Big Data Research 37 (2024) 100481

Contents lists available at ScienceDirect

Big Data Research

journal homepage: www.elsevier.com/locate/bdr

Unmasking hate in the pandemic: A cross-platform study of the COVID-19 infodemic

Fatima Zahrah [a,*], Jason R.C. Nurse [b], Michael Goldsmith [c]

[a] School of Computer Science, University of Bristol, Bristol, United Kingdom
[b] School of Computing, University of Kent, Canterbury, United Kingdom
[c] Department of Computer Science, University of Oxford, Oxford, United Kingdom

ARTICLE INFO

Keywords:
Social media analysis
Cross-platform analysis
Online hate
COVID-19

ABSTRACT

The past few decades have established how digital technologies and platforms have provided an effective medium for spreading hateful content, which has been linked to several catastrophic consequences. Recent academic studies have also highlighted how online hate is a phenomenon that strategically makes use of multiple online platforms. In this article, we seek to advance the current research landscape by harnessing a cross-platform approach to computationally analyse content relating to the 2020 COVID-19 pandemic. More specifically, we analyse content on hate-specific environments from Twitter, Reddit, 4chan and Stormfront. Our findings show how content and posting activity can change across platforms, and how the psychological components of online content can differ depending on the platform being used. Through this, we provide unique insight into the cross-platform behaviours of online hate. We further define several avenues for future research within this field so as to gain a more comprehensive understanding of the global hate ecosystem.

1. Introduction

The ever-increasing role of social media for communication has demonstrated how it has become an integral asset within the home and workplace alike, which has been especially emphasised during the recent COVID-19 pandemic. However, beyond providing a variety of affordances, such technologies have also introduced several implications, such as providing new vehicles for spreading hateful content. Online hate has been linked to several abhorrent real-world events, including the recruitment of extremists [1] as well as inciting mass shootings, stabbings and bombings [2]. The UK government has thus specifically outlined hateful content as one of the primary forms of illegal content online in their Online Harms Paper [3]. Additionally, online hate has emerged as a tool for politically motivated bigotry, xenophobia, homophobia, religious discrimination, and excessive nationalism [4,5]. One instance in which this was exhibited was during the 2016 US presidential election; the narrative of the “Make America Great Again” (MAGA) campaign slogan provided new possibilities for radical nationalist groups to distribute their content more easily and communicate with their audiences at a much larger scale [6,7].

The concept of online hate is still considered a complex phenomenon with an ever-evolving definition, thus, research into online hate is fragmented across numerous disciplines. Given the substantial utilisation of social media platforms as communication channels, there is a growing necessity to employ big data analytics and techniques to extract meaningful insights from hateful content found online. Although various extensive approaches have been proposed within the research landscape to analyse online hate, few studies have investigated how hateful behaviours and content compare across different online platforms [8]. Recent academic discourse has highlighted the interconnected nature of online hate, revealing that it extends beyond individual platforms to form intricate networks spanning multiple platforms, or a global ‘network of networks’ [2]. Despite the ample evidence indicating the strategic utilisation of various platforms by hate groups, research into this phenomenon remains limited.

This paper will aim to build on this particular line of research by harnessing a cross-platform approach to analysis, so as to gain a clearer understanding of the dynamics of the global hate ecosystem. In particular, this research will make use of data collected over the course of the COVID-19 pandemic from four different social-media platforms – Twitter, Reddit, 4chan and Stormfront – to investigate how hateful

* Corresponding author.
E-mail address: fatima.zahrah@bristol.ac.uk (F. Zahrah).

https://doi.org/10.1016/j.bdr.2024.100481
Received 8 January 2024; Received in revised form 9 April 2024; Accepted 12 June 2024
Available online 25 June 2024

content and narratives compare across multiple platforms. More specifically, this study will harness various computational methods and big data techniques, including topic modelling, and linguistic and sentiment analysis, to explore the type of content that is promoted on each platform. Through this, we aim to gain an understanding on how online platforms are used for the different functionalities they offer, and how specific platforms can play a different role within the greater hate ecosystem. We believe that our findings provide novel insight into the cross-platform behaviours of online hate.

The contributions of our work are as follows:

- We collect data from four different online platforms (Twitter, Reddit, 4chan and Stormfront) over a period of 15 months and examine how posting behaviour changes over the course of the COVID-19 pandemic.
- We conduct topic modelling to show how different types of content and narratives are promoted on each platform.
- We perform a deeper study into the linguistic composition of the posts from each platform, and identify distinctions in the type of sentiment and level of emotion used.

The remainder of the paper will be structured as follows. Section 2 will review some of the current literature on online hate. Section 3 will provide a detailed account of our approach and methodology, including the datasets and data-analysis tools used. The results and observations from our findings will be discussed in Section 4. We then present our conclusions and outline avenues for future work in Section 5.

2. Related work

As mentioned previously, online hate is a complex phenomenon which has been subject to increasing scrutiny from academics across various disciplines, including social science and computer science. As a result, research contributions in this field have been fragmented and particularly varied. Recent work on online hate has focused on the targets of hate [9], characterisations of hateful users [10], as well as social-network analysis of online hate networks [2]. Perhaps the largest body of research on online hate in the past decade has been on different approaches for its automatic detection, where a plethora of different machine-learning techniques have been utilised and adapted, including Support Vector Machines (SVM) [11,12], Random Forests [13–15], logistic regression [16–18], Naive Bayes [19,20] and, more recently, deep learning [21,22] with varying levels of accuracy.

Additionally, a number of studies have argued that the various functionalities of social-media platforms also facilitate the development of hateful communities, and thus should also be explored to analyse networks of hate. One such study that used this methodology is conducted by Eddington, where the "Make America Great Again" campaign slogan first used by Donald Trump in the 2016 US presidential elections is investigated on Twitter via the #MAGA hashtag [4]. Though this hashtag was initially introduced to allow social-media users to connect with his campaign, the presence of hate groups connecting with the hashtag also became increasingly apparent. Exploring the networks of this hashtag showed how it was used as a communicative organising site for white-supremacist groups and illuminated the overtly far-right content and hashtag conversation spaces shared and embedded within #MakeAmericaGreatAgain [4]. This demonstrates the importance of establishing the various ways in which networks of hateful content can be formed and organised on social media to gain a more complete idea of the structures of these communities.

Despite all the extensive approaches proposed to analyse online hate within the research landscape, limited studies have investigated how hateful behaviours and content compare across different online platforms [23,24]. As mentioned earlier, academic literature has only recently recognised that online hate branches across several platforms [25]. These networks formed by hate groups have proven to be remarkably resilient, and have increasingly shown to migrate across various platforms and other networks, maintaining and often expanding their connections in the process [2].

One of the first studies to explore how various web communities influence each other was carried out by Zannettou et al. [26]. In that study, the authors investigated how mainstream and alternative news propagate across multiple online communities, whilst measuring the influence that each community has on each other. Using a statistical model, they highlighted that small "fringe" online communities within Reddit and 4chan can have a substantial impact on large mainstream online communities like Twitter. The authors demonstrate how such online platforms are clearly not independent; while they do exhibit different behaviours and internal influence, they are also affected by each other, as well as by the greater Web [26].

Although, research within this aspect of online hate is scarce, in the last year, a few studies have realised the importance of the insights that can be gained from cross-platform analysis. With this motivation, Phadke and Chandaluri conducted a preliminary study where they collected data from the Twitter and Facebook accounts of various hate groups, and explored how content is framed and shared across both platforms [8]. Through this, the authors highlight some differences in the way both platforms were used by hate groups, where Facebook seemed to be used for group radicalisation and recruitment, and Twitter was mainly used to reach a diverse follower base.

More recently, Hitkul et al. [27] conducted a comparative study of Twitter and Parler content during the aftermath of the 2021 Capitol riots. Though this study was not focused on hateful content, it still provided some insight into how sentiment and narratives can differ across platforms. Similarly, Murdock et al. conducted a multi-platform study of fraud and protest-related posts on Twitter, Facebook and Reddit during the aftermath of the 2020 US election [28]. Our study builds on such findings by exploring both mainstream platforms, such as Twitter and Reddit, as well as non-moderated fringe platforms, like 4chan and Stormfront. A further difference of our work is that we only collect data from hate-specific environments, more details of which are given in Section 3.1.

The findings detailed in our paper will therefore aim to further fill the gap currently within this research landscape by providing more extensive empirical and statistical insight into the cross-platform behaviours of online hate on both mainstream and fringe communities, within the context of the COVID-19 pandemic. We provide understanding into the type of content promoted on each platform and the linguistic composition of their posts, as well as some initial observations on any cross-platform activity.

3. Methodology

Our cross-platform analysis of online hate during the COVID-19 pandemic on Twitter, Reddit, 4chan and Stormfront was largely focused on content from white-supremacist ideologies. We focus on online hate from white-supremacist groups and users as previous studies have highlighted how far-right and white supremacist ideologies were used to propagate hate against various minority groups during the COVID-19 pandemic [29,30]. More specifically, this study was carried out with particular regard to the following research questions:

- **RQ1:** How do the participation and posting trends compare across all four platforms over the course of the pandemic?
- **RQ2:** What are the main topics of discussion for hateful users on each of the platforms?
- **RQ3:** Are there any differences in the linguistic composition or general sentiment of the posts from the four platforms?

Our approach therefore comprises three stages: (1) collecting the cross-platform data from all four platforms and observing the posting behaviours, (2) conducting topic modelling on each corpus of posts from the four platforms, and (3) carrying out a more in-depth linguistic analysis of the collected posts to examine their structural properties. Further details on the methods that were used at each stage are provided below.

3.1. Data collection

With the aim of capturing a comprehensive overview of the online conversations across all four platforms during the COVID-19 pandemic, data was collected from January 2020 to March 2021. A majority of all the significant events that had a profound impact on the online discourse related to the pandemic were included within this time frame. More specifically, COVID-19 was declared as a pandemic by the World Health Organisation in January 2020 [31], and was still very much ongoing as of March 2021, where the roll-out of the COVID-19 vaccines was well underway, though some lockdowns and restrictions were being lifted [32]. This time frame for data collection therefore captures the majority of the duration of the pandemic, including its early stages and its ongoing impact. To ensure that only content related to COVID-19 was collected, a list of key terms related to the pandemic and white-supremacist movements, which have also been used in previous studies into online hate during COVID-19 [33,34], was used to filter content during data collection. The terms used in this case are: "covid", "corona", "pandemic", "virus", "lockdown" and "mask". All four of the datasets were filtered using these terms during the collection process.

We chose to analyse Twitter, Reddit, 4chan, and Stormfront within our study as they each offer distinct types of social-media platforms. Twitter is the largest and most mainstream platform of the four, which offers the largest and most diverse online audience. It is also the platform that moderates content the most, and therefore explicit content is often removed fairly quickly. Reddit is another mainstream platform, though with a smaller audience size than Twitter. Content moderation is also carried out by Reddit, although not as much as Twitter, where hateful communities are often cultivated on particular subreddits, but such subreddits have also been removed from the platform if they are increasingly linked to hateful events, such as the subreddits r/fatpeoplehate and r/CoonTown [35].

Both 4chan and Stormfront, on the other hand, represent fringe communities with more specific audiences. 4chan is an anonymous imageboard platform with no content moderation, which has been linked to several offline crimes and extremist attacks [2]. Stormfront is distinct from these platforms in that it prides itself in being "the first White Nationalist forum on the Web", where the platform actively tries to amplify white-supremacist voices and opinions. It has also been linked to several violent acts of extremism, including the mass killing of 77 people in Norway in 2011 [23]. Each of these four online platforms therefore provides a distinct set of functionalities and audiences. We aim to understand the extent to which these platforms can play individual roles and serve different purposes within the wider ecosystem of online hate.

Our Twitter dataset was collected using the official Twitter API.[1] In order to adhere to collection limits, we made use of further filtering using the COVID-19-related keywords listed above. Since Twitter is a more moderated platform in general, it may be argued that only a small percentage of content will be identified as hateful (or the period in which hateful content is present on Twitter will be less). To ensure only hateful content from predominantly white-supremacist users were selected, we collected tweets from the accounts of white-supremacist groups and their supporters. These groups were identified from a list of hate groups published by the Southern Poverty Law Center (SPLC).[2]

This list contains approximately 300 hate organisations from various ideologies, of which 84 are groups supporting white supremacy; it should be noted here that SPLC identifies these groups as white nationalist, neo-Nazi and neo-confederate, but we combined these groups since they have shared views on extreme-right ideology and reported hatred for other races [8]. From these 84 hate groups, we found the associated Twitter accounts of 48 groups. We then used a two-step snowball-sampling approach to identify other hateful groups and users from the follower and followee lists of these accounts. Through this, we identified 478 hateful Twitter accounts, from which we collected tweets relating to COVID-19 over the course of the collection period.

In order to collect the relevant data from both Reddit and 4chan, we made use of the 4CAT Capture and Analysis Toolkit [36]. The Reddit posts were collected from the r/donaldtrump subreddit, which was linked to spreading online hate during 2020, and was consequently banned by Reddit in the aftermath of the 2021 Capitol riots [37]. Similarly, we collected 4chan posts from the Politically Incorrect (/pol/) board, which has also been identified as a key platform for spreading online hate, and has been linked to several violent acts of extremism including the 2019 Christchurch shooting [2].

For our Stormfront dataset, we made use of the "ExtremeBB" dataset provided by the Cambridge Cybercrime Centre,[3] which is a comprehensive collection of data from various extreme forums online. We only made use of the collection of Stormfront posts from this dataset, which we further filtered to include only content posted over the course of our collection period. After collecting all our datasets, we measured the participation trends by observing the frequency of posts being published online, in answer to RQ1. The results from this will be discussed in Section 4.1.

The sizes of the four datasets are as follows:

1. Twitter COVID-19 dataset: 1,361,580 posts
2. Reddit COVID-19 dataset: 46,977 posts
3. 4chan COVID-19 dataset: 845,982 posts
4. Stormfront COVID-19 dataset: 12,281 posts

3.2. Computational analyses

3.2.1. Identifying topics of discussion

In order to identify the main topics of discussion during the course of the COVID-19 pandemic in answer to RQ2, we conducted topic modelling on each of our datasets. We found that using the Latent Dirichlet Allocation (LDA) topic detection model worked better overall on all four datasets than other models, like Non-Negative Matrix Factorisation (NMF), even though NMF usually works better with shorter texts [38]. LDA topic modelling has been used to identify topics within social media posts in many previous studies [39,40], and works under the assumption that a document is comprised of a collection of latent topics [41]. The model uses probabilistic assignments of terms to a user specified number of topics. From this, each unique term in the corpus is assigned a probability distribution relative to the number of topics, indicating for each topic the probability that the term occurs within it, thus providing a distribution of topics over documents.

As LDA topic modelling requires a user specified number of topics, we experimented with our topic model with different numbers of topics across each of our datasets [42]. From this experimentation, we found that the number of topics that produced the most distinct topics in all our datasets was 5, thus this is the final number of topics we identify in all four datasets in our topic analysis. We further explore these findings by assessing the extent to which each topic is discussed in every dataset, where we find the dominant topic in each post and then extract the proportion of posts containing reference to that topics.

A series of pre-processing steps were carried out before the linguistic analysis to clean the posts in each dataset and prepare them for further linguistic analysis (we did not pre-process the datasets when observing the frequency of posting over the course of the pandemic). These steps

[1] https://developer.twitter.com/en/docs/twitter-api.
[2] https://www.splcenter.org/hate-map.
[3] https://www.cambridgecybercrime.uk/datasets.html.

included: (1) Removing any duplicate posts from the datasets to reduce the levels of noise. (2) Removing all punctuation marks. (3) Removing any URLs. (4) Removing any short posts (those less than 5 tokens). (5) Removing any platform-specific noise, for instance 'RT' for the Twitter dataset. We then tokenized all of the posts and created a term-frequency inverse-document frequency (TF-IDF) array to fit the LDA model, which has been suggested by previous work to yield more accurate topics [43]. This analysis is carried out using the Pandas[4] data-analysis library and the 'Natural Language Toolkit' (NLTK)[5] provided by the Python programming language, where the LDA topic modelling was conducted with a Gibbs sampler using the Python Gensim wrapper.

3.2.2. Analysing sentiment and linguistic composition

In order to answer RQ3 and further linguistically analyse our four datasets, we used the programmatically coded dictionary from the Linguistic Inquiry and Word Count (LIWC 2015) [44] analysis tool to automate the process of extracting further information on linguistic structures and psychological meaning from textual content. LIWC is a widely used tool in lexical approaches for personality measurement, and statistically analyses textual content based on 81 different categories by calculating the percentage of words in the input text that match predefined words in a given category [44]. Many previous studies from various disciplines have utilised LIWC to gain a more in-depth understanding of the structural and functional constructs used within language [45,46], as well as to get insight into the psychological meaning of textual content [47].

LIWC is used in our approach to both examine and compare the functional composition of the posts collected from each platform, as well as to extract psychological meaning and sentiment from the datasets. To do this, we analysed each dataset of posts with all 81 LIWC categories. This analysis focuses particularly on the four summary linguistic variables ('analytical thinking', 'clout', 'authenticity', and 'emotional tone'), and 10 more detailed variables that reflect the psychological states, linguistic dimensions, personal concerns, and informal language within each dataset. More specifically, this involves the usage of pronouns ('i', 'we', 'you', 'they'), which have often been identified as a discursive tool used to persuade audiences, as well as emotive language, which used the LIWC categories 'positive emotion', 'negative emotion', 'anger' and 'anxiety'. This is used in our study to identify the types of narratives that are promoted on each platform, as well as to gain further insight into the target audience that each platform addresses.

3.3. Ethics

In this work, we focus on the overall online behaviours of individuals and groups driven by hateful ideologies. All of the data we use in our study is publicly available, including the list of hate groups published by the SPLC, and can be extracted without having to create an account with the platforms. Additionally, we do not include any account handles or organisation names within our study, and do not quote any posts within our paper that could be used to identify accounts and therefore potentially people. Instead we post aggregate findings from our analysis of the posts. The Stormfront data provided by the Cambridge Cybercrime Centre was used in line with a data sharing agreement, where the main purpose of collecting and using this data is to find, understand, investigate and counter political extremism. No data from this dataset is published in our paper and, again, only findings from our analysis are included here.

[4] https://pandas.pydata.org/.

[5] https://www.nltk.org/.

4. Results and discussion

4.1. Participation trends

To address RQ1, the analysis framework first examines the frequency of content posted during the COVID-19 pandemic across the collection period in each of the four datasets. Fig. 1 illustrates that Twitter has significantly more content than the other platforms, with 4chan also having a considerably high volume of posts related to the pandemic. In contrast, it is clear that content is posted much less frequently on Stormfront.

To show how the amount of participation compares across all four platforms over the course of the collection time frame, standard score normalisation [48] is used to create a graph of all the normalised data from each dataset. Here, posting frequency is measured on a monthly basis to assess how real-time developments during the pandemic affected online discourse. Overall, similar trends can be seen in the amount of participation on each platform over the collection period during COVID-19. The most steep peak can be seen on all four platforms around March 2020. This marked when COVID-19 was first declared as a pandemic [31], as well as the introduction of severe restrictions and regulations throughout much of the world, such as lockdowns and travel bans.

The Twitter, Reddit and 4chan datasets show a significant drop in posting activity shortly after this period, but also start to peak again around November 2020 and January 2021. This could indicate the time when lockdowns were once again enforced across many parts of the world, particularly during the Thanksgiving and Christmas holiday season, or when the Omicron variant of COVID-19 emerged, causing infections and death rates to reach unprecedented levels. These peaks could also be attributed to the 2020 US election held in November, where the COVID-19 response and policies continued to be an important subject in political debates and campaign agendas.

The r/donaldtrump subreddit was banned due to its involvement during the January 2021 Capitol riots. As a result, there is a sudden halt in Reddit posts during the final months of the data collection period. Although there is a peak in the frequency of posts during the initial stages of the pandemic, posting behaviour on Stormfront is mostly steady. This suggests that significant offline events do not have a considerable influence on the online behaviour of users on this particular platform. Instead users remain mostly consistent with their posting behaviour.

Therefore, in answer to RQ1, the participation trends across all four platforms seem to be similar, in that they generally peak in posting frequency at around the same time periods following key real-time developments over the course of the COVID-19 pandemic. However, such events seem to control the discourse on platforms with larger audience sizes like Twitter, Reddit and 4chan, more than it does on smaller, underground platforms like Stormfront, where posting frequency is generally more consistent despite offline events.

4.2. Topic analysis

To gain further insight into the key subjects of discussion within each dataset, a topic model, using the LDA topic-detection model, is applied to all four datasets, the findings from which are used to address RQ2. The five identified topics and the percentage of posts containing them are listed in Table 1; this percentage is derived by extracting the dominant topic in each post, and the total sum of the number of posts for each topic was then represented as a percentage of the total posts in the dataset. The topic model shows that the most common topic in COVID-19-related posts from all four platforms is directing a majority of the blame for the outbreak towards China. Users on each of the platforms often express this by frequently using the terms "Chinese virus" and "Wuhan virus" to refer to the pandemic.

Twitter, Reddit and 4chan also communicate particular frustration with China through aggressive terms. Such topics were discussed in

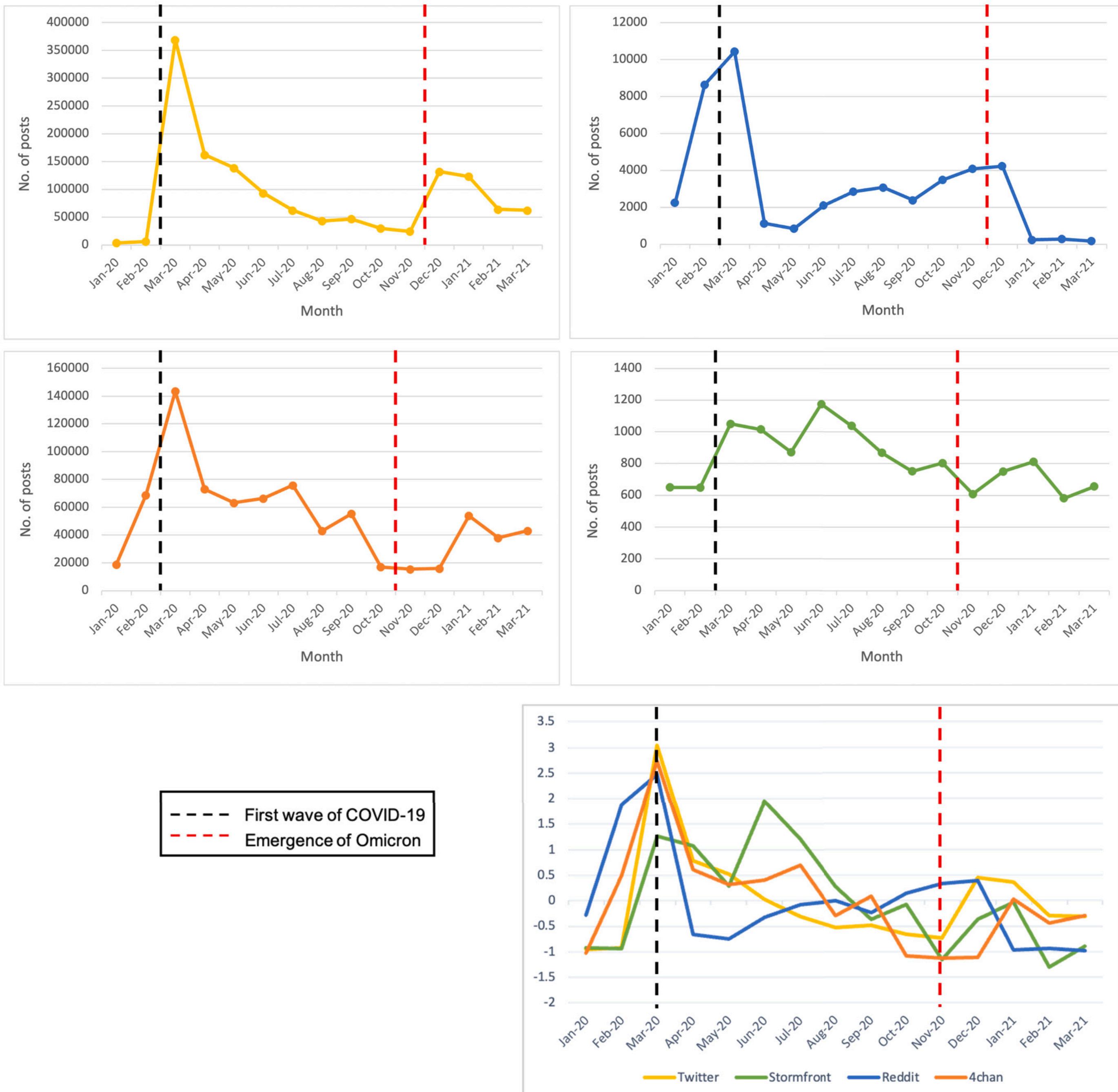

Fig. 1. Graphs showing the frequency of posts across each dataset over the course of the COVID-19 pandemic: Twitter (top-left), Reddit (top-right), 4chan (bottom-left), and Stormfront (bottom-right). A graph with the normalised data from all four datasets is shown at the bottom.

around 51% of the Reddit dataset and around 41% of the 4chan dataset, where the Reddit posts often include phrases like "f*** China" in COVID-19 related discussions (Topic #5), while 4chan posts would even go as far as suggesting to "nuke China" (Topic #5). Twitter users would also use slogans like "make China pay" frequently along with discussing how China has been unpunished, so as to demand that China be held accountable for the pandemic (Topic #4). This indicates that Twitter is often used to petition for various social movements and calls for action. In addition to this, conspiracy theories blaming China for the origin of the virus and for silencing "whistleblowers" speaking out against China's role in the pandemic are also prevalent on Twitter (Topic #2).

The topic model for the Reddit and 4chan posts show that, overall, similar topics were discussed on both these platforms. For instance, resistance to wearing face masks is shown to be a common theme on Reddit (Topic #3) and 4chan (Topic #1), along with discussions about the death rates of the pandemic (Reddit Topic #4, 4chan Topic #3). However it should be noted that the topic model displays how 4chan frequently uses discriminatory and hateful language, particularly when degrading or complaining about various minority groups. Here, the lack of moderation along with the culture of trolling and shock value on 4chan has resulted in the regular usage of hate speech and discriminatory language.

The majority of topics identified in the Stormfront dataset are centred around pro-white and neo-Nazi ideologies, where a strong sense of white identity and membership to the online community is evident amongst the users; around 41% of Stormfront posts contain topics re-

Table 1
A topic model of the most discussed topics during the COVID-19 pandemic and the percentage of posts containing them.

	Twitter	Reddit	4chan	Stormfront
Topic#1	chinese virus, uk coronavirus, stay home new york, national response, failed national, china, lockdown extension **(19%)**	china, china virus, china as*****, russia, communist, china flu **(11%)**	mask, wear, wearing mask, don't wear, f*** masks, face mask **(18%)**	jew, like jews, media, trump, jew york, jew owned, owned media **(21%)**
Topic#2	china, virus, world outbreak, help contain, refused help, whistleblowers china, silenced whistleblowers, virus originate **(23%)**	trump, people, like, just think, Biden, won't **(26%)**	virus, corona virus, fake, chinese virus, corona spread, flue, vaccine, wuhan **(25%)**	people, white people, white hate, blacks, black people, think, racist **(23%)**
Topic#3	agenda21, plandemic, reclassified, recovery rate, flawed, deaths, reclassified **(21%)**	mask, wear, wearing mask, don't wear, face cover **(19%)**	covid 19, deaths, died, coronavirus, cases, flu, vaccine **(15%)**	covid, covid19, vaccine, news, coronavirus, positive, trump, world, covid vaccine **(17%)**
Topic#4	outbreak, enormity, covering severity, communist party, china unpunished, criminal drtedros, make china pay **(22%)**	covid, covid19, deaths, died, covid deaths **(23%)**	asian, white, asian women, black, white men, ch*** **(19%)**	china virus, coronavirus, jews, flu, chinese virus, trump, ccp virus **(19%)**
Topic#5	document revelations, significant document, chinese virus, crisis, wuhan coronavirus **(15%)**	f*** china, china flu, wihan virus, f*** chinese, spread, deadly, coronavirus **(21%)**	chinese virus, jews, communist, ccp, chinese government, f*** china, nuke china, war **(23%)**	anti white, white hate, racist, racism, media, racist media. **(20%)**

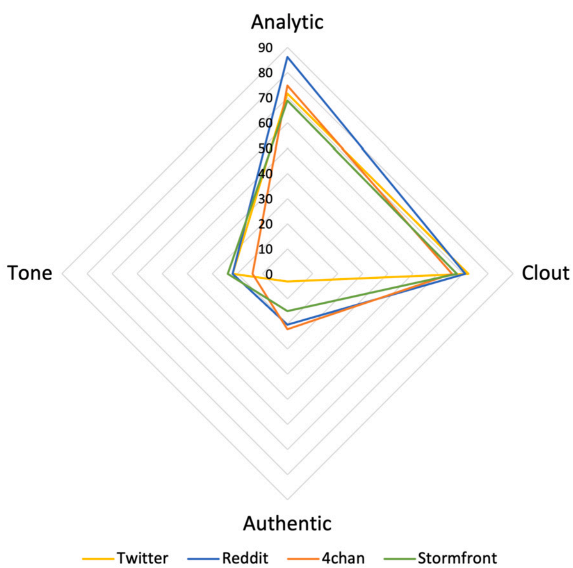

Fig. 2. A comparison of the summary language LIWC categories across all four COVID-19-related datasets.

lated to such themes. Many posts feature antisemitic narratives and conspiracies, including the idea that white people are under threat from non-white groups, and that this threat is being orchestrated by a Jewish conspiracy [49]. This can be seen in Topic #1 and Topic #4 in the topic model, which promote the belief that Jews control the media, government and financial institutions, and are using their power to undermine the white race. Additionally, users often victimise themselves and advocate against white hate while referring to groups of 'others', particularly "blacks" and "Jews", as the enemy (Topic #2).

Within the context of COVID-19-related posts, the topic model shows that Twitter hosts several conspiracy theories, highlighting the impact of misinformation on social-media discourse surrounding the pandemic (Topic #2, Topic #3, Topic #4). This is especially evident through frequent usage of the terms "agenda 21" and "plandemic". These often suggest that COVID-19 is being used as a pretext to enforce government control over citizens through the various restrictions and regulations [50]. Similarly, the term "plandemic" has often been used by conspiracy theorists to refer to the false and baseless claim that the COVID-19 pandemic was not a naturally occurring event, but rather a planned and intentional event by certain individuals or organisations [51]. Such narratives have contributed to the spread of hateful and discriminatory attitudes and actions towards certain groups, especially minority groups. Over 60% of the Twitter posts would discuss topics related to false narratives and conspiracies.

Table 2
Results from the linguistic analysis of the COVID-19-related datasets using LIWC.

LIWC Category	Twitter	Reddit	4chan	Stormfront
I	0.52	1.64	1.57	2.03
We	0.88	1.00	0.87	0.94
You	0.71	1.73	0.86	1.59
They	0.54	1.42	1.68	1.67
Swear	0.17	0.54	1.29	0.25
Religion	0.28	0.21	0.39	1.30

4.3. Linguistic and sentiment analysis

The next component of the cross-platform analysis comprises of exploring and comparing the sentiment and linguistic composition of the COVID-19-related posts collected from each of the four platforms. Here, LIWC is used to highlight the key differences in the psychological processes and various linguistic dimensions, the findings from which are used to address RQ3. The results from this analysis of all four sets of posts are shown in Fig. 2, Fig. 3, and Table 2, where the mean percentages of all words that fall into a particular LIWC category are presented. Further details on how these categories are calculated as well as example words can be found in [44].

Firstly, the four summary language categories (*analytical thinking*, *clout*, *authenticity*, and *emotional tone*) were compared across each dataset, as shown in Fig. 2. The LIWC analysis with the COVID-19-related posts show that the Reddit dataset has a much higher score for *analytical thinking* than the other platforms ($\mu = 86.16$), indicating that users would post more consistent thoughts and opinions on this platform. Twitter posts had a much lower score in this category ($\mu = 74.72$), suggesting a lower degree in logical and hierarchical thinking. This could partly be due to the fact that, over the course of the pandemic, Twitter posts were shown to host a number of false narratives and conspiracy theories, as shown in the findings from the topic model in the previous section. The degree of *clout* within the Twitter ($\mu = 72.16$) and Reddit ($\mu = 70.96$) datasets are shown to be higher than 4chan ($\mu = 65.88$) and Stormfront ($\mu = 67.58$). These higher *clout* scores demonstrate a stronger sense of authority and confidence [44].

The *authenticity* score for the Twitter dataset ($\mu = 2.98$) is much lower than the scores for the other platforms. The Reddit ($\mu = 20.34$) and 4chan ($\mu = 22.04$) posts, however, have considerably higher scores in this category. One inference that could be made from this is that the posts on these two platforms are more personable and disclosing [52]. The scores for the *emotional tone* of each dataset are all below 50. This indicates that the overall emotions on all four platforms are negative. This is consistent with the findings from the topic analysis detailed in the previous section, where the majority of users discussed frustration with the various actors, particularly China, they blamed for the cause

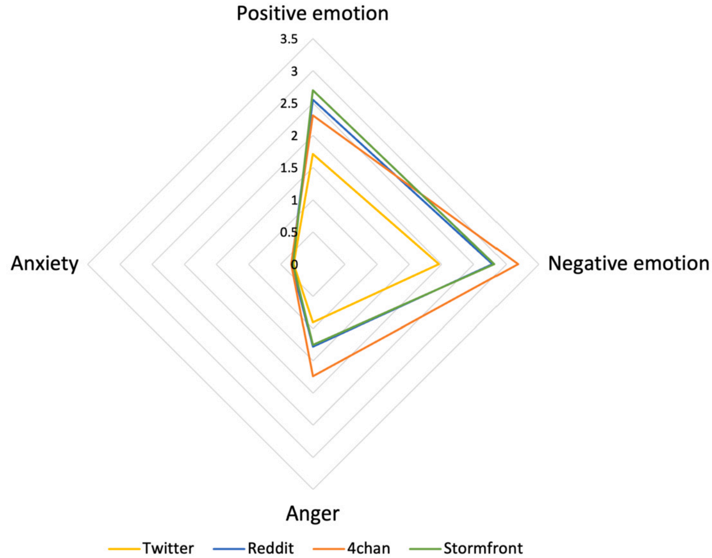

Fig. 3. A comparison of the sentiment LIWC categories across all four COVID-19-related datasets.

of the pandemic. Overall, the 4chan posts were shown to be the most negative ($\mu = 13.90$).

The LIWC sentiment categories (*positive emotion*, *negative emotion*, *anger* and *anxiety*) were then used to further explore the sentiments of the posts from each platform, the results for which can be shown in Fig. 3. Notably, the findings from this component of the LIWC analysis show that *negative emotion* was generally used a lot more than positive emotion across all four platforms, within the context of COVID-19. As expected, 4chan posts have the highest score for *negative emotion* ($\mu = 3.18$). To gain further contextual understanding of the usage of negative emotion, a small sample of posts were also manually analysed. Through this, it was shown that 4chan and Reddit posts complaining about the various regulations put in place to try to control the spread of COVID-19, such as wearing face masks, used more negative emotion. Stormfront users would use negative emotion when promoting hateful narratives against Jews controlling the media, and the victimisation they feel as targets of "white hate". As these were shown to dominate most of the topics of discussion in the topic modelling carried out in the previous section, it is unsurprising that negative emotion was much more present than positive emotion in COVID-19-related posts from all four platforms.

Similar to the results from the analysis of *negative emotion*, the scores for the *anger* LIWC sentiment category show that the 4chan dataset had the highest level of *anger* ($\mu = 1.74$). Again, such language was used to express frustration with regulations put in place, and to put the blame of the creation and spread of the virus on various groups, namely China and Asians, through conspiracy theories and false narratives. 4chan posts were particularly aggressive in expressing this as they would often discuss how to retaliate against the virus by "nuking China", which was previously shown to be a major topic of discussion on this platform in Section 4.2.

The next stage of the LIWC analysis explored the functional composition and linguistic dimensions of the posts within each dataset. In particular, this analysis examined the usage of pronouns, with the results included in Table 2. In general, most of the platforms would use more first-person singular pronouns (such as *I*, *me*, *my*) than first-person plural pronouns (such as *we*, *our*, *us*). In terms of the Twitter dataset, the opposite was true, where posts used less first-person singular pronouns ($\mu = 0.52$) than first-person plural pronouns ($\mu = 0.88$).

The LIWC analysis also shows that third-person plural pronouns (such as *they*, *them*) were used the most in the 4chan ($\mu = 1.68$) and Stormfront ($\mu = 1.67$) datasets. This indicates a larger presence of the "Us vs. Them" mentality [53] on these two platforms than on Twitter and Reddit. This is consistent with the findings gained from the topic analysis carried out in the previous section, where 4chan and Stormfront posts would especially refer to groups of "others", specifically Chinese, Asian people and Jews, in discriminatory and hateful ways.

Previous works have often identified the use of pronouns as a discursive tool used to persuade audiences, partly due to how they can be interpreted by the audience on whether they are inclusive or exclusive of them [54]. In particular, the usage of personal pronouns (such as *we*, *you*, *our*, *us*) is a frequently utilised persuasive tactic that can help make an audience feel more included. The LIWC analysis reveals that this particular approach is used more in Reddit and Stormfront posts compared to the other two platforms, indicating a stronger sense of community on these sites. This finding aligns with previous observations made in the topic analysis, which found that Stormfront users exhibited a stronger sense of shared white identity and membership to their online community. The Twitter posts, however, were shown to be the least inclusive of their audience, which could be attributed to the platform's larger audience size. The platform structure of Reddit and Stormfront as forums that foster online communities may also account for these differences observed in the use of personal pronouns.

5. Conclusions and future work

Ultimately, our work provides some initial comparison across hate-specific online environments on four different platforms within the context of the COVID-19 pandemic. This research builds on previous work by exploring both mainstream platforms, such as Twitter and Reddit, as well as non-moderated fringe platforms, like 4chan and Stormfront. Our study harnessed various computational methods, including topic modelling, linguistic analysis and sentiment analysis to explore the type of content that is promoted on each platform. This cross-platform analysis revealed that the participation trends on all four platforms are generally very similar, with peaks occurring at corresponding times to real-time developments. However, such events seem to control the discourse on platforms with a larger user base, like Twitter, Reddit and 4chan, as compared to smaller, underground platforms like Stormfront.

Through topic modelling, this analysis was able to find that all four platforms would refer to the COVID-19 pandemic as "Chinese virus" or "Wuhan virus", which previous articles have linked to racist and hateful narratives [55]. In this context, Twitter and Stormfront were shown to predominantly promote false and hateful conspiracy theories, whereas 4chan and Reddit users mostly expressed their frustration with regulation and restrictions related to the pandemic. Finally, further sentiment and linguistic analysis showed the use of personal pronouns, which previous literature have shown to be a common persuasive technique in writing, were harnessed by Stormfront users over the course of the pandemic to promote a stronger sense of community and shared identity. The dichotomy mentality of "Us Vs. Them" is reflected strongly in 4chan and Stormfront, which often exhibit hateful narratives regarding groups of "others".

Our study also lays the groundwork for future research endeavours aimed at gaining a deeper understanding of online hate speech dynamics across diverse platforms. While our analysis focused on four platforms (Twitter, Reddit, 4chan, and Stormfront) during the COVID-19 pandemic, there remains ample room for further investigation and exploration. Firstly, expanding the scope of our analysis to include additional platforms beyond the ones studied in this work would provide a more comprehensive understanding of platform-specific content dynamics. Platforms such as Facebook, Instagram, and YouTube, each with their unique user demographics and content moderation policies, warrant examination to ascertain whether online hate manifests differently across these platforms. However, it is essential to acknowledge the challenges posed by restrictive data collection allowances, particularly on platforms like Facebook, which may necessitate the development of innovative research methodologies to overcome these barriers.

Furthermore, future research efforts should employ hate speech classifiers to discern the presence of different hate ideologies across platforms. By leveraging machine learning techniques, researchers can identify patterns and trends in hate speech expression, shedding light on whether certain platforms cater to specific forms of hate or host a broader spectrum of hate ideologies. This nuanced understanding is crucial for devising targeted interventions and platform-specific mitigation strategies. Additionally, extending our analysis beyond the COVID-19 context to encompass other case studies or events would provide insights into the temporal dynamics of online hate. Comparing platform usage patterns and content themes across different events or social phenomena could elucidate whether platform-specific behaviours persist or vary over time and context.

CRediT authorship contribution statement

Fatima Zahrah: Writing – review & editing, Writing – original draft, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Jason R.C. Nurse:** Supervision. **Michael Goldsmith:** Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data.

References

[1] I. Awan, Cyber-extremism: ISIS and the power of social media, Society 54 (2) (2017) 138–149, https://doi.org/10.1007/s12115-017-0114-0, https://doi.org/10.1007/s12115-017-0114-0.

[2] N.F. Johnson, R. Leahy, N.J. Restrepo, N. Velasquez, M. Zheng, P. Manrique, P. Devkota, S. Wuchty, Hidden resilience and adaptive dynamics of the global online hate ecology, Nature 573 (7773) (2019) 261–265, https://doi.org/10.1038/s41586-019-1494-7, https://doi.org/10.1038/s41586-019-1494-7.

[3] Department for Digital Culture Media U.K. Sport, Online Harms White Paper, https://www.gov.uk/government/consultations/online-harms-white-paper, 2019.

[4] S.M. Eddington, The communicative constitution of hate organizations online: a semantic network analysis of "make America great again", Social Media + Society 4 (3) (2018) 2056305118790763, https://doi.org/10.1177/2056305118790763, https://doi.org/10.1177/2056305118790763.

[5] A. Olteanu, C. Castillo, J. Boy, K.R. Varshney, The effect of extremist violence on hateful speech online, in: Proceedings of the International AAAI Conference on Web and Social, Media, 2018.

[6] I. Awan, Islamophobia and Twitter: a typology of online hate against muslims on social media, Policy & Internet 6 (2) (2014) 133–150, https://doi.org/10.1002/1944-2866.POI364, https://doi.org/10.1002/1944-2866.POI364, https://onlinelibrary.wiley.com/doi/abs/10.1002/1944-2866.POI364.

[7] E. Pacheco, N. Melhuish, Online Hate Speech: a Survey on Personal Experiences and Exposure Among Adult New Zealanders, Tech. Rep., SSRN, 2018.

[8] S. Phadke, T. Mitra, Many faced hate: a cross platform study of content framing and information sharing by online hate groups, in: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, CHI'20, Association for Computing Machinery, New York, NY, USA, 2020, pp. 1–13, https://doi.org/10.1145/3313831.3376456.

[9] M. ElSherief, S. Nilizadeh, D. Nguyen, G. Vigna, E. Belding, Peer to peer hate: hate speech instigators and their targets, in: Proceedings of the International AAAI Conference on Web and Social Media, vol. 12, 2018, pp. 52–61.

[10] M. Ribeiro, P. Calais, Y. dos Santos, V. Almeida, W. Meira Jr, Characterizing and detecting hateful users on Twitter, in: Proceedings of the International AAAI Conference on Web and Social Media, vol. 12, 2018, pp. 676–679.

[11] M. Zampieri, S. Malmasi, P. Nakov, S. Rosenthal, N. Farra, R. Kumar, Predicting the type and target of offensive posts in social media, arXiv:1902.09666, 2019.

[12] S. Frenda, B. Ghanem, M. Montes-y Gómez, P. Rosso, D. Pinto, V. Singh, Online hate speech against women: automatic identification of misogyny and sexism on Twitter, Journal of Intelligent & Fuzzy Systems 36 (5) (2019) 4743–4752.

[13] J. Salminen, H. Almerekhi, M. Milenkovic, S.-G. Jung, J. An, H. Kwak, B.J. Jansen, Anatomy of online hate: developing a taxonomy and machine learning models for identifying and classifying hate in online news media, in: Twelfth International AAAI Conference on Web and Social, Media, 2017.

[14] S. Agarwal, A. Sureka, Characterizing linguistic attributes for automatic classification of intent based racist/radicalized posts on tumblr micro-blogging website, arXiv:1701.04931, 2017.

[15] P. Burnap, M.L. Williams, Us and them: identifying cyber hate on Twitter across multiple protected characteristics, EPJ Data Science 5 (1) (2016) 11, https://doi.org/10.1140/epjds/s13688-016-0072-6, https://doi.org/10.1140/epjds/s13688-016-0072-6.

[16] P.S. Br Ginting, B. Irawan, C. Setianingsih, Hate speech detection on Twitter using multinomial logistic regression classification method, in: 2019 IEEE International Conference on Internet of Things and Intelligence System (IoTaIS), 2019, pp. 105–111.

[17] S.S. Aluru, B. Mathew, P. Saha, A. Mukherjee, Deep learning models for multilingual hate speech detection, arXiv:2004.06465, 2020.

[18] Z. Waseem, D. Hovy, Hateful symbols or hateful people? Predictive features for hate speech detection on Twitter, in: Proceedings of the NAACL Student Research Workshop, Association for Computational Linguistics, San Diego, California, 2016, pp. 88–93, https://www.aclweb.org/anthology/N16-2013.

[19] J.Y. Kim, C. Ortiz, S. Nam, S. Santiago, V. Datta, Intersectional bias in hate speech and abusive language datasets, arXiv:2005.05921, 2020.

[20] S. Liu, T. Forss, Combining n-gram based similarity analysis with sentiment analysis in web content classification, in: Proceedings of the International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management - Volume 1, IC3K 2014, SCITEPRESS, Science and Technology Publications, Lda, Setubal, PRT, 2014, pp. 530–537, https://doi.org/10.5220/0005170305300537.

[21] Z. Mossie, J.-H. Wang, Vulnerable community identification using hate speech detection on social media, Information Processing & Management 57 (3) (2020) 102087, https://doi.org/10.1016/j.ipm.2019.102087, https://www.sciencedirect.com/science/article/pii/S0306457318310902.

[22] P. Kapil, A. Ekbal, D. Das, Investigating deep learning approaches for hate speech detection in social media, arXiv:2005.14690, 2020.

[23] M. Conway, Determining the role of the Internet in violent extremism and terrorism: six suggestions for progressing research, Studies in Conflict & Terrorism 40 (1) (2017) 77–98, https://doi.org/10.1080/1057610X.2016.1157408, https://doi.org/10.1080/1057610X.2016.1157408.

[24] F. Zahrah, J.R.C. Nurse, Terrorism and online extremism, in: Elgar Encyclopedia of Technology and Politics, Edward Elgar Publishing, 2022, pp. 62–67.

[25] F. Zahrah, J.R.C. Nurse, M. Goldsmith, A comparison of online hate on reddit and 4chan: a case study of the 2020 US election, in: Proceedings of the 37th ACM/SIGAPP Symposium on Applied Computing, ACM, 2022, pp. 1797–1800.

[26] S. Zannettou, T. Caulfield, E. De Cristofaro, N. Kourtelris, I. Leontiadis, M. Sirivianos, G. Stringhini, J. Blackburn, The web centipede: understanding how web communities influence each other through the lens of mainstream and alternative news sources, in: Proceedings of the 2017 Internet Measurement Conference, IMC'17, Association for Computing Machinery, New York, NY, USA, 2017, pp. 405–417, https://doi.org/10.1145/3131365.3131390.

[27] Hitkul, A. Prabhu, D. Guhathakurta, J. Jain, M. Subramanian, M. Reddy, S. Sehgal, T. Karandikar, A. Gulati, U. Arora, R.R. Shah, P. Kumaraguru, Capitol (pat)riots: a comparative study of Twitter and parler, arXiv:2101.06914, 2021.

[28] I. Murdock, K.M.C. Carley, O. Yagan, Multi-Platform Analysis of 2020 U.S. Election: Fraud and Protest Related Posts, in: IDeaS Conference, the Center for Informed Democracy and Social-Cybersecurity, Pittsburgh, 2021.

[29] V. Louie, A. Viladrich, "divide, divert, & conquer" deconstructing the presidential framing of white supremacy in the covid-19 era, Social Sciences 10 (8) (2021) 280.
[30] A. Elias, J. Ben, F. Mansouri, Y. Paradies, Racism and nationalism during and beyond the covid-19 pandemic, in: Race and Ethnicity in Pandemic Times, Routledge, 2021, pp. 65–75.
[31] WHO, WHO director-general's opening remarks at the media briefing on COVID-19 – 11 March 2020, 2020, https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020.
[32] C. Kantis, S. Kiernan, J.S. Bardi, L. Posner, I. Turilli, UPDATED: timeline of the coronavirus, 2023, https://www.thinkglobalhealth.org/article/updated-timeline-coronavirus.
[33] B. He, C. Ziems, S. Soni, N. Ramakrishnan, D. Yang, S. Kumar, Racism is a virus: anti-Asian hate and counterspeech in social media during the covid-19 crisis, in: Proceedings of the 2021 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ASONAM'21, Association for Computing Machinery, New York, NY, USA, 2021, pp. 90–94, https://doi.org/10.1145/3487351.3488324.
[34] R. Gallotti, F. Valle, N. Castaldo, P. Sacco, M. De Domenico, Assessing the risks of 'infodemics' in response to covid-19 epidemics, Nature Human Behaviour 4 (12) (2020) 1285–1293.
[35] E. Chandrasekharan, M. Samory, A. Srinivasan, E. Gilbert, The bag of communities: identifying abusive behavior online with preexisting Internet data, in: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, CHI'17, Association for Computing Machinery, New York, NY, USA, 2017, pp. 3175–3187, https://doi.org/10.1145/3025453.3026018.
[36] S. Peeters, S. Hagen, 4cat: Capture and analysis toolkit, Computer Software. Vers 1 (0), 2018.
[37] N. Reimann, Reddit bans 'R/donaldtrump' subreddit, https://www.forbes.com/sites/nicholasreimann/2021/01/08/reddit-bans-rdonaldtrump-subreddit/?sh=5980347038b3, 2021.
[38] D. Godfrey, C. Johns, C. Meyer, S. Race, C. Sadek, A case study in text mining: interpreting Twitter data from world cup tweets, arXiv:1408.5427, 2014.
[39] E. Calisir, M. Brambilla, The long-running debate about brexit on social media, in: Proceedings of the International AAAI Conference on Web and Social Media 14 (1), 2020, pp. 848–852, https://aaai.org/ojs/index.php/ICWSM/article/view/7349.
[40] K. Jones, J.R.C. Nurse, S. Li, Behind the mask: a computational study of anonymous' presence on Twitter, Proceedings of the International AAAI Conference on Web and Social Media 14 (1) (2020) 327–338, https://ojs.aaai.org/index.php/ICWSM/article/view/7303.
[41] D.M. Blei, A.Y. Ng, M.I. Jordan, Latent Dirichlet allocation, Journal of Machine Learning Research 3 (2003) 993–1022.
[42] J. Mazarura, A. de Waal, A comparison of the performance of latent Dirichlet allocation and the Dirichlet multinomial mixture model on short text, in: 2016 Pattern Recognition Association of South Africa and Robotics and Mechatronics International Conference (PRASA-RobMech), Institute of Electrical and Electronics Engineers (IEEE), South Africa, 2016, pp. 1–6.
[43] R. Mehrotra, S. Sanner, W. Buntine, L. Xie, Improving lda topic models for microblogs via tweet pooling and automatic labeling, in: Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR'13, Association for Computing Machinery, New York, NY, USA, 2013, pp. 889–892, https://doi.org/10.1145/2484028.2484166.
[44] J. Pennebaker, R. Boyd, K. Jordan, K. Blackburn, The Development and Psychometric Properties of LIWC2015, Tech. Rep, University of Texas, Austin, 2015.
[45] E. Kacewicz, J.W. Pennebaker, M. Davis, M. Jeon, A.C. Graesser, Pronoun use reflects standings in social hierarchies, Journal of Language and Social Psychology 33 (2) (2014) 125–143, https://doi.org/10.1177/0261927X13502654, https://doi.org/10.1177/0261927X13502654.
[46] J.W. Pennebaker, The Secret Life of Pronouns: What Our Words Say About Us, Bloomsbury Press, New York, 2011.
[47] F. Zahrah, J.R. Nurse, M. Goldsmith, #isis vs #actioncountersterrorism: a computational analysis of extremist and counter-extremist Twitter narratives, in: 2020 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW), IEEE, 2020, pp. 438–447.
[48] S. Kumar Nancy, Efficient k-mean clustering algorithm for large datasets using data mining standard score normalization, International Journal on Recent and Innovation Trends in Computing and Communication 2 (2014) 3161–3166.
[49] L. Bowman-Grieve, Exploring "stormfront": a virtual community of the radical right, Studies in Conflict & Terrorism 32 (11) (2009) 989–1007, https://doi.org/10.1080/10576100903259951, https://doi.org/10.1080/10576100903259951.
[50] Reuters, Fact check: agenda 2030 is a global development framework that is not linked to the COVID-19 pandemic, https://www.reuters.com/article/uk-factcheck-agenda-2030-idUSKBN2AN2CQ, 2021.
[51] S. Nazar, T. Pieters, Plandemic revisited: a product of planned disinformation amplifying the COVID-19 "infodemic", https://www.frontiersin.org/articles/10.3389/fpubh.2021.649930, 2021.
[52] W.W. Xu, C. Zhang, Sentiment, richness, authority, and relevance model of information sharing during social crises — the case of #MH370 tweets, Computers in Human Behavior 89 (2018) 199–206, https://doi.org/10.1016/j.chb.2018.07.041, https://www.sciencedirect.com/science/article/pii/S0747563218303637.
[53] M. Cikara, M.M. Botvinick, S.T. Fiske, Us versus them: social identity shapes neural responses to intergroup competition and harm, Psychological Science 22 (3) (2011) 306–313, https://doi.org/10.1177/0956797610397667, https://doi.org/10.1177/0956797610397667.
[54] Y.-J. Zupnik, A pragmatic analysis of the use of person deixis in political discourse, Journal of Pragmatics 21 (4) (1994) 339–383, https://doi.org/10.1016/0378-2166(94)90010-8.
[55] I. Sakki, L. Castrén, Dehumanization through humour and conspiracies in online hate towards Chinese people during the COVID-19 pandemic, British Journal of Social Psychology 61 (4) (2022) 1418–1438, https://doi.org/10.1111/bjso.12543, https://doi.org/10.1111/bjso.12543.