

RESEARCH ARTICLE

What Tweets and YouTube comments have in common? Sentiment and graph analysis on data related to US elections 2020

Alexander Shevtsov^{1,2}, Maria Oikonomidou^{1,2}, Despoina Antonakaki^{1*}, Polyvios Pratikakis^{1,2}, Sotiris Ioannidis^{1,3}

1 Institute of Computer Science, Foundation for Research and Technology, Vassilika Vouton, Heraklion, Crete, Greece, **2** Computer Science Department - University of Crete, Voutes Campus, Heraklion, Crete, Greece, **3** School of Electrical and Computer Engineering, Technical University of Crete, University Campus, Akrotiri, Chania, Greece

* despoina@ics.forth.gr



OPEN ACCESS

Citation: Shevtsov A, Oikonomidou M, Antonakaki D, Pratikakis P, Ioannidis S (2023) What Tweets and YouTube comments have in common? Sentiment and graph analysis on data related to US elections 2020. PLoS ONE 18(1): e0270542. <https://doi.org/10.1371/journal.pone.0270542>

Editor: Daswin De Silva, La Trobe University - Melbourne Campus: La Trobe University, AUSTRALIA

Received: December 11, 2020

Accepted: May 30, 2022

Published: January 31, 2023

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: <https://doi.org/10.1371/journal.pone.0270542>

Copyright: © 2023 Shevtsov et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: In order to obtain the dataset used for the analysis described in this study, we follow the Twitter API restrictions and do

Abstract

Most studies analyzing political traffic on Social Networks focus on a single platform, while campaigns and reactions to political events produce interactions across different social media. Ignoring such cross-platform traffic may lead to analytical errors, missing important interactions across social media that e.g. explain the cause of trending or viral discussions. This work links Twitter and YouTube social networks using cross-postings of video URLs on Twitter to discover the main tendencies and preferences of the electorate, distinguish users and communities' favouritism towards an ideology or candidate, study the sentiment towards candidates and political events, and measure political homophily. This study shows that Twitter communities correlate with YouTube comment communities: that is, Twitter users belonging to the same community in the Retweet graph tend to post YouTube video links with comments from YouTube users belonging to the same community in the YouTube Comment graph. Specifically, we identify Twitter and YouTube communities, we measure their similarity and differences and show the interactions and the correlation between the largest communities on YouTube and Twitter. To achieve that, we have gathered a dataset of approximately 20M tweets and the comments of 29K YouTube videos; we present the volume, the sentiment, and the communities formed in YouTube and Twitter graphs, and publish a representative sample of the dataset, as allowed by the corresponding Twitter policy restrictions.

1 Introduction

Previous studies analyzing political content on Social Networks, mainly focus on a single platform, however, online political campaigns and interactions between users seem to take place across diverse online social networks. Online discourse is scattered across several social networks, sometimes among the same users and communities. The parallel study of the content of these social networks can reveal the political campaigns [1], comments, opinions,

not violate any terms from Twitter Developer Agreement and Policy. According to Twitter Policy, we are not allowed to share the entire dataset, but only 100K user IDs. This dataset is available here: <https://zenodo.org/record/4618233#.YGGJU2Qzada>. The access is open and no approval is required. We provide the directed retweet graph from the Twitter network, all user IDs from the provided retweet graph (89,479 users), all video IDs (vid) extracted from the election related tweets (39,203 video ids) and the directed comment graph.

Funding: This document is the results of the research project co-funded by the European Commission, project CONCORDIA, with grant number 830927 (EUROPEAN COMMISSION Directorate-General Communications Networks, Content and Technology) and by the European Union and Greek national funds through the Operational Program Competitiveness, Entrepreneurship and Innovation, under the call RESEARCH - CREATE - INNOVATE (project ode: T1EDK-02857 and T1EDK-01800). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

tendencies, and beliefs of the electorate [2, 3]. For instance, social media played a very important role during the 2016 US elections [1, 4–6]. However, only a few studies focus on more than one social networks at a time [7, 8], thus possibly missing important interactions.

Users on Twitter and YouTube form tight communities. Twitter homophily has been noticed in communities [9], in hashtags [10], on Twitter lists [11], and modelled in the follow and mention graphs [12, 13]. The analysis of this phenomenon has been studied under the prism of political context as well, like in [14–19], or support homophily between users and social media [20]. Homophily and forming of common communities on social media have been noticed on YouTube as well, as seen in several background studies [21–23].

Social media users with political interests share and seek information about politics on Twitter and there is a high probability these particular users will also search for additional information on other social networks with similar topics. This could indicate potential connections across different social platforms, especially in the case of political discourse. Consequently, the analysis of a single social network cannot reveal the entire picture of the connections across several social networks.

As supported above, online social networks allow users to form groups and communities, where the members are related to the same topic or interest. Based on that, we assume that communities are generated based on the users' preferences. Each social network has its own communities, where the users are already connected to each other, by some hidden layers of each social network. In our study, we provide an analysis of Twitter and YouTube where these layers are consisted of the features like retweets, replies, mentions - in the case of Twitter- and comments and likes -in the case of YouTube-. We already have information about the community structures of each social network, but the correlation of those communities across both of them is still not covered. From our perspective, it is very important to identify if indeed communities with similar interests are connected across different social networks.

In this study, we manage to reveal the connections between the communities across YouTube and Twitter. Through community detection in the two social networks, we reveal the interactions between the comment graph and retweet graph, respectively. To achieve this, initially, we obtain the most popular hashtags related to the US elections and gather a dataset of 19.8M tweets, for a period of four months starting from July 19 of 2020. We extract 28,343 unique YouTube video links contained in the Twitter dataset, as well as their metadata (comments, authors, etc.). We perform volume analysis, identify the main entities in the corpus and perform per-entity sentiment analysis. We study the evolving retweet graph in six sequential time periods from July to November 2020 and show the diffusion of content regarding the two candidates. The results of sentiment analysis give a higher positive sentiment towards Donald Trump on Twitter (35.2%) and YouTube (18%), in comparison to the positive sentiment expressed towards Joe Biden on Twitter (28%) and YouTube (12%). We perform community detection in the retweet graph and YouTube comment graph with the use of Louvain method [24]. With PageRank, we identify the most engaging nodes in both graphs and label the communities that contain these nodes after their name. Finally, we show the interactions between the largest, in size, communities on both social media.

This study makes the following contributions:

- Identifies the connections of the communities affiliated with D.Trump and J.Biden campaign between Twitter and YouTube.
- Reveals the interactions between the YouTube communities in the YouTube-comment graph.
- Presents the Twitter retweet communities generated based on the user retweets.

- Highlights the patterns in the comment graph that supports the linking between the Twitter Retweet graph to the YouTube comment graph.

Analyzing the sentiment is an additional step towards the processing of political content [25–29]. Sentiment analysis has been extensively applied to Twitter traffic, usually regarding a specific event. Through sentiment analysis, we can visualize the variation of the sentiment of the electorate, during a political event [30, 31], a company event [32], a product's reviews [33], or model the public mood and emotion and connect tweets' sentiment features with fluctuations with real events [34]. The main task of these works is to predict elections, classify the electorate and distinguish posts towards one political party or ideology, although it has been addressed in many works that Twitter is not suitable for election prediction [35, 36].

1.1 Background

1.1.1 Homophily, communities on Twitter, YouTube. Existing studies towards the analysis of the interactions across different social networks during the elections period are used for diverse goals. They mainly focus on the observation and investigation of the online discourse, highlighting the main activity of the candidates and the evolution of their online presence [37], evaluate how the candidates are influenced by SNs [38], or investigate whether social networks contribute in the democratization of our political systems, studying online propaganda, or the topics of the content in YouTube videos related to campaigns [39].

There is a plethora of studies on homophily on Twitter and YouTube on political content, analyzing the communities that are formed on social media and content generated by these platforms. For example in [40] authors manage to analyse 8.9 M tweets during the 2015 Spain Elections. This study identifies a specific category of users, the so-called “party evangelists” and explores the activity effect on them regarding the general political conversation. In [41] they analyze the communications of alt-right supporters during the 2018 US-midterm elections. After they obtain a dataset of 52903 tweets posted by 123 alt-right Twitter accounts (from 30/10 until 13/11 2018), they apply community detection (Clusset-Newman-Moore greedy modularity maximization method) in the retweet graph, topic extraction of the communities and demonstrate the communities categorized by the topic discussed among them. In [42] they study homophily in journalists' interactions by analyzing a dataset of 600,000 tweets sent by 2908 Australian journalists in a period of one year. They conclude that journalists mainly interact with other journalists of the same gender, working environment, and location forming a bubble.

In [43] they investigate the structure of the “Occupy Wall Street” movement on Twitter and YouTube through a dataset of 328 identified users including their posts, the number of followers, followees tweets, and favorites of users, as well as the corresponding YouTube videos with the keyword “Occupy Wall Street”. The results show that both social networks were coordinated by US users and specifically, a loosely connected network coordinated around central hub users on Twitter while on YouTube the network was organized by anonymous users. In [44] they analyse a dataset of 238,967 tweets around the #gamergate, associated with an episode “Law & Order: Special Victims Unit” that initiated an online activity and controversy and explore the challenges of accounting on the cultural dynamics of Twitter and YouTube, focusing on the key media objects like images videos and tags.

Community detection has been used towards the discovery of relations between different and heterogeneous social networks like in [7], where they build a heterogeneous network with undirected and directed edges, to simulate a social network and study overlapping community detection. Also, in [45] they apply a method of converting multi-relational networks to single-

relational networks on synthetic and real-world data (DBLP). Background work in community detection across Twitter with other real social network is limited.

1.1.2 Political content analysis on Twitter. The background work specifically during election periods on Twitter includes volume analysis of the number of tweets and comparison with the results of the elections, while they apply topic analysis and show the co-occurrence between terms or candidates within the tweets [3, 46–48]. Other works focus on the classification of social media users to separate Democrats and Republicans by using machine learning techniques on features derived from user information [49], or by automatically constructing user profiles in a dataset of scraped lists of users in the Twitter directories WeFollow and Twel-low [50]. For example in [51] in a categorized dataset of four samples covering the period of January 2013 until December 2014, they apply unsupervised learning to classify the tweets topics posted by the legislators and citizens and reveal that the former follows a discussion of public issues and are potentially more interacting with their supporters than to the general public. In [52] on a dataset of 13 million followers of @realDonaldTrump on 8 November 2016, they apply a clustering analysis and show the following patterns of the electorate.

Studies that focus on network analysis on Twitter, during periods of election study also homophily [13, 14, 53, 54], apply stochastic link structure analysis and show the party interactions [55], apply community detection [56, 57], analyze the networks of mentions and retweets [57–60] and estimate users' partisan preference, while in [60] they study the activity, emotional content, and interactions of political parties and politicians. Finally, in [61] they study the structure of the conversations on Twitter between political, media, and citizen agents, while the analysis shows a decentralized and loosely knit network, during elections in Belgium.

Studies are focusing on the prediction of the election outcome ([62, 63]), like in [48] where there are also studying whether Twitter can be used for political deliberation and mirrors off-line political sentiment, during the German federal election.

There is a plethora of studies that have used sentiment analysis in the political domain [64, 65], either for group polarization [13, 66], study specific events or personalities like the Arab spring [67] or Hugo Chavez [68]. For example, in [30] they study a dataset of 36 million tweets on the 2012 U.S. presidential candidates and apply a real-time analysis of public sentiment. Using the Amazon Mechanical Turk, they label the dataset with the tweets' sentiment (positive, negative, neutral, or unsure), in order to apply statistical classification (Naive Bayes model on uni-gram features) on a training set consisting of nearly 17,000 tweets (16% positive, 56% negative, 18% neutral, 10% unsure). Also, in [69] they attempt to understand the broader picture of how Twitter is used by party candidates, understand the content and the level of interaction by followers. More detailed background on Twitter Sentiment Analysis methods can be found on surveys like [70–73].

Sentiment Analysis on Twitter is not limited to one language; there are works studying multiple language datasets [74–80]. Recent studies in sentiment analysis using Deep Convolutional Neural Networks [74, 81–86]. Our work focuses on a single language (English) which is the dataset based on. Finally, some studies incorporate the use of Twitter features, like emoticons [87–90].

Sentiment in these analyses is represented by a variable with values like 'Positive', 'Negative' and 'Neutral', or even more specific ('Happy', 'Angry', etc.). Each word in the corpus can be assigned with more than one sentiment ('Positive' and 'Negative'). Other metrics that can be measured in this analysis, are 'subjectivity' and 'polarity', where the first one is defined as the ratio of 'positive' and 'negative' tweets to 'neutral' tweets, while the second is defined as the ratio of 'Positive' to 'Negative' tweets.

A necessary step of sentiment analysis is 'text normalization', an initial preprocessing of the corpus to extract the lexical features that can significantly affect the performance [91–93]. The

steps of the preprocessing include tokenization, expansion of abbreviations, and removal of stop words (URLs, mentions, etc.).

A very common step that is also used is to incorporate a lexicon, specially made for the domain of the dataset [3, 94]. In [95] they obtain three different corpora of tweets and explore the usage of linguistic features towards sentiment analysis. In [96] they are adopting a lexicon-based method on diverse Twitter datasets. In [97] they conduct a study of sentiment analysis on a dataset of 26,175 general Bulgarian tweets. Through feature selection and classification (binary SVM), they show that the negative sentiment predominated before and after the election period. In the current study, we are not compiling a specially made lexicon, because we do not have a language barrier or analyzing a plethora of entities; we are just focusing on the two main candidates.

1.1.3 Content analysis on YouTube. YouTube analysis has been used to apply sentiment analysis in the recent US Elections 2020, on limited dataset of approximately 200 comments from YouTube [98], to discover irrelevant and misleading metadata [99], to identify spam campaigns [100], to discover extremists videos and hidden communities [101], to propagate preference information of personalised video [102], to estimate causality between user profiles [103], to spread political advertisement [39, 104, 105], and to apply opinion mining [106]. For example in [104], they explore the Senate Campaign 2008. Among other, they claim that not many people are exposed to television ads, therefore YouTube represents an increase in accessibility of the campaign messages, which favored a group to control the messages on the network. Background work on sentiment analysis on YouTube [107, 108], studies the sentiment on user comments [109, 110], identifies the trends and demonstrates the influence of real events of user sentiments [111], implements model utilizing audio, visual and textual modalities as sources of information [112] and studies the popularity indicators and audience sentiments of videos [113].

Previous studies on both Twitter and YouTube, during the elections period, investigate online propaganda strategies like in [8], or evaluate the influence of candidates via social media [38].

In our study, we focus on the discovery of the relation between YouTube and Twitter via community detection and on the link of Twitter Retweet graph with YouTube comment graph, by using tweets of posted videos, the measurement of their similarity and differences, and the interactions and the correlations between the largest communities on YouTube and Twitter. As seen above, previous work on the analysis of social media corpus during election periods like [98], does not provide this large dataset covering almost three months before the elections of political conversation (in two social media), nor does this extended analysis covering sentiment, volume and graph analysis, as well as the comparison and correlation between two social networks.

2 Dataset

We acknowledge the fact that Twitter and YouTube users do not fully represent the electorate during an election period, which consequently introduces a bias in the electorate dataset. Previous studies have shown that Twitter users belong to a specific age [36], social and ideology demographic group [114]. This means that public opinion is not fully expressed through social media. Regarding YouTube, the usage penetration in the United States 2020, by age group is in better shape than Twitter, since the popularity difference between the age groups differs slightly, although it attracts a younger audience, regardless of the minimum age for using the service is 13 years, in most countries [115], not necessarily part of the audience of our dataset.

Taking into consideration the bias of the dataset, the task of election prediction is challenging. However, the analysis of the specific corpus can shed light on the relationship between Twitter and YouTube. As shown in section 4.4, the task of community detection on both social graphs can contribute to the association between the communities on the YouTube comment graph with the communities on the Twitter retweet graph and the measurement of their similarities and differences.

2.1 Twitter

In this study, we search for all the prevalent hashtags regarding the US elections on the 3rd of November 2020 and obtain the Twitter corpus through streaming Twitter API. The acquisition of the dataset started on 19 July 2020 and finished on 3 November 2020.

This resulted in a dataset of approximately 19.8M tweets, containing 4.5M users, with the most prevalent hashtags being #Trump2020, #vote and #Election2020. In [Table 1](#) we can see the most popular hashtags, sorted by the number of tweets within which they are contained and in [S1 Appendix](#), on Table 5 in [S1 Appendix](#) the whole list of hashtags used in our analysis. In [Fig 1](#) we show the number of daily tweets and comments on YouTube, where we notice a peak on September 29, potentially explained by the first debate [116] and of course the second higher peak for Twitter on the day of the elections.

We considered that 19 July 2020 date was a reasonable starting point for collecting our dataset since the number of tweets in the corresponding hashtags we collect begin to accumulate a significant number, as well as the semantics of the content, started to be more relevant to the conversation related to the elections. Regarding the completeness of the content covered by our dataset, we acknowledge the fact that additional minor hashtags may exist during that period, that were not crawled and included in our corpus. We consider our dataset the complete online discourse, since we got the majority of the hashtags available in that period before the elections. Additionally, there was an overlap between the hashtags in the election discourse and these hashtags were cross-referenced in the tweets. For example, some tweets included the popular hashtags (#Vote) and the minor hashtags were also mentioned in the same tweet. This tweet is also included in our dataset because of #Vote. Additionally, we included only the general hashtags and not the ones that were in favor of a particular candidate. We try to include only two hashtags that are in favor of each candidate, in order to keep a balance between them and not introduce a bias towards one of them. Finally, the main amount of the political conversation was gathered in the popular hashtags and the rest of them do not contain a significant amount of tweets. In the [S1 Appendix](#), in Table 5 of [S1 Appendix](#) we show the list of 20

Table 1. The 10 most popular hashtags in our dataset.

Hashtag	Tweets count
#vote	7.196.981
#trump2020	3.913.969
#election2020	3.535.323
#biden	1.569.981
#trump	959.048
#bidenharris2020	866.120
#debate2020	855.543
#votebluetosaveamerica	837.089
#maga	697.381
#trumphascovid	581.456

<https://doi.org/10.1371/journal.pone.0270542.t001>

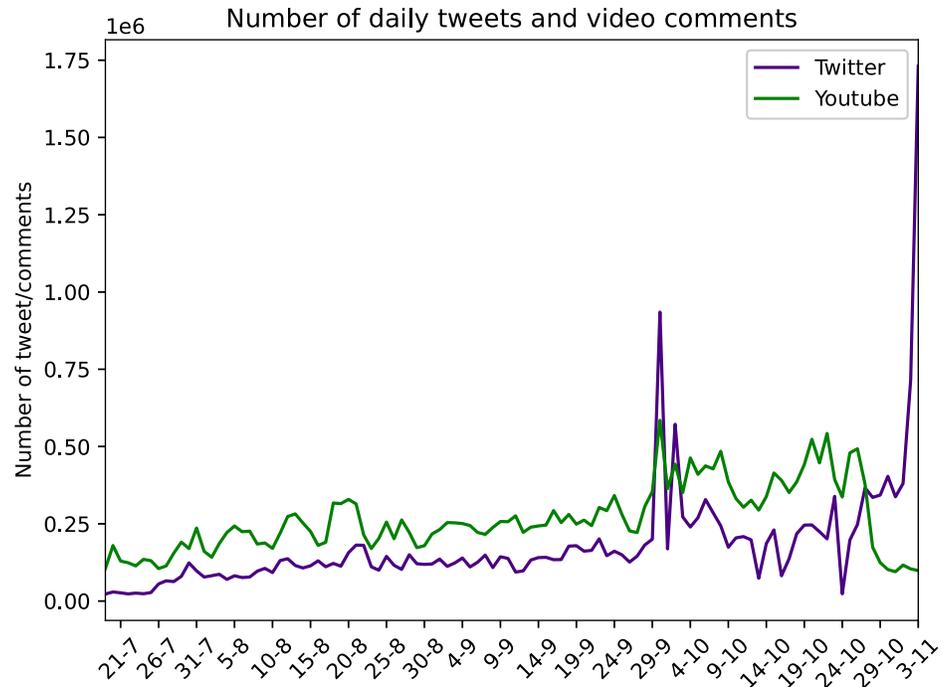


Fig 1. Number of daily tweets and comments on YouTube videos.

<https://doi.org/10.1371/journal.pone.0270542.g001>

most popular hashtags in our dataset. The entire list contains 585,486 unique entries of hashtags, which is too long to be added to the manuscript.

Fig 1 shows the total number of tweets per day, for every hashtag contained in our dataset.

2.2 YouTube

From the election tweets, we extracted all the contained YouTube video links. We followed these links and ended up with 39,203 videos. Through the YouTube Data API, we obtained all the publicly available data regarding these videos. The accessible data used in this study for each video are:

1. Category where it belongs (e.g. News & Politics, Entertainment, Music, etc.),
2. Text and author of each selected comment and,
3. YouTube channel that posted the video.

In Fig 2 we see how many videos belong to each category. In this study, we focus on the election's topic, so we filtered out the videos that do not belong to the following categories: News & Politics, People & Blogs, and Entertainment. The filtering led to a dataset of 28,343 videos.

From the 28,343 videos, we gathered all the comments and their replies generated between 19/7/20 and 3/11/20. This resulted in a dataset of 8,476,193 unique commenters and 39,266,355 comments and replies. Fig 1 shows the total number of comments and replies per day, related to the elections. We notice an increase in the number of comments from July to November, with diurnal patterns, a peak on September 29 on the first debate [116] and of course the second peak for Twitter on the day of the elections.

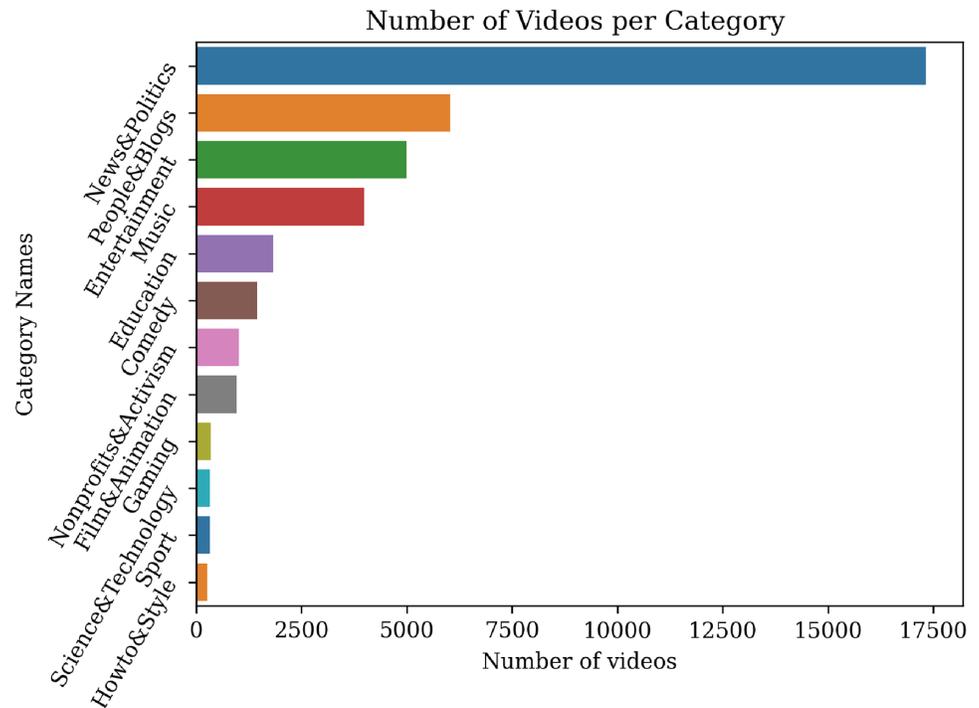


Fig 2. Number of videos per YouTube category.

<https://doi.org/10.1371/journal.pone.0270542.g002>

2.3 Data availability

To obtain the dataset used for the analysis described in this study, we follow the Twitter and YouTube API restrictions and do not violate any terms from the Developer Agreement and Policies. According to Twitter Policy, we are not allowed to share the entire dataset, but only 100K user IDs. This dataset is available here: [zenodo link](#). We provide the directed retweet graph from the Twitter network, all user IDs from the provided retweet graph (89,479 users), all video IDs (vid) extracted from the election-related tweets (39,203 video ids), and the directed comment graph.

The public dataset is composed of the collected data, split into two separate parts, the Twitter and YouTube parts. Each part contains limited information, in order to protect the privacy of users and respect the limitations of the social platform policies.

2.3.1 Twitter. The Twitter part, contains the user_IDs from the retweet graph of the users contained in our dataset, with respect to Twitter policy and limits. This retweet graph can be utilized, in order to identify the retweet relationship between the users during the US Election 2020. In this graph, nodes represent the users in the graph, while edges represent the retweet relationship between two users, accompanied by an edge weight representing the number of retweets between the particular pair of users. Also, the described nodes in the retweet graph are used, in order to collect user information, such as the user posts and user profiles. For these purposes, we used a crawler that respects the Twitter API policy, which is available here: [tWawler](#). This crawler allows the retrieval of the user posts from the user timeline and the profile information based on the provided User_IDs, from Twitter API. In order to keep only the posts that are correlated with the US Election topic, the collection should be filtered based on the hashtags listed in Table 5 in [S1 Appendix](#).

According to Twitter policy, we are not permitted to store and distribute user objects and information that have been removed or suspended from Twitter. Consequently, the shared dataset contains only the tweets that contain the specified hashtags. We do not provide the user information (user objects) and the timeline posts for the accounts, since this information is removed or can be removed in the future by Twitter. Thus, the shared dataset includes only the User_IDs which allows the collection of the publicly available information through the Twitter API, without violating the Twitter policy. For each User_ID we also provide a YouTube video ID that was posted by the user during the elections period. Based on those video IDs, it is possible to collect the users who post the comments and replies to each video with the usage of YouTube API.

2.3.2 YouTube. The YouTube dataset contains a list of video IDs, extracted from the related tweets, described above. For this part, we provide the graph that represents the relationship between two YouTube users commenting on a specific channel. To gather the data needed for this collection part and follow YouTube API's policy restrictions, we used the scripts available at [YouTube Data API](#). Each video on YouTube is categorized based on its content. From the videos obtained from the tweets, we exclude the ones that did not belong to the categories News & Politics, People & Blogs, and Entertainment. For the resulting filtered videos, we gathered all their comments and authors from 19 July 2020 to 3 November 2020, along with the channel that owns them. Initially, to construct the comment graph for each channel, we gather for each filtered video_ID all the user User_IDs from YouTube that have commented on it. Next, we convert the video_IDs to channel_IDs pointing to the channel that owns the video. The result is the graph containing the users and the comments on each channel.

3 Methodology

3.1 Text preprocessing

In order to apply sentiment analysis, we follow a prerequisite set of steps for preprocessing the corpus (tweets - YouTube comments); by removing punctuation symbols, text emojis and URLs; and by modifying mentions and hashtags (remove starting characters of '@' and '#'). This procedure removes the text noise and finally allows the identification of the entity that was discussed by the users. The next step includes the transformation of the text to lower case and tokenization of the collected tweets. Then, we perform the lemmatization of each token. This technique normalizes the inflected word forms. As a result of the previous steps, this sentiment analysis will be performed on lower-cased and normalized sentences.

3.2 Sentiment analysis

There is a plethora of previous studies as mentioned in section 1.1 that analyses the sentiment on Twitter. Vader is broadly used in this domain, as shown in [117–121].

In our implementation of sentiment analysis, we utilize Vader [122] sentiment analysis model from python NLTK library [123]. We choose this particular implementation because it is especially attuned to the sentiment expressed on social media. Vader uses a list of lexical features that are labeled according to their semantic orientation. Thanks to the implementation simplicity, sentiment analysis execution time remains low. Utilizing the already developed sentiment solution, we need to perform text filtering and parse our data to the SentimentIntensityAnalyzer function that returns the scores of positive, negative, and neutral sentiment types. In our analysis, we compute such sentiment for each collected tweet and comment in our database and summarize those sentiment scores daily for each particular user. We compute two values of daily sentiment scores per user: summary, where the daily user sentiment is a sum

product of scores of tweets that was created by the user each day; and the average score where the summary score is divided by the number of tweets where a particular entity was mentioned by the user.

The sentiment analysis of the corpus develops only the emotion vector of a particular sentence and not the emotion of a particular entity. In order to continue our pipeline, we will need the sentiment of each entity. In the next subsection, we explore how we identify the entities.

3.3 The entities and their sentiment

In order to identify the entity that was contained in each tweet, we generate the set of keywords, as shown in Table 2, for the particular dataset entities ('Trump' and 'Biden'). The keywords are being searched in lower case in order to avoid any misspellings and use upper-lower case writing style. We are matching those keywords in each tweet from the Twitter dataset and in each comment in the YouTube dataset, to identify whether the entity was used on the specific tweet/comment text. The entities of J. Biden and D. Trump were identified with a computed set of keywords (including the hashtags enriched with the candidates' names and the parties they are representing). For example hashtag #VoteBlue, as a text does not contain the word Biden but is associated with J. Biden campaign. We manage to recognize it because of our keywords shown in the Table 2. For this reason, we use as keywords the hashtags correlated with each candidate. This approach increases the accuracy of the association of a particular tweet/user with the described entities.

Each time the keyword is found in a particular tweet or comment, we assign the sentiment values on those entities to the user who posted this particular tweet. These entity sentiments are assigned to the users daily, since we are interested in the identification of the user/community dynamically and present how the entity sentiment evolves day by day.

4 Results

4.1 Volume analysis

In this Section, we include several volume measures derived from our dataset. Initially, we perform a volume analysis on the whole corpus of the tweets. In Fig 3 we plot the cumulative distribution of the daily tweets.

In Fig 1 we plot the number of daily tweets and video comments, where we notice the diurnal increases of posts and the massive increase of the number of tweets, on the day of the elections. In Fig 4 we can see the cumulative distribution of the daily tweets per hashtag, where we notice the prevailing hashtags of #Debate2020, #Trump, #BidenHarris2020, #VoteBluetoSaveAmerica and #Biden.

We plot the daily active users per entity, for both Twitter and YouTube, in Fig 5 for each entity of 'Trump' and 'Biden' and we notice that the total number of users is increasing from July to November, the top hashtag is the #Debate 2020 and that the users posting on Twitter and commenting on YouTube for Trump, exceed the users for Biden. The apogee of user activities is presented on the dates of the 29th of September and on the 2nd of October. These user

Table 2. The complete list of all entities with the corresponding keywords that were used for each one.

Entity	Key words
Trump	trump, donald, donaldtrump, trump2020, votetrumpout, trumpislosing, dumptrump, nevertrump, republican
Biden	biden, joseph, bluewave2020, votebluetosaveamerica, voteblue, ridinwithbiden, neverbiden, democrat

<https://doi.org/10.1371/journal.pone.0270542.t002>

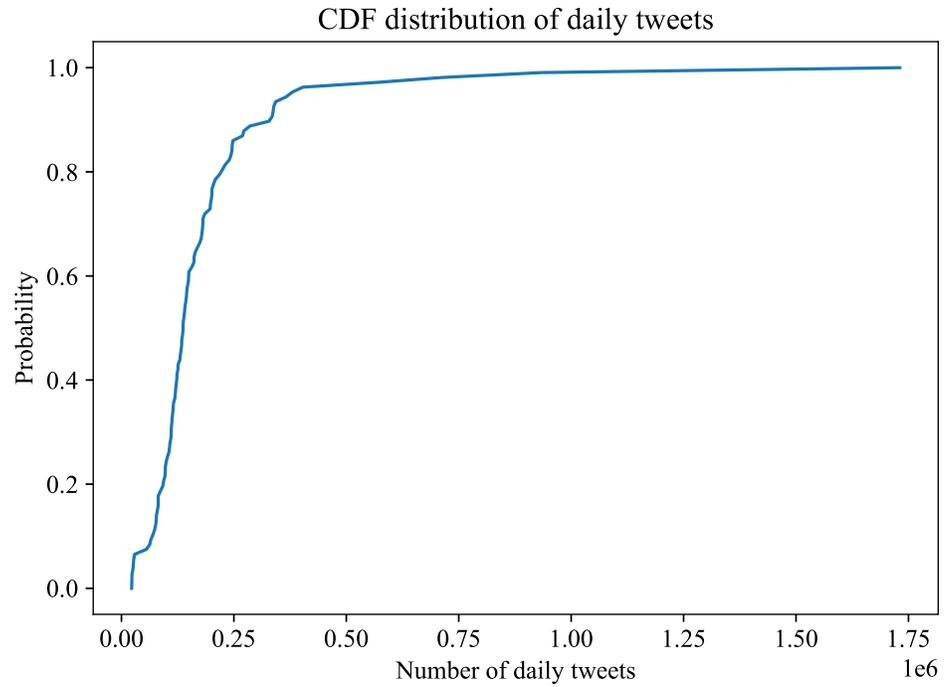


Fig 3. CDF distribution of tweets per day.

<https://doi.org/10.1371/journal.pone.0270542.g003>

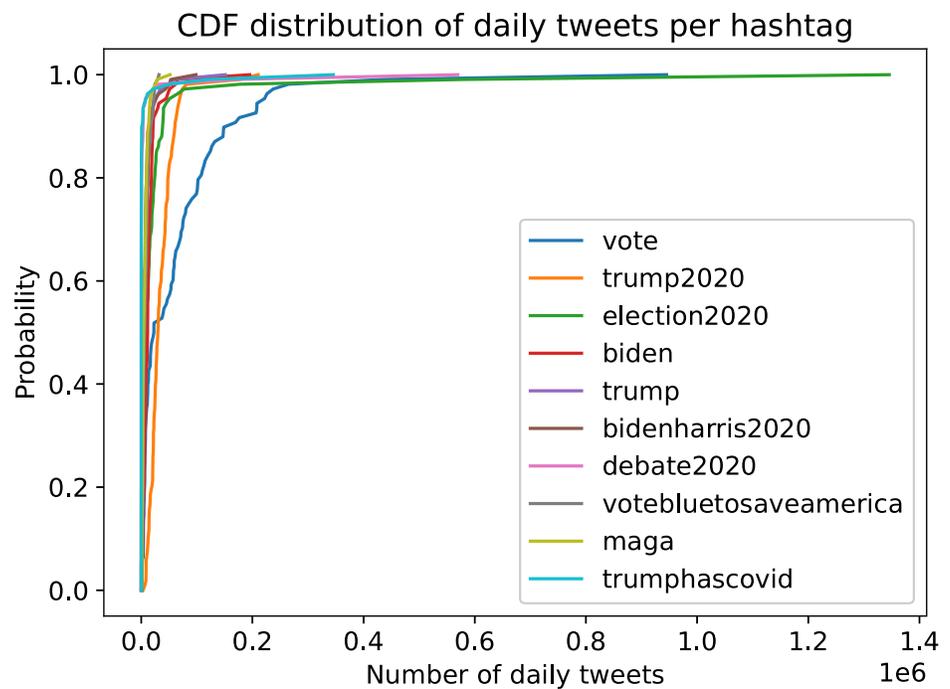


Fig 4. CDF distribution of the daily tweets per hashtag.

<https://doi.org/10.1371/journal.pone.0270542.g004>

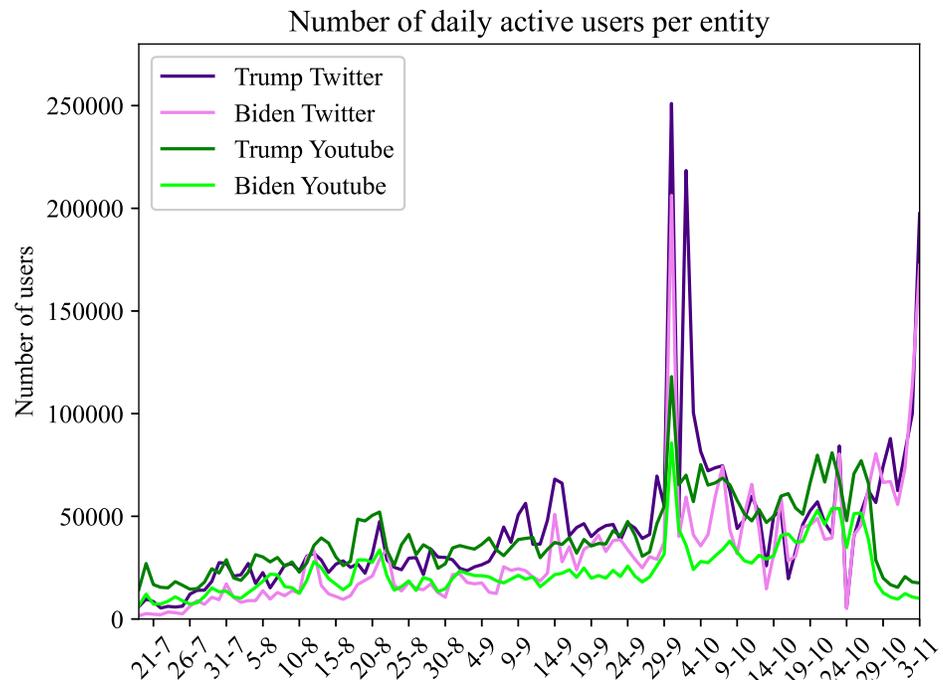


Fig 5. In this figure, we plot the number of daily users that post YouTube comments and tweets where particular entities are discussed.

<https://doi.org/10.1371/journal.pone.0270542.g005>

activity anomalies are explained in more detail in Section: 4.4, where the increased interest of users is correlated with election events. We also notice that the total Twitter traffic overcomes the total YouTube traffic.

4.2 The retweet graph

In Figs 6–11 we plot the retweet graph. This graph is developed based on the retweet relationships between all users in the collected dataset. We represent the retweet relation as a directed edge between two users (nodes). In those directed relations the destination node is the original user who posted the tweet and the source node is the user who actually retweeted that particular tweet. Such representation allows the identification of the user communities that share the same information source. To show how tight the relations are, we assign the number of seen retweets for a particular destination node, as an edge weight. We also perform filtering, by removing the non-significant edges, with weights less than 8. The filtering procedure reduces the noise volume of the non-significant relations and also reduces the number of edges that are not manageable for graph visualization. Before filtering, our retweet graph consisted of 2.874.090 nodes and 9.993.122 edges and after the filtering of the non-significant edges, we reduced the number of nodes to 56.853 and the number of edges to 86.668. For the entity visualization, we use 2 colors for our entities; in red we represent the entity of Trump, and in blue the entity of Biden.

We apply sentiment analysis for each day and we measure the number of tweets a user posts containing a particular entity. We use this counter to provide coloring of the user node by selecting the most popular entity of the user at each particular date. We also develop a plot (Fig 5) with the volume of users that allows us to compare the daily volume per entity. Users that don't use any entities in their tweets remain without a particular coloring.

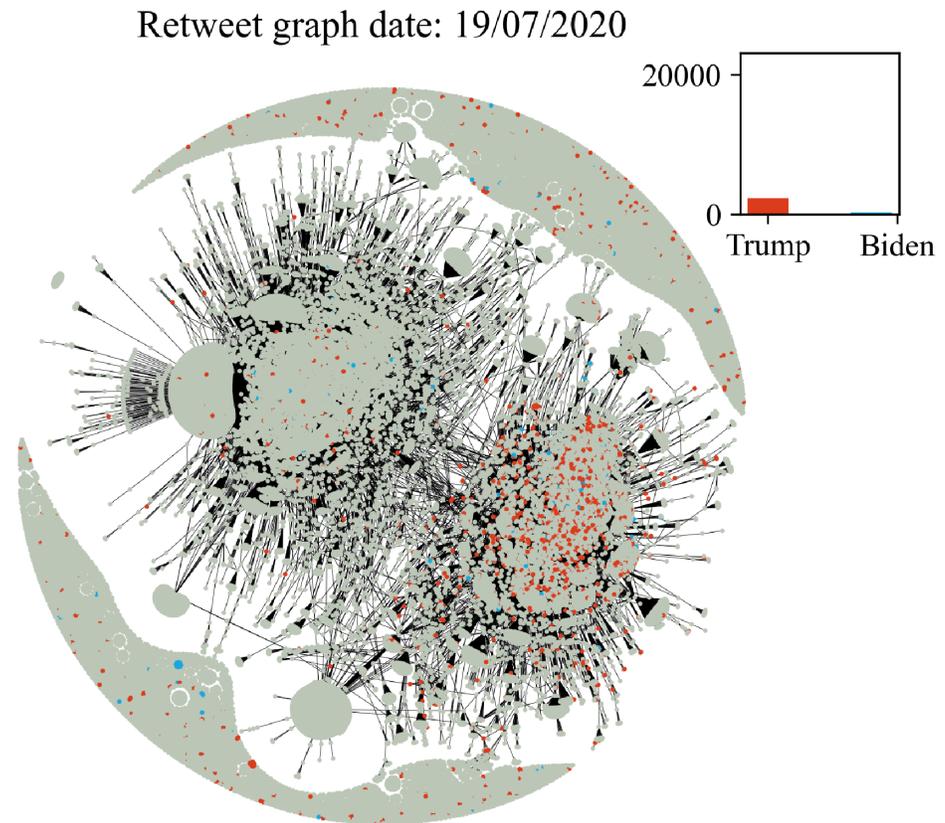


Fig 6. Retweet graph relation between users in our dataset. Colours represent entity activity by users. Red colour entity Trump and blue color entity Biden.

<https://doi.org/10.1371/journal.pone.0270542.g006>

The graph plots in Figs 6–11 were generated with Gephi Furchterman Reingold layout [124], while we export Gephi generated positions and we use them to generate a daily graph with networkX python library [125]. In Figs 6–11 we notice the largest connected component acquires color in the period closer to the elections, while other cores are formed and the mentions of the two main candidates increases or in other periods decrease potentially influenced by real events like the debate (see Section 4.4). We also show how the discussion around those entities obtains more resonance in the period closer to the elections. Also, in Tables 3 and 4 we show the highest retweeted users, with the highest numbers of in-degree and out-degree respectively, with anonymized usernames.

The retweet graphs show the volume of the tweets and the relation between the nodes which represent the users. This apposition of the graphs through the whole pre-election period, demonstrates how the volume of the communication and interaction between the involved users evolves. As we expected, the density of the graph is increasing, which means the volume of the tweets increases, and the discourse is getting more intense, as we approach the period close to the elections.

The discussion around each candidate evolves through time as well. The corresponding hashtag for each candidate is mentioned as long as the electorate references them in the discourse. Of course, this does not necessarily mean a preference towards Trump or Biden. As mentioned in section 2, this analysis does not focus on the prediction of the preference of the electorate or the outcome of the elections; we expect to see how users engage in the online conversation. As we get closer to the election date, we notice some components being formed in

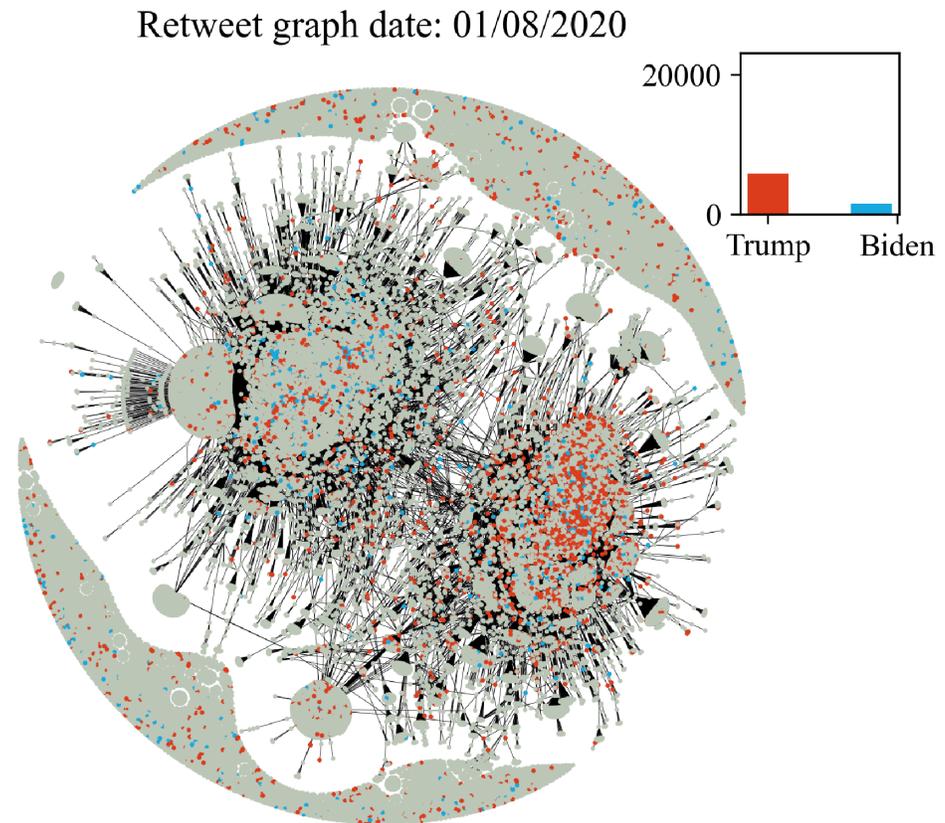


Fig 7. Retweet graph relation between users in our dataset. Colours represent entity activity by users. Red colour entity Trump and blue color entity Biden.

<https://doi.org/10.1371/journal.pone.0270542.g007>

the graph that show the conversations about each candidate and the conversation about both of them in Fig 11 in the last bottom figure (Fig 11).

Additionally, we see two main graph components throughout the whole period that change the colour according to the daily events. For example, in the bottom left subgraph (Fig 10) we notice that the colour is red in both components, while the bottom right Fig 11 one is painted blue (from J. Biden). This happens because the left graph highlights the fact that Trump was infected with COVID-19 on 2/10, and the conversation about Trump was more intense and increased in terms of tweets. Also in the middle right plot we show the day after the debate, whose mixed red-blue colour makes sense if we consider the conversations this particular debate initiated. The colours as expected seem to resume back to normal on the last bottom right (3/11, which corresponds to the main two components blue and red, one for each candidate, since on this date none of the candidates seem to attract any peculiar attention.

4.3 Sentiment analysis

In this Section, we present the results from the sentiment analysis in our corpus, as described in 3.2. In Fig 12 we present the daily average sentiment for the entity 'Biden' and 'Trump'. The solid line is the average sentiment on YouTube comments and the dotted line is the average sentiment in the corpus of the tweets. Below zero we have the negative sentiment and above zero the positive sentiment for each social media. In Fig 13, we plot the daily overall sentiment for entity Biden and 'Trump'.

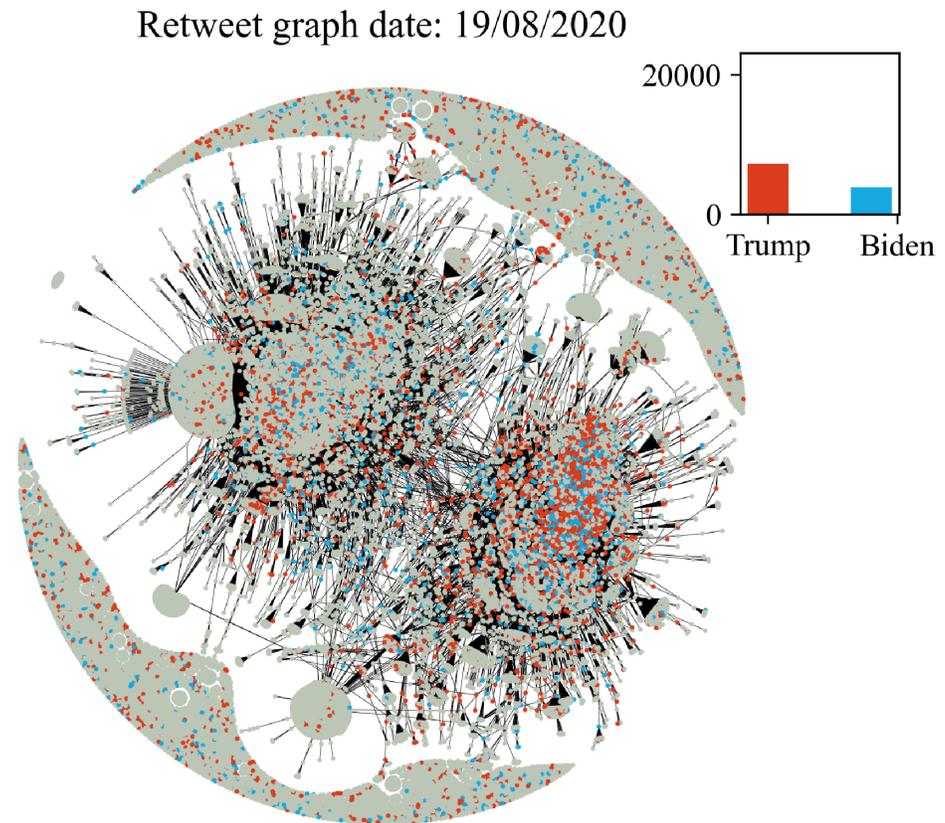


Fig 8. Retweet graph relation between users in our dataset. Colours represent entity activity by users. Red colour entity Trump and blue color entity Biden.

<https://doi.org/10.1371/journal.pone.0270542.g008>

Additionally, in Fig 14 we plot the positive sentiment time series for the two sets of hashtags for each state. In order to identify the location of the user accounts, we extract this information from the corresponding field named 'location' in the Twitter user object. Additionally, we append this information with the location field in the tweet object of the Twitter API. We acknowledge the fact that the first field may be not updated by the user or the second field could be missing because the user does not share the location. This means that the location for many users could be missing. Nevertheless, this is the closest information we have from Twitter about the user location. We notice the daily fluctuations for every state per entity (blue is the entity 'Biden' and red is for 'Trump'). The juxtaposition of the time series in the form resembling an EEG makes it easier to discern localized events from nationwide Twitter traffic. The list of state abbreviations can be found here: [126].

4.4 Event consequences

Since our work is based on the analysis of the 2020 US Presidential elections we monitor real-world events that may trigger significant user interest on social media. Interesting examples of such events are the candidates' debate on TV (September 29 and October 22) and the date when President Donald Trump was diagnosed positively for COVID-19 (October 2). The depiction of the online conversations regarding these events is visible within our analysis. Analyzing the user engagement of these specific periods, one day before and one day after, shows whether such real-world events are connected in the virtual world of social networks.

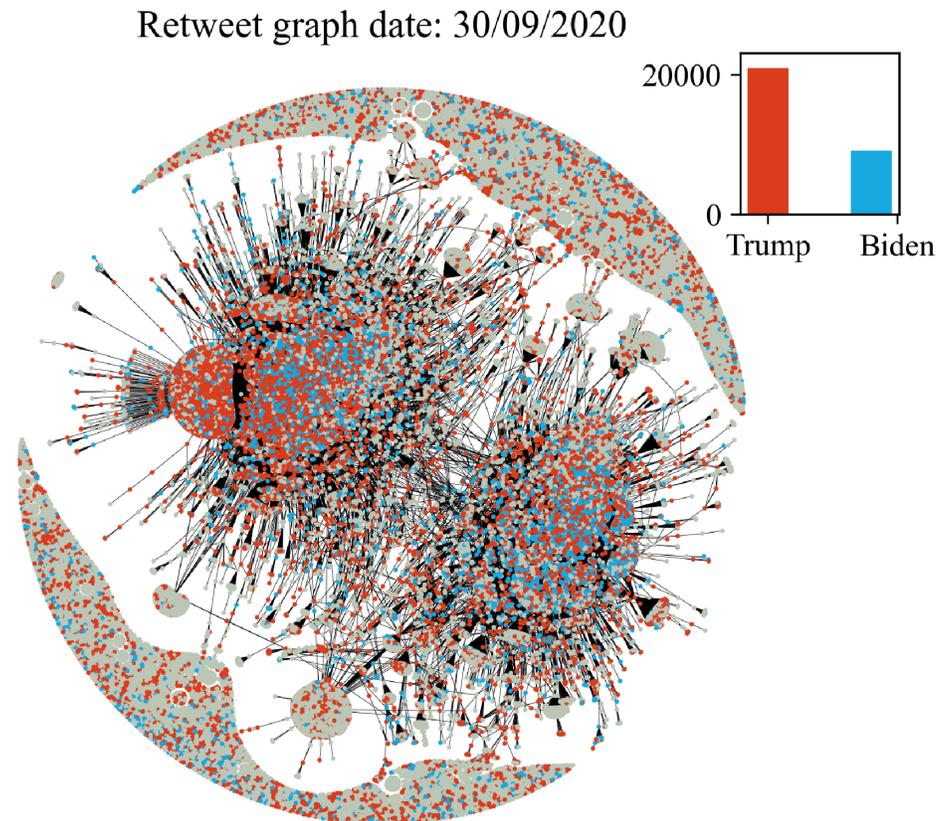


Fig 9. Retweet graph relation between users in our dataset. Colours represent entity activity by users. Red colour entity Trump and blue color entity Biden.

<https://doi.org/10.1371/journal.pone.0270542.g009>

We use sentiment analysis on these specific time points of our dataset timeline to allow us to identify how the social media users react to those occasions, identify the fluctuations of the sentiment and measure the volume around each entity topic.

Our results are presented in Figs 15 and 16, where it is noticeable that the first debate and Donald Trump COVID-19 announcement events generated a high volume of user interest on social media. In the first case (debates) both entities are soared, in comparison with previous dates. At the second event, the volume of the entity ‘Trump’ is taking first place during the user discussions on social media by increasing the volume of tweets for the entity ‘Trump’ and by presenting high dissonance on sentiment values.

5 Linking Twitter and YouTube data

In this section, we explore the differences in the discussion and the community between the two social networks. We perform Louvain community detection on both social graphs, we associate the communities in the YouTube comment graph with the communities in the Twitter retweet graph and measure their similarity and differences. Fig 17 shows the 3-core of the YouTube comment graph, Fig 18 shows the 3-core of the Retweet graph. In both graphs, we have color-coded communities and labeled the top-PageRank nodes of each community.

We used Gephi [124] for the analysis and visualization of Figs 17 and 18.

Fig 17 illustrates the 3-core YouTube comment graph that represents the relation of which channel has commented on which channel. Different colored areas (purple, light grey, green

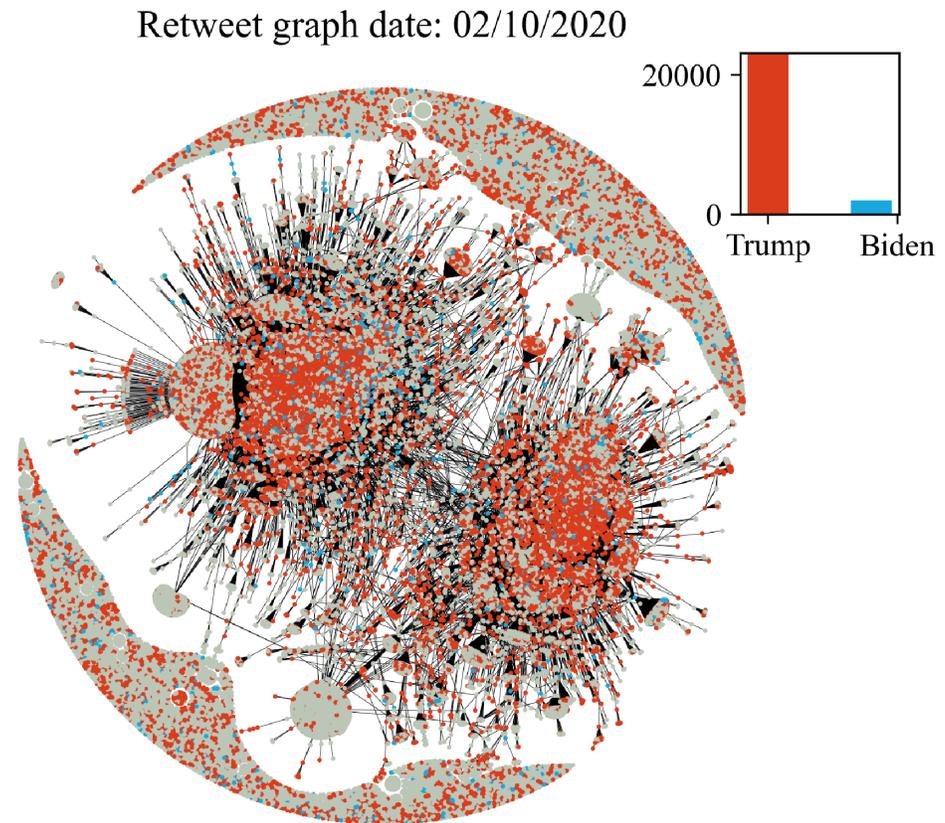


Fig 10. Retweet graph relation between users in our dataset. Colours represent entity activity by users. Red colour entity Trump and blue color entity Biden.

<https://doi.org/10.1371/journal.pone.0270542.g010>

blue) in the graph represent different communities that are formed between the YouTube channels, produced by the Louvain algorithm [24]. For example purple indicates the community between the “Blaze TV”, “The Officer Tatum” and the “Fox News”. However “Fox News” is also between the purple and the grey, which shows the next community (formed by “Fox News” and “Donald J Trump”). These communities do not necessarily correspond to a real life community; they are a result of the Louvain algorithm of Gephi. Regarding the labels, in this figure, we also show the 13 most popular channels after running PageRank.

Similarly, in plot 18 we show the 3-core Retweet graph that represents the relation between users that have retweeted. In the graph we notice seven colored areas of different sizes, indicating different communities, that as in plot 17, do not necessarily correspond to a real life community. Also, the text labels shown, for example, “MaryLTrump” and “DrDenaGrayson” seem to be close, which again cannot be translated to semantic information. It is considered coincidental to be close but not coincidental to be in the same color-area (community).

In both plots (node diagrams Figs 17 and 18) it is not possible to keep readable names of each particular node, through the high density of nodes. The nodes in Fig 18 are the users on Twitter who have posted the tweets included in this plot and similarly, the nodes in Fig 17 are the users of YouTube which are channels. In order to highlight only the important nodes, we manage to provide the names for the most significant nodes based on the PageRank score.

For illustration purposes, nodes in both plots (17,18) with high degrees are shaped like circles (lightly highlighted circle-like scheme) in this plot, but are variables in Gephi that unfortunately cannot be explained semantically in this context and should be ignored.

Retweet graph date: 03/11/2020

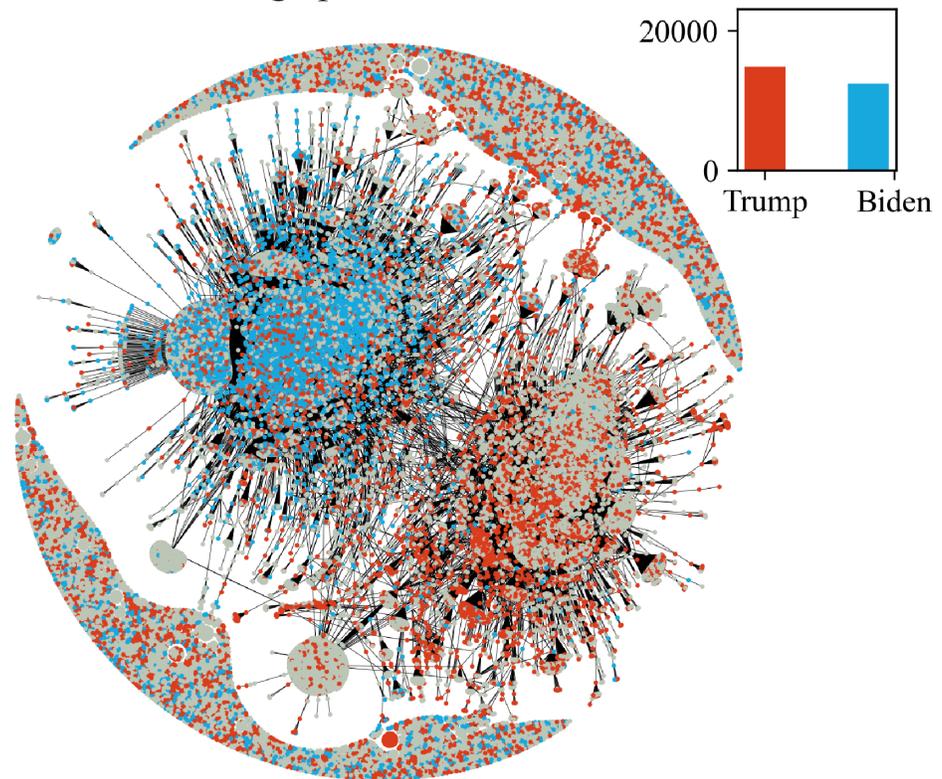


Fig 11. Retweet graph relation between users in our dataset. Colours represent entity activity by users. Red colour entity Trump and blue color entity Biden.

<https://doi.org/10.1371/journal.pone.0270542.g011>

Fig 19 shows the interactions between the 3 largest YouTube communities (top half) in the YouTube-comment graph (YT) and the 6 largest Twitter communities in the Retweet graph (RT). Each community is named after its highest PageRank member (or the second highest, when more clear) in the corresponding graph. The size of each relation depicts the number of

Table 3. Top retweeted users (highest indegree).

User	Number of Followings	Number of Followers
User 1	1.03K	826.4K
User 2	5.1K	2.7M
User 3	97	248.3K

<https://doi.org/10.1371/journal.pone.0270542.t003>

Table 4. Top retweeted users (highest out degree).

User	Number of Followings	Number of Followers
User 1	4.7K	5K
User 2	3.6K	3.6K
User 3	5K	8.1K

<https://doi.org/10.1371/journal.pone.0270542.t004>

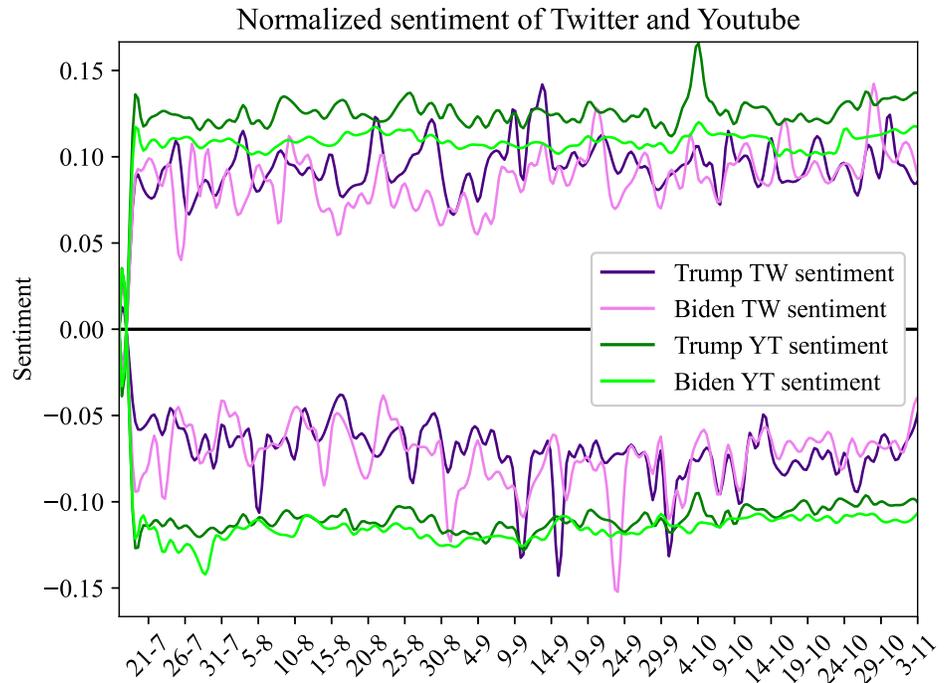


Fig 12. Result of daily average sentiment per entity for Twitter and YouTube collected dataset.

<https://doi.org/10.1371/journal.pone.0270542.g012>

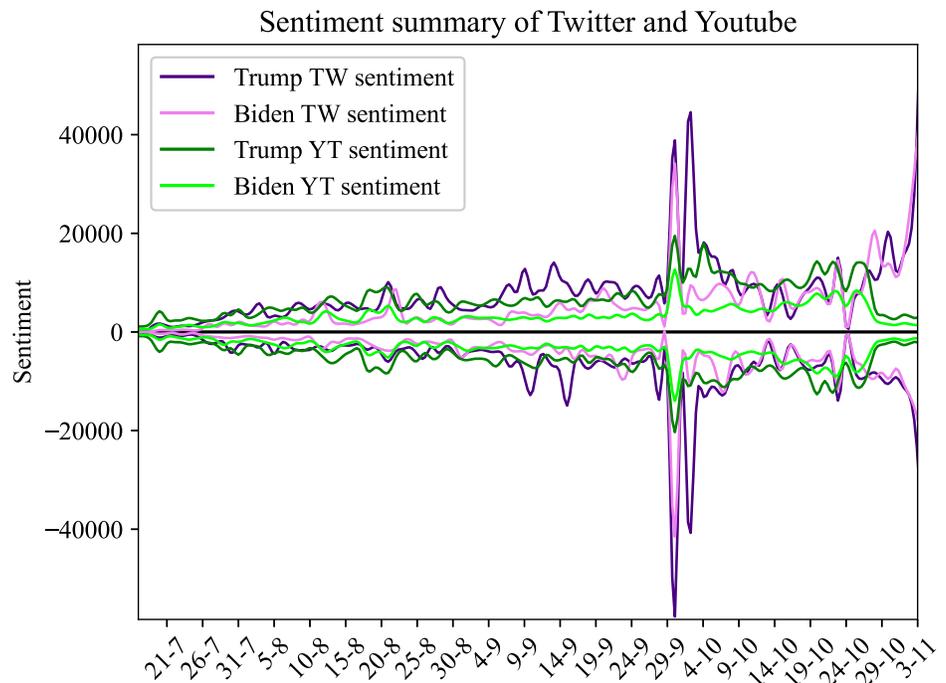


Fig 13. Result of overall sentiment analysis for Twitter and Youtube. Particular sentiment values present the overall sentiment that was posted by users at given dates.

<https://doi.org/10.1371/journal.pone.0270542.g013>

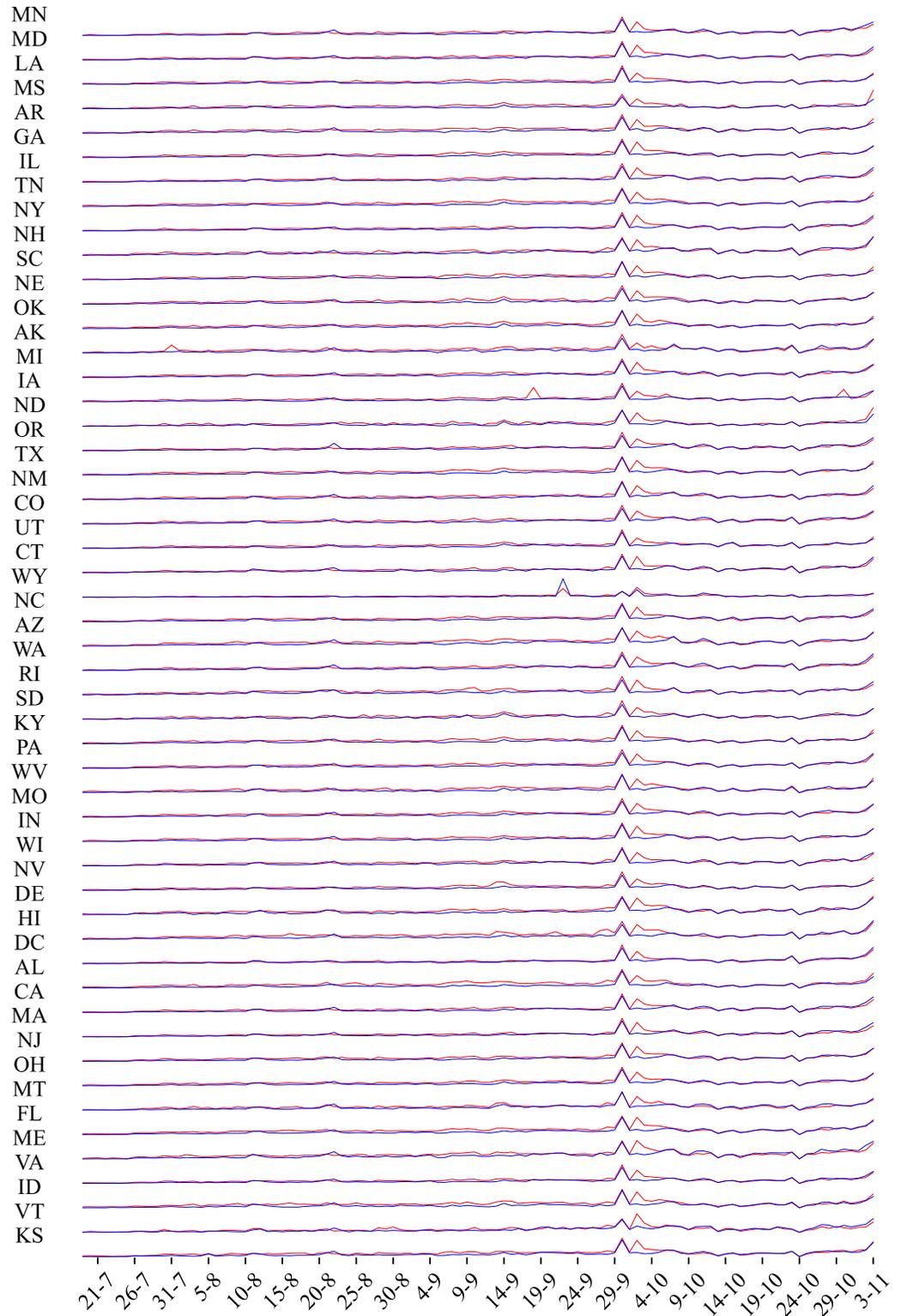


Fig 14. Positive sentiment time series for the two sets of hashtags for each state.

<https://doi.org/10.1371/journal.pone.0270542.g014>

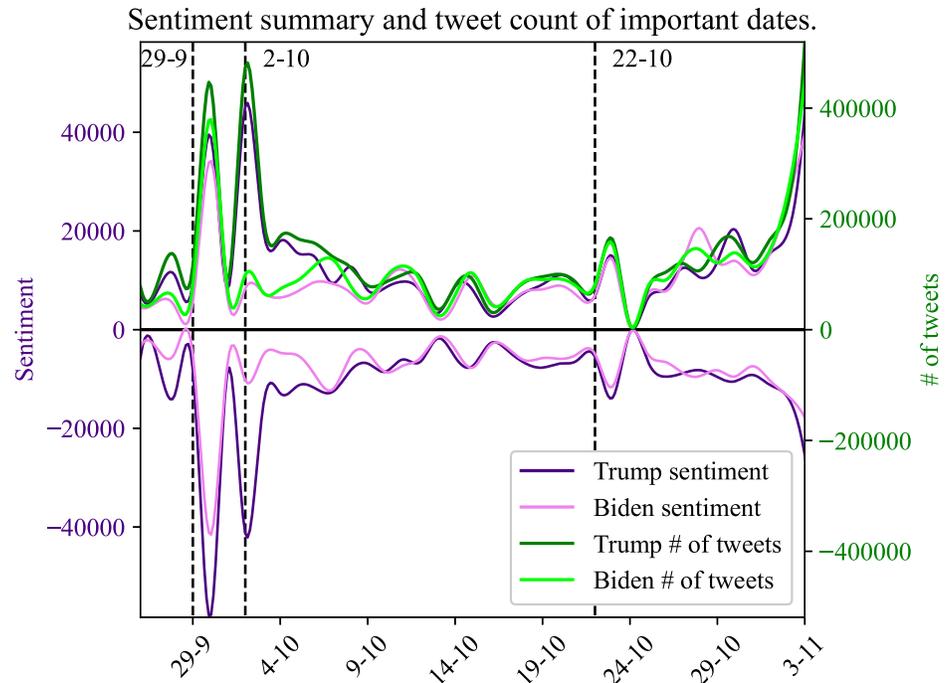


Fig 15. Summary of user tweets sentiment values per each entity followed by the number of tweets. In this plot, we mention 3 important dates in the dataset, where 29/09 is the date of the first debate, 2/10 the date when Donald Trump was tested positive for COVID-19, and 22/10 the date of the last election debate.

<https://doi.org/10.1371/journal.pone.0270542.g015>

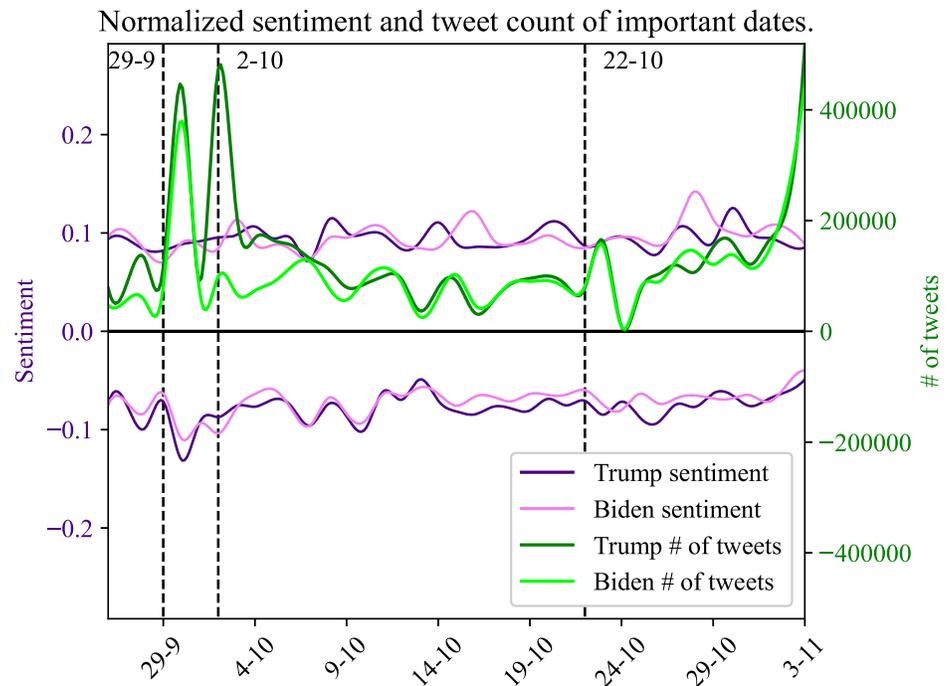


Fig 16. Normalized sentiment values of user tweet per each entity followed by the number of tweets. In this plot, we mention 3 important dates in the dataset, where 29/09 is the date of the first debate, 2/10 the date when Donald Trump was tested positive to COVID-19, and 22/10 the date of the last election debate.

<https://doi.org/10.1371/journal.pone.0270542.g016>

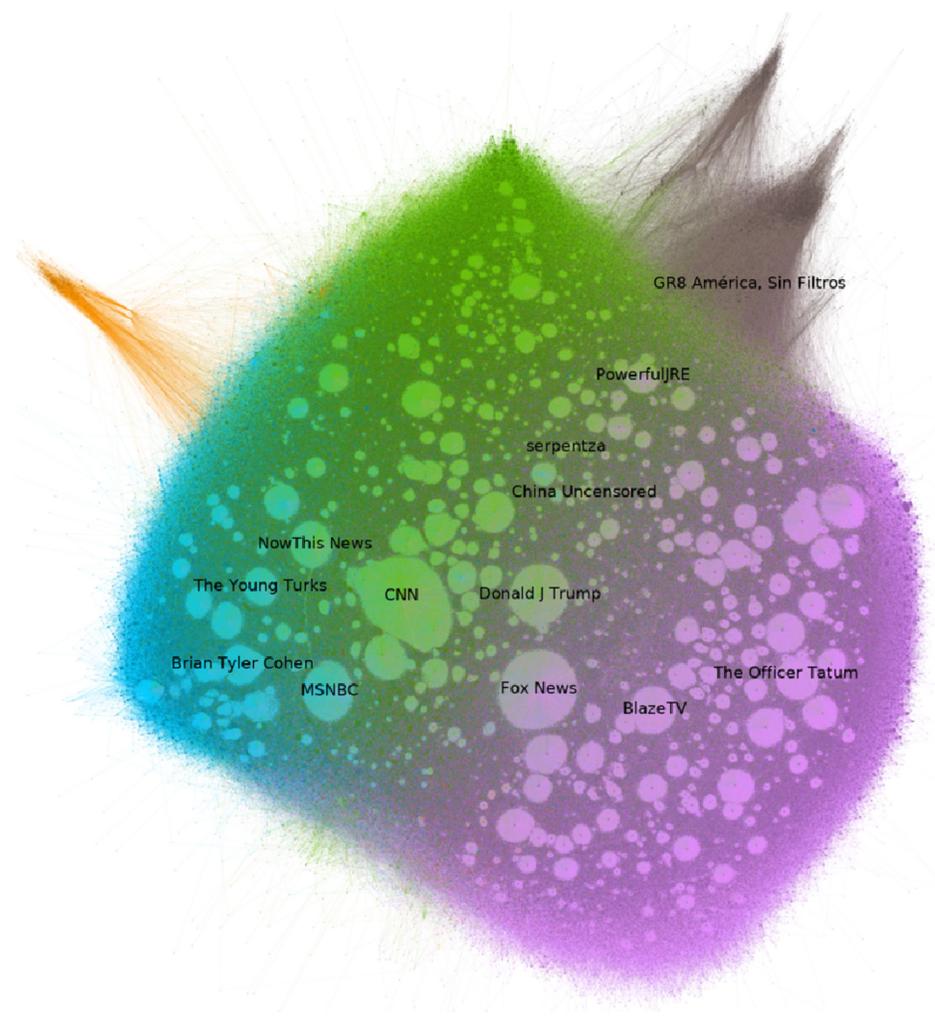


Fig 17. The 3-core of the YouTube comment graph, color-coded by the community.

<https://doi.org/10.1371/journal.pone.0270542.g017>

users in the RT community, that tweeted video URLs from any channel in the YT community. As expected, communities affiliated with the Trump campaign on Twitter, mostly post video links from channels in the Fox News YouTube community, whereas members of the Twitter community mostly affiliated with the Biden campaign post mostly video links from YouTube channels in the CNN YouTube community.

One Twitter community that seems to follow an unexpected pattern is mostly formed by Arabic speakers who, in this setting, tend to use Twitter more as a news media and less as a social network for disseminating political opinions. This observation supports linking the Twitter Retweet graph with the YouTube comment graph using tweets of posted videos.

6 Conclusions

The purpose of this study is to shed light on the online discourse on social media, happening during the US presidential elections of November 2020. We focus on two main social media Twitter and YouTube. This work having obtained the tweets for the most popular hashtags

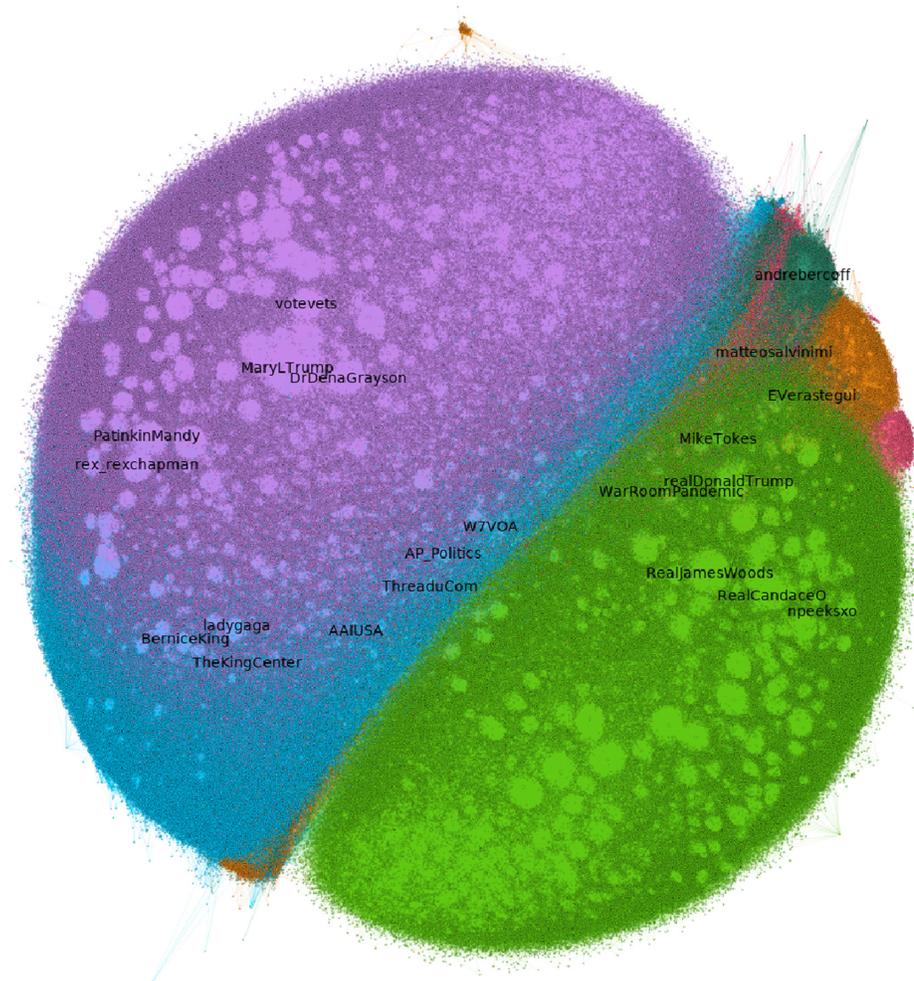


Fig 18. The 3-core of the retweet graph, color-coded by community.

<https://doi.org/10.1371/journal.pone.0270542.g018>

regarding the US elections 2020, as well as the extracted unique YouTube videos, performs an analysis to study the connection between these two social networks, by following a series of steps including: volume analysis of tweets and users, identification of entities, and correlation between the features of the YouTube videos. Next, we apply sentiment analysis on the Twitter corpus and the YouTube metadata and show that the positive sentiment is higher for Donald Trump in comparison with Joe Biden. We identify how real-world events trigger user discussions on social media around the elections.

The next step is to study the Retweet graph across six different time points on the dataset, from July to September 2020, highlight the two main entities ('Biden' and 'Trump') and show how the main connected component becomes denser through time. Finally, the results from the previous steps of the analysis allow us to link the Twitter Retweet graph with the YouTube comment graph and show the interactions between the 3 largest YouTube communities, in the YouTube-comment graph and the 6 largest Twitter communities in the Retweet graph (RT). We include sarcasm detection in our future plans, since it will require a crowd-sourcing technique after the election period, in order to form an adequate ground truth dataset for the training process.

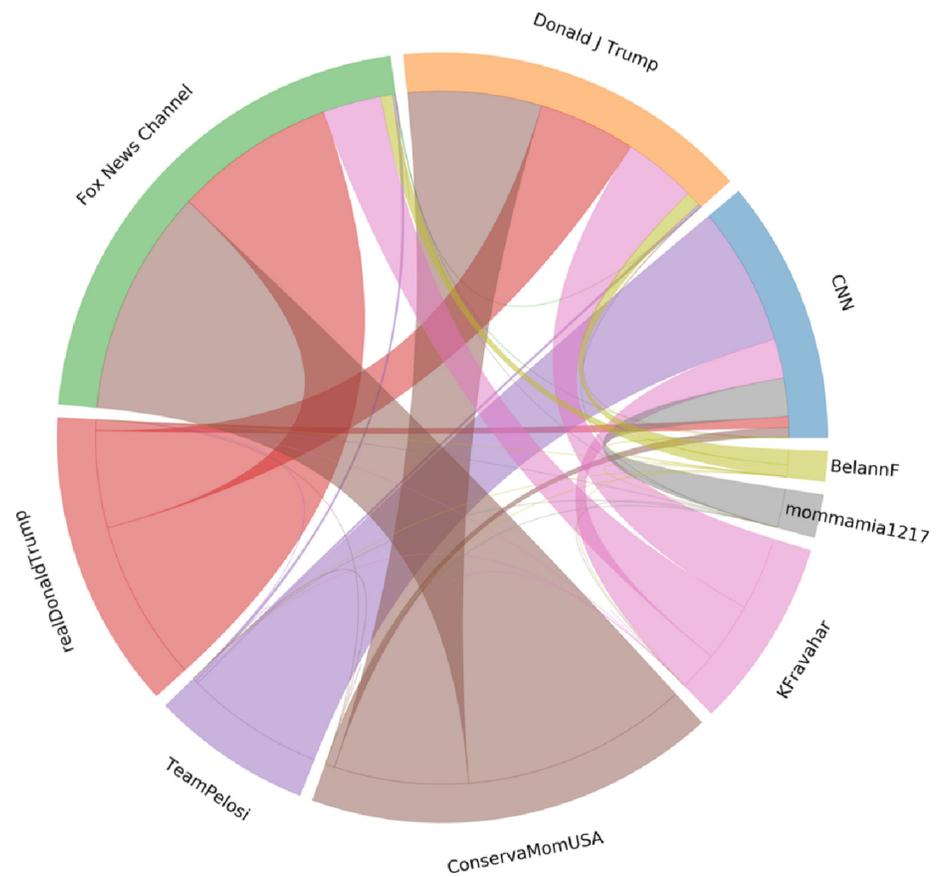


Fig 19. Interactions between largest communities on YouTube and Twitter.

<https://doi.org/10.1371/journal.pone.0270542.g019>

Supporting information

S1 Appendix. List of all hashtags.

(PDF)

S1 Text.

(TXT)

S1 Data.

(XLSX)

Author Contributions

Conceptualization: Alexander Shevtsov, Maria Oikonomidou, Despoina Antonakaki, Polyvios Pratikakis.

Data curation: Alexander Shevtsov, Maria Oikonomidou, Despoina Antonakaki, Polyvios Pratikakis.

Formal analysis: Alexander Shevtsov, Maria Oikonomidou, Despoina Antonakaki, Polyvios Pratikakis.

Funding acquisition: Polyvios Pratikakis, Sotiris Ioannidis.

Investigation: Alexander Shevtsov, Maria Oikonomidou, Despoina Antonakaki, Polyvios Pratikakis.

Methodology: Alexander Shevtsov, Maria Oikonomidou, Despoina Antonakaki, Polyvios Pratikakis.

Project administration: Polyvios Pratikakis, Sotiris Ioannidis.

Resources: Polyvios Pratikakis, Sotiris Ioannidis.

Software: Alexander Shevtsov, Maria Oikonomidou, Polyvios Pratikakis.

Supervision: Despoina Antonakaki, Polyvios Pratikakis, Sotiris Ioannidis.

Validation: Alexander Shevtsov, Maria Oikonomidou, Despoina Antonakaki, Polyvios Pratikakis.

Visualization: Alexander Shevtsov, Maria Oikonomidou, Polyvios Pratikakis.

Writing – review & editing: Alexander Shevtsov, Maria Oikonomidou, Despoina Antonakaki, Polyvios Pratikakis.

References

1. Howard PN, Woolley S, Calo R. Algorithms, bots, and political communication in the US 2016 election: The challenge of automated political communication for election law and administration. *Journal of information technology & politics*. 2018; 15(2):81–93. <https://doi.org/10.1080/19331681.2018.1448735>
2. Baumgartner JC, Mackay JB, Morris JS, Otenyo EE, Powell L, Smith MM, et al. Communicator-in-chief: How Barack Obama used new media technology to win the White House. Lexington Books; 2010.
3. Antonakaki D, Spiliotopoulos D, Samaras V C, Pratikakis P, Ioannidis S, Fragopoulou P. Social media analysis during political turbulence. *PloS one*. 2017; 12(10):e0186836. <https://doi.org/10.1371/journal.pone.0186836> PMID: 29088263
4. Yaqub U, Chun SA, Atluri V, Vaidya J. Analysis of political discourse on twitter in the context of the 2016 US presidential elections. *Government Information Quarterly*. 2017; 34(4):613–626. <https://doi.org/10.1016/j.giq.2017.11.001>
5. Enli G. Twitter as arena for the authentic outsider: exploring the social media campaigns of Trump and Clinton in the 2016 US presidential election. *European Journal of Communication*. 2017; 32(1):50–61. <https://doi.org/10.1177/0267323116682802>
6. Gong Z, Cai T, Thill JC, Hale S, Graham M. Measuring relative opinion from location-based social media: A case study of the 2016 US presidential election. *Plos one*. 2020; 15(5):e0233660. <https://doi.org/10.1371/journal.pone.0233660> PMID: 32442212
7. Huang M, Zou G, Zhang B, Liu Y, Gu Y, Jiang K. Overlapping community detection in heterogeneous social networks via the user model. *Information Sciences*. 2018; 432:164–184. <https://doi.org/10.1016/j.ins.2017.11.001>
8. Golovchenko Y, Buntain C, Eady G, Brown MA, Tucker JA. Cross-platform state propaganda: Russian trolls on twitter and youtube during the 2016 US presidential election. *The International Journal of Press/Politics*. 2020; 25(3):357–389. <https://doi.org/10.1177/1940161220912682>
9. Faralli S, Stilo G, Velardi P. Large scale homophily analysis in twitter using a twixonomy. In: Twenty-Fourth International Joint Conference on Artificial Intelligence; 2015.
10. Hashtag homophily in twitter network: Examining a controversial cause-related marketing campaign. *Computers in Human Behavior*. 2020; 102:87–96. <https://doi.org/10.1016/j.chb.2019.08.006>
11. Kang JH, Lerman K. Using lists to measure homophily on twitter. In: Workshops at the twenty-sixth AAAI conference on artificial intelligence; 2012.
12. Barbieri N, Bonchi F, Manco G. Who to follow and why: link prediction with explanations. In: Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining; 2014. p. 1266–1275.

13. Colleoni E, Rozza A, Arvidsson A. Echo chamber or public sphere? Predicting political orientation and measuring political homophily in Twitter using big data. *Journal of communication*. 2014; 64(2):317–332. <https://doi.org/10.1111/jcom.12084>
14. Himelboim I, Sweetser KD, Tinkham SF, Cameron K, Danelo M, West K. Valence-based homophily on Twitter: Network analysis of emotions and political talk in the 2012 presidential election. *New media & society*. 2016; 18(7):1382–1400. <https://doi.org/10.1177/1461444814555096>
15. Just MR, Crigler AN, Metaxas P, Mustafaraj E. 'It's Trending on Twitter'-An Analysis of the Twitter Manipulations in the Massachusetts 2010 Special Senate Election. In: APSA 2012 Annual Meeting Paper; 2012.
16. Guo L, Rohde A J, Wu HD. Who is responsible for Twitter's echo chamber problem? Evidence from 2016 US election networks. *Information, Communication & Society*. 2020; 23(2):234–251. <https://doi.org/10.1080/1369118X.2018.1499793>
17. Vergeer M. Twitter and political campaigning. *Sociology compass*. 2015; 9(9):745–760. <https://doi.org/10.1111/soc4.12294>
18. Plotkowiak T, Stanoevska-Slabeva K. German politicians and their Twitter networks in the Bundestag election 2009. *First Monday*. 2013;. <https://doi.org/10.5210/fm.v18i5.3816>
19. Nahon K. Where there is social media there is politics. In: *The Routledge companion to social media and politics*. Routledge; 2015. p. 39–55.
20. Vaccari C, Valeriani A, Barberá P, Jost JT, Nagler J, Tucker JA. Of echo chambers and contrarian clubs: Exposure to political disagreement among German and Italian users of Twitter. *Social media+ society*. 2016; 2(3):2056305116664221.
21. Rosenbusch H, Evans AM, Zeelenberg M. Multilevel emotion transfer on YouTube: Disentangling the effects of emotional contagion and homophily on video audiences. *Social Psychological and Personality Science*. 2019; 10(8):1028–1035. <https://doi.org/10.1177/1948550618820309>
22. Ladhari R, Massa E, Skandrani H. YouTube vloggers' popularity and influence: The roles of homophily, emotional attachment, and expertise. *Journal of Retailing and Consumer Services*. 2020; 54:102027. <https://doi.org/10.1016/j.jretconser.2019.102027>
23. Wattenhofer M, Wattenhofer R, Zhu Z. The YouTube social network. In: *Sixth international AAAI conference on weblogs and social media*; 2012.
24. Blondel VD, Guillaume JL, Lambiotte R, Lefebvre E. Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment*. 2008; 2008(10):P10008. <https://doi.org/10.1088/1742-5468/2008/10/P10008>
25. Martínez-Cámara E, Martín-Valdivia MT, Urena-López LA, Montejo-Ráez AR. Sentiment analysis in Twitter. *Natural Language Engineering*. 2014; 20(1):1–28. <https://doi.org/10.1017/S1351324912000332>
26. Giachanou A, Crestani F. Like it or not: A survey of twitter sentiment analysis methods. *ACM Computing Surveys (CSUR)*. 2016; 49(2):1–41. <https://doi.org/10.1145/2938640>
27. Go A, Huang L, Bhayani R. Twitter sentiment analysis. *Entropy*. 2009; 17:252.
28. Mittal A, Goel A. Stock prediction using twitter sentiment analysis. Stanford University, CS229 (2011 <http://cs229.stanford.edu/proj2011/GoelMittal-StockMarketPredictionUsingTwitterSentimentAnalysis.pdf>). 2012;15:2352.
29. Saif H, He Y, Alani H. Alleviating data sparsity for twitter sentiment analysis. In: *CEUR Workshop proceedings*. CEUR Workshop Proceedings (CEUR-WS.org). Lyon, France.: CEUR; 2012. p. 297–312.
30. Wang H, Can D, Kazemzadeh A, Bar F, Narayanan S. A system for real-time twitter sentiment analysis of 2012 US presidential election cycle. In: *Proceedings of the ACL 2012 system demonstrations*. Jeju Island, Korea: ACL; 2012. p. 115–120.
31. Diakopoulos NA, Shamma DA. Characterizing debate performance via aggregated twitter sentiment. In: *Proceedings of the SIGCHI conference on human factors in computing systems*; 2010. p. 1195–1198.
32. Daniel M, Neves RF, Horta N. Company event popularity for financial markets using Twitter and sentiment analysis. *Expert Systems with Applications*. 2017; 71:111–124. <https://doi.org/10.1016/j.eswa.2016.11.022>
33. Mukherjee S, Bhattacharyya P. Feature specific sentiment analysis for product reviews. In: *International Conference on Intelligent Text Processing and Computational Linguistics*. Springer; 2012. p. 475–487.
34. Bollen J, Pepe A, Mao H. Modeling public mood and emotion: Twitter sentiment and socio-economic phenomena. *arXiv preprint arXiv:09111583*. 2009;.

35. Gayo-Avello D. A meta-analysis of state-of-the-art electoral prediction from Twitter data. *Social Science Computer Review*. 2013; 31(6):649–679. <https://doi.org/10.1177/0894439313493979>
36. Gayo-Avello D, Metaxas P, Mustafaraj E. Limits of Electoral Predictions Using Twitter. In: -; 2011. p. 00.
37. Aparaschivei PA, et al. The use of new media in electoral campaigns: Analysis on the use of blogs, Facebook, Twitter and YouTube in the 2009 Romanian presidential campaign. *Journal of Media Research-Revista de Studii Media*. 2011; 4(10):39–60.
38. Effing R, van Hillegersberg J, Huibers T. Social media indicator and local elections in The Netherlands: Towards a framework for evaluating the influence of Twitter, YouTube, and Facebook. In: *Social media and local governments*. Springer; 2016. p. 281–298.
39. Vesnic-Alujevic L, Van Bauwel S. YouTube: A political advertising tool? A case study of the use of YouTube in the campaign for the European Parliament elections. *Journal of Political Marketing*. 2014; 13(3):195–212. <https://doi.org/10.1080/15377857.2014.929886>
40. Baviera T, Sampietro A, García-Ull FJ. Political conversations on Twitter in a disruptive scenario: The role of “party evangelists” during the 2015 Spanish general elections. *The Communication Review*. 2019; 22(2):117–138. <https://doi.org/10.1080/10714421.2019.1599642>
41. Panizo-LLedot A, Torregrosa J, Bello-Orgaz G, Thorburn J, Camacho D. Describing alt-right communities and their discourse on twitter during the 2018 us mid-term elections. In: *International conference on complex networks and their applications*. Springer; 2019. p. 427–439.
42. Hanusch F, Nölleke D. Journalistic Homophily on Social Media: Exploring journalists’ interactions with each other on Twitter. *Digital Journalism*. 2018 02;7:1–23.
43. Park SJ, Lim YS, Park HW. Comparing Twitter and YouTube networks in information diffusion: The case of the “Occupy Wall Street” movement. *Technological forecasting and social change*. 2015; 95:208–217. <https://doi.org/10.1016/j.techfore.2015.02.003>
44. Burgess J, Matamoros-Fernández A. Mapping sociocultural controversies across digital media platforms: One week of# gamergate on Twitter, YouTube, and Tumblr. *Communication Research and Practice*. 2016; 2(1):79–96. <https://doi.org/10.1080/22041451.2016.1155338>
45. Wu Z, Yin W, Cao J, Xu G, Cuzzocrea A. Community detection in multi-relational social networks. In: *International Conference on Web Information Systems Engineering*. Springer; 2013. p. 43–56.
46. Caldarelli G, Chessa A, Pammolli F, Pompa G, Puliga M, Riccaboni M, et al. A multi-level geographical study of Italian political elections from Twitter data. *PloS one*. 2014; 9(5):e95809. <https://doi.org/10.1371/journal.pone.0095809> PMID: 24802857
47. Eom YH, Puliga M, Smailović J, Mozetič I, Caldarelli G. Twitter-based analysis of the dynamics of collective attention to political parties. *PloS one*. 2015; 10(7):e0131184. <https://doi.org/10.1371/journal.pone.0131184> PMID: 26161795
48. Tumasjan A, Sprenger T, Sandner P, Welpel I. Predicting elections with twitter: What 140 characters reveal about political sentiment. In: *Proceedings of the International AAAI Conference on Web and Social Media*. vol. 4; 2010. p. 22.
49. Pennacchiotti M, Popescu AM. A machine learning approach to twitter user classification. In: *Proceedings of the International AAAI Conference on Web and Social Media*. vol. 5; 2011. p. 560–734.
50. Pennacchiotti M, Popescu AM. Democrats, republicans and starbucks aficionados: user classification in twitter. In: *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*; 2011. p. 430–438.
51. Barbera P, Casas A, Nagler J, Egan P, Bonneau R, Jost J, et al. Who Leads? Who Follows? Measuring Issue Attention and Agenda Setting by Legislators and the Mass Public Using Social Media Data. *American Political Science Review*. 2019 07;113. <https://doi.org/10.1017/S0003055419000352> PMID: 33303996
52. Zhang Y, Wells C, Wang S, Rohe K. Attention and amplification in the hybrid media system: The composition and activity of Donald Trump’s Twitter following during the 2016 presidential election. *New Media & Society*. 2018; 20(9):3161–3182. <https://doi.org/10.1177/1461444817744390>
53. Caetano JA, Lima HS, Santos MF, Marques-Neto HT. Using sentiment analysis to define twitter political users’ classes and their homophily during the 2016 American presidential election. *Journal of internet services and applications*. 2018; 9(1):1–15. <https://doi.org/10.1186/s13174-018-0089-0>
54. Halberstam Y, Knight B. Homophily, group size, and the diffusion of political information in social networks: Evidence from Twitter. *Journal of public economics*. 2016; 143:73–88. <https://doi.org/10.1016/j.jpubeco.2016.08.011>
55. Dokoohaki N, Zikou F, Gillblad D, Matskin M. Predicting swedish elections with twitter: A case for stochastic link structure analysis. In: *Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015*; 2015. p. 1269–1276.

56. Chamberlain JM, Spezzano F, Kettler JJ, Dit B. A Network Analysis of Twitter Interactions by Members of the US Congress. *ACM Transactions on Social Computing*. 2021; 4(1):1–22. <https://doi.org/10.1145/3439827>
57. Guerrero-Solé F. Community detection in political discussions on Twitter: An application of the retweet overlap network method to the Catalan process toward independence. *Social science computer review*. 2017; 35(2):244–261. <https://doi.org/10.1177/0894439315617254>
58. Baviera T. Influence in the political Twitter sphere: Authority and retransmission in the 2015 and 2016 Spanish General Elections. *European journal of communication*. 2018; 33(3):321–337. <https://doi.org/10.1177/0267323118763910>
59. Davidson I, Gourru A, Velcin J, Wu Y. Behavioral differences: insights, explanations and comparisons of French and US Twitter usage during elections. *Social Network Analysis and Mining*. 2020; 10(1):1–27. <https://doi.org/10.1007/s13278-019-0611-9>
60. Aragon P, Kappler KE, Kaltenbrunner A, Laniado D, Volkovich Y. Communication dynamics in twitter during political campaigns: The case of the 2011 Spanish national election. *Policy & internet*. 2013; 5(2):183–206. <https://doi.org/10.1002/1944-2866.POI327>
61. D'heer E, Verdegem P. Conversations about the elections on Twitter: Towards a structural understanding of Twitter's relation with the political and the media field. *European journal of communication*. 2014; 29(6):720–734. <https://doi.org/10.1177/0267323114544866>
62. Bekafigo MA, McBride A. Who tweets about politics? Political participation of Twitter users during the 2011 gubernatorial elections. *Social Science Computer Review*. 2013; 31(5):625–643. <https://doi.org/10.1177/0894439313490405>
63. Jungherr A, Jürgens P, Schoen H. Why the pirate party won the german election of 2009 or the trouble with predictions: A response to tumasjan, a., sprenger, to, sander, pg, & welpel, im “predicting elections with twitter: What 140 characters reveal about political sentiment”. *Social science computer review*. 2012; 30(2):229–234.
64. O'Connor B, Balasubramanyan R, Routledge BR, Smith NA. From tweets to polls: Linking text sentiment to public opinion time series. *Tepper School of Business*. 2010; 344:559.
65. Antonakaki D, Spiliotopoulos D, Samaras CV, Ioannidis S, Fragopoulou P. Investigating the complete corpus of referendum and elections tweets. In: 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM). IEEE. San Francisco, CA, USA: IEEE; 2016. p. 100–105.
66. Conover MD, Ratkiewicz J, Francisco MR, Gonçalves B, Menczer F, Flammini A. Political polarization on twitter. *Icwsn*. 2011; 133(26):89–96.
67. Weber I, Garimella VRK, Batayneh A. Secular vs. islamist polarization in egypt on twitter. In: Proceedings of the 2013 IEEE/ACM international conference on advances in social networks analysis and mining. Niagara Ontario Canada: IEEE; 2013. p. 290–297.
68. Morales AJ, Borondo J, Losada JC, Benito RM. Measuring political polarization: Twitter shows the two sides of Venezuela. *Chaos: An Interdisciplinary Journal of Nonlinear Science*. 2015; 25(3):033114. <https://doi.org/10.1063/1.4913758> PMID: 25833436
69. Christensen C. WAVE-RIDING AND HASHTAG-JUMPING. *Information, Communication & Society*. 2013; 16(5):646–666. <https://doi.org/10.1080/1369118X.2013.783609>
70. Bakshi RK, Kaur N, Kaur R, Kaur G. Opinion mining and sentiment analysis. In: 2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom). IEEE. New Delhi, Andaman and Nicobar Islands, India: IEEE; 2016. p. 452–455.
71. Ravi K, Ravi V. A survey on opinion mining and sentiment analysis: tasks, approaches and applications. *Knowledge-Based Systems*. 2015; 89:14–46. <https://doi.org/10.1016/j.knosys.2015.06.015>
72. Serrano-Guerrero J, Olivas JA, Romero FP, Herrera-Viedma E. Sentiment analysis: A review and comparative analysis of web services. *Information Sciences*. 2015; 311:18–38. <https://doi.org/10.1016/j.ins.2015.03.040>
73. Antonakaki D, Fragopoulou P, Ioannidis S. A survey of Twitter research: Data model, graph structure, sentiment analysis and attacks. *Expert Systems with Applications*. 2020; 164:114006. <https://doi.org/10.1016/j.eswa.2020.114006>
74. Wehrmann J, Becker W, Cagnini HE, Barros RC. A character-based convolutional neural network for language-agnostic Twitter sentiment analysis. In: 2017 International Joint Conference on Neural Networks (IJCNN). IEEE. Anchorage, Alaska: IEEE; 2017. p. 2384–2391.
75. Narr S, Hulfenhaus M, Albayrak S. Language-independent twitter sentiment analysis. *Knowledge discovery and machine learning (KDML), LWA*. 2012; 89898:12–14.
76. Davies A, Ghahramani Z. Language-independent Bayesian sentiment mining of Twitter. In: The 5th SNA-KDD Workshop'11 (SNA-KDD'11). University of California: SNA-KDD; 2011. p. 56–58.

77. Guthier B, Ho K, Saddik AE. Language-independent data set annotation for machine learning-based sentiment analysis. In: 2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC). Banff, AB, Canada: SMC; 2017. p. 2105–2110.
78. Saroufim C, Almatarky A, Hady MA. Language independent sentiment analysis with sentiment-specific word embeddings. In: Proceedings of the 9th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis. Brussels, Belgium: ACL; 2018. p. 14–23.
79. Ptáček T, Habernal I, Hong J. Sarcasm detection on czech and english twitter. In: Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers. Dublin, Ireland: ACL; 2014. p. 213–223.
80. Zhang S, Zhang X, Chan J. A Word-Character Convolutional Neural Network for Language-Agnostic Twitter Sentiment Analysis. In: Proceedings of the 22nd Australasian Document Computing Symposium. ADCS 2017. New York, NY, USA: Association for Computing Machinery; 2017. p. 00. Available from: <https://doi.org/10.1145/3166072.3166082>.
81. Severyn A, Moschitti A. Twitter Sentiment Analysis with Deep Convolutional Neural Networks. In: Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval. SIGIR'15. New York, NY, USA: Association for Computing Machinery; 2015. p. 959–962. Available from: <https://doi.org/10.1145/2766462.2767830>.
82. Jianqiang Z, Xiaolin G, Xuejun Z. Deep convolution neural networks for twitter sentiment analysis. IEEE Access. 2018; 6:23253–23260. <https://doi.org/10.1109/ACCESS.2017.2776930>
83. You Q, Luo J, Jin H, Yang J. Robust image sentiment analysis using progressively trained and domain transferred deep networks. arXiv preprint arXiv:150906041. 2015;3:270–279.
84. Dos Santos C, Gatti M. Deep convolutional neural networks for sentiment analysis of short texts. In: Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers. Dublin, Ireland: COLING; 2014. p. 69–78.
85. Alharbi ASM, de Doncker E. Twitter sentiment analysis with a deep neural network: An enhanced approach using user behavioral information. Cognitive Systems Research. 2019; 54:50–61. <https://doi.org/10.1016/j.cogsys.2018.10.001>
86. Severyn A, Moschitti A. Unltn: Training deep convolutional neural network for twitter sentiment classification. In: Proceedings of the 9th international workshop on semantic evaluation (SemEval 2015). Denver, Colorado: SIGLEX — SIGSEM; 2015. p. 464–469.
87. Liu KL, Li WJ, Guo M. Emoticon smoothed language models for twitter sentiment analysis. In: Aaai. vol. 12. Citeseer. Paris, France: Citeseer; 2012. p. 22–26.
88. Wang H, Castanon JA. Sentiment expression via emoticons on social media. In: 2015 IEEE international conference on big data (big data). IEEE. Santa Clara, CA, USA: IEEE; 2015. p. 2404–2408.
89. Zhao J, Dong L, Wu J, Xu K. MoodLens: An Emoticon-Based Sentiment Analysis System for Chinese Tweets. In: Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. KDD'12. New York, NY, USA: Association for Computing Machinery; 2012. p. 1528–1531.
90. Yamamoto Y, Kumamoto T, Nadamoto A. Role of Emoticons for Multidimensional Sentiment Analysis of Twitter. In: Proceedings of the 16th International Conference on Information Integration and Web-Based Applications and Services. iiWAS'14. New York, NY, USA: Association for Computing Machinery; 2014. p. 107–115.
91. Kolchyna O, Souza TT, Treleaven P, Aste T. Twitter sentiment analysis: Lexicon method, machine learning method and their combination. arXiv preprint arXiv:150700955. 2015;5656:33–38.
92. Pak A, Paroubek P. Twitter as a corpus for sentiment analysis and opinion mining. In: LREc. vol. 10. Valletta, Malta: LREC; 2010. p. 1320–1326.
93. Jianqiang Z, Xiaolin G. Comparison research on text pre-processing methods on twitter sentiment analysis. IEEE Access. 2017; 5:2870–2879. <https://doi.org/10.1109/ACCESS.2017.2672677>
94. Ghiassi M, Lee S. A domain transferable lexicon set for Twitter sentiment analysis using a supervised machine learning approach. Expert Systems with Applications. 2018; 106:197–216. <https://doi.org/10.1016/j.eswa.2018.04.006>
95. Kouloumpis E, Wilson T, Moore JD. Twitter sentiment analysis: The good the bad and the omg! Icwsm. 2011; 11(538-541):164.
96. Zhang L, Ghosh R, Dekhil M, Hsu M, Liu B. Combining lexicon-based and learning-based methods for Twitter sentiment analysis. HP Laboratories, Technical Report HPL-2011. 2011; 89.
97. Smailović J, Kranjc J, Grčar M, Žnidaršič M, Mozetič I. Monitoring the Twitter sentiment during the Bulgarian elections. In: 2015 IEEE International Conference on Data Science and Advanced Analytics (DSAA). Paris: IEEE; 2015. p. 1–10.

98. Singh S, Sikka G. YouTube Sentiment Analysis on US Elections 2020. In: 2021 2nd International Conference on Secure Cyber Computing and Communications (ICSCCC). IEEE; 2021. p. 250–254.
99. Bajaj P, Kavidayal M, Srivastava P, Akhtar MN, Kumaraguru P. Disinformation in multimedia annotation: Misleading metadata detection on YouTube. In: Proceedings of the 2016 ACM workshop on Vision and Language Integration Meets Multimedia Fusion; 2016. p. 53–61.
100. O'Callaghan D, Harrigan M, Carthy J, Cunningham P. Network analysis of recurring youtube spam campaigns. arXiv preprint arXiv:12013783. 2012;.
101. Sureka A, Kumaraguru P, Goyal A, Chhabra S. Mining youtube to discover extremist videos, users and hidden communities. In: Asia Information Retrieval Symposium. Springer; 2010. p. 13–24.
102. Baluja S, Seth R, Sivakumar D, Jing Y, Yagnik J, Kumar S, et al. Video suggestion and discovery for youtube: taking random walks through the view graph. In: Proceedings of the 17th international conference on World Wide Web; 2008. p. 895–904.
103. Jang HJ, Sim J, Lee Y, Kwon O. Deep sentiment analysis: Mining the causality between personality-value-attitude for analyzing business ads in social media. *Expert Systems with applications*. 2013; 40(18):7492–7503. <https://doi.org/10.1016/j.eswa.2013.06.069>
104. Klotz RJ. The sidetracked 2008 YouTube senate campaign. *Journal of Information Technology & Politics*. 2010; 7(2-3):110–123. <https://doi.org/10.1080/19331681003748917>
105. Ridout TN, Franklin Fowler E, Branstetter J. Political advertising in the 21st century: The rise of the YouTube ad. In: APSA 2010 Annual Meeting Paper; 2010..
106. Severyn A, Moschitti A, Uryupina O, Plank B, Filippova K. Multi-lingual opinion mining on YouTube. *Information Processing & Management*. 2016; 52(1):46–60. <https://doi.org/10.1016/j.ipm.2015.03.002>
107. Susarla A, Oh JH, Tan Y. Social networks and the diffusion of user-generated content: Evidence from YouTube. *Information Systems Research*. 2012; 23(1):23–41. <https://doi.org/10.1287/isre.1100.0339>
108. Wang X, Wei F, Liu X, Zhou M, Zhang M. Topic sentiment analysis in twitter: a graph-based hashtag sentiment classification approach. In: Proceedings of the 20th ACM international conference on Information and knowledge management; 2011. p. 1031–1040.
109. Thelwall M, Sud P, Vis F. Commenting on YouTube videos: From Guatemalan rock to el big bang. *Journal of the American Society for Information Science and Technology*. 2012; 63(3):616–629. <https://doi.org/10.1002/asi.21679>
110. Lindgren S. 'It took me about half an hour, but I did it!' Media circuits and affinity spaces around how-to videos on YouTube. *European Journal of Communication*. 2012; 27(2):152–170. <https://doi.org/10.1177/0267323112443461>
111. Krishna A, Zambreno J, Krishnan S. Polarity Trend Analysis of Public Sentiment on YouTube. In: Proceedings of the 19th International Conference on Management of Data. COMAD'13. Mumbai, Maharashtra, IND: Computer Society of India; 2013. p. 125–128.
112. Poria S, Cambria E, Howard N, Huang GB, Hussain A. Fusing audio, visual and textual clues for sentiment analysis from multimodal content. *Neurocomputing*. 2016; 174:50–59. <https://doi.org/10.1016/j.neucom.2015.01.095>
113. Amarasekara I, Grant WJ. Exploring the YouTube science communication gender gap: A sentiment analysis. *Public Understanding of Science*. 2019; 28(1):68–84. <https://doi.org/10.1177/0963662518786654> PMID: 29974815
114. Preoțiu-Pietro D, Volkova S, Lampos V, Bachrach Y, Aletras N. Studying user income through language, behaviour and affect in social media. *PloS one*. 2015; 10(9):e0138717. <https://doi.org/10.1371/journal.pone.0138717> PMID: 26394145
115. Araújo CS, Magno G, Meira W, Almeida V, Hartung P, Doneda D. Characterizing videos, audience and advertising in Youtube channels for kids. In: International Conference on Social Informatics. Springer; 2017. p. 341–359.
116. Wikipedia. 2020 United States presidential debates; 2020 (accessed September 30, 2020). Available from: https://en.wikipedia.org/wiki/2020_United_States_presidential_debates.
117. Pano T, Kashef R. A Complete VADER-Based Sentiment Analysis of Bitcoin (BTC) Tweets during the Era of COVID-19. *Big Data and Cognitive Computing*. 2020; 4(4). Available from: <https://www.mdpi.com/2504-2289/4/4/33>. <https://doi.org/10.3390/bdcc4040033>
118. Elbagir S, Yang J. Twitter sentiment analysis using natural language toolkit and VADER sentiment. In: Proceedings of the International MultiConference of Engineers and Computer Scientists. vol. 122; 2019. p. 16.
119. Zahoor S, Rohilla R. Twitter Sentiment Analysis Using Lexical or Rule Based Approach: A Case Study. In: 2020 8th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO); 2020. p. 537–542.

120. Ramteke J, Shah S, Godhia D, Shaikh A. Election result prediction using Twitter sentiment analysis. In: 2016 international conference on inventive computation technologies (ICICT). vol. 1. IEEE; 2016. p. 1–5.
121. Shelar A, Huang CY. Sentiment analysis of twitter data. In: 2018 International Conference on Computational Science and Computational Intelligence (CSCI). IEEE; 2018. p. 1301–1302.
122. Gilbert C, Hutto E. Vader: A parsimonious rule-based model for sentiment analysis of social media text. In: Eighth International Conference on Weblogs and Social Media (ICWSM-14). Available at (20/04/16) <http://comp.social.gatech.edu/papers/icwsm14.vader.hutto.pdf>. vol. 81; 2014. p. 82.
123. Bird S, Klein E, Loper E. Natural Language Processing with Python. O'Reilly Media; 2009.
124. Bastian M, Heymann S, Jacomy M. Gephi: An Open Source Software for Exploring and Manipulating Networks. In: Third international AAAI conference on weblogs and social media; 2009. p. 00. Available from: <http://www.aaai.org/ocs/index.php/ICWSM/09/paper/view/154>.
125. Aric Hagberg DS Pieter Swart. Python networkx library for graph creation/visualization; 2005. Available from: <https://networkx.github.io/>.
126. Wikipedia. List of U.S. state and territory abbreviations; 2020 (accessed September 29, 2020). Available from: https://en.wikipedia.org/wiki/List_of_U.S._state_and_territory_abbreviations.