A Hybrid Approach Combining Network Analysis and Deep Learning for Instagram Political Bot Detection

Mohammadhessan Badami
hesbadami@gmail.com
Brunel University London
Uxbridge, UK

Cristian Mateos
cristian.mateos@isistan.unicen.edu.ar
ISISTAN (UNICEN–CONICET)
Tandil, Buenos Aires, Argentina

Matías Hirsch
matias.hirsch@isistan.unicen.edu.ar
ISISTAN (UNICEN–CONICET)
Tandil, Buenos Aires, Argentina

Tor-Morten Grønli
tor-morten.gronli@kristiania.no
Kristiania University College
Oslo, Norway

Tim A. Majchrzak
tim.majchrzak@ruhr-uni-bochum.de
University of Bochum & Center for Advanced Internet Studies
Bochum, Germany

Gheorghita Ghinea
george.ghinea@brunel.ac.uk
Brunel University London
Uxbridge, UK

Abstract

We propose a hybrid approach combining network analysis with advanced machine learning (ML) techniques to understand the role of social network bots in influencing political discourse. Data from the UK's 2024 public election was collected from Instagram using custom web scraping tools, subsequently anonymized, and preprocessed to ensure ethical compliance. Network analysis was conducted using centrality measures and community detection algorithms, while classic ML models including Random Forest (RF) and XGBoost (XGB) were developed and fine-tuned to detect bots and complemented with a RoBERTa model to detect AI-generated content. Interestingly, experiments show that while bots were present, they did not dominate the discussions. This suggests a strategy of subtle influence rather than overt manipulation. This finding contrasts with existing literature focused on other social media platforms, thus providing a new perspective on bot behavior and insights to mitigate the impact of bots and enhance the integrity of online political engagement.

CCS Concepts

• **Computing methodologies** → **Boosting**; **Information extraction**;
• **Information systems** → **Social networking sites**.

Keywords

Social network, Instagram, bot, political discourse, network analysis, machine learning

ACM Reference Format:
Mohammadhessan Badami, Cristian Mateos, Matías Hirsch, Tor-Morten Grønli, Tim A. Majchrzak, and Gheorghita Ghinea. 2026. A Hybrid Approach Combining Network Analysis and Deep Learning for Instagram Political Bot Detection. In *The 41st ACM/SIGAPP Symposium on Applied Computing (SAC '26), March 23–27, 2026, Thessaloniki, Greece.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3748522.3779720

SAC '26, Thessaloniki, Greece

ACM ISBN 979-8-4007-2294-3/2026/03
https://doi.org/10.1145/3748522.3779720

1 Introduction

The advent of social media has transformed the landscape of political communication. Using platforms like Instagram and Twitter, politicians, parties, and the public engage in debates, share opinions, and mobilize support [17]. A challenge is the rise of automated accounts or *bots* that can manipulate online conversations and influence public opinion. Bots can spread misinformation or amplify certain viewpoints, thereby skewing the perception of political realities [10].

The existing literature highlights the challenges of detecting bots due to their evolving sophistication and the diverse tactics they employ [8]. Accordingly, we propose a hybrid approach that combines network analysis with advanced ML techniques to effectively identify and understand the behavior of bots on Instagram, and evaluate the approach by quantifying the presence of bots within the political discourse on Instagram during the UK's 2024 election.

By developing and applying advanced detection methodologies, our research provides insights into the mechanisms through which bots operate and influence political discourse. The findings have potential implications for policy makers, social media platforms and researchers, offering strategies to mitigate the impact of bots and improve the integrity of online political participation [28]. Furthermore, the proposed hybrid approach can serve as a model for bot detection and analysis on various social media platforms [26]. Eventually, our work could thus pose to be beneficial to end-users.

Section 2 provides an overview of the existing research on the detection of social media bots, their detection methods, and their impact on political discourse. Section 3 outlines the research design, data collection methods, and analytical techniques used. Section 4 presents the results of the data analysis, including the prevalence of bots and their strategies within the Instagram network. Section 5 synthesizes the findings, discusses the implications, and suggests directions for future research.

2 Background and Literature review

2.1 Impact of Bots on Social Media Manipulation

Politics, drawing from modern democracy, stands as one of the most prominent areas revolutionized by online communities. This has led to a growing interest in understanding how data from public opinions on social media can influence political policies and positions. Additionally, there is potential to exploit these data to build

community support in political campaigns – recall the successful use of social media by Barack Obama in the 2010 US Presidential Election [27].

In fact, [15] identified that Twitter can be used to predict election outcomes, since sentiment analysis shows a strong correlation with traditional polls. Results from other contributions also further confirm the capability of social media in predicting election outcomes with varying accuracy, with sentiment analysis identified as the most accurate predictor, and social network analysis providing stable predictions despite surges in political discussions [16]. However, bot detection is still under development. Although some studies have investigated the activity of fake accounts on Instagram [3], only two studies by [1] and [25] have utilized ML approaches for accurately detecting automated Instagram accounts. Accordingly, the former study made significant contributions by creating two labeled public datasets – one for fake accounts (1203 accounts) and another for bots (1400 accounts). However, both datasets had issues, particularly an unnatural bias in the automated accounts dataset.

Having observed this ideal vehicle for boosting community support, and the lack for proper bot detection support, politicians across the world have started to eagerly adopt social bots as part of their strategies. On the path to exploring how political actors globally use social bots to manipulate public opinion and disrupt communication, [28] shows that political bots are widely used by countries like Mexico, Turkey, and Russia to suppress dissent and promote government propaganda, particularly during elections and political crises. He emphasizes the need for better detection methods to combat computational propaganda. [4] also showed significant bot activity in promoting and opposing political candidates during the 2014 presidential elections in Brazil. Stepping beyond presidential elections, [6] showed that social bots were found to amplify divisive narratives (e.g., *anti-mask*) and create echo chambers by sharing content within their political lines during the COVID-19 pandemic.

2.2 Social Media Opinions as a Complex Network

Due to the fundamental complexity of social media on the one hand, and the numerous successes of network science and complex systems scientists on the other hand, a link between opinion mining, social media analysis, and network science has led to the emergence of a branch that studies the social media as a network system. Recent developments in the study of complex networks have been heavily inspired by empirical studies of systems such as the Internet, social networks, and biological networks [22]. Mathematical modelling and statistical analysis of network properties such as the small-world effect and power-law degree distributions can serve as underlying properties across various types of networks [22].

In related work, [11] explain how aggregated data from mobile phone usage can be utilized to create models that predict the spread of information, behaviors, or diseases through a population with an approach based on complex networks. Problems such as spam on Twitter, particularly around the hashtags, have been explored by [29], in which they used a combination of manual coding and algorithmic detection to identify spam. Their use of network analysis tools such as Network Workbench and GUESS identified that 14% of tweets centered around the #robotpickuplines hashtag were spam.

2.3 Hybrid Approach to Social Bot Detection

Recent studies have investigated the combination of ML and network analysis for the problem at hand. A notable effort in this respect is that of [21], which also borrows concepts from statistical physics such as Random Matrix Theory (RMT) to advance bot detection and opinion leader identification in social networks.

The SEGCN model by [19] enhances the expressive power of traditional Graph Convolutional Networks (GCNs) by utilizing subgraph encoding to capture structural features like cycles and triangles. SEGCN incorporates semantic, property, and structural features of social media accounts, increasing bot detection accuracy. Experiments on datasets such as Twibot-20 and Twibot-22 demonstrate that SEGCN outperforms state-of-the-art models. [2] showed that GCNs outperform MLP (Multi-Layer Perceptron), emphasizing the importance of considering both profile features and social graph structures.

Following this, [7] combine graph-based features and SOMs (Self-Organizing Maps) to identify bots by analyzing the topological structure of the network, using features such as in-degree, out-degree, and so on. The approach isolates bots into smaller clusters, enhancing the efficiency of detection. Experiments using the CTU-13 dataset show that the approach can detect various types of bots with high accuracy and low computational cost.

With the goal to capture bots that disguise as genuine users and those that act collectively, [9] present BotRGCN, a model that uses Relational GCNs (R-GCN) to construct a heterogeneous graph from follow relationships and applies multi-modal user information encoded using RoBERTa to improve detection accuracy. BotRGCN achieves state-of-the-art performance on the TwiBot-20 benchmark, outperforming SEGCN and several baseline methods.

2.4 Summary

According to the explored literature, the **selection of ML models** for bot detection in social media is driven by the need for high accuracy and robustness against evolving bot behaviors. Accordingly, ensemble tree methods, such as Random Forest (RF), were the preferred choice of most researchers, constituting the largest share (Figure 1). The success of RF, as demonstrated by [13], who achieved an accuracy of 99.99%, inspired us to employ Random Forests in our methodology while exploring the political sphere on Instagram.

GCNs and Convolutional Neural Networks (CNNs) are also prominently used, which are particularly effective in capturing the complex interactions and relationships within social media networks. Decision Trees, including models like C4.5 and J48, make up 13.3%, highlighting their simplicity and interpretability. Logistic Regression, LSTM (Long Short-Term Memory), and various clustering techniques also feature significantly, reflecting the diverse approaches taken by researchers to tackle the bot detection problem.

Feature selection also plays a crucial role in the effectiveness of bot detection models. We incorporate a comprehensive set of features in various categories, inspired by previous successful implementations. Accordingly, network features are the most frequently used (see Figure 2). These features are critical in understanding the user's position and influence within the social network, which is often indicative of bot activity.

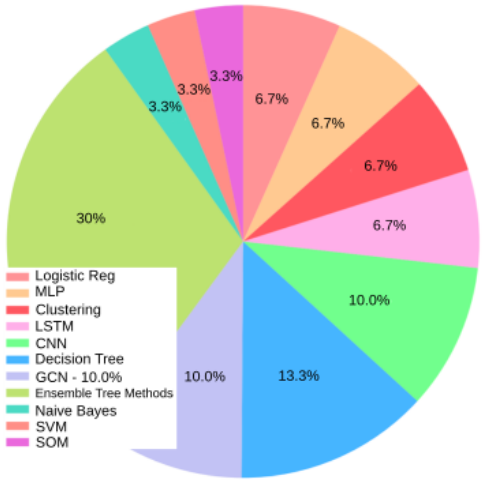

Figure 1: Distribution of used models.

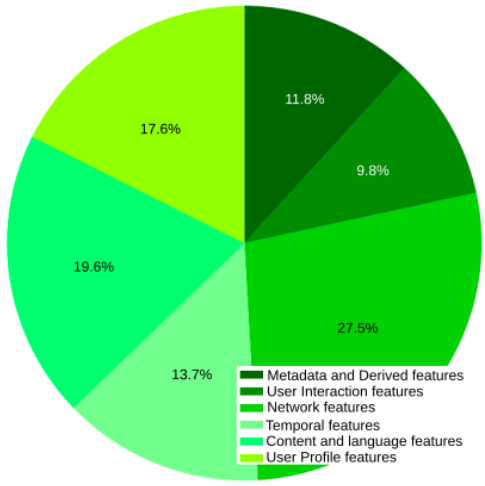

Figure 2: Distribution of feature types used.

Content and language features, including comment length, hashtag usage, and sentiment analysis, offer insights into the behavior and intentions of the user. These features are particularly useful in identifying automated content generation and spam-like behaviors. User profile features, such as account age, number of followers, and verified status, are fundamental in distinguishing bots from genuine users. Temporal features, such as burstiness and diurnal activity patterns, help capture the timing and frequency of user interactions. User interaction features, such as comment and mention networks, provide additional context on the user's engagement with others. Metadata and derived features, such as profile image and bio characteristics, further enhance the detection model by incorporating auxiliary information about the user. Our feature engineering process was heavily influenced by the work of [9], who emphasized the importance of diverse and comprehensive feature sets in improving detection accuracy.

3 Methodology

3.1 Research Objectives

With an overarching claim of studying the 2024 UK general election, our study approaches to gather and analyze the Instagram comments related to the political figures involved in the general election to propose a discussion revolving around two opposing user types: bots and authentic users. We summarize our Research Objectives as follows:

Objective 1 (RO1): Quantify the presence of bots in the political discourse during the UK's 2024 general election process.

Objective 2 (RO2): Analyze the strategies and activities of bots, including their interaction patterns and influence on the network.

For Objective 1, we check for the presence of the bots during the campaign period centered on the three main parties introduced below. As for Objective 2, by studying the network threading through these bots, we aim to see how the bots operate through the political landscape.

The structure of the above can be summarized as follows: We initiated data collection by focusing on the main three candidates of the 2024 UK general election, namely Rishi Sunak of the Conservatives, Keir Starmer of the Labour party, and Ed Davey of the Liberal Democrats, spanning from May 2nd until July 11th, 2024. Subsequently, we constructed a reply network and employed a combination of ML models on user profile features and natural language processing on these opinions to label the opinions in this discourse with a bot score.

3.2 Data Collection and Storage

Through meticulous effort at reverse engineering the Instagram's GraphQL query system, we manually scraped all the user profiles, posts, and comments belonging to the leaders of the three main parties. Over the course of the election campaign (May 2nd, 2024 - July 11th, 2024), all comments were collected and warehoused in an Apache Hive table; a data warehouse software project built on top of the well-known Apache Hadoop middleware. These comments of discourse towards the main parties of the election on Instagram were made by users, the feature set of whom, in coherence to Ethical norms, was purged of all data from which the commentators' and users' identities can be inferred. This means that meta-data such as usernames, profile images, and profile biographies, were dropped in our attempt to explore the viability of maintaining high accuracy within the restrictions of highest ethical standards, relying mostly on engineering textual features from comments and integrating network analysis.

3.3 Supervised Learning of User Profile Features

Available labeled datasets for training bot detection models in Instagram are very limited, with only one peer-reviewed publicly contributed dataset in the scholar landscape from [1] who in their work contributed by generating a labeled dataset for bots (1400 accounts), which however has problems. The dataset has an unnatural bias, as a correction of which they themselves implemented cost-sensitive genetic algorithms. To supplement this lack of data for bots and genuine accounts, we attempted training another model contributing to the verdict, but with a simplifying assumption, that

since all harmful bots possess fake identities, using a dataset of fake accounts, will result in using a set of Instagram accounts which could be considered a superset for Instagram bots. This same assumption was made by [8], when they were preparing the dataset of Twitter bots that was used for developing the famous bot detection tool Botometer [26], a successful model used in a huge percentage of the literature, confirming the validity for our assumption. To account for the user profile features of our dataset, RFs were chosen as the classifiers. Their success, as demonstrated by [13] who achieved an accuracy of 99.99%, was the major contributor to their employment in our methodology of exploring social bots' presence in the political sphere of Instagram.

3.4 AI-Text Classification Task

We complement our two RF models, with an analysis over the comments' text to detect AI-written content. To do this, we fine-tuned and benchmarked a RoBERTa model [20] on HC3 [12], an up-to-date ChatGPT text detection dataset to classify the comment text as bot generated or not. RoBERTa modifies key hyperparameters, removes the next sentence prediction objective, and employs much larger mini-batches and learning rates during training. By eliminating the next-sentence prediction task, it focuses solely on the masked language modeling (MLM) objective. This adjustment, coupled with the use of larger mini-batches and higher learning rates, enables the model to process more data in less time, thereby enhancing its language understanding capabilities. Furthermore, the model is trained on an order of magnitude more data than BERT, including the novel CC-News dataset composed of public news articles. These enhancements result in significant performance improvements across a range of NLP benchmarks, such as MNLI, QNLI, RTE, STS-B, RACE, and the GLUE benchmark [20]. Following the approach of [24], in order to improve on the negative impact of text length on detectors' performance, we labeled machine-generated texts that are overly short and simple as "Unlabeled". Such texts are too short and as a consequence, are highly similar to human, making their detection difficult. As our Instagram comments' texts are shorter and simpler, the "Unlabeled" property dominates them, and in this sense, we modelled our AI-Text classification task as a partial Positive-Unlabeled (PU) problem, and employed the Multiscale Positive-Unlabeled (MPU) training framework, formulated by [24], to boost our detector's performance.

3.5 Ensemble Voting System

Having fine-tuned a RoBERTa model for AI-Text classification and two RF classifiers on feature-rich but relatively small datasets, rigorously cross-validated and benchmarked with 10 folds to make sure each model's generalizability is assessed properly, we developed a voting mechanism, where we utilized a weighted approach based on the Mean Macro F1 score achieved by each model during benchmarking. The use of Macro F1 score has been endorsed in several studies focusing on bot detection and similar classification problems [9, 20].

We labeled all the users in our scraped dataset of Instagram comments based on the aggregated predictions from our ensemble of models. Users identified as bots by the majority vote, weighted by the mean Macro F1 scores, were classified accordingly. This method follows the methodology of combining classifiers to improve detection performance, as demonstrated previously [10]. The weighted voting ensures that models with higher predictive performance have a greater influence on the final decision, improving overall accuracy and robustness. Our approach of employing an ensemble of three models combines the strengths of each individual model, leveraging their unique insights to enhance detection capabilities. The F1 score, which is the harmonic mean of precision and recall, provides a balanced measure of a model's performance, particularly in cases of imbalanced datasets where one class may be under-represented [18]. The Macro F1 score is particularly beneficial in our context as it treats all classes equally, preventing the model from favoring the majority class. This ensures a comprehensive evaluation of the model's ability to detect both bots and genuine users accurately.

3.6 Network Analysis

To understand the engagement strategies of bots within the political discourse, we constructed a user mention network where users are represented as nodes and the mentions between them as directed edges. A mention indicates that one user referenced another in their comment (e.g., @username). The process began with data preparation, where we loaded the dataset containing user comments and extracted mentions using regular expressions to identify patterns like @username. This extraction enabled us to form edges that connect the nodes in the network, signifying the interactions between users.

For constructing the network, we utilized NetworkX, a Python library designed for the creation, manipulation, and study of complex networks. Once the network was constructed, we focused on identifying different aspects of user interactions and influence within the network using several key metrics. Specifically, we calculated degree centrality, betweenness centrality, and eigenvector centrality for each node. The degree centrality measures the number of incoming and outgoing edges for each node, representing how connected a node is within the network, providing a straightforward indicator of a user's activity within the network. Mathematically, the degree centrality $C_D(v)$ of a node v is defined by [5, 22] as:

$$C_D(v) = \frac{deg(v)}{N-1} \tag{1}$$

where $deg(v)$ is the degree of node v and N is the total number of nodes in the network.

Another crucial metric is the Betweenness Centrality, which quantifies the number of times a node acts as a bridge along the shortest path between two other nodes. This measure highlights nodes that are crucial and influential for information flow within the network. Mathematically, the betweenness centrality $C_B(v)$ of a node v is given by:

$$C_B(v) = \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}} \tag{2}$$

where σ_{st} is the total number of shortest paths from node s to node t, and $\sigma_{st}(v)$ is the number of those paths that pass through node v [5, 22].

Eigenvector centrality is another measure that assesses a node's influence based on the influence of its neighbors. It extends the concept of degree centrality by considering the importance of a node's neighbors. A node is considered important if it is connected

to other important nodes. The eigenvector centrality $C_E(v)$ of a node v is the v-th component of the eigenvector corresponding to the largest eigenvalue of the adjacency matrix A which is given by:

$$C_E(v) = \frac{1}{\lambda} \sum_{t=1}^{N} A_{vt} C_E(t) \tag{3}$$

$$AV = \lambda V$$

Here $C_E(t)$ is the eigenvector centrality of the t'th node, A is the adjacency matrix of the network, λ is the largest eigenvalue, and V is the eigenvector [5, 22]. Overall, these methodologies combine network analysis, ML, and natural language processing to provide a comprehensive understanding of bot behaviors and their impact on political discourse.

3.7 Ethical Considerations

To uphold ethical integrity, our research protocol received approval from the Brunel Ethics Committee. This approval underscored our commitment to ethical research practices and provided a structured framework for conducting our study responsibly. To maintain the integrity and ethical standards of our study, several measures were implemented to anonymize and protect user data effectively.

Firstly, to ensure anonymity, we replaced all usernames with unique identifiers, which allowed us to build and analyze the network structure without compromising the identities of the individuals involved. Additionally, we encoded descriptive features from the user's profile autonomously within our data processing pipeline. This approach ensured that features extracted from user profiles, such as the length of the username or the presence of non-ASCII characters, were processed in a manner that precluded the possibility of inferring the commenter's identity. Furthermore, all data scraping activities were confined to publicly available content on Instagram, ensuring compliance with ethical guidelines and respect for user privacy.

In our analysis, we made a deliberate decision to retain the names of prominent political figures and organizational accounts for the top centrality analysis. The justification for this approach is twofold. Firstly, public figures such as politicians operate in a public domain where their interactions and engagements are of legitimate public interest. Their inclusion helps in understanding the broader impact and influence patterns within the network, especially in the context of political discourse. Secondly, maintaining these identifiers allowed for more accurate network analysis and interpretation of the centrality measures, providing critical insights into how influential figures contribute to the overall dynamics of the network.

3.8 Research Limitations

The process of analyzing user data, especially when constructing network graphs, must adhere to strict ethical guidelines to protect user information. This includes anonymizing data where possible and ensuring that data collection and analysis methods do not infringe on individual privacy rights. There also exists a technical limitation of account bans due to scraping activities, which we attempted to mitigate through the use of proxies and varied user agents, as approved by [1] in their approach to fake account detection.

Table 1: Mean Macro F1 Score of Fake and Bot Account Detection using User Profile Features

Classifier	Dataset	Without Over-sampling	With Over-sampling
Random Forest	automated accounts	0.9270	0.9270
	fake accounts	0.9053	0.9642
XGBoost	automated accounts	**0.9291**	**0.9291**
	fake accounts	**0.9137**	**0.9689**

Furthermore, the adaptive nature of bot developers [14] presents another significant drawback. Bots are becoming increasingly sophisticated, employing evolving tactics that may elude existing detection mechanisms. This arms race necessitates a methodology that is not only robust but also flexible, capable of adapting to new strategies employed by inauthentic actors.

Lastly, the technological and resource limitations in processing and analyzing large-scale data sets represent a logistical barrier. Ensuring that our methodology is resource-efficient while still effective in detecting bots across vast networks requires ongoing optimization of computational strategies and potentially leveraging cloud computing resources to manage data scalability.

4 Experimental Evaluation

4.1 Fake and Bot Accounts Detection on Instagram: Model Evaluation

We implemented the RF classifier on user profile features, for two datasets of fake and automated accounts. Upon evaluation, the results, while reasonable, fell short of the performance metrics reported in the literature. Specifically, the mean Macro F1 Score for detecting fake accounts was 0.9053, indicating room for improvement compared to the 0.94 achieved by [1] on the same dataset. Then, we decided to incorporate gradient boosting techniques by implementing the XGB algorithm, which incrementally builds models to correct the mistakes of its predecessors. This approach yielded an improved mean Macro F1 Score of 0.9137 for fake account detection.

To continue improving, we addressed class imbalance in the fake accounts dataset using the Synthetic Minority Over-sampling Technique (SMOTE), which significantly enhanced the model's ability to accurately classify minority classes. Upon evaluation, the model achieved an impressive mean Macro F1 Score of 0.9689. Then, we evaluated the RF and XGB on both tasks: detecting fake accounts and identifying automated (bot) accounts. For each task, we conducted experiments both with and without class balancing to comprehensively assess the models' capabilities.

As can be seen in Table1, the boosted model demonstrated superior performance across both tasks, particularly when class balancing techniques were applied.

4.2 AI Content Detection: Model Evaluation

With the successful evaluation of our user profile models, we focus on training a RoBERTa model to detect AI-generated content,

Table 2: Mean Macro F1 Score of AI Content Detection

Model	Dataset	Mean Macro F1
RoBERTa	HC3-Full	0.9856
Finetuned with MPU	HC3-Sent (short texts)	0.8706

which enhanced our detection capabilities by leveraging natural language processing techniques. We trained our models on ChatGPT corpora, following the setting of HC3 [12] to test the performance of our methods. All texts were reduced into shorter texts for a sentence-level variant.

As observed in Table2 our RoBERTa model fine-tuned with MPU achieved a Mean Macro F1 score of 0.9856 on the HC3-Full dataset, slightly below the benchmark set by [24], who reported a score of 0.9860. However, our model significantly outperformed their results on the HC3-Sent (short texts) dataset, achieving a Mean Macro F1 score of 0.8706 compared to their 0.8531.

This remarkable performance can be attributed to our meticulous hyperparameter fine-tuning and the comprehensive grid search implementation we employed. While we did not surpass the highest benchmark on the HC3-Full dataset, our results still fall within the standard deviation of [24], indicating comparable effectiveness. Having trained our AI Content classifier, we integrated the RoBERTa model with our ensemble models for a final voting system, ensuring the highest accuracy in labelling our Instagram data.

4.3 Descriptive Analysis of Bot Presence (RO1)

The dataset was collected via our method of distributed web scraping software. The collection process executed between 2016 until 2024, resulted in a total of 1,958,056 comments from which 85,874 belonged to the campaign period of which 52,428 were kept after thorough cleaning. Also, during the campaign period which was between 2nd May 2024 until 11th July 2024, 225 posts, 26,588 child comments and a total of 28,847 users were identified. Child comments constitute discussion comments with back and forth mentioning, also known as replies to the parent comments.

The number of comments per day fluctuates between 58 and 719 for the first part of the considered period (see Figure 3), and takes a steep rise to 2,753 with the release of manifestos and the start of debates between June 10th and 19th, jumping to 4,224 on the polling day (July 4th). The number of users follows a similar trend, being almost never lower than 34 and rising as high as 3,467 on the election day. The peak on the 4th of July that can be observed in both panels is in concurrence of the election's polling day.

The average percentage of bots in the comments is around 5.65% for the whole Instagram lifetime of 2016 to 2024 for the three main candidates. Interestingly, the average presence of bots in the dataset increases to 9.71% during the campaign period. Bot activity peaks to 16.28% in the debate week and witnesses another steep rise to 14.57% on July 3rd, the day before the polling day.

The quantity of new bots that appear in the discussion can be seen in panel B of Figure 4. It shows the new bots that appear for the first time in our dataset. The most rapid and large increase of new kind of bots in the dataset appears to fall right before the release of manifestos and start of the heat of debates between the candidates.

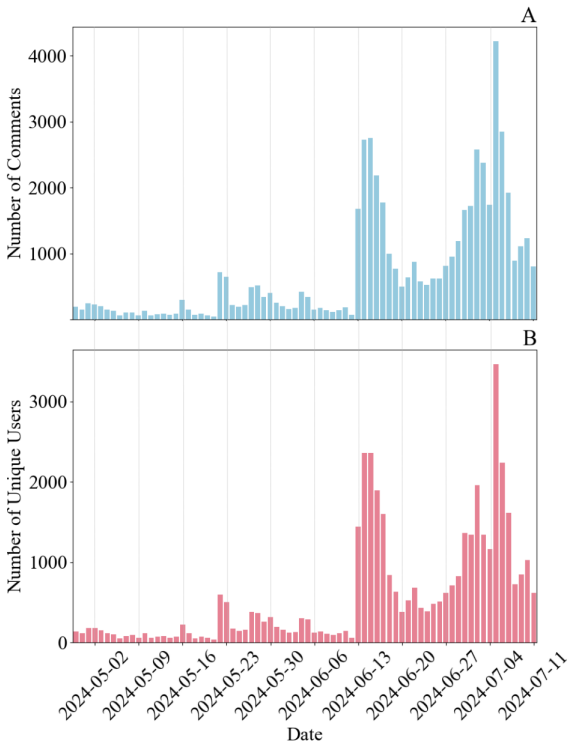

Figure 3: Number of comments (including child comments) (A) and users (B) per day

Additionally, an increase in the percentage of bots comes with the polling day, showing a new wave of bots, different from those from the debates period, entering the discussion just to focus on the final election date.

It is observed that the increase in bots' presence with the start of debates and release of manifestos, comes with a relative decrease in their activity, as per Figure 5. Bots were on average commenting more before May 22nd, 2024 when Prime Minister Rishi Sunak requested a dissolution of parliament from King Charles III and announced the date of polling day for the general election as July 4th. The separation between genuine users and bots increases from June 8th and they start to comment less in the period of the debates and take their steep increase several days before the polling day.

Bot silence during critical discourse stages may reflect strategic planning to target weak nodes and topics before elections. Unlike older bots, newer ones appear more active, likely due to insights into public opinion trends and key points for crowd manipulation.

4.4 Bots Behavior Analysis (RO2)

We now report our results of the behavioral analysis of these bots. Overall, 8.87% of all the comments in the discourse have been posted by bots. Nearly all bot comments (95.72%) are parent comments, indicating bots' tendency to engage less in direct interactions, preferring to post comments that do not address specific users, and in the small percentage of the cases they do, the vast majority of recorded interactions are with genuine users (97.29%), targeting real

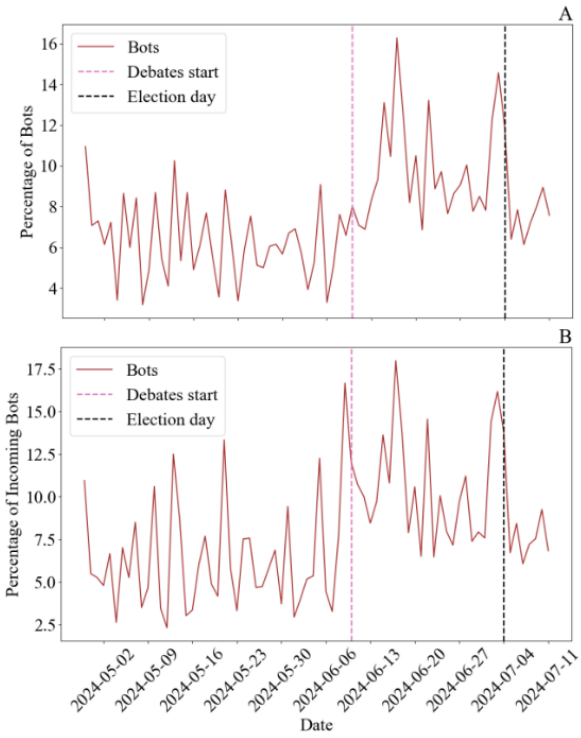

Figure 4: Time series plots of the daily percentage of bots among users (A) and incoming users (B)

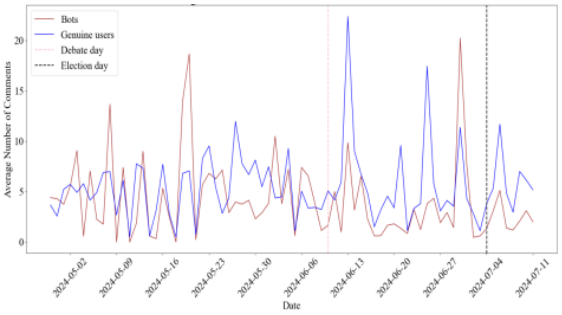

Figure 5: Average daily activity of bots and genuine users in the discourse over time

users to influence or disrupt the discourse. Interestingly, no interactions with verified users (0.00%) are recorded. Bots might not target verified users directly, possibly due to the higher visibility and security of these accounts.

The high percentage of bot comments as parent comments suggests that bots are used more for broadcasting messages rather than engaging in conversations. Additionally, the exclusive interaction of bots with genuine users indicates a targeted strategy to influence or disrupt human conversations. The lack of bot-to-bot interactions suggests that bots are not designed to converse with each other, or such behavior is not significant in this dataset.

We conducted several network analyses focused on user mention network, centrality measures, and interaction patterns. By examining

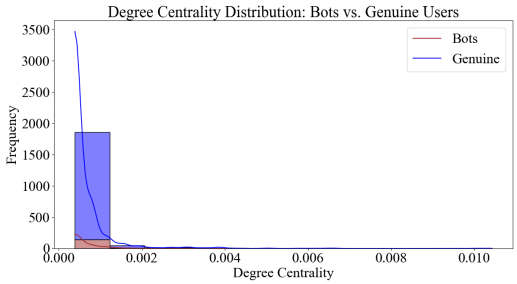

Figure 6: Degree Centrality Distribution: Bots vs. Genuine Users

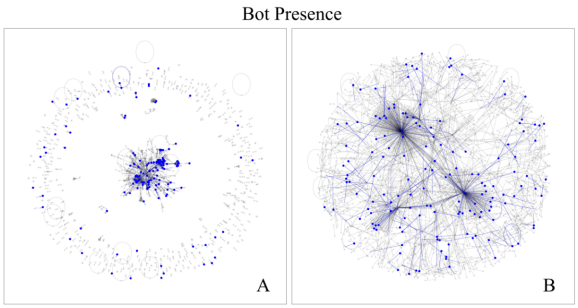

Figure 7: Network visualizations highlighting bot presence (blue nodes)

these aspects, we aimed to uncover the engagement strategies of bots and their influence within the network.

User Mention Network: We constructed a user mention network where nodes represented users and directed edges represented mentions (e.g., username) in comments. This network allowed us to visualize and analyze how users, particularly bots, engaged with others in the discourse. This process resulted in a directed graph, which we subsequently converted to an undirected graph to facilitate community detection using the Louvain algorithm.

Centrality Measures: We calculated key centrality measures to identify influential users within the network. Degree centrality, betweenness centrality, and eigenvector centrality were computed to gauge the prominence and influence of each user. Our analysis revealed that genuine users dominated the higher centrality scores, indicating their significant role in the discourse. The degree centrality distribution plot (Figure 6) illustrates this skew, with genuine users exhibiting higher connectivity compared to bots.

Figure 7 highlights the presence of bots within the user mention network. The blue nodes represent bots, and the gray nodes represent other users. In the network, bots are distributed both centrally and peripherally, indicating varied roles and engagement levels. Several bots are positioned at the core of the network, suggesting they are actively involved in the main discourse and have a high degree of interaction with other users. These central bots may be influential in steering conversations, spreading information, or amplifying certain narratives.

Conversely, the presence of bots at the periphery suggests some are less engaged or operate in more niche, isolated parts of the

network. These peripheral bots might be involved in targeted interactions or specific discussions, possibly aiming to influence smaller sub-communities or specific user groups. Comparing the positions of bots with non-bot users reveals that bots are well-integrated into the network's structure. They do not form isolated clusters but rather intersperse with genuine users. This integration allows bots to blend in and potentially have a more significant impact on the discourse by engaging with genuine users directly.

5 Conclusion

Our hybrid approach using network analysis combined with ML proved effective in identifying and analyzing bot activity. Experiments in the context of the UK's 2024 general election provided several key insights into the role of bots in political discourse on Instagram. Although bots are active, their influence is limited compared to genuine users, especially influential political figures, which contrasts with studies on other platforms, such as Twitter, where bots have been shown to have a more substantial impact [10, 23]. Their role might be more about creating noise or attempting to sway undecided voters subtly rather than overtly shifting public opinion. The findings imply that human influence, particularly from well-connected and prominent figures, plays a more critical role in shaping online political discourse.

One significant limitation of our approach is the reliance on publicly available data, which may not capture the full extent of bot activity. Private messages and interactions, which are not accessible through our data collection methods, could provide additional insights into bot strategies. Another limitation is the focus on the 2024 UK general election. While this provides a specific and relevant context, the findings may not be generalizable to other elections or political contexts. Lastly, while our use of RF and XGB provided robust results, exploring newer models and techniques could further enhance detection capabilities.

References

[1] Fatih Cagatay Akyon and Murat Emre Kalfaoglu. 2019. Instagram fake and automated account detection. In *2019 Innovations in Intelligent Systems and Applications Conference (ASYU)*. IEEE, IEEE, 1–7. doi:10.1109/ASYU48272.2019.8946364

[2] Seyed Ali Alhosseini, Rifat Bin Tareaf, Payam Najafi, and Christoph Meinel. 2019. Detect me if you can: Spam bot detection using inductive representation learning. In *Companion Proceedings of the 2019 World Wide Web Conference*. 148–153.

[3] Mohammed Aljabri, Rania Zagrouba, S. M. Shaahid, Fahad Alnasser, Abdullah Saleh, and Dhafer M. Alomari. 2023. Machine learning-based social media bot detection: a comprehensive literature review. *Social Network Analysis and Mining* 13, 1 (2023), 20. doi:10.1007/s13278-023-00975-3

[4] D Arnaudo. 2017. Computational propaganda in Brazil: social bots during elections. https://ora.ox.ac.uk/objects/uuid:e88de32c-baaa-4835-bb76-e00473457f46/files/mfb987109cf3f289e7126c993a4bdb7de. Accessed: 2025-06-11.

[5] Albert-László Barabási and Márton Pósfai. 2016. *Network science*. Cambridge University Press.

[6] H. C. H. Chang, E. Chen, M. Zhang, G. Muric, and E. Ferrara. 2021. Social bots and social media manipulation in 2020: The year in review. In *Handbook of Computational Social Science, Volume 1*, Ferruh Yilmaz and Jason Reifler (Eds.). Routledge, 304–323.

[7] Sarwar Chowdhury, Mohammad Khanzadeh, Ramesh Akula, Fei Zhang, Shuang Zhang, H Medal, and L Bian. 2017. Botnet detection using graph-based feature clustering. *Journal of Big Data* 4 (2017), 1–23.

[8] Stefano Cresci, Roberto Di Pietro, Marinella Petrocchi, Angelo Spognardi, and Maurizio Tesconi. 2017. Social fingerprinting: Detection of spambot groups through DNA-inspired behavioral modeling. *IEEE Transactions on Dependable and Secure Computing* 15, 4 (2017), 561–576. doi:10.1109/TDSC.2017.2681672

[9] Sheng Feng, Hao Wan, Ning Wang, and Meng Luo. 2021. BotRGCN: Twitter bot detection with relational graph convolutional networks. In *Proc. of the 2021 IEEE/ACM Int. Conf. on Advances in Social Networks Analysis and Mining*. ACM, 236–239.

[10] Emilio Ferrara, Onur Varol, Clayton Davis, Filippo Menczer, and Alessandro Flammini. 2016. The rise of social bots. *Commun. ACM* 59, 7 (2016), 96–104.

[11] Marta C. González and Albert-László Barabási. 2007. From data to models. *Nature Physics* 3, 4 (2007), 224–225. doi:10.1038/nphys566

[12] Biyang Guo, Xin Zhang, Ziyuan Wang, Minqi Jiang, Jinran Nie, Yuxuan Ding, Jianwei Yue, and Yupeng Wu. 2023. How close is chatgpt to human experts? comparison corpus, evaluation, and detection. *arXiv preprint arXiv:2301.07597* (2023).

[13] Md Alamgir Hossain and Md Saiful Islam. 2023. A novel hybrid feature selection and ensemble-based machine learning approach for botnet detection. *Scientific Reports* 13, 1 (2023), 21207.

[14] Sofia Hurtado, Poushali Ray, and Radu Marculescu. 2019. Bot detection in reddit political discussion. In *Proceedings of the fourth international workshop on social sensing*. 30–35.

[15] Kevin Jahanbakhsh and Yong-Yeol Moon. 2014. The predictive power of social media: On the predictability of US presidential elections using Twitter. arXiv:1407.0622 [cs.SI] arXiv preprint.

[16] Kokil Jaidka, Saifuddin Ahmed, Marko Skoric, and Martin Hilbert. 2019. Predicting elections from social media: A three-country, three-method comparative study. *Asian Journal of Communication* 29, 3 (2019), 252–273. doi:10.1080/01292986.2019.1583193

[17] Jan H. Kietzmann, Kristopher Hermkens, Ian P. McCarthy, and Bruno S. Silvestre. 2011. Social media? Get serious! Understanding the functional building blocks of social media. *Business Horizons* 54, 3 (2011), 241–251. doi:10.1016/j.bushor.2011.01.005

[18] Zachary C Lipton, Charles Elkan, and Balakrishnan Naryanaswamy. 2014. Optimal thresholding of classifiers to maximize F1 measure. In *Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2014, Nancy, France, September 15-19, 2014. Proceedings, Part II 14*. Springer, 225–239.

[19] Feng Liu, Zhiguang Li, Chao Yang, Dong Gong, Hu Lu, and Feng Liu. 2024. SEGCN: a subgraph encoding based graph convolutional network model for social bot detection. *Scientific Reports* 14, 1 (2024), 4122. doi:10.1038/s41598-024-16427-7

[20] Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692* (2019).

[21] Saeed Mohammadi, Pejman Moradi, Andrey Trufanov, and Gholam Reza Jafari. 2023. A structural approach to detecting opinion leaders in Twitter by random matrix theory. *Scientific Reports* 13, 1 (2023), 21788. doi:10.1038/s41598-023-48277-z

[22] M. E. J. Newman. 2003. The structure and function of complex networks. *SIAM Rev.* 45, 2 (2003), 167–256. doi:10.1137/S003614450342480

[23] Chengcheng Shao, Giovanni Luca Ciampaglia, Onur Varol, Kai-Cheng Yang, Alessandro Flammini, and Filippo Menczer. 2018. The spread of low-credibility content by social bots. *Nature Communications* 9, 1 (2018), 4787. doi:10.1038/s41467-018-06930-7

[24] Y Tian, H Chen, X Wang, Z Bai, Q Zhang, R Li, C Xu, and Y Wang. [n. d.]. Multiscale positive-unlabeled detection of ai-generated texts, CoRR abs/2305.18149 (2023). *URL: https://doi. org/10.48550/arXiv* 2305 ([n. d.]).

[25] Ümmügülsüm Tunç, Ebru Atalar, Mehmet Salih Gargı, and Zeynep Elif Aydın. 2022. Classification of fake, bot, and real accounts on Instagram using machine learning. *Politeknik Dergisi* 27, 2 (2022), 479–488. doi:10.2339/politeknik.850971

[26] Onur Varol, Emilio Ferrara, Clayton Davis, Filippo Menczer, and Alessandro Flammini. 2017. Online human-bot interactions: Detection, estimation, and characterization. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 11. AAAI, 280–289.

[27] Sunil Wattal, David Schuff, Munir Mandviwalla, and C. Britt Williams. 2010. Web 2.0 and politics: The 2008 US presidential election and an e-politics research agenda. *MIS Quarterly* 34, 4 (2010), 669–688.

[28] Samuel C. Woolley and Philip N. Howard. 2016. Automation, algorithms, and politics | Political communication, computational propaganda, and autonomous agents—Introduction. *International Journal of Communication* 10 (2016), 4882–4890.

[29] Sarita Yardi, Daniel Romero, and Grant Schoenebeck. 2010. Detecting spam in a Twitter network. *First Monday* 15, 1 (2010). https://firstmonday.org/ojs/index.php/fm/article/view/2793 Accessed: 2025-06-11.