An Examination of the Efficacy of Sentiment Analysis and Social Network Analysis in Unveiling Disinformation Campaigns during the 2022 Kenyan General Elections

Rose Kwamboka
Strathmore University
rose.kwamboka@strathmore

ABSTRACT
This paper explored the use of social network analysis (SNA) as a tool for identifying disinformation/hate speech actors on Twitter. With the increasing prevalence of disinformation in online communication, there is a growing need to identify the sources and spread of such false information including speed and virality of disinformation campaigns, and ultimately take necessary action on the actors. SNA is a valuable tool in this context as it allows for the analysis of social networks to identify key actors, their relationships and virality of their campaigns.

The paper begins by providing an overview of SNA and its applications in different fields, including the study of online communication. It then examines the limitations of this study. Next, the paper presents a case study that illustrates the use of SNA to identify and map disinformation actors in a real-world context.

The case study of Kenya's 2022 general election, demonstrates how the analysis of network structure, connectivity, and centrality can help identify important actors who are likely to play a critical role in the spread of disinformation. It also addresses how the disinformation campaigns can be countered through the use of data analysis techniques. The paper concludes by exploring the role of sentiment analysis in determining the polarity of content on social media.

1. INTRODUCTION
Social network analysis (SNA) is a promising approach that researchers can use to identify disinformation actors on social media. With billions of users using various social media platforms to share information and communicate with others, social media has recently emerged as an important information source. However the spread of disinformation and misinformation on social media has become a huge problem for academics, policymakers, and the general public. A promising strategy for locating and mapping disinformation actors on social media is SNA. In this study, we explore the use of SNA to identify and map disinformation actors during the 2022 election period in Kenya.

We also develop a regression model with the goal of predicting sentiments based on the tweets.

1.1 Problem Statemnet
The proliferation of disinformation on social media platforms has become a critical issue in recent years.

Disinformation actors use these platforms to spread false information and influence public opinion, potentially causing harm to individuals, organisations, and society as a whole. While efforts have been made to identify and combat disinformation, traditional methods such as content analysis - which can identify the content of disinformation, but may not be able to identify the the actors behind the disinformation campaigns - have proven to be inadequate, given the scale and complexity of social media platforms.

Thus, there is a need for more sophisticated tools and techniques, such as social network analysis, as used in this paper, to identify disinformation actors and their networks, in order to effectively address this problem.

1.2 Objectives
The objectives of this study are to:

- Identify and map virality of disinformation actors and their networks, including their behaviour, content, and interactions.

- Analyse how accurately a regression model can predict the sentiment of tweets

1.3 Justification
This paper will be of interest to researchers, policymakers, and practitioners who are concerned with disinformation and its impact on society.

1.4 Assumptions and scope
This study collected sample tweets from January 2022 until December 2022. The search terms were based on a hatelex - a political hate speech lexicon - developed by the Mapema Coalition.

The analysis was limited to text-based content - tweets - and did not investigate other types of media such as images and videos, that could be used in disseminating disinformation.

2. LITERATURE REVIEW

[7] explored the use of social media to detect and analyse the dynamics of the dissemination of false information on the internet. He suggested that social media data analysis can be a useful tool in identifying disinformation actors and patterns.

[8] added to [7]'s work that explored the use of social network analysis to identify the spread of false information and to distinguish it from legitimate information sources.

[5] investigated the dynamics of disinformation in social media including identifying disinformation actors and patterns, as well as identifying the spread of false information and its impact on the public.

Meanwhile, [14] identified disinformation campaigns on Twitter and analysed the spread of false information on the same platform and its impact on public opinion.

Particularly, [6] as well as [9] investigated the use of social network analysis to identify disinformation campaigns in political communication. They suggested that social network analysis can be used to identify disinformation actors and their relationships, as well as to analyse the spread of false information and its impact on public opinion.

[11] identified the propagation path of fake news content on social media by constructing a social network graph. They used a three level based, top-down tree structured networks.

Meanwhile, [2]investigated how false news spreads on Twitter. They further used social network analysis to study the role of bots in the 2016 US Presidential election. They found that bots were responsible for a significant portion of election-related tweets and that their activity was largely concentrated on pro-Trump and anti-Clinton messages.

Still on bots,[4] conducted a critical review of the literature on bots, echo chambers, and disinformation. They found that social network analysis is a useful tool for identifying the sources of disinformation but noted that it has limitations in detecting coordinated campaigns.

[15] examined the different types of bots, their impact on social media analysis, and detection methods, using Botometer as a case study. The study found that bot detection has become an essential part of the social media experience for many users, but misinterpretation of the results is a problem. It also found that botnets can have a significant impact on social network analysis, leading to overestimations of the influence of certain actors.

However,[13] conducted a DARPA Twitter bot challenge to evaluate the ability of algorithms to detect bots on social media platforms. They found that social network analysis is a useful tool for identifying bots, but noted that it is important to consider the social context in which the bots are operating.

Further,[12] studied the participatory nature of disinformation campaigns, highlighting the importance of identifying the human actors behind these campaigns. They found that social network analysis is a useful tool for identifying the structure of disinformation campaigns but noted that it may not be sufficient for identifying the actors behind them.

In a study on model development [10] discuss the importance of selecting the appropriate functional form for the regression model, as well as the use of regularization techniques to avoid overfitting. Another important consideration is the selection of the most relevant independent variables. The authors [16] propose a variable selection method based on the least absolute shrinkage and selection operator (LASSO) to identify the most important predictors in a regression model. By considering these factors, researchers can develop more accurate and robust regression models that can inform decision-making in various fields.

3. METHODOLOGY

To collect tweets for this study, I used Meltwater API. I created a search query using a[1] - a 23-term lexicon of hate speech in Kenya, which was flagged in 2022 by The National Cohesion and Integration Commission (NCIC) in collaboration with the Mapema Coalition to track down online hate speech and digital incitement aimed at the 2022 elections. By using date range, location, and language filters, I was able to refine my search results. I filtered out irrelevant tweets using Meltwater's advanced search operators. The advanced search operators use boolean logic to narrow down the specifics of one's search criteria. Utilising Meltwater has the added benefit of offering a large number of additional search operators in addition to the most popular ones like AND, OR, and NOT.

Quotation marks are used to enclose keywords or phrases in this boolean search. For example "Chunga Kura". In a more sophisticated search, multiple phrases are grouped together using parentheses, as is the case in ("Madoadoa" OR "Madoa doa") AND "Azimio". This will retrieve documents containing either "Madoadoa Azimio" OR "Madoa doa Azimio".

Other operators available for use include: Source Operators; Proximity Operators such as INGRESS which matches documents that contain the key phrase entered in the first paragraph; Social Media Operators like @ (Twitter handle) which matches documents where a Twitter handle was mentioned; Complex Operators like >= which matches documents with a reach greater than or equal to the given value; Entity Operators; and Visual Analytics Operators, for instance, imageText enables you to match text (which must be specified in all lowercase) inside an image by using optical character recognition on images. imageText: stop would correspond to documents where the word stop is contained within the image, such as on a stop sign.

This yielded a total of about 101,000 publicly available and relevant tweets for the period between January 1, 2022 and December 31, 2022. Of these tweets, a random sample of 20,000 tweets were made available. The data had the following 37 variables: Date; Headline; URL; Opening Text; Hit Sentence; Source; Influencer; Country; Subregion; Language; Reach; Desktop Reach; Mobile Reach; Twitter Social

```
(Uncircumcised OR Fumigation OR Eliminate OR Kill OR Kaffir OR Madoadoa OR "Madoa doa" OR "Chunga Kura"
OR "Kama noma noma" OR Kwekwe OR Mende OR Hatupangwingwi OR "Operation Linda kura" OR "Watu wa kurusha
mawe" OR "Watajua hawajui" OR "Wabara waende kwao" OR Wakuja OR "Uthamaki ni witu" OR "Kimurkeldet" OR
"Otutu Labatonik") AND ((Raila OR Ruto OR Uhuru OR Kenyatta OR Musalia OR Wetangula OR (ODM) OR (Jubilee
AND Party) OR (UDA AND Party) OR Azimio OR "Azimio La Umoja" OR Kalonzo OR "Kalonzo Musyoka" OR (Wiper
AND Party) OR "Wiper Democratic Movement"))
```

Figure 1: Hatelex used in the Boolean search

Echo; Facebook Social Echo; Reddit Social Echo; National Viewership; Engagement; AVE; Sentiment; Key Phrases; Input Name; Keywords; Twitter Authority; Tweet Id; Twitter Id; Twitter Client; Twitter Screen Name User; Profile Url; Twitter Bio; Twitter Followers; Twitter Following; Alternate Date Format; Time; State City; Document Tags.

The dates were chosen to capture data for the entire election period which includes; pre-election which includes the period before the official campaign date, which was declared on May 29, 2022; the campaign period leading to the August 9, 2022 general elections; the election period and the post-election period.

A wordcloud visualised the most frequent words from the tweets.

We then developed a logistic model by first pre-processing the data. Here, we renamed some columns for clarity and converted the user names to lower case after removing all punctuation. In the tweet column, we replaced any Twitter usernames (mentions), any URLs, any non-ASCII including punctuation, numbers, and special characters - , with a space character. We further replaced any words with length of 1 to 4 characters in order to remove stop words or common words, such as "the", "and", "or", and "for", which may not be relevant or may interfere with certain text analysis techniques. We then replaced any sequence of one or more spaces in the string variable "Tweet_Texts_Cleaned" with a single space. We finally lemmatised the corpus. We split the data into training and testing sets.

The most relevant features to include in the model were then picked on which we trained the logistic model using the training data. Hyperparemmeter tuning was done on the validation dataset and the performace of the model evaluated using the testing data. Here, we calculate metrics accuracy, precision, recall, and F1 score to assess the model's ability to make accurate predictions and thus make informed decisions.

Finally, we identified and mapped a coordinated network of accounts on Twitter amplifying false information during the 2022 election period. Gephi was used for visualising and analysing the social networks. In Gephi, you can Visualise networks in a variety of ways, including force-directed layouts, circular layouts, and hierarchical layouts and customize network visualisations by adjusting node and edge size, color, and shape. We can further analyse network properties such as node and edge betweenness centrality, clustering coefficients, and degree distribution. To successfully visualise a network graph using Twitter data in Gephi, you would typically need to collect a dataset that includes usernames, mentions, retweets (to identify influential users in your network) and timestamps (to create a time-series

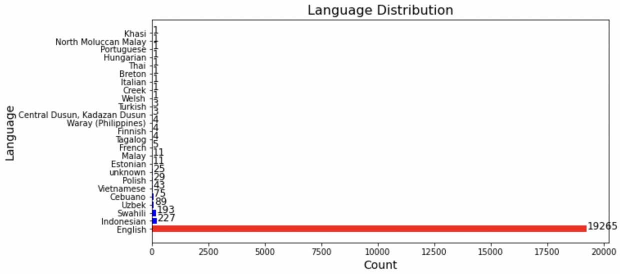

Figure 2: Language Analysis: Most frequently used language

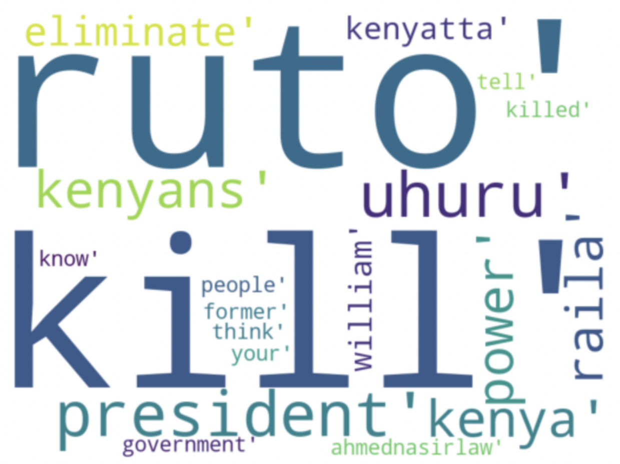

Figure 3: A word cloud showing the most frequently used words

analysis of your network).

4. RESULTS AND ANALYSIS

English was the most frequently used language with Indonesia coming at a distant second.

The word cloud summarises the most frequent words in the Twitter hate speech corpus incuding 'kill', 'ruto', 'president', and 'uhuru', were the top 5 most frequently used words.

Prevalence of hate speech in Kenya between January 1, 2022 and December 31, 2022. Naturally, we expect the hate speech to be high during the election period. However, based on Code for Africa investigations which sought to pull down hate speech posts from Twitter during the election period(dates), most posts were deleted from the social media platform thus the reduced number of hate speech posts.

The logistic regression model, which was trained to predict one of three possible classes: Negative, Neutral, or Positive, performed well on unseen data, with high precision, recall, and F1-score values for each class, and an overall accuracy of 0.91. The metrics used to evaluate the performance of the model are precision, recall, and F1-score.

Precision measures the proportion of instances that were correctly classified as belonging to a certain class, out of all instances that were predicted to belong to that class. Re-

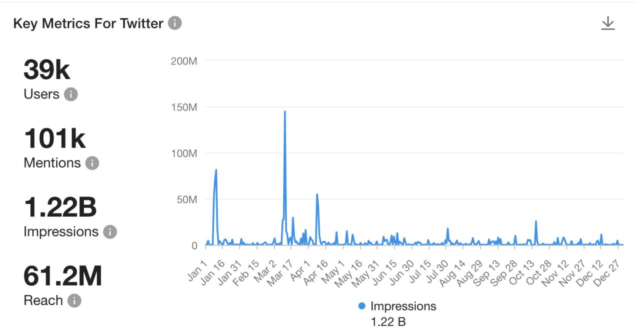

Figure 4: Prevalence of hate speech in Kenya between January 1, 2022 and December 31, 2022

call measures the proportion of instances that were correctly classified as belonging to a certain class, out of all instances that actually belong to that class, while, F1-score is the harmonic mean of precision and recall.

Looking at the results, we can see that the model performs best on the Negative class, with a precision of 0.97 and a recall of 0.90, indicating that the model correctly identifies a high proportion of negative instances, and the majority of the instances it identifies as negative are indeed negative. The Neutral and Positive classes also have high precision and recall values, indicating that the model performs well on these classes, although not as well as on the Negative class.

It is important to note that we encountered the common challenge of imbalanced data, where the number of observations in negative category was larger than that of the positive and neutral categories. We found that our logistic model was biased towards the Negative Sentiment category during training, leading to the highest accuracy for that category during evaluation. However, this biased evaluation could be misleading in assessing the overall model performance. To mitigate this, we used other evaluation metrics, such as precision, recall, and F1 score, to assess the model's performance across the different categories. By using these metrics, we were able to identify potential issues related to imbalanced data and gain a more comprehensive understanding of our model's performance. Overall, we found that considering the balance of the dataset is crucial for accurate evaluation and avoiding potential biases in machine learning models.

In mapping the coordinated network of accounts on Twitter, we found that the network used a copy and paste strategy. For instance, 1618 accounts re-posted the tweet from usename ahmednasirlaw, 'When I used to tell Kenyans/world that former President Uhuru's Government used to kill Kenyans as an official governmental policy you used to think I was an alarmist.' This indicates a level of coordination that could be financially motivated.

The top amplifiers include amounomalande, makaralph, bwasexo, clementboursin and afro_samuraii.

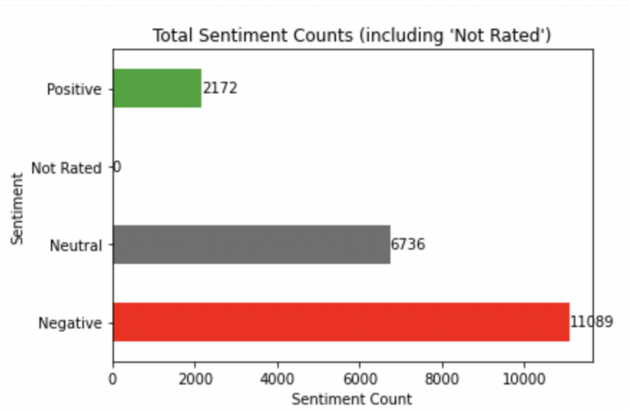

Figure 5: A horizontal bar chart showing the count of occurrences for each category in the "Sentiment" column including the outlier named 'Not Rated'. This was to show the imbalance in the dataset.

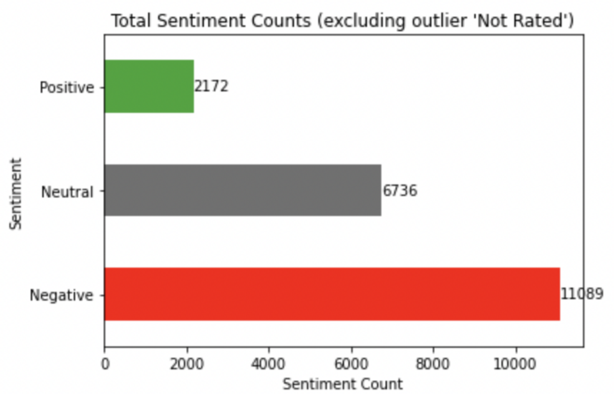

Figure 6: A horizontal bar chart showing the count of occurrences for each category in the "Sentiment" column excluding the outlier named 'Not Rated'.

	precision	recall	f1-score	support
Negative	0.90	0.96	0.93	2290
Neutral	0.91	0.85	0.88	1454
Positive	0.96	0.85	0.90	456
accuracy			0.91	4200
macro avg	0.93	0.89	0.90	4200
weighted avg	0.91	0.91	0.91	4200

Figure 7: Performance of the logistic regression model on the train dataset

	precision	recall	f1-score	support
Negative	0.97	0.90	0.93	3618
Neutral	0.83	0.92	0.87	1813
Positive	0.83	0.95	0.89	569
accuracy			0.91	6000
macro avg	0.88	0.92	0.90	6000
weighted avg	0.91	0.91	0.91	6000

Figure 8: Performance of the logistic regression model on the test dataset

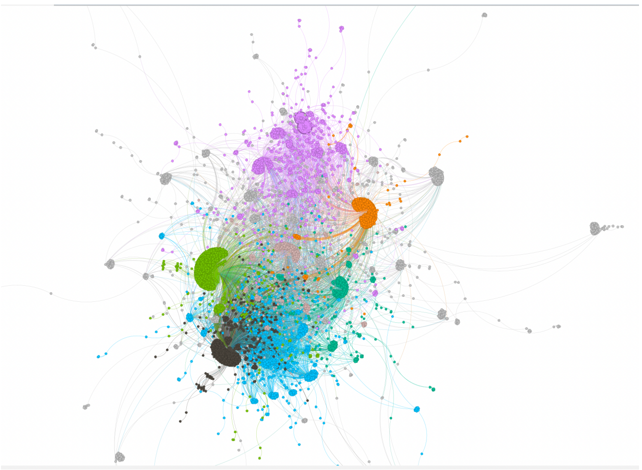

Figure 9: A social network graph showing the entire network as clusters of Twitter accounts using the hatelex.

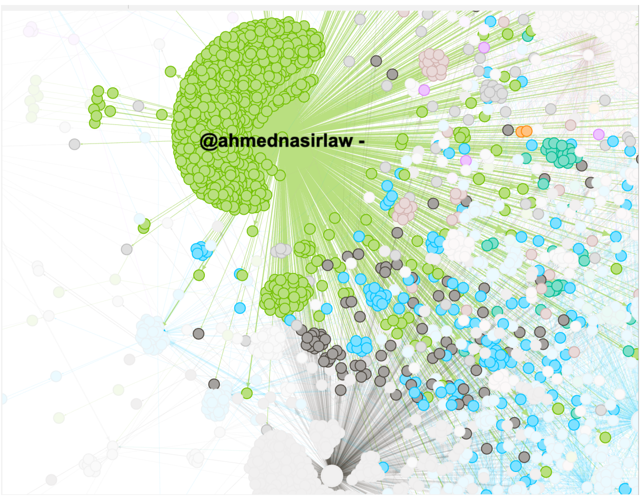

Figure 10: A social network map showing clusters of Twitter accounts aamplifyig hate speech content. Key 'seed nodes' such as ahmednasirlaw tweets was amplified by other accounts retweeting them.

The seeders included afropages, mihou, mtchadiens, peterbradyakewa and munkengl.

5. CONCLUSION

The study conducted a social network analysis to identify disinformation actors on 20,000 tweets scraped from Twitter during the 2022 election period. The social network analysis brought to light the copy-and-paste strategy as an approach to spreading specific narratives intended to impact public opinion during the 2022 Kenya General Elections. These strategies, storylines, and approaches attempted to sway public opinion and eventually the outcome of the election through propaganda, hate speech, coordinated inauthentic behavior (CIB), misinformation, and incitement.

Even if the Kenyan elections are over, the effects will endure for a while, so steps must be taken to stop election-related poisonous content and methods in future elections globally.

[3]

References

[1] 2022. Boolet email copy 07042022 - cohesion.or.ke. *National Cohesion and Integration Commission.*

[2] 2016. The rise of social bots | communications of the ACM. Communications of the ACM.

[3] .

[4] Bastos, M.T. and Mercea, D. 2017. The brexit botnet and user-generated hyperpartisan news. *SSRN Electronic Journal.* (2017).

[5] Corman, S.R. et al. 2017. Detecting and mitigating disinformation in social media: Trends, challenges, and opportunities. *Proceedings of the 50th Hawaii International Conference on System Sciences (2017).* (2017).

[6] Han, H. et al. 2017. Detection of online rumor based on propagation mechanism analysis and social network theory. *Research Gate.* (2017), 1213–1221.

[7] Lawton, G. 2011. In the news. *IEEE Intelligent Systems.* 26, (May 2011), 5–9.

[8] Monteiro, J.M. et al. 2023. Peer-to-peer video streaming. *Digital Arts and Entertainment.* (2023), 169–228.

[9] Piña Iturbe, A. et al. 2020. Identifying disinformation campaigns on twitter: Social network analysis of spanish political parties. Research Gate.

[10] Royston, P. and Sauerbrei, W. 2009. *Multivariable model-building: A pragmatic approach to regression analysis based on fractional polynomials for modelling continuous variables.* Wiley.

[11] Sivasankari, S. and Vadivu, G. 2021. Tracing the fake news propagation path using social network analysis. *Soft Computing.* 26, (Jul. 2021), 12883–12891.

[12] Starbird, K. et al. 2019. Disinformation, "fake news," and influence campaigns on twitter. *Data and Society.* (2019), 1–58.

[13] Subrahmanian, V.S. et al. 2016. The DARPA twitter bot challenge. *Computer.* 49, (Jun. 2016), 38–46.

[14] Xiong, Dr.H. et al. eds. 2013. TIST: Vol 5, no 1. ACM Transactions on Intelligent Systems; Technology; Association for Computing MachineryNew YorkNYUnited States.

[15] Yang, K. et al. 2019. Arming the public with artificial intelligence to counter social bots. *Human Behavior and Emerging Technologies*. 1, (Jan. 2019), 48–61.

[16] Zou, C. et al. 2010. Lasso-based multivariate linear profile monitoring - annals of operations research. *SpringerLink*. Springer US.