AMSI VACATIONRESEARCH SCHOLARSHIPS 2020–21

Get a Thirst for Research this Summer

Understanding How Influence Affects Information Flow In Online Social Networks

Bridget Smart

Supervised by Dr Lewis Mitchell

The University of Adelaide

Vacation Research Scholarships are funded jointly by the Department of Education, Skills and Employment and the Australian Mathematical Sciences Institute.

AMSI AUSTRALIAN MATHEMATICAL SCIENCES INSTITUTE

Australian Government Department of Education, Skills and Employment

Abstract

Online social networks are increasingly becoming an influencing factor in how the public are reacting to and forming opinion around topics from politics to public health. Understanding how we can better capture and understand behaviour online is key to developing strategies which counteract malicious activity, ensuring productive dialogue in our online communities, to support cybersecurity and transparency in online social networks. By developing a framework to identify sources of influence online we can begin to capture the features which contribute to the success of phenomena including disinformation campaigns and "echo chambers".

Using 1.8 million Tweets from the 6th of January 2021, all containing the keyword 'trump', this work develops a strategy to identify patterns of behaviour online, gauging public opinion and response to events both internal and external to the online network. In Washinton, on the 6th of January 2021 a riot occurred to prevent the counting of the Electoral College votes, with rioters eventually breaching the Capitol Building. By combining existing sentiment analysis and network analytic tools, we are able to identify several key events throughout the day, and better understand how they have influenced the online network.

1 Introduction

Social media is increasingly being recognised as a source of information, and recent global and political events have highlighted the role of influence and interference on social media (Ardevol-Abreu, 2015). These events have raised concerns about the quality of information available on social media networks, and the perception of opinion as many people turn to social media to connect and learn. Understanding information flow across social media networks is key to understanding and soothing these concerns, and developing mechanisms which can restore trust and control to users.

Online social networks provide a forum for individuals and businesses to produce and curate content, and the sheer volume of data available can provide valuable insights, particularly when these large data sets are understood and managed efficiently. The data is readily available from social media platforms such as Twitter proves a valuable recourse to understand and analyse information flow over online networks. Researchers have used this data to make inference from public online activity to the private habits of an individual, to predict and explore the occurrence of "echo chambers", and to understand and measure public opinion on specific topics (Galuba et al, 2011) (Hofman et al, 2011). Considering what influence social media users can have on each other provides an opportunity to understand phenomenon such as "echo chambers", fake news and disinformation campaigns. However, this data can also be used maliciously to target disinformation campaigns and advertising [15]. This influence is increasingly becoming a factor in how the public are reacting to and forming opinion around topics from politics to public health. Understanding how we can better capture and understand behaviour online is key to developing strategies which counteract malicious activity, ensuring productive dialogue in our online communities, to support cybersecurity and transparency in online social networks.

This work uses 1.8 million Tweets from the 6th of January all containing the keyword 'trump', to develops a strategy to identify patterns of behaviour online, gauging public opinion and response to events both internal and external to the online network. On the 6th of January a riot was lead by Trump, with rioters aiming to cause the Electoral College to reject the result of the election. The rioters moved to the United States Capitol Building, eventually breaching police lines and gaining entry to the building, an event which ultimately cost five lives.

We begin by cleaning our data and undertaking a general exploratory data analysis. This revealed spikes in the frequency of certain topics and users in conjunction with several key events which occurred throughout the day, including the start of Trump's Save America Rally, when the rioters reached police lines at the Capitol, and when the Rioters breached police lines, gaining entry into the Capitol. Using the lexicon based sentiment analysis tool VADER, we will consider how sentiment changes across the day, identifying general responses to several key events and individuals. A brief network analysis is considered, to show the potential for combining the two techniques in future work.

Statement of Authorship

The workload was divided as follows:

- Bridget Smart collected the data using the Twitter API, analysed the data using Python, produced the results and reported and interpreted the results and wrote this report.
- Dr Lewis Mitchell supervised the work, assisted with interpreting the results and proofread this report.

2 Data

Malicious activity online has previously been recorded alongside political events in the United States of America. This has previously manifested in phenomena including fake news, bots and misinformation, particularly around political elections. To build on this existing literature, the data used for this project was collected on the 6th of January 2021, the day of the Electoral College vote count. In the lead up to this event, tensions were building among conservative networks in response to claims of voter fraud from Donald Trump and other members of the Republican Party. On the 6th of January, a riot was lead by Trump with rioters aiming to cause the Electoral College to reject the result of the election. The rioters moved to the United States Capitol Building, eventually breaching police lines and gaining entry to the building, an event which ultimately cost five lives.

Twitter data was used to conduct this investigation, due to the popularity of the social media platform in the United States of America; approximately 68.7 million Twitter users are based in the United States. Using Twitter data will provide a balanced picture of the day's events, as it is a platform which individuals from across the political spectrum employ to share *Tweets*, which are short text segments up to 280 characters long, which can be shared in response to another user or topic, or as a standalone message.

Our data set comprised of 1,874,367 tweets which were captured between 4:22 am and 2:47 pm, Washington time on the 6th of January 2021, encompassing both the lead up to and the start of the riot on the United States Capitol. Tweets which contained the keyword 'trump' were collected. Using Twitter limited our data set to only include individuals who use Twitter, and shared a tweet which met the search criteria. This may exclude certain individuals, but as we are aiming to understand influence on online social networks, so this is an inherent feature of the problem.

For each tweet, a number of features were collected. These included the following fields which remained empty for a given Tweet if the data was not available (Table 1).

2.1 Collecting the data

To collect the data, the Twitter *API* was used. An Application Programming Interface (API) is a tool which handles interaction between software intermediaries. The Twitter API is a tool to read and write Twitter data. A short script was written using the Python programming language which searches Tweets as they are posted, saving those which meet predefined criteria. By collecting the Tweets live, issues surrounding duplicates in the database or limits on the number of results returned in a search. However, collecting data live meant the data collection was subject to a rate limit, of 3000 Tweets per minute. This lead to limited volume of data later in the data set, and not all Tweets which matched the search criteria were collected.

Field name	Field Type	Description
Text	String	Contents of the Tweet. This is usually a short message comprised of words, emojis or other characters.
Hashtags	Vector of strings	This vector contains each hashtag which the Tweet contains. Hashtags are a word of set of words preceded by the hash symbol, #, which link a Tweet to a particular topic.
Username	String	This is the display name which the user who posted the Tweet has selected.
Timestamp	String	Date and time in Coordinated Universal Time. This was converted to a Datetime format and into Washington time for analysis.
Location	String	This gives the latitude and longitude coordinate for Tweets which have been geotagged.
Replied to	String	If the Tweet is a reply or direct response to another user, the other user's Twitter handle will appear.
Language	String	Short code giving the language of the Tweet. Here, we only collect Tweets tagged as being in English.
Verified	Boolean	Logical variable describing if the user who posted is verified.
Followers	Integer	Number of followers of the user who posted.
Friends	Integer	Number of friends of the user who posted.
Statuses	Integer	Total number of statuses posted by the original user.
Retweeted	Integer	Total of number of retweets this Tweet has received. This is equal to 0 for all collected Tweets, as they have been collected in real-time so have nott be retweeted.

Table 1: Data Features

The data was saved to a JSON file before analysis occurred.

Another limitation from the dataset was the collection of username rather than screen name for the user who originally shared the Tweet. The username is a user chosen tag which is subject to change and is chosen by the user themselves. There are no restrictions to prevent duplicate usernames, so it is not possible to exactly determine which accounts the Tweets originated from. For further work, care should be taken to ensure that the screen name is collected. The screen name is the users Twitter *handle*, a unique identifier for each account. This is subject to change, but knowing when the Tweets were collected will enable the originating account to be identified.

The Tweepy library, which was used to handle the search queries has an inbuilt function which allows a screen name to be converted to a user ID or username. As the screen name was collected for any user which the original Tweet was in reply to, these names were converted to user names to enable our network analysis to occur. This ensured that Tweets from one account were identified by the same name, although does lead to the assumptions that a screen name does not change across the dataset for a given user, and that these user names are unique.

The data also included Tweets from accounts which have since been deleted, including the account realDonaldTrump which was suspended by Twitter in response to the events of the day. This lead to issues automatically resolving screen names into usernames, although this only occurred for two cases which were manually handled.

3 Identifying influence

When measuring influence online, there are a number of metrics which can be considered. This work will consider two methods of tracking influence, with a focus on how information and intention spreads across a network. The first, using easy to track metrics including the count of Tweets containing a hashtag, or in response to a given user across time is used to explore the data and better understand how the network reacts to events both external to and internal to the network. Then, a sentiment analysis tool is used to track the spread of sentiment across a network. The aim of this is to identify abnormal trends or users who drive a change in sentiment across a network.

This work begins by tracking the spread of a hashtag or phrase across the online social network, to track the spread of a topic across the network. Considering the occurrence of the top five hashtags across the day. Hashtags are included in Tweets to link a Tweet to a particular movement or topic, so hashtags are a natural way to quickly understand how topics of conversation are changing across time.

Figure 1 shows the occurrence of the 5 most popular hash tags across our data set with some major events from the day super imposed.

The first red line shows the time when Trump's Save America Rally begun near the White House, where Donald Trump, then his lawyer Rudy Giuliani spoke, encouraging the rioters. After he spoke, the rioters began to make their way toward the Capitol. Shown in green, the frequency of the hashtag 'CapitolRiots', is non zero early in the day, followed by a large spike just after the rally begun. This indicates that the Capitol Riots

Figure 1: Frequency of Tweets containing one of the top 5 most frequent hashtags across the day, all times are local to Washington DC. The Tweets are grouped into 5 minute periods, with the first and last groups removed.

became an increasingly popular topic amongst Tweets containing the keyword 'trump'.

The next major marked event is when rioters and police began to clash at the capitol and finally when the rioters breached the police lines. Preceding this event, there is a spike in the occurrence of the hashtag '25thAmendmentNow'. This refers to an Amendment in The Constitution of the United States which was viewed as a tool to remove Trump from power. The hashtag 'BREAKING' is often used by media organisations to indicate that a Tweet contains breaking news, and we can see spikes in the frequency of this hashtag corresponding to all marked events, with a large spike occurring in conjunction with the rioters breaching police lines and entering the Capitol.

To understand how different users are interacting with popular users, we also consider *replies*, Tweets which are in reply to another user's Tweet. This links the two Tweets together on the platform, and is often used to reinforce or dispute other's positions. As the data set is large, we only consider the 14 most frequently replied to users (Figure 2). When a spike occurs, this indicates that replied to a user have become significantly more common. In Figure 2, two notable spikes are those to Representative Mo Brooks (RepMoBrooks), who released a public statement condemning the riots, and to Rudy Giuliani, who left a voicemail on the wrong phone, in which he urges Republican Senators to delay the counting of senate votes.

Combining these two analytic tools, allow responses to events to be identified and isolated, without any knowledge of the events of the day. All of the events which have been discussed were external to the Twitter

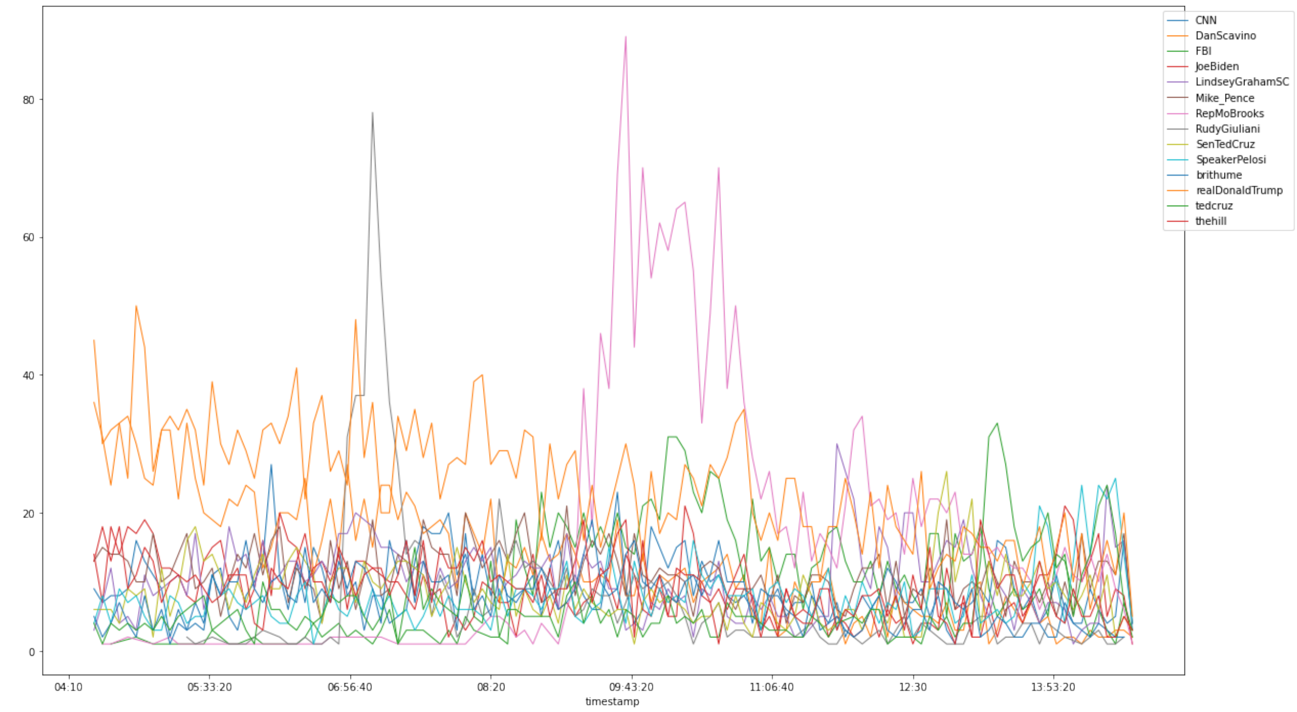

Figure 2: Frequency of Tweets which are in reply to one of the fourteen most replied to users. All times are local to Washington DC. The Tweets are grouped into 5 minute periods, with the first and last groups removed.

network, shared on the network by media organisations or other Twitter user's who had found out the news from an alternative source.

3.1 Sentiment Analysis

To gain a deeper understanding into events with the network, for example users sharing an opinion of a certain sentiment, a sentiment analysis tool will be used to determine a sentiment for each Tweet. This will allow us to measure public perception more analytically, looking for spikes in positive or negative sentiment which will correspond to key events throughout the day or other unusual behaviour in the network.

Denote each sequence of words as X, a vector where each entry represents a single word. Let the i^{th} word in the vector be given by x_i, for $i = 1, 2, ..., n$ for a sequence of words of length n.

VADER (Valence Aware Dictionary for Sentiment Reasoning), is a rule and lexicon based tool, so is a sentiment analysis tool which uses a set of rules and predefined cases to assign a sentiment score to each word. VADER is well suited to Twitter data as it has word lists which include emoji, emoticons and slang, which frequently occur. VADER also includes a number of phrases, including 'under the weather' which is counted as a negative phrase. It calculates a sentiment score for each word, taking into account any words which came before and if this word appears in any of the predefined cases. Capitalisation is also able to affect the sentiment of a word, increasing or decreasing how positive, neutral or negative the sentiment associated with that word is.

To perform sentiment analysis, VADER calculates a *valence score* for each word, which gives each word a value between -4 and 4, where -4 corresponds to a very negative word, 0 to a neutral word and 4 to a positive word.

For each phrase, passed in as a vector of words, VADER calculates four different scores. These are the *compound score, positive score, neutral score or negative score.* The positive, neutral and negative scores will sum to 1 and give the proportion of words which fall into each category, weighted by the valence score for each word. The compound score is computed by summing the valence scores of each word in the data, which are then normalised to return a value between -1 and 1.

Using a rule and lexicon based sentiment analysis tool is also beneficial as it does not require any training, which requires labelled data and was infeasible given the constraints of this project.

These four sentiment scores were calculated for all tweets in the database, with the average compound score across time shown in Figure 3.

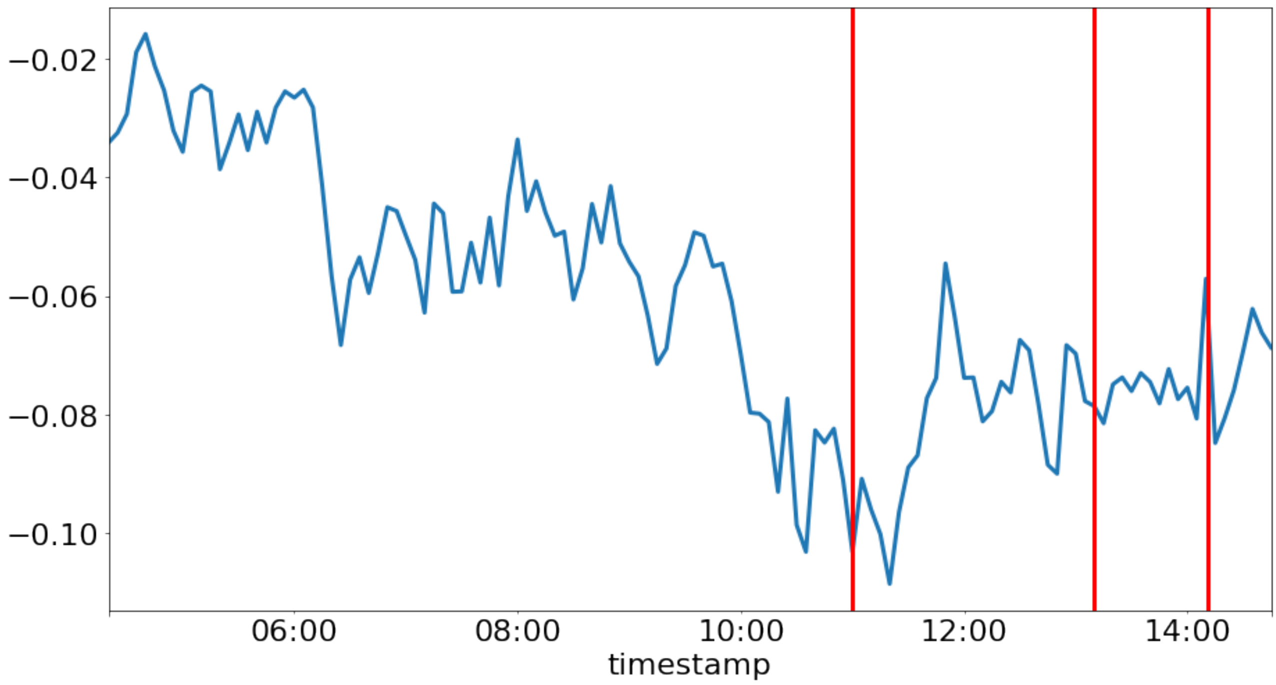

Figure 3: Average Compound Score. All times are local to Washington DC and the times of the three events previously outlined are shown in red. The Tweets are grouped into 5 minute periods, with the first and last groups removed.

4 Network Analysis

There is a large volume of existing work which uses network analysis techniques to understand social networks. These tools were briefly explored in this work, to gain a full understanding of existing tools, and to explore the potential for combining several of these techniques into a comprehensive tool.

A network is a directed graph is comprised of a set of *nodes* and *edges*. A node is an object and a connection between two nodes is an edge.

The *degree* of a node refers to the number of edges which connect to the node. For the networks we generate, the degree gives the number of Tweets shared or in response to a given user, represented by a node.

There are a number of centrality measures which can be used to measure a network, and understand the connections which a network has. For the purposes of this investigation, only the degree of nodes are considered, as issues relating to identifying unique users mean other techniques may give misleading insights.

4.1 Network Results

The data was filtered to only include Tweets which were in response to the top 100 most frequently replied to users, reducing our data set to 45,863 Tweets. This made the constructed networks more manageable.

To generate networks, the Python library Networkx was used.

Isolating the users which have the highest degree across 10,000 Tweet sections of the filtered dataset, reveal similar patterns to earlier analysis Figure 2.

Tweets in filtered data set	0-10,000	20,000-30,000	35,863 - 45,863
Highest degree	Donald Trump	Mo Brooks	Donald Trump
	Dan Scavino	Donald Trump	Senator Ted Cruz
	CNN	FBI	Lindsey Graham
	Lindsey Graham	Mike Pence	Joe Biden
	Joe Biden	Dan Scavino	Rep Mo Brooks

Table 2: Top 5 Users with the Highest Degree

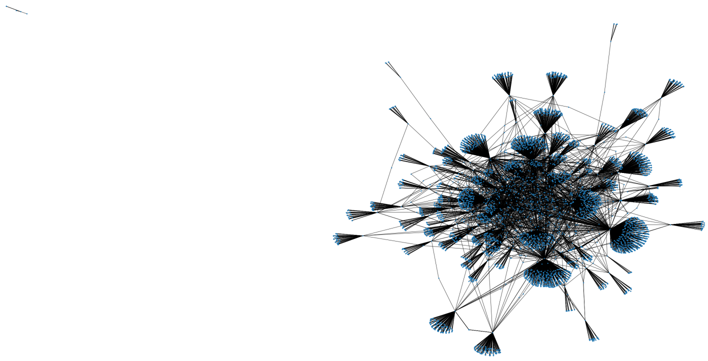

Figure 4: First 10,000 Tweets in Filtered Data set

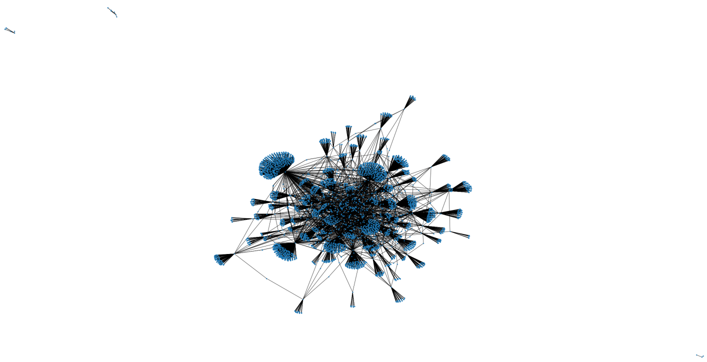

Figure 5: Tweets 20,000-30,000 in Filtered Data set

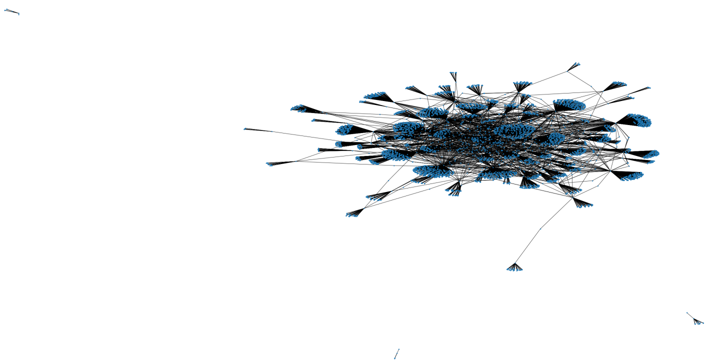

Figure 6: Last 10,000 Tweets in filtered data set

5 Conclusion

Twitter data is readily available and proves a valuable recourse to understand and analyse information flow over online social networks. This data can be used to explore how influence online can be measured and assessed, and this report aims to explore a number of techniques to do so.

This work uses 1.8 million Tweets from the 6th of January all containing the keyword 'trump', exploring a number of different ways of measuring influence and identifying drivers within the online network revealed a number of tools which gave varied, but corresponding insights. These Tweets were collected on the 6th of January 2021, the day which the Counting of the Electroral College votes occured for the 2020 United States Presidential Election. On that day a riot was lead by Trump, with rioters aiming to cause the Electoral College to reject the result of the election. The rioters moved to the United States Capitol Building, eventually breaching police lines and gaining entry to the building, an event which ultimately cost five lives.

After the data was collected using the Twitter API, the data was cleaned, working around issues created by the collection of a username rather than a screen name. When the data was collected, the username of the user who originally posted the Tweet was saved, but this is a non-unique identifier, preventing Tweets from being ascribed to a single individual.

Preliminary data analysis and looking at frequency of several metrics across the data set revealed interesting patterns, including an increase of frequency of hashtags or users which preceded or followed several key events throughout the day.

Considering the five most common hashtags in the data, there is a spike in the frequency of the hashtag 'CapitolRiots' after Trump's Save America Rally begun, where Donald Trump, then his lawyer Rudy Giuliani spoke, encouraging the rioters. Interestingly, the occurance of this hashtag is non zero throughout the entre day, indicating that it was a topic of discussion, appearing within our dataset.

When rioters and police began to clash at the capitol there was a spike in the occurrence of the hashtag '25thAmendmentNow'. This refers to an Amendment in The Constitution of the United States which was viewed as a tool to remove Trump from power. Another interesting feature is the hashtag 'BREAKING', often used by media organisations to indicate that a Tweet contains breaking news, for which spikes in the frequency of this hashtag correspond to all major events discussed, with a large spike occurring in conjunction with the rioters breaching police lines and entering the Capitol.

To understand how different users are interacting with popular users, we also consider the 14 most frequently replied to users (Figure 2). When there is a large number of replies to a specific user, this indicates that a user, or their actions are driving discussion. In Figure 2, two notable spikes are those to Representative Mo Brooks (RepMoBrooks), who released a public statement condemning the riots, and to Rudy Giuliani, who left a voicemail on the wrong phone, in which he urges Republican Senators to delay the counting of senate votes.

These findings were supported by the sentiment analysis, performed using the lexicon and rule based tool VADER. VADER allowed a compound sentiment score to be calculated for each Tweet, plotting this metric against time revealed that average sentiment across the data set reached it's most negative subsequent to the beginning of the riot, peaking to a local maxima before remaining steady throughout the remainder of the data.

Another avenue explored was network analysis tools. This was performed by first filtering the data, to only include Tweets which were in reply to one of the 100 most replied to users, reducing our data set to 45,863 Tweets. Networks were constructed for sets of 10,000 Tweets, to represent the change of the network across time. The users which have the highest node correspond to the most replied to users identified in the initial frequency based approach.

5.1 Future Work

Our work revealed interesting findings using a variety of tools. Future work will focus on combining these approaches, using sentiment data to filter people in the network to identify users of interest, who drive changes in sentiment with network analysis techniques to model and understand how these sentiments move across a network.

To do this, a more sophisticated sentiment analysis tool will need to be used which is able to distinguish sentiment with higher resolution, to better identify when a new sentiment occurs in the network.

The data set used was well suited to this task, as it corresponded with several large events which occurred in the U.S, which involved Donald Trump. This ensured that the results could be verified against real-world findings, although the techniques discussed can be refined to apply to smaller events which are both external to or internal to the network. To generalise these results, another tool will need to be developed which is able to identify when a significant event is detected.

The project achieved its objective, comparing and exploring a number of techniques to better understand and quantify influence in online social networks.

References

Ardevol-Abreu A, Gil de Zuniga, H, Weeks, E, 2015,‘ Online Influence? Social Media Use, Opinion Leadership, and Political Persuasion’, *International Journal of Public Opinion Research*, 29(2):214–239.

Axelrod, R 1997, ‘The Dissemination of Culture: A Model with Local Convergence and Global Polarization’, *Nature Human Behaviour*, 41(2):203–226.

Brian, W, Ardèvol-Abreu, A, Gil de Zúñiga, H, 2017, ‘Online influence? Social media use, opinion leadership, and political persuasion.’, *International Journal of Public Opinion Research*, 29.2 : 214-239.

Cover, T, Thomas, J, 1999, ‘Elements of information theory’, *John Wiley Sons*.

Frenkel, S, 2021, ‘How the Storming of Capitol Hill Was Organised on Social Media’, *The New York Times*, accessed at https://www.nytimes.com/2021/01/06/us/politics/protesters-storm-capitol-hill-building.html.

Galuba, W, Asur S, Huberman, A, Romero, D, 2011, ‘Influence and passivity in social media’ .*Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 18–33.

Haddadi, H, Benevenuto, F, Cha, M, Gummadi, P, 2010, ‘Measuring user influence in Twitter: The million follower fallacy’,*International Conference on Weblogs and Social Media ICWSM*, 10(10-17):30.

Hofman, J, Mason, W, Bakshy, E Watts, D, 2011, ‘Everyone’s an influencer: quantifying influence on Twitter’, *In Proceedings of the fourth ACM international conference on Web search and data mining*, pages 65–74.

Hutto, C, Gilbert, E, 2014, ‘VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text’, *Eighth International Conference on Weblogs and Social Media (ICWSM-14)*, Ann Arbor, MI.

Liu, X, Mitchell, L, Bagrow, J, 2019, ‘Information flow reveals prediction limits in online social activity’, *Nature Human Behaviour*, pages 122–128.

Minaeian, S, Paquet-Labelle, A, Beaudry, S, Dubois, E, 2020, ‘Who to Trust on Social Media: How Opinion Leaders and Seekers Avoid Disinformation and Echo Chambers’, *Social Media + Society*, 6(2).

Petras, G, Loehrke, J, Padilla, R, Zarracina, J, Borresen, J, 2021, ‘Timeline: How the storming of the U.S. Capitol unfolded on Jan. 6, 2021’, *USA Today*, accessed at https://www.usatoday.com/in-depth/news/2021/01/06/dc-protests-capitol-riot-trump-supporters-electoral-college-stolen-election/6568305002/.