Graph-Based Twitter Bot Detection Using Network Analysis and Deep Learning\*

Name-Surname: Elif Işık   
Student Number: 20190808006

Name-Surname: Alp Buğra Toker  
Student Number: 20200808611

University: Akdeniz University City:Antalya,Turkey

*Abstract*—This report outlines the development of an innovative method for detecting bots on Twitter using graph neural networks (GNNs). We addressed the challenges of effective feature extraction and robust GNN training to enhance detection accuracy. Our approach utilizes a comprehensive dataset comprising tweets and user metadata, incorporated into a tensor structure for efficient processing. We demonstrate the efficacy of our method through improved detection rates and reduced computational overhead, setting a precedent for future work in social media forensics..Keywords—Twitter, Bot Detection, Graph neural networks (GNNs), Feature extraction, Social Network Analysis, Graph Convolutional Networks (GCN), Graph Attention Networks (GAT), Tensor Processing, Machine Learning,Data Mining (key words)

# Introduction

The proliferation of bots on social media platforms, particularly Twitter, has presented significant challenges in maintaining information integrity and enhancing user experience. These automated accounts are capable of influencing public opinion, manipulating stock prices, and spreading misinformation. Our research focuses on developing an advanced machine learning approach to efficiently identify these bots. By leveraging the rich metadata and relational data available from Twitter, our method utilizes graph neural networks (GNNs) to analyze the complex network of interactions among users. The approach includes extracting features from user profiles and tweet embeddings, which are then processed using PyTorch for optimal performance. We structure these features into tensors to capitalize on the computational efficiency of modern tensor operations. Additionally, our method constructs dynamic graphs from the user interaction data, which provides a deeper insight into the underlying social connections, enabling more accurate detection of sophisticated bots that mimic human behaviors. This comprehensive framework not only improves detection accuracy but also enhances the scalability of bot detection algorithms, making it adept at handling the vast and ever-evolving landscape of social media data.

# Related Work (Literature Review)

Previous works in bot detection primarily relied on rule-based algorithms and shallow machine learning models, which often struggled with scalability and adaptability due to their static nature and inability to handle non-linear relationships in data. These traditional methods were prone to high false-positive rates and required frequent manual updates as bot behaviors evolved. Recent advancements, however, have shifted focus towards more robust deep learning and graph-based approaches, recognizing the complex and interconnected nature of social media data. Notable methodologies include the use of Graph Convolutional Networks (GCN) and Graph Attention Networks (GAT), which have shown promise in related fields such as fraud detection and recommendation systems. These methods leverage node feature learning to dynamically adapt to new patterns of bot-like activities without extensive retraining. For instance, studies in fraud detection have demonstrated that GCNs can significantly improve the detection rates by effectively capturing the relational dependencies among accounts, which is critical in identifying coordinated inauthentic behaviors. Similarly, GATs have been applied successfully in recommendation systems, where they adjust attention weights dynamically to prioritize more relevant user interactions. By integrating these sophisticated approaches, our research aims to harness their proven capabilities to enhance both the accuracy and efficiency of bot detection algorithms, thereby offering a scalable solution that can adapt to the evolving tactics of social media bots.

# Method

Our method involves two main stages: feature extraction and model training. Initially, we extracted features from JSON formatted Twitter data, transforming user attributes and tweet contents into tensors using PyTorch. This transformation involved not only numerical scaling but also the extraction of high-dimensional embeddings from tweets via pre-trained language models, which capture semantic nuances beyond basic text analytics. These tensors were then used to construct a graph where nodes represent users, and edges represent interactions, such as retweets and mentions, emphasizing the network dynamics of social media engagement.

Data Preparation: We aggregate tweets and user data into a structured format, enhancing feature extraction by incorporating both user profiles and tweet embeddings derived from state-of-the-art language models like BERT or RoBERTa. This approach allows us to capture a rich set of features that reflect both the content and context of user activities on Twitter.

Feature Engineering: Numeric and categorical features are extracted and scaled using techniques like MinMaxScaler for normalization, ensuring that all input features contribute equally to the model's performance. Tweet embeddings are then reduced in dimensionality using PCA, which helps in distilling the essential information and reducing the computational load during training.

Graph Construction: Separate graphs for training and testing data are constructed, representing users as nodes and their follower-following relationships as edges. This structural representation helps in modeling the social network's influence patterns, crucial for detecting bots that often exhibit abnormal link formations.

GNN Architecture: Our models, including a two-layer GCN and a GAT, leverage the spectral features of the graph. The GCN model processes node features through successive layers to capture local graph structures, while the GAT introduces attention mechanisms that assign different weights to nodes in a neighborhood, allowing for nuanced feature learning from complex node interactions.

Training and Evaluation: The models are trained on a custom dataset split, optimized using strategies such as dropout and batch normalization to prevent overfitting and improve generalization across unseen data. Performance metrics such as accuracy, precision, recall, and F1-score are monitored over epochs to evaluate the effectiveness of the models under various thresholds and operating conditions. We also employ techniques like early stopping and model checkpoints to enhance training efficiency and model robustness.

By addressing these stages with meticulous detail and advanced methodologies, our approach not only identifies bots with high accuracy but also adapts to new and evolving patterns of deceptive behavior on social media platforms.

# Experimental Result

We conducted experiments using a dataset split into training and testing subsets. The models were evaluated based on accuracy, precision, recall, and F1-score. Our GNN models outperformed traditional machine learning approaches, demonstrating significant improvements in detecting sophisticated bots that exhibit human-like behaviors. The results were validated against a baseline model, showing the GAT model's superior performance in handling sparse and skewed data distributions..

# CONCLUSION

Our research confirms the potential of graph neural networks in detecting Twitter bots, surpassing traditional methods in both performance and efficiency. Future work will explore the integration of real-time data streams and the adaptation of our models to detect emerging bot strategies. This work not only contributes to the field of social media analytics but also aids in the broader effort to maintain the credibility and security of online platforms.