



Cross-platform information spread during the January 6th capitol riots

Lynnette Hui Xian Ng¹ · Iain J. Cruickshank¹ · Kathleen M. Carley¹

Received: 24 January 2022 / Revised: 12 July 2022 / Accepted: 15 July 2022

© The Author(s), under exclusive licence to Springer-Verlag GmbH Austria, part of Springer Nature 2022

Abstract

Social media has become an integral component of the modern information system. An average person typically has multiple accounts across different platforms. At the same time, the rise of social media facilitates the spread of online mis/disinformation narratives within and across these platforms. In this study, we characterize the coordinated information dissemination of information laden with mis- and disinformation narratives within and across two platforms, Parler and Twitter, during the online discourse surrounding the January 6th 2021 Capitol Riots event. Through the use of username similarity, we discover joint theme endorsements between both platforms. Using anomalously high volume of shared-link matches of external websites and YouTube videos, we discover separate information consumption habits between both platforms, with very few common sources of information between users of the different platforms. However, through analyzing the similarity of the texts with Locality Sensitive Hashing of constructed text vectors, we identify similar narratives between the platforms despite separate consumption of external websites, highlighting the similarities and differences of information spread within and between the two social media environments.

Keywords Social media coordination · Coordinated information dissemination · Parler · Twitter · Natural language processing

1 Introduction

Social media has become an integral component of the modern information system. Its usage has been expanding at an astonishing rate worldwide and shows no indication of decline (Center 2021). There is now a plethora of social media platforms, and the average person has multiple social media accounts (Index 2021). Social media usage has become increasingly cross-platform, with many users drawing information and interacting with more than one social media platform. While there is much research into social

media, most research focuses on user behavior on a particular platform; there has been considerably less research on cross-platform usage of social media platforms.

The rise of social media facilitates the spread of online mis/disinformation campaigns both within and across these social media platforms (Ng and Taeihagh 2021). These online information campaigns have been widely recognized and shown to have substantial impacts on society, ranging from politics (Pierri et al. 2020; Golovchenko et al. 2020) to vaccination reception (Swire-Thompson and Lazer 2020; Ng and Loke 2021) to protests (Ng and Carley 2022; Magelinski et al. 2022).

Due to the impact of online mis/disinformation, new fields like Social Cybersecurity have emerged to develop means of mitigating and fighting the information threat (Carley 2020). Social media conflict studies have developed methods to extract inter-community conflict behaviors (Datta and Adar 2019) and show that cross-communities activity between Reddit communities is related to mobilization of conflict (Kumar et al. 2018), an event that which can escalate into large-scale offline violence. These form the need for the emerging field of social cybersecurity field which aims to aid building healthy online communities. Recent research in the field has identified

Iain J. Cruickshank and Kathleen M. Carley have contributed equally to this study.

✉ Lynnette Hui Xian Ng
lynnetteng@andrew.cmu.edu

Iain J. Cruickshank
icruicks@andrew.cmu.edu

Kathleen M. Carley
kathleen.carley@cs.cmu.edu

¹ Center for Computational Analysis of Social and Organizational Systems, Carnegie Mellon University, 4665 Forbes Avenue, Pittsburgh, PA 15213, USA

the use of external content like URLs (Horawalavithana et al. 2021; Giglietto et al. 2020; Cruickshank et al. 2021) and cross-platform behaviors (Ng et al. 2021; Kin and Sameera 2021; Iamnitchi et al. 2020) in the spread of mis/disinformation.

On January 6th 2021, supporters of then-President Donald Trump stormed the US Capitol to protest the result of the 2020 US Presidential elections. This event, later known as the January 6th Capitol Riots, has been recognized for the enabling role that social media played in organizing the riots (Frenkel 2021; Timberg et al. 2021). However, it is not well understood how multiple social media platforms are linked and the role of media like websites and videos play in information dissemination during such events.

In this study, we further the analysis of social media discourse surrounding the January 6th Capitol Riots by tackling the topic of cross-platform information spread between two social media platforms: Parler and Twitter. This work extends ideas from earlier preliminary analyses of narrative clusters in Parler Ng et al. (2021) and joins ideas on external link sharing behavior on Twitter Cruickshank and Carley (2020). We gathered data from Parler and Twitter surrounding the Capitol Riots incident and analyzed the coordinated dissemination behavior of external websites and YouTube videos and their content. We analyze the cross-platform information environment and further analyze the groups and narratives with a focus on user ideological affiliations by categorizing the users into military, QAnon and Patriot users. These user classes were based on news of the profiles of people that were actively involved in the January 6th events NPR Staff (2021).

In particular, we ask the following Research Questions with respect to the January 6th Capitol Riots:

1. RQ1: Are there similar users or themes of users within and across the platforms?
2. RQ2: What are the information dissemination patterns within and across platforms for YouTube videos and websites?
3. RQ3: Are there similar narratives that are shared across platforms, and what are their patterns of dissemination through YouTube videos and websites?

To answer these research questions, we propose a methodology that studies the problem by building five different similarity graphs based on usernames, URLs, website content and YouTube video transcripts. Using these graphs, we make further investigation into the narratives shared within and between these social media platforms through external links, pointing to external websites and YouTube videos. We use the graphs constructed by the similarity of textual website content and the transcription of the YouTube audio to infer the trend of the combination of audio and visual narrative dissemination during the event.

In this study, we make the following contributions:

1. We develop a systematic methodology for analyzing coordinated information dissemination through common external links pointing to websites and YouTube videos shared between social media users. We enhance this methodology toward understanding shared narratives through the content of the links using natural language processing methods.
2. For RQ1, we give an overview of the types of information users with different affinities spread by characterizing the users through user name similarity. The results demonstrate how users on both platforms endorse similar themes in their usernames.
3. For RQ2, we analyze the interplay between external link sources (website content, YouTube video content) and social media platforms. Links are shared more within than across platforms.
4. For RQ3, we analyze within- and across-platform coordinated information dissemination in Parler and Twitter around the January 6th 2021 Capitol Riots event by comparing textual similarities. We observe separate information consumption sites between the two social media platforms, yet the same narratives are disseminated.

2 Related work

Several recent works have examined the spread of external content in social media discussions. Since the number of characters in a post on social media platforms is often limited, for users to disseminate longer narratives, they typically place their content on external websites and leverage social media as content distributors (Hounsel et al. 2020). These external websites shared are often along ideological or political lines on social media platforms like Twitter Kuzma et al. (2021) and have been observed to share deceptive content in a coordinated manner (i.e., during the 2019 European elections (Pierri et al. 2020)). External websites have also shown to be an important part of misinformation-laden conversations like the Twitter discourse around the COVID-19 vaccination and are often used to spread misinformation (Cruickshank et al. 2021; Horawalavithana et al. 2021).

Additionally, YouTube videos have been heavily spread on social media platforms and used in a wide range of mis/disinformation campaigns. The video hosting site has been used to strategically coordinate an information operations campaign using narratives during the White Helmets campaign in Syria Kin et al. (2021); Choudhury et al. (2020); Pacheco et al. (2020); Iamnitchi et al. (2020). YouTube videos have also been employed for the spread of propaganda by far-right extremists (Squire 2021), state-sponsored trolls

(Golovchenko et al. 2020) and religious extremists (Hussain et al. 2018; Klausen et al. 2012).

Analysis of link sharing behavior has also given rise to observations of coordinated link sharing behavior. Recent work highlights that mis/disinformation campaigns not only spread external websites on social media platforms but also do so in a coordinated manner (Pacheco et al. 2021; Horawalavithana et al. 2021; Giglietto et al. 2020). This coordinated link sharing behavior further spreads the mis/disinformation on the external website by getting the website to trend higher on the social media platform and hence artificially boosting its popularity (Giglietto et al. 2020).

Beyond just link sharing behavior, coordinated behavior on social media presents in other forms as a means of spreading mis/disinformation (Pacheco et al. 2020; Nizzoli et al. 2020; Ng et al. 2021). This coordinated behavior can be detected through the users' sharing of the same—or nearly the same—text, website URLs, or social media artifacts over the course of a short period of time (Magelinski et al. 2022; Pacheco et al. 2021).

Finally, recent research has also centered on mis/disinformation campaigns on the social media platform Parler. Parler positions itself as a free speech social network and has been recently scrutinized as a platform that facilitated the coordination of the January 6, 2021, Capitol Riots (Munn 2021), where hundreds stormed the US Capitol Hill building, calling out themes of election fraud. Past work have found coordination between Parler users in spreading disinformation-laden content on the platform (Ng et al. 2021). However, it is less clear the nature of cross-platform information spread or user interaction in such mis/disinformation-laden events.

3 Terminology

This study performs a multi-platform examination of coordinated information dissemination through website contents and YouTube videos across two social media platforms: Parler and Twitter. In this section, we define some of the terminology used in this study.

A post on the Parler platform is called a *Parley*, which can range in length up to 1000 text characters. A post on the Twitter platform is a *Tweet*, which ranges in length up to 280 characters. We collectively use the term *Post* to refer to both Parleys and Tweets, especially when highlighting analysis between the two platforms. A *website URL* is the complete web address to a website. The text on the website is referred to in this study as *website content*. The *full URL* is the entire URL. Sometimes the URL includes a *search term string* which is used to identify search queries; a *base URL* the URL that does not include the search term; and the *URL domain* the web domain identifier for the website. For example, in the string

Table 1 Summary of data terms used in this study and their relationships

Source	Descriptor	Object type(s)	Data type(s)
Parler	Parley	Post	Text up to 1000 characters
Twitter	Tweet	Post	Text up to 280 characters
Website	URL	Domain	Unique text string
		Webpage	Text of unspecified length from content of webpage indicated by URL
YouTube	Video	Video ID	Unique text string
		Transcript	Text of unspecified length from audio transcribed from the content of video

“https://twitter.com/search?q=social%20media&src=typed_query”

the full URL is the entire string; the base URL is “https://twitter.com/search”; the search term string is “q=social%20media&src=typed_query”; and the URL domain is “https://twitter.com.”

A *YouTube link* refers to the web address of a YouTube video. A *YouTube Transcript* refers to the text encoding of what is said in a YouTube video.

We use the phrase *link* to collectively refer to both website URL and YouTube link. We use the phrase *text* to collectively refer to both URL content and YouTube transcript.

Crossover refers to the links that connect between both Parler and Twitter social media platforms.

Table 1 contains a summary of the various terms used and their relationships to each other.

4 Data collection

Due to the unique nature of the setup of each social media platform and the extracted data, we use a multitude of techniques for data collection and processing. We describe the data collection and processing techniques in this section. A diagrammatic overview of the data processing framework is presented in 1.

4.1 Parler data

We obtained data on Parler surrounding the Capitol Riot event from a previous study on coordinating texts surrounding the Capitol Riots (Ng et al. 2021). This dataset consists of a partial HTML scrape of Parleys shortly after the Capitol Riots when Internet users sought to preserve the data from the social network when Amazon Web Services banned Parler from being hosted on its service (Lyons 2021). In

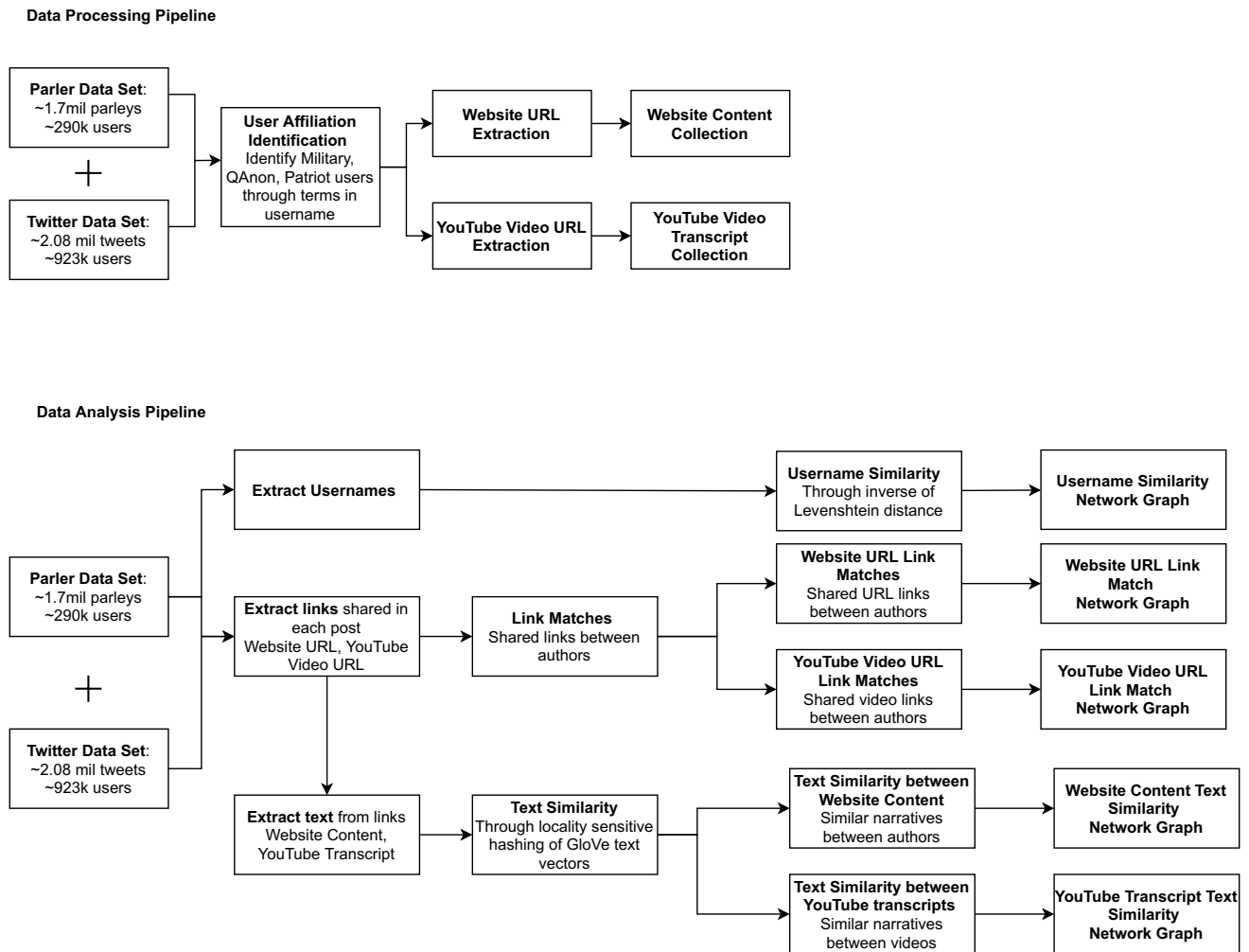


Fig. 1 Pipeline for Data Processing and Analysis

total, the dataset consists of 1.7 million posts from 290,000 unique users between January 3rd to January 10th of 2021.

4.2 Twitter data

Twitter data was obtained from a previous study on Twitter discourse surrounding the Capitol Riots (Ng et al. 2021). The data was collected with Twitter V2 REST API using the following well-known hashtags associated with the events of January 6th: #stopthesteal, #stopthefraud, #marchfortrump, #marchtosaveamerica, #magacivilwar, #saveamericarally and #wildprotest. This dataset in total consists of 2.08 million tweets from 923,008 users from January 3rd to January 12th of 2021.

Typically, we collect Twitter data for one week after the event occurrence due to the half-life of tweets about an event (Alperin et al. 2019). That is, the majority of the tweets of the event are superseded in 8.1 ± 2.2 days. Due to this, the Twitter data has two more days of data as compared to the Parler

data—from January 10th to January 12th 2021. Since both data collection were opportunistic samples to capture the discourse of the event, we decided not to remove any data because much of the data has since been removed from both social media platforms.

4.3 User affiliation identification

The Parler dataset had segmented users into three affiliations: “Military,” “QAnon” and “Patriot” users, based on string identifiers found in their user account handle or user text description [?]. We adopt the same segmentation in our study. We adopt the same process to segment Twitter users into the three groups, using the same string identifiers that were used in the Parler user affiliation segmentation found in their user name or user text description. The user affiliation segregation will be used to further enhance our analysis

into user coordination behavior. The full list of terms used for this user affiliation segregation is presented in Table 3.

Past work has shown that different online communities have different linguistic markers, and community affiliation correlates with user activity level (Wang et al. 2016). Building on this notion, we will analyze user information dissemination activity in terms of their affiliated community, investigating how the narratives of the Capitol Riots differ for different communities.

4.4 URL identification and website content data collection

We begin this section by extracting the external website URL(s) shared in each post. For Parleys, external website URLs are extracted by first finding *href* tags in the HTML pages of the Parler scrape, followed by extracting the links from those tags. For tweets, external website URLs are extracted from the *ents* field from the tweet data.

Next, we collect the website content from the URL. We check if the URL was shortened using the a link shortener by identifying if it contains any known link shortener services as subphrases (i.e., bit.ly, ow.ly, etc.). If the URL was previously shortened when posted, we expand it using the Unshortenit Python package¹. We also remove the query terms at the end of all URLs, except YouTube URLs. The processing of YouTube URLs are described in the next section. These query terms are typically prefixed by “?=,” and we thus retrieve the *base URL* upon removal.

Using the full URL, we use the Selenium Python package with a Chrome driver² to scrape the page’s HTML. We further parse the retrieved HTML using the BeautifulSoup Python library³ to obtain only the page content. We skip over pages that cannot be scraped in this manner. In total, 1.8 million URLs were shared during the event. Of these URLs, only 80,733 URLs were unique. At the time of scraping, 56,863 URLs remained and were therefore retrieved.

Lastly, we match the users to the URLs they shared. For each user, we annotate a list of base URLs and URL domains shared by the user. This information will be used in subsequent steps, detailed in Sect. 5.1.1.

4.4.1 YouTube transcript data collection

To understand the narratives spread within the audio content of YouTube videos shared during the Capitol Riots event, we turn to the YouTube transcripts. We extract YouTube links shared in each post using regex by identifying variations

of the YouTube domain name (“youtube.com/,” “youtu.be.com/,” “m.youtube.com/”). This domain name is then suffixed by a string of eleven alphanumeric characters representing the YouTube video identification (ID) code. We use the YouTube Data API to retrieve the video transcript, which provides a text transcription of the spoken material. In total, 111,146 unique YouTube video URLs were shared during the event. We were only able to collect data from 53,061 videos. The other videos were removed at the time of collection.

5 Methodology

To better understand the shared content between users, we created novel algorithms to construct network graphs and represent similarities between users in terms of shared content. A diagrammatic overview of the data processing framework is presented in Fig. 1.

We compared the similarity of users across three dimensions: username similarity, URL similarity and text similarity. Each dimension gives us a different perspective of coordinated information dissemination. Username similarity reflects the joint identification toward an affiliation; URL similarity reflects the joint referral of an external link; and text similarity reflects the joint amplification of a narrative, either through external website content or a YouTube video. After obtaining the similarity values through these three dimensions, we construct network graphs to visualize the user similarities by linking users through their presence and strength of similarity in each dimension, further aiding in user interaction analysis.

5.1 Username similarity

As part of our investigation into RQ1, we look into username similarity to identify similar users and themes of users based on their expression of these attributes in their public username.

We identified group identity clusters of users. These were clusters of individuals that signal the same group identity through their username. We identify such clusters by comparing the similarity of their usernames. Previous work has observed that users use phrases within their usernames to identify as belonging to a certain group. In particular, a survey on Twitter media perception reveals that cues such as avatar construct and username are likely to signal identity and lead to differential responses to political information (Cooks and Bolland 2021). In addition, having very similar usernames across platforms can signal that both accounts belong to the same person, as username is a key feature for social profile identity mapping (Correa et al. 2012). We incorporate these ideas and leverage username similarities

¹ <https://pypi.org/project/unshortenit/>.

² <https://selenium-python.readthedocs.io/>.

³ <https://www.crummy.com/software/BeautifulSoup/bs4/>.

to highlight these professed user identities and, by doing so, find communities of users.

To do this, we used the Levenshtein distance between the usernames of all users in the dataset. The Levenshtein distance is a metric measure of two word strings through the minimum of single-character edits required to change one word string to another. These edits can be in the form of insertions, deletions or substitutions (Levenshtein 1966). We disregarded usernames that are less than three characters long. For the rest of the names, we perform a pairwise comparison of all the usernames across the dataset, obtaining a distance measure of how similar the two usernames are.

We then transform the Levenshtein distance metric into a username similarity metric. We use min–max normalization to scale all the distances collected in the dataset and define the username similarity metric as $(1 - \text{normalized distance})$. This measure places a higher weight on username pairs that require lesser character edits to change one username string to another, hence increasing their similarity score.

In a later step, we will overlay the user affiliation information with the username similarity metric to construct network graphs representing user similarity. The graph construction is detailed in Sect. 5.1.3.

5.1.1 URL and YouTube links matches

As part of our investigation of RQ2 to understand information dissemination patterns, we perform website URL and YouTube link matching, both within and across platforms. URL and YouTube link matching between users represent sharing the same content between and across platforms. This allows inference of the coordinated effort by multiple users to amplify particular sets of links. The link sharing similarity between any two users is calculated by the number of times each pair of users shared the same base URL in their post. We annotate each pair of users by the total number of same base URLs shared.

5.1.2 Text similarity

To understand cross-platform information spread (RQ3), we studied the presence of similar narratives in the link content shared on both platforms. We adopt this method of analyzing information spread based on the common external links because, in our dataset, there are no Tweets with external links that are Parleys and no Parleys with external links that are Tweets. However, this does not mean that information and narratives do not spread between both platforms; nor does that mean that there are no groups of actors that communicate the same messaging between platforms.

Therefore, to identify the extent of information spread between and across platforms, we turn to text similarity methods which identify posts that are similar in their texts.

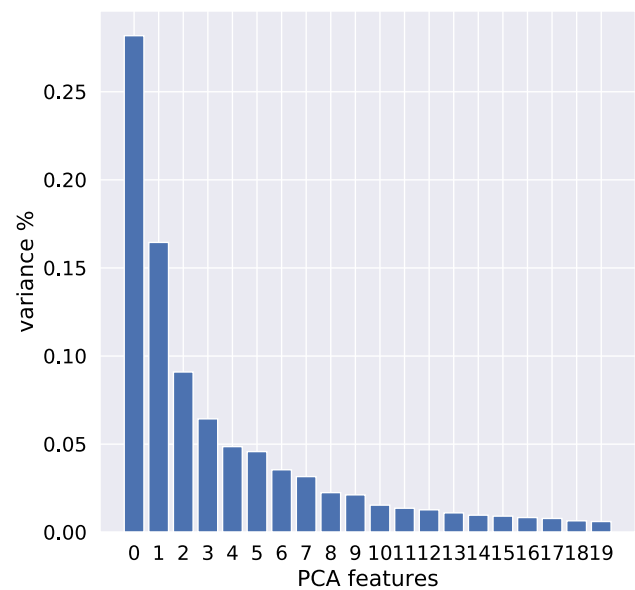


Fig. 2 Principal Component Analysis chart of the features of all the text vectors

With similar posts identified, we construct network graphs to observe the patterns of narrative dissemination through YouTube videos and websites.

Text similarity represents the similar narratives that are being shared among users of the two social media platforms. We form one text similarity network for website content and another text similarity network for YouTube transcripts. To measure text similarity, we perform a k -nearest neighbors (k NN) approximate search of text vectors formed via hash-based indexing (Sugawara et al. 2016).

We first preprocess the texts to remove punctuation, stop-words and other social media artifacts (hashtags, @-mentions, URLs, etc.). We also remove texts that the web scraper only returned the phrase “Advertisements...” Next, for each text, we formulate a 300-dimensional document–word embedding representation of the text using GLoVe vectors. GLoVe vectors are pre-trained word embeddings trained on 6 billion Wikipedia words, resulting in 300-dimensional real-valued vectors (Penningto et al. 2014). The document is represented as a bag of words and then embedded as GLoVe vectors. We opted to use this method because we are dealing with extremely short texts. After preprocessing the texts, the mean number of words in each text is 10.6 ± 6.3 .

Next, we reduce the 300-dimensional vector into a 20-dimensional vector. We choose 20 dimensions after performing principal component analysis (PCA) on the entire text 300-dimensional vector space (Fig. 2). An analysis of the PCA chart shows the variance of the features tends to 0 after 20 features. Hence we select 20 dimensions and perform dimensionality reduction of the vectors. This compresses the vector into the top 20 salient components, which

aids in vector comparison. We opted to use a dimensionality reduction technique on the vectors because subsequently we will need to perform an all-pairs cosine similarity search for to find similar texts. This is a computationally expensive operation of the complexity of $O(n^2)$. Given that we have over 3.7 million texts to compare against, the reduced dimension serves to speed up the calculation. This compute was ran on a 16-core Xeon-R 3.3GHz Windows desktop with an NVIDIA 3090 GPU. The time taken for one iteration of the all-pairs search for the reduced vector is 2.76 seconds, while the time taken for the 300-dimensional vector is 1 minute 9 seconds. We also manually verified that this improvement in performance gain does not affect the results of the text similarity by looking through the top ten matched posts. For example, the text “Facebook announced Monday would remove content includes phrase stopthesteal” matched with “Facebook banning use phrase stopthesteal.”

Next, we use the hash-based technique of Locality Sensitive Hashing and transform the 20-dimensional word vectors into a 20-dimensional binary hash (Gioni et al. 1999). This step allows for an efficient search of the closest vectors within the reduced space.

We index the 20-dimensional hash vectors using the FAISS (Johnson et al. 2017) library for Locality Sensitive Hashing. The FAISS library utilizes Graphics Processing Units (GPUs) to parallelize the similarity search for vectors, shortening the computation time compared to performing processing on the Central Processing Unit. We then perform an all-pairs cosine similarity search to determine the top $k = \log_2 N$ closest vector to each text vector, where N is the total number of texts. The user information for each of the top k texts are stored for use in the network construction (Sect. 5.1.3).

5.1.3 Similarity network construction

Using the similarity measures described in the previous sections—username, URL, and text similarity—we constructed network graphs to visualize the similarities between the users. Separate networks were constructed for each similarity dimension. In each network graph, a node represented a user on either the Parler or Twitter social media platform. Two nodes were joined with a link if they were similar enough in that dimension. The links for each of the networks were variously weighted. For username similarity, the links were weighted according to the username similarity metric, a transformation of the Levenshtein distance. For URL similarity, the links were weighted by the number of times the two users posted the same URL. For text similarity, the links were weighted by the number of times the users posted a similar text.

To extract the core structure of the graph, we threshold the graph networks via link weights as was done in previous

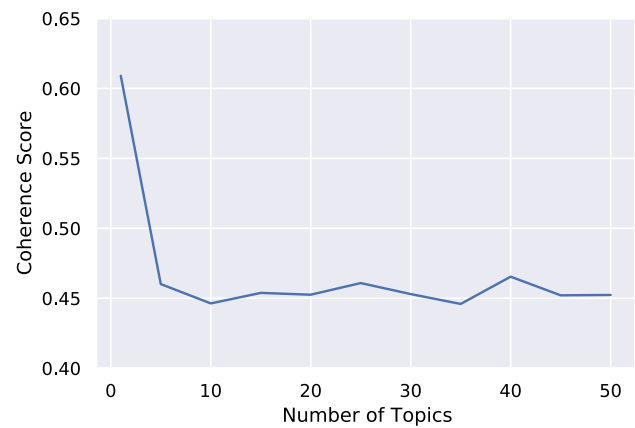


Fig. 3 Coherence Scores plot for LDA against increasing number of topics. We use five topics for determining the LDA in our clusters due to the stabilized coherence scores after five topics

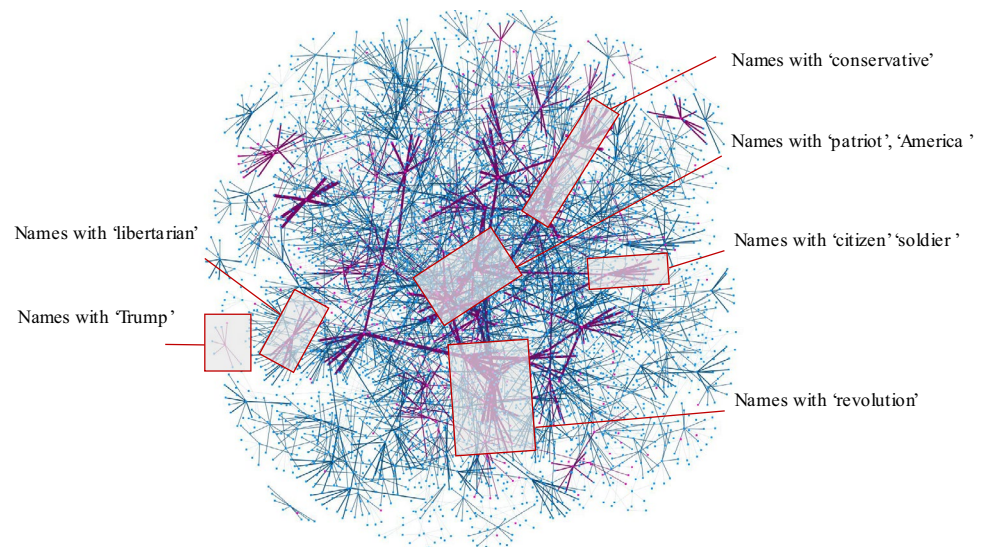
studies which constructed networks of similar properties—textual similarity networks from social media texts (Ng et al. 2021). This thresholding method on textual similarity methods has been demonstrated to sieve out the core members of the graph structure, reducing clutter and noise for the next step of the analysis, yet preserving the nodes and links where the connecting links were above the mean plus one standard deviation value of all the links in the graph.

To understand the themes in each similarity graph, we segment visual groups of user clusters. For the username similarity graph, we retrieved the usernames from each of the clusters and inspect them for similarities, e.g., common phrases used in the names. For the YouTube transcript and Website Content graphs, we retrieve the corresponding transcript and website content. We perform Latent Dirichlet Allocation (LDA) on the text arising from the website content and YouTube transcripts of the clusters using the Gensim Python library⁴. We analyze the coherence scores, the measure of semantic similarity between words in the topic, produced by the LDA algorithm across an increasing number of topics. The output of the analysis is shown in Fig. 3. We utilize the elbow rule in this figure and select five as a suitable number of topics, after which the coherence scores stabilize as the number of topics increases. Using this result, we obtain five topics for all the clusters. We then manually interpret these common topics of discourse within the cluster, joining topics together where appropriate, and report the results. Trends observed in the graphs are summarized in Sect. 6.1.

To determine clusters from the dense thresholded graph, we used the Louvain network clustering technique to segment the clusters based on their node connection (Blondel et al. 2008). We then retrieve the usernames from the

⁴ <https://radimrehurek.com/gensim/models/ldamodel.html>.

Fig. 4 Username similarity graph, highlighting notable group identity clusters. The width of the links represents the strength of username similarity calculated by the Levenshtein distance. Pink nodes are Parler users, Blue nodes are Twitter users. Line color represents the platforms the pair of users come from; purple lines indicate that a Parler and a Twitter user have extremely similar username. We highlight clusters of common group identity terms reflected by the users in their usernames, not the actual usernames



clusters. For a more in-depth analysis of the clusters, we retrieved the URLs and YouTube links shared by the users identified by the retrieved usernames.

6 Results

In this section, we detail the results of the network graphs constructed through the similarity metrics.

6.1 Username similarity shows joint theme endorsement

Social identity theory provides a basis for human perception and interaction within groups. One way of identifying with a group is through self-classification (Tajfel et al. 1971), which is observed on social media through the presence of an identity or a group name in the username. A name can contribute to a post being successful, among other factors (i.e., content) (Lakkaraju et al. 2013).

The username similarity graph, shown in Fig. 4, has some distinct users and groups of users that declare their support of certain ideologies or group identities through their username. We zoom into the clusters with high link weights between the users, that is, the graph shows thick link widths and have links across both platforms. We manually inspect the user groups. Users with high link weights represent that their usernames are very similar to each other. We then inspect the metadata of these users and noted common themes in the user names as the identities these groups of users affiliate with.

Through this method, we identify many user clusters whose members identify with similar themes. In Fig. 4, this is seen by the purple clusters of links and nodes, which represents that the pair of users with vastly similar usernames

come from two different platforms: Parler (pink), Twitter (blue). We highlight six group identities of note that are related to the Capitol Riots event. We collectively interpreted their texts by manually inspecting all their posts (Tweets/ Parleys) and present them as follows:

1. Names with “libertarian”: This cluster called for a party “to end the duopoly.” This is a theme that first surfaced in the 2016 elections, calling out the weaknesses of the two main parties in the USA, the Republicans and the Democrats. The “End the Duopoly” campaign seeks political realignment through a third party (Benn 2017).
2. Names with “Trump”: This cluster called for protests downtown against electoral fraud and called then Vice President Mike Pence a traitor when he refused to turn over the 2020 election results.
3. Names with “conservative”: This cluster mainly reported news surrounding the Capitol riots. It also spread a disinformation narrative that “Antifa terrorists” were bussed to the Capitol to the #stopthesteal rally.
4. Names with “patriot” and “America”: This cluster talked about the Air Force veteran who was shot during the riots and also spread the narrative of an “Antifa terrorist” being behind the January 6th events.
5. Names with “citizen” and “soldier”: This cluster voiced its support for the #stopthesteal rally and “vow to never concede defeat.”
6. Names with “revolution”: This cluster called out for voter fraud with the mail-in ballots and examined leaked emails of presidential candidates and the Democratic National Committee, spreading them through Wikileaks file URLs and Dropbox links.

Table 2 Proportion of the top ten URLs shared in each platform, measured against the total number of URLs shared in the respective platforms. Those websites identified as having a politically far-right bias by third-party sources are denoted with an *

URL	Proportion of URLs (%)
Parler	
thegatewaypundit.com*	16.83
uscouriertoday.com	5.11
americanmilitarynews.com	5.01
vocaroo.com*	4.94
noqreport.com*	2.90
zerohedge.com*	2.81
neonrevolt.com*	2.20
wnd.com*	2.10
breitbart.com*	1.49
wanynedupree.com*	1.43
Twitter	
theverge.com	5.97
dailybeast.com	3.98
thegatewaypundit.com*	3.41
news.yahoo.com	2.63
rawstory.com	2.55
wsj.com	1.65
cnn.com	1.65
reut.rs	1.39
washingtonpost.com	1.14
vox.com	1.41

We note that all these six clusters echoed the #stopthesteal hashtag and associated narratives. This particular hashtag and its associated narratives advocated for overturning the 2020 US Election results in favor of then-President Donald Trump. It has also been a key phrase associated with the disinformation campaign calling for voter fraud and delegitimizing the 2020 US Elections.

These clusters of users mobilize other users to their cause by developing a sense of identity among participants through common themes in their usernames and effectively communicating their goals of the movement through calls to action campaigns, such as “End the Duopoly.” This is similar to the observations made on the subreddit r/The_Donald regarding the 2016 US Election users that had clear calls to action and a distinct identity retained and mobilized the largest number of participants (Flores-Saviaga et al. 2018).

6.2 Website URL sharing shows separate information consumption sites

The website sharing network constructed with website URL matches and YouTube video matches is shown in Fig. 5. We present Parler and Twitter users in the same graph to depict

the information dissemination across and within the platforms. In these graphs, we observe that each platform has a different information ecosystem, stemming from the fact that clusters of link-sharing activity are within platforms. 94.4% of the users share common URLs/Links among other users of both platforms, while very few users (5.6%) share URLs/Links across platforms.

In cross-platform link sharing, website URLs present a more platform-isolated view while YouTube Links are more intermixed. This indicates that a higher percentage of YouTube Links are shared across users of both platforms. More specifically, 0.38% of the website URL matches are shared by users of both platforms, while almost twice (0.72%) of YouTube Links are shared by users of both platforms.

Parler users spread news from websites mostly associated with the right-wing conservatives: “vocaroo.com,” “breitbart.com,” “waynedupree.com,” “noqreport.com” and so forth. We determine the political slant of these websites (right/left) through the Wikipedia information about them. On the other hand, a large percentage of Twitter links shared come from more reputable websites like Wall Street Journal, The Verge, CNN, and Yahoo. Other websites shared in smaller proportion include Facebook posts (0.43%), “raw-story.com” (2.55%) and “lailluminator.com” (1.45%). We present the proportion of the top 10 URLs shared in each platform in Table 2.

One website base URL that is commonly spread between the users in both platforms is “thegatewaypundit.” The Gateway Pundit is a known far-right news website with the slogan “where hope finally makes a comeback.” It has been known to publish fake news and hoaxes by MediaBiasFactCheck, a crowdsourced media bias rating website The Gateway Pundit (2021).

The top themes shared based on the URL content derived from the thresholded website URL graph include: the police began shooting the protesters despite the protesters not having any weapons; the need to flood to the capitol at Washington DC to stop certifying the 2020 election; the lack of Republican observers present at some election tabulation centers; and the presence of Antifa domestic terrorists ahead of the #stopthesteal march.

Analyzing the users that share at least one YouTube video link in the dataset, we observe that Twitter users shared more videos than to Parler users. Comparing the ratio of the number of times a video was shared on the platform against the number of posts in the platform, we observe that this ratio is $2.06e^{-4}$ for Twitter and $6.96e^{-5}$ for Parler. This indicates that, on average, Twitter users share more YouTube videos compared to Parler users, after accounting for the larger user base on Twitter as compared to Parler. Similarly, the ratios of the number of unique links to the total number of posts for Twitter and Parler are 0.16 and 0.08, respectively. This

indicates that Twitter users share twice as many unique links compared to Parler users.

The proportion of unique links shared to the number of users are 12.7% and 0.7% on Parler and Twitter, respectively. This indicates that individual Parler users engage more with external content, leading to them sharing more unique URLs on average. While narrative themes between sites shared by Twitter and Parler users are similar, the origination of the common content are different. The users on each platform draw on different information sources, demonstrated by the small proportion of common URLs (8.76% of unique URLs were shared between both platforms) between the users of both platforms. YouTube links to videos that contained the live streaming of the Capitol Riots event were most commonly shared among both platforms, up to 2,857 times. The top themes shared among YouTube videos include videos of the Capitol Riots event itself and videos explaining voter fraud, which are determined by the most common terms from the YouTube video transcripts.

In investigating the number of users that have shared links to another user across platforms (i.e., crossover links), we report the following observations. In order to understand those users who have a link to a user from then other platform we looked at the known, January 6th-related user affiliations of QAnon, Patriot and Military. We found 64.6% of all Patriot users contain a crossover link, 26.3% of Military users have crossover links and then 9.1% of QAnon users have crossover links.

In terms of YouTube URLs, we find that 62.1% of QAnon users have crossover links, followed by 21.8% of Patriot users and 16.1% of Military users that exhibit the crossover linking behavior.

These observations show that more Patriot users crossover in cases of website URLs while a larger proportion of QAnon users crossover for YouTube URLs. This also gives clues to the key mediums in which these user groups use to disseminate their information across platforms.

6.3 Text similarity shows similar narratives

We investigated the narratives presented in the texts within the website and YouTube video URLs. We found that the text narratives in both website content and YouTube content are remarkably similar, despite the two user bases having shared URLs. The text similarity graphs are presented in Fig. 6. Both the website content and YouTube transcript similarity graphs show that only one core of nodes, linking outwards to other nodes, indicating a core set of narratives with evolved narratives.

To better understand the similarity structure present in the text, we then examined the nodes of the website content similarity graph (Fig. 6a). The most recurring URL domains

within Parler data are from “waynedupree,” “thescoopus” and “gatewaypundit.” The most recurring URL domains within Twitter data are from “theverge,” “yahooonews,” “theepochtimes” and “cnn.” Three key ideas surfaced within the content of these websites:

- (1) Voter fraud: This narrative tried to establish the presence of voter fraud in the 2020 US Elections due to mail-in ballot. It also provided a testimony of a team of data scientists claiming that more than 17,650 votes were changed against the Republican candidate Donald Trump.
- (2) Black Lives Matter and Antifa: This disinformation narrative advocates the idea that the police used weapons against Trump supporters to “protect the Antifa and Black Lives Matter militants” who were gathered at the Capitol.
- (3) Chinese intervention: This narrative argues that the Chinese are “actively engaged in various types of warfare against the United States” and cited examples of money and power scandals of the “Silk Road investigators” who fought back against the “offensive actions” of the Chinese.

We then examined the nodes of the YouTube transcript similarity graph (Fig. 6b). Among all the texts of the nodes shared, four key themes emerged: voter fraud, streaming of the elections, weapons used by police at the Capitol, which resulted in a death of an Air Force veteran, and that “democracy is at stake”; hence, the people should “rise to maintain our democracy.”

Overall, the content similarity graphs in Fig. 6 show that most communication via URLs is almost exclusively within the platforms. YouTube transcripts have much higher similarity than website content, which we posit could be of the nature that spoken language during the high-tempo event requires less thought than written words, leading to more common words used. We grouped the graph into Parler URLs only, Twitter URLs only and, Both URLs (i.e., URLs that are shared across both platforms). Investigating the connections between these three clusters, we observe that for both website and YouTube content similarity, the strongest connection is between the Both and the Parler clusters. Parler users share more unique content in the websites and YouTube videos. Coupled with the observations of Fig. 5, where Twitter users match in URLs more, we infer that Twitter users predominantly create links between communities by sharing URLs to disseminate information, but typically have lesser unique content.

Fig. 5 Similar link sharing networks. Pink nodes are Parler users, Blue nodes are Twitter users. The width of the links represents the strength of the link similarity, calculated by the number of links shared between both users

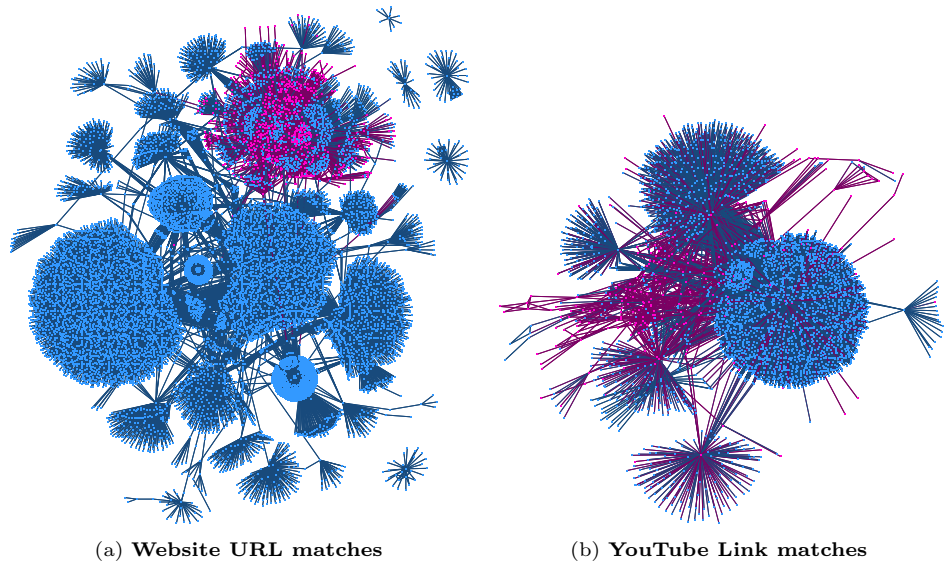


Fig. 6 Text Similarity. Pink nodes are Parler URLs, blue nodes are Twitter URLs and green nodes are URLs that occur within both Parler and Twitter. The width of the links represents the strength of the text similarity between the two websites. The nodes are sized according to the number of times the URLs are shared on the sites

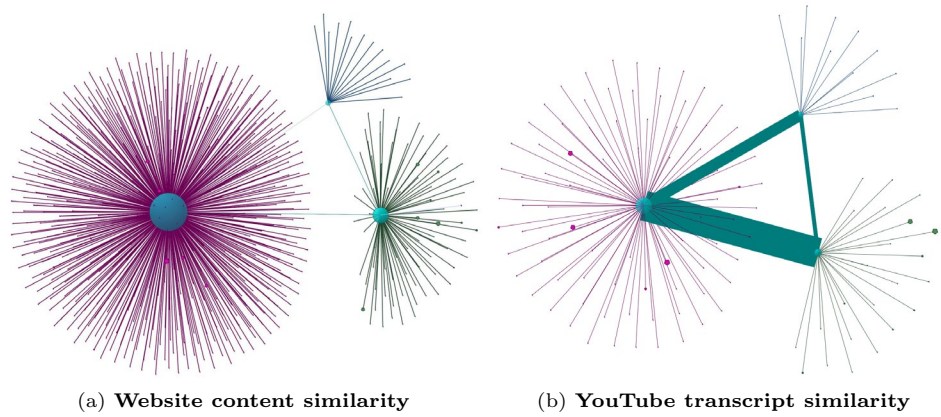
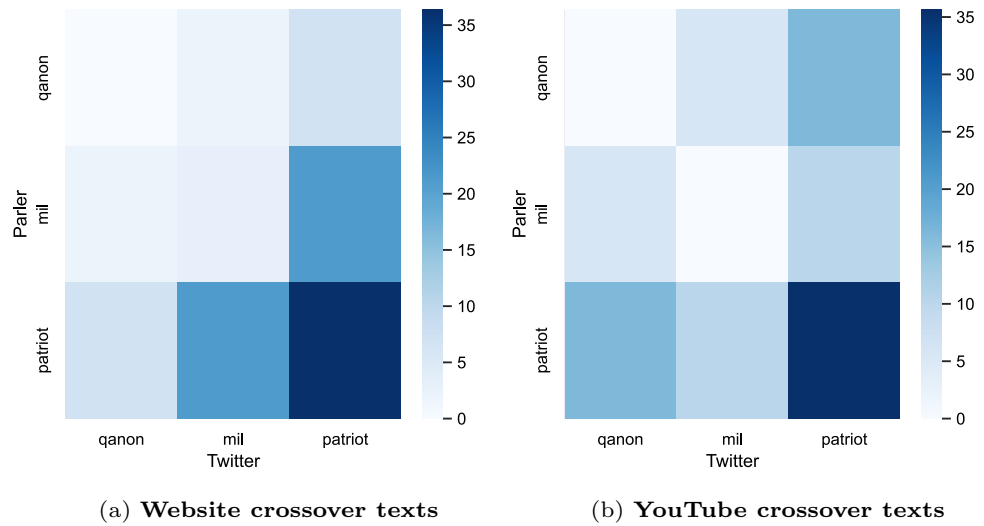


Fig. 7 Heatmap representing the percentage distribution of crossovers between Parler and Twitter platforms. These represent the users in each group that crossover both platforms as a percentage of the total number of users that crossover between both platforms



6.4 Crossover text similarity across affiliations

To better understand the nature of cross-platform information spread, we then analyzed the types of accounts between platforms that shared similar external content. More specifically, we binned the different types of users between both platforms by their labels (i.e., “Patriot,” QAnon, etc.) We counted the number of shared similar texts between the bins of users. Figure 7, displays the counts of shared similar texts between users of the different platforms. In both website and YouTube crossover texts, the largest proportion of users that have similar texts are the patriot users on both platforms. Military users tend to post more URLs with similar texts, leading to a higher website crossover percentage, while QAnon users post more links to similar YouTube videos.

7 Discussion

In this work, we analyzed coordinating information dissemination within and across Parler and Twitter during the January 6th Capitol Riot events. We also looked at the online activity of three main user groups associated with the January 6th Capitol Riots: users that present a military or veteran affiliation, users that use the user “patriot” as part of their user name and users that identify with QAnon-related terms. Our general observations from the username similarities reveal that users among both platforms express their endorsement for themes by including the theme as part of their usernames. Link matches of external websites and YouTube videos reveal separate information sources between both platforms. Despite the differences in shared external content between the platforms, the narratives disseminated within the separate information ecosystems echo similar themes based on those sources’ textual content.

User Affiliation Parler users were more willing to affiliate with a certain group than Twitter users. 8.57% of the Parler users identified themselves with one of the three, salient affiliations. However, it is not as common on Twitter. Many of the Twitter user name representations (user name, screen names, description) did not openly affiliate themselves with a single group and we were only able to find 2.76% of affiliated users. In total, 1.06 million users in our dataset had expressed affiliation with one of the three groups of interest. Of the users that expressed an affiliation, the most common affiliation expressed is toward “patriot,” in which 57.5% of users expressed that affiliation

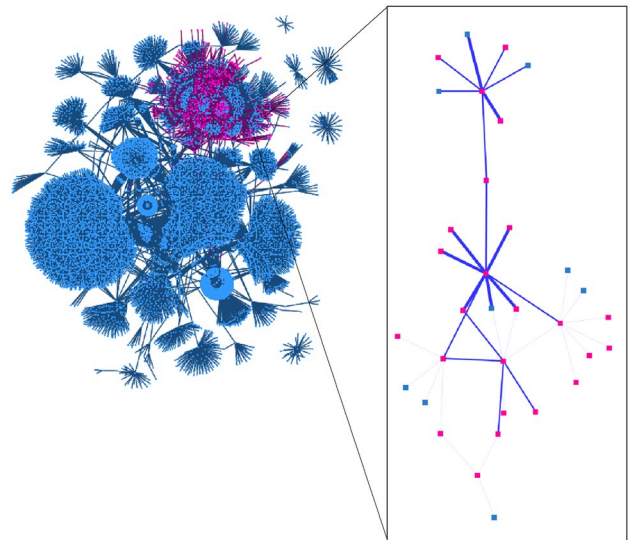


Fig. 8 A cluster of users identified through strong ties in Website URL matches and similar usernames. Pink nodes are Parler users, Blue nodes are Twitter users. The width of the links represents the number of website URL links shared between the users. The user names have been redacted to preserve user privacy

across both platforms. This is followed by an expressed affiliation to QAnon with 28.9% of users then military affiliation with 13.6% of users that expressed affiliation.

RQ1: Are there similar users or themes of users within and across the platforms?

In terms of discovering groups of identity clusters through username similarity, we observe that the network graphs connect users from both platforms, with no clear segregation of users of a single platform. This goes to show that users across both platforms jointly identify with particular communities and hence express it publicly in their usernames. Signaling theory shows that people declare their associations to recruit others or communicate perspective (Tajfel 1974). By putting their affiliation in their username handle, these users are sending a signal about their identity and drawing attention to their group. This result suggests that different communities of users have not fragmented across social media platforms and that there is a cross-platform presence of communities, like those that support QAnon conspiracy theories or identify with a military identity.

RQ2: What are the information dissemination patterns within and across platforms for YouTube videos and websites?

We observed that the two platforms appear to have dissimilar URLs and links shared among the users. Users mostly share the same links with other users in the same platform than across platforms, suggesting users from the two platforms have separate information sources. Only 9% of users share a link that was shared on both platforms with the website URL perspective and 11.1% of users share a YouTube link that was shared on both platforms. Only 2% of the YouTube video URLs and 1% of website URLs are shared among both platforms. So, the external content used in both platforms has little overlap, indicating different information consumption and dissemination habits between the platforms.

In the link sharing similarity graphs, users have weighted relationships and some links are stronger than others. Pairs of users with stronger links more frequently post common URL/YouTube videos across/between platforms. From a cohesion perspective, individuals who share the same information (via links) are more likely to be in the same group. If their usernames are also similar, this strengthens the cross-platform group relationship, suggesting the presence of groups of users that disseminate information across platforms.

After overlaying the Website URL and username similarity graph, we observe a group of users with high URL sharing activity between them, and all their names are either prefixed with “Patriot” or suffixed with “_republic” or “America.” This is represented in the network graph in Fig. 8. This result also suggests that using social media artifacts like external URLs or videos, combined with user expressions could be used as means of understanding cross-platform, online communities.

RQ3: Are there similar narratives that are shared across platforms, and what are their patterns of dissemination through YouTube videos and websites?

Despite the different information dissemination patterns, the content shared within both platforms from the text similarity graphs presents a different story: there is a central core of narratives perpetuating that is shared among the users of both platforms. This highlights that although both platforms have separate information ecosystems and consume information via different sets of websites and YouTube videos, the content they consume are similar. This may be platform-specific, in which information content producers tune their information to suit the tastes of the user base of the different social media platform.

The use of external links as a method of sharing and spreading information on social media has been much studied (Cruickshank et al. 2021; Kumpel et al. 2015). This study builds on the past work on social media link sharing behavior to examine the presence of coordinated narrative spread within and across platforms during the January 6th Capitol Riots. While an individual’s link sharing behavior

may be independent of each other as independent posts, but the synchrony of link sharing together with the other factors that we have examined—user affiliations, username similarity, and common narratives within the links—points to evidence of a picture of some degree of coordinated information spread between the two platforms.

In this study, we observe that users first make a deliberate act of signaling their affiliation and identity through terms embedded in their username handles, resulting in groups of user identities (e.g., “military,” “libertarian”). We also focus on the similarities and differences of the nature in information spread between the two social media environments. Despite being on different platforms, these users spread common narratives through external links and YouTube videos, suggesting the coordination of parallel information spread across both platforms. Lastly, we do not observe that users with different affiliations (i.e., “Patriot” and “QAnon” users) spread similar narratives, suggesting that the signaling of the group identity through username handles affects the narrative consumption.

Limitations and Future Work. First, our dataset contains only a partial scrape of the posts posted on the social media platforms during the event; the Parler dataset was an incomplete collection of posts by virtue of the rush of the researchers to archive the data before the platform shut down. Additionally, Twitter’s streaming API provides a 1% sampling of Tweets, so the collection of Tweets is also incomplete. While we believe the findings of this study generalize to the full collection of data, we cannot be sure without an analysis of all of the data being produced during the events under study.

Second, a significant proportion of URLs and YouTube videos had been removed from when the Capitol Riot events occurred to when we collected data from them. Thus, we were unable to obtain the content and make an assessment of all of the links being shared. This might have distorted our analysis to be biased toward the content that is still available.

Determining the core cluster size through thresholding by (mean + standard deviation) can at times be an arbitrary threshold and is dependent on the data collected. Future work calls for more stable thresholding algorithms that are invariant to the density of the network graphs.

This work studies Parleys and Tweets within a one-week time window of the Capitol Riots event. This study is largely opportunistic and observational. While there are some temporal aspects to the dynamics of cross-platform information spread, we do not really have the data to support temporal analysis. There are exact time stamps for Twitter data but because Parler data is obtained from a data dump does not have specific time stamps except “4h ago”/ “3h ago,” by which we are unsure if the reference points are the same. Additionally, the short timeframe where the data burst occurs means there is not a natural way to break up the data temporally. Future work should further develop this

Table 3 Terms used to heuristically define user affiliation within Parler and Twitter data

User category	Terms used to define that user category
Military/Veteran	Army, navy, air force, airforce, marine, veteran, military, servicemember, coastguard, coast guard, soldier, infantry, sergeant
Patriot	Patriot
QAnon	Qanon, wwg1wga, Q, thegreatawakening, thestorm, theplan

method for larger-scale and longer events to include temporal changes and temporal dependency of posts in analyzing the dynamics of information spread across platforms.

Despite these limitations, we hope our work provides a methodology for characterizing content spread within and across platforms, and a had provided a glimpse into the cross-platform information spread between Parler and Twitter. This work is, to the best of our knowledge this work is, the first attempt to link the two platforms Parler and platforms Twitter through investigating information spread via coordinated narratives.

8 Conclusion

In this study, we studied the coordinated information dissemination across Parler and Twitter through three dimensions: username similarity, link matching similarity of external websites, and YouTube videos and textual similarity of the links identified. We discovered that users across platforms jointly endorse similar themes in their usernames. The users share very little common links across platforms but huge amounts of common links among users of the same platforms. This suggests they consume their information through separate ecosystems, i.e., share links from different groups of websites. If the users largely consume news/information through these platforms, it would mean that the two groups of users are reading different sets of websites. Strikingly, the narrative content of these links and YouTube videos are extremely similar across platforms, perpetuating a few key narratives, suggesting coordinated information dissemination across platforms, yet tailored to the tastes of each specific platform. The early detection of coordinated narratives spreading across platforms in an effort to organize movements may present the possibility of stopping offline violence in the real world. We hope the techniques here can be used to better analyze coordinated information dissemination within and across platforms and streamline network structures into core components for further investigation.

Appendix 1 Determining user affiliation

Acknowledgements The research for this paper was supported in part by the Knight Foundation and the Office of Naval Research Grant (N000141812106) by the center for Informed Democracy and Social-cybersecurity (IDeaS) and the center for Computational Analysis of Social and Organizational Systems (CASOS) at Carnegie Mellon University. The views and conclusions are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Knight Foundation, Office of Naval Research or the US Government.

Author Contributions All authors contributed to the conception of the study. IC: performed the data collection, while LN: performed the analysis. All authors contributed to writing of the first draft. All authors read and approved the final manuscript.

Funding The research for this paper was supported in part by the Knight Foundation and the Office of Naval Research Grant (N000141812106) by the center for Informed Democracy and Social-cybersecurity (IDeaS) and the center for Computational Analysis of Social and Organizational Systems (CASOS) at Carnegie Mellon University.

Data availability Data can be made available upon reasonable requests in accordance to the Terms of Conditions of Twitter and Parler. Please contact the authors for data requests.

Code availability Code will be made available upon publication of this paper at <https://github.com/quarby/coordinating-narratives-information-spread>.

Declarations

Conflict of interest All authors declare that they have no conflict of interest.

Ethical approval We did not conduct any experimental research involving humans or animals. All data used in this study was open and freely available online and its use is in accordance with university ethical standards on using Internet-available data. All procedures performed in studies involving social media data were in accordance with the ethical standards of the university.

References

- Alperin JP, Gomez CJ, Haustein S (2019) Identifying diffusion patterns of research articles on twitter: a case study of online engagement with open access articles. *Public Underst Sci* 28(1):2–18
- Benn J (2017) End the duopoly: America needs a major political realignment. *HuffPost* https://www.huffpost.com/entry/end-the-duopoly-america-n_b_9463274
- Blondel VD, Guillaume J-L, Lambiotte R, Lefebvre E (2008) Fast unfolding of communities in large networks. *J Stat Mech: Theory Exp* 2008(10):10008
- Carley KM (2020) Social cybersecurity: an emerging science. *Comput Math Organ Theory* 26(4):365–381
- Center PR (2021) Social media fact sheet. <https://www.pewresearch.org/internet/fact-sheet/social-media/>
- Choudhury N, Ng KW, Iamnitchi A (2020) Strategic information operation in youtube: the case of the white helmets. In: Thomson R, Bisgin H, Dancy C, Hyder A, Hussain M (eds) *Social, cultural, and behavioral modeling*. Springer, Cham, pp 318–328
- Cooks EJ, Bolland A (2021) Hostile media perception on twitter: the effect of mediated social identity cues on biased perception. *J Broadcast Electron Media* 65(4):435–456
- Correa D, Sureka A, Sethi R (2012) Whacky!-what anyone could know about you from twitter. In: 2012 10th annual international conference on privacy, security and trust, pp 43–50 IEEE
- Cruickshank I, Ginossar T, Sulskis J, Zheleva E, Berger-Wolf T (2021) Content and dynamics of websites shared over vaccine-related tweets in covid-19 conversations: a computational analysis. *J Med Internet Res* 23(12):e29127
- Cruickshank IJ, Carley KM (2020) Clustering analysis of website usage on twitter during the covid-19 pandemic. In: annual international conference on information management and big data, Springer, pp. 384–399
- Datta S, Adar E (2019) Extracting inter-community conflicts in reddit. In: Proceedings of the international AAAI conference on web and social media 13:146–157
- Flores-Saviaga CI, Keegan BC, Savage S (2018) Mobilizing the trump train: understanding collective action in a political trolling community. In: Twelfth international AAAI conference on web and social media
- Frenkel S (2021) The storming of Capitol Hill was organized on social media. *The New York Times* <https://www.nytimes.com/2021/01/06/us/politics/protesters-storm-capitol-hill-building.html>
- Giglietto F, Righetti N, Rossi L, Marino G (2020) Coordinated link sharing behavior as a signal to surface sources of problematic information on facebook. In: international conference on social media and society. *SMSociety'20*, pp 85–91. Association for Computing Machinery, New York, NY, USA <https://doi.org/10.1145/3400806.3400817>
- Gionis A, Indyk P, Motwani R (1999) Similarity search in high dimensions via hashing. In: Proceedings of the 25th international conference on very large data bases. *VLDB '99*, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, pp. 518–529
- Golovchenko Y, Buntain C, Eady G, Brown MA, Tucker JA (2020) Cross-platform state propaganda: Russian trolls on twitter and youtube during the 2016 us presidential election. *Int J Press/Polit* 25(3):357–389
- Horawalavithana S, De Silva R, Nabeel M, Elvitigala C, Wijesekara P, Iamnitchi A (2021) Malicious and low credibility urls on twitter during the astrazeneca covid-19 vaccine development. In: Thomson R, Hussain MN, Dancy C, Pyke A (eds) *Social, cultural, and behavioral modeling*. Springer, Cham, pp 3–12
- Hounsel A, Holland J, Kaiser B, Borgolte K, Feamster N, Mayer J (2020) Identifying disinformation websites using infrastructure features. In: 10th {USENIX} workshop on free and open communications on the internet (FOCI 20)
- Hussain MN, Tokdemir S, Agarwal N, Al-Khateeb S (2018) Analyzing disinformation and crowd manipulation tactics on youtube. In: 2018 IEEE/ACM international conference on advances in social networks analysis and mining (ASONAM), pp 1092–1095 IEEE
- Iamnitchi A, Ng K, Horawalavithana S (2020) Twitter is the megaphone of cross-platform messaging on the white helmets. In: Proceedings of the 13th international conference in social, cultural, and behavioral modeling <https://www.springerprofessional.de/en/twitter-is-the-megaphone-of-cross-platform-messaging-on-the-whit/18470618>
- Index GW (2021) Social media marketing trends in. <https://www.gwi.com/reports/social>
- Johnson J, Douze M, Jégou H (2017) Billion-scale similarity search with gpus. arXiv preprint [arXiv:1702.08734](https://arxiv.org/abs/1702.08734)
- Kin Wai Ng AI, Sameera Horawalavithana (2021) Multi-platform information operations: Twitter, facebook and youtube against the white helmets. In: Proceedings of the 15th international AAAI conference on web and social media <https://doi.org/10.36190/2021.36>. http://workshop-proceedings.icwsm.org/abstract?id=2021_36
- Kin Wai N, Horawalavithana S, Iamnitchi A (2021) Multi-platform information operations: twitter, facebook and youtube against the white helmets
- Klausen J, Barbieri ET, Reichlin-Melnick A, Zelin AY (2012) The youtube jihadists: a social network analysis of al-muhajiroun's propaganda campaign. *Perspect Terror* 6(1):36–53
- Kumar S, Hamilton WL, Leskovec J, Jurafsky D (2018) Community interaction and conflict on the web. In: Proceedings of the 2018 world wide web conference, pp 933–943
- Kuzma R, Cruickshank IJ, Carley KM (2021) Analysis of external content in the vaccination discussion on twitter. *CoRR abs/2107.09183* 2107.09183
- Kümpel AS, Karnowski V, Keyling T (2015) News sharing in social media: a review of current research on news sharing users, content, and networks. *Social Media Society* 1(2):2056305115610141
- Lakkaraju H, McAuley J, Leskovec J (2013) What's in a name? Understanding the interplay between titles, content, and communities in social media. In: Proceedings of the international AAAI conference on web and social media, 7, pp 311–320
- Levenshtein VI *et al.* (1966) Binary codes capable of correcting deletions, insertions, and reversals. In: *Soviet Physics Doklady*, vol. 10, pp 707–710 Soviet Union
- Lyons K (2021) Parler posts, some with GPS data, have been archived by an independent researcher. *The Verge* <https://www.theverge.com/2021/1/11/22224689/parler-posts-gps-data-archived-independent-researchers-amazon-apple-google-capitol>
- Magelinski T, Ng L, Carley K (2022) A synchronized action framework for detection of coordination on social media. *J Online Trust Saf.* <https://doi.org/10.54501/jots.v1i2.30>
- Munn L (2021) More than a mob: parler as preparatory media for the u.s. capitol storming. *First Monday.* <https://doi.org/10.5210/firstmonday.11574>
- NPR Staff (2021) The capitol siege: the arrested and their stories. <https://www.npr.org/2021/02/09/965472049/the-capitol-siege-the-arrested-and-their-stories> Accessed April 30, 2021
- Ng LHX, Loke JY (2021) Analyzing public opinion and misinformation in a covid-19 telegram group chat. *IEEE Internet Comput* 25(2):84–91. <https://doi.org/10.1109/MIC.2020.3040516>
- Ng LHX, Cruickshank I, Carley KM (2021) Coordinating narratives and the capitol riots on parler
- Ng LH, Taihagh (2021) A How does fake news spread? understanding pathways of disinformation spread through apis. *Policy Internet*

- Ng LHX, Carley KM (2022) Online coordination: methods and comparative case studies of coordinated groups across four events in the united states. In: 14th ACM Web Science Conference 2022. WebSci '22, pp 12–21. Association for Computing Machinery, New York, NY, USA . <https://doi.org/10.1145/3501247.3531542>
- Nizzoli L, Tardelli S, Avvenuti M, Cresci S, Tesconi M (2020) Coordinated behavior on social media in 2019 uk general election. arXiv preprint [arXiv:2008.08370](https://arxiv.org/abs/2008.08370)
- Pacheco D, Hui P-M, Torres-Lugo C, Truong BT, Flammini A, Menczer F (2021) Uncovering coordinated networks on social media: methods and case studies. Proc Int AAAI Conf Web Soc Media 15(1):455–466
- Pacheco D, Flammini A, Menczer F (2020) Unveiling coordinated groups behind white helmets disinformation. In: companion proceedings of the web conference 2020. WWW '20, pp. 611–616. Association for Computing Machinery, New York, NY, USA <https://doi.org/10.1145/3366424.3385775>
- Pacheco D, Hui P, Torres-Lugo C, Truong BT, Flammini A, Menczer F (2020) Uncovering coordinated networks on social media. CoRR [abs/2001.05658](https://arxiv.org/abs/2001.05658)[arXiv:2001.05658](https://arxiv.org/abs/2001.05658)
- Pennington J, Socher R, Manning CD (2014) Glove: global vectors for word representation. In: empirical methods in natural language processing (EMNLP), pp 1532–1543 <http://www.aclweb.org/anthology/D14-1162>
- Pierri F, Artoni A, Ceri S (2020) Investigating italian disinformation spreading on twitter in the context of 2019 european elections. PLoS ONE 15(1):1–23. <https://doi.org/10.1371/journal.pone.0227821>
- Squire M (2021) Monetizing propaganda: how far-right extremists earn money by video streaming. In: 13th ACM web science conference 2021, pp 158–167
- Sugawara K, Kobayashi H, Iwasaki M (2016) On approximately searching for similar word embeddings. In: Proceedings of the 54th annual meeting of the association for computational linguistics (Vol.1: Long Papers), pp 2265–2275
- Swire-Thompson B, Lazer D (2020) Public health and online misinformation: challenges and recommendations. Annu Rev Public Health 41:433–451. <https://doi.org/10.1146/annurev-publhealth-040119-094127>
- Tajfel H (1974) Social identity and intergroup behaviour. Soc Sci Inf 13(2):65–93
- Tajfel H, Billig MG, Bundy RP, Flament C (1971) Social categorization and intergroup behaviour. Eur J Soc Psychol 1(2):149–178
- The Gateway Pundit (2021) <https://mediabiasfactcheck.com/the-gateway-pundit/>
- Timberg C, Dwoskin E, Albergotti R (2021) Inside facebook, Jan. 6 violence fueled anger, regret over missed warning signs. WP Company <https://www.washingtonpost.com/technology/2021/10/22/jan-6-capitol-riot-facebook/>
- Wang A, Hamilton WL, Leskovec J (2016) Learning linguistic descriptors of user roles in online communities. In: Proceedings of the first workshop on nlp and computational social science, pp 76–85

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.