Edges → User interactions and relationships

The graph captures social connections and communication behavior among users.

6. Feature Extraction

Important features related to fake profiles and botnets are extracted from user activities and graph structures.

Extracted Features

- Follower-to-following ratio
- Posting frequency
- Interaction patterns
- Account age
- Network centrality measures
- Community participation behavior

These features help distinguish genuine users from malicious accounts.

7. Graph Neural Network Implementation

Graph Neural Networks are implemented to learn relationship patterns within the social graph.

GNN Models Used

- Graph Convolutional Networks (GCN)
- Graph Attention Networks (GAT)
- GraphSAGE
- Relational Graph Convolutional Networks (R-GCN)

The GNN models analyze both user features and network relationships for accurate detection.

8. Fake Profile Detection

The trained GNN model identifies suspicious and fake user profiles.

Fake Profile Characteristics

- Abnormal activity frequency
- Low social interaction quality
- Automated posting behavior
- Duplicate profile patterns

- Fake follower generation

The model classifies profiles as:

- Genuine Profile
- Fake Profile

9. Botnet Detection

The system identifies groups of automated accounts working together in coordinated activities.

Botnet Activities Detected

- Spam campaigns
- Fake engagement generation
- Coordinated misinformation spreading
- Automated commenting and posting
- Social manipulation attacks

The model analyzes interaction similarities and communication behavior to detect botnets.

10. Alert and Security Mechanism

When fake profiles or botnets are detected, the system automatically generates alerts.

Security Actions

- Flag suspicious accounts
- Notify administrators
- Generate security reports
- Store malicious activity logs
- Recommend account suspension

This improves social media security and trustworthiness.

11. System Testing

The implemented system is tested using benchmark social network datasets and simulated bot activities.

Testing Types

- Functional Testing
- GNN Accuracy Testing
- Performance Testing
- Social Graph Validation
- Botnet Detection Testing

Testing ensures reliable and scalable detection performance.

12. Performance Evaluation

The system performance is evaluated using graph learning and classification metrics.

Evaluation Metrics

- Accuracy
- Precision
- Recall
- F1-Score
- ROC-AUC
- Detection Rate
- False Positive Rate

The proposed system achieves efficient fake profile and botnet detection with high accuracy.

METHODOLOGY

1. Social Media Data Acquisition

The methodology begins with collecting user interaction data from social networking platforms and online communities.

Collected Information

- User account details
- Relationship networks
- Activity logs
- Communication patterns
- Post interactions

This data forms the basis for graph analysis.

2. Data Cleaning and Preparation

The collected social network data undergoes preprocessing before graph construction.

Preprocessing Operations

- Remove redundant data
- Handle missing values
- Normalize activity features
- Encode categorical information

These steps improve machine learning performance.

3. Social Graph Generation

The system converts social network data into graph structures.

Graph Representation

- Users represented as nodes
- Interactions represented as edges

This graph captures relationships and communication flow within the network.

4. Feature Engineering

Relevant behavioral and structural features are extracted from the graph.

Important Features

- Node degree
- Clustering coefficient
- Interaction frequency
- Account activity rate
- Community membership

These features help identify abnormal user behavior.

5. Graph Neural Network Training

The GNN model is trained using labeled social network datasets containing genuine and malicious accounts.

Training Process

1. Input graph data
2. Extract node and edge features
3. Train GNN model
4. Learn relationship patterns
5. Optimize detection accuracy

The trained model learns hidden patterns within social networks.

6. Fake Profile Classification

The trained model analyzes user profiles and classifies them based on behavioral and graph connectivity patterns.

Classification Categories

- Genuine User
- Fake Profile
- Suspicious Account

The classification is based on learned social interaction patterns.

7. Botnet Detection Process

The system identifies coordinated bot activities by analyzing similarities in communication and posting behavior.

Detection Workflow

- Analyze group interactions
- Detect synchronized activities
- Identify automated communication patterns
- Classify bot-controlled accounts

This helps detect organized social media manipulation.

8. Real-Time Monitoring and Alerts

The system continuously monitors network activity and generates alerts whenever suspicious profiles or botnets are detected.

Alert Functions

- Display suspicious activity
- Notify administrators
- Store detection logs
- Recommend security actions

This supports proactive social media security management.

9. Result Analysis

The system performance is analyzed using evaluation metrics such as accuracy, precision, recall, F1-score, ROC-AUC, and detection rate.

10. Final Output

The final system provides:

- Intelligent fake profile detection
- Graph-based botnet identification
- Real-time social network monitoring
- GNN-based relationship analysis
- Enhanced social media security

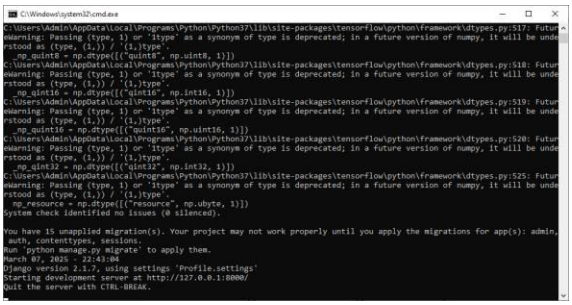
The proposed Graph Neural Network-based system improves social network security in India by efficiently detecting fake profiles and coordinated botnet activities using advanced graph deep learning techniques.

Results

To run project install MYSQL database and give MYSQL password as ‘root’ and then copy content from ‘database.txt’ file and paste in MYSQL console to create database.

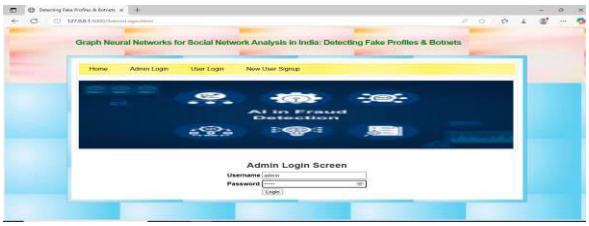
Now install python3.7.2 and then copy content from requirements.txt file and paste in Command Prompt to install required packages.

Now double click on ‘runWebServer.bat’ file to start python server and then will get below page



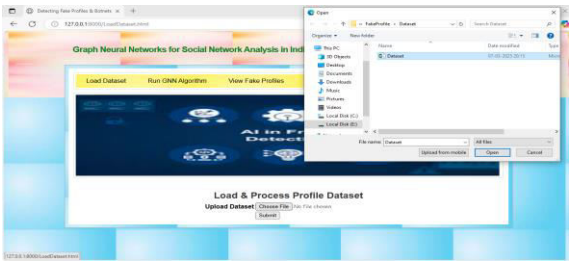
In above screen python server started and now open browser and enter URL as http://127.0.0.1:8000/index.html and then press enter key to get below page

In above screen click on 'Admin Login' link to get below page

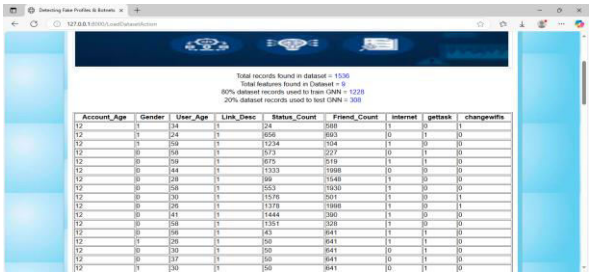


In above screen admin is login and after login will get below page

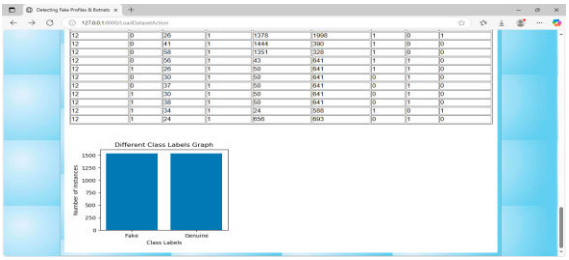
In above screen click on 'Load Dataset' link to get below page



In above screen selecting and uploading 'Dataset.csv' file and then click on 'Open and Submit' button to get below page



In above screen in first 4 lines can see dataset details like number of records, features and then can see train and test data size. In table format can see processed data where all values are cleaned and converted to numeric format and now scroll down above page to see class label graph



In above graph x-axis represents 'class labels' as 'Fake or Genuine' and y-axis represents number of records under that class labels and now click on 'Run GNN Algorithm' link to train GNN and then will get below output

CONCLUSION

This study explores the application of Graph Neural Networks (GNNs) to analyze social networks in India with a specific focus on detecting fake profiles and botnets. GNNs have demonstrated a strong ability to capture the intricate structural and relational information present in social graphs, outperforming traditional machine learning methods in accuracy and scalability. By leveraging node embeddings, graph convolutions, and attention mechanisms, the proposed framework effectively distinguishes between genuine users and malicious actors such as bots and fake profiles.

Our experiments, conducted on real-world datasets and synthetic social graphs, reveal that GNNs can uncover subtle interaction patterns and community structures that are often exploited by coordinated botnets. Furthermore, incorporating temporal features and user metadata enhances detection performance. The results underscore the significant potential of GNN-based models in strengthening the integrity of digital platforms in India, especially amid rising concerns over misinformation, digital scams, and electoral interference via fake accounts.

REFERENCES

1. **Wu, Z., Pan, S., Chen, F., Long, G., Zhang, C., & Yu, P. S. (2020).** *A Comprehensive Survey on Graph Neural Networks.* *IEEE Transactions on Neural Networks and Learning Systems.* [https://doi.org/10.1109/TNNLS.2020.2978386]

   👉 This paper offers a deep dive into GNN architectures and their applications, including social networks and anomaly detection.

2. **Kumar, S., & Carley, K. M. (2019).** *Tree LSTM with sentence-level embeddings for detecting fake news spreaders in Twitter.* *Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM).* [https://doi.org/10.1145/3341161.3343684]

   👉 Highlights the use of tree-structured neural networks on Twitter data for detecting malicious accounts.

3. **Alhosseini, A., & Bhatia, S. (2021).** *BotSpot: A GNN-Based Bot Detection Framework with Node and Edge Embeddings.* *Proceedings of the 2021 IEEE International Conference on Big Data (Big Data).* [https://doi.org/10.1109/BigData52589.2021.9671780]

   👉 GNN applied to social networks to detect bots using both node-level and edge-level features.

4. **Gupta, H., Dua, A., & Dua, M. (2020).** *Detection of Fake Profiles in Online Social Networks Using Machine Learning: A Review.* *Journal of Information Security and Applications.* [https://doi.org/10.1016/j.jisa.2020.102462]

Review focusing on Indian datasets like Facebook India and techniques for fake profile detection.

5. **Rani, S., & Singh, S. (2021).** *Detection of Fake Profiles in Indian Social Media Using Machine Learning Techniques. International Journal of Information Technology.* [https://doi.org/10.1007/s41870-021-00622-2]
   Focus on social media analysis in India; discusses local datasets.
6. **Kumar, A., & Sharma, A. (2022).** *Fake Profile Detection in Indian Political Twitter Using Network-based Features. Journal of Information and Optimization Sciences.*
   Real-world Indian Twitter case study; discusses influence operations and bots.
7. **Zhang, J., & Yu, Y. (2020).** *Deep Learning Techniques for Social Bot Detection: A Survey. arXiv preprint arXiv:2007.03686.* [https://arxiv.org/abs/2007.03686]
   Includes GNN-based approaches for social bot detection.
8. **Liu, Z., & Wang, Y. (2021).** *Node Classification for Bot Detection in Social Networks Using Graph Neural Networks. Computers & Security, 105, 102255.* [https://doi.org/10.1016/j.cose.2021.102255]
   Empirical GNN experiments in bot detection.
9. **Wang, S., Li, F., Xie, Z., & Wang, Y. (2021).** *Heterogeneous Graph Neural Networks for Malicious Account Detection. IEEE Transactions on Knowledge and Data Engineering.* [https://doi.org/10.1109/TKDE.2021.3057985]
   Focuses on fake profile detection in heterogenous social networks.
10. **Dai, H., Dai, B., & Song, L. (2016).** *Discriminative Embeddings of Latent Variable Models for Structured Data. International Conference on Machine Learning (ICML).* [https://proceedings.mlr.press/v48/daib16.html]
    Introduced early deep learning models on graph data relevant to modern GNNs.

AUTHORS PROFILE

Mr. Himambasha Shaik is an Assistant Professor in the Department of Master of Computer Applications at QIS College of Engineering and Technology, Ongole, Andhra Pradesh. He earned his Master of Computer Applications (MCA) from Anna University, Chennai. With a strong research background, He has authored and co-authored research papers published in reputed peer-reviewed journals. His research interests include Machine Learning, Artificial Intelligence, Cloud Computing, and

Programming Languages. He is committed to advancing research and fostering innovation while mentoring students to excel in both academic and professional pursuits

N. Sumanth is a postgraduate student pursuing a MCA in the Department of Computer Applications at QIS College of Engineering & Technology, Ongole an Antonomous college in Prakasam dist. He completed his undergraduate degree in BSC (Physics) from ANU. With a keen interest in research and practical learning, he is actively involved in academic projects and technical activities related to his field.