Botbook: A Graph-Based Framework for Classification of Bot Accounts on Facebook

Rona Domantay[1], Leovide Daniel Bato[2], Ezrha Leigh Dangilan[3], Geoff Anthony Dulnuan[4], Kaizer Cyn Gura[5], Ava Narag[6], Bernard Carlo Pacis[7], Genrev Roque[8], and Jaymee Sofia Surro[9]

[1] Computer Science - Computer Applications Department, Saint Louis University - Baguio
redomantay@slu.edu.ph
[2] Computer Science - Computer Applications Department, Saint Louis University - Baguio
2223451@slu.edu.ph
[3] Computer Science - Computer Applications Department, Saint Louis University - Baguio
2223144@slu.edu.ph
[4] Computer Science - Computer Applications Department, Saint Louis University - Baguio
2226098@slu.edu.ph
[5] Computer Science - Computer Applications Department, Saint Louis University - Baguio
2221097@slu.edu.ph
[6] Computer Science - Computer Applications Department, Saint Louis University - Baguio
2224512@slu.edu.ph
[7] Computer Science - Computer Applications Department, Saint Louis University - Baguio
2223202@slu.edu.ph
[8] Computer Science - Computer Applications Department, Saint Louis University - Baguio
2223129@slu.edu.ph
[9] Computer Science - Computer Applications Department, Saint Louis University - Baguio
2212173@slu.edu.ph

Abstract: **This study addresses the rise of malicious social bots on Facebook, and the serious threats it presents to online discourse, emphasizing the requirement for efficient detection methods on social media platforms. Existing methods of bot detection struggle to respond to evolving bot behaviors and platform-specific dynamics. To address these current challenges, this research introduces Botbook, a graph-based framework for identifying and categorizing malicious social bots. It is designed to detect and classify bot accounts—specifically distinguishing humans from Trolls, Cyborgs, and Spam Bots—through the integration of behavioral, linguistic, and interaction-based features. Utilizing a static graph model built from user comment data, the study constructs feature-rich profiles and applies machine learning classifiers to label users, with Random Forest achieving the highest classification performance. Subsequently, Graph Neural Networks, such as Graph Convolutional Network (GCN) and Graph Attention Network (GAT), are employed to analyze user interactions, with GAT demonstrating an accuracy rate of 90.7% and superior F1 and AUC-ROC scores, providing effective identification of minority class instances and contributing to better overall model generalization. While this**

study is limited to static graph analysis and platform-specific data constraints, the study's findings contribute to the development of scalable and context-aware bot detection systems, offering critical insights for enhancing social media integrity and safeguarding democratic processes in the Philippines and beyond.

***Keyword:* Bots, Social Media, Facebook, Online Discourse, Graph Neural Networks**

I. INTRODUCTION

For the past years, social media platforms have become a cornerstone of communication. Its ability to provide many avenues of connection has made it invaluable in information dissemination. However, this platform has been increasingly targeted by the proliferation of automated accounts known as social bots.

Social bots are internet-based applications that automatically generate content and mimic human behavior [21]. In 2023, it was reported that almost half of global Internet traffic is bot activities [32]. These social bots can have varying purposes that can range from benign or malicious. Benign social bots include chatbots that respond to frequently asked questions, and knowbots that automatically acquire helpful information from the Internet [7]. On the other hand, malicious social bots aim to disrupt online ecosystems by manipulating and deceiving real human accounts. Commonly, malicious social bots are employed to commit political manipulation, distribute false news, proliferate spam, and infiltrate online spaces [2]. Malicious bots encompass a range of types, including spam bots, which aim to spread unwelcome content such as advertisements, phishing links, and explicit material; sock puppets, which use hoax identities to interact with real users; trolls, who are politically motivated to engage with other accounts; cyborgs, which are accounts partially controlled by automated programs; influence expansion bots, which are created to increase social connections to establish online influence; and fake followers, which are accounts managed by entities paid to follow other users to increase perceived popularity and credibility while forming a mutual following network to avoid detection [21].

With the continuous advancements made in creating malicious social bots, an effective detection method is paramount to preserve the safety of online ecosystems. However, despite advancements in detecting technologies, existing methods continue to possess several significant gaps.

A. Literature Review

The presence of social media bots poses harm to the integrity and security of social media platforms like Facebook, through their ability to manipulate, deceive, and spread misinformation. As bots evolve, detection methods must evolve alongside them to accurately distinguish them from genuine users. This literature review discusses existing bot detection techniques and identifies the current limitations faced in the bot detection landscape.

A.1. Machine Learning-Based Bot Detection

Machine Learning (ML) has become an important tool in bot detection methods. Various studies have employed various ML models to enhance detection accuracy against evolving bot behaviors. Studies such as [15] integrated metadata and non-text features in bot detection through application of text mining and an attention-based Bi-LSTM model with their 30,000-tweet dataset. The model achieved 79.7% accuracy in distinguishing fake news from bots and humans, demonstrating its effectiveness in filtering harmful social-bot-generated misinformation. Using sentiment-based features, this study [18] achieved an 87% accuracy rate in detecting bots from Dutch tweets. Further, there are models such as Bot-DenseNet that integrates text embeddings and metadata, and provides a better trade-off between performance and feasibility than typical transformer models like RoBERTa [25]. DABot, a bot detection model for Weibo, utilized metadata-based, interaction-based, content-based, and timing-based features, outperforming other models with its 98.87% accuracy [37].

Despite these high performance rates, ML-based methods struggle in generalizing towards multilingual datasets. Studies underscore the need for improved feature engineering and call for more sophisticated approaches in countering the use of generative Artificial Intelligence (AI) in social media bot content. According to this study [35], there are current difficulties in detecting AI-generated text, mostly used by bot accounts. Moreover, real-time detection is emphasized in this study [8] to decrease the ability of bots to spread false information.

A.2. Graph-Based Bot Detection

Graph-based models utilize user networks to uncover bot behavior, incorporating topological, temporal, and frequency-based features to augment its detection accuracy.

Several graph-based models exist such as BGSRD, which combines BERT for textual analysis, with a Graph Convolutional Network (GCN) for structural learnings [13], MSGS, a Graph Neural Network (GNN) that analyzes connections with the use of attention mechanisms to improve detection accuracy [32], and SEGCN, which analyzes entire sub-networks, allowing it to capture more complex patterns and improve bot detection accuracy [20]. Another notable graph-based detection tool is BotDGT, a dynamic graph transformer model that captures temporal changes in bot behavior, providing superior adaptability over static models [17].

Graph-based techniques show strong potential, but challenges remain in scalability, real-time processing, and adapting to platform-specific network structures. Future research should explore more dynamic graph modeling techniques and adversarial learning to counter increasingly sophisticated bot networks.

A.3. Hybrid and Other Approaches to Bot Detection

Other detection tools combine ML, linguistic analysis, and graph-based methods to enhance detection accuracy and resilience. This study [34] adopted linguistic features for a Spanish-based corpus, achieving 90% accuracy for their NaiveBayes and SimpleLogistic classifiers. In another study [29], researchers opted for a one class classification model, instead of the usual binary classification model, for detection of novel bot behaviors and achieved an AUC score of 0.9.

On the other hand, this study [31] focused on text features, timestamps and other information-based features to define a commercially viable bot detection tool for Facebook. Their experimental design yielded an F1 score of 0.71 and found that bots account form only a small percentage of total users but generate disproportionately large amounts of content.

BotWise model detects Twitter bots through behavioral analysis and social media vulnerabilities, focusing on the bots' effectiveness in disseminating misinformation, and was able to correctly recognize 93% of the human accounts and 84% of manually labeled bot accounts [4]. Lastly, user features, topic modeling, and sentiment analysis were all incorporated in this model, detecting 15% of their dataset as bots spreading COVID-19 misinformation [10].

A.4. Challenges and Limitations in Bot Detection

Despite the growing literature in bot detection, key limitations can still be seen. Most bot detection tools are focused only on specific social media platforms such as X, formerly known as Twitter, neglecting other platforms that contain a significant presence of social media bots (e.g. Instagram and Facebook). To add, most datasets are skewed towards the English language, and are limited to tackling bots under financial and election-related contexts [16]. Some detection models find themselves at risk of being obsolete due to evolving bot behaviors [9], and some sentiment-based models can fail to accurately understand real-world context and nuance [26].

Future research should utilize diverse datasets across platforms, languages, and regions. Comparative studies on evolving bots can refine classification techniques and expanding beyond text to behavioral modeling and multimodal data could enhance detection of social media bots

B. Problem Statement

Bot detection is critical due to the great impact of social bots on political, societal, and informational spheres. Bots are proven to amplify or suppress specific views [6], spread misinformation and narrative manipulation [30], influence political debates [23], increase negative sentiments in political movements [19], and have grown to be more sophisticated to reinforce echo chambers and worsen the spread of misinformation [24]. As a consequence, bot detection is especially important to implement in regions where the prevalence of bot accounts is significant, such as the Philippines [22], where the presence of bots can initiate information disorders that can undermine critical political processes such as elections.

Furthermore, the collected literature for bot detection are mostly focused on the social media platform X, formerly known as Twitter. Harmful behaviors caused by bots are also visible in other social media platforms. Without bot detection models for those platforms, the prevalence of bots causing harm to users will be inevitable.

Lastly, there is a glaring gap in the existing literature on bot detection frameworks and tools fitted to the Philippine setting, despite the country being rife with social bots influencing political and societal conversations [3], [5], [27].

C. Objectives

As highlighted in the previous section, addressing the problem of malicious bots and their effects is crucial to ensuring safe and trustworthy online spaces. This study seeks to address these problems by exploring a graph-based neural network approach, aimed at detecting and classifying malicious social bots. Specifically, this study aims to accomplish the following:

- Extend the scope of detection to less observed platforms such as Facebook.
- Create a classification criteria for humans and bot types
- Create a graph-based detection framework that can distinguish between human and bot accounts, and further classify bot accounts, based on their behaviors and purposes.

D. Scope and Limitations

This study focuses on detecting and classifying bots in online social networks using static graph-based methods. The user network is modeled as a snapshot, with nodes representing users and edges as their interactions, with a strong focus on textual and structural properties to identify bot behavior. This study focuses specifically on political discussions during the 2022 Philippine Presidential Elections, where high frequency of political bots is observed. Furthermore, the graph interactions are based on comment relationships on selected political discussions and not on social network interactions between accounts, due to Facebook's privacy policy regarding user data.

However, the study has several limitations. A static graph is unable to account for temporal changes or evolving behaviors within the network, rendering this approach less effective to bots with adaptive or burst behavior. As a result, bots that alter their activity or interact differently over time might not be accurately detected, reducing the method's overall effectiveness. Sophisticated bots designed to mimic human behavior may blend into the network without raising obvious red flags. Furthermore, the reliance on specific graph-based features could lead to a feature selection bias, as other essential aspects of bot behavior, such as content or interaction patterns, may be overlooked, limiting the generalizability of the method to other platforms.

Despite these limitations, this study provides a focused analysis of bot detection and classification using static graph techniques, though limited by the static nature of the analysis, dataset, and privacy restrictions.

E. Significance

As bots become more adept at mimicking human behavior and continue to proliferate political manipulation and misinformation, detection systems must be able to adapt to their behavior to mitigate their effects across online platforms. This study positions itself as an innovative groundwork for detection systems and offers nuanced observations into benign and harmful bot behaviors. Moreover, this study's contextual focus on the Philippines is paramount to understanding and addressing regional challenges, such as bot-driven misinformation that can undermine electoral processes.

With the results of this study, the proponents aim to inform public policy and platform-specific strategies to mitigate malicious bots, ensure a healthier online ecosystem, and safeguard democratic processes and the integrity of online political discussions. As such, this study stands to contribute to creating a robust, scalable, and contextually relevant bot detection framework that can be utilized to create moderation tools or plugins specified for Facebook, ultimately fostering safer and more reliable online spaces.

II. METHODS

A. Data Collection

The dataset consists of comments from Facebook posts related to the previous 2022 Philippine Election, sourced from prominent news outlets in the Philippines such as ABS-CBN News, GMA News, INQUIRER.net, Philippine Star, Manila Bulletin, and Rappler.

Initially, comments were manually collected in a JSON file. However, this method proved to be tedious and time-consuming, thus the researchers have decided to explore automated means. Several online scraping tools were tested, including Apify, Bright Data, ScrapingBot, Crawlbase, and Nimbleway, with Apify producing the best results.

Using Apify's Facebook Comments Scraper, comments were gathered from posts meeting these criteria: at least 100 comments, election-related content, discussions on presidential, vice-presidential, and senatorial candidates, and posts created between April 26, 2022, and May 9, 2022. Due to Apify's free version limitations, multiple scraping sessions were conducted, and the data was stored in JSON format.

Ethical considerations were given paramount priority, ensuring only publicly available comments were collected as per Apify's Terms of Service and Meta's data usage and privacy policies. Personally identifiable information, such as user names and profile links were removed during preprocessing and all user IDs were anonymized. Since certain fields like post date and time were not scraped, they were manually added during preprocessing. Python was used to restructure the raw JSON from Apify, grouping users and their comments.

Each JSON entry represents a unique user and their associated comments, with each comment containing details such as the Facebook post content, the news outlet, the post's date and time, the comment ID, the comment text, the comment's date and time, the presence of an attachment, its reply status, the parent comment ID, the comment depth, and the number of likes. Figure 1 provides a sample snippet of the cleaned JSON structure.

The final dataset contained 735,883 comments and 325,320 unique users, collected from 350 posts in total, to be used for data analysis and feature extraction.

```
{
    "userID": "3613768",
    "comments": [
      {
        "facebookPost": "Presidential aspirant Vice President Leni Robredo received a warm
welcome from a …",
        "newsOutlet": "Philippine Star",
        "postDate": "2022-04-30",
        "postTime": "13:31:00",
        "commentID":
"Y29tbWVudDoyNjc5Mzc4ODA1NTQ5MTE3XzM4ODAzODQ5NjU1MDg1Nw==",
        "commentText": "Pano naging massive?",
        "commentDate": "2022-05-02",
        "commentTime": "06:53:59",
        "attachment": null,
        "isReply": false,
        "parentCommentID": null,
        "depth": 0,
        "likes": 0
      }
    ]
}
```

Fig. 1: Sample Snippet of Preprocessed JSON

B. Annotation

To build the foundation of the graph-based model, an annotated subset of the data is constructed through manually labeling comments from users. This manually labeled subset represents 0.26% of the entire dataset and serves as the ground truth for training and evaluation. Since each user has varied comment volumes, stratified sampling was employed. User comment counts were computed and log-transformed to normalize the distribution. Users were then grouped into 10 log-spaced bins to ensure balanced representation across activity levels. A stratified sample consisting of 840 users was drawn by selecting an equal number of users from each bin, preventing over-representation of low-activity users and capturing diverse interaction patterns.

Three annotators from the group independently reviewed the stratified sample and annotated each user based on their comment patterns, following a bot annotation protocol established in prior research [31]. This independent labeling process was designed to reduce individual bias and allow for assessment of annotation consistency. Each user was categorized into one of four groups: human, troll, spam bot or, cyborg. Humans engage in genuine discussions with natural posting patterns, making them easy to identify [20], [11]. Spam bots are fully automated, posting repetitive content like ads or malicious links, detectable through activity patterns. Cyborgs blend automation with human input, mimicking real users while promoting agendas, making detection harder. Trolls, whether automated or manual, provoke reactions and spread misinformation, requiring intent analysis for identification [20], [12].

After all three annotators completed their reviews, inter-rater reliability was assessed using two established metrics: Krippendorff's Alpha and Fleiss' Kappa, both yielding a score of 0.371, indicating fair agreement among annotators. To finalize the labels, a majority vote computation was used to assign a definitive label to each user.

Based on the finalized annotations, 53.86% of users were identified as Human (453 out of 840), 24.49% as Trolls (206 out of 840), 18.55% as Cyborgs (156 out of 840), and 3.09% as Spam Bots (26 out of 840).

C. Features Development

Features extracted from gathered data provide significant information that can determine human and bot-like behaviors. Using the feature development framework delineated by a previous Facebook detection framework [31] and features that can be extracted from the existing characteristics in the dataset, this work aims to extract properties derived from text, engagement, media, and timestamps. Thus, this directed to the proponents' approach to the following measures from the dataset:

- Average Text Length: The mean length of a user's comments , providing insights into user engagement and verbosity.
- Innovation Rate: The rate at which new words appear in a sequence of comments, capturing linguistic diversity and novelty. A higher innovation rate indicates organic discussions, while a lower rate suggests repetitive or automated behavior.
- Average Links Usage: The prevalence of hyperlinks within comments, which can indicate information sharing, promotional content, or potential automated activity.

- Average Response Time: The mean time interval between commenting on a post or replying to a comment, reflecting user engagement dynamics.
- Thread Deviation: The average difference between the time a user posts a comment and the average response time of all comments in the thread, measuring how consistently users engage over time.
- Engagement: The cumulative number of likes received across all comments, serving as an indicator of user influence and content popularity.
- Number of Comments: Total number of comments made by a user.
- Average Pairwise Text Dissimilarity: The measure of how similar or dissimilar a user's comments are to one another. Values closer to zero indicate repetitive behavior, which may suggest automated or scripted content generation.
- Average Likes per Comment: The mean number of likes per comment by a user, indicating user influence and popularity.
- Reply Ratio: The proportion of a user's comments that are replies to others, reflecting possible interaction behaviors.
- Comment Time Variance: The extent at which a user's comments are spread out over the course of a day. Low variance indicates commenting at regular intervals, while higher variance indicates bursty commenting behavior.
- Number of Unique Posts Commented On: The total number of distinct posts a user has commented on.
- Number of Unique News Outlets Commented On: The count of different news outlets on which the user has commented.
- Average Comments per Post: The average number of comments a user leaves on each post they engage with.
- Duplicate Comment Rate: The proportion of a user's comments that are exact duplicates.
- Average Media Usage: The average number of times a user includes media (e.g., photos, videos, GIFs) in their comments.

These features are computed for each user, forming a behavioral profile that aids in distinguishing human and bot-like activity. Through examination of these features, a structured and measurable approach to capturing patterns is created and a classification criteria for human and bot types can be established.

With the extracted features and manually annotated data, the proponents proceeded with classifying the remaining unlabeled users from the dataset. To achieve this, three classification models were implemented: Support Vector Machine (SVM), Random Forest Classifier (RFC), and Extreme Gradient Boosting (XGBoost). Since the dataset exhibited class imbalance, the Synthetic Minority Over-sampling Technique (SMOTE) was applied to ensure a more balanced distribution across categories.

For model optimization, hyperparameter tuning was performed for all three classifiers. The Random Forest model and XGBoost model were fine-tuned using GridSearchCV, allowing for an exhaustive search of the best hyperparameters. Meanwhile, the SVM classifier underwent tuning via RandomizedSearchCV, which provided an efficient approach by selecting random combinations of hyperparameters. After training, all models were evaluated using classification accuracy and additional performance metrics to assess their effectiveness in user classification. The better performing model is chosen to determine labels for the unlabeled users in the dataset.

D. Graph Neural Network for Classification

A graph is constructed to represent relationships in the dataset, with nodes representing users and edges capturing interactions. Edges specifically define user interactions such as user-comment edges linking users who reply to each other and co-comment edges linking users who commented on the same post. Each node is embedded with a 16-dimensional feature vector, comprising behavioral, linguistic, engagement, and temporal attributes derived from the user's activity to inform the classification process.

The proponents aim to compare two Graph Neural Networks (GNNs), namely Graph Convolutional Networks (GCNs) and Graph Attention Network (GATs) to determine the best performing model. GCNs are able to collate neighborhood information in a graph, making them effective in dealing with large-scale graphs [1]. On the other hand, GATs enhance this aggregated information through introducing attention mechanisms that highlight important connections within the graph, helping with capturing subtle differences in bot behavior [36].

These GNNs aim to classify users by analyzing their interactions and embedded features. This undertaking follows a supervised learning approach, where the models are trained on a labeled dataset to classify users into the following categories: Humans, Spam Bots, Cyborgs, or Trolls. Through leveraging these models, the researchers aim to determine which model architecture is best suited for detecting social media bots from genuine users, prompting better accuracy when detecting and more robust applications.

E. Evaluation of Graph Neural Networks

In this study, the model evaluation process is designed to assess the performance of both model architectures, integrating feature-based and graph-based approaches–ensuring effective detection of bot-like behavior across varied user interactions on Facebook. The evaluation utilizes a robust 10-fold cross-validation scheme, with each serving as validation once while the remaining folds train the model. Performance is measured using precision, recall, and F1-score, and Area Under the ROC Curve (AUC-ROC), ensuring a balanced assessment of bot detection accuracy.

III. RESULTS

A. Features Development

The results of the feature extraction yielded a classification criteria distinguishing four types of Facebook users: Humans, Spam Bots, Cyborgs, and Trolls. Each group exhibited distinct behavioral patterns across a variety of features, such as average comment length, engagement levels, response time, and linguistic dissimilarity.

TABLE I: SUMMARY OF EXTRACTED FEATURES BY USER TYPE

Feature	Humans	Spam Bots	Cyborgs	Trolls
Avg. Comment Length	85.75	89.65	184.25	96.24
Avg. Comments per	12.83	104	63.65	26.21

User				
Avg. Response Time	2200.15	3185.78	4989.61	9152.55
Avg. Engagement	41.42	41.69	251.29	85.17
Avg. Media Usage (per comment)	0.1	0.12	0.15	0.2
Avg. Link Usage (per comment)	0.48	0.6	0.78	0.42
Avg. Duplicate Comment Rate	0.02	0.5	0.1	0.03
Avg. Pairwise Dissimilarity	0.66	0.4	0.69	0.89
Avg. Thread Deviation	10,512.00	22,305.50	17,111.25	34,270.00
Avg. Innovation Rate	0.35	0.68	0.53	0.49
Comment Time Variance	3.15	5.7	6.9	7.28
Reply Ratio	0.28	0.17	0.35	0.22

Through data analysis, Cyborgs were characterized by significantly high average comment lengths, (x̄ = 184.25), high engagement values (x̄ = 251.29), and exceptionally high average response times (x̄ = 4989.61). This bot type also showed the highest average number of comments per user, at 63.65, high pairwise comment similarity at 0.69, and variability in comment times, exhibiting bursty but variable activity.

Humans, on the other hand, had moderate values for the majority of features. They were observed to have much fewer comments on average (x̄ =12.83), shorter average comment lengths (x̄ = 85.75), and low average link and media usage. Moreover, humans displayed low rates of duplicated comments and a very high pairwise dissimilarity mean.

Spam Bots demonstrated shorter comment durations, significantly higher comment volumes, and high duplicate comment rate. Engagement and reply ratio are lower than humans and cyborgs. Further, Spam Bots posted frequently across multiple outlets, displaying high thread deviation and innovation rates despite having lower complexity in comment content.

Lastly, Trolls maintained average values for most interaction-based features, but particularly noteworthy for their average media usage (x̄ = 0.20), engagement (x̄ = 85.17), and comment dissimilarity (x̄ = 0.89). Their comments frequently had comparatively long response durations (x̄ = 9152.55), considerable thread deviation (x̄ = 34,270), and the biggest variance in comment time (x̄ = 7.28).

B. Data Annotation

The evaluation of the classification algorithms with the help of the development of features revealed that the Random Forest classifier outperformed SVM and XGBoost in labeling the remaining unlabeled users into the four classifications (Humans, Spam Bots, Cyborgs, and Trolls). Each model's performance was assessed using classification accuracy, precision, recall, and F1-score. The results are summarized in Table II.

TABLE II: PERFORMANCE METRICS OF CLASSIFICATION MODELS

Model	Accuracy	Macro Precision	Macro Recall	Macro F1-score
SVM	0.79	0.8	0.79	0.79
RFC	0.84	0.84	0.84	0.84
XGBoost	0.83	0.83	0.83	0.83

The SVM classifier achieved an overall accuracy of 79%, with strong performance in classifying the Cyborg class, but showed weaker performance in Human and Spam Bot classification.The RFC classifier outperformed the two other models across all metrics, achieving an 84% rate and a macro F1 score off 0.84. The classifier demonstrated high precision and recall across all user types, especially Spam Bots. The XGBoost classifier closely followed with an accuracy rate of 83% and a macro F1 score of 0.83. It achieved perfect precision and recall for the Spam Bots class, but had slightly lower scores than the RFC model for CYborg and Troll classes.

Based on these results, the proponents selected the Random Forest model to classify the unlabeled users from the data set due to its superior performance and balanced classification capability across all user types.

C. Graph Neural Network for Classification

The constructed graph for this undertaking consists of 325,320 nodes and 6,112,052 edges, where each node represents an 16-dimensional feature vector, consisting of the same features discussed and analyzed in Chapter 2, Section B. Moreover, the graph also includes labels for each node which indicate their classification as either Human, Spam Bot, Cyborg, and Troll.

TABLE III: VISUALIZATION OF LARGE-SCALE GRAPH SUBSTRUCTURES FOR NODE CLASSIFICATION

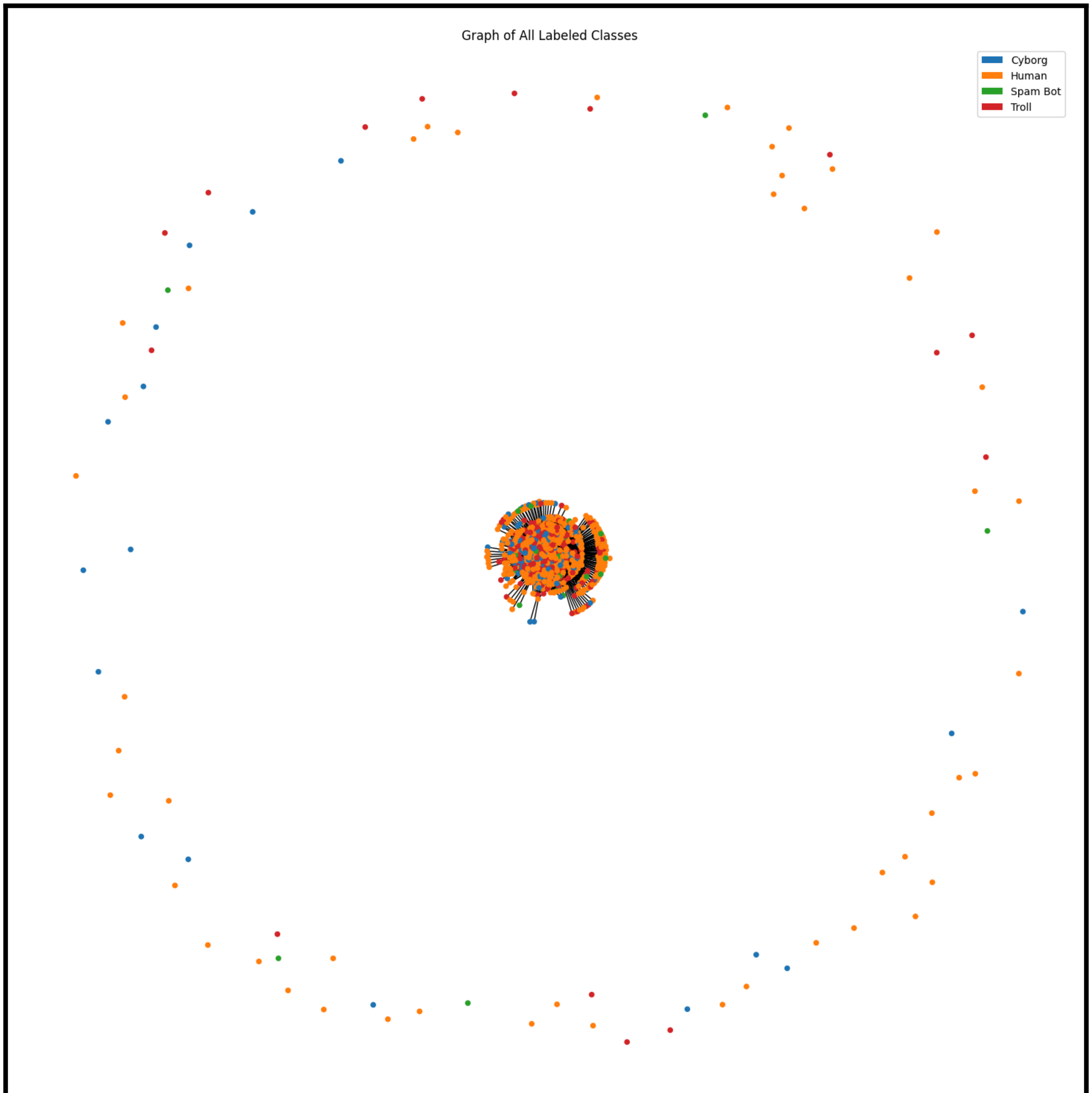


Fig. 2: Graph of Labeled Classes

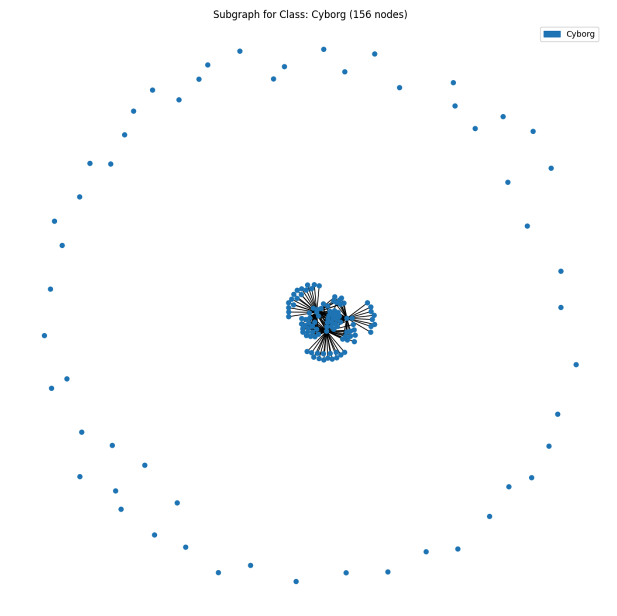


Fig. 3: Subgraph of Cyborg Class

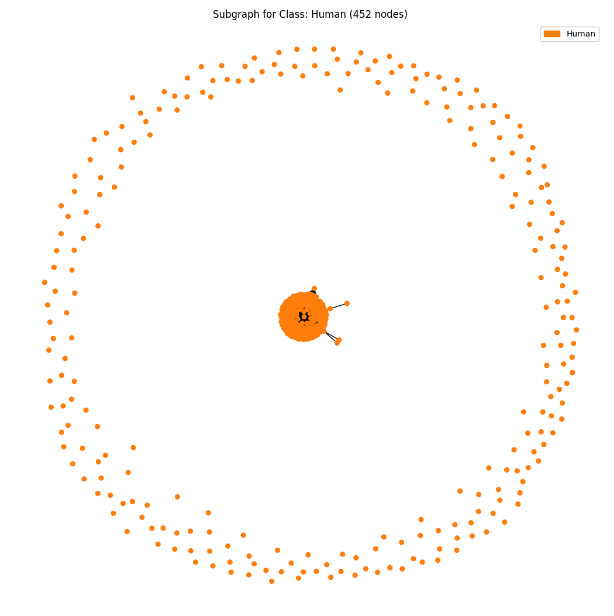


Fig. 4: Subgraph of Human Class

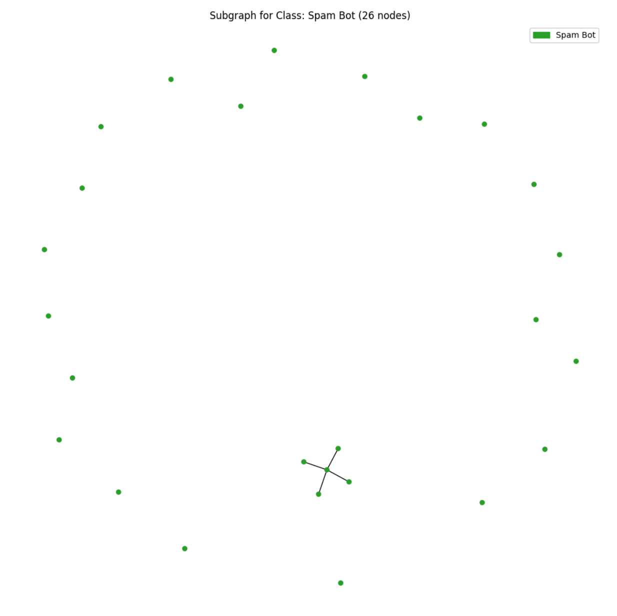


Fig. 5: Subgraph of Spam bot Class

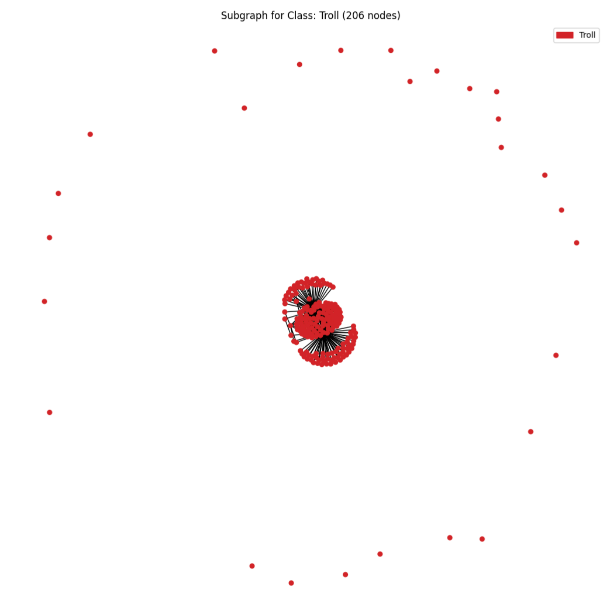


Fig. 6: Subgraph of Troll Class

The figures presented in Table III visualizes the distribution within the large-scale graph structure of the study's sample data. Figure 2 displays the entire labeled graph, where the four classes—Human, Spam Bot, Cyborg, and Troll—are distinctly colored to reflect their structural separations and clustering patterns within the graph. This serves as a macro-level overview of the node distributions and interconnections.

Figures 3 to 6 present isolated subgraphs for each specific class. Figure 3 shows the Cyborg class, which appears moderately connected and displays mixed characteristics between human and bot-like behaviors. Figure 4, representing the Human class, highlights a more densely clustered and interconnected structure, suggesting strong relational properties among genuine user nodes. Figure 5 focuses on the Spam Bot class, which is characterized by a more dispersed and loosely connected pattern, reflecting automated or script-based behavior with limited interaction. Lastly, Figure 6 visualizes the Troll class, which shows sporadic clustering, indicative of targeted disruptive behavior and isolated group formations.

These collectively provide valuable insights into the topological behavior of each class, reinforcing the role of structural features in enhancing node classification performance. They also serve as qualitative support for the effectiveness of Graph Neural Networks (GCN and GAT), which leverage these substructures to learn meaningful representations for accurate classification.

To model these patterns, both the GCN and GAT architectures employed consist of two layers. The first layer performs a graph convolution operation in GCN and a multi-head attention mechanism in GAT, enabling the models to aggregate information from neighboring nodes in structurally meaningful ways. The second layer produces the final classification outputs. Additionally, the GCN model uses a ReLU activation function, while the GAT model utilizes an ELU activation. Dropout layers and optimizers are integrated to enhance generalization and prevent overfitting, contributing to improved predictive performance.

D. Evaluation of Graph Neural Networks

Table IV and V present the performance metrics of the GCN and GAT model respectively, assessed using a 10-fold cross-validation scheme. Both models have strong accuracy rates at 90% and precision scores across all-folds. With the GAT model attaining 90.70% and GCN slightly higher at 90.71%, the average accuracy for both models is strikingly close. The slight variation implies that the two models are comparable in predicting user types.

Likewise, the two model’s precision scores over folds remain almost the same. However, it is noted that while GCN exhibits a higher mean precision score, its low recall affects its overall meaning and suggests a failure to identify positive instances of a class effectively.

TABLE IV: PERFORMANCE METRICS FOR 10-FOLD CROSS-VALIDATION OF GCN MODEL

Fold	Accuracy	Precision	Recall	F-1 score	AUC-ROC
1	90.92%	0.9773	0.2500	0.2381	0.4989
2	91.02%	0.9775	0.2500	0.2382	0.4751
3	90.79%	0.9770	0.2500	0.2379	0.4920
4	90.73%	0.9768	0.2500	0.2378	0.4963
5	90.31%	0.9758	0.2500	0.2373	0.4952
6	90.46%	0.9762	0.2500	0.2375	0.4592
7	90.76%	0.9769	0.2500	0.2379	0.4766
8	90.71%	0.9768	0.2500	0.2378	0.4653

9	90.76%	0.9769	0.2500	0.2379	0.5007
10	90.60%	0.9765	0.2500	0.2377	0.4819
Average	90.71%	0.9768	0.2500	0.2378	0.4841

In contrast, the GAT model has a much greater F1 score and recall than the GCN model, illustrating a better balance between recall and precision. This proves better generalization capability for the GAT model, especially when identifying minority classes such as Troll and Spam Bot. The GAT model's consistency in class distinction is further demonstrated by its higher AUC-ROC score of 0.7645 as opposed to GCN's 0.4841.

In conclusion, the GAT model shows a more balanced and efficient classification performance overall.

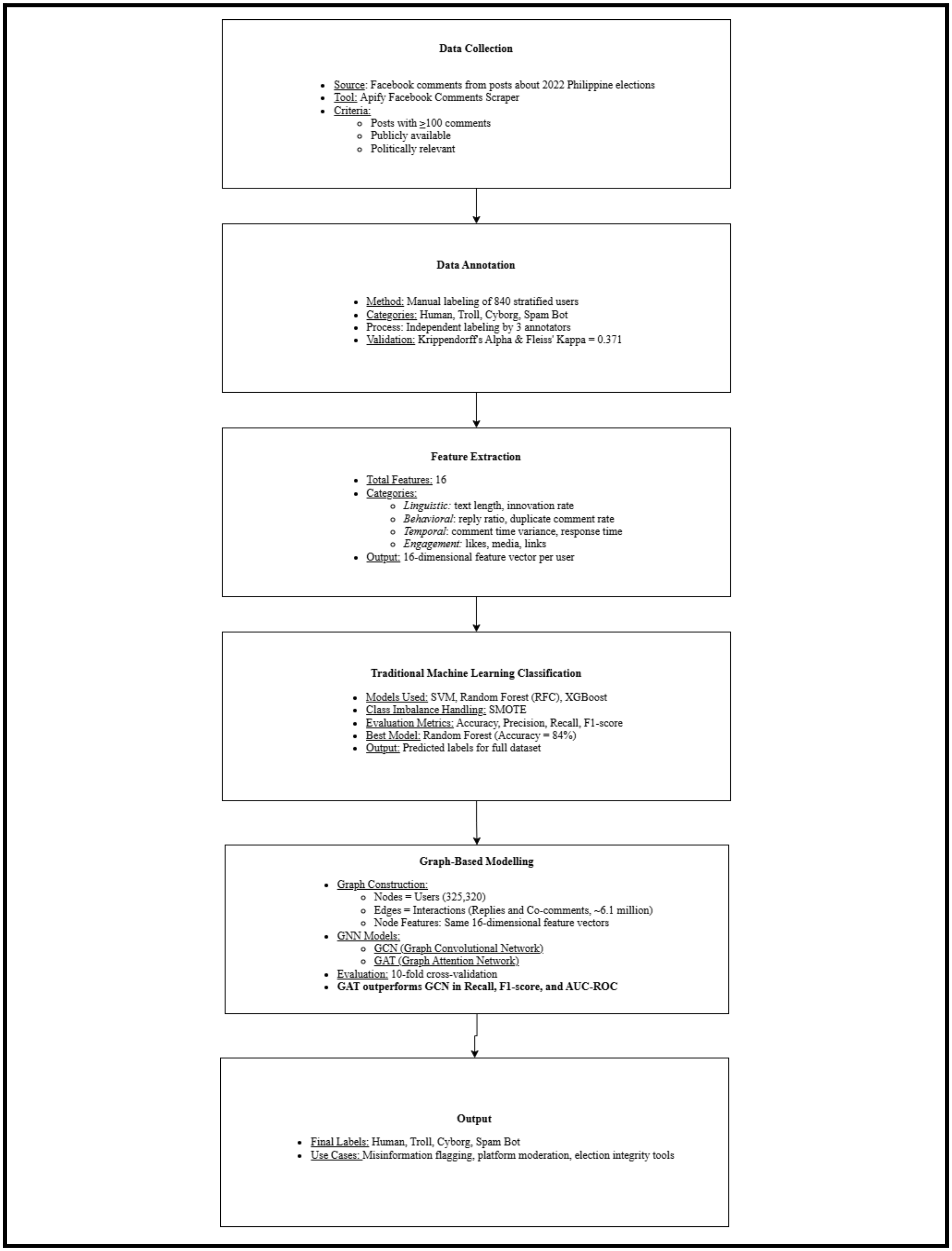
TABLE V: PERFORMANCE METRICS FOR 10-FOLD CROSS-VALIDATION OF GAT MODEL

Fold	**Accuracy**	**Precision**	**Recall**	**F-1 score**	**AUC-ROC**
1	90.92%	0.9773	0.5064	0.5833	0.7806
2	91.02%	0.9775	0.4616	0.5216	0.7927
3	90.79%	0.9770	0.4241	0.4553	0.7839
4	90.73%	0.9768	0.4710	0.5265	0.7745
5	90.31%	0.9758	0.4748	0.5398	0.7800
6	90.46%	0.9762	0.4236	0.4584	0.7209
7	90.66%	0.5148	0.4298	0.4667	0.734
8	90.71%	0.9768	0.4807	0.5465	0.8020
9	90.76%	0.9769	0.4562	0.5087	0.7426
10	90.60%	0.9765	0.4255	0.4646	0.7337
Average	90.70%	0.9306	0.4554	0.5071	0.7645

E. Framework Overview and Integration

This undertaking introduced Botbook, a graph-based framework designed to classify bot accounts on Facebook. Unlike single-model approaches, Botbook comprises interconnected stages, from data collection, feature engineering, classification, and graph-based modelling. It allows for a flexible and scalable solution for social bot classification in context-rich environments.

Figure 7 illustrates the complete pipeline of the Botbook framework, showcasing the step-by-step integration of data collection, annotation, feature engineering, traditional machine learning, and graph-based modeling for bot detection on Facebook. This modular flowchart demonstrates how each stage contributes to a unified detection system: beginning with ethically sourced, election-related Facebook comments, progressing through carefully validated manual annotations, and culminating in a hybrid classification process that combines behavioral profiling with graph neural networks.



Data Collection

- Source: Facebook comments from posts about 2022 Philippine elections
- Tool: Apify Facebook Comments Scraper
- Criteria:
  - Posts with ≥100 comments
  - Publicly available
  - Politically relevant

Data Annotation

- Method: Manual labeling of 840 stratified users
- Categories: Human, Troll, Cyborg, Spam Bot
- Process: Independent labeling by 3 annotators
- Validation: Krippendorff's Alpha & Fleiss' Kappa = 0.371

Feature Extraction

- Total Features: 16
- Categories:
  - *Linguistic:* text length, innovation rate
  - *Behavioral*: reply ratio, duplicate comment rate
  - *Temporal*: comment time variance, response time
  - *Engagement:* likes, media, links
- Output: 16-dimensional feature vector per user

Traditional Machine Learning Classification

- Models Used: SVM, Random Forest (RFC), XGBoost
- Class Imbalance Handling: SMOTE
- Evaluation Metrics: Accuracy, Precision, Recall, F1-score
- Best Model: Random Forest (Accuracy = 84%)
- Output: Predicted labels for full dataset

Graph-Based Modelling

- Graph Construction:
  - Nodes = Users (325,320)
  - Edges = Interactions (Replies and Co-comments, ~6.1 million)
  - Node Features: Same 16-dimensional feature vectors
- GNN Models:
  - GCN (Graph Convolutional Network)
  - GAT (Graph Attention Network)
- Evaluation: 10-fold cross-validation
- **GAT outperforms GCN in Recall, F1-score, and AUC-ROC**

Output

- Final Labels: Human, Troll, Cyborg, Spam Bot
- Use Cases: Misinformation flagging, platform moderation, election integrity tools

Fig. 7: Overview of the Botbook Detection Framework

The framework follows a modular pipeline that supports both traditional and graph-based machine learning methods. Starting from the collection of Facebook comment data from major Philippine news outlets during the 2022 Presidential elections. The collected comments capture a critical period of political discourse and activity, in which bot activity is recorded to be high. To ensure ethical standards and data validity, only publicly available comments were collected, with all personally identifiable information obfuscated.

A stratified sample set of users were manually annotated into four distinct categories—Human, Troll, Spam Bot, and Cyborg— based on commenting behavior and following annotation protocols established by previous researchers. These labels serve as the ground truth for supervised learning and basis for the construction of a robust classification model.

Using the labeled data, the framework extracted a diverse set of 16 features capturing user behavior across linguistic, behavioral, temporal, and engagement-based data. These features allowed for the development of a feature-rich behavioral profile for each user. Three classifiers, namely Random Forest Classification, Support Vector Machine, and XGBoost,were trained and compared using SMOTE to handle class imbalance. Upon comparison, the Random Forest model achieved the highest performance across all metrics and was chosen to label the remainder of the dataset.

Following classification, a static interaction graph was constructed, where each node represents a user and edges signify interaction patterns such as co-commenting and replying. Each node was embedded with a 16-dimensional feature vector, preserving the behavioral profile derived from earlier stages.

Two Graph Neural Network architectures were implemented: Graph Convolutional Network (GCN) and Graph Attention Network (GAT). While both models achieved similar accuracy (~90.7%), GAT demonstrated superior performance in recall, F1-score, and AUC-ROC, highlighting its strength in detecting minority classes and subtle patterns in user interactions. These results affirm the value of attention mechanisms in graph-based learning and validate the framework's choice of architecture for nuanced classification tasks.

Botbook serves as a unified framework rather than a collection of independent methods. It is distinguished by its integration of data preprocessing, human-in-the-loop annotation, traditional machine learning, graph construction, and advanced deep learning models into a unified pipeline. Each component builds upon the previous stage, and together, they form a system capable of adapting to new datasets, extending across different platforms, and evolving with bot behavior.

IV. DISCUSSION

The analysis of features reveals distinct behaviors exhibited by the four user categories. Cyborgs show a blend of human-like behavior and automation as seen by their high levels of engagement, lengthy comments, and bursty yet variable activity. Their rapid response speed and pairwise comment similarity point to a semi-automated interaction that maximizes output while preserving plausibility. Spam Bots, on the other hand, can be identified by their high content volumes, frequent cross-page activity, and high duplicate rations. These bots are likely to prioritize volume over quality in terms of content, with the objective of dominating discussions or spreading promotional content. However, despite their content simplicity, innovation and thread deviation scores of Spam Bots

Trolls are identified by their disruptive activity, shown in their high thread deviation scores, delayed response patterns, and strong media usage. Their significantly high dissimilarity scores and erratic response rates point to an intention to elicit, mislead, or disrupt online conversations. Finally, Humans exhibit expected patterns of organic interaction as seen in their moderate engagement, low frequency of comments, and high content dissimilarity.

These classifications provide valuable insights into the behaviors of Facebook users and can help identify non-human influences in online environments. The differentiation between user types offers a basis for developing classification criteria that can detect and categorize users based on their activity patterns.

For the models employed in this study, both GCN and GAT models were compared using key classification metrics: accuracy, precision, recall, F1-score, and AUC-ROC. Both models achieved similar accuracy, with the GCN model at 90.71% and the GAT model at 90.70%. This indicates that, in general, both models performed well in terms of overall classification. However, differences became apparent when examining the other metrics that provide more insight into the models' effectiveness in handling imbalanced data.

The GCN model, while maintaining high accuracy and precision, demonstrates consistently low recall (0.2500) and F1-score (0.2378). This suggests that the model is primarily learning to predict the majority class correctly while failing to identify minority class instances, a common issue in imbalanced datasets. The low AUC-ROC (0.4841) supports this interpretation, indicating poor discrimination capability between classes.

On the other hand, the GAT model showed a significantly higher F1-score (0.5071) and recall (0.4554), while keeping accuracy and precision at a respectable level. This suggests that GAT's attention method improves representation learning by enabling the model to pay greater attention to informative nodes and the context of their neighborhoods. Because of this, GAT is better at identifying positive examples, even when there is a class imbalance.

When evaluating the AUC-ROC, the higher average score of 0.7645 for GAT, as opposed to 0.4841 for GCN, suggests that GAT has a better overall ability to distinguish between classes. This reinforces the argument that GAT provides a more generalizable and robust performance, especially in detection applications where minority class detection is often critical. Although accuracy and precision are often emphasized, these findings underscore the importance of a holistic evaluation using recall, F1-score, and AUC-ROC. In scenarios where false negatives are costly, such as in safety-critical or security-sensitive tasks, GAT would be the more suitable model despite the similar performance on surface-level metrics like accuracy.

A. Conclusion

In conclusion, this paper presented Botbook as a unified, graph-based framework for social bot detection that reflects a deliberate integration of traditional and modern machine learning strategies, tailored to the unique challenges posed by Facebook comment data in the context of the 2022 Philippine elections. Beyond its technical architecture, the framework offers critical insights into how modular, context-aware detection systems can operate effectively in environments marked by data imbalance, privacy constraints, and evolving adversarial behaviors.

The comparative analysis of Graph Attention Networks (GAT) and Graph Convolutional Networks (GCN) based on 10-fold cross-validation provided comparable average accuracy (90.7%) and precision (0.97), further examination revealed that GAT consistently outperformed GCN in recall, F1-score, and AUC-ROC metrics. These results suggest that the attention mechanism in GAT enables more effective identification of minority class instances, contributing to better overall model generalization.

The findings underscore the importance of evaluating graph-based models using a comprehensive set of performance metrics, particularly in imbalanced classification tasks. Misclassifying users as bots can negatively impact user experience and business operations, leading to financial losses and degraded trust in platforms [14], [28]. Therefore, reducing false positives is critical for a bot detection framework. Given its improved balance between precision and recall, GAT demonstrates greater suitability for real-world applications where detecting minority class instances is essential.

The modular design of Botbook also enhances its adaptability. Each stage—data collection, feature engineering, user labeling, graph construction, and GNN training—operates independently, allowing the framework to evolve with the platform, dataset, or detection objectives. For instance, future iterations could incorporate dynamic graphs to capture temporal changes in behavior, or integrate multimodal features such as profile metadata or media content. This flexibility positions Botbook as more than just a static solution; it becomes a reusable and extensible architecture for detecting complex malicious actors across online platforms.

B. Recommendations for Future Research

With the existing limitations of this research, future undertakings for researchers to improve and expand upon the current work can focus on areas such as the following:

- Application of the developed model to social media platforms (namely, Facebook) for real-time detection and flagging of misleading or malicious content: Given that the current study employed static graph analysis and acknowledged that a static graph is unable to account for temporal changes or evolving behaviors, future research should prioritize (a) developing dynamic graph models that can capture the temporal evolution of bot behaviors and (b) investigating methods for real-time feature extraction and analysis to enable timely detection and flagging of malicious content.
- Expanding the data set to include the user profile's background and relationships between other users to improve the model's graph construction: Considering Facebook's privacy policy regarding user data, future work would focus on (a) investigating alternative data sources or anonymization techniques, (b) exploring the use of additional user features beyond comments, such as profile information, posting history, etc. and (c) considering the incorporation of multimodal data, such as images or videos, to provide further insights into bot behavior. Additionally, this includes exploring contexts beyond the 2022 Philippine Election to improve the refinement of classification techniques for evolving bots.

Future researchers can significantly refine bot detection methodologies and broaden the spectrum of analysis to contribute to safer and more trustworthy online spaces by more effectively detecting and classifying malicious bots in everyday conversations.

REFERENCES

[1] Mohammad Abrar, Md Habibur Rahman, Md Abdul Aziz, Jung-In Baik, Young-Hwan You, and Hyoung-Kyu Song. 2023. Graph Convolutional Network Design for Node Classification

Accuracy Improvement. *Mathematics* 11, 17 (August 2023), 3680–3680. DOI:https://doi.org/10.3390/math11173680

[2] Muhammad Abulaish and Mohd Fazil. 2020. Socialbots: Impacts, Threat-Dimensions, and Defense Challenges. *IEEE Technology and Society Magazine* 39, 3 (September 2020), 52–61. DOI:https://doi.org/10.1109/mts.2020.3012327

[3] Eric Blancaflor, James V Taylor, Caitlin D Datu, Jared Karll, and Kennichi O Nitta. 2023. Bots Gone Rogue: Exploring the Negative Outcomes in the Philippines. (November 2023). DOI:https://doi.org/10.1109/icecet58911.2023.10389532

[4] Braeden Bowen. 2021. "IT DOESN'T MATTER NOW WHO'S RIGHT AND WHO'S NOT:" a Model to Evaluate and Detect Bot Behavior on Twitter. *OhioLink ETD Center*. Retrieved from https://www.academia.edu/95172692/_IT_DOESN_T_MATTER_NOW_WHO_S_RIGHT_AND_WHO_S_NOT_A_Model_to_Evaluate_and_Detect_Bot_Behavior_on_Twitter

[5] Sky Chatuchinda. 2022. Social Media and Disinformation : Fake Accounts, Bots, and Trolls: How Social Media Influences Philippine's Future. *Friedrich Naumann Foundation for Freedom*. Retrieved from https://www.freiheit.org/southeast-and-east-asia/fake-accounts-bots-and-trolls-how-social-media-influences-philippines

[6] Chun Cheng, Yun Luo, and Changbin Yu. 2020. Dynamic mechanism of social bots interfering with public opinion in network. *Physica A: Statistical Mechanics and its Applications* 551, (August 2020), 124163. DOI:https://doi.org/10.1016/j.physa.2020.124163

[7] Ashkan Dehghan, Kinga Siuta, Agata Skorupka, Akshat Dubey, Andrei Betlen, D H Miller, Wei Xu, Bogumił Kamiński, and Paweł Prałat. 2023. Detecting bots in social-networks using node and structural embeddings. *Journal of Big Data* 10, 1 (July 2023). DOI:https://doi.org/10.1186/s40537-023-00796-3

[8] Phillip George Efthimion, Scott Payne, and Nicholas Proferes. 2018. Supervised Machine Learning Bot Detection Techniques to Identify Social Twitter Bots. *SMU Scholar*. Retrieved from https://scholar.smu.edu/datasciencereview/vol1/iss2/5/

[9] Zineb Ellaky, Faouzia Benabbou, and Sara Ouahabi. 2023. Systematic Literature Review of Social Media Bots Detection Systems. *Journal of King Saud University - Computer and Information Sciences* 35, 5 (May 2023), 101551–101551. DOI:https://doi.org/10.1016/j.jksuci.2023.04.004

[10] Edmund Genfi. 2021. Detecting Bots Using a Hybrid Approach. *Theses, Dissertations and Culminating Projects* (May 2021). Retrieved May 21, 2022 from https://digitalcommons.montclair.edu/etd/736/

[11] Steven Gianvecchio, Mengjun Xie, Zhenyu Wu, and Haining Wang. 2011. Humans and Bots in Internet Chat: Measurement, Analysis, and Automated Classification. *IEEE/ACM Transactions on Networking* 19, 5 (October 2011), 1557–1571. DOI:https://doi.org/10.1109/tnet.2011.2126591

[12] Robert Gorwa and Douglas Guilbeault. 2018. Unpacking the Social Media Bot: A Typology to Guide Research and Policy. *Policy & Internet* 12, 2 (August 2018), 225–248. DOI:https://doi.org/10.1002/poi3.184

[13] Qinglang Guo, Haiyong Xie, Yangyang Li, Wen Ma, and Chao Zhang. 2021. Social Bots Detection via Fusing BERT and Graph Convolutional Networks. *Symmetry* 14, 1 (December 2021), 30. DOI:https://doi.org/10.3390/sym14010030

[14] Patience Haggin. 2025. Efforts to Weed out Fake Users for Online Advertisers Fall Short. *The Wall Street Journal*. Retrieved April 3, 2025 from https://www.wsj.com/business/media/efforts-to-weed-out-fake-users-for-online-advertisers-fall-short-0a5ec1a6

[15] Nick Hajli, Usman Saeed, Mina Tajvidi, and Farid Shirazi. 2021. Social Bots and the Spread of Disinformation in Social Media: the Challenges of Artificial Intelligence. *British Journal of Management* 33, 3 (October 2021). DOI:https://doi.org/10.1111/1467-8551.12554

[16] Kadhim Hayawi, Susmita Saha, Mohammad Mehedy Masud, Sujith Samuel Mathew, and Mohammed Kaosar. 2023. Social media bot detection with deep learning methods: a systematic review. *Neural Computing and Applications* (March 2023). DOI:https://doi.org/10.1007/s00521-023-08352-z

[17] Buyun He, Yingguang Yang, Qi Wu, Hao Liu, Renyu Yang, Hao Peng, Xiang Wang, Yong Liao, and Pengyuan Zhou. 2024. Dynamicity-aware Social Bot Detection with Dynamic Graph Transformers. *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence (IJCAI '24)* (July 2024), 5844–5852. DOI:https://doi.org/10.24963/ijcai.2024/646

[18] Maryam Heidari, James H Jr Jones, and Ozlem Uzuner. 2021. An Empirical Study of Machine Learning Algorithms for Social Media Bot Detection. *IEEE Xplore*, 1–5. DOI:https://doi.org/10.1109/IEMTRONICS52119.2021.9422605

[19] Linda Li, Orsolya Vásárhelyi, and Balázs Vedres. 2024. Social bots spoil activist sentiment without eroding engagement. *Scientific Reports* 14, 1 (November 2024). DOI:https://doi.org/10.1038/s41598-024-74032-0

[20] Feng Liu, Zhenyu Li, Chunfang Yang, Daofu Gong, Haoyu Lu, and Fenlin Liu. 2024. SEGCN: a Subgraph Encoding Based Graph Convolutional Network Model for Social Bot Detection. *Scientific Reports* 14, 1 (February 2024), 4122. DOI:https://doi.org/10.1038/s41598-024-54809-z

[21] Salvador Lopez-Joya, J. Angel Diaz-Garcia, M Ruiz Dolores, and Maria J. Martin-Bautista. 2023. Bot Detection in Twitter: an Overview. *Lecture Notes in Computer Science* (September 2023), 131–144. DOI:https://doi.org/10.1007/978-3-031-42935-4_11

[22] Maria Clarisse J. Loro. 2023. *DECODING THE DIGITAL FRONTIER: LEGAL ANALYSIS OF MICROTARGETING, BOTS, AND AI-POWERED ELECTION PROPAGANDA IN THE PHILIPPINES*. Retrieved from https://libpros.com/wp-content/uploads/2024/09/LORO-Ma.-Clarisse_Dissertation-Ma.-Clarisse-Loro.pdf

[23] Luca Luceri, Felipe Cardoso, and Silvia Giordano. 2021. Down the bot hole: Actionable insights from a one-year analysis of bot activity on Twitter. *First Monday* (February 2021). DOI:https://doi.org/10.5210/fm.v26i3.11441

[24] Wentao Ma. 2024. LLM Echo Chamber: personalized and automated disinformation. *arXiv.org*. Retrieved December 12, 2024 from https://arxiv.org/abs/2409.16241

[25] David Martin-Gutierrez, Gustavo Hernandez-Penaloza, Alberto Belmonte Hernandez, Alicia Lozano-Diez, and Federico Alvarez. 2021. A Deep Learning Approach for Robust Detection of Bots in Twitter Using Transformers. *IEEE Access* 9, (2021), 54591–54601. DOI:https://doi.org/10.1109/access.2021.3068659

[26] Mariam Orabi, Djedjiga Mouheb, Zaher Al Aghbari, and Ibrahim Kamel. 2020. Detection of Bots in Social Media: A Systematic Review. *Information Processing & Management* 57, 4 (July 2020), 102250. DOI:https://doi.org/10.1016/j.ipm.2020.102250

[27] Derrick Paulo. 2022. Trolls for hire in Philippines: The concealed political weapon used in a social media war. *Channel News Asia*. Retrieved from https://www.channelnewsasia.com/cna-insider/paid-troll-army-hire-philippines-social-media-elections-influencers-2917556

[28] Adrian Rauchfleisch and Jonas Kaiser. 2020. The False Positive Problem of Automatic Bot Detection in Social Science Research. *PLOS ONE* 15, 10 (October 2020), e0241045. DOI:https://doi.org/10.1371/journal.pone.0241045

[29] Jorge Rodríguez-Ruiz, Javier Israel Mata-Sánchez, Raúl Monroy, Octavio Loyola-González, and Armando López-Cuevas. 2020. A one-class classification approach for bot detection on Twitter. *Computers & Security* 91, 0167-4048 (April 2020), 101715. DOI:https://doi.org/10.1016/j.cose.2020.101715

[30] Giancarlo Ruffo, Alfonso Semeraro, Anastasia Giachanou, and Paolo Rosso. 2023. Studying fake news spreading, polarisation dynamics, and manipulation by bots: A tale of networks and language. *Computer Science Review* 47, (February 2023), 100531. DOI:https://doi.org/10.1016/j.cosrev.2022.100531

[31] Giovanni C. Santia, Munif Ishad Mujib, and Jake Ryland Williams. 2019. Detecting Social Bots on Facebook in an Information Veracity Context. *Proceedings of the International AAAI Conference on Web and Social Media* 13, (July 2019), 463–472. DOI:https://doi.org/10.1609/icwsm.v13i01.3244

[32] Shuhao Shi, Kai Qiao, Zhengyan Wang, Jie Yang, Baojie Song, Jian Chen, and Bin Yan. 2023. Muti-scale Graph Neural Network with Signed-attention for Social Bot Detection: A Frequency Perspective. *ArXiv (Cornell University)* (January 2023). DOI:https://doi.org/10.48550/arxiv.2307.01968

[33] Thales. 2024. Bots Now Make up Nearly Half of All Internet Traffic Globally. *Thales Group*. Retrieved from https://www.thalesgroup.com/en/worldwide/security/press_release/bots-now-make-nearly-half-all-internet-traffic-globally

[34] María José Díaz Torres and Antonio Rico-Sulayes . 2021. Detection of bot accounts in a Twitter corpus: Author profiling of social media users as human vs. nonhuman}. *Lengua y Habla* 25, (2021), 76--86.

[35] Julien Tourille, Babacar Sow, and Adrian Popescu. 2022. Automatic Detection of Bot-generated Tweets. *Association for Computing Machinery* (June 2022). DOI:https://doi.org/10.1145/3512732.3533584

[36] Petar Velickovic, Guillem Cucurull, Arantxa Casanova, Adriana Romero, P. Lio', and Yoshua Bengio. 2018. Graph Attention Networks. *ICLR* (2018). DOI:https://doi.org/10.17863/CAM.48429

[37] Yuhao Wu, Yuzhou Fang, Shuaikang Shang, Jing Jin, Lai Wei, and Haizhou Wang. 2021. A Novel Framework for Detecting Social Bots with Deep Neural Networks and Active Learning. *Knowledge-Based Systems* 211, (January 2021), 106525. DOI:https://doi.org/10.1016/j.knosys.2020.106525